

Э. Б. Винберг

Курс алгебры

МЦНМО

Э. Б. Винберг

КУРС АЛГЕБРЫ

Электронное издание

Москва
Издательство МЦНМО
2014

УДК 512
ББК 22.14
В48

Винберг Э. Б.
Курс алгебры
Электронное издание
М.: МЦНМО, 2014
590 с.
ISBN 978-5-4439-2013-9

Книга представляет собой расширенный вариант курса алгебры, читаемого в течение трех семестров на математических факультетах. В нее включены такие дополнительные разделы, как элементы коммутативной алгебры (в связи с аффинной алгебраической геометрией), теории Галуа, теории конечномерных ассоциативных алгебр и теории групп Ли. Это позволяет использовать книгу не только как учебник по общему курсу алгебры, но и как пособие для тех, кто желает углубить свои познания в алгебре. Изложение иллюстрируется большим количеством примеров и сопровождается задачами, часто содержащими дополнительный материал.

Книга предназначена для математиков и физиков — студентов, аспирантов, преподавателей и научных работников.

Подготовлено на основе книги: Э. Б. Винберг. Курс алгебры. — 2-е изд., стереотип. — М.: МЦНМО, 2013.

Издательство Московского центра
непрерывного математического образования
119002, Москва, Большой Власьевский пер., 11. Тел. (499) 241-74-83
<http://www.mccme.ru>

ISBN 978-5-4439-2013-9

© Винберг Э. Б., 2011.
© МЦНМО, 2014.

Оглавление

Предисловие	6
Глава 1. Алгебраические структуры	9
§ 1. Введение	9
§ 2. Абелевы группы	12
§ 3. Кольца и поля	17
§ 4. Поле комплексных чисел	22
§ 5. Кольца вычетов	28
§ 6. Векторные пространства	34
§ 7. Алгебры	38
§ 8. Алгебра матриц	41
Глава 2. Начала линейной алгебры	48
§ 1. Системы линейных уравнений	48
§ 2. Базис и размерность векторного пространства	58
§ 3. Ранг матрицы	68
§ 4. Определители	74
§ 5. Некоторые приложения определителей	88
Глава 3. Начала алгебры многочленов	92
§ 1. Построение и основные свойства алгебры многочленов	92
§ 2. Общие свойства корней многочленов	99
§ 3. Основная теорема алгебры комплексных чисел	106
§ 4. Корни многочленов с вещественными коэффициентами	110
§ 5. Теория делимости в евклидовых кольцах	117
§ 6. Многочлены с рациональными коэффициентами	123
§ 7. Многочлены от нескольких переменных	127
§ 8. Симметрические многочлены	132
§ 9. Кубические уравнения	140
§ 10. Поле рациональных дробей	147
Глава 4. Начала теории групп	154
§ 1. Определение и примеры	154
§ 2. Группы в геометрии и физике	162
§ 3. Циклические группы	166
§ 4. Системы порождающих	173

§ 5. Разбиение на смежные классы	175
§ 6. Гомоморфизмы	183
Глава 5. Векторные пространства	192
§ 1. Подпространства	192
§ 2. Линейные отображения	197
§ 3. Сопряженное пространство	205
§ 4. Билинейные и квадратичные функции	209
§ 5. Евклидово пространство	221
§ 6. Эрмитовы пространства	230
Глава 6. Линейные операторы	234
§ 1. Матрица линейного оператора	234
§ 2. Собственные векторы	240
§ 3. Линейные операторы и билинейные функции в евклидовом пространстве	246
§ 4. Жорданова форма	258
§ 5. Функции от линейного оператора	265
Глава 7. Аффинные и проективные пространства	277
§ 1. Аффинные пространства	277
§ 2. Аффинные отображения	283
§ 3. Выпуклые множества	290
§ 4. Евклидовы аффинные пространства	302
§ 5. Квадрики	309
§ 6. Проективные пространства	323
Глава 8. Тензорная алгебра	338
§ 1. Тензорное произведение векторных пространств	338
§ 2. Тензорная алгебра векторного пространства	346
§ 3. Симметрическая алгебра	353
§ 4. Алгебра Грассмана	360
Глава 9. Коммутативная алгебра	372
§ 1. Конечно порожденные абелевы группы	372
§ 2. Идеалы и факторкольца	386
§ 3. Модули над кольцами главных идеалов	395
§ 4. Нётеровы кольца	403
§ 5. Алгебраические расширения	407

§ 6. Конечно порожденные алгебры и аффинные алгебраические многообразия	420
§ 7. Разложение на простые множители	431
Глава 10. Группы	441
§ 1. Прямые и полупрямые произведения	441
§ 2. Коммутант	448
§ 3. Действия	451
§ 4. Теоремы Силова	458
§ 5. Простые группы	461
§ 6. Расширения Галуа	465
§ 7. Основная теорема теории Галуа	471
Глава 11. Линейные представления и ассоциативные алгебры	478
§ 1. Инвариантные подпространства	478
§ 2. Полная приводимость линейных представлений конечных и компактных групп	491
§ 3. Конечномерные ассоциативные алгебры	496
§ 4. Линейные представления конечных групп	504
§ 5. Инварианты	516
§ 6. Алгебры с делением	523
Глава 12. Группы Ли	537
§ 1. Определение и простейшие свойства групп Ли	537
§ 2. Экспоненциальное отображение	545
§ 3. Касательная алгебра Ли и присоединенное представление	549
§ 4. Линейные представления групп Ли	555
Ответы к задачам	563
Словарь сокращений английских слов, употребляемых в обозначениях	568
Список литературы	570
Указатель обозначений	572
Предметный указатель	575

Предисловие

Поводом для написания настоящего учебника послужил двухгодичный курс алгебры, прочитанный мною в Математическом колледже Независимого московского университета (НМУ) в 1992—1994 гг. Энтузиазм слушателей и относительно малое их число позволили мне читать курс на более высоком уровне, чем это принято на механико-математическом факультете МГУ (мехмате), и затронуть ряд тем, не входящих в курс алгебры мехмата. Однако при написании учебника я использовал свой опыт преподавания на мехмате, и его окончательный вариант имеет лишь отдаленное сходство с курсом, прочитанным в НМУ.

В издательстве «Факториал» книга переиздавалась три раза — в 1999, 2000 и 2002 годах. Новое издание вышло в издательстве МЦНМО в 2011 году.

По содержанию гл. 1—4 примерно соответствуют курсу алгебры первого семестра мехмата, а гл. 5—7 и отчасти гл. 8 — курсу линейной алгебры и геометрии второго семестра. Оставшиеся главы значительно перекрывают курс алгебры третьего семестра. Они адресованы в первую очередь тем студентам, которые хотят стать алгебраистами.

Глава 7 посвящена геометрии евклидовых, аффинных и проективных пространств. Однако ее ни в коей мере нельзя считать полноценным учебным пособием по геометрии; скорее это алгебраический взгляд на геометрию.

В первых четырех главах я постарался сделать изложение настолько подробным, насколько это может быть разумно, если иметь в виду такого читателя, как студент первого семестра мехмата. (Впрочем, язык множеств и отображений используется с самого начала без каких-либо объяснений.) Однако затем я начинаю позволять себе опускать некоторые легко восполнимые детали, считая, что читатель постепенно набирается математической культуры.

В книге почти нет технически сложных доказательств. В соответствии со своим взглядом на математику я стремился заменять выкладки и сложные рассуждения идеями. Кому-то это может показаться трудным, но усилия, потраченные на усвоение идей, окупят-

ся возможностью самостоятельно решать задачи, не рассматриваемые в учебнике.

Приведенный в конце книги список литературы на русском языке, которая, на мой взгляд, может быть полезной читателю, безусловно, далеко не полон и даже до некоторой степени случаен.

Основные изменения, сделанные при переизданиях, имели целью упростить изложение в техническом и идейном плане. В частности, с этой целью полностью переписана глава «Тензорная алгебра». Дано изложение теории абелевых групп, независимое от общей теории модулей над кольцами главных идеалов и подготавливающее читателя к восприятию этой общей теории, если он захочет это сделать.

С целью облегчить жизнь начинающему читателю аксиоматические определения поля комплексных чисел и определителей даны лишь после их конструктивных определений. Понятие линейного отображения и весь относящийся к нему материал перенесены из гл. 2 в гл. 5. Дано более простое доказательство существования жорданова базиса для нильпотентного линейного оператора.

В то же время при переизданиях было сделано несколько небольших добавлений. Так, дано доказательство неприводимости многочлена деления круга на любое число частей; описано приложение теории абелевых групп к исследованию симметрии кристаллов; добавлены некоторые сведения о (тензорных) произведениях и симметрических степенях линейных представлений групп с примером, иллюстрирующим применение этих понятий к физике. Добавлены задачи, содержащие существенную дополнительную информацию о линейных представлениях групп, увеличено число примеров групп Ли.

Я искренне благодарен всем бывшим и нынешним сотрудникам кафедры высшей алгебры мехмата, в общении с которыми сложились мои представления о преподавании алгебры.

Я благодарю редактора первого издания учебника Г. М. Цукерман, в результате тщательной работы которой было обнаружено большое количество неточностей и опечаток, а также сотрудников издательства МЦНМО Ю. Н. Торхова и В. В. Шувалова, чьи энтузиазм и самоотверженность немало способствовали улучшению качества учебника.

Я выражаю благодарность всем людям, указавшим мне на опечатки и неточности, в особенности И. В. Аржанцеву, А. П. Мишиной,

А. Д. Свердлову, а также профессору Скипу Гарибальди из университета Эмори (США).

Рисунок на переплете, выполненный на компьютере Ф. Э. Винбергом, иллюстрирует гомоморфизм $SU_2 \rightarrow SO_3$.

О нумерации. Теоремы нумеруются в пределах параграфа. При ссылке на теорему другого параграфа той же главы первая цифра означает номер параграфа, при ссылке на теорему другой главы первая цифра означает номер главы, вторая — номер параграфа. Так, теорема 2 — это теорема 2 того же параграфа, теорема 3.2 — это теорема 2 § 3 той же главы, а теорема 6.3.2 — это теорема 2 § 3 гл. 6. То же относится к параграфам, предложениям, примерам, задачам и замечаниям. Формулы и рисунки нумеруются в пределах главы.

Э. Б. Винберг

Алгебраические структуры

Когда вы знакомитесь с новыми людьми, вы прежде всего запоминаете их имена и внешность. После этого, встречаясь с ними в разных ситуациях, вы постепенно узнаете их лучше и некоторые из них, может быть, становятся вашими друзьями.

В первой главе состоится лишь внешнее знакомство читателя со многими из алгебраических структур, рассматриваемых в этой книге. Более глубокое их понимание будет приходить в процессе дальнейшего чтения книги и решения задач.

§ 1. Введение

Если вообще можно четко определить предмет алгебры, то это изучение алгебраических структур — множеств с определенными в них операциями. Под операцией в множестве M понимается любое отображение

$$M \times M \rightarrow M,$$

т. е. правило, по которому из любых двух элементов множества M получается некоторый элемент этого же множества. Элементами множества M могут быть как числа, так и объекты другого рода.

Хорошо известными и важными примерами алгебраических структур являются следующие числовые множества с операциями сложения и умножения:

\mathbb{N} — множество натуральных чисел,

\mathbb{Z} — множество всех целых чисел,

$\mathbb{Z}_+ = \mathbb{N} \cup \{0\}$ — множество неотрицательных целых чисел,

\mathbb{Q} — множество рациональных чисел,

\mathbb{R} — множество всех вещественных (= действительных) чисел,

\mathbb{R}_+ — множество неотрицательных вещественных чисел.

Подчеркнем, что операции сложения и умножения определены далеко не на всяком числовом множестве. Например, в множестве отрицательных чисел не определена операция умножения, так как

произведение двух отрицательных чисел является положительным числом. В множестве иррациональных чисел не определены ни сложение, ни умножение, так как сумма и произведение двух иррациональных чисел могут быть рациональными.

Приведем примеры алгебраических структур, состоящих не из чисел.

Пример 1. Пусть M, N, P — какие-то множества и

$$f: N \rightarrow M, \quad g: P \rightarrow N$$

— какие-то отображения. Произведением, или композицией, отображений f и g называется отображение

$$fg: P \rightarrow M,$$

определяемое формулой

$$(fg)(a) = f(g(a)) \quad \forall a \in P,$$

т. е. результат последовательного выполнения сначала отображения g , а потом f . (Обычно, если это не может привести к недоразумению, произведение отображений записывают без какого-либо специального знака, т. е. пишут просто fg : ср. обозначение $\ln \sin x$ в анализе.) В частности, при $M = N = P$ мы получаем таким образом операцию на множестве всех отображений множества M в себя. Эта операция дает много важных примеров алгебраических структур, называемых группами. Так, например, согласно аксиоматике евклидовой геометрии, произведение двух движений плоскости есть также движение. Рассматривая в множестве всех движений плоскости операцию умножения, мы получаем алгебраическую структуру, называемую группой движений плоскости.

Пример 2. Множество векторов пространства с операциями сложения и векторного умножения является примером алгебраической структуры с двумя операциями. Кстати, отметим, что скалярное умножение векторов не является операцией в определенном выше смысле, так как его результат не есть элемент того же множества. Подобные более общие операции также рассматриваются в алгебре, но мы пока не будем об этом думать.

Все приведенные выше примеры являются естественными в том смысле, что они были открыты в результате изучения реального мира и внутреннего развития математики. В принципе можно рассматривать любые операции в любых множествах. Например, можно

рассматривать операцию в множестве \mathbb{Z}_+ , ставящую в соответствие любым двум числам число совпадающих цифр в их десятичной записи. Однако лишь немногие алгебраические структуры представляют реальный интерес.

Следует уточнить, что алгебраиста интересуют только те свойства алгебраических структур и составляющих их элементов, которые могут быть выражены в терминах заданных операций. Этот подход находит свое выражение в понятии изоморфизма.

Определение 1. Пусть M — множество с операцией \circ , а N — множество с операцией $*$. Алгебраические структуры (M, \circ) и $(N, *)$ называются *изоморфными*, если существует такое биективное отображение

$$f: M \rightarrow N,$$

что

$$f(a \circ b) = f(a) * f(b)$$

для любых $a, b \in M$. В этом случае пишут $(M, \circ) \simeq (N, *)$. Само отображение f называется *изоморфизмом* структур (M, \circ) и $(N, *)$.

Аналогичным образом определяется изоморфизм алгебраических структур с двумя или бóльшим числом операций.

Пример 3. Отображение

$$a \mapsto 2^a$$

является изоморфизмом множества всех вещественных чисел с операцией сложения и множества положительных чисел с операцией умножения, поскольку

$$2^{a+b} = 2^a 2^b.$$

Вместо основания 2 можно было бы взять любое положительное основание, отличное от 1. Это показывает, что между изоморфными алгебраическими структурами может существовать много различных изоморфизмов.

Пример 4. Пусть V — множество векторов плоскости, а T — множество параллельных переносов. Для любого вектора a обозначим через t_a параллельный перенос на вектор a . (Если $a = 0$, то t_a — это тождественное преобразование.) Легко видеть, что

$$t_a \circ t_b = t_{a+b},$$

где \circ обозначает умножение (композицию) параллельных переносов, а $+$ обозначает сложение векторов (определяемое по прави-

лу параллелограмма). Следовательно, отображение $a \mapsto t_a$ является изоморфизмом алгебраических структур $(V, +)$ и (T, \circ) .

Ясно, что если две алгебраические структуры изоморфны, то любое утверждение, формулируемое только в терминах заданных операций, будет справедливым в одной из этих структур тогда и только тогда, когда оно справедливо в другой.

Например, операция \circ в множестве M называется *коммутативной*, если

$$a \circ b = b \circ a$$

для любых $a, b \in M$. Если структура (M, \circ) изоморфна структуре $(N, *)$ и операция \circ в множестве M коммутативна, то и операция $*$ в множестве N коммутативна.

Таким образом, в принципе все равно, какую из изоморфных друг другу алгебраических структур изучать: все они являются различными моделями одного и того же объекта. Однако выбор модели может оказаться небезразличным для фактического решения какой-либо задачи. Определенная модель может представить для этого наибольшее удобство. Например, если какая-то модель имеет геометрический характер, то она позволяет применить геометрические методы.

§ 2. Абелевы группы

Сложение вещественных чисел обладает следующими свойствами:

$$(C1) \quad a + b = b + a \text{ (коммутативность);}$$

$$(C2) \quad (a + b) + c = a + (b + c) \text{ (ассоциативность);}$$

$$(C3) \quad a + 0 = a;$$

$$(C4) \quad a + (-a) = 0.$$

Из этих свойств чисто логическим путем могут быть получены и другие свойства, например, наличие операции вычитания, обратной к сложению. Это означает, что для любых a, b уравнение

$$x + a = b$$

имеет единственное решение. Докажем, что это так. Если c — решение данного уравнения, т. е. $c + a = b$, то

$$(c + a) + (-a) = b + (-a).$$

Пользуясь свойствами (C2)—(C4), получаем

$$(c + a) + (-a) = c + (a + (-a)) = c + 0 = c.$$

Таким образом,

$$c = b + (-a).$$

Это показывает, что если решение существует, то оно единственно и равно $b + (-a)$. С другой стороны, подстановка $x = b + (-a)$ в рассматриваемое уравнение показывает, что $b + (-a)$ действительно является решением:

$$(b + (-a)) + a = b + ((-a) + a) = b + (a + (-a)) = b + 0 = b.$$

Умножение вещественных чисел обладает аналогичными свойствами:

(У1) $ab = ba$ (коммутативность);

(У2) $(ab)c = a(bc)$ (ассоциативность);

(У3) $a1 = a$;

(У4) $aa^{-1} = 1$ при $a \neq 0$.

Свойства (У1)—(У4) лишь формой записи отличаются от свойств (C1)—(C4), с единственной оговоркой, что в (У4) мы предполагаем, что $a \neq 0$, в то время как в (C4) никаких ограничений на a нет. Поэтому приведенный выше вывод из свойств (C1)—(C4) наличия операции вычитания, будучи переведен на язык умножения, даст вывод из свойств (У1)—(У4) наличия операции деления, обратной к умножению. Более точно, таким путем доказывается, что для любого $a \neq 0$ и любого b уравнение $xa = b$ имеет единственное решение, равное ba^{-1} .

Все эти рассуждения приведены здесь не для того, чтобы читатель узнал что-либо новое о вещественных числах, а чтобы подвести его к важной для алгебры идее. Эта идея есть аксиоматический метод в алгебре. Он состоит в одновременном изучении целых классов алгебраических структур, выделяемых теми или иными аксиомами, представляющими собой какие-то свойства операций в этих структурах. При этом совершенно не важно, как в каждом конкретном случае эти операции определяются. Коль скоро выполнены аксиомы, справедлива и любая теорема, полученная логическим путем из этих аксиом.

Конечно, лишь немногие системы аксиом действительно интересны. Невозможно придумать «из головы» такую систему аксиом,

которая привела бы к содержательной теории. Все системы аксиом, рассматриваемые в современной алгебре, имеют длительную историю и являются результатом анализа алгебраических структур, возникших естественным путем. Таковы системы аксиом группы, кольца, поля, векторного пространства и другие, с которыми читатель познакомится в этом курсе.

Свойства (C1)—(C4), а также (У1)—(У4) являются по сути дела системой аксиом абелевой группы. Перед тем как привести точные формулировки этих аксиом, скажем несколько слов о терминологии. Названия и обозначения операций в алгебраических структурах не имеют принципиального значения, однако чаще всего они называются сложением или умножением и обозначаются соответствующим образом. Это позволяет использовать разработанную терминологию и систему обозначений, относящиеся к операциям над вещественными числами, а также вызывает полезные ассоциации.

Приведем вначале определение абелевой группы, использующее язык сложения.

Определение 1. (Аддитивной) абелевой группой называют множество A с операцией сложения, обладающей следующими свойствами:

- 1) $a + b = b + a$ для любых $a, b \in A$ (коммутативность);
- 2) $(a + b) + c = a + (b + c)$ для любых $a, b, c \in A$ (ассоциативность);
- 3) в A существует такой элемент 0 (нуль), что $a + 0 = a$ для любого $a \in A$;
- 4) для любого элемента $a \in A$ существует такой элемент $-a \in A$ (противоположный элемент), что $a + (-a) = 0$.

Выведем некоторые простейшие следствия из этих аксиом.

- 1) Нуль единствен. В самом деле, пусть 0_1 и 0_2 — два нуля. Тогда

$$0_1 = 0_1 + 0_2 = 0_2.$$

- 2) Противоположный элемент единствен. В самом деле, пусть $(-a)_1$ и $(-a)_2$ — два элемента, противоположных a . Тогда

$$(-a)_1 = (-a)_1 + (a + (-a)_2) = ((-a)_1 + a) + (-a)_2 = (-a)_2.$$

- 3) Для любых a, b уравнение $x + a = b$ имеет единственное решение, равное $b + (-a)$. Доказательство см. выше. Это решение называется разностью элементов b и a и обозначается $b - a$.

Из свойства ассоциативности нетрудно вывести (попробуйте сделать это), что сумма произвольного числа (а не только трех) элементов не зависит от расстановки скобок. Пользуясь этим, скобки обычно вообще опускают.

Пример 1. Числовые множества \mathbb{Z} , \mathbb{Q} , \mathbb{R} являются абелевыми группами относительно обычной операции сложения.

Пример 2. Множество векторов (плоскости или пространства) является абелевой группой относительно обычного сложения векторов.

Пример 3. Последовательность из n чисел назовем *строкой* длины n . Множество всех строк длины n , составленных из вещественных чисел, обозначим через \mathbb{R}^n . Определим сложение строк по правилу

$$(a_1, a_2, \dots, a_n) + (b_1, b_2, \dots, b_n) = (a_1 + b_1, a_2 + b_2, \dots, a_n + b_n).$$

Очевидно, что множество \mathbb{R}^n является абелевой группой относительно этой операции. Ее нулем служит *нулевая строка*

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

Пример 4. Множество всех функций, определенных на заданном подмножестве числовой прямой, является абелевой группой относительно обычного сложения функций.

Приведем теперь определение абелевой группы, использующее язык умножения.

Определение 1'. (*Мультипликативной*) абелевой группой называют множество A с операцией умножения, обладающей следующими свойствами:

- 1) $ab = ba$ для любых $a, b \in A$ (*коммутативность*);
- 2) $(ab)c = a(bc)$ для любых $a, b, c \in A$ (*ассоциативность*);
- 3) в A существует такой элемент e (*единица*), что $ae = a$ для любого $a \in A$;
- 4) для любого элемента $a \in A$ существует такой элемент $a^{-1} \in A$ (*обратный элемент*), что $aa^{-1} = e$.

Единица мультипликативной абелевой группы иногда обозначается символом 1.

Простейшие следствия аксиом абелевой группы, полученные выше на аддитивном языке, на мультипликативном языке выглядят следующим образом.

- 1) Единица единственна.

2) Обратный элемент единствен.

3) Для любых a, b уравнение $xa = b$ имеет единственное решение, равное ba^{-1} . Оно называется *частным* от деления b на a (или *отношением* элементов b и a) и обозначается $\frac{b}{a}$ (или b/a).

Пример 5. Числовые множества $\mathbb{Q}^* = \mathbb{Q} \setminus \{0\}$ и $\mathbb{R}^* = \mathbb{R} \setminus \{0\}$ являются абелевыми группами относительно обычной операции умножения.

В дальнейшем мы познакомимся с общим понятием группы (не обязательно абелевой), которое не включает требования коммутативности операции.

Читатель, наверное, заметил, что некоторые из рассмотренных выше абелевых групп содержатся в других, причем операция в «маленькой» группе определяется так же, как в «большой». Это приводит нас к понятию подгруппы.

Вообще, пусть M — множество с операцией \circ и N — какое-либо его подмножество. Говорят, что N *замкнуто относительно операции* \circ , если

$$a, b \in N \Rightarrow a \circ b \in N.$$

В этом случае операция \circ определена в множестве N и превращает его в некоторую алгебраическую структуру. Если операция \circ в M обладает каким-то свойством, имеющим характер тождественного соотношения (например, свойством коммутативности или ассоциативности), то она, очевидно, обладает этим свойством и в N . Однако другие свойства операции \circ могут не наследоваться подмножеством N .

Так, подмножество аддитивной абелевой группы, замкнутое относительно сложения, не обязано быть абелевой группой, так как оно может не содержать нуля или элемента, противоположного какому-либо его элементу. Например, подмножество \mathbb{Z}_+ замкнуто относительно сложения в абелевой группе \mathbb{Z} , но не является абелевой группой (и вообще группой), так как не содержит противоположного элемента ни к одному своему элементу, кроме нуля.

Определение 2. Подмножество B аддитивной абелевой группы A называется *подгруппой*, если

- 1) B замкнуто относительно сложения;
- 2) $a \in B \Rightarrow -a \in B$;
- 3) $0 \in B$.

Замечание 1. Легко видеть, что если B непусто, то из первых двух условий вытекает третье. Поэтому третье условие может быть заменено условием непустоты.

Очевидно, что всякая подгруппа аддитивной абелевой группы сама является абелевой группой относительно той же операции.

Пример 6. В аддитивной группе \mathbb{R} имеется следующая цепочка подгрупп:

$$\mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}.$$

Пример 7. В аддитивной группе векторов пространства множество векторов, параллельных заданной плоскости или прямой, является подгруппой.

В любой аддитивной абелевой группе имеются две «тривиальные» подгруппы: вся группа и подгруппа, состоящая только из нуля.

Задача 1. Доказать, что всякая подгруппа группы \mathbb{Z} имеет вид $n\mathbb{Z}$, где $n \in \mathbb{Z}_+$ (решение этой задачи можно найти в § 4.3).

Приведем мультипликативный вариант предыдущего определения.

Определение 2. Подмножество B мультипликативной абелевой группы A называется *подгруппой*, если

- 1) B замкнуто относительно умножения;
- 2) $a \in B \Rightarrow a^{-1} \in B$;
- 3) $e \in B$.

Пример 8. В группе \mathbb{R}^* имеется следующая цепочка подгрупп:

$$\{\pm 1\} \subset \mathbb{Q}^* \subset \mathbb{R}^*.$$

§ 3. Кольца и поля

В отличие от групп кольца и поля — это алгебраические структуры с двумя операциями, называемыми обычно сложением и умножением. Их аксиомы, как и аксиомы абелевой группы, подсказаны свойствами операций над вещественными числами. При этом аксиомы кольца — это разумный минимум требований относительно свойств операций, позволяющий охватить и другие важные примеры алгебраических структур, из которых мы пока можем привести только уже упоминавшееся множество векторов пространства с операциями сложения и векторного умножения.

Определение 1. *Кольцом называется множество K с операциями сложения и умножения, обладающими следующими свойствами:*

1) относительно сложения K есть абелева группа (называемая *аддитивной группой кольца K*);

2) $a(b + c) = ab + ac$ и $(a + b)c = ac + bc$ для любых $a, b, c \in K$ (*дистрибутивность умножения относительно сложения*).

Выведем некоторые следствия аксиом кольца, не входящие в число следствий аксиом аддитивной абелевой группы, перечисленных в § 2.

1) $a0 = 0a = 0$ для любого $a \in K$. В самом деле, пусть $a0 = b$. Тогда

$$b + b = a0 + a0 = a(0 + 0) = a0 = b,$$

откуда

$$b = b - b = 0.$$

Аналогично доказывается, что $0a = 0$.

2) $a(-b) = (-a)b = -ab$ для любых $a, b \in K$. В самом деле,

$$ab + a(-b) = a(b + (-b)) = a0 = 0$$

и, аналогично, $ab + (-a)b = 0$.

3) $a(b - c) = ab - ac$ и $(a - b)c = ac - bc$ для любых $a, b, c \in K$. В самом деле,

$$a(b - c) + ac = a(b - c + c) = ab$$

и, аналогично, $(a - b)c + bc = ac$.

Кольцо K называется *коммутативным*, если умножение в нем коммутативно, т. е.

$$ab = ba \quad \forall a, b,$$

и *ассоциативным*, если умножение в нем ассоциативно, т. е.

$$(ab)c = a(bc) \quad \forall a, b, c.$$

Элемент 1 кольца называется *единицей*, если

$$a1 = 1a = a \quad \forall a.$$

Так же, как в случае мультипликативной абелевой группы, доказыва-
ется, что в кольце не может быть двух различных единиц (но может
не быть ни одной).

Замечание 1. Если $1 = 0$, то для любого a имеем

$$a = a1 = a0 = 0,$$

т. е. кольцо состоит из одного нуля. Таким образом, если кольцо
содержит более одного элемента, то $1 \neq 0$.

Замечание 2. При наличии коммутативности из двух тождеств
дистрибутивности, входящих в определение кольца, можно оста-
вить лишь одно. Аналогичное замечание относится к определению
единицы.

Пример 1. Числовые множества \mathbb{Z} , \mathbb{Q} , \mathbb{R} являются коммутатив-
ными ассоциативными кольцами с единицей относительно обык-
новенных операций сложения и умножения.

Пример 2. Множество $2\mathbb{Z}$ четных чисел является коммутатив-
ным ассоциативным кольцом без единицы.

Пример 3. Множество всех функций, определенных на задан-
ном подмножестве числовой прямой, является коммутативным ас-
социативным кольцом с единицей относительно обычных операций
сложения и умножения функций.

Пример 4. Множество векторов пространства с операциями сло-
жения и векторного умножения является некоммутативным и неас-
социативным кольцом. Однако в нем выполняются следующие тож-
дества, которые в некотором смысле заменяют коммутативность
и ассоциативность:

$$a \times b + b \times a = 0 \quad (\text{антикоммутативность}),$$

$$(a \times b) \times c + (b \times c) \times a + (c \times a) \times b = 0 \quad (\text{тождество Якоби}).$$

Антикоммутативность очевидна в силу определения векторного ум-
ножения. По поводу проверки тождества Якоби см. пример 7.5.

Задача 1. Пусть X — какое-либо множество и 2^X — множество
всех его подмножеств. Доказать, что 2^X — кольцо относительно опе-
раций симметрической разности

$$M \Delta N = (M \setminus N) \cup (N \setminus M)$$

и пересечения, взятых в качестве сложения и умножения соответ-
ственно. Доказать, что это кольцо коммутативно, ассоциативно и об-
ладает единицей.

Элемент a^{-1} кольца с единицей называется *обратным* к элементу a , если

$$aa^{-1} = a^{-1}a = 1.$$

(В коммутативном кольце достаточно требовать, чтобы $aa^{-1} = 1$.) Так же, как в случае мультипликативной абелевой группы, доказывается, что в ассоциативном кольце с единицей никакой элемент не может иметь двух различных обратных элементов (но может не иметь ни одного). Элемент, имеющий обратный, называется *обратимым*.

Определение 2. *Поле* называется коммутативное ассоциативное кольцо с единицей, в котором всякий ненулевой элемент обратим.

Замечание 3. Кольцо, состоящее из одного нуля, не считается полем.

Примерами полей служат поле рациональных чисел \mathbb{Q} и поле вещественных чисел \mathbb{R} . Кольцо \mathbb{Z} не является полем: в нем обратимы только ± 1 .

Задача 2. Доказать, что существует поле, состоящее из двух элементов. (Очевидно, что один из этих элементов должен быть нулем поля, а другой — его единицей.)

Любое поле обладает следующим важным свойством:

$$ab = 0 \Rightarrow a = 0 \text{ или } b = 0.$$

В самом деле, если $a \neq 0$, то, умножая обе части равенства $ab = 0$ на a^{-1} , получаем $b = 0$.

Существуют и другие кольца, обладающие этим свойством, например, кольцо \mathbb{Z} . Они называются *кольцами без делителей нуля*. В кольце без делителей нуля возможно сокращение:

$$\{ac = bc \text{ (или } ca = cb) \text{ и } c \neq 0\} \Rightarrow a = b.$$

В самом деле, равенство $ac = bc$ может быть переписано в виде $(a - b)c = 0$, откуда при $c \neq 0$ получаем $a - b = 0$, т. е. $a = b$.

Приведем пример коммутативного ассоциативного кольца с делителями нуля.

Пример 5. В кольце функций на подмножестве X числовой прямой (см. пример 3) есть делители нуля, если только X содержит более одной точки. В самом деле, разобьем X на два непустых под-

множества X_1 и X_2 и положим при $i = 1, 2$

$$f_i(x) = \begin{cases} 1 & \text{при } x \in X_i, \\ 0 & \text{при } x \notin X_i. \end{cases}$$

Тогда $f_1, f_2 \neq 0$, но $f_1 f_2 = 0$.

Отсутствие делителей нуля в поле означает, что произведение любых двух ненулевых элементов также является ненулевым элементом. Ненулевые элементы поля K образуют абелеву группу относительно умножения. Она называется *мультипликативной группой поля K* и обозначается через K^* .

Аналогично понятию подгруппы абелевой группы вводится понятие подкольца.

Определение 3. Подмножество L кольца K называется *подкольцом*, если

- 1) L является подгруппой аддитивной группы кольца K ;
- 2) L замкнуто относительно умножения.

Очевидно, что всякое подкольцо само является кольцом относительно тех же операций. При этом оно наследует такие свойства, как коммутативность и ассоциативность.

Пример 6. Цепочка подгрупп аддитивной группы \mathbb{R} , приведенная в примере 1, является в то же время цепочкой подколец.

Пример 7. При любом $n \in \mathbb{Z}_+$ множество $n\mathbb{Z}$ является подкольцом кольца \mathbb{Z} . (Ср. задачу 2.1.)

Задача 3. Доказать, что все конечные подмножества множества X образуют подкольцо кольца 2^X из задачи 1.

Определение 4. Подмножество L поля K называется *подполем*, если

- 1) L является подкольцом кольца K ;
- 2) $a \in L, a \neq 0 \Rightarrow a^{-1} \in L$;
- 3) $1 \in L$.

Очевидно, что всякое подполе является полем относительно тех же операций.

Пример 8. Поле \mathbb{Q} является подполем поля \mathbb{R} .

Задача 4. Доказать, что подмножество L поля K является подполем тогда и только тогда, когда

- 1) L замкнуто относительно вычитания и деления;
- 2) $L \ni 0, 1$.

Задача 5. Доказать, что поле \mathbb{Q} не имеет нетривиальных (т.е. отличных от него самого) подполей.

§4. Поле комплексных чисел

Подобно тому как невозможность деления в кольце целых чисел приводит к необходимости расширить его до поля рациональных чисел, невозможность извлечения квадратных корней из отрицательных чисел в поле вещественных чисел приводит к необходимости расширить его до большего поля, называемого полем комплексных чисел.

Для того чтобы прийти к определению комплексных чисел естественным путем, проведем вначале некоторый анализ. А именно, предположим, что уже имеется некоторое поле \mathbb{C} , содержащее поле \mathbb{R} вещественных чисел и некий элемент i , квадрат которого равен -1 , и посмотрим, как оно должно быть устроено.

Наряду с элементом i поле \mathbb{C} должно содержать элементы $a + bi$, где a и b — любые вещественные числа. Докажем, что все эти элементы различны. Пусть $a_1 + b_1i = a_2 + b_2i$, $a_1, b_1, a_2, b_2 \in \mathbb{R}$. Тогда

$$a_1 - a_2 = (b_2 - b_1)i.$$

Возводя это равенство в квадрат, получаем

$$(a_1 - a_2)^2 = -(b_2 - b_1)^2,$$

откуда

$$a_1 - a_2 = b_2 - b_1 = 0,$$

т. е. $a_1 = a_2, b_1 = b_2$, что и требовалось доказать.

Далее, из свойств операций в поле и соотношения $i^2 = -1$ следует, что

$$(a_1 + b_1i) + (a_2 + b_2i) = (a_1 + a_2) + (b_1 + b_2)i, \quad (1)$$

$$(a_1 + b_1i)(a_2 + b_2i) = (a_1a_2 - b_1b_2) + (a_1b_2 + b_1a_2)i. \quad (2)$$

Это показывает, что подмножество $K = \{a + bi : a, b \in \mathbb{R}\} \subset \mathbb{C}$ замкнуто относительно сложения и умножения. Из формулы (1) следует, что

$$-(a + bi) = (-a) + (-b)i \in K, \quad (3)$$

а из формулы (2) — что

$$(a + bi)(a - bi) = a^2 + b^2 \in \mathbb{R}$$

и, значит,

$$(a + bi)^{-1} = \frac{a}{a^2 + b^2} + \left(-\frac{b}{a^2 + b^2}\right)i \in K \quad \text{при } a^2 + b^2 \neq 0. \quad (4)$$

Следовательно, K — подполе поля \mathbb{C} . Так как поле K уже содержит поле вещественных чисел и квадратный корень из -1 (а значит, и квадратный корень из любого отрицательного числа), то нам нет необходимости рассматривать какое-то большее поле, т. е. можно считать, что $\mathbb{C} = K$.

Предыдущее исследование подсказывает, как можно построить поле комплексных чисел. Рассмотрим множество \mathbb{C} пар (a, b) , где $a, b \in \mathbb{R}$. Определим в нем сложение и умножение по формулам

$$\begin{aligned} (a_1, b_1) + (a_2, b_2) &= (a_1 + a_2, b_1 + b_2), \\ (a_1, b_1)(a_2, b_2) &= (a_1a_2 - b_1b_2, a_1b_2 + b_1a_2), \end{aligned}$$

подсказанным формулами (1) и (2). Очевидно, что \mathbb{C} является абелевой группой относительно сложения (ср. пример 2.3) и что умножение дистрибутивно относительно сложения и коммутативно. Непосредственной выкладкой проверяется ассоциативность умножения:

$$\begin{aligned} ((a_1, b_1)(a_2, b_2))(a_3, b_3) &= (a_1a_2 - b_1b_2, a_1b_2 + b_1a_2)(a_3, b_3) = \\ &= (a_1a_2a_3 - a_1b_2b_3 - b_1a_2b_3 - b_1b_2a_3, b_1a_2a_3 + a_1b_2a_3 + a_1a_2b_3 - b_1b_2b_3) = \\ &= (a_1, b_1)(a_2a_3 - b_2b_3, a_2b_3 + b_2a_3) = (a_1, b_1)((a_2, b_2)(a_3, b_3)). \end{aligned}$$

(О том, как можно избежать этих вычислений, см. пример 7.4.) Таким образом, \mathbb{C} — коммутативное ассоциативное кольцо.

Так как

$$(a, b)(1, 0) = (a, b),$$

то элемент $(1, 0)$ — единица кольца \mathbb{C} . Формула (4) подсказывает, как должен выглядеть элемент, обратный к (a, b) при $a^2 + b^2 \neq 0$. И действительно, непосредственная проверка показывает, что

$$(a, b) \left(\frac{a}{a^2 + b^2}, -\frac{b}{a^2 + b^2} \right) = (1, 0).$$

Следовательно, \mathbb{C} — поле.

Далее,

$$\begin{aligned} (a_1, 0) + (a_2, 0) &= (a_1 + a_2, 0), \\ (a_1, 0)(a_2, 0) &= (a_1a_2, 0), \end{aligned}$$

т. е. операции над парами вида $(a, 0)$ сводятся к соответствующим операциям над их первыми компонентами. Условимся отождествлять пару $(a, 0)$ с вещественным числом a . Тогда мы можем сказать, что построенное поле \mathbb{C} содержит поле \mathbb{R} в качестве подполя.

Положим $i = (0, 1)$; тогда

$$i^2 = (-1, 0) = -1,$$

$$a + bi = (a, b) \quad \text{при } a, b \in \mathbb{R}.$$

Таким образом, каждый элемент поля \mathbb{C} однозначно представляется в виде $a + bi$, где $a, b \in \mathbb{R}$.

Построенное поле \mathbb{C} и называется *полем комплексных чисел*.

Представление комплексного числа $c \in \mathbb{C}$ в виде $a + bi$ ($a, b \in \mathbb{R}$) называется его *алгебраической формой*; при этом число a называется *вещественной частью* числа c и обозначается $\operatorname{Re} c$, а число b называется *мнимой частью* числа c и обозначается $\operatorname{Im} c$. Комплексные числа, не являющиеся вещественными, называются *мнимыми*; числа вида bi , где $b \in \mathbb{R}$, называются *чисто мнимыми*.

Так как при выводе формул (1) и (2) мы использовали только то свойство элемента i , что $i^2 = -1$, а элемент $i' = -i$ также обладает этим свойством, то эти формулы остаются верными при замене i на i' . Это означает, что отображение

$$c = a + bi \mapsto \bar{c} = a - bi \quad (a, b \in \mathbb{R}),$$

является изоморфизмом поля \mathbb{C} на себя. Оно называется *комплексным сопряжением*. Вообще, изоморфизм какой-либо алгебраической структуры на себя называется ее *автоморфизмом*. Таким образом, комплексное сопряжение $c \mapsto \bar{c}$ есть автоморфизм поля комплексных чисел. Очевидно, что $\bar{\bar{c}} = c$.

Вещественные числа характеризуются тем, что они совпадают со своими сопряженными. Отсюда следует, что для любого $c \in \mathbb{C}$ числа $c + \bar{c}$ и $c\bar{c}$ вещественны. Более точно, если $c = a + bi$ ($a, b \in \mathbb{R}$), то

$$c + \bar{c} = 2a, \quad c\bar{c} = a^2 + b^2. \quad (5)$$

Комплексные числа можно изображать точками или векторами на плоскости. А именно, число $c = a + bi$ изображается точкой или вектором с декартовыми координатами (a, b) (рис. 1). Иногда удобнее представление комплексных чисел точками, иногда —

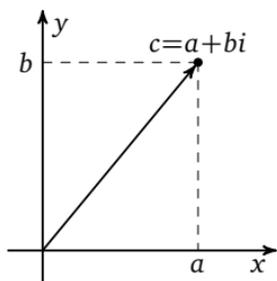


Рис. 1

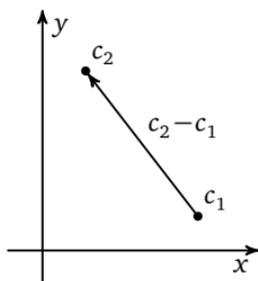


Рис. 2

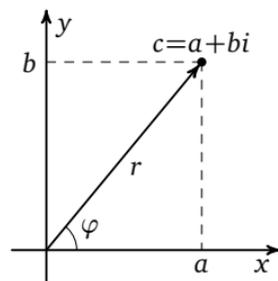


Рис. 3

векторами. При векторном представлении сложению комплексных чисел соответствует обычное сложение векторов по правилу параллелограмма (или эквивалентному ему правилу треугольника).

Отметим, что разность комплексных чисел c_2 и c_1 представляется вектором, соединяющим точки, изображающие c_1 и c_2 (рис. 2).

Вместо декартовых координат на плоскости иногда бывает удобно использовать полярные. Это приводит к следующим понятиям.

Модулем комплексного числа $c = a + bi$ называется длина вектора, изображающего это число. Модуль числа c обозначается через $|c|$. Очевидно, что

$$|c| = \sqrt{a^2 + b^2}.$$

Аргументом комплексного числа называется угол, образуемый соответствующим вектором с положительным направлением оси абсцисс. Аргумент определен с точностью до прибавления целого кратного 2π . Аргумент числа 0 не определен. Аргумент числа c обозначается через $\arg c$.

Пусть r и φ — модуль и аргумент числа c (рис. 3). Очевидно, что

$$a = r \cos \varphi, \quad b = r \sin \varphi,$$

откуда

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

Такое представление комплексного числа называется его *тригонометрической формой*. Так как тригонометрическая форма данного комплексного числа определена однозначно с точностью до прибав-

ления к φ целого кратного 2π , то при $r_1, r_2 > 0$

$$r_1(\cos \varphi_1 + i \sin \varphi_1) = r_2(\cos \varphi_2 + i \sin \varphi_2) \Leftrightarrow \\ \Leftrightarrow \{r_1 = r_2, \varphi_1 = \varphi_2 + 2\pi k, k \in \mathbb{Z}\}.$$

Тригонометрическая форма комплексных чисел хорошо приспособлена к таким операциям, как умножение, деление, возведение в степень и извлечение корня. А именно, из формул для косинуса и синуса суммы двух углов следует, что

$$r_1(\cos \varphi_1 + i \sin \varphi_1) \cdot r_2(\cos \varphi_2 + i \sin \varphi_2) = \\ = r_1 r_2 (\cos(\varphi_1 + \varphi_2) + i \sin(\varphi_1 + \varphi_2)),$$

т. е. при умножении комплексных чисел их модули перемножаются, а аргументы складываются. Отсюда вытекают следующие формулы для деления и возведения в степень:

$$\frac{r_1(\cos \varphi_1 + i \sin \varphi_1)}{r_2(\cos \varphi_2 + i \sin \varphi_2)} = \frac{r_1}{r_2} (\cos(\varphi_1 - \varphi_2) + i \sin(\varphi_1 - \varphi_2)),$$

$$[r(\cos \varphi + i \sin \varphi)]^n = r^n (\cos n\varphi + i \sin n\varphi) \quad (\text{формула Муавра}).$$

Извлечение корня n -й степени из комплексного числа $c = r(\cos \varphi + i \sin \varphi)$ есть решение уравнения $z^n = c$. Пусть $|z| = s$, $\arg z = \psi$; тогда $s^n = r$, $n\psi = \varphi + 2\pi k$ ($k \in \mathbb{Z}$). Следовательно,

$$s = \sqrt[n]{r} \quad (\text{арифметическое значение корня}), \quad \psi = \frac{\varphi + 2\pi k}{n}.$$

Окончательно получаем

$$z = \sqrt[n]{r} \left(\cos \frac{\varphi + 2\pi k}{n} + i \sin \frac{\varphi + 2\pi k}{n} \right).$$

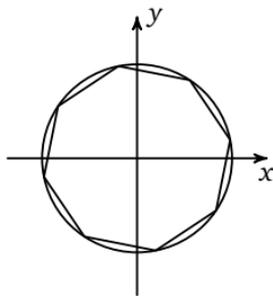


Рис. 4

Одинаковые значения z получаются по этой формуле тогда и только тогда, когда в качестве k берутся числа, сравнимые по модулю n . Отсюда следует, что при $c \neq 0$ уравнение $z^n = c$ имеет ровно n решений, получаемых, например, при $k = 0, 1, \dots, n-1$. В геометрическом изображении эти числа располагаются в вершинах правильного n -угольника с центром в начале координат (см. рис. 4, где изображен случай $n = 8$).

Вместо того чтобы называть комплексным числом пару вещественных чисел, как мы делали выше, можно было бы назвать комплексным числом точку (или вектор) плоскости и соответствующим образом определить операции сложения и умножения. Существуют и другие способы построения поля комплексных чисел: см., в частности, примеры 7.4 и 9.2.14. Как «примирить» все эти, казалось бы, различные определения?

Для того чтобы лучше понять, что такое поле комплексных чисел, нужно прежде подумать над тем, что такое поле вещественных чисел. Строгое построение поля вещественных чисел обычно приводится в курсе анализа. Мы не будем входить в его детали. Однако заметим, что имеется несколько определений вещественных чисел: как бесконечных десятичных дробей, как сечений Дедекинда множества рациональных чисел и т. д. Формально говоря, при этом получаются различные поля. Какое из них является «настоящим» полем вещественных чисел? Ответ на этот вопрос состоит в том, что все они изоморфны и их следует рассматривать просто как различные модели одного и того же объекта, называемого полем вещественных чисел.

Наиболее удовлетворительным в подобной ситуации всегда является аксиоматический подход, при котором сначала формулируются в виде аксиом свойства, которыми должен обладать искомый объект, а затем доказываются, что этими свойствами он определяется однозначно с точностью до изоморфизма, и с помощью какой-либо конструкции доказываются его существование. В случае поля вещественных чисел такими аксиомами (помимо аксиом поля) могут быть аксиомы порядка, аксиома Архимеда и аксиома непрерывности.

Замечание 1. Нетрудно доказать, что любые две модели поля вещественных чисел не просто изоморфны, но между ними имеется *единственный* изоморфизм. (Доказательство сводится к доказательству того, что всякий изоморфизм поля \mathbb{R} на себя тождествен, и основано на соображении, что неотрицательные числа при любом изоморфизме должны переходить в неотрицательные, так как они и только они являются квадратами в поле \mathbb{R} .) Это означает, что каждый элемент поля \mathbb{R} имеет свою индивидуальность, т. е. в любой модели могут быть идентифицированы числа 10 , $\sqrt{2}$, π и т. д.

Дадим теперь аксиоматическое определение поля комплексных чисел.

Определение 1. *Поле комплексных чисел* называется всякое поле \mathbb{C} , обладающее следующими свойствами:

- 1) оно содержит в качестве подполя поле \mathbb{R} вещественных чисел;
- 2) оно содержит такой элемент i , что $i^2 = -1$;
- 3) оно минимально среди полей с этими свойствами, т. е. если $K \subset \mathbb{C}$ — какое-либо подполе, содержащее \mathbb{R} и i , то $K = \mathbb{C}$.

Замечание 2. Из равенства $x^2 + 1 = (x - i)(x + i)$ следует, что уравнение $x^2 = -1$ имеет в \mathbb{C} ровно 2 решения: i и $-i$. Если какое-либо подполе содержит одно из этих решений, то оно содержит и другое.

Построенное выше поле \mathbb{C} обладает этими свойствами. Если теперь \mathbb{C}' — другое поле комплексных чисел и $i' \in \mathbb{C}'$ — такой элемент, что $(i')^2 = -1$, то, поскольку формулы (1) и (2) остаются справедливыми при замене i на i' , отображение

$$f: \mathbb{C} \rightarrow \mathbb{C}', \quad a + bi \mapsto a + bi' \quad (a, b \in \mathbb{R}),$$

является изоморфизмом поля \mathbb{C} на поле \mathbb{C}' .

Таким образом, поле \mathbb{C} , удовлетворяющее приведенным выше аксиомам, существует и единственно с точностью до изоморфизма.

§ 5. Кольца вычетов

Расширения кольца целых чисел приводят к цепочке колец

$$\mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C},$$

в которую, как мы позже увидим, можно вставить и другие звенья (в том числе и продолжить ее вправо). Кольца вычетов определяют также на основе целых чисел, но идея их определения совершенно иная. Это часто используемый в математике прием «склейки» — образования фактормножества по отношению эквивалентности.

Пусть M — какое-либо множество. Всякое подмножество $R \subset M \times M$ называется *отношением* на множестве M . Если $(a, b) \in R$, то говорят, что элементы a и b находятся в отношении R , и пишут aRb .

Приведем примеры отношений.

Пример 1. M — множество людей; aRb , если a знает b .

Пример 2. M то же самое; aRb , если a и b знакомы.

Пример 3. M то же самое; aRb , если a и b живут в одном доме.

Пример 4. $M = \mathbb{R}$; aRb , если $a \leq b$.

Пример 5. M — множество окружностей на плоскости; aRb , если окружности a и b равны, т. е. переводятся одна в другую движением.

Отношение R называется *отношением эквивалентности*, если оно обладает следующими свойствами:

- 1) aRa (*рефлексивность*);
- 2) $aRb \Rightarrow bRa$ (*симметричность*);
- 3) aRb и $bRc \Rightarrow aRc$ (*транзитивность*).

Из приведенных выше примеров отношений только третье и пятое являются отношениями эквивалентности: первое и четвертое не симметричны, а второе симметрично, но не транзитивно.

Отношение эквивалентности обычно записывается как $a \sim_R b$ или просто $a \sim b$.

Пусть R — отношение эквивалентности на множестве M . Для каждого $a \in M$ положим

$$R(a) = \{b \in M : a \sim_R b\}.$$

Из свойств отношений эквивалентности легко выводится, что $a \in \in R(a)$ и

$$R(a) \cap R(b) \neq \emptyset \Rightarrow R(a) = R(b).$$

Таким образом, подмножества $R(a)$ образуют разбиение множества M , т. е. покрывают его и попарно не пересекаются. Они называются *классами эквивалентности отношения R* . Два элемента эквивалентны тогда и только тогда, когда они принадлежат одному классу.

Множество, элементами которого являются классы эквивалентности отношения R , называется *фактормножеством* множества M по отношению эквивалентности R и обозначается через M/R . Отображение

$$M \rightarrow M/R, \quad a \mapsto R(a),$$

называется *отображением факторизации*.

Так, в третьем из приведенных выше примеров классы эквивалентности — это множества жильцов одного дома. Фактормножество можно отождествить с множеством домов; тогда отображение факторизации — это отображение, ставящее в соответствие каждому человеку дом, в котором он живет. В пятом примере классы эквивалентности — это множества окружностей одного радиуса, фактормножество отождествляется с множеством положительных чисел, а отображение факторизации — это отображение, ставящее в соответствие каждой окружности ее радиус.

Пусть в множестве M задана некоторая операция $(x, y) \mapsto x * y$. Отношение эквивалентности R в множестве M называется *согласованным* с операцией $*$, если

$$a \sim_R a', \quad b \sim_R b' \Rightarrow a * b \sim_R a' * b'.$$

В этом случае на фактормножестве M/R также можно определить операцию $*$ по правилу

$$R(a) * R(b) = R(a * b). \quad (6)$$

В словесном выражении это определение выглядит так: чтобы произвести операцию над какими-либо двумя классами эквива-

лентности, надо выбрать в них произвольных представителей, произвести операцию над ними и взять тот класс, в котором будет лежать получившийся элемент. Тот факт, что этот класс не будет зависеть от выбора указанных представителей, как раз и обеспечивается согласованностью отношения эквивалентности с операцией.

Очевидно, что все свойства операции в M , имеющие характер тождества, например коммутативность и ассоциативность, наследуются определенной таким образом операцией в M/R . То же самое можно сказать о наличии нуля (единицы) и противоположного (обратного) элемента. Более точно, если, скажем, операция в M называется сложением и в M имеется нулевой элемент 0 относительно этой операции, то $R(0)$ — нулевой элемент в M/R ; если $-a$ — элемент, противоположный элементу a в M , то $R(-a)$ — элемент, противоположный элементу $R(a)$ в M/R .

Приступим теперь к построению колец вычетов. Пусть n — фиксированное натуральное число. Рассмотрим в множестве \mathbb{Z} целых чисел следующее отношение сравнимости по модулю n : a сравнимо с b по модулю n (обозначение: $a \equiv b \pmod{n}$), если $a - b$ делится на n или, что равносильно, если a и b дают одинаковые остатки при делении на n .

Очевидно, что это отношение эквивалентности, причем классы эквивалентности могут быть занумерованы числами $0, 1, \dots, n - 1$ таким образом, что r -й класс состоит из всех целых чисел, дающих при делении на n остаток r .

Класс эквивалентности, содержащий целое число a , называется *вычетом* числа a по модулю n и обозначается через $[a]_n$ или просто через $[a]$, если ясно, какое n имеется в виду.

Фактормножество множества \mathbb{Z} по отношению сравнимости по модулю n обозначается через \mathbb{Z}_n . Мы можем написать, что

$$\mathbb{Z}_n = \{[0]_n, [1]_n, \dots, [n - 1]_n\},$$

но следует понимать, что каждый элемент множества \mathbb{Z}_n можно обозначать по-разному. Так, элемент $[1]_n$ может быть с таким же успехом обозначен через $[2n + 1]_n$, $[-(n - 1)]_n$ и т. д.

Докажем теперь, что отношение сравнимости по модулю n согласовано с операциями сложения и умножения в \mathbb{Z} . Пусть

$$a \equiv a' \pmod{n}, \quad b \equiv b' \pmod{n}$$

Тогда

$$a + b \equiv a' + b \equiv a' + b' \pmod{n}$$

и, аналогично,

$$ab \equiv a'b \equiv a'b' \pmod{n}.$$

Таким образом, мы можем определить в множестве \mathbb{Z}_n операции сложения и умножения по формулам

$$[a]_n + [b]_n = [a + b]_n, \quad [a]_n [b]_n = [ab]_n$$

(справедливым для любых $a, b \in \mathbb{Z}$). Тем самым \mathbb{Z}_n превращается в коммутативное ассоциативное кольцо с единицей. Оно называется *кольцом вычетов по модулю n* .

Пример 6. Ниже приведены таблицы сложения и умножения в кольце \mathbb{Z}_5 . При этом ради простоты квадратные скобки в обозначениях элементов этого кольца опущены.

+	0	1	2	3	4
0	0	1	2	3	4
1	1	2	3	4	0
2	2	3	4	0	1
3	3	4	0	1	2
4	4	0	1	2	3

×	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

Мы видим, в частности, что элементы 2 и 3 взаимно обратны, а элемент 4 обратен сам себе.

Пример 7. Вычислим $[2]^{100}$ в кольце \mathbb{Z}_{125} :

$$[2]^7 = [128] = [3], \quad [2]^{35} = ([2]^7)^5 = [3]^5 = [243] = [-7],$$

$$[2]^{50} = [2]^{35} ([2]^7)^2 [2] = [-7][3]^2 [2] = [-126] = [-1],$$

$$[2]^{100} = ([2]^{50})^2 = [1].$$

Полученный результат означает, что

$$2^{100} \equiv 1 \pmod{125}.$$

Учитывая, что 2^{100} делится на 8, получаем

$$2^{100} \equiv 376 \pmod{1000},$$

т. е. десятичная запись числа 2^{100} оканчивается на 376.

Кольцо \mathbb{Z}_n обладает всеми свойствами поля, кроме, быть может, обратимости ненулевых элементов. Очевидно, что \mathbb{Z}_2 — поле из двух элементов, о котором шла речь в задаче 3.2. Рассмотрение приведенной выше таблицы умножения в кольце \mathbb{Z}_5 показывает, что \mathbb{Z}_5 — также поле. С другой стороны, \mathbb{Z}_4 — не поле, так как элемент $[2]$ в этом кольце необратим.

Теорема 1. *Кольцо \mathbb{Z}_n является полем тогда и только тогда, когда n — простое число.*

Доказательство. 1) Пусть n составное, т. е. $n = kl$, где $1 < k, l < n$. Тогда $[k]_n, [l]_n \neq 0$, но

$$[k]_n[l]_n = [kl]_n = [n]_n = 0.$$

Таким образом, в кольце \mathbb{Z}_n имеются делители нуля и, значит, оно не является полем.

2) Пусть, напротив, n — простое число и $[a]_n \neq 0$, т. е. a не делится на n . Будем искать элемент, обратный к $[a]_n$, подбором, т. е. умножая $[a]_n$ по очереди на все элементы кольца. Получим элементы

$$[0]_n, [a]_n, [2a]_n, \dots, [(n-1)a]_n. \quad (7)$$

Докажем, что все они различны. В самом деле, если $[ka]_n = [la]_n$ ($0 \leq k < l \leq n-1$), то $[(l-k)a]_n = 0$, т. е. $(l-k)a$ делится на n , что невозможно, так как ни $l-k$, ни a на n не делятся. (Здесь мы использовали то, что n простое.) Следовательно, в последовательности элементов (7) встречаются все элементы кольца \mathbb{Z}_n , в том числе $[1]_n$, а это и означает, что элемент $[a]_n$ обратим. \square

Задача 1. Доказать, что при любом n элемент $[k]_n$ обратим в кольце \mathbb{Z}_n тогда и только тогда, когда n и k взаимно просты.

В полях вычетов мы встречаемся с новым явлением, не имевшим места в числовых полях (подполях поля комплексных чисел). А именно, в поле \mathbb{Z}_n (n простое) выполняется равенство

$$\underbrace{1 + 1 + \dots + 1}_n = 0. \quad (8)$$

(Конечно, это верно и в кольце \mathbb{Z}_n при любом n .) Это приводит к некоторым особенностям алгебраических преобразований в этом поле, о которых мы скажем ниже.

Пусть, вообще, K — произвольное поле. Наименьшее натуральное n , для которого в поле K выполняется равенство (8), называется *характеристикой* этого поля; если такого n не существует, то

говорят, что K — поле нулевой характеристики. Таким образом, \mathbb{Z}_n (n простое) — поле характеристики n , а числовые поля имеют нулевую характеристику. Характеристика поля K обозначается через $\text{char } K$.

Если $\text{char } K = n$, то для любого $a \in K$

$$\underbrace{a + a + \dots + a}_n = \underbrace{(1 + 1 + \dots + 1)}_n a = 0a = 0.$$

Характеристика поля, если она положительна, всегда является простым числом. В самом деле, пусть $\text{char } K = n = kl$ ($1 < k, l < n$). Тогда

$$\underbrace{1 + 1 + \dots + 1}_n = \underbrace{(1 + 1 + \dots + 1)}_k \underbrace{(1 + 1 + \dots + 1)}_l = 0$$

и, значит, либо $\underbrace{1 + 1 + \dots + 1}_k = 0$, либо $\underbrace{1 + 1 + \dots + 1}_l = 0$, что противоречит определению характеристики.

Большинство формул элементарной алгебры справедливы в любом поле, так как при их выводе используются только те свойства операций сложения и умножения, которые входят в число аксиом поля или являются их следствием. Особенность полей положительной характеристики проявляется только в тех формулах, которые содержат умножение или деление на натуральные числа.

Рассмотрим, например, формулу

$$(a + b)^2 = a^2 + 2ab + b^2.$$

Она справедлива в любом поле, если понимать $2ab$ как $ab + ab$. Однако в поле характеристики 2 она принимает более простой вид

$$(a + b)^2 = a^2 + b^2.$$

Более общо, в поле характеристики p справедливо тождество

$$(a + b)^p = a^p + b^p.$$

В самом деле, по формуле бинома Ньютона

$$(a + b)^p = \sum_{k=0}^p C_p^k a^{p-k} b^k.$$

Однако при $0 < k < p$

$$C_p^k = \frac{p(p-1)\dots(p-k+1)}{k!}$$

(число сочетаний из p по k), очевидно, делится на p . Следовательно, все слагаемые формулы бинорма Ньютона, кроме первого и последнего, в рассматриваемом случае равны нулю.

Задача 2. Вывести отсюда, что в поле \mathbb{Z}_p справедливо тождество $a^p = a$. (Другое доказательство последнего факта, называемого малой теоремой Ферма, будет дано в § 4.5.)

Хуже обстоит дело, когда приходится делить на натуральное число, например, когда мы находим выражение для ab из выписанной выше формулы квадрата суммы. Для того чтобы придать смысл этому делению в любом поле, можно рассматривать умножение на натуральное число k как умножение на элемент $\underbrace{1 + 1 + \dots + 1}_k$ данно-

го поля; тогда деление на k можно понимать как деление на этот элемент. Однако если k делится на характеристику поля, то этот элемент равен нулю и деление невозможно.

Так, формула для решения квадратного уравнения, содержащая деление на 2, применима в указанном смысле в любом поле характеристики $\neq 2$, но в поле характеристики 2 она не работает.

Пример 8. Решим квадратное уравнение

$$x^2 + x - 1 = 0$$

в поле \mathbb{Z}_{11} . По обычной формуле находим:

$$x_{1,2} = \frac{[-1] \pm \sqrt{[5]}}{[2]}.$$

Так как $[5] = [16] = [4]^2$, то можно считать, что $\sqrt{[5]} = [4]$ (одно из значений квадратного корня). Следовательно,

$$x_1 = \frac{[-1] + [4]}{[2]} = \frac{[3]}{[2]} = \frac{[14]}{[2]} = [7], \quad x_2 = \frac{[-1] - [4]}{[2]} = \frac{[-5]}{[2]} = \frac{[6]}{[2]} = [3].$$

§ 6. Векторные пространства

Векторы, рассматриваемые в элементарной геометрии, можно не только складывать, но и умножать на числа. Анализ свойств этих двух операций приводит к понятию векторного пространства.

Прежде чем мы дадим определение, необходимо отметить, что здесь мы выходим за рамки того понимания операции на множестве, которое принималось до сих пор. Умножение вектора на число

не есть операция над двумя элементами одного и того же множества. Это операция, которая каждой паре (число, вектор) ставит в соответствие вектор. В общем определении векторного пространства дело обстоит так же, однако вещественные числа заменяются элементами произвольного (но фиксированного) поля.

Определение 1. Векторным (или линейным) пространством над полем K называется множество V с операциями сложения и умножения на элементы поля K , обладающими следующими свойствами:

- 1) относительно сложения V есть абелева группа;
- 2) $\lambda(a + b) = \lambda a + \lambda b$ для любых $\lambda \in K$, $a, b \in V$;
- 3) $(\lambda + \mu)a = \lambda a + \mu a$ для любых $\lambda, \mu \in K$, $a \in V$;
- 4) $(\lambda\mu)a = \lambda(\mu a)$ для любых $\lambda, \mu \in K$, $a \in V$;
- 5) $1a = a$ для любого $a \in V$.

Элементы векторного пространства называются *векторами*. Элементы поля K , в отличие от векторов, мы будем иногда, допуская вольность речи, называть числами, даже если K не есть числовое поле.

Векторы в смысле элементарной геометрии мы будем отныне называть *геометрическими векторами*. Операции над ними удовлетворяют всем аксиомам векторного пространства, что, собственно, и послужило основой для данного выше определения. Пространство геометрических векторов евклидовой плоскости (соответственно трехмерного евклидова пространства) мы будем обозначать через E^2 (соответственно через E^3). Подчеркнем, что это векторное пространство над полем \mathbb{R} . Приведем другие важные примеры векторных пространств.

Пример 1. Множество K^n строк длины n с элементами из поля K является векторным пространством над K относительно операций, определенных формулами

$$(a_1, a_2, \dots, a_n) + (b_1, b_2, \dots, b_n) = (a_1 + b_1, a_2 + b_2, \dots, a_n + b_n),$$

$$\lambda(a_1, a_2, \dots, a_n) = (\lambda a_1, \lambda a_2, \dots, \lambda a_n).$$

Пример 2. Множество $F(X, K)$ всех функций на множестве X со значениями в поле K является векторным пространством относительно обычных операций над функциями:

$$(f + g)(x) = f(x) + g(x), \quad (\lambda f)(x) = \lambda f(x).$$

Пример 3. Пусть K — подполе поля L . Тогда L можно рассматривать как векторное пространство над K , определив умножение элементов из L на элементы из K просто как умножение в L . В частности, поле \mathbb{C} есть в этом смысле векторное пространство над \mathbb{R} .

Укажем некоторые следствия аксиом векторного пространства, не являющиеся следствиями только аксиом абелевой группы. Все они доказываются аналогично похожим на них следствиям аксиом кольца (см. § 3). Символом 0 обозначается как нуль поля K , так и нулевой вектор, т. е. нуль аддитивной группы V ; читатель увидит, что это не приводит к путанице.

- 1) $\lambda 0 = 0$ для любого $\lambda \in K$ (здесь 0 — нулевой вектор).
- 2) $\lambda(-a) = -\lambda a$ для любых $\lambda \in K, a \in V$.
- 3) $\lambda(a - b) = \lambda a - \lambda b$ для любых $\lambda \in K, a, b \in V$.
- 4) $0a = 0$ для любого $a \in V$ (здесь 0 слева — число, справа — вектор).

$$5) (-1)a = -a \text{ для любого } a \in V.$$

$$6) (\lambda - \mu)a = \lambda a - \mu a \text{ для любых } \lambda, \mu \in K, a \in V.$$

Определение 2. Подмножество U векторного пространства V называется *подпространством*, если

- 1) U является подгруппой аддитивной группы V ;
- 2) $a \in U \Rightarrow \lambda a \in U$ для любого $\lambda \in K$.

Замечание 1. В определении подгруппы требуется, чтобы

$$a \in U \Rightarrow -a \in U.$$

При наличии условия 2) это свойство выполняется автоматически, так как $-a = (-1)a$.

Подпространство векторного пространства само является векторным пространством относительно тех же операций.

Пример 4. В пространстве E^3 множество векторов, параллельных заданной плоскости или прямой, является подпространством.

Пример 5. В пространстве $F(X, \mathbb{R})$ всех функций на заданном промежутке X числовой прямой множество непрерывных функций является подпространством.

В каждом векторном пространстве V есть два «тривиальных» подпространства: само пространство V и нулевое подпространство (состоящее из одного нулевого вектора). Последнее мы будем обозначать символом 0 .

Определение 3. Векторные пространства V и U над полем K называются *изоморфными*, если существует такое биективное отобра-

ражение

$$\varphi: V \rightarrow U,$$

что

1) $\varphi(a + b) = \varphi(a) + \varphi(b)$ для любых $a, b \in V$;

2) $\varphi(\lambda a) = \lambda \varphi(a)$ для любых $\lambda \in K, a \in V$.

Само отображение φ называется при этом *изоморфизмом* пространств V и U .

Как мы увидим в § 2.2, описание векторных пространств с точностью до изоморфизма весьма просто. В частности, все так называемые конечномерные векторные пространства, с которыми мы в основном и будем иметь дело в этом курсе, изоморфны пространствам K^n . Ключевым понятием этой теории является понятие базиса.

Всякое выражение вида

$$\lambda_1 a_1 + \lambda_2 a_2 + \dots + \lambda_n a_n \quad (\lambda_1, \lambda_2, \dots, \lambda_n \in K)$$

называется *линейной комбинацией* векторов $a_1, a_2, \dots, a_n \in V$. Говорят, что вектор b *линейно выражается* через векторы a_1, a_2, \dots, a_n , если он равен некоторой их линейной комбинации.

Определение 4. Система векторов $\{e_1, e_2, \dots, e_n\} \subset V$ называется *базисом* векторного пространства V , если каждый вектор $a \in V$ единственным образом линейно выражается через e_1, e_2, \dots, e_n . Коэффициенты этого выражения называются *координатами* вектора a в базисе $\{e_1, e_2, \dots, e_n\}$.

Пример 6. Из геометрии известно, что любые два неколлинеарных вектора e_1, e_2 составляют базис пространства E^2 (рис. 5). Аналогично любые три некомпланарных вектора составляют базис пространства E^3 .

Пример 7. *Единичные строки*

$$e_1 = (1, 0, \dots, 0),$$

$$e_2 = (0, 1, \dots, 0),$$

.....

$$e_n = (0, 0, \dots, 1)$$

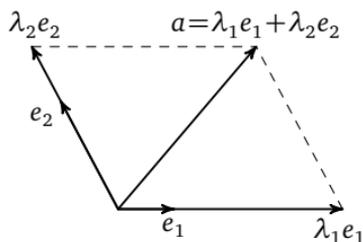


Рис. 5

составляют базис пространства K^n . Координатами строки $a = (a_1, a_2, \dots, a_n)$ в этом базисе служат числа a_1, a_2, \dots, a_n . Конечно, в пространстве K^n имеются и другие базисы.

Пример 8. В качестве базиса поля \mathbb{C} как векторного пространства над \mathbb{R} (см. пример 3) можно взять $\{1, i\}$. Координатами комплексного числа в этом базисе служат его вещественная и мнимая части.

Предложение 1. *Всякое векторное пространство V над полем K , имеющее базис из n векторов, изоморфно пространству K^n .*

Доказательство. Пусть $\{e_1, e_2, \dots, e_n\}$ — базис пространства V . Рассмотрим отображение

$$\varphi: V \rightarrow K^n,$$

ставящее в соответствие каждому вектору строку из его координат в базисе $\{e_1, e_2, \dots, e_n\}$. Очевидно, что это биективное отображение. Далее, если

$$a = a_1 e_1 + a_2 e_2 + \dots + a_n e_n, \quad b = b_1 e_1 + b_2 e_2 + \dots + b_n e_n,$$

то

$$\begin{aligned} a + b &= (a_1 + b_1)e_1 + (a_2 + b_2)e_2 + \dots + (a_n + b_n)e_n, \\ \lambda a &= (\lambda a_1)e_1 + (\lambda a_2)e_2 + \dots + (\lambda a_n)e_n. \end{aligned}$$

Отсюда следует, что φ — изоморфизм. □

Пример 9. Пространство E^2 (соответственно E^3) изоморфно \mathbb{R}^2 (соответственно \mathbb{R}^3).

§ 7. Алгебры

Ввиду крайней простоты своего строения векторные пространства не интересны сами по себе, но они служат необходимым фоном для многих алгебраических (и не только алгебраических) теорий. Так, комбинируя понятия векторного пространства и кольца, мы приходим к важному понятию алгебры.

Определение 1. *Алгеброй над полем K называется множество A с операциями сложения, умножения и умножения на элементы поля K , обладающими следующими свойствами:*

- 1) относительно сложения и умножения на элементы поля A есть векторное пространство;
- 2) относительно сложения и умножения A есть кольцо;
- 3) $(\lambda a)b = a(\lambda b) = \lambda(ab)$ для любых $\lambda \in K, a, b \in A$.

Замечание 1. Термин «алгебра», употреблявшийся нами до сих пор только как название одного из разделов математики, в этом определении имеет, естественно, другой смысл.

Пример 1. Всякое поле L , содержащее K в качестве подполя, можно рассматривать как алгебру над K . В частности, поле \mathbb{C} есть алгебра над \mathbb{R} .

Пример 2. Пространство E^3 есть алгебра относительно операции векторного умножения.

Пример 3. Множество $F(X, K)$ функций на множестве X со значениями в поле K (см. пример 6.2) является алгеброй над K относительно обычных операций сложения и умножения функций и умножения функции на число. Эта алгебра коммутативна, ассоциативна и обладает единицей (каковой является функция, тождественно равная единице).

Задача 1. Доказать, что кольцо 2^X из задачи 3.1 превращается в алгебру над полем \mathbb{Z}_2 , если определить в нем умножение на элементы этого поля по правилам

$$0M = \emptyset, \quad 1M = M \quad \forall M \in 2^X.$$

Предположим, что алгебра A обладает базисом $\{e_1, e_2, \dots, e_n\}$ как векторное пространство над K , и пусть

$$a = a_1e_1 + a_2e_2 + \dots + a_n e_n = \sum_{i=1}^n a_i e_i,$$

$$b = b_1e_1 + b_2e_2 + \dots + b_n e_n = \sum_{i=1}^n b_i e_i$$

— два произвольных элемента этой алгебры. Тогда из дистрибутивности умножения относительно сложения и свойства 3) в определении алгебры следует, что

$$ab = \sum_{i=1}^n a_i (e_i b) = \sum_{i=1}^n a_i \left(\sum_{j=1}^n b_j (e_i e_j) \right) = \sum_{i,j=1}^n a_i b_j (e_i e_j).$$

Это показывает, что умножение в алгебре A полностью определяется произведениями базисных векторов.

Если умножение базисных векторов коммутативно, т. е.

$$e_i e_j = e_j e_i \quad \forall i, j,$$

то и умножение в алгебре A в целом коммутативно. В самом деле, для любых $a, b \in A$ мы тогда в предыдущих обозначениях получаем

$$ab = \sum_{i,j} a_i b_j (e_i e_j) = \sum_{i,j} b_j a_i (e_j e_i) = ba.$$

Аналогично доказывается, что если умножение базисных векторов ассоциативно, т. е.

$$(e_i e_j) e_k = e_i (e_j e_k) \quad \forall i, j, k,$$

то и умножение в алгебре A в целом ассоциативно.

С другой стороны, если V — какое-то векторное пространство с базисом $\{e_1, e_2, \dots, e_n\}$ и e_{ij} ($i, j = 1, 2, \dots, n$) — произвольные векторы этого пространства, то мы можем определить операцию умножения в V по правилу

$$ab = \sum_{i,j} a_i b_j e_{ij}$$

и тем самым превратить V в алгебру. Иначе говоря, если мы не требуем, чтобы умножение обладало какими-нибудь дополнительными свойствами (например, было коммутативным), таблица умножения базисных векторов алгебры может быть совершенно произвольной.

Пример 4. Поле \mathbb{C} как алгебра над \mathbb{R} задается следующей таблицей умножения базисных векторов:

\times	1	i
1	1	i
i	i	-1

Это можно принять за определение поля \mathbb{C} . Проверка коммутативности и ассоциативности умножения сводится тогда к тривиальной проверке коммутативности и ассоциативности умножения элементов 1 и i .

Пример 5. В ортонормированном (т.е. состоящем из ортогональных единичных векторов) базисе $\{i, j, k\}$ пространства E^3 таблица векторного умножения выглядит следующим образом:

\times	i	j	k
i	0	k	$-j$
j	$-k$	0	i
k	j	$-i$	0

Это умножение антикоммумутативно и удовлетворяет тождеству Якоби (см. пример 3.4). Последнее тождество достаточно проверить для базисных векторов, что не составляет труда (проделайте это!).

Пример 6. Алгебра кватернионов \mathbb{H} задается базисом $\{1, i, j, k\}$ со следующей таблицей умножения:

\times	1	i	j	k
1	1	i	j	k
i	i	-1	k	$-j$
j	j	$-k$	-1	i
k	k	j	$-i$	-1

Эта алгебра ассоциативна (проверьте это!), но не коммутативна. Она содержит в качестве подалгебры (см. определение ниже) алгебру комплексных чисел. Позже мы увидим, что в алгебре \mathbb{H} , как и в поле, всякий ненулевой элемент обратим. Таким образом, это «некоммутативное поле».

Задача 2. Доказать, что двумерная алгебра над полем \mathbb{Z}_2 с базисом $\{1, \alpha\}$ и таблицей умножения

\times	1	α
1	1	α
α	α	$1 + \alpha$

является полем (из 4 элементов).

Подмножество алгебры называется *подалгеброй*, если оно одновременно является подпространством и подкольцом. Отображение алгебр называется *изоморфизмом*, если оно одновременно является изоморфизмом векторных пространств и колец.

§ 8. Алгебра матриц

Матрицей размера $m \times n$ над полем K называется прямоугольная таблица из элементов поля K , имеющая m строк и n столбцов. В буквенной записи элементы матрицы обычно обозначаются одной и той же буквой с двумя индексами, первый из которых есть

номер строки, а второй — номер столбца:

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}.$$

Иногда ради краткости мы будем писать просто $A = (a_{ij})$.

Суммой матриц $A = (a_{ij})$ и $B = (b_{ij})$ одинакового размера называется матрица

$$A + B = (a_{ij} + b_{ij}).$$

Произведением матрицы $A = (a_{ij})$ на элемент $\lambda \in K$ называется матрица

$$\lambda A = (\lambda a_{ij}).$$

Относительно этих двух операций все матрицы размера $m \times n$ образуют векторное пространство, которое мы будем обозначать $K^{m \times n}$. По сути дела оно не отличается от пространства строк K^{mn} . Специфика матриц проявляется при определении их умножения.

Произведением матрицы $A = (a_{ij})$ размера $m \times n$ и матрицы $B = (b_{jk})$ размера $n \times p$ называется матрица $AB = (c_{ik})$ размера $m \times p$, элементы которой находятся по формулам

$$c_{ik} = \sum_{j=1}^n a_{ij} b_{jk}.$$

Иными словами, если определить скалярное произведение строки длины n на столбец высоты n как сумму произведений их соответствующих элементов, то можно сказать, что c_{ik} есть скалярное произведение i -й строки матрицы A на k -й столбец матрицы B . Подчеркнем, что произведение двух матриц определено только тогда, когда их размеры согласованы, а именно, когда число столбцов первой матрицы равно числу строк второй.

Это определение мотивируется следующим образом. Будем говорить, что переменные y_1, \dots, y_m (принимаяющие значения в поле K) линейно выражаются через переменные x_1, \dots, x_n , если существуют такие $a_{ij} \in K$, что

$$y_i = \sum_{j=1}^n a_{ij} x_j \quad (i = 1, 2, \dots, m).$$

Предположим, что переменные x_1, \dots, x_n , в свою очередь, линейно выражаются через переменные z_1, \dots, z_p :

$$x_j = \sum_{k=1}^p b_{jk} z_k \quad (j = 1, 2, \dots, n).$$

Тогда, подставив эти выражения в предыдущие, мы получим линейные выражения переменных y_1, \dots, y_m через z_1, \dots, z_p :

$$y_i = \sum_{k=1}^p c_{ik} z_k \quad (i = 1, 2, \dots, m),$$

где $c_{ik} = \sum_{j=1}^n a_{ij} b_{jk}$, т. е. матрица $C = (c_{ik})$ есть произведение матриц $A = (a_{ij})$ и $B = (b_{jk})$ в смысле данного выше определения.

Пример 1.

$$\begin{pmatrix} 1 & 0 & 2 \\ 0 & -1 & 3 \end{pmatrix} \begin{pmatrix} 2 & -1 \\ 0 & 5 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 \cdot 2 + 0 \cdot 0 + 2 \cdot 1 & 1 \cdot (-1) + 0 \cdot 5 + 2 \cdot 1 \\ 0 \cdot 2 + (-1) \cdot 0 + 3 \cdot 1 & 0 \cdot (-1) + (-1) \cdot 5 + 3 \cdot 1 \end{pmatrix} = \begin{pmatrix} 4 & 1 \\ 3 & -2 \end{pmatrix}.$$

Пример 2.

$$\begin{aligned} & \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix} \begin{pmatrix} \cos \beta & -\sin \beta \\ \sin \beta & \cos \beta \end{pmatrix} = \\ & = \begin{pmatrix} \cos \alpha \cos \beta - \sin \alpha \sin \beta & -\cos \alpha \sin \beta - \sin \alpha \cos \beta \\ \sin \alpha \cos \beta + \cos \alpha \sin \beta & -\sin \alpha \sin \beta + \cos \alpha \cos \beta \end{pmatrix} = \\ & = \begin{pmatrix} \cos(\alpha + \beta) & -\sin(\alpha + \beta) \\ \sin(\alpha + \beta) & \cos(\alpha + \beta) \end{pmatrix}. \end{aligned}$$

Умножение матриц ассоциативно в том смысле, что

$$(AB)C = A(BC), \quad (9)$$

если только размеры матриц A, B, C согласованы таким образом, что указанные произведения имеют смысл.

В самом деле, пусть

$$(AB)C = (u_{il}), \quad A(BC) = (v_{il}).$$

Имеем тогда

$$\begin{aligned} u_{il} &= \sum_k \left(\sum_j a_{ij} b_{jk} \right) c_{kl} = \sum_{j,k} a_{ij} b_{jk} c_{kl}, \\ v_{il} &= \sum_j a_{ij} \left(\sum_k b_{jk} c_{kl} \right) = \sum_{j,k} a_{ij} b_{jk} c_{kl}, \end{aligned}$$

так что $u_{il} = v_{il}$.

Матрица размера $n \times n$ называется *квадратной матрицей* порядка n . Квадратная матрица имеет две диагонали. Одна из них, ведущая из левого верхнего угла в правый нижний, называется *главной диагональю*, или просто *диагональю*, а другая — *побочной диагональю*. Квадратная матрица называется *диагональной*, если все ее элементы, находящиеся вне (главной) диагонали, равны нулю. Умножение на диагональные матрицы выглядит особенно просто:

$$\begin{pmatrix} a_1 & 0 & \dots & 0 \\ 0 & a_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & a_n \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1p} \\ b_{21} & b_{22} & \dots & b_{2p} \\ \dots & \dots & \dots & \dots \\ b_{n1} & b_{n2} & \dots & b_{np} \end{pmatrix} = \begin{pmatrix} a_1 b_{11} & a_1 b_{12} & \dots & a_1 b_{1p} \\ a_2 b_{21} & a_2 b_{22} & \dots & a_2 b_{2p} \\ \dots & \dots & \dots & \dots \\ a_n b_{n1} & a_n b_{n2} & \dots & a_n b_{np} \end{pmatrix}$$

(каждая строка второй матрицы умножается на соответствующий диагональный элемент первой матрицы) и, аналогично,

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} b_1 & 0 & \dots & 0 \\ 0 & b_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & b_n \end{pmatrix} = \begin{pmatrix} a_{11} b_1 & a_{12} b_2 & \dots & a_{1n} b_n \\ a_{21} b_1 & a_{22} b_2 & \dots & a_{2n} b_n \\ \dots & \dots & \dots & \dots \\ a_{m1} b_1 & a_{m2} b_2 & \dots & a_{mn} b_n \end{pmatrix}$$

(каждый столбец первой матрицы умножается на соответствующий диагональный элемент второй матрицы). Диагональную матрицу

$$\begin{pmatrix} a_1 & 0 & \dots & 0 \\ 0 & a_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & a_n \end{pmatrix}$$

мы будем обозначать $\text{diag}(a_1, a_2, \dots, a_n)$.

Диагональная матрица вида

$$E = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

называется *единичной матрицей*. Из предыдущих формул следует, что для любой матрицы A размера $t \times n$

$$AE = A, \quad EA = A, \tag{10}$$

где E в первом случае обозначает единичную матрицу порядка n , а во втором — единичную матрицу порядка t .

Следующие очевидные свойства связывают операцию умножения матриц с другими операциями:

$$A(B + C) = AB + AC, \quad (A + B)C = AC + BC, \quad (11)$$

$$(\lambda A)B = A(\lambda B) = \lambda(AB) \quad \forall \lambda \in K. \quad (12)$$

(Как и в свойстве ассоциативности, здесь предполагается, что размеры матриц согласованы таким образом, что все указанные действия имеют смысл.)

Сумма и произведение квадратных матриц одного и того же порядка n определены и также являются квадратными матрицами порядка n . Свойства (9)—(12) показывают, что все квадратные матрицы порядка n образуют ассоциативную алгебру с единицей. Мы будем обозначать ее $L_n(K)$.¹

Отметим некоторые «отрицательные» свойства алгебры $L_n(K)$ при $n \geq 2$. (Алгебра $L_1(K)$ есть поле K .)

1) Алгебра $L_n(K)$ не коммутативна. При $n = 2$ это можно продемонстрировать на следующем примере:

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Аналогичные примеры можно привести и при $n > 2$.

2) Алгебра $L_n(K)$ имеет делители нуля. Это показывает, например, второе из приведенных выше равенств. Более того, существуют такие ненулевые матрицы, квадрат которых равен нулю, например,

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}^2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

3) Не всякий ненулевой элемент алгебры $L_n(K)$ обратим. Это следует из наличия делителей нуля и того факта, что делитель нуля не может быть обратим (см. доказательство отсутствия делителей нуля в поле, данное в § 3). Так, например, матрицы $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ и $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ необратимы в $L_2(K)$.

Задача 1. Матрица E_{ij} , у которой на (i, j) -м месте стоит 1, а на остальных местах — нули, называется *матричной единицей* (не путать с единичной матрицей!). Матричные единицы E_{ij} ($i, j = 1, \dots, n$)

¹Буква «L» в нашем обозначении — первая буква слова «linear»; это связано с тем, что матрицы можно интерпретировать как линейные отображения (см. § 5.2). Другое часто встречающееся обозначение для этой алгебры — $M_n(K)$.

образуют базис векторного пространства $L_n(K)$. Выписать таблицу умножения алгебры $L_n(K)$ в этом базисе.

Задача 2. Матрицы вида λE ($\lambda \in K$) называются *скалярными*. Очевидно, что всякая скалярная матрица перестановочна со всеми квадратными матрицами того же порядка. Доказать обратное: всякая квадратная матрица, перестановочная со всеми квадратными матрицами того же порядка, скалярна.

Задача 3. Доказать, что в алгебре $L_2(\mathbb{R})$ матрицы вида

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$$

образуют подалгебру, изоморфную алгебре комплексных чисел.

Задача 4. Доказать, что в алгебре $L_2(\mathbb{C})$, рассматриваемой как алгебра над \mathbb{R} , матрицы вида

$$\begin{pmatrix} a & -\bar{b} \\ b & \bar{a} \end{pmatrix}$$

образуют подалгебру, изоморфную алгебре кватернионов (см. пример 7.6).

Для каждой матрицы

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

определим *транспонированную матрицу*

$$A^T = \begin{pmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \dots & \dots & \dots & \dots \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{pmatrix},$$

строками которой служат столбцы матрицы A , а столбцами — строки матрицы A . Если (i, j) -й элемент транспонированной матрицы обозначить через a_{ij}^T , то

$$a_{ij}^T = a_{ji}.$$

Очевидно, что

$$(A^T)^T = A,$$

$$(A+B)^T = A^T + B^T,$$

$$(\lambda A)^T = \lambda A^T \quad \forall \lambda \in K.$$

Докажем, что

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

В самом деле, пусть $AB = C = (c_{ik})$; тогда

$$c_{ki}^T = c_{ik} = \sum_j a_{ij} b_{jk} = \sum_j b_{kj}^T a_{ji}^T,$$

откуда видно, что $C^T = B^T A^T$.

Замечание 1. Читатель может проследить, что все построения последних трех параграфов проходят без изменений, если в качестве K взять произвольное коммутативное ассоциативное кольцо с единицей, например, кольцо целых чисел или кольцо вычетов. Единственное отличие является терминологическим: вместо термина «векторное пространство» в этой более общей ситуации употребляется термин «модуль». (См. § 9.3.)

Система уравнений называется *совместной*, если она имеет хотя бы одно решение, и *несовместной* в противном случае. Совместная система может иметь одно или более решений. Решить систему уравнений — это значит найти все ее решения.

Подчеркнем, что одно решение системы уравнений с n неизвестными — это упорядоченный набор из n чисел, т. е. элемент пространства K^n .

Существует простой общий метод решения систем линейных уравнений, называемый *методом Гаусса*. Его идея состоит в приведении любой системы линейных уравнений с помощью некоторых специальных преобразований, называемых элементарными, к эквивалентной системе некоторого простого вида, все решения которой легко найти. Напомним, что две системы уравнений называются *эквивалентными*, если множества их решений совпадают, т. е. если каждое решение первой из них является решением второй и наоборот.

Определение 1. *Элементарными преобразованиями системы линейных уравнений называются преобразования следующих трех типов:*

- 1) прибавление к одному уравнению другого, умноженного на число;
- 2) перестановка двух уравнений;
- 3) умножение одного уравнения на число, отличное от нуля.

Подчеркнем, что при элементарном преобразовании первого типа изменяется только одно уравнение — то, к которому прибавляется другое, умноженное на число.

Очевидно, что всякое решение исходной системы уравнений является решением новой системы, полученной элементарным преобразованием. С другой стороны, исходная система уравнений может быть получена из новой системы подходящим элементарным преобразованием того же типа. Так, если мы прибавим к первому уравнению второе, умноженное на c , то можно вернуться назад, прибавив к первому уравнению новой системы ее второе уравнение (оно такое же, как у исходной системы), умноженное на $-c$. Поэтому при любом элементарном преобразовании мы получаем систему уравнений, эквивалентную исходной.

Так как нам удобнее работать не с самими системами линейных уравнений, а с их (расширенными) матрицами, дадим соответствующее определение для матриц.

равными нулю. Мы получим матрицу вида

$$\left(\begin{array}{ccc|cccc} 0 & \dots & 0 & a_{1j_1} & \dots & \dots \\ \hline & & \mathbf{0} & & \boxed{A_1} & \end{array} \right).$$

Поступая таким же образом с матрицей A_1 , мы в конце концов получим матрицу вида (2). \square

Замечание 1. В этом доказательстве мы обошлись без элементарных преобразований третьего типа. Однако на практике они могут быть полезны.

Пример 1. Приведем к ступенчатому виду матрицу

$$\begin{pmatrix} 1 & 2 & 1 & 0 & 2 \\ 1 & 3 & 2 & -1 & 4 \\ 2 & 1 & -1 & 3 & -2 \\ 2 & 0 & -2 & 3 & 1 \end{pmatrix}.$$

Вычитая из 2-й, 3-й и 4-й строк 1-ю строку, умноженную на 1, 2 и 2 соответственно, получаем матрицу

$$\begin{pmatrix} 1 & 2 & 1 & 0 & 2 \\ 0 & 1 & 1 & -1 & 2 \\ 0 & -3 & -3 & 3 & -6 \\ 0 & -4 & -4 & 3 & -3 \end{pmatrix}.$$

Далее, прибавляя к 3-й и 4-й строкам 2-ю строку, умноженную на 3 и 4 соответственно, получаем матрицу

$$\begin{pmatrix} 1 & 2 & 1 & 0 & 2 \\ 0 & 1 & 1 & -1 & 2 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 5 \end{pmatrix}.$$

Наконец, переставляя 3-ю и 4-ю строки, получаем ступенчатую матрицу

$$\begin{pmatrix} 1 & 2 & 1 & 0 & 2 \\ 0 & 1 & 1 & -1 & 2 \\ 0 & 0 & 0 & -1 & 5 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Замечание 2. Предыдущий пример специально подобран таким образом, чтобы j_1, j_2, \dots, j_r не были просто первыми r членами на-

турального ряда. Такая ситуация является в определенном смысле исключительной. Например, $j_1 \neq 1$ только при условии, что первый столбец исходной матрицы нулевой. Как правило,

$$j_1 = 1, \quad j_2 = 2, \quad \dots, \quad j_r = r.$$

В этом случае матрица (2) называется *трапецидальной*.

Применим доказанную теорему к решению систем линейных уравнений.

Определение 3. Система линейных уравнений называется *ступенчатой*, если ее расширенная матрица ступенчатая.

Из теоремы 1 следует, что всякую систему линейных уравнений с помощью элементарных преобразований можно привести к ступенчатому виду. Поэтому нам достаточно научиться решать ступенчатые системы.

Введем некоторую терминологию. Квадратная матрица $A = (a_{ij})$ называется (*верхней*) *треугольной*, если $a_{ij} = 0$ при $i > j$, и *строго* *треугольной*, если, кроме того, $a_{ii} \neq 0$ при всех i . Система линейных уравнений называется (*строго*) *треугольной*, если ее матрица коэффициентов (*строго*) *треугольна*.

Замечание 3. Квадратная матрица $A = (a_{ij})$ называется *нижней* *треугольной*, если $a_{ij} = 0$ при $i < j$.

Рассмотрим теперь произвольную ступенчатую систему линейных уравнений. Пусть число ненулевых строк (число ступенек) ее матрицы коэффициентов равно r , а число ненулевых строк расширенной матрицы равно \tilde{r} . Очевидно, что $\tilde{r} = r$ или $r + 1$.

Возможны следующие три принципиально разных случая.

1-й случай. $\tilde{r} = r + 1$. В этом случае система содержит уравнение вида

$$0x_1 + 0x_2 + \dots + 0x_n = b,$$

где $b \neq 0$, и, следовательно, несовместна.

2-й случай. $\tilde{r} = r = n$. В этом случае после отбрасывания нулевых уравнений получается строго треугольная система. Из ее последнего уравнения однозначно определяется x_n , затем из предпоследнего уравнения определяется x_{n-1} и т. д. Следовательно, система имеет единственное решение.

3-й случай. $\tilde{r} = r < n$. Пусть в этом случае j_1, j_2, \dots, j_r — номера ведущих коэффициентов ненулевых уравнений системы. Извест-

Считая неизвестные x_1, x_2, x_4 главными, а неизвестное x_3 — свободным, перепишем систему в виде

$$\begin{cases} x_1 + 2x_2 &= -x_3 + 2, \\ x_2 - x_4 &= -x_3 + 2, \\ -x_4 &= 5. \end{cases}$$

Решая ее относительно x_1, x_2, x_4 , находим общее решение

$$\begin{cases} x_1 = x_3 + 8, \\ x_2 = -x_3 - 3, \\ x_4 = -5. \end{cases}$$

Замечание 4. Для единообразия можно считать, что в случае определенной системы все неизвестные являются главными, а свободные неизвестные отсутствуют. Общее решение есть тогда единственное решение системы.

Строго треугольную матрицу можно путем элементарных преобразований строк привести к единичной матрице. Для этого нужно сначала к каждой строке, кроме последней, прибавить последнюю строку с таким коэффициентом, чтобы элемент последнего столбца стал равным нулю, затем аналогичным образом, прибавляя предпоследнюю строку, сделать равными нулю все элементы предпоследнего столбца, кроме диагонального, и т. д. В результате мы получим диагональную матрицу. Умножая ее строки на подходящие числа, мы получим единичную матрицу. Пользуясь этим, можно при решении системы линейных уравнений не останавливаться на ступенчатом виде, а, продолжив преобразования, привести матрицу коэффициентов при главных неизвестных к единичной матрице. Тогда общее решение просто считывается с полученной матрицы. Эта процедура называется *обратным ходом метода Гаусса*.

Пример 3. Продолжим преобразование примера 1, предварительно отбросив нулевую строку. Вычтя из 2-й строки 3-ю, получим матрицу

$$\begin{pmatrix} 1 & 2 & 1 & 0 & 2 \\ 0 & 1 & 1 & 0 & -3 \\ 0 & 0 & 0 & -1 & 5 \end{pmatrix}.$$

Вычтя из 1-й строки удвоенную 2-ю и умножив 3-ю строку на -1 , получим матрицу

$$\begin{pmatrix} 1 & 0 & -1 & 0 & 8 \\ 0 & 1 & 1 & 0 & -3 \\ 0 & 0 & 0 & 1 & -5 \end{pmatrix}.$$

Таким образом, система уравнений из примера 2 эквивалентна системе

$$\begin{cases} x_1 - x_3 = 8, \\ x_2 + x_3 = -3, \\ x_4 = -5. \end{cases}$$

Переносим члены с x_3 в правую часть, получаем уже найденное выше общее решение.

Система однородных линейных уравнений всегда совместна, так как она имеет нулевое решение. Если она определена, то она имеет только нулевое решение, если неопределена, то имеет хотя бы одно ненулевое решение (и даже бесконечно много таких решений, если поле K бесконечно). В предыдущих обозначениях, последний случай имеет место, если $r < n$. Пользуясь тем, что всегда $r \leq m$, мы приходим к следующей теореме, которая является важным теоретическим следствием метода Гаусса.

Теорема 2. *Всякая система однородных линейных уравнений, число уравнений которой меньше числа неизвестных, имеет ненулевое решение.*

Совокупность всех решений совместной системы линейных уравнений с n неизвестными не может быть произвольным подмножеством пространства K^n . Читателю, вероятно, известно, что нетривиальная совместная система линейных уравнений задает в E^2 точку или прямую, а в E^3 — точку, прямую или плоскость. Следующая теорема является алгебраической версией этих утверждений, справедливой в любой размерности и для любого поля.

Теорема 3. 1) *Совокупность всех решений системы однородных линейных уравнений с n неизвестными является подпространством пространства K^n .*

2) *Совокупность всех решений произвольной совместной системы линейных уравнений есть сумма какого-либо одного ее решения и подпространства решений системы однородных линейных уравнений с той же матрицей коэффициентов.*

на два вопроса: какие векторные пространства обладают базисом и не может ли в векторном пространстве быть двух базисов, состоящих из разного числа векторов.

Чтобы ответить на эти вопросы, нам понадобится ввести некоторые понятия и доказать ряд утверждений, которые важны и сами по себе.

Пусть V — векторное пространство над полем K .

Линейная комбинация

$$\lambda_1 a_1 + \lambda_2 a_2 + \dots + \lambda_n a_n \quad (\lambda_1, \lambda_2, \dots, \lambda_n \in K)$$

векторов $a_1, a_2, \dots, a_n \in V$ называется *тривиальной*, если $\lambda_1 = \lambda_2 = \dots = \lambda_n = 0$, и *нетривиальной* в противном случае.

Определение 1. Векторы a_1, a_2, \dots, a_n называются *линейно зависимыми*, если существует их нетривиальная линейная комбинация, равная нулю, и *линейно независимыми* в противном случае.

Подчеркнем, что понятие линейной зависимости (или независимости) относится не к отдельным векторам, а к их совокупностям или, как говорят, системам векторов.

Замечание 1. Понятие системы векторов отличается от понятия множества векторов тем, что, во-первых, векторы системы предполагаются занумерованными и, во-вторых, среди них могут быть равные. Таким образом, система из n векторов — это, в сущности, отображение множества $\{1, 2, \dots, n\}$ в пространство V . Заметим, однако, что свойство системы векторов быть линейно зависимой или независимой не зависит от нумерации векторов в ней.

Замечание 2. Термин «линейная комбинация» на самом деле употребляется в двух смыслах: как указание действий, которые производятся над данными векторами, что равносильно заданию коэффициентов $\lambda_1, \lambda_2, \dots, \lambda_n$, и как результат этих действий. В выражении «нетривиальная линейная комбинация данных векторов равна нулю» нетривиальность понимается в первом смысле, а равенство нулю — во втором.

Линейная независимость векторов a_1, a_2, \dots, a_n означает, иными словами, что равенство

$$\lambda_1 a_1 + \lambda_2 a_2 + \dots + \lambda_n a_n = 0$$

выполняется только при $\lambda_1 = \lambda_2 = \dots = \lambda_n = 0$.

Пример 1. Система, состоящая из одного вектора, линейно зависима тогда и только тогда, когда этот вектор нулевой.

Пример 2. Система, состоящая из двух векторов, линейно зависима тогда и только тогда, когда эти векторы пропорциональны.

Пример 3. Три геометрических вектора линейно зависимы тогда и только тогда, когда они компланарны (параллельны одной плоскости).

Очевидно, что если система векторов содержит линейно зависимую подсистему, то она сама линейно зависима. Так, например, всякая система векторов, содержащая пропорциональные векторы, линейно зависима.

Лемма 1. Векторы a_1, a_2, \dots, a_n ($n > 1$) линейно зависимы тогда и только тогда, когда хотя бы один из них линейно выражается через остальные.

Доказательство. 1) Пусть, например,

$$a_1 = \mu_2 a_2 + \dots + \mu_n a_n.$$

тогда

$$a_1 - \mu_2 a_2 - \dots - \mu_n a_n = 0,$$

что показывает линейную зависимость векторов a_1, a_2, \dots, a_n .

2) Обратно, пусть

$$\lambda_1 a_1 + \lambda_2 a_2 + \dots + \lambda_n a_n = 0,$$

где не все коэффициенты $\lambda_1, \lambda_2, \dots, \lambda_n$ равны нулю. Допустим для определенности, что $\lambda_1 \neq 0$. Тогда

$$a_1 = -\frac{\lambda_2}{\lambda_1} a_2 - \dots - \frac{\lambda_n}{\lambda_1} a_n,$$

т. е. a_1 линейно выражается через a_2, \dots, a_n . □

Замечание 3. Неверно, что любой вектор линейно зависимой системы линейно выражается через остальные. Пусть, например, a — какой-нибудь ненулевой вектор. Система $\{a, 0\}$ линейно зависима, так как

$$0a + 1 \cdot 0 = 0,$$

но вектор a , очевидно, не выражается через нулевой вектор.

Лемма 2. Пусть векторы a_1, a_2, \dots, a_n линейно независимы. Вектор b линейно выражается через a_1, a_2, \dots, a_n тогда и только тогда, когда векторы a_1, a_2, \dots, a_n, b линейно зависимы.

Доказательство. Если вектор b линейно выражается через a_1, a_2, \dots, a_n , то a_1, a_2, \dots, a_n, b линейно зависимы согласно предыдущей лемме. Обратно, пусть

$$\lambda_1 a_1 + \lambda_2 a_2 + \dots + \lambda_n a_n + \mu b = 0,$$

причем не все коэффициенты $\lambda_1, \lambda_2, \dots, \lambda_n, \mu$ равны нулю. Можно утверждать, что $\mu \neq 0$: в противном случае мы получили бы линейную зависимость векторов a_1, a_2, \dots, a_n , что противоречит условию. Но тогда

$$b = -\frac{\lambda_1}{\mu} a_1 - \frac{\lambda_2}{\mu} a_2 - \dots - \frac{\lambda_n}{\mu} a_n. \quad \square$$

Лемма 3. Пусть вектор b линейно выражается через векторы a_1, a_2, \dots, a_n . Это выражение единственно тогда и только тогда, когда векторы a_1, a_2, \dots, a_n линейно независимы.

Доказательство. 1) Пусть вектор b допускает два различных выражения через a_1, a_2, \dots, a_n :

$$b = \lambda_1 a_1 + \lambda_2 a_2 + \dots + \lambda_n a_n = \lambda'_1 a_1 + \lambda'_2 a_2 + \dots + \lambda'_n a_n.$$

тогда

$$(\lambda'_1 - \lambda_1) a_1 + (\lambda'_2 - \lambda_2) a_2 + \dots + (\lambda'_n - \lambda_n) a_n = 0$$

есть линейная зависимость между a_1, a_2, \dots, a_n .

2) Обратно, пусть

$$\mu_1 a_1 + \mu_2 a_2 + \dots + \mu_n a_n = 0$$

есть линейная зависимость между a_1, a_2, \dots, a_n . Тогда если

$$b = \lambda_1 a_1 + \lambda_2 a_2 + \dots + \lambda_n a_n,$$

то также

$$b = (\lambda_1 + \mu_1) a_1 + (\lambda_2 + \mu_2) a_2 + \dots + (\lambda_n + \mu_n) a_n,$$

что дает другое выражение b через a_1, a_2, \dots, a_n . □

Предложение 1 (основная лемма о линейной зависимости). Если векторы b_1, b_2, \dots, b_m линейно выражаются через векторы a_1, a_2, \dots, a_n , причем $m > n$, то векторы b_1, b_2, \dots, b_m линейно зависимы.

Доказательство. Пусть

$$\begin{aligned} b_1 &= \mu_{11}a_1 + \mu_{12}a_2 + \dots + \mu_{1n}a_n, \\ b_2 &= \mu_{21}a_1 + \mu_{22}a_2 + \dots + \mu_{2n}a_n, \\ &\dots\dots\dots \\ b_m &= \mu_{m1}a_1 + \mu_{m2}a_2 + \dots + \mu_{mn}a_n. \end{aligned}$$

Для любых $\lambda_1, \lambda_2, \dots, \lambda_m \in K$ получаем

$$\begin{aligned} \lambda_1 b_1 + \lambda_2 b_2 + \dots + \lambda_m b_m &= (\lambda_1 \mu_{11} + \lambda_2 \mu_{21} + \dots + \lambda_m \mu_{m1})a_1 + \\ &+ (\lambda_1 \mu_{12} + \lambda_2 \mu_{22} + \dots + \lambda_m \mu_{m2})a_2 + \\ &\dots\dots\dots \\ &+ (\lambda_1 \mu_{1n} + \lambda_2 \mu_{2n} + \dots + \lambda_m \mu_{mn})a_n. \end{aligned}$$

Рассмотрим систему n однородных линейных уравнений с m неизвестными

$$\begin{cases} \mu_{11}x_1 + \mu_{21}x_2 + \dots + \mu_{m1}x_m = 0, \\ \mu_{12}x_1 + \mu_{22}x_2 + \dots + \mu_{m2}x_m = 0, \\ \dots\dots\dots \\ \mu_{1n}x_1 + \mu_{2n}x_2 + \dots + \mu_{mn}x_m = 0. \end{cases}$$

Если $(\lambda_1, \lambda_2, \dots, \lambda_m)$ — произвольное решение этой системы, то

$$\lambda_1 b_1 + \lambda_2 b_2 + \dots + \lambda_m b_m = 0.$$

С другой стороны, по теореме 1.2 эта система имеет ненулевое решение. Следовательно, векторы b_1, b_2, \dots, b_m линейно зависимы. \square

Пусть $S \subset V$ — какое-то подмножество. Совокупность всевозможных (конечных) линейных комбинаций векторов из S называется *линейной оболочкой* множества S и обозначается через $\langle S \rangle$. Это наименьшее подпространство пространства V , содержащее S (проверьте это!). Говорят, что пространство V *порождается* множеством S , если $\langle S \rangle = V$.

Определение 2. Векторное пространство называется *конечномерным*, если оно порождается конечным числом векторов, и *бесконечномерным* в противном случае.

Ввиду леммы 3 определение 1.6.4 базиса векторного пространства можно переформулировать следующим образом.

Определение 3. *Базисом* векторного пространства V называется всякая линейно независимая система векторов, порождающая пространство V .

Теорема 1. *Всякое конечномерное векторное пространство V обладает базисом. Более точно, из всякого конечного порождающего множества $S \subset V$ можно выбрать базис пространства V .*

Доказательство. Если множество S линейно зависимо, то по лемме 1 в нем найдется вектор, линейно выражающийся через остальные. Выкидывая этот вектор, мы получаем порождающее множество из меньшего числа векторов. Продолжая так дальше, мы в конце концов получим линейно независимое порождающее множество, т. е. базис. \square

Теорема 2. *Все базисы конечномерного векторного пространства V содержат одно и то же число векторов.*

Это число называется *размерностью* пространства V и обозначается $\dim V$.

Доказательство. Если бы в пространстве V существовали два базиса из разного числа векторов, то, согласно предложению 1, тот из них, в котором больше векторов, был бы линейно зависим, что противоречит определению базиса. \square

Замечание 4. Нулевое векторное пространство (состоящее из одного нулевого вектора) считается обладающим «пустым базисом»; в соответствии с этим его размерность считается равной нулю.

Пример 4. Пространство E^2 (соответственно E^3) имеет размерность 2 (соответственно 3).

Пример 5. Ввиду примера 1.6.7 пространство K^n имеет размерность n .

Пример 6. Поле комплексных чисел как векторное пространство над \mathbb{R} имеет размерность 2, а алгебра кватернионов (см. пример 1.7.6) — размерность 4.

Пример 7. Если X — конечное множество из n элементов, то векторное пространство $F(X, K)$ всех функций на X со значениями в K (см. пример 1.6.2) имеет размерность n . В самом деле, рассмотрим так называемые δ -функции δ_a ($a \in X$), определяемые формулами

$$\delta_a(x) = \begin{cases} 1, & \text{если } x = a, \\ 0, & \text{если } x \neq a. \end{cases}$$

Очевидно, что любая функция $\varphi \in F(X, K)$ единственным образом выражается через δ -функции, а именно,

$$\varphi = \sum_{a \in X} \varphi(a) \delta_a.$$

Следовательно, функции δ_a , $a \in X$, составляют базис пространства $F(X, K)$, причем координатами функции в этом базисе служат ее значения. Если множество X бесконечно, то для любого n в пространстве $F(X, K)$ имеется n линейно независимых векторов, например, $\delta_{a_1}, \delta_{a_2}, \dots, \delta_{a_n}$, где $a_1, a_2, \dots, a_n \in X$ различны, и, следовательно, пространство $F(X, K)$ бесконечномерно.

Пример 8. Поле \mathbb{R} как векторное пространство над \mathbb{Q} бесконечномерно. В самом деле, если бы оно было конечномерным, то вещественное число определялось бы конечным набором рациональных чисел — своих координат в некотором базисе этого пространства. Но тогда множество всех вещественных чисел было бы счетным, что неверно.

Задача 1. Найти число векторов n -мерного векторного пространства над конечным полем из q элементов.

Задача 2. Доказать, что пространство всех непрерывных функций на любом промежутке числовой прямой бесконечномерно.

Из основной леммы о линейной зависимости (предложение 1) следует, что любые $m > n$ векторов n -мерного векторного пространства V линейно зависимы и, значит, в любом (конечном или бесконечном) множестве $S \subset V$ имеется максимальное линейно независимое подмножество, т. е. такое линейно независимое подмножество, которое становится линейно зависимым при добавлении к нему любого из оставшихся векторов множества S . Более того, любое линейно независимое подмножество множества S можно дополнить до максимального линейно независимого подмножества.

Предложение 2. Всякое максимальное линейно независимое подмножество $\{e_1, \dots, e_k\}$ множества S является базисом линейной оболочки $\langle S \rangle$ этого множества.

Доказательство. Нужно доказать, что каждый вектор из $\langle S \rangle$ линейно выражается через e_1, \dots, e_k . По определению линейной оболочки каждый вектор из $\langle S \rangle$ линейно выражается через векторы из S . Поэтому достаточно доказать, что каждый вектор $a \in S$ линейно выражается через e_1, \dots, e_k . Для $a \in \{e_1, \dots, e_k\}$ это очевидно. Для $a \notin \{e_1, \dots, e_k\}$ это следует из леммы 2. \square

Применяя высказанные соображения к $S = V$, мы получаем следующую теорему.

Теорема 3. Всякую линейно независимую систему векторов конечномерного векторного пространства V можно дополнить до базиса.

В частности, любой ненулевой вектор можно включить в базис, а любые n линейно независимых векторов n -мерного векторного пространства уже составляют базис.

Задача 3. Найти число базисов n -мерного векторного пространства над полем из q элементов.

Следующая теорема устанавливает свойство монотонности размерности.

Теорема 4. *Всякое подпространство U конечномерного векторного пространства V также конечномерно, причем $\dim U \leq \dim V$. Более того, если $U \neq V$, то $\dim U < \dim V$.*

Доказательство. Пусть $\{e_1, e_2, \dots, e_k\}$ — максимальная линейно независимая система векторов подпространства U . Согласно предложению 2, $\{e_1, e_2, \dots, e_k\}$ — базис этого подпространства. Следовательно, $\dim U = k$. Линейно независимую систему $\{e_1, e_2, \dots, e_k\}$ можно дополнить до базиса всего пространства V . Следовательно, если $U \neq V$, то $\dim V > k$. \square

Задача 4. Найти число k -мерных подпространств n -мерного векторного пространства над полем из q элементов.

Следующая теорема дает исчерпывающее описание всех конечномерных векторных пространств.

Теорема 5. *Конечномерные векторные пространства над одним и тем же полем изоморфны тогда и только тогда, когда они имеют одинаковую размерность.*

Доказательство. Если $f: V \rightarrow U$ — изоморфизм векторных пространств и $\{e_1, e_2, \dots, e_n\}$ — базис пространства V , то $\{f(e_1), f(e_2), \dots, f(e_n)\}$ — базис пространства U , так что $\dim V = \dim U$. Обратно, согласно предложению 1.6.1, всякое n -мерное векторное пространство над полем K изоморфно K^n ; следовательно, все такие пространства изоморфны между собой. \square

Таким образом, в любом рассуждении мы вправе заменить произвольное n -мерное векторное пространство над полем K пространством строк K^n . В пространстве K^n имеется «привилегированный» базис, состоящий из единичных строк (см. пример 1.6.7). С другой стороны, если в каком-либо n -мерном векторном пространстве V задан базис, то сопоставление каждому вектору строки его координат (как в доказательстве предложения 1.6.1) определяет канонический изоморфизм пространства V и пространства K^n , при котором векторам заданного базиса соответствуют единичные строки. В этом смысле можно сказать, что пространство строк — это не что

иное, как конечномерное векторное пространство с выделенным базисом.

Выясним, как связаны между собой координаты вектора в разных базисах. Пусть $\{e_1, \dots, e_n\}$ и $\{e'_1, \dots, e'_n\}$ — два базиса векторного пространства V . Выразим векторы второго базиса через первый базис:

$$e'_j = \sum_i e_i c_{ij} \quad (j = 1, \dots, n). \quad (7)$$

Квадратная матрица $C = (c_{ij})$ называется *матрицей перехода* от базиса $\{e_1, \dots, e_n\}$ к базису $\{e'_1, \dots, e'_n\}$. Согласно этому определению, j -й столбец матрицы C есть столбец координат вектора e'_j в базисе $\{e_1, \dots, e_n\}$. Если распространить правило умножения матриц на случай, когда элементами одной из них являются векторы (что имеет смысл ввиду операций, определенных в векторном пространстве), то равенства (7) могут быть переписаны в следующей матричной форме:

$$(e'_1, \dots, e'_n) = (e_1, \dots, e_n)C. \quad (8)$$

Пусть $x \in V$ — какой-либо вектор. Разложим его по базисам $\{e_1, \dots, e_n\}$ и $\{e'_1, \dots, e'_n\}$:

$$x = x_1 e_1 + \dots + x_n e_n = x'_1 e'_1 + \dots + x'_n e'_n.$$

Положим

$$X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad X' = \begin{pmatrix} x'_1 \\ \vdots \\ x'_n \end{pmatrix}.$$

Тогда

$$x = (e'_1, \dots, e'_n)X' = (e_1, \dots, e_n)CX',$$

откуда получается следующая формула *преобразования координат* при переходе от базиса $\{e_1, \dots, e_n\}$ к базису $\{e'_1, \dots, e'_n\}$:

$$X = CX' \quad (9)$$

или, более подробно,

$$x_i = \sum_j c_{ij} x'_j \quad (i = 1, \dots, n). \quad (10)$$

Понятия базиса и размерности могут быть распространены на бесконечномерные векторные пространства. Чтобы это сделать, надо определить, что такое линейная комбинация бесконечной систе-

мы векторов. В чисто алгебраической ситуации нет иного выхода, кроме как ограничиться рассмотрением линейных комбинаций, в которых лишь конечное число коэффициентов отлично от нуля.

Пусть $\{a_i : i \in I\}$ — система векторов, занумерованных элементами бесконечного множества I . *Линейной комбинацией* векторов a_i , $i \in I$, называется выражение вида $\sum_{i \in I} \lambda_i a_i$, в котором лишь конечное число коэффициентов λ_i отлично от нуля, так что сумма фактически является конечной и, таким образом, имеет смысл. На основе этого определения линейной комбинации точно так же, как в случае конечных систем векторов, определяются понятия линейной выразимости, линейной зависимости и базиса.

Векторное пространство, обладающее счетным базисом, называется *счетномерным*.

Пример 9. Очевидно, что множество всех последовательностей (строк бесконечной длины) из элементов поля K является векторным пространством относительно операций сложения и умножения на элементы поля K , определяемых так же, как для строк конечной длины. Последовательность называется *финитной*, если лишь конечное число ее членов отлично от нуля. Финитные последовательности образуют подпространство в пространстве всех последовательностей. Обозначим его через K^∞ . В качестве его базисных векторов можно взять последовательности вида

$$e_i = (0, \dots, 0, 1, 0, \dots) \quad (i = 1, 2, \dots)$$

(единица стоит на i -м месте). Таким образом, пространство K^∞ счетномерно.

Так же, как предложение 1.6.1, доказывается тот факт, что всякое счетномерное векторное пространство над K изоморфно K^∞ .

Задача 5. Доказать, что поле \mathbb{R} как векторное пространство над \mathbb{Q} не является счетномерным.

Задача 6. Доказать, что из всякого счетного порождающего множества векторного пространства можно выбрать базис (конечный или счетный).

Задача 7. Доказать, что любое несчетное множество векторов в счетномерном векторном пространстве линейно зависимо (и, следовательно, любой базис счетен).

Задача 8. Доказать, что всякую (конечную или счетную) линейно независимую систему векторов счетномерного векторного пространства можно дополнить до базиса.

Задача 9. Доказать, что всякое подпространство счетномерного векторного пространства не более чем счетномерно (т. е. счетномерно или конечномерно). Привести пример счетномерного подпространства счетномерного векторного пространства, не совпадающего со всем пространством.

Задачи 6—9 представляют собой аналоги теорем 1—4 для счетномерных векторных пространств. Аналогичные утверждения могут быть доказаны и для несчетномерных пространств, но для этого требуется привлечение аппарата канторовской теории множеств (трансфинитной индукции или леммы Цорна). С другой стороны, такой чисто алгебраический подход имеет ограниченную сферу применения. Обычно несчетномерное векторное пространство снабжается топологией, которая позволяет придавать смысл бесконечным суммам векторов.

§ 3. Ранг матрицы

На основе понятия размерности векторного пространства вводятся понятия ранга системы векторов и ранга матрицы.

Определение 1. Рангом системы векторов называется размерность ее линейной оболочки. Рангом матрицы называется ранг системы ее строк.

Ранг матрицы A обозначается через $\text{rk } A$.

Системы векторов $\{a_1, a_2, \dots, a_n\}$ и $\{b_1, b_2, \dots, b_m\}$ называются эквивалентными, если каждый из векторов b_j линейно выражается через a_1, a_2, \dots, a_n и, наоборот, каждый из векторов a_i линейно выражается через b_1, b_2, \dots, b_m . Это, очевидно, равносильно совпадению линейных оболочек:

$$\langle a_1, a_2, \dots, a_n \rangle = \langle b_1, b_2, \dots, b_m \rangle.$$

Поэтому ранги эквивалентных систем векторов равны.

Из определения элементарных преобразований следует, что строки матрицы A' , полученной из матрицы A каким-либо элементарным преобразованием, линейно выражаются через строки матрицы A . Но так как матрица A может быть получена из A' обратным элементарным преобразованием, то и, наоборот, ее строки линейно выражаются через строки матрицы A' . Таким образом, системы строк матриц A и A' эквивалентны и, следовательно, ранги этих матриц равны.

Этим можно воспользоваться для вычисления ранга матрицы.

Предложение 1. Ранг матрицы равен числу ненулевых строк любой ступенчатой матрицы, к которой она приводится элементарными преобразованиями строк.

Доказательство. Так как ранг матрицы не меняется при элементарных преобразованиях, то нам достаточно доказать, что ранг ступенчатой матрицы равен числу ее ненулевых строк. Для этого, в свою очередь, достаточно доказать, что ненулевые строки ступенчатой матрицы линейно независимы.

Предположим, что линейная комбинация ненулевых строк ступенчатой матрицы (2) с коэффициентами $\lambda_1, \lambda_2, \dots, \lambda_r$ равна нулю. Рассматривая j_1 -ю координату этой линейной комбинации, находим, что $\lambda_1 a_{1j_1} = 0$, откуда $\lambda_1 = 0$. Рассматривая, далее, j_2 -ю координату с учетом того, что $\lambda_1 = 0$, находим, что $\lambda_2 a_{2j_2} = 0$, откуда $\lambda_2 = 0$. Продолжая так дальше, получаем, что все коэффициенты $\lambda_1, \lambda_2, \dots, \lambda_r$ равны нулю, что и требовалось доказать. \square

В частности, какую бы последовательность элементарных преобразований, приводящих заданную матрицу к ступенчатому виду, мы ни выбрали, число ненулевых строк полученной ступенчатой матрицы будет одним и тем же.

С учетом предложения 1 результаты, полученные в §1 при анализе ступенчатых систем линейных уравнений, приводят к следующим теоремам.

Теорема 1 (теорема Кронекера—Капелли). Система линейных уравнений совместна тогда и только тогда, когда ранг матрицы ее коэффициентов равен рангу расширенной матрицы.

Теорема 2. Совместная система линейных уравнений является определенной тогда и только тогда, когда ранг матрицы ее коэффициентов равен числу неизвестных.

Следующая теорема является ответом на вопрос о «степени неопределенности» системы линейных уравнений, поставленный в §1.

Теорема 3. Размерность пространства решений системы однородных линейных уравнений с n неизвестными и матрицей коэффициентов A равна $n - \text{rk } A$.

Доказательство. Рассмотрим систему уравнений (4). С помощью элементарных преобразований приведем ее к ступенчатому виду. В силу предложения 1 число ненулевых уравнений в этом ступенчатом виде будет равно $r = \text{rk } A$. Поэтому общее решение будет содержать r главных неизвестных и с точностью до перенумерации

ментарные преобразования строк матрицы A соответствуют элементарным преобразованиям этой системы, при которых ее решения не меняются. Отсюда получаем

Предложение 2. *Линейные зависимости между столбцами матрицы не меняются при элементарных преобразованиях строк.*

Следствие. *При элементарных преобразованиях строк матрицы ранг системы ее столбцов не меняется.*

Доказательство. Линейная зависимость между какими-то столбцами матрицы может пониматься как линейная зависимость между всеми ее столбцами, в которую остальные столбцы входят с нулевыми коэффициентами. Следовательно, если какие-то столбцы матрицы линейно зависимы, то они останутся линейно зависимыми после любых элементарных преобразований строк. Так как элементарные преобразования обратимы, то и наоборот: если какие-то столбцы матрицы линейно независимы, то они и останутся линейно независимыми. Значит, если какие-то столбцы матрицы составляют максимальную линейно независимую систему ее столбцов, то после любых элементарных преобразований строк столбцы с теми же номерами будут составлять максимальную линейно независимую систему столбцов полученной матрицы, и поэтому ранг матрицы не изменится. \square

Так как элементарными преобразованиями строк любую матрицу можно привести к ступенчатому виду, то для нахождения ранга системы столбцов достаточно научиться это делать для ступенчатых матриц.

Предложение 3. *Ранг системы столбцов ступенчатой матрицы равен числу ее ненулевых строк.*

Доказательство. Пусть A — ступенчатая матрица, число ненулевых строк которой равно r . Ясно, что при выкидывании нулевых строк линейные зависимости между столбцами сохраняются и, значит, ранг системы столбцов не меняется. Поэтому можно считать, что у матрицы A нет нулевых строк, т. е. имеется всего r строк. Но тогда ранг системы ее столбцов не превосходит r (размерности пространства всех столбцов высоты r). Покажем, что столбцы, проходящие через ведущие элементы строк (углы ступенек), линейно независимы. Для этого рассмотрим систему однородных линейных уравнений с матрицей коэффициентов A . Наличие нетривиальной линейной зависимости между указанными выше столбцами означало бы, что эта система имеет ненулевое решение, в котором все сво-

бодные неизвестные равны нулю, что невозможно. Следовательно, ранг матрицы A равен r . \square

Так как ранг системы строк ступенчатой матрицы также равен числу ее ненулевых строк, то из доказанных предложений вытекает

Теорема 4. *Ранг системы строк любой матрицы равен рангу системы ее столбцов.*

Иными словами, ранг матрицы не меняется при транспонировании.

Проведенные рассуждения одновременно дают удобный способ нахождения максимальной линейно независимой системы столбцов. А именно, если путем элементарных преобразований строк матрица A приведена к ступенчатому виду, в котором ведущие элементы ненулевых строк имеют номера j_1, \dots, j_r , то столбцы (исходной!) матрицы A с номерами j_1, \dots, j_r составляют максимальную линейно независимую систему ее столбцов. Заметим, что эта процедура не дает способа найти максимальную линейно независимую систему строк матрицы A . Дело в том, что, хотя ранг системы строк при элементарных преобразованиях строк не меняется, линейные зависимости между строками (в отличие от столбцов) не сохраняются.

Замечание 1. Элементарные преобразования столбцов матрицы определяются аналогично элементарным преобразованиям строк. Они равносильны умножению на элементарные матрицы справа. Из доказанного выше следует, что ранг матрицы не меняется и при элементарных преобразованиях столбцов.

Теорема 5. *Ранг произведения матриц не превосходит ранга каждого из множителей.*

Доказательство. Пусть $A = (a_{ij})$, $B = (b_{jk})$ и $AB = C = (c_{ik})$. Согласно определению умножения матриц

$$c_{ik} = \sum_j a_{ij} b_{jk}.$$

Рассматривая эти равенства при фиксированном i , мы видим, что i -я строка матрицы C есть линейная комбинация строк матрицы B с коэффициентами из i -й строки матрицы A . Следовательно, линейная оболочка строк матрицы C содержится в линейной оболочке строк матрицы B и, значит, $\text{rk } C \leq \text{rk } B$.

Аналогично, рассматривая те же равенства при фиксированном k , мы видим, что k -й столбец матрицы C есть линейная ком-

бинация столбцов матрицы A с коэффициентами из k -го столбца матрицы B . Следовательно, линейная оболочка столбцов матрицы C содержится в линейной оболочке столбцов матрицы A и, значит, $\text{rk } C \leq \text{rk } A$. \square

Задача 1. Доказать, что ранг суммы матриц не превосходит сумму их рангов. Привести пример, когда имеет место равенство.

Рассмотрим отдельно случай квадратных матриц.

Определение 2. Квадратная матрица A порядка n называется невырожденной, если $\text{rk } A = n$.

Иными словами, матрица A невырожденна, если ее строки линейно независимы, или, что эквивалентно, если ее столбцы линейно независимы.

Теорема 6. Квадратная система линейных уравнений определена (т. е. имеет единственное решение) тогда и только тогда, когда ее матрица коэффициентов невырожденна.

Доказательство. В самом деле, квадратная система линейных уравнений с матрицей коэффициентов A определена тогда и только тогда, когда элементарными преобразованиями строк матрица A приводится к строго треугольному виду, т. е. когда $\text{rk } A = n$. \square

Согласно общему определению обратного элемента в кольце с единицей (см. §1.3) матрицей, обратной матрице A , называется такая матрица A^{-1} , что $AA^{-1} = A^{-1}A = E$. Если обратная матрица существует, то она единственна.

Теорема 7. Квадратная матрица обратима тогда и только тогда, когда она невырожденна.

Доказательство. Из теоремы 5 следует, что произведение двух квадратных матриц может быть невырожденным, только если оба множителя невырожденны. Очевидно, что единичная матрица E невырожденна. Поэтому всякая обратимая матрица невырожденна.

Обратно, пусть матрица A невырожденна. Будем искать обратную матрицу как решение матричного уравнения $AX = E$. Для элементов k -го столбца матрицы X это дает систему линейных уравнений с матрицей коэффициентов A и столбцом свободных членов, равным k -му столбцу матрицы E . В силу невырожденности матрицы A каждая из этих систем и, тем самым, исходное матричное уравнение имеет единственное решение. Обозначим решение этого матричного уравнения через B .

Аналогично, рассмотрим матричное уравнение $YA = E$. Оно равносильно уравнению $A^T Y^T = E$ и потому также имеет единственное

решение (так как матрица A^T также невырожденна). Обозначим это решение через C .

Пользуясь ассоциативностью умножения матриц, получаем

$$B = (CA)B = C(AB) = C.$$

Таким образом, $B = C$ — матрица, обратная матрице A . □

Доказательство этой теоремы дает и практический способ нахождения обратной матрицы. А именно, матрица, обратная невырожденной матрице A порядка n , может быть найдена как решение матричного уравнения $AX = E$, которое сводится к решению n систем линейных уравнений с общей матрицей коэффициентов A , столбцы свободных членов которых составляют матрицу E . Эти системы могут решаться одновременно методом Гаусса. После приведения матрицы коэффициентов к единичной матрице (что возможно в силу ее невырожденности) преобразованные столбцы свободных членов составят искомую матрицу A^{-1} .

Пример 1. Найдем матрицу, обратную к матрице

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 5 \end{pmatrix}.$$

Для этого сделаем следующие элементарные преобразования:

$$\left(\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 3 & 5 & 0 & 1 \end{array} \right) \longrightarrow \left(\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 0 & -1 & -3 & 1 \end{array} \right) \longrightarrow \left(\begin{array}{cc|cc} 1 & 0 & -5 & 2 \\ 0 & 1 & 3 & -1 \end{array} \right).$$

Таким образом,

$$A^{-1} = \begin{pmatrix} -5 & 2 \\ 3 & -1 \end{pmatrix}.$$

Замечание 2. На самом деле приведенное доказательство теоремы 7 доказывает следующее более сильное утверждение: если матрица A вырожденна, то ни одно из уравнений $AX = E$ и $YA = E$ не имеет решения, а если она невырожденна, то каждое из этих уравнений имеет единственное решение и эти решения совпадают. (Априори можно было бы допустить возможность, что одно из этих уравнений имеет решение, а другое не имеет.)

§ 4. Определители

Вопрос о невырожденности квадратной матрицы или, что равносильно, о линейной независимости n векторов n -мерного пространства в каждом конкретном случае можно решить приведением

матрицы к ступенчатому виду элементарными преобразованиями строк. Однако представляет интерес нахождение общего условия, которому должны удовлетворять элементы матрицы для того, чтобы она была невырожденной. Поясним идею получения такого условия на примере геометрических векторов.

Пусть l — ориентированная прямая на плоскости. Для всякого вектора a обозначим через $f(a)$ его проекцию на l (взятую с соответствующим знаком). Из определения операций над векторами следует, что

$$f(a + b) = f(a) + f(b) \text{ для любых векторов } a, b,$$

$$f(\lambda a) = \lambda f(a) \text{ для любого вектора } a \text{ и любого числа } \lambda.$$

Всякая функция векторного аргумента, обладающая этими свойствами, называется *линейной*.

Пара неколлинеарных векторов $a_1, a_2 \in E^2$ называется *ориентированной положительно*, если поворот от a_1 к a_2 (на угол, меньший π) происходит в положительном направлении. Для любых векторов a_1, a_2 обозначим через $\text{area}(a_1, a_2)$ ориентированную площадь параллелограмма, натянутого на эти векторы, т. е. площадь, взятую со знаком плюс, если пара $\{a_1, a_2\}$ ориентирована положительно, и со знаком минус в противном случае; если векторы a_1 и a_2 коллинеарны, то положим $\text{area}(a_1, a_2) = 0$. Величина $|\text{area}(a_1, a_2)|$ может служить мерой линейной независимости векторов a_1 и a_2 .

Функция $\text{area}(a_1, a_2)$ векторных аргументов a_1 и a_2 обладает следующими свойствами:

- 1) она линейна по a_1 и по a_2 ;
- 2) $\text{area}(a_2, a_1) = -\text{area}(a_1, a_2)$;
- 3) если $\{e_1, e_2\}$ — положительно ориентированный ортонормированный базис, то $\text{area}(e_1, e_2) = 1$.

Последние два свойства очевидны. Для доказательства первого представим площадь параллелограмма как произведение основания на высоту. Мы получим тогда

$$\text{area}(a_1, a_2) = |a_1| h_2,$$

где $|a_1|$ — длина вектора a_1 , а h_2 — проекция вектора a_2 на прямую, ортогональную a_1 (рис. 1). В силу сказанного выше h_2 есть линейная функция от a_2 . Отсюда следует линей-

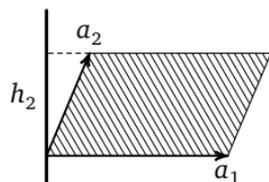


Рис. 1

ность $\text{area}(a_1, a_2)$ по a_2 . Аналогично, взяв за основание a_2 , можно доказать линейность по a_1 .

Свойств 1)–3) достаточно для вычисления $\text{area}(a_1, a_2)$. Выразим векторы a_1, a_2 через положительно ориентированный ортонормированный базис $\{e_1, e_2\}$:

$$\begin{aligned} a_1 &= a_{11}e_1 + a_{12}e_2, \\ a_2 &= a_{21}e_1 + a_{22}e_2. \end{aligned}$$

Тогда

$$\begin{aligned} \text{area}(a_1, a_2) &= \text{area}(a_{11}e_1 + a_{12}e_2, a_{21}e_1 + a_{22}e_2) = \\ &= a_{11}a_{21} \text{area}(e_1, e_1) + a_{11}a_{22} \text{area}(e_1, e_2) + a_{12}a_{21} \text{area}(e_2, e_1) + \\ &\quad + a_{12}a_{22} \text{area}(e_2, e_2) = a_{11}a_{22} - a_{12}a_{21}. \end{aligned}$$

Выражение $a_{11}a_{22} - a_{12}a_{21}$ называется определителем матрицы $A = (a_{ij})$ порядка 2. Из предыдущего следует, что векторы a_1 и a_2 линейно независимы тогда и только тогда, когда определитель матрицы, составленной из их координат, отличен от нуля.

Аналогичным образом можно доказать, что ориентированный объем $\text{vol}(a_1, a_2, a_3)$ параллелепипеда, натянутого на векторы a_1, a_2, a_3 , обладает следующими свойствами:

- 1) он линеен по каждому из трех аргументов a_1, a_2, a_3 ;
- 2) он меняет знак при перестановке любых двух аргументов;
- 3) если $\{e_1, e_2, e_3\}$ — положительно ориентированный ортонормированный базис, то $\text{vol}(e_1, e_2, e_3) = 1$.

(Тройка $\{a_1, a_2, a_3\}$ считается ориентированной положительно, если поворот от a_1 к a_2 со стороны a_3 происходит в положительном направлении.)

Пользуясь этими свойствами, можно получить следующее выражение для $\text{vol}(a_1, a_2, a_3)$ через координаты векторов a_1, a_2, a_3 в положительно ориентированном ортонормированном базисе (проделайте это!):

$$\begin{aligned} \text{vol}(a_1, a_2, a_3) &= a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - \\ &\quad - a_{11}a_{23}a_{32} - a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33}. \end{aligned}$$

Выражение, стоящее в правой части этого равенства, называется определителем матрицы $A = (a_{ij})$ порядка 3. Таким образом, векторы a_1, a_2, a_3 линейно независимы тогда и только тогда, когда определитель матрицы, составленной из их координат, отличен от нуля.

Определитель матрицы $A = (a_{ij})$ порядка 3 представляет собой алгебраическую сумму всевозможных произведений трех элементов матрицы, взятых по одному из каждой строки и из каждого столбца. На рис. 2 схематически изображено, какие из этих произведений берутся со знаком плюс и какие — со знаком минус.

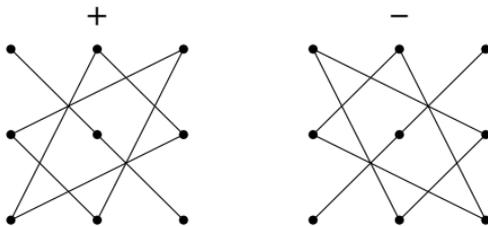


Рис. 2

Определитель матрицы A обозначается либо через $\det A$, либо путем замены круглых скобок, заключающих в себе матрицу, вертикальными чертами.

Пример 1.

$$\begin{vmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{vmatrix} = \cos^2 \alpha + \sin^2 \alpha = 1.$$

Пример 2.

$$\begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{vmatrix} = 1 \cdot 5 \cdot 9 + 2 \cdot 6 \cdot 7 + 3 \cdot 4 \cdot 8 - 3 \cdot 5 \cdot 7 - 2 \cdot 4 \cdot 9 - 1 \cdot 6 \cdot 8 = \\ = 45 + 84 + 96 - 105 - 72 - 48 = 0.$$

В случае произвольной размерности и произвольного поля, когда мы не располагаем такими понятиями, как площадь или объем, естественно попытаться ввести определитель как функцию, обладающую свойствами, аналогичными свойствам 1)–3). Дадим необходимые для этого определения.

Пусть V — векторное пространство над произвольным полем K и $f(a_1, a_2, \dots, a_m)$ — функция от m векторов пространства V , принимающая значения в K .

Определение 1. Функция $f(a_1, a_2, \dots, a_m)$ называется *полилинейной* (или, точнее, *m -линейной*), если она линейна по каждому аргументу.

Например, линейность по первому аргументу означает, что

$$\begin{aligned} f(a'_1 + a''_1, a_2, \dots, a_m) &= f(a'_1, a_2, \dots, a_m) + f(a''_1, a_2, \dots, a_m), \\ f(\lambda a_1, a_2, \dots, a_m) &= \lambda f(a_1, a_2, \dots, a_m). \end{aligned}$$

Определение 2. Полилинейная функция $f(a_1, a_2, \dots, a_m)$ называется *кососимметрической*, если при перестановке любых двух аргументов она умножается на -1 .

Важное свойство кососимметрической полилинейной функции состоит в том, что, если только $\text{char } K \neq 2$, она обращается в нуль всякий раз, когда какие-либо два аргумента принимают одинаковые значения. В самом деле, при перестановке этих двух аргументов значение функции не изменится, но, с другой стороны, оно должно умножиться на -1 ; следовательно, оно равно нулю.

Замечание 1. Если $\text{char } K = 2$, то последнее свойство следует принять за определение кососимметричности. Докажем, что из него, наоборот, вытекает кососимметричность в определенном выше смысле. Поскольку при проверке кососимметричности по каким-либо двум аргументам значения остальных аргументов следует считать фиксированными (хотя и любыми), достаточно рассмотреть случай билинейной (т. е. 2-линейной) функции. Пусть f — билинейная функция, обращающаяся в нуль при одинаковых значениях аргументов. Тогда для любых $a, b \in V$ имеем

$$0 = f(a + b, a + b) = f(a, a) + f(a, b) + f(b, a) + f(b, b) = f(a, b) + f(b, a),$$

откуда $f(b, a) = -f(a, b)$.

Теперь введем понятия, необходимые для описания явного аналитического выражения определителя матрицы порядка n , подобно тем, которые были получены при $n = 2$ и 3 .

Последовательность (k_1, k_2, \dots, k_n) чисел $1, 2, \dots, n$, расположенных в каком-либо порядке, называется *перестановкой* из n элементов. Так как k_1 может принимать n различных значений, k_2 при заданном k_1 может принимать $n - 1$ значений, k_3 при заданных k_1 и k_2 может принимать $n - 2$ значений и т. д., то имеется всего

$$n(n - 1)(n - 2) \cdot \dots \cdot 2 \cdot 1 = n!$$

перестановок из n элементов. Перестановка $(1, 2, \dots, n)$ называется *тривиальной*.

Замечание 2. Слово «перестановка» в математической литературе (в частности, в этой книге) иногда употребляется в общечело-

веческом смысле как изменение порядка каких-либо объектов (например, перестановка слов в предложении).

Говорят, что пара чисел образует *инверсию* в заданной перестановке, если большее из них стоит левее меньшего. Перестановка называется *четной* (соответственно *нечетной*), если число инверсий в ней четно (соответственно нечетно). Наряду с этим определяется *знак* перестановки, равный 1, если перестановка четна, и -1 , если она нечетна. Знак перестановки (k_1, k_2, \dots, k_n) обозначается через $\text{sgn}(k_1, k_2, \dots, k_n)$.

Пример 3. При $n = 3$ четные перестановки — это $(1, 2, 3)$ (нет инверсий), $(2, 3, 1)$ (две инверсии) и $(3, 1, 2)$ (две инверсии), нечетные — $(1, 3, 2)$ (1 инверсия), $(3, 2, 1)$ (3 инверсии) и $(2, 1, 3)$ (1 инверсия).

Пример 4. Тривиальная перестановка не имеет инверсий и поэтому четна. Напротив, в перестановке $(n, n - 1, \dots, 2, 1)$ любая пара чисел образует инверсию. Поэтому число инверсий в этой перестановке равно

$$C_n^2 = \frac{n(n-1)}{2} \equiv \left[\frac{n}{2} \right] \pmod{2}.$$

Следовательно,

$$\text{sgn}(n, n - 1, \dots, 2, 1) = (-1)^{n(n-1)/2} = (-1)^{\lfloor n/2 \rfloor}.$$

Перемена местами двух элементов в перестановке называется *транспозицией* этих элементов.

Предложение 1. При любой транспозиции четность перестановки меняется.

Доказательство. При транспозиции соседних элементов меняется взаимное расположение только этих элементов, так что число инверсий изменяется (увеличивается или уменьшается) на 1; следовательно, четность меняется. Транспозиция элементов i и j , разделенных s другими элементами, может быть осуществлена путем $2s + 1$ последовательных транспозиций соседних элементов: сначала переставляем i со всеми промежуточными элементами и с j , затем переставляем j со всеми промежуточными элементами. Каждый раз знак перестановки будет меняться по доказанному выше. Так как это произойдет нечетное число раз, то в результате знак перестановки изменится на противоположный. \square

Следствие. При $n > 1$ число четных перестановок из n элементов равно числу нечетных.

Доказательство. Выпишем все четные перестановки и в каждой из них произведем транспозицию первых двух элементов. Тогда мы получим, причем по одному разу, все нечетные перестановки. \square

Теперь мы в состоянии дать определение определителя квадратной матрицы любого порядка.

Определение 3. *Определителем* квадратной матрицы $A = (a_{ij})$ порядка n называется число

$$\det A = \sum_{(k_1, k_2, \dots, k_n)} \operatorname{sgn}(k_1, k_2, \dots, k_n) a_{1k_1} a_{2k_2} \dots a_{nk_n}. \quad (12)$$

При $n = 2$ и 3 мы получаем выражения, приведенные в начале этого параграфа.

Теорема 1. 1) *Определитель является кососимметрической полилинейной функцией строк матрицы.*

2) *Всякая функция f на множестве квадратных матриц порядка n , являющаяся кососимметрической полилинейной функцией строк матрицы, имеет вид*

$$f(A) = f(E) \det A.$$

В частности, если $f(E) = 1$, то $f = \det$ (т. е. $f(A) = \det A$ для любой матрицы A).

Доказательство. Будем обозначать через a_1, a_2, \dots, a_n строки матрицы A . Если f — какая-либо функция на множестве матриц, то, рассматривая ее как функцию строк матрицы, будем писать

$$f(A) = f(a_1, a_2, \dots, a_n).$$

1) Линейность определителя по каждой из строк матрицы вытекает из того, что для любого i его можно представить в виде

$$\det A = \sum_j a_{ij} u_j,$$

где u_1, u_2, \dots, u_n не зависят от элементов i -й строки матрицы.

Для проверки кососимметричности посмотрим, что происходит при перестановке i -й и j -й строк матрицы. Разобьем множество всех перестановок на пары перестановок, получаемых друг из друга транспозицией k_i и k_j . Согласно предложению 1, произведения $a_{1k_1} a_{2k_2} \dots a_{nk_n}$, соответствующие перестановкам из одной такой пары, входят в выражение (12) с противоположными знаками. При перестановке i -й и j -й строк они меняются ролями и, следовательно, все выражение умножается на -1 .

Замечание 3. Если $\text{char } K = 2$, то кососимметричность следует понимать в смысле замечания 1. Ее доказательство в этом случае состоит в том, что в случае равенства i -й и j -й строк матрицы A члены выражения (12), соответствующие перестановкам каждой из описанных выше пар, взаимно уничтожаются.

2) Предположим, что f — полилинейная кососимметрическая функция строк матрицы. Пусть e_1, e_2, \dots, e_n — единичные строки. Тогда

$$\begin{aligned} f(A) &= f(a_1, a_2, \dots, a_n) = f\left(\sum_{k_1} a_{1k_1} e_{k_1}, \sum_{k_2} a_{2k_2} e_{k_2}, \dots, \sum_{k_n} a_{nk_n} e_{k_n}\right) = \\ &= \sum_{k_1, k_2, \dots, k_n} a_{1k_1} a_{2k_2} \dots a_{nk_n} f(e_{k_1}, e_{k_2}, \dots, e_{k_n}). \end{aligned}$$

В силу кососимметричности функции f , если какие-то два из чисел k_1, k_2, \dots, k_n равны, то $f(e_{k_1}, e_{k_2}, \dots, e_{k_n}) = 0$. Если все они различны, то

$$f(e_{k_1}, e_{k_2}, \dots, e_{k_n}) = \text{sgn}(k_1, k_2, \dots, k_n) f(e_1, \dots, e_n).$$

В самом деле, если это равенство верно для какой-то перестановки (k_1, k_2, \dots, k_n) , то оно верно и для любой перестановки, получаемой из нее транспозицией, так как при транспозиции обе части равенства умножаются на -1 . Очевидно, что оно верно для тривиальной перестановки. Так как любую перестановку можно получить из тривиальной последовательными транспозициями, то доказываемое равенство верно для любой перестановки. Учитывая, что $f(e_1, e_2, \dots, e_n) = f(E)$, мы получаем доказываемое утверждение. \square

При $n \geq 4$ вычисление определителя непосредственно по формуле (12) в общем случае весьма затруднительно. Существуют значительно более простые способы вычисления определителей. Они основаны на свойствах определителей, доказываемых ниже.

Предложение 2. *Определитель матрицы не изменяется при элементарном преобразовании строк первого типа.*

Доказательство. Пусть, скажем, к 1-й строке матрицы A прибавляется 2-я строка, умноженная на c . Полученную матрицу обозначим через A' . Имеем (в обозначениях доказательства теоремы 1):

$$\begin{aligned} \det A' &= \det(a_1 + ca_2, a_2, \dots, a_n) = \\ &= \det(a_1, a_2, \dots, a_n) + c \det(a_2, a_2, \dots, a_n) = \det A. \end{aligned} \quad \square$$

При перестановке двух строк определитель, как мы знаем, умножается на -1 , а при умножении какой-либо строки на число он умножается на это число. Таким образом мы можем проследить за изменением определителя при любых элементарных преобразованиях строк матрицы. Так как любую матрицу с помощью элементарных преобразований строк можно привести к ступенчатому виду, а всякая ступенчатая квадратная матрица является треугольной (но, может быть, не строго треугольной), то нам остается научиться вычислять определитель треугольной матрицы.

Предложение 3. *Определитель треугольной матрицы равен произведению ее диагональных элементов.*

Доказательство. Произведение диагональных элементов входит в выражение (12) определителя любой матрицы со знаком плюс, так как соответствует тривиальной перестановке. В случае треугольной матрицы все остальные члены этого выражения равны нулю. В самом деле, если $a_{1k_1} a_{2k_2} \dots a_{nk_n} \neq 0$, то

$$k_1 \geq 1, \quad k_2 \geq 2, \quad \dots, \quad k_n \geq n;$$

но так как

$$k_1 + k_2 + \dots + k_n = 1 + 2 + \dots + n,$$

то это возможно только при

$$k_1 = 1, \quad k_2 = 2, \quad \dots, \quad k_n = n. \quad \square$$

Помимо того, что они дают практический способ вычисления определителей, предложения 2 и 3 позволяют нам ответить на вопрос, ради которого мы и ввели понятие определителя.

Теорема 2. *Квадратная матрица A невырождена тогда и только тогда, когда $\det A \neq 0$.*

Доказательство. С помощью элементарных преобразований строк приведем матрицу A к ступенчатому виду. Если при этом использовались элементарные преобразования второго или третьего типов, то определитель может измениться, но, во всяком случае, его равенство нулю или отличие от нуля сохранится. Матрица A невырождена тогда и только тогда, когда полученная ступенчатая матрица является строго треугольной; но это равносильно тому, что ее определитель отличен от нуля. \square

Продолжим изучение свойств определителей.

Лемма 1. Пусть (i_1, i_2, \dots, i_n) и (j_1, j_2, \dots, j_n) — произвольные перестановки. Произведение $a_{i_1 j_1} a_{i_2 j_2} \dots a_{i_n j_n}$ входит в выражение (12) определителя матрицы A со знаком $\operatorname{sgn}(i_1, i_2, \dots, i_n) \operatorname{sgn}(j_1, j_2, \dots, j_n)$.

Доказательство. Для того чтобы выяснить, с каким знаком входит в $\det A$ рассматриваемое произведение, нужно расположить его сомножители по порядку номеров строк. Этого можно достичь, последовательно меняя местами два сомножителя. При каждой такой перемене в перестановках, образуемых номерами строк и столбцов, одновременно происходят транспозиции, так что произведение их знаков не меняется. Таким образом, если полученное в результате произведение будет иметь вид $a_{1k_1} a_{2k_2} \dots a_{nk_n}$, то

$$\operatorname{sgn}(k_1, k_2, \dots, k_n) = \operatorname{sgn}(i_1, i_2, \dots, i_n) \operatorname{sgn}(j_1, j_2, \dots, j_n),$$

а это и означает, что рассматриваемое произведение входит в $\det A$ с указанным знаком. \square

Теорема 3. $\det A^T = \det A$.

Доказательство. Определитель матрицы A^T , как и определитель матрицы A , есть алгебраическая сумма всевозможных произведений n элементов матрицы A , взятых по одному из каждой строки и из каждого столбца. Единственное, за чем надо проследить, — это то, что одинаковые произведения входят в $\det A$ и $\det A^T$ с одинаковыми знаками. При переходе от матрицы A к матрице A^T у каждого элемента номера строки и столбца меняются местами, и, соответственно, у каждого произведения, входящего в определитель, меняются местами перестановки, составленные из номеров строк и столбцов. Предыдущая лемма показывает, что при этом знак, с которым входит данный член в определитель, не меняется. \square

Из этой теоремы следует, что всякое свойство определителей остается справедливым, если заменить в нем строки столбцами, а столбцы — строками. В частности, мы таким образом получаем

Следствие. Определитель есть кососимметрическая полилинейная функция столбцов матрицы.

Теорема 4 (об определителе матрицы с углом нулей). Пусть матрица A имеет вид

$$A = \begin{pmatrix} B & D \\ 0 & C \end{pmatrix},$$

где B и C — квадратные матрицы. Тогда

$$\det A = \det B \cdot \det C.$$

Доказательство. При фиксированных B и D определитель матрицы A является кососимметрической полилинейной функцией ее последних строк и, тем самым, кососимметрической полилинейной функцией строк матрицы C . Согласно теореме 1, получаем отсюда

$$\det A = \det \begin{pmatrix} B & D \\ 0 & E \end{pmatrix} \cdot \det C.$$

Первый множитель, в свою очередь, при фиксированной матрице D является кососимметрической полилинейной функцией столбцов матрицы B , откуда

$$\det \begin{pmatrix} B & D \\ 0 & E \end{pmatrix} = \det \begin{pmatrix} E & D \\ 0 & E \end{pmatrix} \cdot \det B = \det B$$

(поскольку матрица $\begin{pmatrix} E & D \\ 0 & E \end{pmatrix}$ треугольная с единицами на диагонали). \square

Ввиду теоремы 3 аналогичная формула верна и для матриц с правым верхним углом нулей.

Пример 5. Вычислим так называемый *определитель Вандермонда*

$$V(x_1, x_2, \dots, x_n) = \begin{vmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \dots & x_2^{n-1} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_n & x_n^2 & \dots & x_n^{n-1} \end{vmatrix}.$$

Вычитая из каждого столбца, начиная с последнего, предыдущий столбец, умноженный на x_1 , и применяя теорему 4, получаем

$$\begin{aligned} V(x_1, x_2, \dots, x_n) &= \begin{vmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & x_2 - x_1 & x_2(x_2 - x_1) & \dots & x_2^{n-2}(x_2 - x_1) \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_n - x_1 & x_n(x_n - x_1) & \dots & x_n^{n-2}(x_n - x_1) \end{vmatrix} = \\ &= (x_2 - x_1) \dots (x_n - x_1) V(x_2, \dots, x_n). \end{aligned}$$

Продолжая так дальше, в конце концов получаем

$$V(x_1, x_2, \dots, x_n) = \prod_{i>j} (x_i - x_j). \quad (13)$$

Пусть A — произвольная (не обязательно квадратная) матрица. Всякая матрица, составленная из элементов матрицы A , находящихся на пересечении каких-то выбранных строк и каких-то выбранных

столбцов, называется *подматрицей* матрицы A . Подчеркнем, что выбираемые строки и столбцы не обязаны идти подряд.

Определитель квадратной подматрицы порядка k называется *минором* порядка k матрицы A . Иногда, допуская вольность речи, саму квадратную подматрицу также называют минором. В частности, если A — квадратная матрица порядка n , то минор порядка $n - 1$, получаемый вычеркиванием i -й строки и j -го столбца, называется *дополнительным минором* элемента a_{ij} и обозначается через M_{ij} .
Число

$$A_{ij} = (-1)^{i+j} M_{ij}$$

называется *алгебраическим дополнением* элемента a_{ij} . Смысл алгебраического дополнения ясен из следующей леммы.

Лемма 2.

$$\begin{vmatrix} a_{11} & \dots & a_{1j} & \dots & a_{1n} \\ \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & a_{ij} & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ a_{n1} & \dots & a_{nj} & \dots & a_{nn} \end{vmatrix} = a_{ij} A_{ij}.$$

(В левой части стоит определитель матрицы, полученной из матрицы $A = (a_{ij})$ заменой нулями всех элементов i -й строки, кроме a_{ij} .)

Доказательство. Поменяем местами i -ю строку со всеми предыдущими строками и j -й столбец со всеми предыдущими столбцами. При этом мы будем $i - 1$ раз менять местами строки и $j - 1$ раз столбцы, так что определитель умножится на

$$(-1)^{i-1+j-1} = (-1)^{i+j}.$$

В результате получится определитель вида

$$\begin{vmatrix} a_{ij} & 0 & \dots & 0 \\ a_{1j} & a_{11} & \dots & a_{1n} \\ \dots & \dots & \dots & \dots \\ a_{nj} & a_{n1} & \dots & a_{nn} \end{vmatrix},$$

где в правом нижнем углу стоит дополнительный минор элемента a_{ij} . По теореме об определителе матрицы с углом нулей этот определитель равен $a_{ij} M_{ij}$. С учетом предыдущего знака отсюда и получается доказываемое равенство. \square

Теорема 5. Для любой квадратной матрицы A

$$\det A = \sum_j a_{ij} A_{ij} = \sum_i a_{ij} A_{ij}.$$

Первая из этих формул называется *формулой разложения определителя по i -й строке*, вторая — *формулой разложения определителя по j -му столбцу*.

Доказательство. Так как каждый член выражения (12) для $\det A$ содержит ровно один элемент из i -й строки, то предыдущая лемма означает, что сумма тех членов, которые содержат a_{ij} , равна $a_{ij}A_{ij}$. Отсюда вытекает формула разложения по строке. Аналогично доказывается формула разложения по столбцу. \square

Замечание 4. Знаки $(-1)^{i+j}$ чередуются в матрице в шахматном порядке, причем на главной диагонали стоят плюсы.

Пример 6. Вычисление определителя Δ из примера 2 разложением по 2-й строке дает

$$\Delta = -4 \begin{vmatrix} 2 & 3 \\ 8 & 9 \end{vmatrix} + 5 \begin{vmatrix} 1 & 3 \\ 7 & 9 \end{vmatrix} - 6 \begin{vmatrix} 1 & 2 \\ 7 & 8 \end{vmatrix} = -4 \cdot (-6) + 5 \cdot (-12) - 6 \cdot (-6) = 0.$$

Пример 7. Вычислим определитель порядка n вида

$$\Delta_n = \begin{vmatrix} 2 & 1 & 0 & \dots & 0 & 0 \\ 1 & 2 & 1 & \dots & 0 & 0 \\ 0 & 1 & 2 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 2 & 1 \\ 0 & 0 & 0 & \dots & 1 & 2 \end{vmatrix}.$$

Разлагая его по 1-й строке и затем второй из полученных определителей по 1-му столбцу, получаем

$$\Delta_n = 2\Delta_{n-1} - \begin{vmatrix} 1 & 1 & \dots & 0 & 0 \\ 0 & 2 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 2 & 1 \\ 0 & 0 & \dots & 1 & 2 \end{vmatrix} = 2\Delta_{n-1} - \Delta_{n-2},$$

откуда

$$\Delta_n - \Delta_{n-1} = \Delta_{n-1} - \Delta_{n-2}.$$

Это означает, что последовательность $(\Delta_1, \Delta_2, \Delta_3, \dots)$ есть арифметическая прогрессия. Так как $\Delta_1 = 2$, $\Delta_2 = 3$, то ее разность равна 1 и

$$\Delta_n = n + 1.$$

Теорема 6. Для любых квадратных матриц A и B (одного порядка)

$$\det AB = \det A \cdot \det B.$$

Доказательство. Легко видеть, что строки c_1, \dots, c_n матрицы AB получаются из строк a_1, \dots, a_n матрицы A умножением на B :

$$c_i = a_i B \quad (i = 1, \dots, n).$$

Отсюда следует, что при фиксированной матрице B определитель $\det AB$ есть кососимметрическая полилинейная функция строк матрицы A . В самом деле, пусть, например, $a_1 = a'_1 + a''_1$, где a'_1, a''_1 — какие-то строки; тогда

$$\begin{aligned} \det(a_1 B, a_2 B, \dots, a_n B) &= \det((a'_1 + a''_1) B, a_2 B, \dots, a_n B) = \\ &= \det(a'_1 B + a''_1 B, a_2 B, \dots, a_n B) = \\ &= \det(a'_1 B, a_2 B, \dots, a_n B) + \det(a''_1 B, a_2 B, \dots, a_n B). \end{aligned}$$

Остальные свойства проверяются аналогично. После этого, применяя следствие теоремы 1, получаем:

$$\det AB = \det EB \cdot \det A = \det A \cdot \det B. \quad \square$$

Пример 8. Выразим неориентированный объем V параллелепипеда, натянутого на векторы $a_1, a_2, a_3 \in E^3$, через длины $|a_1|, |a_2|, |a_3|$ его ребер и плоские углы (см. рис. 3)

$$\alpha_1 = \widehat{a_2 a_3}, \quad \alpha_2 = \widehat{a_3 a_1}, \quad \alpha_3 = \widehat{a_1 a_2}.$$

Пусть $A = (a_{ij})$ — матрица, составленная из координат векторов a_i в ортонормированном базисе. Мы знаем (см. начало параграфа), что $V = \pm \det A$. Поэтому

$$V^2 = (\det A)^2 = \det A \cdot \det A^T = \det AA^T.$$

Из правила умножения матриц следует, что (i, j) -й элемент матрицы AA^T есть скалярное произведение

$$(a_i, a_j) = |a_i| |a_j| \cos \widehat{a_i a_j}.$$

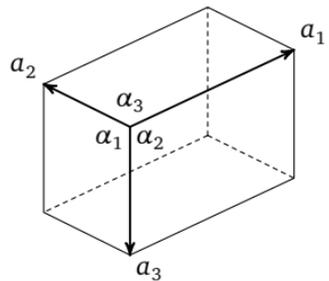


Рис. 3

Обозначим через A ее матрицу коэффициентов и через A_i ($i = 1, 2, \dots, n$) матрицу, полученную из A заменой ее i -го столбца столбцом свободных членов.

Теорема 1. Если $\det A \neq 0$, то система (14) имеет единственное решение, которое может быть найдено по формулам

$$x_i = \frac{\det A_i}{\det A} \quad (i = 1, 2, \dots, n).$$

Эти формулы называются *формулами Крамера*.

Доказательство. При любом элементарном преобразовании системы (14) в матрицах A и A_i ($i = 1, 2, \dots, n$) одновременно происходит соответствующее элементарное преобразование строк и, следовательно, отношения, стоящие в правых частях формул Крамера, не изменяются. С помощью элементарных преобразований строк матрицу A можно привести к единичной матрице. Поэтому достаточно доказать теорему в том случае, когда $A = E$.

Если $A = E$, то система имеет вид

$$\begin{cases} x_1 & & & = b_1, \\ & x_2 & & = b_2, \\ & & \ddots & \vdots \\ & & & x_n = b_n. \end{cases}$$

Она, очевидно, имеет единственное решение $x_i = b_i$ ($i = 1, 2, \dots, n$). С другой стороны,

$$\det A = \det E = 1, \quad \det A_i = \begin{vmatrix} 1 & 0 & \dots & b_1 & \dots & 0 & 0 \\ 0 & 1 & \dots & b_2 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & b_i & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & b_{n-1} & \dots & 1 & 0 \\ 0 & 0 & \dots & b_n & \dots & 0 & 1 \end{vmatrix} = b_i,$$

так что формулы Крамера в этом случае действительно верны. \square

Если $\det A = 0$, то ступенчатый вид матрицы A не будет строго треугольным и, следовательно, система (14) либо несовместна, либо неопределенна. Опасно в этом случае пытаться как-то трактовать формулы Крамера. Они просто не применимы (ведь они доказывались в предположении, что $\det A \neq 0$), и надо действовать как-то иначе.

Задача 1. Доказать, что если $\det A = 0$, но $\det A_i \neq 0$ для какого-либо i , то система (14) несовместна.

Задача 2. Показать, что если

$$\det A = \det A_1 = \dots = \det A_n = 0,$$

то система (14) может быть как несовместной, так и неопределенной. (Привести примеры, показывающие, что обе возможности реализуются.)

Отметим, что формулы Крамера — это далеко не лучший способ для практического решения систем линейных уравнений, за исключением, быть может, случая $n = 2$. Они имеют в основном теоретическое значение. В частности, они позволяют получить следующие явные формулы для элементов обратной матрицы.

Теорема 2. Пусть $A = (a_{ij})$ — невырожденная квадратная матрица. Тогда

$$A^{-1} = \frac{1}{\det A} \begin{pmatrix} A_{11} & A_{21} & \dots & A_{n1} \\ A_{12} & A_{22} & \dots & A_{n2} \\ \dots & \dots & \dots & \dots \\ A_{1n} & A_{2n} & \dots & A_{nn} \end{pmatrix}.$$

(Через A_{ij} обозначается алгебраическое дополнение к элементу a_{ij} ; см. § 4.)

Доказательство. Матрица A^{-1} является решением матричного уравнения

$$AX = E.$$

Это уравнение рассыпается на n уравнений относительно столбцов X_1, X_2, \dots, X_n матрицы X :

$$AX_j = E_j, \quad (15)$$

где E_j — j -й столбец матрицы E .

В координатной записи уравнение (15) представляет собой систему n линейных уравнений относительно элементов $x_{1j}, x_{2j}, \dots, x_{nj}$ столбца X_j . Матрицей коэффициентов этой системы служит матрица A , а столбцом свободных членов — столбец E_j . По формулам Крамера

$$x_{ij} = \frac{1}{\det A} \begin{vmatrix} a_{11} & \dots & 0 & \dots & a_{1n} \\ \dots & \dots & \dots & \dots & \dots \\ a_{j1} & \dots & 1 & \dots & a_{jn} \\ \dots & \dots & \dots & \dots & \dots \\ a_{n1} & \dots & 0 & \dots & a_{nn} \end{vmatrix} = \frac{A_{ji}}{\det A},$$

что и требовалось доказать. □

Пример 1. Для невырожденной матрицы порядка 2

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

получаем

$$A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Эту простую формулу имеет смысл запомнить.

Задача 3. Пусть A — невырожденная целочисленная (т. е. состоящая из целых чисел) квадратная матрица. Доказать, что матрица A^{-1} является целочисленной тогда и только тогда, когда $\det A = \pm 1$.

Наконец, нахождение ранга любой матрицы также может быть сведено к вычислению определителей.

Теорема 3 (о ранге матрицы). *Ранг матрицы равен наибольшему порядку ее миноров, отличных от нуля.*

Доказательство. Пусть ранг матрицы A равен r , и пусть $s > r$. Тогда любые s строк матрицы A линейно зависимы и, тем более, линейно зависимы строки любой квадратной подматрицы порядка s , представляющие собой части соответствующих строк матрицы A . Следовательно, любой минор порядка s равен нулю. Далее, рассмотрим подматрицу, образованную какими-либо r линейно независимыми строками матрицы A . Ее ранг, очевидно, также равен r и, значит, среди ее столбцов найдется r линейно независимых. Минор порядка r , образованный этими столбцами, не равен нулю. \square

Задача 4. Доказать теорему о ранге матрицы в следующей более сильной форме: если в матрице A имеется минор порядка r , отличный от нуля, а все миноры порядка $r + 1$, получаемые приписыванием к нему одной строки и одного столбца (так называемые *окаймляющие миноры*), равны нулю, то $\text{rk } A = r$.

Задача 5. Доказать, что в матрице ранга r любой минор порядка r , образуемый пересечением r линейно независимых строк с r линейно независимыми столбцами, отличен от нуля.

Начала алгебры многочленов

§1. Построение и основные свойства алгебры многочленов

Функция вещественной переменной x называется *многочленом*, если она может быть представлена в виде

$$f(x) = a_0 + a_1x + a_2x^2 + \dots + a_nx^n,$$

где $a_0, a_1, a_2, \dots, a_n$ — какие-то вещественные числа (некоторые из них или даже все могут равняться нулю). Можно доказать, и мы это сделаем ниже в более общей ситуации, что такое представление единственно с точностью до приписывания членов с нулевыми коэффициентами, т. е. если

$$a_0 + a_1x + a_2x^2 + \dots + a_nx^n = b_0 + b_1x + b_2x^2 + \dots + b_nx^n \quad \forall x \in \mathbb{R},$$

то $a_k = b_k$ при $k = 0, 1, 2, \dots, n$.

Очевидно, что сумма и произведение многочленов, а также произведение многочлена на любое число, также являются многочленами. Это означает, что многочлены образуют подалгебру в алгебре всех функций вещественной переменной (см. пример 1.7.3). Эта подалгебра называется *алгеброй многочленов над \mathbb{R}* и обозначается $\mathbb{R}[x]$.

Из предыдущего следует, что многочлены $1, x, x^2, \dots$ образуют базис алгебры $\mathbb{R}[x]$. Таблица умножения для этого базиса выглядит весьма просто:

$$x^k x^l = x^{k+l}.$$

Если попытаться аналогичным образом трактовать многочлены над любым полем K , то возникает трудность, состоящая в том, что формально различные многочлены могут быть тождественно равны при всех значениях переменной. Например, многочлены x и x^2 над полем \mathbb{Z}_2 оба принимают значение 0 при $x = 0$ и 1 при $x = 1$. В то же время хотелось бы рассматривать их как разные многочлены.

Выход состоит в формальном определении, при котором многочлен фактически отождествляется с последовательностью его коэффициентов.

Рассмотрим векторное пространство K^∞ финитных последовательностей элементов поля K (см. пример 2.2.9). Условимся нумеровать члены последовательностей, начиная с нуля, и пусть e_k ($k = 0, 1, 2, \dots$) обозначает последовательность, k -й член которой равен 1, а все остальные члены равны 0. Последовательности e_0, e_1, e_2, \dots образуют базис пространства K^∞ .

Превратим пространство K^∞ в алгебру, определив умножение базисных векторов по правилу

$$e_k e_l = e_{k+l}.$$

Из коммутативности и ассоциативности сложения целых чисел следует, что умножение базисных векторов, а значит, и любых элементов полученной алгебры, коммутативно и ассоциативно. Элемент e_0 является ее единицей. Эта алгебра называется *алгеброй многочленов* над K и обозначается $K[x]$ (вместо x может использоваться любая другая буква).

Для того чтобы перейти к привычному представлению многочленов, условимся, во-первых, отождествлять элементы вида $a e_0$ ($a \in K$) алгебры $K[x]$ с соответствующими элементами поля K и, во-вторых, элемент e_1 обозначим через x (здесь проявляется роль выбранной буквы x). Тогда в соответствии с определением операций в $K[x]$ мы получаем, что $e_k = x^k$ и

$$\begin{aligned} (a_0, a_1, a_2, \dots, a_n, 0, \dots) &= a_0 e_0 + a_1 e_1 + a_2 e_2 + \dots + a_n e_n = \\ &= a_0 + a_1 x + a_2 x^2 + \dots + a_n x^n. \end{aligned}$$

Числа a_0, a_1, a_2, \dots называются *коэффициентами* многочлена. Последний из ненулевых коэффициентов называется *старшим коэффициентом*, а его номер — *степенью* многочлена. Степень многочлена f обозначается через $\deg f$. Степень нулевого многочлена не определена, однако иногда удобно считать, что она равна $-\infty$.

Легко видеть, что

$$\deg(f + g) \leq \max\{\deg f, \deg g\}, \quad (1)$$

$$\deg fg = \deg f + \deg g. \quad (2)$$

Докажем, например, последнее равенство. Пусть

$$\begin{aligned} f &= a_0 + a_1x + \dots + a_nx^n \quad (a_n \neq 0), \\ g &= b_0 + b_1x + \dots + b_mx^m \quad (b_m \neq 0). \end{aligned}$$

Тогда при перемножении f и g получается только один член степени $n + m$, а именно, $a_nb_mx^{n+m}$, а членов большей степени не получается вообще. Так как в поле нет делителей нуля, то $a_nb_m \neq 0$ и, стало быть,

$$\deg fg = n + m = \deg f + \deg g.$$

Предыдущее рассуждение показывает, что в алгебре $K[x]$ нет делителей нуля. Из него же следует, что обратимыми элементами в этой алгебре являются только многочлены нулевой степени, т. е. ненулевые элементы поля K .

Замечание 1. Многочлен можно обозначать $f(x)$ или просто f , если из контекста ясно, какой буквой обозначается «переменная».

Замечание 2. Часто бывает удобнее располагать многочлен не по возрастающим, а по убывающим степеням x :

$$f = a_0x^n + a_1x^{n-1} + \dots + a_{n-1}x + a_n.$$

Замечание 3. В качестве K можно взять любое коммутативное ассоциативное кольцо с единицей (ср. замечание 1.8.1). В этом случае все предыдущее остается без изменений, за исключением последней части, связанной с формулой (2), где нужно дополнительно потребовать, чтобы в кольце K не было делителей нуля.

Замечание 4. Произведение финитных последовательностей (a_0, a_1, a_2, \dots) и (b_0, b_1, b_2, \dots) в кольце $K[x]$ есть последовательность (c_0, c_1, c_2, \dots) , члены которой находятся по формулам

$$c_k = \sum_{l=0}^k a_l b_{k-l}.$$

Эти формулы имеют смысл и для любых (не обязательно финитных) последовательностей. Таким образом получается коммутативная ассоциативная алгебра с единицей, называемая *алгеброй формальных степенных рядов* над K и обозначаемая $K[[x]]$. Ее элементы обычно записывают как формальные бесконечные суммы вида

$$a_0 + a_1x + a_2x^2 + \dots$$

Алгебра $K[[x]]$, как и $K[x]$, не имеет делителей нуля, но доказывается это по-другому (попробуйте это сделать!).

Каждый многочлен

$$f = a_0 + a_1x + a_2x^2 + \dots + a_nx^n \quad (3)$$

определяет функцию на K со значениями в K , значение которой в точке $c \in K$ по определению равно

$$f(c) = a_0 + a_1c + a_2c^2 + \dots + a_nc^n.$$

Так как сумма и произведение многочленов, а также произведение многочлена на число приводятся к каноническому виду (3) преобразованиями, использующими только свойства операций в $K[x]$, справедливые и в поле K , то мы придем к одному и тому же результату, сделав подстановку $x = c$ до или после этих преобразований. Это означает, что

$$(f + g)(c) = f(c) + g(c), \quad (fg)(c) = f(c)g(c), \quad (\lambda f)(c) = \lambda f(c),$$

т. е. операции над многочленами приводят к таким же операциям над соответствующими функциями.

Как мы показали на примере в начале параграфа, разные многочлены могут определять одну и ту же функцию. Оказывается, однако, что такое возможно, только если поле K конечно.

Теорема 1. *Если поле K бесконечно, то разные многочлены над K определяют разные функции.*

Доказательство. Пусть многочлены $f, g \in K[x]$ определяют одну и ту же функцию. Тогда их разность $h = f - g$ определяет нулевую функцию, т. е. $h(c) = 0$ для всех $c \in K$. Предположим, что $h \neq 0$, и пусть

$$h = a_0 + a_1x + a_2x^2 + \dots + a_{n-1}x^{n-1} \quad (a_{n-1} \neq 0).$$

Возьмем различные $x_1, x_2, \dots, x_n \in K$ (здесь используется бесконечность поля K). Совокупность верных равенств

$$\begin{cases} a_0 + a_1x_1 + a_2x_1^2 + \dots + a_{n-1}x_1^{n-1} = 0, \\ a_0 + a_1x_2 + a_2x_2^2 + \dots + a_{n-1}x_2^{n-1} = 0, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ a_0 + a_1x_n + a_2x_n^2 + \dots + a_{n-1}x_n^{n-1} = 0 \end{cases}$$

будем рассматривать как (квадратную) систему однородных линейных уравнений относительно $a_0, a_1, a_2, \dots, a_{n-1}$. Определитель мат-

рицы коэффициентов этой системы есть определитель Вандермонда $V(x_1, x_2, \dots, x_n)$ (см. пример 2.4.5) и потому отличен от нуля. Следовательно, система имеет только нулевое решение, что противоречит нашему предположению. \square

Замечание 5. Даже если поле K конечно, то множество всех многочленов над K бесконечно (но счетно). Однако множество всех функций на K со значениями в K в этом случае конечно, и поэтому обязательно должны существовать разные многочлены, определяющие одну и ту же функцию. Тем не менее теорема 1 и ее доказательство остаются в силе для многочленов, степень которых меньше числа элементов поля K .

Задача 1. Так называемая *задача интерполяции* состоит в нахождении многочлена степени $< n$, принимающего в заданных (различных) точках $x_1, x_2, \dots, x_n \in K$ заданные значения $y_1, y_2, \dots, y_n \in K$. (В частности, при $n = 2$ это называется линейной интерполяцией.) Доказать, что задача интерполяции имеет единственное решение при любых x_1, x_2, \dots, x_n и y_1, y_2, \dots, y_n .

Деление одного многочлена на другой в обычном смысле слова в алгебре $K[x]$, как правило, невозможно. Однако возможно так называемое деление с остатком, похожее на деление с остатком в кольце целых чисел.

Теорема 2. Пусть $f, g \in K[x]$, причем $g \neq 0$. Тогда существуют такие многочлены q и r , что $f = qg + r$ и либо $r = 0$, либо $\deg r < \deg g$. Многочлены q и r определены этими условиями однозначно.

Нахождение таких многочленов q и r и называется *делением с остатком* многочлена f на g . При этом q называется *неполным частным*, а r — *остатком* от деления f на g . Многочлен f делится на g в алгебре $K[x]$ тогда и только тогда, когда $r = 0$.

Доказательство. 1) Докажем возможность деления с остатком. Если $\deg f < \deg g$, то можно взять $q = 0$, $r = f$. Если $\deg f \geq \deg g$, то q и r находятся обычной процедурой «деления уголком». А именно, пусть

$$\begin{aligned} f &= a_0x^n + a_1x^{n-1} + \dots + a_{n-1}x + a_n, \\ g &= b_0x^m + b_1x^{m-1} + \dots + b_{m-1}x + b_m, \end{aligned}$$

где $a_0, b_0 \neq 0$. Рассмотрим многочлен

$$f_1 = f - \frac{a_0}{b_0}x^{n-m}g.$$

Его степень меньше, чем степень многочлена f . Если $\deg f_1 < \deg g$, то мы можем взять

$$q = \frac{a_0}{b_0} x^{n-m}, \quad r = f_1.$$

В противном случае поступаем с многочленом f_1 так же, как с f . В конце концов мы получим такой многочлен

$$q = c_0 x^{n-m} + c_1 x^{n-m-1} + \dots + c_{n-m},$$

что $\deg(f - qg) < \deg g$. Это и будет неполное частное от деления f на g , а многочлен $r = f - qg$ будет остатком.

2) Докажем, что многочлены q и r определены условиями теоремы однозначно. Пусть

$$f = q_1 g + r_1 = q_2 g + r_2,$$

где $\deg r_1 < \deg g$ и $\deg r_2 < \deg g$. Тогда

$$r_1 - r_2 = (q_2 - q_1)g$$

и, если $q_1 \neq q_2$, то

$$\deg(r_1 - r_2) = \deg(q_2 - q_1) + \deg g \geq \deg g,$$

что, очевидно, неверно. Следовательно, $q_1 = q_2$ и $r_1 = r_2$. \square

Особое значение имеет деление с остатком на линейный двучлен $x - c$. В этом случае остаток имеет степень < 1 , т. е. является элементом поля K . Таким образом, результат деления с остатком многочлена f на $x - c$ имеет вид

$$f(x) = (x - c)q(x) + r \quad (r \in K).$$

Отсюда следует, что

$$f(c) = r,$$

т. е. остаток равен значению многочлена f в точке c . Это утверждение называется *теоремой Безу*.

Деление с остатком на $x - c$ осуществляется по замечательно простой схеме, называемой *схемой Горнера*.

А именно, пусть

$$\begin{aligned} a_0 x^n + a_1 x^{n-1} + \dots + a_{n-1} x + a_n = \\ = (x - c)(b_0 x^{n-1} + b_1 x^{n-2} + \dots + b_{n-2} x + b_{n-1}) + r. \end{aligned}$$

Приравнивая коэффициенты при соответствующих степенях x , получаем цепочку равенств

$$\begin{aligned} a_0 &= b_0, \\ a_1 &= b_1 - cb_0, \\ a_2 &= b_2 - cb_1, \\ &\dots\dots\dots \\ a_{n-1} &= b_{n-1} - cb_{n-2}, \\ a_n &= r - cb_{n-1}, \end{aligned}$$

откуда находим следующие рекуррентные формулы для b_0, b_1, \dots, b_{n-1} и r :

$$\begin{aligned} b_0 &= a_0, \\ b_1 &= a_1 + cb_0, \\ b_2 &= a_2 + cb_1, \\ &\dots\dots\dots \\ b_{n-1} &= a_{n-1} + cb_{n-2}, \\ r &= a_n + cb_{n-1}. \end{aligned}$$

Исходные данные и результаты вычислений удобно расположить в виде таблицы:

	a_0	a_1	a_2	\dots	a_{n-1}	a_n
c	b_0	b_1	b_2	\dots	b_{n-1}	r

Каждое число во второй строке этой таблицы, начиная с b_1 , находится как сумма числа, стоящего над ним, и числа, стоящего слева, умноженного на c .

В частности, это дает очень эффективный способ вычисления значений многочлена.

Пример 1. Найдем значение многочлена

$$f = 2x^6 - 11x^4 - 19x^3 - 7x^2 + 8x + 5$$

в точке $x = 3$. По схеме Горнера получаем:

	2	0	-11	-19	-7	8	5
3	2	6	7	2	-1	5	20

Таким образом, $f(3) = 20$.

§ 2. Общие свойства корней многочленов

Элемент c поля K называется *корнем* многочлена $f \in K[x]$ (или соответствующего алгебраического уравнения $f(x)=0$), если $f(c)=0$. Из теоремы Безу (см. предыдущий параграф) следует

Теорема 1. Элемент c поля K является корнем многочлена $f \in K[x]$ тогда и только тогда, когда f делится на $x - c$.

Этим можно воспользоваться для доказательства следующей теоремы.

Теорема 2. Число корней ненулевого многочлена не превосходит его степени.

Доказательство. Пусть c_1 — корень многочлена f . Тогда

$$f = (x - c_1)f_1 \quad (f_1 \in K[x]).$$

Пусть c_2 — корень многочлена f_1 . Тогда

$$f_1 = (x - c_2)f_2 \quad (f_2 \in K[x])$$

и, значит,

$$f = (x - c_1)(x - c_2)f_2.$$

Продолжая так дальше, мы в конце концов представим многочлен f в виде

$$f = (x - c_1)(x - c_2)\dots(x - c_m)g, \quad (4)$$

где многочлен $g \in K[x]$ не имеет корней. Числа c_1, c_2, \dots, c_m — это все корни многочлена f . В самом деле, для любого $c \in K$ имеем

$$f(c) = (c - c_1)(c - c_2)\dots(c - c_m)g(c)$$

и, так как $g(c) \neq 0$, то $f(c) = 0$, только если $c = c_i$ для некоторого i . Таким образом, число корней многочлена f не превосходит m (оно может быть меньше m , поскольку не исключено, что среди чисел c_1, c_2, \dots, c_m есть одинаковые); но

$$m = \deg f - \deg g \leq \deg f. \quad \square$$

Замечание 1. Эта теорема фактически уже была нами доказана другим способом в процессе доказательства теоремы 1.1. С другой стороны, из нее можно получить доказательство теоремы 1.1, не использующее теории линейных уравнений. А именно, если различные многочлены f и g над бесконечным полем K определяют одну

и ту же функцию, то все элементы поля K являются корнями ненулевого многочлена $h = f - g$, что противоречит только что доказанной теореме.

Доказательство предыдущей теоремы наводит на мысль, что некоторые корни правильнее было бы считать несколько раз. Этому можно придать точный смысл.

Корень c многочлена f называется *простым*, если f не делится на $(x - c)^2$, и *кратным* в противном случае. *Кратностью* корня c называется наибольшее из таких k , что f делится на $(x - c)^k$. Таким образом, простой корень — это корень кратности 1. Иногда удобно считать, что число, не являющееся корнем, — это корень кратности 0.

Очевидно, что c является корнем кратности k многочлена f тогда и только тогда, когда

$$f = (x - c)^k g, \quad (5)$$

где $g(c) \neq 0$.

Теперь мы докажем уточнение теоремы 2.

Теорема 3. Число корней ненулевого многочлена с учетом их кратностей (т. е. если каждый корень считается столько раз, какова его кратность) не превосходит степени многочлена, причем равенство имеет место тогда и только тогда, когда этот многочлен разлагается на линейные множители.

Доказательство. Перепишем равенство (4), объединив одинаковые множители:

$$f = (x - c_1)^{k_1} (x - c_2)^{k_2} \dots (x - c_s)^{k_s} g \quad (6)$$

(c_1, c_2, \dots, c_s различны). Ясно, что c_1, c_2, \dots, c_s — это все корни многочлена f . Далее, выделяя в (6) множитель $(x - c_i)^{k_i}$, мы можем написать

$$f = (x - c_i)^{k_i} h_i, \quad \text{где } h_i(c_i) \neq 0.$$

Следовательно, c_i — корень кратности k_i .

Таким образом, число корней многочлена f с учетом их кратностей равно

$$k_1 + k_2 + \dots + k_s = \deg f - \deg g,$$

откуда и вытекают все утверждения теоремы. \square

Замечание 2. Условно считается, что многочлен нулевой степени разлагается в произведение пустого множества линейных множителей.

Если многочлен $f = a_0x^n + a_1x^{n-1} + \dots + a_{n-1}x + a_n$ разлагается на линейные множители, то это разложение может быть записано в виде $f = a_0(x - c_1)(x - c_2)\dots(x - c_n)$, где c_1, c_2, \dots, c_n — корни многочлена f , причем каждый из них повторен столько раз, какова его кратность. Приравнивая коэффициенты при соответствующих степенях x в этих двух представлениях многочлена f , мы получаем следующие формулы Виета:

$$\begin{aligned} c_1 + c_2 + \dots + c_n &= -\frac{a_1}{a_0}, \\ c_1c_2 + c_1c_3 + \dots + c_{n-1}c_n &= \frac{a_2}{a_0}, \\ &\dots\dots\dots \\ \sum_{i_1 < i_2 < \dots < i_k} c_{i_1}c_{i_2}\dots c_{i_k} &= (-1)^k \frac{a_k}{a_0}, \\ &\dots\dots\dots \\ c_1c_2\dots c_n &= (-1)^n \frac{a_n}{a_0}. \end{aligned}$$

В левой части k -й формулы Виета стоит сумма всевозможных произведений k корней многочлена f . С точностью до множителя $(-1)^k$ это коэффициент при x^{n-k} в произведении $(x - c_1)(x - c_2)\dots(x - c_n)$.

Пример 1. Комплексные корни 5-й степени из 1

$$\varepsilon_k = \cos \frac{2\pi k}{5} + i \sin \frac{2\pi k}{5} \quad (k=0, 1, 2, 3, 4)$$

(рис. 1) суть корни многочлена $x^5 - 1$. По первой из формул Виета их сумма равна нулю. Приравнивая нулю сумму их вещественных

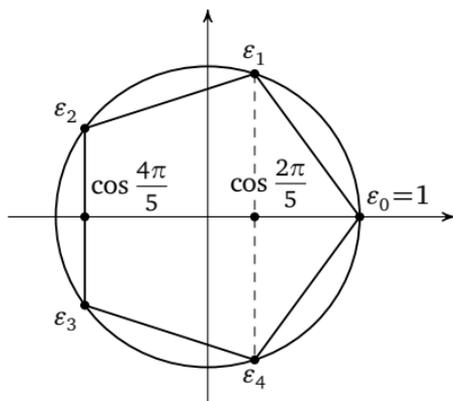


Рис. 1

частей, получаем

$$2 \cos \frac{4\pi}{5} + 2 \cos \frac{2\pi}{5} + 1 = 0.$$

Пусть $\cos \frac{2\pi}{5} = x$; тогда $\cos \frac{4\pi}{5} = 2x^2 - 1$, так что

$$4x^2 + 2x - 1 = 0,$$

откуда

$$\cos \frac{2\pi}{5} = \frac{\sqrt{5}-1}{4}, \quad \cos \frac{4\pi}{5} = -\frac{\sqrt{5}+1}{4}.$$

Задача 1. Пусть n — простое число. Пользуясь задачей 1.5.2 и последней из формул Виета, доказать *теорему Вильсона*:

$$(n-1)! \equiv -1 \pmod{n}.$$

Многочлен f называется *нормированным* (или *приведенным*), если $a_0 = 1$. Формулы Виета позволяют выразить коэффициенты нормированного многочлена через его корни (при условии, что число корней с учетом кратностей равно степени многочлена).

Пример 2. Найдем нормированный многочлен 4-й степени

$$f = x^4 + a_1x^3 + a_2x^2 + a_3x + a_4,$$

имеющий двукратный корень 1 и простые корни 2, 3. По формулам Виета

$$-a_1 = 1 + 1 + 2 + 3 = 7,$$

$$a_2 = 1 \cdot 1 + 1 \cdot 2 + 1 \cdot 3 + 1 \cdot 2 + 1 \cdot 3 + 2 \cdot 3 = 17,$$

$$-a_3 = 1 \cdot 1 \cdot 2 + 1 \cdot 1 \cdot 3 + 1 \cdot 2 \cdot 3 + 1 \cdot 2 \cdot 3 = 17,$$

$$a_4 = 1 \cdot 1 \cdot 2 \cdot 3 = 6.$$

Таким образом,

$$f = x^4 - 7x^3 + 17x^2 - 17x + 6.$$

Кратность корня многочлена может быть истолкована и другим способом, по крайней мере в случае $\text{char } K = 0$. Для этого нужно определить дифференцирование многочленов.

Из правил дифференцирования функций вещественной переменной следует, что производная многочлена есть также многочлен. Обозначим через D отображение алгебры $\mathbb{R}[x]$ в себя, ставящее в со-

ответствие каждому многочлену его производную. Отображение D обладает следующими свойствами:

- 1) оно линейно;
- 2) $D(fg) = (Df)g + f(Dg)$;
- 3) $Dx = 1$.

Эти наблюдения позволяют определить дифференцирование многочленов над любым полем K , когда определение производной, даваемое в анализе, не имеет смысла.

Предложение 1. *Отображение $D: K[x] \rightarrow K[x]$, обладающее свойствами 1)–3), существует и единственно.*

Доказательство. Пусть D — такое отображение. Тогда

$$D1 = D(1 \cdot 1) = (D1) \cdot 1 + 1 \cdot (D1) = D1 + D1,$$

откуда $D1 = 0$. Докажем по индукции, что $Dx^n = nx^{n-1}$. При $n = 1$ это верно по предположению, а переход от $n - 1$ к n делается выкладкой

$$Dx^n = D(x^{n-1}x) = (Dx^{n-1})x + x^{n-1}(Dx) = (n-1)x^{n-2} \cdot x + x^{n-1} = nx^{n-1}.$$

Тем самым отображение D однозначно определено на базисных векторах $1, x, x^2, \dots$, а значит, и на всем пространстве $K[x]$.

Обратно, построим линейное отображение $D: K[x] \rightarrow K[x]$, задав его на базисных векторах формулами

$$D1 = 0, \quad Dx^n = nx^{n-1} \quad (n = 1, 2, \dots),$$

и проверим, что оно обладает свойством 2). В силу линейности достаточно проверить это свойство для базисных векторов. Имеем

$$D(x^m x^n) = Dx^{m+n} = (m+n)x^{m+n-1},$$

$$(Dx^m)x^n + x^m(Dx^n) = mx^{m-1}x^n + nx^m x^{n-1} = (m+n)x^{m+n-1}. \quad \square$$

Многочлен Df называется *производной* многочлена f и обозначается, как обычно, через f' .

Сделав в многочлене $f \in K[x]$ замену $x = c + y$, где $c \in K$, мы можем представить его в виде многочлена (той же степени) от $y = x - c$ или, как говорят, разложить по степеням $x - c$:

$$f = b_0 + b_1(x - c) + b_2(x - c)^2 + \dots + b_n(x - c)^n. \quad (7)$$

Очевидно, что если c — корень многочлена f , то его кратность равна номеру первого отличного от нуля коэффициента этого разложения.

Предложение 2. Если $\text{char } K = 0$, то коэффициенты разложения многочлена $f \in K[x]$ по степеням $x - c$ могут быть найдены по формулам

$$b_k = \frac{f^{(k)}(c)}{k!}.$$

(Здесь $f^{(k)}$, как обычно, обозначает k -ю производную многочлена f .)

Доказательство. Продифференцируем равенство (7) k раз и подставим $x = c$. □

Таким образом,

$$f = f(c) + \frac{f'(c)}{1!}(x - c) + \frac{f''(c)}{2!}(x - c)^2 + \dots + \frac{f^{(n)}(c)}{n!}(x - c)^n.$$

Эта формула называется *формулой Тейлора* для многочленов.

Из формулы Тейлора и сделанного выше замечания следует

Теорема 4. При условии, что $\text{char } K = 0$, кратность корня c многочлена $f \in K[x]$ равна наименьшему порядку производной многочлена f , не обращающейся в нуль в точке c .

Следствие. При том же условии всякий k -кратный корень многочлена f является $(k - 1)$ -кратным корнем его производной.

Замечание 3. В случае $\text{char } K > 0$ кратность корня c может быть меньше указанного в теореме 4 числа. Более того, такого числа может вообще не существовать. Так, например, если n — простое число, то первая, а значит, и все последующие производные многочлена $x^n \in \mathbb{Z}_n[x]$, имеющего n -кратный корень 0, являются нулевыми многочленами.

В случае $K = \mathbb{R}$ теорема 4 позволяет истолковать кратность геометрически. А именно, если кратность корня c многочлена $f \in K[x]$ равна k , то вблизи точки c многочлен f ведет себя как $b(x - c)^k$ ($b \neq 0$). Это означает, что его график в точке c при $k = 1$ просто пересекает ось x , а при $k > 1$ имеет с ней касание $(k - 1)$ -го порядка. Кроме того, знак многочлена $f(x)$ при прохождении точки c при нечетном k меняется, а при четном k не меняется (см. рис. 2).

Коэффициенты разложения (7), а значит, и значения производных многочлена f в точке c (в случае $\text{char } K = 0$) могут быть найдены последовательными делениями с остатком многочлена f на $x - c$. А именно, при первом делении получается остаток b_0 и неполное частное

$$f_1 = b_1 + b_2(x - c) + \dots + b_n(x - c)^{n-1};$$

при делении f_1 на $x - c$ получается остаток b_1 и т. д.

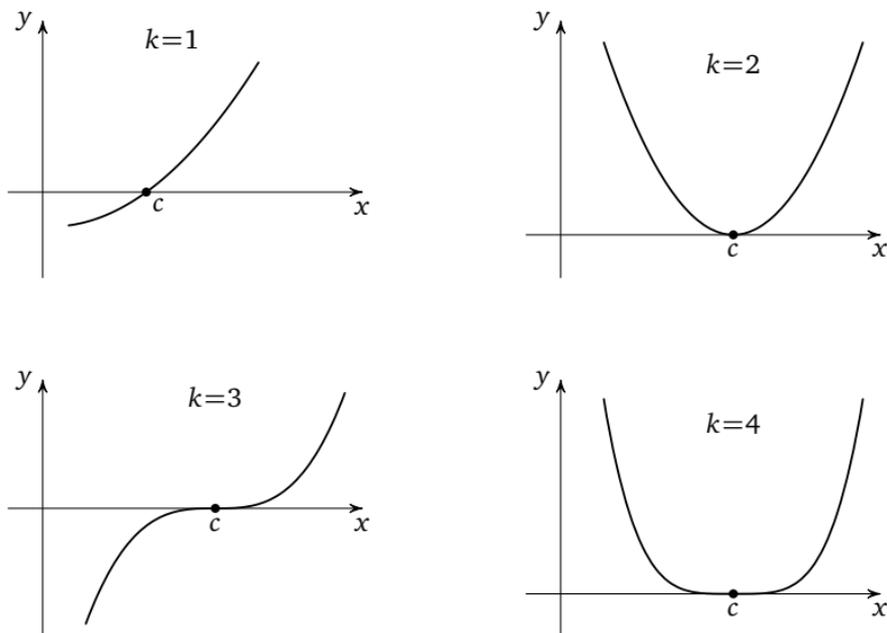


Рис. 2

Пример 3. Разложим указанным способом по степеням $x - 2$ многочлен

$$f = x^5 - 5x^4 + 7x^3 - 2x^2 + 4x - 8 \in \mathbb{R}[x].$$

Последовательные деления с остатком на $x - 2$ будем проводить по схеме Горнера, используя строку результатов каждого деления как строку исходных данных для следующего деления:

	1	-5	7	-2	4	-8
2	1	-3	1	0	4	0
	1	-1	-1	-2	0	
	1	1	1	0		
	1	3	7			
	1	5				
	1					

Таким образом,

$$f = 7(x - 2)^3 + 5(x - 2)^4 + (x - 2)^5.$$

В частности, мы видим, что 2 — трехкратный корень многочлена f . Кроме того,

$$f'''(2) = 3! \cdot 7 = 42, \quad f^{IV}(2) = 4! \cdot 5 = 120, \quad f^V(2) = 5! \cdot 1 = 120.$$

§ 3. Основная теорема алгебры комплексных чисел

Оценка сверху числа корней многочлена, полученная в предыдущем параграфе, ничего не говорит о наличии хотя бы одного корня. И действительно, существуют многочлены положительной степени, не имеющие корней, например, многочлен $x^2 + 1$ над полем \mathbb{R} вещественных чисел. Именно это обстоятельство послужило поводом для построения поля \mathbb{C} комплексных чисел. Если бы и над полем \mathbb{C} существовали многочлены положительной степени, не имеющие корней, это привело бы к необходимости его дальнейшего расширения. Однако, к счастью, это не так. Это составляет содержание теоремы, которую называют *основной теоремой алгебры комплексных чисел*.

Теорема 1. *Всякий многочлен положительной степени над полем комплексных чисел имеет корень.*

Поле, над которым всякий многочлен положительной степени имеет хотя бы один корень, называется *алгебраически замкнутым*. Таким образом, теорема 1 означает, что поле \mathbb{C} комплексных чисел алгебраически замкнуто.

Существует несколько доказательств этой теоремы. Любое из них включает элементы анализа, так как оно должно как-то использовать определение поля вещественных чисел, которое не является чисто алгебраическим. Доказательство, приводимое ниже, является почти полностью аналитическим.

Нам понадобится понятие предела последовательности комплексных чисел. Перед тем как дать соответствующее определение, напомним, что модуль $|z|$ комплексного числа z есть длина вектора, изображающего это число. Отсюда следует, что $|z_1 - z_2|$ есть расстояние между точками, изображающими числа z_1 и z_2 . Известные из геометрии неравенства показывают (см. рис. 3 и 4), что

$$|z_1 + z_2| \leq |z_1| + |z_2|, \quad ||z_1| - |z_2|| \leq |z_1 - z_2|.$$

(Равенства могут иметь место, когда соответствующий треугольник вырождается в отрезок.)

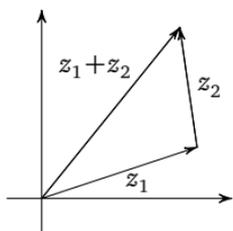


Рис. 3

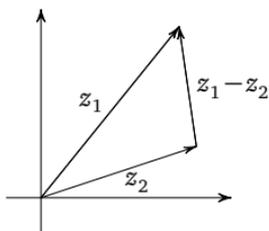


Рис. 4

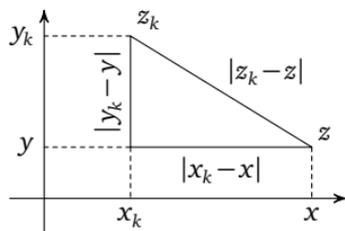


Рис. 5

Определение 1. Последовательность комплексных чисел z_k ($k \in \mathbb{N}$) называется сходящейся к комплексному числу z (обозначение: $z_k \rightarrow z$), если $|z_k - z| \rightarrow 0$.

Лемма 1. Пусть $z_k = x_k + y_k i$, $z = x + y i$ ($x_k, y_k, x, y \in \mathbb{R}$). Тогда

$$z_k \rightarrow z \Leftrightarrow x_k \rightarrow x \text{ и } y_k \rightarrow y.$$

Доказательство. Имеем (см. рис. 5)

$$|z_k - z| = \sqrt{|x_k - x|^2 + |y_k - y|^2},$$

так что

$$x_k \rightarrow x \text{ и } y_k \rightarrow y \Rightarrow z_k \rightarrow z.$$

Обратная импликация вытекает из неравенств

$$|x_k - x| \leq |z_k - z|, \quad |y_k - y| \leq |z_k - z|. \quad \square$$

Лемма 2. $z_k \rightarrow z \Rightarrow |z_k| \rightarrow |z|$.

Доказательство следует из того, что

$$||z_k| - |z|| \leq |z_k - z|. \quad \square$$

Лемма 3. $z_k \rightarrow z$ и $w_k \rightarrow w \Rightarrow z_k + w_k \rightarrow z + w$ и $z_k w_k \rightarrow zw$.

Доказательство такое же, как для последовательностей вещественных чисел:

$$|(z_k + w_k) - (z + w)| = |(z_k - z) + (w_k - w)| \leq |z_k - z| + |w_k - w| \rightarrow 0,$$

$$|z_k w_k - zw| = |(z_k - z)w_k + z(w_k - w)| \leq |z_k - z||w_k| + |z||w_k - w| \rightarrow 0. \quad \square$$

Следствие. Пусть $z_k \rightarrow z$ и $f \in \mathbb{C}[z]$ — любой многочлен. Тогда $f(z_k) \rightarrow f(z)$.

(Здесь мы допускаем вольность в обозначениях, обычную в анализе, когда значение переменной обозначается так же, как сама переменная.)

Определение 2. Последовательность комплексных чисел z_k ($k \in \mathbb{N}$) называется сходящейся к бесконечности (обозначение: $z_k \rightarrow \infty$), если $|z_k| \rightarrow \infty$.

Лемма 4. Из всякой последовательности комплексных чисел z_k можно выбрать либо подпоследовательность, сходящуюся к некоторому комплексному числу z_0 , либо подпоследовательность, сходящуюся к бесконечности.

Доказательство. Если последовательность модулей чисел z_k неограниченна, то из нее можно выбрать подпоследовательность, сходящуюся к бесконечности; но тогда соответствующая подпоследовательность самих чисел z_k согласно определению будет стремиться к бесконечности.

Пусть теперь последовательность модулей ограничена, т. е. существует такое $C > 0$, что

$$|z_k| \leq C \quad \forall k.$$

Представим z_k в алгебраической форме:

$$z_k = x_k + y_k i.$$

Тогда

$$|x_k| \leq |z_k| \leq C, \quad |y_k| \leq |z_k| \leq C.$$

По теореме Больцано—Вейерштрасса из последовательности x_k можно выбрать сходящуюся подпоследовательность. Перейдя к этой подпоследовательности и изменив обозначения, можно считать, что

$$x_k \rightarrow x_0.$$

Аналогичным образом, перейдя еще раз к подпоследовательности, можно считать, что

$$y_k \rightarrow y_0.$$

Но тогда по лемме 1

$$z_k \rightarrow z_0 = x_0 + y_0 i. \quad \square$$

Лемма 5. Если $z_k \rightarrow \infty$ и $f \in \mathbb{C}[z]$ — многочлен положительной степени, то $f(z_k) \rightarrow \infty$.

Доказательство. Пусть

$$f = a_0 z^n + a_1 z^{n-1} + \dots + a_{n-1} z + a_n \quad (a_0 \neq 0);$$

тогда

$$|f(z_k)| = |z_k|^n \left| a_0 + \frac{a_1}{z_k} + \dots + \frac{a_{n-1}}{z_k^{n-1}} + \frac{a_n}{z_k^n} \right|.$$

Сумма, стоящая под знаком модуля, стремится к a_0 , так как все слагаемые, кроме a_0 , стремятся к 0. Так как $|z_k|^n \rightarrow \infty$, то и $|f(z_k)| \rightarrow \infty$. \square

Следующая лемма является ключевой для доказательства основной теоремы.

Лемма 6 (лемма Даламбера). Пусть $f \in \mathbb{C}[z]$ — многочлен положительной степени и $f(z_0) \neq 0$. Тогда сколь угодно близко к z_0 можно найти такое z , что $|f(z)| < |f(z_0)|$.

Доказательство. Разложим f по степеням $z - z_0$ и разделим на $f(z_0)$. Учитывая, что несколько первых коэффициентов разложения, следующих за свободным членом, могут оказаться равными нулю, запишем результат в виде

$$\frac{f(z)}{f(z_0)} = 1 + c_p(z - z_0)^p + c_{p+1}(z - z_0)^{p+1} + \dots + c_n(z - z_0)^n \quad (c_p \neq 0).$$

Нам нужно доказать существование такого z , что

$$\left| \frac{f(z)}{f(z_0)} \right| < 1.$$

Идея доказательства состоит в том, что, если выбирать z достаточно близким к z_0 , выполнение этого неравенства будет зависеть только от суммы первых двух членов предыдущего разложения.

Будем искать z в виде

$$z = z_0 + tz_1$$

(см. рис. 6), где $t \in (0, 1)$, а z_1 — комплексное число, удовлетворяющее условию $c_p z_1^p = -1$.

Имеем тогда

$$\frac{f(z)}{f(z_0)} = 1 - t^p + t^{p+1} \varphi(t),$$

где φ — некоторый многочлен степени $n - p - 1$ (с комплексными коэффициентами). Если C — максимум модулей коэффициентов многочлена φ , то

$$|\varphi(t)| \leq A = (n - p)C$$

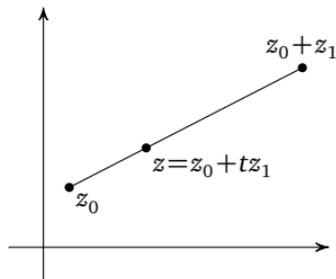


Рис. 6

и, следовательно,

$$\left| \frac{f(z)}{f(z_0)} \right| \leq 1 - t^p + At^{p+1} = 1 - t^p(1 - At) < 1$$

при $t < 1/A$. □

Доказательство теоремы 1. Пусть $f \in \mathbb{C}[z]$ — многочлен положительной степени. Положим

$$M = \inf_z |f(z)|.$$

Из определения нижней грани следует, что существует такая последовательность комплексных чисел z_k , что

$$|f(z_k)| \rightarrow M. \quad (8)$$

Перейдя к подходящей подпоследовательности, мы можем согласно лемме 4 считать, что либо $z_k \rightarrow \infty$, либо $z_k \rightarrow z_0 \in \mathbb{C}$. В первом случае в силу леммы 5 мы придем в противоречие с (8). Во втором случае $f(z_k) \rightarrow f(z_0)$ и по лемме 2

$$|f(z_k)| \rightarrow |f(z_0)| = M.$$

Если $M > 0$, то лемма Даламбера приводит нас в противоречие с определением M . Следовательно, $M = 0$, т. е. $f(z_0) = 0$. □

Следствие 1. В алгебре $\mathbb{C}[x]$ всякий ненулевой многочлен разлагается на линейные множители.

В самом деле, в силу доказанной теоремы многочлен g в разложении (4) должен иметь нулевую степень, т. е. быть просто числом.

В силу теоремы 2.3 получаем отсюда

Следствие 2. Всякий многочлен степени n над \mathbb{C} имеет n корней (с учетом кратностей).

§ 4. Корни многочленов с вещественными коэффициентами

Многочлен степени n с вещественными коэффициентами может иметь $< n$ (в частности, вообще не иметь) вещественных корней, но, как и всякий многочлен с комплексными коэффициентами, он всегда имеет ровно n комплексных корней (с учетом кратностей). Мнимые корни многочлена с вещественными коэффициентами обладают специальным свойством.

Теорема 1. Если c — мнимый корень многочлена $f \in \mathbb{R}[x]$, то \bar{c} также является корнем этого многочлена, причем той же кратности, что и c .

Доказательство. Пусть

$$f = a_0x^n + a_1x^{n-1} + \dots + a_{n-1}x + a_n \quad (a_0, a_1, \dots, a_n \in \mathbb{R}).$$

Если $f(c) = 0$, то, поскольку комплексное сопряжение является автоморфизмом поля \mathbb{C} (см. § 1.4),

$$\begin{aligned} f(\bar{c}) &= a_0\bar{c}^n + a_1\bar{c}^{n-1} + \dots + a_{n-1}\bar{c} + a_n = \\ &= \bar{a}_0\bar{c}^n + \bar{a}_1\bar{c}^{n-1} + \dots + \bar{a}_{n-1}\bar{c} + \bar{a}_n = \overline{f(c)} = \bar{0} = 0, \end{aligned}$$

т. е. \bar{c} — также корень многочлена f . Аналогично доказывается, что

$$f^{(k)}(c) = 0 \Leftrightarrow f^{(k)}(\bar{c}) = 0.$$

Следовательно, кратности корней c и \bar{c} одинаковы. \square

Следствие. В алгебре $\mathbb{R}[x]$ всякий ненулевой многочлен разлагается на линейные множители и квадратичные множители с отрицательным дискриминантом.

Доказательство. Заметим, что если c — мнимое число, то квадратный трехчлен

$$(x - c)(x - \bar{c}) = x^2 - (c + \bar{c})x + c\bar{c}$$

имеет вещественные коэффициенты; его дискриминант, очевидно, отрицателен.

Пусть теперь

$$c_1, \dots, c_s, c_{s+1}, \dots, c_{s+t}, \bar{c}_{s+1}, \dots, \bar{c}_{s+t}$$

— это все (различные) комплексные корни многочлена $f \in \mathbb{R}[x]$, причем

$$c_1, \dots, c_s \in \mathbb{R}, \quad c_{s+1}, \dots, c_{s+t} \notin \mathbb{R}.$$

Если кратность корня c_i равна k_i , то

$$\begin{aligned} f &= a_0(x - c_1)^{k_1} \dots (x - c_s)^{k_s} \times \\ &\times [(x - c_{s+1})(x - \bar{c}_{s+1})]^{k_{s+1}} \dots [(x - c_{s+t})(x - \bar{c}_{s+t})]^{k_{s+t}}, \end{aligned}$$

(где a_0 — старший коэффициент многочлена f). Перемножая линейные множители в квадратных скобках, получаем искомое разложение. \square

Пример 1.

$$\begin{aligned}
 x^5 - 1 &= (x - 1) \left(x - \left(\cos \frac{2\pi}{5} + i \sin \frac{2\pi}{5} \right) \right) \left(x - \left(\cos \frac{2\pi}{5} - i \sin \frac{2\pi}{5} \right) \right) \times \\
 &\quad \times \left(x - \left(\cos \frac{4\pi}{5} + i \sin \frac{4\pi}{5} \right) \right) \left(x - \left(\cos \frac{4\pi}{5} - i \sin \frac{4\pi}{5} \right) \right) = \\
 &= (x - 1) \left(x^2 - 2x \cos \frac{2\pi}{5} + 1 \right) \left(x^2 - 2x \cos \frac{4\pi}{5} + 1 \right) = \\
 &= (x - 1) \left(x^2 - \frac{\sqrt{5}-1}{2}x + 1 \right) \left(x^2 + \frac{\sqrt{5}+1}{2}x + 1 \right)
 \end{aligned}$$

(см. пример 2.1).

Пример 2. Для многочлена f из примера 2.3 разложение, о котором идет речь, имеет вид

$$f = (x - 2)^3(x^2 + x + 1).$$

Из теоремы 1 также следует, что любой многочлен $f \in \mathbb{R}[x]$ нечетной степени имеет хотя бы один вещественный корень. Впрочем, это легко доказать и по-другому. А именно, если старший коэффициент многочлена f положителен, то

$$\lim_{x \rightarrow +\infty} f(x) = +\infty, \quad \lim_{x \rightarrow -\infty} f(x) = -\infty$$

и, значит, многочлен f принимает как положительные, так и отрицательные значения. По теореме о промежуточном значении непрерывной функции отсюда следует, что в некоторой точке он обращается в нуль.

Понятно, что представляет интерес определение точного числа вещественных корней. Вычисляя значение многочлена в отдельных точках, мы можем обнаружить, что в каких-то точках a и b он принимает значения разных знаков. Отсюда следует, что в интервале (a, b) находится по меньшей мере один корень, а точнее — нечетное число корней (с учетом их кратностей). Таким образом мы можем оценить снизу число вещественных корней.

Пример 3. Для многочлена

$$f = x^4 + x^2 - 4x + 1$$

находим

$$f(0) = 1 > 0, \quad f(1) = -1 < 0, \quad f(2) = 13 > 0.$$

Следовательно, f имеет корни в каждом из интервалов $(0, 1)$ и $(1, 2)$. Нетрудно показать, что $f(x) > 0$ при $x \leq 0$, а также при $x \geq 2$. Следовательно, все вещественные корни многочлена f лежат в интервале

(0, 2). Однако точное их число остается неизвестным, так как в одном из интервалов (0, 1) и (1, 2) может быть три корня.

Существуют методы, которые в принципе позволяют определить как общее число вещественных корней любого многочлена с вещественными коэффициентами, так и число его корней в любом промежутке числовой прямой. Однако их практическое применение связано с довольно большими вычислениями. Мы приведем здесь одну теорему, которая хотя и не всегда дает полный ответ, но зато не требует никаких вычислений. Она говорит не просто о числе всех вещественных корней, но о числе положительных (или отрицательных) корней и является обобщением следующего тривиального утверждения: если все коэффициенты многочлена неотрицательны, то он не имеет положительных корней.

Для формулировки этой теоремы нам понадобится одно вспомогательное понятие.

Пусть имеется конечная последовательность вещественных чисел

$$a_0, a_1, a_2, \dots, a_n.$$

Говорят, что на k -м месте в этой последовательности имеется *перемена знака*, если $a_k \neq 0$ и знак числа a_k противоположен знаку последнего из предшествующих ему ненулевых членов последовательности. (Если a_k — первый из ненулевых членов последовательности, то на k -м месте перемены знака нет.)

Теорема 2 (теорема Декарта). *Число положительных корней (с учетом их кратностей) многочлена $f \in \mathbb{R}[x]$ не превосходит числа перемен знака в последовательности его коэффициентов и сравнимо с ним по модулю 2; если же все (комплексные) корни многочлена f вещественны, то эти числа равны.*

Обозначим через $N(f)$ число положительных корней многочлена f и через $L(f)$ число перемен знака в последовательности его коэффициентов. Очевидно, что эти числа не изменяются при умножении f на -1 ; поэтому всегда можно считать, что старший коэффициент многочлена f положителен. Кроме того, если 0 является k -кратным корнем многочлена f , то при делении f на x^k эти числа также не изменяются; поэтому можно считать, что свободный член многочлена f отличен от нуля.

Лемма 1. $N(f) \equiv L(f) \pmod{2}$.

Доказательство. Пусть

$$f = a_0 x^n + a_1 x^{n-1} + \dots + a_{n-1} x + a_n \quad (a_0 > 0, a_n \neq 0).$$

Тогда $f(0) = a_n$ и $f(x) > 0$ при достаточно больших x . Когда мы движемся вправо по числовой прямой, то при прохождении каждого простого корня $f(x)$ меняет знак, а при прохождении k -кратного корня знак $f(x)$ умножается на $(-1)^k$, т. е. как бы меняется k раз. Поэтому $N(f)$ четно, если $a_n > 0$, и нечетно, если $a_n < 0$. То же самое можно сказать и об $L(f)$. \square

Лемма 2. $N(f) \leq N(f') + 1$.

Доказательство. По теореме Ролля между любыми двумя корнями многочлена f лежит корень его производной. Кроме того, каждый k -кратный корень многочлена f является $(k - 1)$ -кратным корнем его производной (следствие теоремы 2.4). Отсюда получаем, что $N(f') \geq N(f) - 1$. \square

Лемма 3. $L(f') \leq L(f)$.

Доказательство очевидно. \square

Число отрицательных корней многочлена f равно числу положительных корней многочлена

$$\bar{f}(x) = (-1)^n f(-x).$$

Лемма 4. $L(f) + L(\bar{f}) \leq n = \deg f$.

Доказательство. Коэффициенты многочлена \bar{f} получаются из коэффициентов многочлена f попеременным умножением на ± 1 . Предположим вначале, что все коэффициенты a_0, a_1, \dots, a_n многочлена f отличны от нуля. Тогда если на k -м месте в последовательности a_0, a_1, \dots, a_n имеется перемена знака, то на том же месте в последовательности коэффициентов многочлена \bar{f} перемены знака нет, и наоборот. Поэтому в этом случае $L(f) + L(\bar{f}) = n$. В общем случае, когда среди коэффициентов a_0, a_1, \dots, a_n могут быть нули, при их замене произвольными числами, отличными от нуля, числа $L(f)$ и $L(\bar{f})$ могут только увеличиться. Так как после этого их сумма по доказанному станет равной n , то $L(f) + L(\bar{f}) \leq n$. \square

Доказательство теоремы 2. Докажем неравенство $N(f) \leq L(f)$ индукцией по $\deg f$. Если $\deg f = 0$, то $N(f) = L(f) = 0$. Пусть $\deg f = n > 0$. Тогда $\deg f' = n - 1$. Пользуясь леммами 2 и 3 и предположением индукции, получаем

$$N(f) \leq N(f') + 1 \leq L(f') + 1 \leq L(f) + 1.$$

Однако равенство $N(f) = L(f) + 1$ невозможно ввиду леммы 1. Следовательно, $N(f) \leq L(f)$.

Пусть теперь известно, что все корни многочлена f вещественны. Мы можем считать, что 0 не является корнем. Имеем тогда в силу уже доказанного неравенства и леммы 4

$$n = N(f) + N(\bar{f}) \leq L(f) + L(\bar{f}) \leq n,$$

откуда

$$N(f) = L(f), \quad N(\bar{f}) = L(\bar{f}). \quad \square$$

Пример 4. Для многочлена f из примера 3 имеем $L(f) = 2$, так что $N(f) \leq 2$. Но мы уже установили, что $N(f) \geq 2$. Следовательно, $N(f) = 2$.

Пример 5. Многочлен $f = x^2 - x + 1$ не имеет положительных (и вообще вещественных) корней, но $L(f) = 2$, так что в этом случае $N(f) < L(f)$.

Применяя теорему Декарта к многочлену

$$g(x) = f(c+x) = f(c) + \frac{f'(c)}{1!}x + \frac{f''(c)}{2!}x^2 + \dots + \frac{f^{(n)}(c)}{n!}x^n,$$

мы получаем информацию о числе корней многочлена f в промежутке $(c, +\infty)$. В частности, если все коэффициенты многочлена g неотрицательны, то он не имеет положительных корней (тривиальный случай теоремы Декарта), а это означает, что все вещественные корни многочлена f не превосходят c .

Пример 6. Найдем границы вещественных корней многочлена

$$f = x^5 - 5x^3 - 10x^2 + 2.$$

Пользуясь схемой Горнера, вычислим $f(3)$:

$$\begin{array}{r|rrrrrr} & 1 & 0 & -5 & -10 & 0 & 2 \\ 3 & 1 & 3 & 4 & 2 & 6 & 20 \end{array}$$

Мы видим, что $f(3) = 20 > 0$. Более того, все коэффициенты неполного частного оказались положительными. Поэтому все производные многочлена f при $x = 3$ также положительны (см. пример 2.3) и, значит, все его вещественные корни меньше 3. Рассмотрим теперь многочлен

$$\bar{f}(x) = -f(-x) = x^5 - 5x^3 + 10x^2 - 2.$$

Вычислим значения многочлена \bar{f} и его производных при $x = 1$:

	1	0	-5	10	0	-2
1	1	1	-4	6	6	4
	1	2	-2	4	10	
	1	3	1	5		

Мы видим, что

$$\bar{f}(1) = 4 > 0, \quad \bar{f}'(1) = 10 > 0, \quad \bar{f}''(1) = 2 \cdot 5 > 0.$$

Значения следующих производных также положительны, поскольку последняя строка таблицы состоит только из положительных чисел. Следовательно, все вещественные корни многочлена \bar{f} меньше 1, а это означает, что все вещественные корни многочлена f больше -1 . Таким образом, все вещественные корни многочлена f лежат в интервале $(-1, 3)$.

Задача 1. Исследовав производную многочлена \bar{f} , доказать, что многочлен f из предыдущего примера имеет только один отрицательный корень.

Обратимся теперь к вопросу о приближенном вычислении корней.

Если известно, что многочлен $f \in \mathbb{R}[x]$ имеет ровно один корень в каком-то интервале, то этот корень может быть в принципе найден с любой степенью точности с помощью вычисления значений многочлена в подходящим образом подобранных точках. Поясним это на следующем примере.

Пример 7. Как мы показали (см. пример 4), многочлен f из примера 3 имеет ровно один корень в интервале $(1, 2)$. Найдем значение этого корня с точностью до 0,01. Мы видели, что $f(1) < 0$. Вычисляя $f(x)$ при $x = 1,1; 1,2; 1,3$, мы обнаруживаем, что

$$f(1,2) < 0, \quad f(1,3) > 0.$$

Следовательно, корень лежит в интервале $(1,2; 1,3)$. Вычисляя $f(x)$ при $x = 1,21; 1,22; 1,23; 1,24; 1,25$, находим, что

$$f(1,24) < 0, \quad f(1,25) > 0.$$

Следовательно, искомый корень лежит в интервале $(1,24; 1,25)$.

Конечно, существуют гораздо более совершенные методы приближенного вычисления корней. Они применимы к алгебраиче-

ским уравнениям любой степени, а некоторые из них — и к трансцендентным уравнениям. Однако изложение этих методов выходит за рамки нашего курса: они относятся скорее к вычислительной математике, чем к алгебре.

Замечание 1. Если многочлен имеет кратный корень, но его коэффициенты даны нам лишь приближенно, хотя бы и с любой степенью точности, то мы в принципе не можем доказать наличие этого кратного корня, так как при сколь угодно малом изменении коэффициентов многочлена он может либо рассыпаться на простые корни, либо вообще перестать существовать. Так, в случае двукратного корня мы никогда не сможем сделать выбор между ситуациями, изображенными на рис. 7, а), а в случае трехкратного — между ситуациями, изображенными на рис. 7, б).

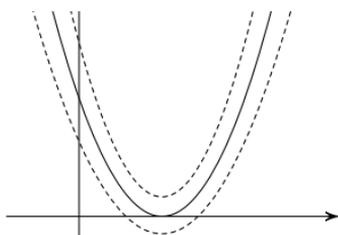


Рис. 7, а)

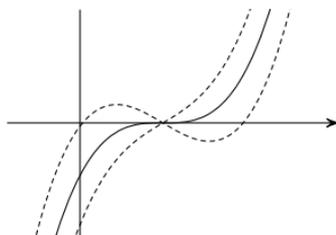


Рис. 7, б)

§ 5. Теория делимости в евклидовых кольцах

Разложение многочленов над \mathbb{C} на линейные множители и многочленов над \mathbb{R} на линейные и квадратичные множители аналогично разложению целых чисел на простые множители. Для многочленов над произвольным полем также имеется подобное разложение, но его множители могут иметь любую степень. Задачу отыскания такого разложения можно рассматривать как обобщение задачи отыскания корней многочлена (которой она равносильна в случае многочленов над \mathbb{C}). Она не имеет общего решения, пригодного для любого поля. В этом параграфе мы докажем единственность указанного разложения. Одновременно мы докажем единственность разложения целого числа на простые множители — факт широко известный, но не доказываемый в средней школе.

Для того чтобы охватить единым рассуждением оба случая, введем некоторые общие понятия.

Определение 1. Коммутативное ассоциативное кольцо с единицей и без делителей нуля называется *целостным кольцом* (или *областью целостности*).

Так, кольцо \mathbb{Z} целых чисел и кольцо $K[x]$ многочленов над любым полем K являются целостными кольцами. Более того, кольцо многочленов над любым целостным кольцом также является целостным кольцом (см. замечание 1.3).

Замечание 1. Кольцо, состоящее из одного нуля, не считается целостным.

Пусть A — целостное кольцо. Говорят, что элемент $a \in A$ делится на элемент $b \in A$ (обозначение: $a : b$) или, иначе, что b делит a (обозначение: $b | a$), если существует такой элемент $q \in A$, что $a = qb$. Элементы a и b называются *ассоциированными* (обозначение: $a \sim b$), если выполняется любое из следующих эквивалентных условий:

- 1) $b | a$ и $a | b$;
- 2) $a = cb$, где c — обратимый элемент.

В следующем определении аксиоматизируется то общее, что есть у кольца многочленов над полем и кольца целых чисел, — возможность деления с остатком.

Определение 2. Целостное кольцо A , не являющееся полем, называется *евклидовым*, если существует функция

$$N: A \setminus \{0\} \rightarrow \mathbb{Z}_+$$

(называемая *нормой*), удовлетворяющая следующим условиям:

- 1) $N(ab) \geq N(a)$, причем равенство имеет место только тогда, когда элемент b обратим;
- 2) для любых $a, b \in A$, где $b \neq 0$, существуют такие $q, r \in A$, что $a = qb + r$ и либо $r = 0$, либо $N(r) < N(b)$.

Замечание 2. Условие 2) означает возможность «деления с остатком». Его единственности (т. е. однозначной определенности пары (q, r)) не требуется.

Замечание 3. Вторая часть условия 1) на самом деле может быть выведена из остальных условий. В самом деле, пусть элемент b необратим. Тогда a не делится на ab . Разделим a на ab с остатком:

$$a = q(ab) + r.$$

Так как $r = a(1 - qb)$, то

$$N(a) \leq N(r) < N(ab).$$

Основными для нас примерами евклидовых колец являются кольцо \mathbb{Z} целых чисел и кольцо $K[x]$ многочленов над полем K . В качестве нормы в первом случае можно взять модуль целого числа, а во втором — степень многочлена.

Существуют и другие евклидовы кольца.

Пример 1. Комплексные числа вида $c = a + bi$, где $a, b \in \mathbb{Z}$, называются *целыми гауссовыми числами*. Они образуют подкольцо в \mathbb{C} , обозначаемое через $\mathbb{Z}[i]$. Кольцо $\mathbb{Z}[i]$ является евклидовым относительно нормы

$$N(c) = |c|^2 = a^2 + b^2.$$

В самом деле, очевидно, что $N(cd) = N(c)N(d)$ и, поскольку $N(1) = 1$, обратимые элементы кольца $\mathbb{Z}[i]$ — это элементы с нормой 1, т.е. ± 1 и $\pm i$. Отсюда следует, что выполнено условие 1) в определении евклидова кольца. Докажем возможность деления с остатком.

Пусть $c, d \in \mathbb{Z}[i]$, $d \neq 0$. Рассмотрим целое гауссово число q , ближайшее к $\frac{c}{d}$. Легко видеть, что $\left| \frac{c}{d} - q \right| \leq 1/\sqrt{2}$ (см. рис. 8). Положим $r = c - qd$. Тогда $c = qd + r$ и

$$N(r) = |c - qd|^2 = \left| \frac{c}{d} - q \right|^2 |d|^2 \leq \frac{1}{2} N(d) < N(d).$$

Задача 1. Доказать, что кольцо рациональных чисел вида $2^{-n}m$ ($m \in \mathbb{Z}$, $n \in \mathbb{Z}_+$) является евклидовым.

Определение 3. Наибольшим общим делителем элементов a и b целостного кольца называется их общий делитель, делящийся на все их общие делители. Он обозначается через (a, b) или $\text{НОД}\{a, b\}$.

Наибольший общий делитель, если он существует, определен однозначно с точностью до ассоциированности. Однако его может не существовать. Например, элементы x^5 и x^6 в кольце многочленов без линейного члена не имеют наибольшего общего делителя.

Теорема 1. В евклидовом кольце для любых элементов a, b существует наибольший общий делитель d , и он может быть представлен в виде $d = au + bv$, где u, v — какие-то элементы кольца.

Доказательство. Если $b = 0$, то $d = a = a \cdot 1 + b \cdot 0$. Если a делится на b , то $d = b = a \cdot 0 + b \cdot 1$. В противном случае разделим c с остатком a на b , затем b на полученный остаток, затем первый остаток на

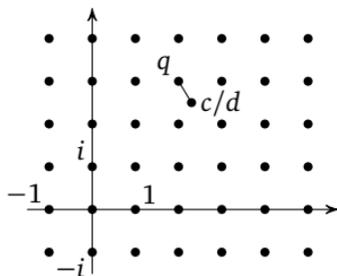


Рис. 8

второй остаток и т. д. Так как нормы остатков убывают, то в конце концов деление произойдет без остатка. Получим цепочку равенств:

$$\begin{aligned} a &= q_1 b + r_1, \\ b &= q_2 r_1 + r_2, \\ r_1 &= q_3 r_2 + r_3, \\ &\dots\dots\dots \\ r_{n-2} &= q_n r_{n-1} + r_n, \\ r_{n-1} &= q_{n+1} r_n. \end{aligned}$$

Докажем, что последний ненулевой остаток r_n и есть наибольший общий делитель элементов a и b .

Двигаясь по выписанной цепочке равенств снизу вверх, получаем последовательно

$$r_n | r_{n-1}, \quad r_n | r_{n-2}, \quad \dots, \quad r_n | r_1, \quad r_n | b, \quad r_n | a.$$

Таким образом, r_n — общий делитель элементов a и b .

Двигаясь по той же цепочке равенств сверху вниз, получаем последовательно

$$\begin{aligned} r_1 &= au_1 + bv_1, \\ r_2 &= au_2 + bv_2, \\ r_3 &= au_3 + bv_3, \\ &\dots\dots\dots \\ r_n &= au_n + bv_n, \end{aligned}$$

где u_i, v_i ($i = 1, \dots, n$) — какие-то элементы кольца (например, $u_1 = 1, v_1 = -q_1$). Таким образом, r_n можно представить в виде $au + bv$. Отсюда, в свою очередь, следует, что r_n делится на любой общий делитель элементов a и b . \square

Процедура нахождения наибольшего общего делителя, использованная в этом доказательстве, называется *алгоритмом Евклида*. Элементы $a, b \in A$ называются *взаимно простыми*, если $(a, b) = 1$. В этом случае, согласно доказанной теореме, существуют такие $u, v \in A$, что

$$au + bv = 1.$$

Перейдем теперь к вопросу о разложении на простые множители.

Определение 4. Необратимый ненулевой элемент p целостного кольца называется *простым*, если он не может быть представлен в виде $p = ab$, где a и b — необратимые элементы.

Иначе говоря, элемент p простой, если всякий его делитель ассоциирован либо с 1, либо с p . Простые элементы кольца \mathbb{Z} в этом смысле — это числа вида $\pm p$, где p — простое число.

Простые элементы кольца $K[x]$, где K — поле, по традиции называются *неприводимыми многочленами*. Таким образом, неприводимый многочлен — это такой многочлен положительной степени, который не может быть разложен в произведение двух многочленов положительной степени.

Очевидно, что всякий многочлен первой степени неприводим. Из основной теоремы алгебры комплексных чисел вытекает, что неприводимые многочлены над \mathbb{C} — это только многочлены первой степени, а из следствия теоремы 4.1 — что неприводимые многочлены над \mathbb{R} — это многочлены первой степени и многочлены второй степени с отрицательным дискриминантом. В следующем параграфе мы обсудим вопрос о неприводимых многочленах над \mathbb{Q} и, в частности, увидим, что они могут иметь любую степень.

Пусть теперь A — любое евклидово кольцо.

Лемма 1. Если простой элемент p кольца A делит произведение $a_1 a_2 \dots a_n$, то он делит хотя бы один из сомножителей a_1, a_2, \dots, a_n .

Доказательство. Докажем это утверждение индукцией по n . При $n = 2$ предположим, что p не делит a_1 . Тогда $(p, a_1) = 1$ и, значит, существуют такие $u, v \in A$, что $pu + a_1v = 1$. Умножая это равенство на a_2 , получаем

$$pua_2 + a_1a_2v = a_2,$$

откуда следует, что p делит a_2 .

При $n > 2$ представим произведение $a_1 a_2 \dots a_n$ в виде $a_1(a_2 \dots a_n)$. По доказанному $p \mid a_1$ или $p \mid a_2 \dots a_n$. Во втором случае по предположению индукции $p \mid a_i$, где i — один из индексов $2, \dots, n$. \square

Теорема 2. В евклидовом кольце всякий необратимый ненулевой элемент может быть разложен на простые множители, причем это разложение единственно с точностью до перестановки множителей и умножения их на обратимые элементы.

Замечание 4. Говоря о разложении на простые множители, мы не исключаем разложения, состоящего только из одного множителя.

Доказательство. Назовем необратимый ненулевой элемент $a \in A$ хорошим, если он может быть разложен на простые множители. Предположим, что существуют плохие элементы. Выберем из них элемент с наименьшей нормой. Пусть это будет элемент a . Он не может быть простым. Следовательно, $a = bc$, где b и c — необратимые элементы. Имеем $N(b) < N(a)$ и $N(c) < N(a)$ и, значит, b и c — хорошие элементы; но тогда, очевидно, и a — хороший элемент, что противоречит нашему предположению. Таким образом, всякий необратимый ненулевой элемент кольца A может быть разложен на простые множители.

Докажем теперь индукцией по n , что если

$$a = p_1 p_2 \dots p_n = q_1 q_2 \dots q_m, \quad (9)$$

где p_i, q_j — простые элементы, то $m = n$ и, после подходящей перенумерации множителей, $p_i \sim q_i$ при $i = 1, 2, \dots, n$.

При $n=1$ это утверждение очевидно. При $n > 1$ имеем $p_1 | q_1 q_2 \dots q_m$ и по лемме 1 существует такой номер i , что $p_1 | q_i$. Тогда $p_1 \sim q_i$. Можно считать, что $i = 1$ и $p_1 = q_1$. Сокращая равенство (9) на p_1 , получаем

$$p_2 \dots p_n = q_2 \dots q_m.$$

По предположению индукции отсюда следует, что $m = n$ и, после подходящей перенумерации, $p_i \sim q_i$ при $i = 2, \dots, n$. Тем самым утверждение доказано. \square

Следствие. Пусть $a = p_1^{k_1} \dots p_s^{k_s}$ — разложение элемента $a \in A$ на простые множители, причем $p_i \not\sim p_j$ при $i \neq j$. Тогда всякий делитель d элемента a имеет вид

$$d = c p_1^{l_1} \dots p_s^{l_s},$$

где $0 \leq l_i \leq k_i$ ($i = 1, \dots, s$), а c — обратимый элемент.

Доказательство. Пусть $a = qd$. Разложим q и d на простые множители. Перемножив эти разложения, мы получим разложение a на простые множители. Сравнив его с данным разложением, получим требуемый результат. \square

Задача 2. Доказать, что в евклидовом кольце

а) $b | a, c | a$ и $(b, c) = 1 \Rightarrow bc | a$;

б) $c | ab$ и $(b, c) = 1 \Rightarrow c | a$.

Задача 3. Наименьшим общим кратным элементов a и b целостного кольца называется их общее кратное (т. е. элемент, делящийся на a и на b), делящее все их общие кратные. Оно обозначается

через $[a, b]$ или $\text{НОК}\{a, b\}$). Доказать, что в евклидовом кольце для любых элементов a, b существует наименьшее общее кратное $[a, b]$, причем

$$(a, b)[a, b] \sim ab.$$

Задача 4. В кольце $\mathbb{Z}[i]$ (см. пример 1) разложить на простые множители числа 2, 3 и 5 и подумать, в чем принципиальная разница между этими тремя случаями.

Известно, что простых чисел бесконечно много. Напомним рассуждение, которое это доказывает. Предположим, что p_1, p_2, \dots, p_n — это все простые числа. Тогда число $p_1 p_2 \dots p_n + 1$ не делится ни на одно из них, что, очевидно, невозможно. Точно такое же рассуждение показывает бесконечность числа нормированных неприводимых многочленов над любым полем K . Если поле K бесконечно, то этот результат не представляет интереса, так как в этом случае имеется бесконечно много нормированных многочленов первой степени. Однако если поле K конечно, то этот результат означает, что имеются неприводимые многочлены сколь угодно высокой степени. На самом деле в этом случае имеются неприводимые многочлены любой степени.

Задача 5. Перечислить неприводимые многочлены степеней ≤ 4 над полем \mathbb{Z}_2 и доказать, что существует ровно 6 неприводимых многочленов степени 5.

§ 6. Многочлены с рациональными коэффициентами

Из однозначности разложения целого числа на простые множители вытекает

Теорема 1. Если многочлен

$$f = a_0 x^n + a_1 x^{n-1} + \dots + a_{n-1} x + a_n \in \mathbb{Z}[x]$$

имеет рациональный корень $\frac{u}{v}$, где $u, v \in \mathbb{Z}$, $(u, v) = 1$, то $u \mid a_n$, $v \mid a_0$.

Доказательство. Имеем

$$0 = v^n f\left(\frac{u}{v}\right) = a_0 u^n + a_1 u^{n-1} v + \dots + a_{n-1} u v^{n-1} + a_n v^n.$$

Все слагаемые в правой части, кроме последнего, делятся на u . Следовательно, и последнее слагаемое должно делиться на u . Но так как u и v взаимно просты, то a_n делится на u (см. задачу 5.2, б)). Аналогично доказывается, что a_0 делится на v . \square

Следствие. Если нормированный многочлен с целыми коэффициентами имеет рациональный корень, то этот корень целый.

Очевидно, что всякий многочлен с рациональными коэффициентами пропорционален многочлену с целыми коэффициентами. Поэтому теорема 1 позволяет путем конечного числа испытаний найти все рациональные корни любого многочлена с рациональными коэффициентами. Конечно, таких корней, как правило, нет. Приводимый ниже специально подобранный пример относится к разряду тех исключений, которые подтверждают правило.

Пример 1. Рациональными корнями многочлена

$$f = 2x^4 - 7x^3 + 4x^2 - 2x - 3$$

согласно теореме 1 могут быть только

$$\pm \frac{1}{2}, \quad \pm 1, \quad \pm \frac{3}{2}, \quad \pm 3.$$

Испытания дают 2 корня:

$$x_1 = 3, \quad x_2 = -\frac{1}{2}.$$

Следующая теорема может рассматриваться как обобщение теоремы 1.

Теорема 2 (лемма Гаусса). Если многочлен с целыми коэффициентами разлагается в произведение двух многочленов с рациональными коэффициентами, то он разлагается в произведение двух пропорциональных им многочленов с целыми коэффициентами.

Иначе говоря, если $f \in \mathbb{Z}[x]$ и $f = gh$, где $g, h \in \mathbb{Q}[x]$, то существует такое $\lambda \in \mathbb{Q}^*$, что $\lambda g, \lambda^{-1}h \in \mathbb{Z}[x]$.

Перед тем, как доказывать эту теорему, введем некоторые вспомогательные понятия.

Многочлен $f \in \mathbb{Z}[x]$ называется *примитивным*, если его коэффициенты взаимно просты в совокупности, т. е. не имеют общего простого делителя. Если такой делитель есть, то его можно вынести за скобки. Поэтому всякий ненулевой многочлен с целыми коэффициентами, а значит, и всякий ненулевой многочлен с рациональными коэффициентами, пропорционален некоторому примитивному многочлену (определенному однозначно с точностью до умножения на ± 1).

Пусть p — какое-нибудь простое число. Определим *редукцию по модулю p* многочлена

$$f = a_0x^n + a_1x^{n-1} + \dots + a_{n-1}x + a_n \in \mathbb{Z}[x]$$

как многочлен

$$[f]_p = [a_0]_p x^n + [a_1]_p x^{n-1} + \dots + [a_{n-1}]_p x + [a_n]_p \in \mathbb{Z}_p[x],$$

коэффициенты которого суть вычеты по модулю p коэффициентов многочлена f . Из определения операций над вычетами следует, что

$$[f + g]_p = [f]_p + [g]_p,$$

$$[fg]_p = [f]_p [g]_p$$

для любых многочленов $f, g \in \mathbb{Z}[x]$.

Доказательство теоремы 2. Пусть $f \in \mathbb{Z}[x]$ и $f = gh$, где $g, h \in \mathbb{Q}[x]$. Согласно предыдущему, многочлены g и h пропорциональны каким-то примитивным многочленам g_1 и h_1 . Имеем

$$f = \mu g_1 h_1, \quad \mu \in \mathbb{Q}.$$

Пусть $\mu = \frac{u}{v}$, где $u, v \in \mathbb{Z}$, $(u, v) = 1$. Докажем, что $v = \pm 1$, откуда будет следовать утверждение теоремы. Если это не так, то пусть p — какой-нибудь простой делитель числа v . В равенстве

$$vf = \mu g_1 h_1$$

сделаем редукцию по модулю p . Мы получим

$$0 = [u]_p [g_1]_p [h_1]_p.$$

Однако $[u]_p \neq 0$, так как u и v по предположению взаимно просты. В то же время $[g_1]_p \neq 0$ и $[h_1]_p \neq 0$, так как g_1 и h_1 — примитивные многочлены и, следовательно, все их коэффициенты не могут делиться на p . Это противоречит отсутствию делителей нуля в кольце $\mathbb{Z}_p[x]$. \square

Следствие. Если многочлен $f \in \mathbb{Z}[x]$ допускает разложение в произведение двух многочленов положительной степени в кольце $\mathbb{Q}[x]$, то он допускает такое разложение и в кольце $\mathbb{Z}[x]$.

Это существенно облегчает доказательство неприводимости многочленов над \mathbb{Q} .

Пример 2. Пусть p — простое число. Докажем неприводимость над \mathbb{Q} «многочлена деления круга на p частей»

$$f = x^{p-1} + x^{p-2} + \dots + x + 1 = \frac{x^p - 1}{x - 1}.$$

(Комплексными корнями этого многочлена являются все нетривиальные корни p -й степени из 1, которые вместе с 1 делят окруж-

ность $|z| = 1$ на p равных частей.) Как следует из формулы бинома Ньютона (см. § 1.5), в кольце $\mathbb{Z}_p[x]$ имеет место равенство

$$x^p - 1 = (x - 1)^p,$$

так что

$$[f]_p = (x - 1)^{p-1}.$$

Если $f = gh$, где $g, h \in \mathbb{Z}[x]$ — многочлены положительной степени, то $[f]_p = [g]_p [h]_p$ и, значит,

$$[g]_p = (x - 1)^k, \quad [h]_p = (x - 1)^l \quad (k, l > 0, k + l = p - 1).$$

Следовательно,

$$[g(1)]_p = [g]_p(1) = 0, \quad [h(1)]_p = [h]_p(1) = 0,$$

т. е. $g(1)$ и $h(1)$ делятся на p . Но тогда $f(1) = g(1)h(1)$ делится на p^2 , что не соответствует действительности, ибо $f(1) = p$.

Имеется способ, принадлежащий Кронекеру, который в принципе позволяет для любого многочлена с целыми коэффициентами определить, приводим он или неприводим над \mathbb{Q} . Он основывается на следующих соображениях.

Пусть $f \in \mathbb{Z}[x]$ — многочлен степени n , не имеющий целых корней. Предположим, что он разлагается в $\mathbb{Z}[x]$ в произведение двух многочленов положительной степени:

$$f = gh.$$

Тогда степень одного из них, скажем, g , не превосходит $m = \left\lfloor \frac{n}{2} \right\rfloor$.

Будем придавать переменной x различные целые значения x_0, x_1, \dots, x_m . Из равенств

$$f(x_i) = g(x_i)h(x_i)$$

следует, что $g(x_i) \mid f(x_i)$ при $i = 0, 1, \dots, m$. Многочлен g однозначно определяется своими значениями в точках x_0, x_1, \dots, x_m . Выбирая всевозможные наборы делителей d_0, d_1, \dots, d_m целых чисел $f(x_0), f(x_1), \dots, f(x_m)$ и находя для каждого из них интерполяционный многочлен степени $\leq m$, принимающий в точках x_0, x_1, \dots, x_m значения d_0, d_1, \dots, d_m , можно найти всех кандидатов на роль g (их будет конечное число). Те из них, которые имеют дробные коэффициенты, следует сразу отбросить. Испытав оставшиеся многочлены, можно определить, имеются ли среди них делители многочлена f , в зависимости от чего и будет решен вопрос о приводимости последнего.

§7. Многочлены от нескольких переменных

Функция вещественных переменных x_1, x_2, \dots, x_n называется *многочленом*, если она может быть представлена в виде

$$f(x_1, x_2, \dots, x_n) = \sum_{k_1, k_2, \dots, k_n} a_{k_1 k_2 \dots k_n} x_1^{k_1} x_2^{k_2} \dots x_n^{k_n}, \quad (10)$$

где суммирование в правой части происходит по конечному числу наборов (k_1, k_2, \dots, k_n) неотрицательных целых чисел. (Формально можно считать, что суммирование происходит по всем таким наборам, но лишь конечное число коэффициентов $a_{k_1 k_2 \dots k_n}$ отлично от нуля.) Многочлены образуют подалгебру в алгебре всех функций от x_1, x_2, \dots, x_n . Она называется алгеброй многочленов от x_1, x_2, \dots, x_n над \mathbb{R} и обозначается $\mathbb{R}[x_1, x_2, \dots, x_n]$.

Можно показать (см. теорему 1 ниже), что представление многочлена над \mathbb{R} в виде (10) однозначно, т. е. коэффициенты многочлена определяются его значениями.

При определении алгебры многочленов от n переменных над любым полем K возникает такая же трудность, как и в случае одной переменной. Это приводит к необходимости формального определения, которое может быть дано, например, следующим образом.

Рассмотрим бесконечномерную алгебру над K с базисом

$$\{e_{k_1 k_2 \dots k_n} : k_1, k_2, \dots, k_n \in \mathbb{Z}_+\}$$

и таблицей умножения

$$e_{k_1 k_2 \dots k_n} e_{l_1 l_2 \dots l_n} = e_{k_1+l_1, k_2+l_2, \dots, k_n+l_n}.$$

Очевидно, что эта алгебра коммутативна и ассоциативна и что элемент $e_{00\dots 0}$ является ее единицей. Она называется *алгеброй многочленов* над K и обозначается через $K[x_1, x_2, \dots, x_n]$.

Условимся отождествлять элементы вида $ae_{00\dots 0}$ ($a \in K$) с соответствующими элементами поля K и введем обозначения

$$e_{10\dots 0} = x_1,$$

$$e_{01\dots 0} = x_2,$$

.....

$$e_{00\dots 1} = x_n.$$

Тогда

$$e_{k_1 k_2 \dots k_n} = x_1^{k_1} x_2^{k_2} \dots x_n^{k_n}$$

и любой элемент

$$\sum_{k_1, k_2, \dots, k_n} a_{k_1 k_2 \dots k_n} e_{k_1 k_2 \dots k_n} \in K[x_1, x_2, \dots, x_n]$$

записывается в обычном виде (10).

Многочлен (10) называется *однородным* степени d , если

$$a_{k_1 k_2 \dots k_n} = 0 \quad \text{при } k_1 + k_2 + \dots + k_n \neq d.$$

Однородные многочлены заданной степени d образуют конечномерное подпространство, так как имеется лишь конечное число наборов (k_1, k_2, \dots, k_n) целых неотрицательных чисел, удовлетворяющих условию

$$k_1 + k_2 + \dots + k_n = d.$$

Задача 1. Доказать, что размерность пространства однородных многочленов степени d от n переменных равна

$$CC_n^d = \frac{n(n+1)\dots(n+d-1)}{d!}$$

(число сочетаний с повторениями из n по d).

Любой многочлен однозначно представляется в виде суммы однородных многочленов степеней $0, 1, 2, \dots$, называемых его *однородными компонентами*. (Лишь конечное число из них отлично от нуля.)

Степенью (по совокупности переменных) ненулевого многочлена называется максимальная из степеней его ненулевых членов или, что то же самое, максимальная из степеней его ненулевых однородных компонент. Степень многочлена f обозначается через $\deg f$. Справедливы следующие соотношения:

$$\deg(f + g) \leq \max\{\deg f, \deg g\}, \quad (11)$$

$$\deg(fg) = \deg f + \deg g. \quad (12)$$

Первое из них очевидно, второе мы докажем чуть позже.

С другой стороны, каждый многочлен $f \in K[x_1, x_2, \dots, x_n]$ однозначно представляется в виде

$$f(x_1, x_2, \dots, x_n) = \sum_{k=0}^{\infty} f_k(x_2, \dots, x_n) x_1^k, \quad (13)$$

где f_0, f_1, f_2, \dots — какие-то многочлены от x_2, \dots, x_n , лишь конечное число которых отлично от нуля. Наибольший из номеров многочленов f_k , отличных от нуля, называется степенью многочлена f по x_1 и обозначается через $\deg_{x_1} f$.

Пользуясь представлением (13), можно рассматривать кольцо $K[x_1, x_2, \dots, x_n]$ как кольцо многочленов от x_1 с коэффициентами из $K[x_2, \dots, x_n]$:

$$K[x_1, x_2, \dots, x_n] = K[x_2, \dots, x_n][x_1]. \quad (14)$$

Замечание 1. Мы говорим о кольцах, а не об алгебрах, так как по определению $K[x_1, x_2, \dots, x_n]$ есть алгебра над K , в то время как $K[x_2, \dots, x_n][x_1]$ есть алгебра над $K[x_2, \dots, x_n]$. Однако если рассматривать $K[x_2, \dots, x_n][x_1]$ как алгебру над K (пользуясь тем, что $K[x_2, \dots, x_n] \supset K$), то можно говорить о равенстве алгебр.

Предложение 1. Алгебра $K[x_1, x_2, \dots, x_n]$ не имеет делителей нуля.

Доказательство. В §1 было фактически доказано (см. замечание 1.3), что кольцо многочленов от одной переменной над целостным кольцом также является целостным кольцом (в частности, не имеет делителей нуля). Поэтому равенство (14) позволяет доказать наше утверждение индуктивным путем, начиная с поля K . \square

Теперь мы в состоянии доказать соотношение (12). Разложим многочлены f и g на однородные компоненты:

$$\begin{aligned} f &= f_0 + f_1 + \dots + f_d \quad (\deg f_k = k, f_d \neq 0), \\ g &= g_0 + g_1 + \dots + g_e \quad (\deg g_k = k, g_e \neq 0). \end{aligned}$$

Ясно, что при их перемножении не появится членов степени $> d + e$, а сумма всех членов степени $d + e$ будет равна $f_d g_e$. По доказанному $f_d g_e \neq 0$. Следовательно,

$$\deg fg = d + e = \deg f + \deg g.$$

Как и в случае $n = 1$, всякий многочлен от n переменных над полем K определяет функцию на K^n со значениями в K .

Теорема 1. Если поле K бесконечно, то разные многочлены от n переменных над K определяют разные функции.

Доказательство. Как и в случае многочленов от одной переменной (см. доказательство теоремы 1.1), достаточно доказать, что ненулевой многочлен определяет ненулевую функцию. Докажем это индукцией по n .

При $n = 1$ это составляет содержание теоремы 1.1. Предположим теперь, что многочлен $f \in K[x_1, x_2, \dots, x_n]$ ($n > 1$) определяет нулевую функцию. Представим его в виде (13) и придадим какие-то значения переменным x_2, \dots, x_n . Мы получим многочлен от одной переменной x_1 с коэффициентами из K , обращающийся в нуль при любом значении x_1 . По теореме 1.1 все его коэффициенты равны нулю. Таким образом, каждый из многочленов $f_k \in K[x_2, \dots, x_n]$ обращается в нуль при любых значениях x_2, \dots, x_n , т. е. определяет нулевую функцию. По предположению индукции отсюда следует, что $f_k = 0$ при любом k ; но тогда и $f = 0$. \square

Иногда бывает полезно следующее более сильное утверждение, которое легко выводится из доказанной теоремы.

Следствие. Пусть поле K бесконечно и $h \in K[x_1, x_2, \dots, x_n]$ — какой-либо ненулевой многочлен. Если многочлены $f, g \in K[x_1, x_2, \dots, x_n]$ принимают одинаковые значения при всех значениях переменных x_1, x_2, \dots, x_n , при которых многочлен h не обращается в нуль, то они равны.

Доказательство. При указанных условиях многочлены fh и gh принимают одинаковые значения вообще при всех значениях переменных и, значит, $fh = gh$. Так как в алгебре многочленов нет делителей нуля (предложение 1), то отсюда следует, что $f = g$. \square

Замечание 2. Если поле K конечно, то теорема и ее доказательство тем не менее остаются в силе для многочленов, степень которых по каждому из переменных меньше числа элементов поля K (см. замечание 1.5).

Задача 2. Доказать, что если поле K содержит q элементов, то функции, определяемые одночленами $x_1^{k_1} \dots x_n^{k_n}$ с $k_1, \dots, k_n < q$, составляют базис в пространстве всех функций на K^n со значениями в K .

При $n > 1$ члены многочлена от n переменных нельзя, вообще говоря, однозначно упорядочить по их степеням, поскольку может быть несколько членов одинаковой степени. Между тем какое-то упорядочение иногда бывает полезно. В этих случаях обычно используют *лексикографическое* (т. е. подобное упорядочению слов в словаре) *упорядочение*, при котором вначале сравниваются показатели при x_1 , затем, если они равны, показатели при x_2 и т. д. Если одночлен u лексикографически старше одночлена v , то мы будем писать $u \succ v$. Согласно определению, это означает, что первая из переменных, которая входит в u и v с разными показателями, входит в u с бóльшим показателем, чем в v .

Предложение 2. *Отношение лексикографического упорядочения одночленов обладает следующими свойствами:*

- 1) *если $u \succ v$ и $v \succ w$, то $u \succ w$ (транзитивность);*
- 2) *если $u \succ v$, то $uw \succ vw$ для любого одночлена w ;*
- 3) *если $u_1 \succ v_1$ и $u_2 \succ v_2$, то $u_1 u_2 \succ v_1 v_2$.*

Первое из этих свойств, собственно, и дает основание называть отношение « \succ » упорядочением.

Доказательство. 1) Пусть первая переменная, которая не входит во все одночлены u, v, w с одним и тем же показателем, входит в них с показателями k, l, m соответственно. Тогда

$$k \geq l \geq m,$$

причем хотя бы в одном из двух случаев имеет место строгое неравенство. Следовательно, $k > m$, а это и означает, что $u \succ w$.

2) При умножении на w к показателям, с которыми каждая из переменных входит в u и v , добавляется одно и то же число, и знак неравенства (или равенства) между этими показателями не меняется, а только эти неравенства и имеют значение при сравнении одночленов.

3) Пользуясь предыдущим свойством, получаем

$$u_1 u_2 \succ v_1 u_2 \succ v_1 v_2. \quad \square$$

Пример 1. Следующий многочлен расположен по лексикографическому убыванию членов:

$$x_1^2 x_2 + x_1 x_2^2 x_3 + 2x_1 x_3^2 + x_2 x_3^3 - x_2 x_3^2 + 3.$$

Обратите внимание на то, что член $x_1 x_2^2 x_3$ лексикографически младше $x_1^2 x_2$, хотя его степень больше.

Среди ненулевых членов любого ненулевого многочлена $f \in K[x_1, x_2, \dots, x_n]$ имеется единственный, который лексикографически старше всех остальных. Он называется *старшим членом* многочлена f .

Предложение 3. *Старший член произведения ненулевых многочленов равен произведению их старших членов.*

Доказательство. Достаточно доказать это утверждение для двух многочленов. Пусть f_1, f_2 — ненулевые многочлены, u_1, u_2 — их старшие члены и v_1, v_2 — какие-то их члены. Если $v_1 \neq u_1$ или $v_2 \neq u_2$, то в силу предложения 2

$$u_1 u_2 \succ v_1 v_2.$$

Следовательно, после приведения подобных членов в произведении $f_1 f_2$ произведение $u_1 u_2$ сохранится в качестве ненулевого члена, который старше всех остальных. \square

§ 8. Симметрические многочлены

Определение 1. Многочлен $f \in K[x_1, x_2, \dots, x_n]$ называется *симметрическим*, если он не изменяется ни при каких перестановках переменных.

Так как любая перестановка может быть осуществлена путем последовательных перестановок двух элементов, то многочлен является симметрическим, если он не изменяется при перестановке любых двух переменных.

Очевидно, что каждая однородная компонента симметрического многочлена также является симметрическим многочленом.

Пример 1. *Степенные суммы*

$$s_k = x_1^k + x_2^k + \dots + x_n^k \quad (k = 1, 2, \dots),$$

очевидно, являются симметрическими многочленами.

Пример 2. Следующие симметрические многочлены называются *элементарными симметрическими многочленами*:

$$\sigma_1 = x_1 + x_2 + \dots + x_n,$$

$$\sigma_2 = x_1 x_2 + x_1 x_3 + \dots + x_{n-1} x_n,$$

.....

$$\sigma_k = \sum_{i_1 < i_2 < \dots < i_k} x_{i_1} x_{i_2} \dots x_{i_k},$$

.....

$$\sigma_n = x_1 x_2 \dots x_n.$$

Пример 3. *Определитель Вандермонда*

$$V(x_1, x_2, \dots, x_n) = \prod_{i>j} (x_i - x_j)$$

(см. пример 2.4.5), представляющий собой произведение разностей всевозможных пар переменных, при перестановках переменных может только умножиться на ± 1 за счет того, что в некоторых случаях уменьшаемое и вычитаемое поменяются ролями. Число таких слу-

чаев равно числу инверсий в соответствующей перестановке. Следовательно,

$$V(x_{k_1}, x_{k_2}, \dots, x_{k_n}) = \operatorname{sgn}(k_1, k_2, \dots, k_n)V(x_1, x_2, \dots, x_n).$$

Таким образом, сам определитель Вандермонда не является симметрическим многочленом, но таковым является его квадрат

$$V(x_1, x_2, \dots, x_n)^2 = \prod_{i>j} (x_i - x_j)^2.$$

Пример 4. При любых перестановках переменных x_1, x_2, x_3, x_4 многочлены

$$h_1 = x_1x_2 + x_3x_4, \quad h_2 = x_1x_3 + x_2x_4, \quad h_3 = x_1x_4 + x_2x_3$$

переставляются между собой. Поэтому любой симметрический многочлен от них будет симметрическим многочленом от x_1, x_2, x_3, x_4 . В частности, таковым является их произведение

$$h_1h_2h_3 = (x_1x_2 + x_3x_4)(x_1x_3 + x_2x_4)(x_1x_4 + x_2x_3).$$

Задача 1. Доказать, что многочлен

$$(x_1 + x_2 - x_3 - x_4)(x_1 - x_2 + x_3 - x_4)(x_1 - x_2 - x_3 + x_4)$$

является симметрическим.

Симметрические многочлены находят применение в исследовании алгебраических уравнений с одним неизвестным благодаря формулам Виета (см. § 2), которые выражают элементарные симметрические многочлены от корней алгебраического уравнения через его коэффициенты (при условии, что число корней уравнения в рассматриваемом поле равно его степени). Ясно, что только симметрические многочлены от корней уравнения однозначно определены: значение любого другого многочлена, вообще говоря, зависит от нумерации корней. С другой стороны, мы покажем, что любой симметрический многочлен от корней алгебраического уравнения может быть выражен через коэффициенты этого уравнения.

Пример 5. Многочлен $s_2 = x_1^2 + x_2^2 + \dots + x_n^2$ является симметрическим. Легко видеть, что

$$s_2 = \sigma_1^2 - 2\sigma_2. \quad (15)$$

Поэтому сумма квадратов корней алгебраического уравнения

$$x^n + a_1x^{n-1} + a_2x^{n-2} + \dots + a_{n-1}x + a_n = 0$$

равна $a_1^2 - 2a_2$.

Очевидно, что сумма и произведение симметрических многочленов, а также произведение симметрического многочлена на число являются симметрическими многочленами. Иными словами, симметрические многочлены образуют подалгебру в алгебре всех многочленов.

Следовательно, если $F \in K[X_1, X_2, \dots, X_m]$ — произвольный многочлен от m переменных и $f_1, f_2, \dots, f_m \in K[x_1, x_2, \dots, x_n]$ — какие-то симметрические многочлены, то $F(f_1, f_2, \dots, f_m)$ — также симметрический многочлен от x_1, x_2, \dots, x_n . Естественно поставить вопрос, нельзя ли найти такие симметрические многочлены f_1, f_2, \dots, f_m , чтобы всякий симметрический многочлен можно было выразить через них указанным способом. Оказывается, что в качестве таких многочленов можно взять элементарные симметрические многочлены $\sigma_1, \sigma_2, \dots, \sigma_n$.

Теорема 1. *Всякий симметрический многочлен единственным образом представляется в виде многочлена от элементарных симметрических многочленов.*

Доказательству теоремы предположим две леммы.

Лемма 1. *Пусть $u = ax_1^{k_1} x_2^{k_2} \dots x_n^{k_n}$ — старший член симметрического многочлена f . Тогда*

$$k_1 \geq k_2 \geq \dots \geq k_n. \quad (16)$$

Доказательство. Предположим, что $k_i < k_{i+1}$ для некоторого i . Наряду с членом u многочлен f должен содержать член

$$u' = ax_1^{k_1} \dots x_i^{k_{i+1}} x_{i+1}^{k_i} \dots x_n^{k_n},$$

получающийся из u перестановкой x_i и x_{i+1} . Легко видеть, что $u' \succ u$. Это противоречит тому, что u — старший член многочлена f . \square

Лемма 2. *Для любого одночлена $u = x_1^{k_1} x_2^{k_2} \dots x_n^{k_n}$, показатели которого удовлетворяют неравенствам (16), существуют такие неотрицательные целые числа l_1, l_2, \dots, l_n , что старший член многочлена $\sigma_1^{l_1} \sigma_2^{l_2} \dots \sigma_n^{l_n}$ совпадает с u . Числа l_1, l_2, \dots, l_n определены этим условием однозначно.*

Доказательство. Старший член многочлена σ_k равен $x_1 x_2 \dots x_k$. В силу предложения 7.3 старший член многочлена $\sigma_1^{l_1} \sigma_2^{l_2} \dots \sigma_n^{l_n}$ равен

$$x_1^{l_1} (x_1 x_2)^{l_2} \dots (x_1 x_2 \dots x_n)^{l_n} = x_1^{l_1 + l_2 + \dots + l_n} x_2^{l_2 + \dots + l_n} \dots x_n^{l_n}.$$

Приравнявая его одночлену u , получаем систему линейных уравнений

$$\begin{cases} l_1 + l_2 + \dots + l_n = k_1, \\ l_2 + \dots + l_n = k_2, \\ \dots\dots\dots \\ l_n = k_n, \end{cases}$$

которая, очевидно, имеет единственное решение

$$l_i = k_i - k_{i+1} \quad (i = 1, 2, \dots, n-1), \quad l_n = k_n. \quad (17)$$

Из условия леммы следует, что определенные таким образом числа l_1, l_2, \dots, l_n неотрицательны. \square

Замечание 1. Уравнение $l_1 + l_2 + \dots + l_n = k_1$ показывает, что степень одночлена $X_1^{l_1} X_2^{l_2} \dots X_n^{l_n}$ по совокупности переменных равна степени одночлена u по x_1 .

Доказательство теоремы 1. Пусть $f \in K[x_1, x_2, \dots, x_n]$ — симметрический многочлен. Нам нужно найти такой многочлен $F \in K[X_1, X_2, \dots, X_n]$, что

$$F(\sigma_1, \sigma_2, \dots, \sigma_n) = f.$$

Если $f = 0$, то можно взять $F = 0$. В противном случае пусть $u_1 = \alpha x_1^{k_1} x_2^{k_2} \dots x_n^{k_n}$ — старший член многочлена f . По лемме 1 выполняются неравенства (16). По лемме 2 существует такой одночлен $F_1 \in K[X_1, X_2, \dots, X_n]$, что старший член многочлена $F_1(\sigma_1, \sigma_2, \dots, \sigma_n)$ равен u_1 . Рассмотрим симметрический многочлен

$$f_1 = f - F_1(\sigma_1, \sigma_2, \dots, \sigma_n).$$

Если $f_1 = 0$, то можно взять $F = F_1$. В противном случае пусть u_2 — старший член многочлена f_1 . Ясно, что он младше, чем u_1 . Существует такой одночлен $F_2 \in K[X_1, X_2, \dots, X_n]$, что старший член многочлена $F_2(\sigma_1, \sigma_2, \dots, \sigma_n)$ равен u_2 . Рассмотрим симметрический многочлен

$$f_2 = f_1 - F_2(\sigma_1, \sigma_2, \dots, \sigma_n).$$

Если $f_2 = 0$, то можно взять $F = F_1 + F_2$. В противном случае, продолжая процесс, получаем последовательность симметрических многочленов f, f_1, f_2, \dots , старшие члены которых удовлетворяют неравенствам

$$u_1 \succ u_2 \succ \dots$$

По лемме 1 показатель при любой переменной в любом из одночленов u_m не превосходит показателя при x_1 в этом одночлене, а он, в свою очередь, не превосходит k_1 . Поэтому для наборов показателей одночленов u_m имеется лишь конечное число возможностей, так что описанный выше процесс должен оборваться. Это означает, что $f_M = 0$ для некоторого M . В качестве F можно тогда взять $F_1 + F_2 + \dots + F_M$.

Докажем теперь, что многочлен F определен однозначно. Предположим, что F и G — такие многочлены, что

$$F(\sigma_1, \sigma_2, \dots, \sigma_n) = G(\sigma_1, \sigma_2, \dots, \sigma_n).$$

Рассмотрим их разность $H = F - G$. Тогда

$$H(\sigma_1, \sigma_2, \dots, \sigma_n) = 0.$$

Нам нужно доказать, что $H = 0$. Предположим, что это не так, и пусть H_1, H_2, \dots, H_s — все ненулевые члены многочлена H . Обозначим через w_i ($i = 1, 2, \dots, s$) старший член многочлена

$$H_i(\sigma_1, \sigma_2, \dots, \sigma_n) \in K[x_1, x_2, \dots, x_n].$$

В силу леммы 2 среди одночленов w_1, w_2, \dots, w_s нет пропорциональных. Выберем из них старший. Пусть это будет w_1 . По построению одночлен w_1 старше всех остальных членов многочлена $H_1(\sigma_1, \sigma_2, \dots, \sigma_n)$ и всех членов многочленов $H_i(\sigma_1, \sigma_2, \dots, \sigma_n)$ ($i = 2, \dots, s$). Поэтому после приведения подобных членов в сумме

$$\begin{aligned} H_1(\sigma_1, \sigma_2, \dots, \sigma_n) + H_2(\sigma_1, \sigma_2, \dots, \sigma_n) + \dots + H_s(\sigma_1, \sigma_2, \dots, \sigma_n) = \\ = H(\sigma_1, \sigma_2, \dots, \sigma_n) \end{aligned}$$

член w_1 сохранится, так что эта сумма не будет равна нулю, что противоречит нашему предположению. \square

Замечание 2. Согласно замечанию 1, для любого m

$$\deg F_m = \deg_{\mathfrak{G}_{x_1}} u_m \leq \deg_{\mathfrak{G}_{x_1}} u_1 = \deg_{\mathfrak{G}_{x_1}} f (= k_1).$$

Следовательно,

$$\deg F = \deg_{\mathfrak{G}_{x_1}} f. \quad (18)$$

Следуя доказательству этой теоремы, можно в принципе найти выражение любого конкретного симметрического многочлена через $\sigma_1, \sigma_2, \dots, \sigma_n$.

Пример 6. Выразим через $\sigma_1, \sigma_2, \dots, \sigma_n$ многочлен

$$f = s_3 = x_1^3 + x_2^3 + \dots + x_n^3.$$

Представим вычисления в виде таблицы.

m	u_m	$F_m(\sigma_1, \sigma_2, \dots, \sigma_n)$	f_m
1	x_1^3	$\sigma_1^3 = \sum_i x_i^3 + 3 \sum_{i \neq j} x_i^2 x_j + 6 \sum_{i < j < k} x_i x_j x_k$	$-3 \sum_{i \neq j} x_i^2 x_j - 6 \sum_{i < j < k} x_i x_j x_k$
2	$-3x_1^2 x_2$	$-3\sigma_1 \sigma_2 = -3 \sum_{i \neq j} x_i^2 x_j - 9 \sum_{i < j < k} x_i x_j x_k$	$3 \sum_{i < j < k} x_i x_j x_k$
3	$3x_1 x_2 x_3$	$3\sigma_3 = 3 \sum_{i < j < k} x_i x_j x_k$	0

Таким образом,

$$s_3 = \sigma_1^3 - 3\sigma_1 \sigma_2 + 3\sigma_3 \quad (19)$$

На практике для однородных симметрических многочленов удобнее применять другой способ, который мы поясним на следующем примере.

Пример 7. Выразим через $\sigma_1, \sigma_2, \sigma_3, \sigma_4$ многочлен

$$f = (x_1 x_2 + x_3 x_4)(x_1 x_3 + x_2 x_4)(x_1 x_4 + x_2 x_3)$$

из примера 4. В обозначениях доказательства теоремы 1 имеем $u_1 = x_1^3 x_2 x_3 x_4$. Не производя вычислений, можно найти с точностью до коэффициентов возможных кандидатов на роль одночленов u_2, u_3, \dots . Во-первых, их показатели должны удовлетворять неравенствам леммы 1. Во-вторых, поскольку f — однородный многочлен степени 6, сумма их показателей должна равняться 6. В-третьих, они должны быть младше u_1 . Выпишем в таблицу все наборы показателей одночленов, удовлетворяющих этим условиям, в порядке лексикографического убывания, начиная с набора показателей одночлена u_1 , а справа выпишем соответствующие произведения элементарных симметрических многочленов, найденные по формулам (17):

3	1	1	1	$\sigma_1^2 \sigma_4$
2	2	2	0	σ_3^2
2	2	1	1	$\sigma_2 \sigma_4$

Итак, мы можем утверждать, что

$$f = \sigma_1^2 \sigma_4 + a \sigma_3^2 + b \sigma_2 \sigma_4.$$

Для того чтобы найти коэффициенты a и b , будем придавать в этом равенстве переменным x_1, x_2, x_3, x_4 какие-нибудь выбранные значения. Представим вычисления в виде таблицы, в правом столбце которой будем выписывать получаемые уравнения:

x_1	x_2	x_3	x_4	σ_1	σ_2	σ_3	σ_4	f	
1	1	1	0	3	3	1	0	1	$a = 1$
1	1	-1	-1	0	-2	0	1	8	$-2b = 8$

Таким образом, $a = 1$ и $b = -4$, так что

$$f = \sigma_1^2 \sigma_4 + \sigma_3^2 - 4 \sigma_2 \sigma_4.$$

В случае неоднородного симметрического многочлена этот способ можно применить к каждой его однородной компоненте и полученные выражения сложить.

Замечание 3. Изложенная теория без всяких изменений переносится на более общий случай, когда K — произвольное коммутативное ассоциативное кольцо с единицей. Так, в случае $K = \mathbb{Z}$ получается следующий результат: всякий симметрический многочлен с целыми коэффициентами представляется в виде многочлена с целыми коэффициентами от элементарных симметрических многочленов.

Доказанная теорема в сочетании с формулами Виета позволяет найти любой симметрический многочлен от корней заданного алгебраического уравнения. А именно, пусть $f \in K[x_1, x_2, \dots, x_n]$ — симметрический многочлен и $F \in K[X_1, X_2, \dots, X_n]$ — такой многочлен, что

$$f = F(\sigma_1, \sigma_2, \dots, \sigma_n).$$

Пусть, далее, c_1, c_2, \dots, c_n — корни алгебраического уравнения

$$a_0 x^n + a_1 x^{n-1} + \dots + a_{n-1} x + a_n = 0 \quad (a_0 \neq 0).$$

Тогда

$$f(c_1, c_2, \dots, c_n) = F\left(-\frac{a_1}{a_0}, \frac{a_2}{a_0}, \dots, (-1)^n \frac{a_n}{a_0}\right). \quad (20)$$

Замечание 4. Пусть $\deg_{x_1} f = k$. Тогда $\deg F = k$ (см. замечание 2) и, домножив равенство (20) на a_0^k , мы получим в правой части однородный многочлен степени k от $a_0, a_1, a_2, \dots, a_n$.

Пример 8. Пусть c_1, c_2, c_3, c_4 — корни уравнения

$$x^4 + px^2 + qx + r = 0. \quad (21)$$

Найдем уравнение 3-й степени, корнями которого являются числа

$$d_1 = c_1c_2 + c_3c_4, \quad d_2 = c_1c_3 + c_2c_4, \quad d_3 = c_1c_4 + c_2c_3.$$

Запишем его в виде

$$y^3 + a_1y^2 + a_2y + a_3 = 0.$$

Согласно формулам Виета

$$a_1 = -(d_1 + d_2 + d_3), \quad a_2 = d_1d_2 + d_1d_3 + d_2d_3, \quad a_3 = -d_1d_2d_3.$$

Имеем $d_i = h_i(c_1, c_2, c_3, c_4)$, где h_1, h_2, h_3 — многочлены из примера 4. Находим:

$$\begin{aligned} h_1 + h_2 + h_3 &= \sigma_2, \\ h_1h_2 + h_1h_3 + h_2h_3 &= \sum_{\substack{i \neq j, k \\ j < k}} x_i^2 x_j x_k = \sigma_1 \sigma_3 - 4\sigma_4, \\ h_1h_2h_3 &= \sigma_1^2 \sigma_4 + \sigma_3^2 - 4\sigma_2 \sigma_4. \end{aligned}$$

(Последнее равенство есть результат примера 7.) По формулам Виета

$$\begin{aligned} \sigma_1(c_1, c_2, c_3, c_4) &= 0, \\ \sigma_2(c_1, c_2, c_3, c_4) &= p, \\ \sigma_3(c_1, c_2, c_3, c_4) &= -q, \\ \sigma_4(c_1, c_2, c_3, c_4) &= r. \end{aligned}$$

Следовательно,

$$a_1 = -p, \quad a_2 = -4r, \quad a_3 = 4pr - q^2,$$

т. е. искомое уравнение имеет вид

$$y^3 - py^2 - 4ry + (4pr - q^2) = 0. \quad (22)$$

Задача 2. В обозначениях предыдущего примера доказать, что

$$\begin{aligned} (c_1 + c_2 - c_3 - c_4)^2 &= 4(d_1 - p), \\ (c_1 - c_2 + c_3 - c_4)^2 &= 4(d_2 - p), \\ (c_1 - c_2 - c_3 + c_4)^2 &= 4(d_3 - p) \end{aligned}$$

и, кроме того,

$$(c_1 + c_2 - c_3 - c_4)(c_1 - c_2 + c_3 - c_4)(c_1 - c_2 - c_3 + c_4) = -8q \quad (23)$$

(см. задачу 1).

Пользуясь результатами этой задачи, можно свести решение уравнения (21) к решению уравнения (22) (при условии что $\text{char } K \neq 2$). А именно, складывая с подходящими знаками равенства

$$\begin{aligned} c_1 + c_2 + c_3 + c_4 &= 0, \\ c_1 + c_2 - c_3 - c_4 &= 2\sqrt{d_1 - p}, \\ c_1 - c_2 + c_3 - c_4 &= 2\sqrt{d_2 - p}, \\ c_1 - c_2 - c_3 + c_4 &= 2\sqrt{d_3 - p}, \end{aligned}$$

получаем

$$c_{1,2,3,4} = \frac{1}{2} \left(\pm \sqrt{d_1 - p} \pm \sqrt{d_2 - p} \pm \sqrt{d_3 - p} \right),$$

где число минусов должно быть четно. Исходные значения квадратных корней здесь следует выбирать таким образом, чтобы их произведение равнялось $-q$ (см. формулу (23)).

Уравнение (22) называется *кубической резольвентой* уравнения (21).

§ 9. Кубические уравнения

При решении квадратных уравнений ключевую роль играет дискриминант. По его обращению в нуль можно судить о наличии кратного корня, а по его знаку (в случае поля вещественных чисел) — о числе вещественных корней.

Выясним смысл дискриминанта $D(\varphi)$ квадратного трехчлена

$$\varphi = a_0x^2 + a_1x + a_2 \in \mathbb{C}[x].$$

Пусть c_1, c_2 — корни этого трехчлена. Тогда

$$\begin{aligned} D(\varphi) &= a_1^2 - 4a_0a_2 = a_0^2 \left[\left(\frac{a_1}{a_0} \right)^2 - \frac{4a_2}{a_0} \right] = \\ &= a_0^2 [(c_1 + c_2)^2 - 4c_1c_2] = a_0^2 (c_1 - c_2)^2. \end{aligned}$$

В случае когда $a_0, a_1, a_2 \in \mathbb{R}$, полученная формула хорошо объясняет ту связь между дискриминантом и свойствами корней, о кото-

рой говорилось выше. А именно, имеются следующие три возможности:

- 1) $c_1, c_2 \in \mathbb{R}$, $c_1 \neq c_2$; тогда $c_1 - c_2$ — отличное от нуля вещественное число и $D(\varphi) > 0$;
- 2) $c_1 = c_2 \in \mathbb{R}$; тогда $c_1 - c_2 = 0$ и $D(\varphi) = 0$;
- 3) $c_1 = \bar{c}_2 \notin \mathbb{R}$; тогда $c_1 - c_2$ — отличное от нуля чисто мнимое число и $D(\varphi) < 0$.

Что еще более важно, эта формула подсказывает, как можно определить дискриминант любого многочлена

$$\varphi = a_0 x^n + a_1 x^{n-1} + \dots + a_{n-1} x + a_n \in K[x] \quad (a_0 \neq 0).$$

Предположим вначале, что многочлен φ имеет n корней $c_1, c_2, \dots, c_n \in K$. Определим тогда его *дискриминант* $D(\varphi)$ по формуле

$$D(\varphi) = a_0^{2n-2} \prod_{i>j} (c_i - c_j)^2. \quad (24)$$

(Показатель при a_0 не так важен; почему мы выбрали его именно таким, будет ясно из дальнейшего.)

Иными словами, $D(\varphi)$ есть умноженное на a_0^{2n-2} значение симметрического многочлена (см. пример 8.3)

$$f = \prod_{i>j} (x_i - x_j)^2$$

от корней многочлена φ . Описанная в § 8 процедура позволяет выразить $D(\varphi)$ через коэффициенты многочлена φ . Так как

$$\deg_{x_1} f = 2n - 2,$$

то в силу замечания 8.4 это выражение будет представлять собой некоторый однородный многочлен Δ степени $2n - 2$ от a_0, a_1, \dots, a_n :

$$D(\varphi) = \Delta(a_0, a_1, \dots, a_n). \quad (25)$$

Для нахождения многочлена Δ нет необходимости знать, что многочлен φ имеет n корней в K . Это позволяет определить дискриминант любого многочлена φ по формуле (25).

Замечание 1. Так как f имеет целые коэффициенты, то и Δ имеет целые коэффициенты (см. замечание 8.3).

Замечание 2. Можно доказать (см. теорему 9.5.6), что для любого многочлена $\varphi \in K[x]$ степени n существует расширение L поля K , в котором φ имеет n корней. (Например, если $K = \mathbb{R}$, то можно взять $L = \mathbb{C}$.) Так как

описанная выше процедура вычисления дискриминанта не зависит от того, над каким полем рассматривается многочлен φ (лишь бы его коэффициенты лежали в этом поле), то для $D(\varphi)$ будет справедлива формула (24), если в качестве c_1, c_2, \dots, c_n взять корни многочлена φ в поле L .

Из определения (24) дискриминанта ясно, что многочлен $\varphi \in \mathbb{C}[x]$ имеет кратные корни тогда и только тогда, когда $D(\varphi) = 0$. Это показывает, что наличие кратных корней является исключительным обстоятельством: если выбирать коэффициенты многочлена наудачу, то вероятность того, что он будет иметь кратные корни, равна нулю.

Пусть теперь φ — кубический многочлен с вещественными коэффициентами и c_1, c_2, c_3 — его комплексные корни. Тогда

$$D(\varphi) = a_0^4 (c_1 - c_2)^2 (c_1 - c_3)^2 (c_2 - c_3)^2.$$

Имеются следующие три возможности (с точностью до перенумерации корней):

- 1) c_1, c_2, c_3 — различные вещественные числа; тогда $D(\varphi) > 0$;
- 2) $c_1, c_2, c_3 \in \mathbb{R}, c_2 = c_3$; тогда $D(\varphi) = 0$;
- 3) $c_1 \in \mathbb{R}, c_2 = \bar{c}_3 \notin \mathbb{R}$; тогда

$$D(\varphi) = a_0^4 [(c_1 - c_2)(c_1 - \bar{c}_2)]^2 (c_2 - \bar{c}_2)^2 = a_0^4 |c_1 - c_2|^4 (c_2 - \bar{c}_2)^2 < 0.$$

Таким образом, мы приходим к тому же выводу, что и в случае квадратного трехчлена: все корни многочлена φ вещественны тогда и только тогда, когда $D(\varphi) \geq 0$.

Задача 1. Доказать, что если φ — многочлен любой степени с вещественными коэффициентами, не имеющий кратных комплексных корней, то

$$\operatorname{sgn} D(\varphi) = (-1)^t,$$

где t — число пар комплексно-сопряженных мнимых корней многочлена φ .

Мы найдем теперь явное выражение дискриминанта кубического многочлена через его коэффициенты, но перед этим сделаем некоторые общие замечания, позволяющие упростить вычисления.

Любой многочлен можно нормировать, разделив на старший коэффициент, что не изменит его корней. Далее, любой нормированный многочлен

$$\varphi = x^n + a_1 x^{n-1} + a_2 x^{n-2} + \dots + a_{n-1} x + a_n$$

над полем нулевой характеристики (или, более общо, над полем, характеристика которого не делит n) с помощью замены

$$x = y - \frac{a_1}{n}$$

приводится к многочлену

$$\psi = y^n + b_2 y^{n-2} + \dots + b_{n-1} y + b_n,$$

в котором коэффициент при y^{n-1} равен нулю. Многочлен такого вида называется *неполным*. При $n = 2$ именно таким способом получается формула решения квадратного уравнения. При $n > 2$ эта замена не решает дела, но, во всяком случае, может упростить задачу.

Найдем дискриминант неполного кубического многочлена

$$\varphi = x^3 + px + q. \quad (26)$$

Следуя способу, изложенному в примере 8.7, будем искать выражение симметрического многочлена

$$f = (x_1 - x_2)^2 (x_1 - x_3)^2 (x_2 - x_3)^2$$

через элементарные симметрические многочлены $\sigma_1, \sigma_2, \sigma_3$. Многочлен f является однородным степени 6, и его старший член равен $x_1^4 x_2^2$. Выпишем наборы показателей старших членов симметрических многочленов, которые могут встретиться в процессе, описанном в доказательстве теоремы 8.1, и соответствующие им произведения элементарных симметрических многочленов:

4	2	0	$\sigma_1^2 \sigma_2^2$
4	1	1	$\sigma_1^3 \sigma_3$
3	3	0	σ_2^3
3	2	1	$\sigma_1 \sigma_2 \sigma_3$
2	2	2	σ_3^2

Мы видим, что

$$f = \sigma_1^2 \sigma_2^2 + a \sigma_1^3 \sigma_3 + b \sigma_2^3 + c \sigma_1 \sigma_2 \sigma_3 + d \sigma_3^2. \quad (27)$$

Для вычисления $D(\varphi)$ мы должны будем сделать в выражении (27) подстановку

$$\sigma_1 = 0, \quad \sigma_2 = p, \quad \sigma_3 = -q.$$

Поэтому коэффициенты a и c не будут влиять на окончательный результат, и мы можем их не находить.

Для нахождения b и d будем в равенстве (27) придавать переменным x_1, x_2, x_3 значения, указанные в следующей таблице, в правом столбце которой выписаны получаемые при этом уравнения:

x_1	x_2	x_3	σ_1	σ_2	σ_3	f	
1	-1	0	0	-1	0	4	$-b=4$
2	-1	-1	0	-3	2	0	$-27b+4d=0$

Таким образом, $b = -4$, $d = -27$ и

$$D(\varphi) = -4p^3 - 27q^2. \quad (28)$$

Пример 1. Найдем число вещественных корней многочлена

$$\varphi = x^3 - 0,3x^2 - 4,3x + 3,9.$$

С помощью замены

$$y = x - 0,1$$

приводим его к неполному многочлену (коэффициенты которого могут быть найдены по схеме Горнера, как в примере 2.3)

$$\psi = y^3 - 4,33y + 3,468.$$

Теперь

$$D(\varphi) = D(\psi) = 4 \cdot 4,33^3 - 27 \cdot 3,468^2 = 0,0013 > 0.$$

Следовательно, многочлен φ имеет 3 различных вещественных корня.

Замечание 3. Дискриминант кубического многочлена общего вида

$$\varphi = a_0x^3 + a_1x^2 + a_2x + a_3$$

равен

$$D(\varphi) = a_1^2a_2^2 - 4a_1^3a_3 - 4a_0a_2^3 + 18a_0a_1a_2a_3 - 27a_0^2a_3^2.$$

Изложим теперь способ решения кубического уравнения.

Предположим, что основное поле K содержит нетривиальный (т. е. отличный от 1) кубический корень из единицы, скажем, ω . Тогда $1, \omega, \omega^{-1}$ — это все кубические корни из единицы, и по формуле Виета получаем

$$\omega + \omega^{-1} = -1. \quad (29)$$

Рассмотрим линейные многочлены

$$h_1 = x_1 + \omega x_2 + \omega^{-1} x_3, \quad h_2 = x_1 + \omega^{-1} x_2 + \omega x_3.$$

При перестановке x_2 и x_3 они меняются местами, а при перестановке x_1 и x_2 многочлен h_1 переходит в ωh_2 , а h_2 — в $\omega^{-1} h_1$. Отсюда следует, что многочлены

$$f = h_1^3 + h_2^3, \quad g = h_1 h_2$$

являются симметрическими. Выражая их через элементарные симметрические многочлены, получаем

$$f = 2\sigma_1^3 - 9\sigma_1\sigma_2 + 27\sigma_3, \quad g = \sigma_1^2 - 3\sigma_2.$$

Пусть теперь c_1, c_2, c_3 — корни многочлена (26). Положим

$$d_1 = c_1 + \omega c_2 + \omega^{-1} c_3, \quad d_2 = c_1 + \omega^{-1} c_2 + \omega c_3.$$

Из предыдущего следует, что

$$d_1^3 + d_2^3 = -27q, \quad d_1 d_2 = -3p$$

и, значит,

$$d_1^3 d_2^3 = -27p^3.$$

Таким образом, d_1^3 и d_2^3 — это корни квадратного уравнения

$$x^2 + 27qx - 27p^3 = 0.$$

Решая его, находим

$$d_1^3 = 27 \left(-\frac{q}{2} + \sqrt{\frac{p^3}{27} + \frac{q^2}{4}} \right), \quad (30)$$

$$d_2^3 = 27 \left(-\frac{q}{2} - \sqrt{\frac{p^3}{27} + \frac{q^2}{4}} \right). \quad (31)$$

Заметим, что выражение, стоящее под знаком радикала, лишь множителем $-\frac{1}{108}$ отличается от дискриминанта многочлена (26).

Складывая равенства

$$\begin{aligned} c_1 + c_2 + c_3 &= 0, \\ c_1 + \omega c_2 + \omega^{-1} c_3 &= d_1, \\ c_1 + \omega^{-1} c_2 + \omega c_3 &= d_2, \end{aligned}$$

с учетом соотношения (29) получаем

$$c_1 = \frac{1}{3}(d_1 + d_2).$$

Поскольку нумерация корней может быть произвольной, эта формула на самом деле дает все три корня, если в качестве d_1 и d_2 выбирать всевозможные значения кубических корней из выражений (30) и (31), связанные полученным выше соотношением

$$d_1 d_2 = -3p. \quad (32)$$

Таким образом, мы приходим к следующей окончательной формуле

$$c_{1,2,3} = \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{p^3}{27} + \frac{q^2}{4}}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{p^3}{27} + \frac{q^2}{4}}},$$

называемой *формулой Кардано*.

Замечание 4. Формула Кардано имеет смысл, если извлекаются входящие в нее квадратные и кубические корни. В частности, если мы решаем по этой формуле кубическое уравнение с вещественными коэффициентами, то нам, вообще говоря, придется работать с комплексными числами, даже если нас интересуют только вещественные корни. Именно так обстоит дело в случае положительного дискриминанта, когда все три корня вещественны: в этом случае число, стоящее под знаком квадратного радикала, отрицательно.

Пример 2. Найдем корни многочлена ψ из примера 1. Имеем

$$\frac{p^3}{27} + \frac{q^2}{4} = -\frac{1}{108}D(\psi) \approx -0,0000120,$$

так что под знаком одного из кубических радикалов в формуле Кардано будет стоять число

$$\begin{aligned} -\frac{q}{2} + \sqrt{\frac{p^3}{27} + \frac{q^2}{4}} &\approx -1,734 + 0,00347i \approx \\ &\approx 1,73400[\cos(\pi - 0,00200) + i \sin(\pi - 0,00200)]. \end{aligned}$$

Под знаком другого кубического радикала будет стоять комплексно-сопряженное число. Условие (32) означает в данном случае, что при извлечении кубических корней следует комбинировать их комплексно-сопряженные значения. При сложении комплексно-сопряженных чисел получается их удвоенная вещественная часть.

Таким образом,

$$c_1 \approx 2 \sqrt[3]{1,73400} \cos \frac{\pi - 0,00200}{3} \approx 1,20278,$$

$$c_2 \approx 2 \sqrt[3]{1,73400} \cos \frac{\pi + 0,00200}{3} \approx 1,20001,$$

$$c_3 \approx -2 \sqrt[3]{1,73400} \cos \frac{0,00200}{3} \approx -2,40277.$$

§ 10. Поле рациональных дробей

Таким же образом, как кольцо целых чисел расширяется до поля рациональных чисел, любое целостное кольцо можно расширить до поля.

Пусть A — целостное кольцо. Рассмотрим множество пар (a, b) , где $a, b \in A$, $b \neq 0$, и определим в нем отношение эквивалентности по правилу

$$(a_1, b_1) \sim (a_2, b_2) \Leftrightarrow a_1 b_2 = a_2 b_1.$$

Рефлексивность и симметричность этого отношения очевидны; докажем его транзитивность. Если $(a_1, b_1) \sim (a_2, b_2)$ и $(a_2, b_2) \sim (a_3, b_3)$, то

$$a_1 b_2 b_3 = a_2 b_1 b_3 = a_3 b_1 b_2,$$

откуда после сокращения на b_2 получаем

$$a_1 b_3 = a_3 b_1,$$

т. е. $(a_1, b_1) \sim (a_3, b_3)$.

Из данного определения следует, что

$$(a, b) \sim (ac, bc) \tag{33}$$

для любого $c \neq 0$. С другой стороны, как показывает следующая ниже цепочка эквивалентностей, любая эквивалентность $(a_1, b_1) \sim (a_2, b_2)$ является следствием эквивалентностей типа (33):

$$(a_1, b_1) \sim (a_1 b_2, b_1 b_2) = (a_2 b_1, b_1 b_2) \sim (a_2, b_2).$$

(Мы сначала умножили оба члена пары (a_1, b_1) на b_2 , а затем сократили оба члена получившейся пары на b_1 .)

Определим теперь сложение и умножение пар по правилам

$$(a_1, b_1) + (a_2, b_2) = (a_1 b_2 + a_2 b_1, b_1 b_2),$$

$$(a_1, b_1)(a_2, b_2) = (a_1 a_2, b_1 b_2).$$

Докажем, что определенное выше отношение эквивалентности согласовано с этими операциями. В силу предыдущего достаточно показать, что при умножении обоих членов одной из пар (a_1, b_1) и (a_2, b_2) на элемент $c \neq 0$ сумма и произведение этих пар заменятся эквивалентными им парами; но очевидно, что при такой операции оба члена суммы и произведения умножатся на тот же элемент c .

Класс эквивалентности, содержащий пару (a, b) , условимся записывать как «дробь» $\frac{a}{b}$ или a/b (пока это просто символ, не подразумевающий фактического деления). Ввиду доказанного выше операции сложения и умножения пар определяют операции сложения и умножения дробей, осуществляемые по обычным правилам:

$$\frac{a_1}{b_1} + \frac{a_2}{b_2} = \frac{a_1 b_2 + a_2 b_1}{b_1 b_2}, \quad \frac{a_1}{b_1} \frac{a_2}{b_2} = \frac{a_1 a_2}{b_1 b_2}.$$

Докажем, что относительно этих операций дроби образуют поле.

Любое конечное множество дробей можно привести к общему знаменателю, а сложение дробей с одинаковыми знаменателями сводится к сложению их числителей. Поэтому сложение дробей коммутативно и ассоциативно. Дробь $\frac{0}{1}$ ($= \frac{0}{b}$ при любом $b \neq 0$) служит нулем для операции сложения дробей, а дробь $-\frac{a}{b}$ противоположна дроби $\frac{a}{b}$. Таким образом, дроби образуют абелеву группу относительно сложения.

Коммутативность и ассоциативность умножения очевидны. Следующая цепочка равенств доказывает дистрибутивность умножения дробей относительно сложения:

$$\left(\frac{a_1}{b} + \frac{a_2}{b}\right) \frac{a_3}{b_3} = \frac{(a_1 + a_2)a_3}{bb_3} = \frac{a_1 a_3 + a_2 a_3}{bb_3} = \frac{a_1}{b} \frac{a_3}{b_3} + \frac{a_2}{b} \frac{a_3}{b_3}.$$

Дробь $\frac{1}{1}$ служит единицей для операции умножения дробей, а при $a \neq 0$ дробь $\frac{b}{a}$ обратна дроби $\frac{a}{b}$.

Построенное поле называется *полем отношений* (или *полем дробей*) кольца A и обозначается через $\text{Quot } A$.

Сложение и умножение дробей вида $\frac{a}{1}$ сводятся к соответствующим операциям над их числителями. Кроме того, $\frac{a}{1} = \frac{b}{1}$ только при $a = b$. Следовательно, дроби такого вида образуют подкольцо, изоморфное A . Условившись отождествлять дробь вида $\frac{a}{1}$ с элементом a кольца A , мы получим вложение кольца A в поле $\text{Quot } A$. Далее,

поскольку

$$\frac{a}{b} \frac{b}{1} = \frac{a}{1},$$

дробь $\frac{a}{b}$ равна отношению элементов a и b кольца A в поле $\text{Quot } A$. Таким образом, обозначение $\frac{a}{b}$ можно теперь понимать содержательным образом.

В силу (33) дробь не изменится, если ее числитель и знаменатель умножить или разделить (если это возможно) на один и тот же элемент кольца A . Если A — евклидово кольцо, то путем сокращения числителя и знаменателя на их наибольший общий делитель любая дробь приводится к виду $\frac{a}{b}$, где $(a, b) = 1$. Такой вид дроби называется *несократимым*. (Допуская вольность речи, обычно говорят просто о несократимой дроби.)

Предложение 1. *Любой вид дроби над евклидовым кольцом получается из любого ее несократимого вида умножением числителя и знаменателя на один и тот же элемент.*

Доказательство. Пусть $\frac{a}{b} = \frac{a_0}{b_0}$, причем $(a_0, b_0) = 1$. Из равенства $ab_0 = a_0b$ следует, что $b_0 \mid a_0b$ и, значит, $b_0 \mid b$. Пусть $b = cb_0$; ясно, что тогда $a = ca_0$. \square

Следствие. *Несократимый вид дроби над евклидовым кольцом определен однозначно с точностью до умножения числителя и знаменателя на один и тот же обратимый элемент.*

Поле отношений кольца \mathbb{Z} целых чисел есть поле \mathbb{Q} рациональных чисел. Поле отношений кольца $K[x]$ многочленов над полем K называется *полем рациональных дробей* (или *рациональных функций*) над полем K и обозначается через $K(x)$.

Каждая рациональная дробь определяет функцию на K со значениями в K , определенную там, где ее знаменатель (в несократимой записи) не обращается в нуль. А именно, значением дроби $\frac{f}{g}$ ($f, g \in K[x]$) в точке $c \in K$ называется число $\frac{f(c)}{g(c)}$. Легко видеть, что операции сложения и умножения дробей соответствуют таким же операциям над определяемыми ими функциями в их общей области определения.

Задача 1. Доказать, что если рациональные дроби $\frac{f_1}{g_1}$ и $\frac{f_2}{g_2}$ над бесконечным полем K определяют функции, совпадающие в их общей области определения, то $\frac{f_1}{g_1} = \frac{f_2}{g_2}$.

Рациональная дробь $\frac{f}{g}$ называется *правильной*, если $\deg f < \deg g$. Очевидно, что сумма и произведение правильных дробей являются правильными дробями.

Предложение 2. *Всякая рациональная дробь единственным образом разлагается в сумму многочлена и правильной дроби.*

Доказательство. Пусть $f, g \in K[x]$, $g \neq 0$. Разделим f на g с остатком в кольце $K[x]$:

$$f = qg + r \quad (q, r \in K[x], \deg r < \deg g). \quad (34)$$

Тогда

$$\frac{f}{g} = q + \frac{r}{g}, \quad (35)$$

причем $\frac{r}{g}$ — правильная дробь.

Пусть теперь

$$\frac{f}{g} = q_1 + \frac{r_1}{g_1}$$

— какое-нибудь другое разложение дроби $\frac{f}{g}$ в сумму многочлена и правильной дроби. Тогда

$$q - q_1 = \frac{r_1}{g_1} - \frac{r}{g},$$

и мы приходим к противоречию, так как ненулевой многочлен не может равняться правильной дроби. \square

Многочлен q из равенства (35) называется *целой частью* дроби $\frac{f}{g}$.

Предложение 3. *Всякая правильная рациональная дробь вида*

$$\frac{f}{g_1 g_2 \dots g_s},$$

где g_1, g_2, \dots, g_s попарно взаимно просты, разлагается в сумму правильных дробей со знаменателями g_1, g_2, \dots, g_s .

Доказательство. Докажем это утверждение индукцией по s . Положим $g = g_1 g_2 \dots g_s$. При $s = 2$, согласно теореме 5.1, существуют такие многочлены u_1 и u_2 , что $g_1 u_1 + g_2 u_2 = f$. Разделив это равенство на g , получим

$$\frac{f}{g} = \frac{u_2}{g_1} + \frac{u_1}{g_2}.$$

Так как дробь $\frac{f}{g}$ правильная, то сумма целых частей дробей $\frac{u_2}{g_1}$ и $\frac{u_1}{g_2}$ должна быть равна нулю. Выделив их, мы получим разложение дроби $\frac{f}{g}$ в сумму правильных дробей со знаменателями g_1 и g_2 .

При $s > 2$ заметим, что многочлены g_1 и $g_2 \dots g_s$ взаимно просты, и по доказанному дробь $\frac{f}{g}$ разлагается в сумму правильных дробей со знаменателями g_1 и $g_2 \dots g_s$. В свою очередь, вторая из этих дробей по предположению индукции разлагается в сумму правильных дробей со знаменателями g_2, \dots, g_s . \square

Задача 2. Доказать, что разложение, о котором идет речь в предложении 3, единственно.

Изложим теперь теорию, используемую в математическом анализе при интегрировании рациональных функций.

Определение 1. Рациональная дробь $\frac{f}{g}$ над полем K называется *простейшей*, если $g = p^k$, где $p \in K[x]$ — неприводимый многочлен, и $\deg f < \deg p$.

В частности, всякая дробь вида

$$\frac{a}{(x-c)^k} \quad (a, c \in K)$$

является простейшей. В случае $K = \mathbb{C}$ дробями такого вида исчерпываются все простейшие дроби. В случае $K = \mathbb{R}$ имеются еще простейшие дроби вида

$$\frac{ax+b}{(x^2+px+q)^k} \quad (a, b, p, q \in \mathbb{R}),$$

где $p^2 - 4q < 0$.

Теорема 1. Всякая правильная рациональная дробь $\frac{f}{g}$ разлагается в сумму простейших дробей. Более точно, если $g = p_1^{k_1} p_2^{k_2} \dots p_s^{k_s}$ — разложение многочлена g на неприводимые множители, то дробь $\frac{f}{g}$ разлагается в сумму простейших дробей со знаменателями

$$p_1, p_1^2, \dots, p_1^{k_1}, p_2, p_2^2, \dots, p_2^{k_2}, \dots, p_s, p_s^2, \dots, p_s^{k_s}.$$

Доказательство. Ввиду предложения 3 дробь $\frac{f}{g}$ разлагается в сумму правильных дробей со знаменателями $p_1^{k_1}, p_2^{k_2}, \dots, p_s^{k_s}$. Поэтому нам достаточно доказать теорему в случае, когда $g = p^k$, где p —

неприводимый многочлен. В этом случае, разделив f на p с остатком, мы получим

$$\frac{f}{p^k} = \frac{f_1}{p^{k-1}} + \frac{r}{p^k}, \quad \deg r < \deg p.$$

Второе из слагаемых является простейшей дробью, а первое является правильной дробью как разность правильных дробей. Продолжая эту процедуру, мы в конце концов разложим дробь $\frac{f}{p^k}$ в сумму простейших дробей со знаменателями p, p^2, \dots, p^k . \square

Замечание 1. В силу задачи 2 разложение, о котором идет речь в теореме, единственно.

Пример 1. Предположим, что

$$g = (x - c_1)(x - c_2) \dots (x - c_n),$$

где c_1, c_2, \dots, c_n различны. Тогда

$$\frac{f}{g} = \frac{a_1}{x - c_1} + \frac{a_2}{x - c_2} + \dots + \frac{a_n}{x - c_n},$$

где $a_1, a_2, \dots, a_n \in K$. Для нахождения a_i умножим обе части предыдущего равенства на g и положим $x = c_i$. Мы получим тогда

$$f(c_i) = a_i(c_i - c_1) \dots (c_i - c_{i-1})(c_i - c_{i+1}) \dots (c_i - c_n) = a_i g'(c_i),$$

откуда

$$a_i = \frac{f(c_i)}{g'(c_i)}.$$

Итак,

$$\frac{f}{g} = \sum_{i=1}^n \frac{f(c_i)}{g'(c_i)(x - c_i)} \quad (36)$$

(при условии, что $\deg f < \deg g$). Интересно отметить, что, умножив обе части этого равенства на g , мы получим *интерполяционную формулу Лагранжа*

$$f = \sum_{i=1}^n b_i \frac{(x - c_1) \dots (x - c_{i-1})(x - c_{i+1}) \dots (x - c_n)}{(c_i - c_1) \dots (c_i - c_{i-1})(c_i - c_{i+1}) \dots (c_i - c_n)},$$

которая задает многочлен f степени $< n$, принимающий в точках c_1, c_2, \dots, c_n значения b_1, b_2, \dots, b_n .

Задача 3. Доказать равенство

$$\frac{1}{x^n - 1} = \frac{1}{n} \sum_{i=0}^{n-1} \frac{\varepsilon_i}{x - \varepsilon_i},$$

где $\varepsilon_0, \varepsilon_1, \dots, \varepsilon_{n-1}$ — комплексные корни n -й степени из единицы.

Задача 4. Разложить в сумму простейших дробей над полем \mathbb{Z}_p (p простое) дробь $\frac{1}{x^p - x}$.

Пример 2. Метод неопределенных коэффициентов, использованный в предыдущем примере, разумно применять и в более общей ситуации. Разложим, например, в сумму простейших дробей над \mathbb{R} рациональную дробь

$$\frac{x}{(x+1)(x^2+1)^2}.$$

Имеем, согласно теореме 1,

$$\frac{x}{(x+1)(x^2+1)^2} = \frac{a}{x+1} + \frac{bx+c}{x^2+1} + \frac{dx+e}{(x^2+1)^2},$$

где a, b, c, d, e — какие-то вещественные числа. Для их нахождения умножим предыдущее равенство на $(x+1)(x^2+1)^2$:

$$x = a(x^2+1)^2 + (bx+c)(x+1)(x^2+1) + (dx+e)(x+1).$$

Положив в этом равенстве последовательно $x = -1$ и $x = i$, получим $-1 = 4a$, $i = (di+e)(i+1) = (e-d) + (d+e)i$, откуда

$$a = -\frac{1}{4}, \quad d = e = \frac{1}{2}.$$

Далее, сравнив свободные члены и коэффициенты при x^4 , получим $0 = a + c + e$, $0 = a + b$, откуда

$$b = \frac{1}{4}, \quad c = -\frac{1}{4}.$$

Таким образом,

$$\frac{x}{(x+1)(x^2+1)^2} = -\frac{1}{4(x+1)} + \frac{x-1}{4(x^2+1)} + \frac{x+1}{2(x^2+1)^2}.$$

Начала теории групп

§ 1. Определение и примеры

В первой главе читатель познакомился с понятием абелевой группы. Абелевыми группами являются, в частности, аддитивная группа любого кольца, мультипликативная группа любого поля и аддитивная группа любого векторного пространства. Важнейшие примеры неабелевых групп появляются как группы преобразований.

Назовем *преобразованием* множества X всякое его отображение в себя.

Определение 1. *Группой преобразований* множества X называется всякая совокупность G его биективных преобразований, удовлетворяющая следующим условиям:

- 1) если $\varphi, \psi \in G$, то $\varphi\psi \in G$;
- 2) если $\varphi \in G$, то $\varphi^{-1} \in G$;
- 3) $\text{id} \in G$.

(Здесь $\varphi\psi$ обозначает произведение (композицию) преобразований φ и ψ , а id — тождественное преобразование.)

Пример 1. Совокупность $S(X)$ всех биективных преобразований множества X является группой преобразований. Если множество X бесконечно, эта группа слишком велика, чтобы быть интересной. Если X конечно, то можно считать, что $X = \{1, 2, \dots, n\}$; в этом случае группа $S(X)$ называется *группой подстановок* или *симметрической группой* степени n и обозначается через S_n . Подстановка $\sigma \in S_n$ может быть записана в виде таблицы

$$\sigma = \begin{pmatrix} i_1 & i_2 & \dots & i_n \\ j_1 & j_2 & \dots & j_n \end{pmatrix},$$

в первой строке которой выписаны в каком-то порядке числа $1, 2, \dots, n$, а во второй строке — их образы, т. е. $j_k = \sigma(i_k)$. Фиксируя расположение чисел в первой строке (например, располагая их в порядке возрастания), мы видим, что число подстановок равно числу перестановок (см. § 2.4), т. е. $n!$. При этом каждая подстановка может быть записана $n!$ способами. Приведем пример на умножение

подстановок:

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 4 & 3 & 2 & 1 \\ 2 & 1 & 4 & 3 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{pmatrix}.$$

(Здесь мы сначала для удобства переписали первую подстановку таким образом, чтобы первая строка в ее записи совпала со второй строкой в записи второй подстановки.)

Пример 2. Движения евклидовой плоскости E^2 (соответственно евклидова пространства E^3) образуют группу преобразований, обозначаемую через $\text{Isom } E^2$ (соответственно $\text{Isom } E^3$).

Замечание 1. В предыдущих главах мы обозначали через E^2 (соответственно E^3) множество векторов евклидовой плоскости (соответственно пространства). Здесь же символ E^2 (соответственно E^3) использован для обозначения самой евклидовой плоскости (соответственно пространства). Впрочем, если в плоскости (соответственно в пространстве) фиксирована некоторая точка o (которую мы будем в дальнейшем называть *началом отсчета*), то можно договориться отождествлять точки с их радиусами-векторами относительно точки o . Это соглашение часто будет подразумеваться в дальнейшем.

Замечание 2. В той версии аксиоматики евклидовой геометрии, которая берет за основу понятие расстояния между точками, движение определяется как преобразование, сохраняющее расстояния, и сформулированное в примере 2 свойство является очевидной теоремой. В другой версии, в которой понятие движения является одним из неопределяемых понятий, это свойство является аксиомой, а утверждение о том, что всякое преобразование, сохраняющее расстояния, есть движение, является (несложной) теоремой.

Пример 3. Пусть $A = (a_{ij})$ — квадратная матрица порядка n с элементами из поля K . Отображение

$$\varphi_A: K^n \rightarrow K^n, \quad x = (x_1, \dots, x_n) \mapsto y = (y_1, \dots, y_n),$$

$$y_i = \sum_{j=1}^n a_{ij} x_j \quad (i = 1, \dots, n),$$

называется *линейным преобразованием* пространства K^n . В матричной форме, если представлять x и y как столбцы, можно записать это определение как $y = Ax$. Если A и B — две матрицы, то $(AB)x = A(Bx)$, то есть

$$\varphi_{AB} = \varphi_A \varphi_B.$$

Очевидно, что φ_E — это тождественное преобразование. Поэтому, если матрица A невырождена, преобразование $\varphi_{A^{-1}}$ обратное преобразованию φ_A (откуда, в частности, следует, что φ_A биективно). Линейное преобразование, определяемое невырожденной матрицей, называется *невырожденным*. Из вышесказанного следует, что невырожденные линейные преобразования образуют группу преобразований пространства K^n . Отметим, что всякое линейное преобразование $\varphi = \varphi_A$ обладает следующими очевидными свойствами:

$$\varphi(x' + x'') = \varphi(x') + \varphi(x''), \quad \varphi(\lambda x) = \lambda \varphi(x)$$

(ср. определение линейной функции векторного аргумента в § 2.4).

Замечание 3. Легко показать (см. § 5.2), что эти свойства можно принять за определение линейного преобразования, т. е. всякое преобразование пространства K^n , обладающее этими свойствами, имеет вид φ_A для некоторой матрицы A .

Пример 4. Назовем *параллельным переносом* векторного пространства V на вектор $a \in V$ преобразование

$$t_a: x \mapsto x + a.$$

Легко видеть, что

$$t_a t_b = t_{a+b}, \quad t_a^{-1} = t_{-a}, \quad \text{id} = t_0. \quad (1)$$

Эти формулы показывают, что совокупность $\text{Trans}(V)$ всех параллельных переносов пространства V является группой преобразований.

Задача 1. Доказать, что совокупность всех возрастающих непрерывных функций на отрезке $[0, 1]$, удовлетворяющих условиям $f(0) = 0$, $f(1) = 1$, является группой преобразований отрезка $[0, 1]$.

Анализируя свойства операции умножения в группах преобразований, мы приходим к следующему понятию группы, которое отличается от понятия абелевой группы отсутствием требования коммутативности.

Определение 2. Группой называется множество G с операцией умножения, обладающей следующими свойствами:

- 1) $(ab)c = a(bc)$ для любых $a, b, c \in G$ (*ассоциативность*);
- 2) существует такой элемент $e \in G$ (*единица*), что $ae = ea = a$ для любого $a \in G$;
- 3) для всякого элемента $a \in G$ существует такой элемент $a^{-1} \in G$ (*обратный элемент*), что $aa^{-1} = a^{-1}a = e$.

Группа называется *абелевой* или *коммутативной*, если

$$ab = ba \quad \forall a, b \in G.$$

Данное определение группы использует мультипликативную терминологию. Аддитивная терминология обычно используется только для абелевых групп (хотя в принципе операция в группе может называться и обозначаться как угодно).

Аналогично тому, как это было сделано для абелевых групп, доказывается единственность единицы и обратного элемента в любой группе. Что касается деления, то в неабелевой группе следует различать левое и правое деления. А именно, для любых $a, b \in G$ уравнение $ax = b$ имеет единственное решение, равное $a^{-1}b$, а уравнение $xa = b$ имеет единственное решение, равное ba^{-1} .

В любой группе

$$(ab)^{-1} = b^{-1}a^{-1}.$$

В самом деле,

$$(ab)(b^{-1}a^{-1}) = a(bb^{-1})a^{-1} = aa^{-1} = e.$$

Всякая группа преобразований является группой относительно операции умножения преобразований. Действительно, ассоциативность этой операции известна, единицей служит тождественное преобразование, а обратным элементом — обратное преобразование.

Пример 5. Невырожденные квадратные матрицы порядка n над полем K образуют группу по умножению, обозначаемую $GL_n(K)$. Поскольку имеется взаимно однозначное соответствие между невырожденными квадратными матрицами порядка n и невырожденными линейными преобразованиями пространства K^n , причем умножению матриц соответствует умножение линейных преобразований, группа $GL_n(K)$ изоморфна группе невырожденных линейных преобразований пространства K^n . В дальнейшем мы будем иногда, говоря о группе $GL_n(K)$, рассматривать ее именно как группу линейных преобразований.

Группа $GL_n(K)$ есть группа обратимых элементов кольца $L_n(K)$ всех матриц. Если A — любое ассоциативное кольцо с единицей, то множество его обратимых элементов также является группой по умножению. Мы будем обозначать эту группу через A^* . Частным случаем является мультипликативная группа K^* поля K (состо-

ящая из всех ненулевых элементов этого поля). Заметим, что $K^* = \text{GL}_1(K)$.

Пример 6. Как показывают формулы (1), группа $\text{Trans}(V)$ изоморфна аддитивной группе пространства V .

Пример 7. Конечная группа может быть задана своей таблицей умножения. Так, множество $G = \{e, a, b, c\}$ с таблицей умножения

	e	a	b	c
e	e	a	b	c
a	a	e	c	b
b	b	c	e	a
c	c	b	a	e

является абелевой группой. В самом деле, элемент e служит ее единицей и каждый элемент обратен сам себе. Далее, легко видеть, что любая перестановка элементов a, b, c является автоморфизмом множества G с указанной операцией. Поэтому, если исключить тривиальные случаи с участием единицы и принять во внимание коммутативность, доказательство ассоциативности сводится к проверке следующих соотношений:

$$a^2b = a(ab) = b, \quad (ab)c = a(bc) = e.$$

Задача 2. Доказать, что множество $G = \{A, B, B, \Gamma, Д, E\}$ с операцией, заданной таблицей

	A	B	B	Г	Д	E
A	E	Д	Г	B	B	A
B	B	Г	Д	E	A	B
B	B	A	E	Д	Г	B
Г	Д	E	A	B	B	Г
Д	Г	B	B	A	E	Д
E	A	B	B	Г	Д	E

является группой, изоморфной S_3 .

Задача 3. Доказать, что если в множестве G с ассоциативной операцией существует такой элемент e (правая единица), что $ae = a$ для любого $a \in G$, и для любого $a \in G$ существует такой элемент a^{-1} (правый обратный элемент), что $aa^{-1} = e$, то G — группа.

Определение 3. Подгруппой группы G называется всякое подмножество $H \subset G$, удовлетворяющее следующим условиям:

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

Замечание 4. Так как $aa^{-1} = e$, условие 3) можно заменить требованием непустоты подмножества H .

Очевидно, что всякая подгруппа сама является группой относительно той же операции.

Сравнивая определения 1 и 3, мы видим, что группа преобразований множества X — это не что иное, как подгруппа группы $S(X)$.

Пример 8. Пусть f — какой-либо многочлен от n переменных. Тогда

$$\text{Sym } f = \{ \sigma \in S_n : f(x_{\sigma(1)}, x_{\sigma(2)}, \dots, x_{\sigma(n)}) = f(x_1, x_2, \dots, x_n) \}$$

есть подгруппа группы S_n . В самом деле, пусть $\sigma, \tau \in \text{Sym } f$. Положим $x_{\sigma(i)} = y_i$; тогда

$$\begin{aligned} f(x_{\sigma\tau(1)}, x_{\sigma\tau(2)}, \dots, x_{\sigma\tau(n)}) &= f(y_{\tau(1)}, y_{\tau(2)}, \dots, y_{\tau(n)}) = f(y_1, y_2, \dots, y_n) = \\ &= f(x_{\sigma(1)}, x_{\sigma(2)}, \dots, x_{\sigma(n)}) = f(x_1, x_2, \dots, x_n). \end{aligned}$$

Остальные две аксиомы подгруппы выполнены очевидным образом. В частности, многочлен f является симметрическим тогда и только тогда, когда $\text{Sym } f = S_n$. В качестве примера многочлена с менее богатой, но не тривиальной симметрией рассмотрим многочлен $f = x_1x_2 + x_3x_4$ (от 4 переменных). Легко видеть, что группа $\text{Sym } f$ состоит из 8 подстановок, сохраняющих разбиение множества $\{1, 2, 3, 4\}$ на два подмножества $\{1, 2\}$ и $\{3, 4\}$. (Допускается перестановка этих подмножеств и перестановка элементов в каждом из них; см. по этому поводу также пример 5.11).

Пример 9. Аналогично, невырожденные линейные преобразования пространства K^n , сохраняющие какой-либо заданный многочлен от n переменных, образуют подгруппу группы $\text{GL}_n(K)$. Невырожденные линейные преобразования пространства \mathbb{R}^n , сохраняющие многочлен $x_1^2 + x_2^2 + \dots + x_n^2$, называются *ортогональными преобразованиями*; они образуют подгруппу группы $\text{GL}_n(\mathbb{R})$, которая называется *ортогональной группой* и обозначается через O_n . Так как в декартовых координатах пространства E^2 (соответственно E^3) многочлен $x^2 + y^2$ (соответственно $x^2 + y^2 + z^2$) выражает квадрат

длины вектора, то преобразования из группы O_2 (соответственно O_3) могут пониматься как линейные преобразования, сохраняющие длину вектора. Дадим явное описание группы O_2 . Условие

$$\varphi = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in O_2$$

означает, что

$$(ax + by)^2 + (cx + dy)^2 = x^2 + y^2,$$

т. е.

$$a^2 + c^2 = b^2 + d^2 = 1, \quad ab + cd = 0. \quad (2)$$

Из уравнения $a^2 + c^2 = 1$ следует, что существует такой угол α , что

$$a = \cos \alpha, \quad c = \sin \alpha.$$

Оставшиеся два уравнения показывают, что

$$b = \pm \sin \alpha, \quad d = \mp \cos \alpha.$$

Таким образом,

$$\varphi = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix} \quad (3)$$

или

$$\varphi = \begin{pmatrix} \cos \alpha & \sin \alpha \\ \sin \alpha & -\cos \alpha \end{pmatrix}. \quad (4)$$

Геометрический смысл этих преобразований становится ясным, если ввести комплексную координату $z = x + yi$. Тогда в случае (3) преобразование φ есть просто умножение z на $\cos \alpha + i \sin \alpha$, т. е. прибавление к аргументу z числа α , что геометрически описывается как поворот на угол α . В случае (4) преобразование φ есть композиция поворота на угол α и зеркального отражения относительно вещественной оси (комплексного сопряжения), что, как легко видеть, есть отражение относительно прямой, составляющей с вещественной осью угол $\alpha/2$. Отметим, что в первом случае преобразование φ сохраняет ориентацию плоскости, а во втором — меняет ее.

Пример 10. Движения евклидовой плоскости, оставляющие на месте начало отсчета o , образуют подгруппу группы $\text{Isom } E^2$. Обозначим ее через H . Пусть e_1 и e_2 — координатные векторы. Из аксиом евклидовой геометрии следует, что для любых перпендикулярных единичных векторов f_1 и f_2 существует единственное движение φ ,

оставляющее точку o на месте и переводящее векторы e_1, e_2 в векторы f_1, f_2 соответственно. Если пара $\{f_1, f_2\}$ ориентирована положительно (т. е. f_2 получается из f_1 поворотом на $\pi/2$ против часовой стрелки), то φ есть поворот вокруг точки o на некоторый угол α ; в противном случае φ есть отражение относительно некоторой прямой l , проходящей через точку o (см. рис. 1). Но, как мы показали в предыдущем примере, это есть в точности ортогональные преобразования. Таким образом, $H = O_2$.

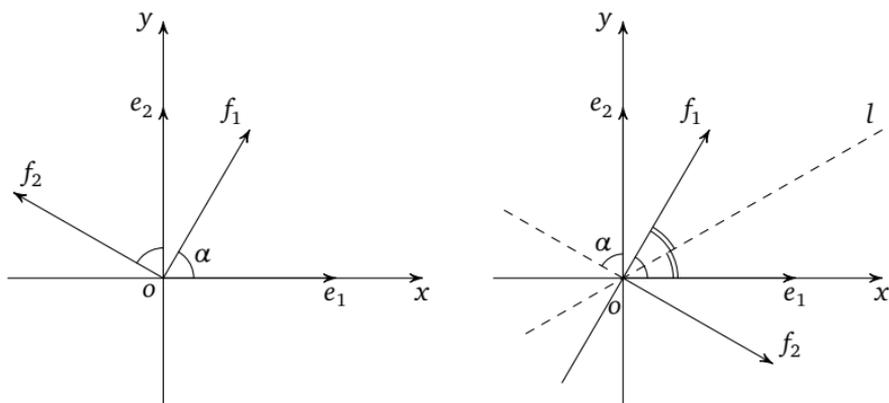


Рис. 1

Пример 11. Пусть F — какая-либо фигура на евклидовой плоскости. Тогда

$$\text{Sym } F = \{\varphi \in \text{Isom } E^2 : \varphi(F) = F\}$$

есть подгруппа группы $\text{Isom } E^2$; она называется *группой симметрии* фигуры F . Так, группа симметрии окружности с центром в начале отсчета o есть группа O_2 . Группа симметрии правильного n -угольника с центром в точке o есть подгруппа группы O_2 , состоящая из поворотов вокруг точки o на углы, кратные $2\pi/n$, и отражений относительно прямых, проходящих через o и одну из вершин или середину одной из сторон. Таким образом, эта группа содержит $2n$ элементов (n поворотов и n отражений); она называется *группой диэдра* и обозначается через D_n .

Пример 12. В силу формулы умножения определителей матрицы с определителем 1 образуют подгруппу в группе $\text{GL}_n(K)$. Эта подгруппа называется *унимодулярной группой* и обозначается через $\text{SL}_n(K)$.

Пример 13. Целочисленные матрицы с определителем 1 образуют подгруппу в группе $SL_n(\mathbb{R})$, обозначаемую через $SL_n(\mathbb{Z})$ (см. задачу 2.5.3).

Пример 14. Множество невырожденных диагональных матриц порядка n является абелевой подгруппой группы $GL_n(K)$.

Задача 4. Доказать, что множество строго треугольных квадратных матриц порядка n является подгруппой группы $GL_n(K)$.

§ 2. Группы в геометрии и физике

Цель этого параграфа — дать общее представление о роли групп в геометрии и физике.

В XIX в. математики осознали, что евклидова геометрия не является единственной мыслимой геометрией. Даже если принять, что «пространство, в котором мы живем», подчиняется законам евклидовой геометрии (что на самом деле верно лишь в первом приближении), имеет смысл изучать геометрию и других пространств, которые возникают в результате математических построений. В связи с этим возникает вопрос, что же в таком случае следует понимать под геометрией. Обобщая различные понятия, рассматриваемые в евклидовой геометрии, можно сформулировать различные ответы на этот вопрос.

В частности, обобщая понятие группы движений евклидовой геометрии, немецкий математик Клейн в своей лекции 1872 г., получившей известность под названием «Эрлангенская программа», дал определение геометрии как науки, изучающей свойства фигур, инвариантные относительно заданной группы преобразований.

Более подробно, пусть задано некоторое множество X и некоторая группа G его преобразований. Фигуру $F_1 \subset X$ будем считать эквивалентной (или равной, как говорят в элементарной геометрии) фигуре $F_2 \subset X$ относительно группы G и писать $F_1 \sim_G F_2$, если существует такое преобразование $\varphi \in G$, что $F_2 = \varphi(F_1)$. Проверим, что это действительно отношение эквивалентности:

- 1) $F \sim_G F$, так как $F = \text{id}(F)$ и $\text{id} \in G$;
- 2) если $F_1 \sim_G F_2$, т. е. $F_2 = \varphi(F_1)$, где $\varphi \in G$, то $F_2 \sim_G F_1$, так как $F_1 = \varphi^{-1}(F_2)$ и $\varphi^{-1} \in G$;
- 3) если $F_1 \sim_G F_2$ и $F_2 \sim_G F_3$, т. е. $F_2 = \varphi(F_1)$ и $F_3 = \psi(F_2)$, где $\varphi, \psi \in G$, то $F_1 \sim_G F_3$, так как $F_3 = \psi\varphi(F_1)$ и $\psi\varphi \in G$.

Мы видим, таким образом, что три аксиомы отношения эквивалентности в точности соответствуют трем аксиомам группы преобразований.

Одной из задач геометрии является нахождение необходимых и достаточных условий эквивалентности фигур (вспомните признаки равенства треугольников в евклидовой геометрии). Этой цели служат величины, инвариантные относительно преобразований из группы G (такие, как расстояние между точками или мера угла в евклидовой геометрии). Соотношения между этими инвариантами суть геометрические теоремы (например, теорема Пифагора или теорема о том, что медианы треугольника пересекаются в одной точке).

Конечно, далеко не любая группа преобразований приводит к интересной и важной для приложений геометрии. Все такие геометрии связаны с достаточно богатыми группами преобразований, которых не так много. Минимальным требованием здесь является транзитивность.

Определение 1. Группа G преобразований множества X называется *транзитивной*, если для любых $x, y \in X$ существует такое преобразование $\varphi \in G$, что $y = \varphi(x)$.

(Это означает, что в соответствующей геометрии все точки эквивалентны в смысле данного выше определения эквивалентности фигур.)

Пример 1. Группа $\text{Trans}(K^n)$ параллельных переносов пространства K^n (см. пример 1.4) транзитивна. В самом деле, для любых $x, y \in K^n$ имеем

$$y = t_{y-x}x.$$

Однако группа $\text{Trans}(K^n)$ все еще слишком мала, чтобы определять интересную геометрию. В качестве примера интересной геометрии, отличной от евклидовой, приведем аффинную геометрию.

Пусть $\varphi \in \text{GL}_n(K)$ и $a \in K^n$. Тогда

$$\varphi t_a \varphi^{-1} = t_{\varphi(a)}. \quad (5)$$

В самом деле, для любого $x \in K^n$ имеем:

$$(\varphi t_a \varphi^{-1})(x) = \varphi(\varphi^{-1}(x) + a) = x + \varphi(a) = t_{\varphi(a)}x.$$

Предложение 1. Для любой подгруппы $G \subset \text{GL}_n(K)$ множество

$$\text{Trans}(K^n) \cdot G = \{t_a \varphi : a \in K^n, \varphi \in G\}$$

является транзитивной группой преобразований пространства K^n .

Доказательство. При $a, b \in K^n$, $\varphi, \psi \in \text{GL}_n(K)$ имеем, согласно формулам (1) и (5),

$$(t_a \varphi)(t_b \psi) = t_a(\varphi t_b \varphi^{-1})\varphi\psi = t_{a+\varphi(b)}\varphi\psi \in \text{Trans}(K^n) \cdot G.$$

Отсюда следует, что

$$(t_a \varphi)^{-1} = t_{-\varphi^{-1}(a)}\varphi^{-1} \in \text{Trans}(K^n) \cdot G.$$

Таким образом, $\text{Trans}(K^n) \cdot G$ — группа преобразований. Она транзитивна, поскольку уже ее подгруппа $\text{Trans}(K^n)$ транзитивна. \square

В частности, мы можем взять $G = \text{GL}_n(K)$. Полученная группа

$$\text{GA}_n(K) = \text{Trans}(K^n) \cdot \text{GL}_n(K) \quad (6)$$

называется *полной аффинной группой* пространства K^n , а ее элементы — (биективными) *аффинными преобразованиями*. Связанная с ней геометрия называется *аффинной геометрией*.

В случае $K = \mathbb{R}$, $n = 2$ мы получаем аффинную геометрию евклидовой плоскости.

Предложение 2. *Группа движений евклидовой плоскости есть подгруппа группы $\text{GA}_2(\mathbb{R})$, равная $\text{Trans}(\mathbb{R}^2) \cdot \text{O}_2$.*

Доказательство. Прежде всего, заметим, что все параллельные переносы и все ортогональные преобразования являются движениями (см. пример 1.10). Пусть теперь f — какое-либо движение. Положим $a = f(o)$. Тогда движение $\varphi = t_a^{-1}f$ оставляет на месте точку o и, значит, принадлежит группе O_2 . Таким образом,

$$f = t_a \varphi \in \text{Trans}(\mathbb{R}^2) \cdot \text{O}_2. \quad \square$$

Аналогичным образом описывается группа движений евклидова пространства.

Следствие. *Если фигуры $F_1, F_2 \subset E^2$ равны в евклидовой геометрии, то они равны и в аффинной геометрии.*

Группа $\text{GA}_2(\mathbb{R})$ не совпадает с группой движений. Примером аффинного преобразования, не являющегося движением, может служить гомотетия (с коэффициентом $\neq \pm 1$) или растяжение вдоль какой-либо оси. Таким образом, группа $\text{GA}_2(\mathbb{R})$ богаче группы движений, и фигуры, не равные в евклидовой геометрии, могут оказаться равными в аффинной геометрии. Так, в аффинной геометрии все окружности равны.

Задача 1. Доказать, что в аффинной геометрии все треугольники равны.

В аффинной геометрии отсутствует понятие расстояния между точками. Однако, как показывает следующая задача, имеется инвариант трех точек, лежащих на одной прямой.

Задача 2. Доказать, что при аффинных преобразованиях сохраняется отношение, в котором точка делит отрезок.

В рамках группового подхода могут быть построены также проективная и конформная геометрии, геометрия Лобачевского и другие геометрии, используемые в математике и ее приложениях.

Группы преобразований в физике описывают симметрию физических законов, в частности симметрию пространства-времени.

Точка пространства-времени задается тремя пространственными координатами x, y, z и временной координатой t , так что пространство-время с фиксированной системой отсчета может быть отождествлено с \mathbb{R}^4 . Переход к другой системе отсчета означает некоторое преобразование пространства \mathbb{R}^4 . Как в классической, так и в релятивистской механике (точнее, в специальной теории относительности) существует понятие инерциальных систем отсчета, в которых все законы механики имеют одинаковый вид. Переходы от одной инерциальной системы отсчета к другим составляют некоторую группу преобразований пространства \mathbb{R}^4 . Эта группа однозначно определяет законы механики. Отличие релятивистской механики от классической обусловлено тем, что она берет за основу другую группу преобразований.

Группа симметрии пространства-времени в классической механике есть группа Галилея, описываемая следующим образом:

$$G = \text{Trans}(\mathbb{R}^4) \cdot H \cdot O_3,$$

где O_3 — группа ортогональных преобразований пространственных координат, а H — группа преобразований вида

$$(x, y, z, t) \mapsto (x + at, y + bt, z + ct, t),$$

соответствующих переходу к новой системе отсчета, равномерно и прямолинейно движущейся относительно старой. Из этого описания группы Галилея видно, что в классической механике время абсолютно в том смысле, что разность временных координат двух событий одинакова во всех инерциальных системах отсчета.

Согласно представлению релятивистской механики, группа симметрии пространства-времени есть группа Пуанкаре

$$G = \text{Trans}(\mathbb{R}^4) \cdot O_{3,1},$$

где $O_{3,1}$ — группа линейных преобразований, сохраняющих многочлен

$$x^2 + y^2 + z^2 - t^2$$

(в системе единиц, в которой скорость света равна 1). Группа $O_{3,1}$ содержит группу O_3 , не затрагивающую временной координаты. Нетривиальным примером преобразований из $O_{3,1}$ могут служить преобразования Лоренца

$$(x, y, z, t) \mapsto (x, y, z \operatorname{ch} a + t \operatorname{sh} a, z \operatorname{sh} a + t \operatorname{ch} a),$$

перемешивающие пространственные и временные координаты. Вид этих преобразований показывает, что в релятивистской механике время не абсолютно.

Группа Пуанкаре была описана в работах Лоренца и Пуанкаре как группа симметрии законов электродинамики (уравнений Максвелла). Заслуга Эйнштейна состояла в том, что он имел смелость сделать вывод, что и законы механики должны иметь ту же группу симметрии.

Группы преобразований лежат также в основе кристаллографии и теории элементарных частиц. Так, в кристаллографии они описывают симметрию кристаллических структур и, тем самым, физических свойств кристаллов. (См. рис. 2, где изображены кристаллические структуры поваренной соли, алмаза и графита.)

§ 3. Циклические группы

Так же как в группе \mathbb{R}^* , в любой группе G могут быть определены степени элемента $g \in G$ с целыми показателями:

$$g^k = \begin{cases} \underbrace{gg \dots g}_k, & \text{если } k > 0, \\ e, & \text{если } k = 0, \\ \underbrace{g^{-1}g^{-1} \dots g^{-1}}_{-k}, & \text{если } k < 0. \end{cases}$$

Имеет место свойство

$$g^k g^l = g^{k+l}. \quad (7)$$

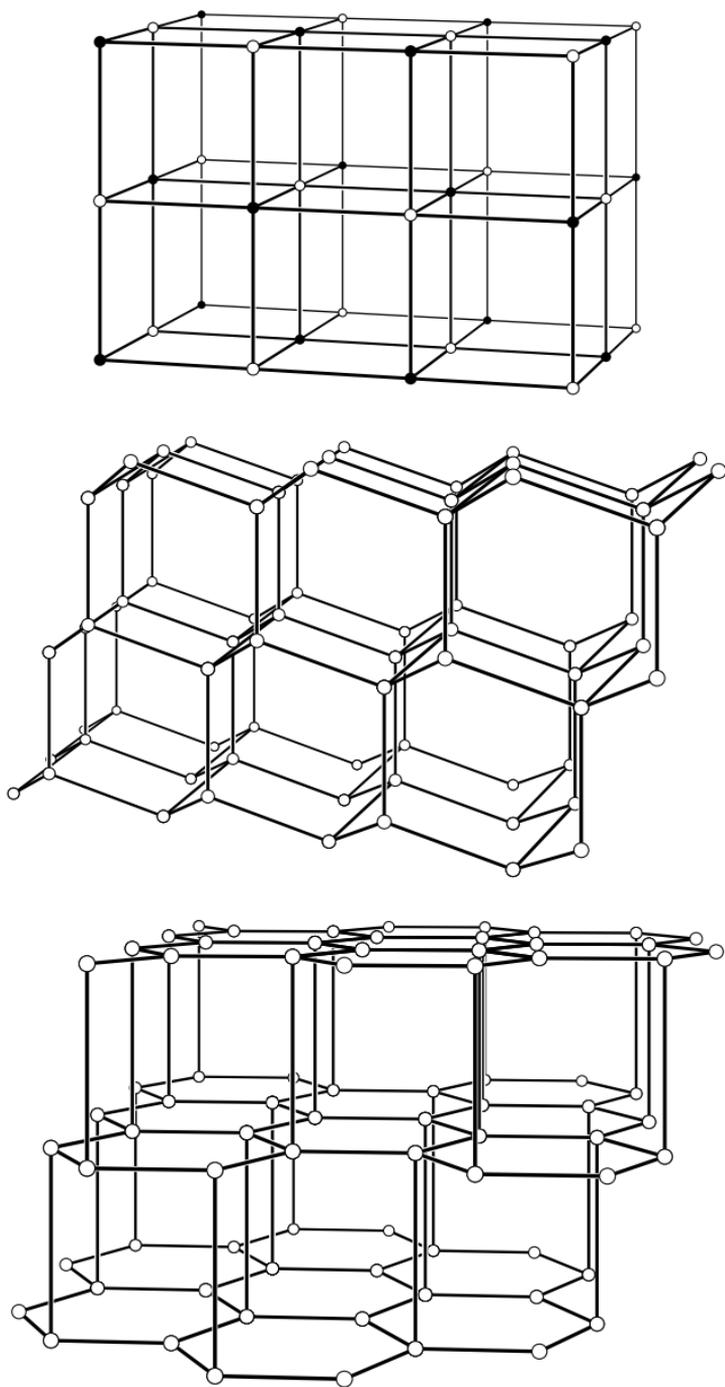


Рис. 2

Это очевидно, если $k, l > 0$. Рассмотрим случай, когда $k > 0, l < 0, k + l > 0$. Тогда

$$g^k g^l = \underbrace{gg \dots g}_k \underbrace{g^{-1} g^{-1} \dots g^{-1}}_{-l} = \underbrace{gg \dots g}_{k+l} = g^{k+l}.$$

Аналогично рассматриваются остальные случаи.

Из (7) следует, что

$$(g^k)^{-1} = g^{-k}.$$

Кроме того, $e = g^0$ по определению. Таким образом, степени элемента g образуют подгруппу в группе G . Она называется *циклической подгруппой, порожденной элементом g* , и обозначается через $\langle g \rangle$.

Возможны два принципиально разных случая: либо все степени элемента g различны, либо нет. В первом случае подгруппа $\langle g \rangle$ бесконечна. Рассмотрим более подробно второй случай.

Пусть $g^k = g^l, k > l$; тогда $g^{k-l} = e$. Наименьшее из натуральных чисел m , для которых $g^m = e$, называется в этом случае *порядком* элемента g и обозначается через $\text{ord } g$.

Предложение 1. Если $\text{ord } g = n$, то

- 1) $g^m = e \Leftrightarrow n \mid m$;
- 2) $g^k = g^l \Leftrightarrow k \equiv l \pmod{n}$.

Доказательство. 1) Разделим m на n с остатком:

$$m = qn + r, \quad 0 \leq r < n.$$

Тогда в силу определения порядка

$$g^m = (g^n)^q \cdot g^r = g^r = e \Leftrightarrow r = 0.$$

2) В силу предыдущего

$$g^k = g^l \Leftrightarrow g^{k-l} = e \Leftrightarrow n \mid (k-l) \Leftrightarrow k \equiv l \pmod{n}. \quad \square$$

Следствие. Если $\text{ord } g = n$, то подгруппа $\langle g \rangle$ содержит n элементов.

Доказательство. Действительно,

$$\langle g \rangle = \{e, g, g^2, \dots, g^{n-1}\}, \quad (8)$$

причем все перечисленные элементы различны. □

В том случае, когда не существует такого натурального m , что $g^m = e$ (т. е. имеет место первый из описанных выше случаев), полагают $\text{ord } g = \infty$. Отметим, что $\text{ord } e = 1$; порядки же всех остальных элементов группы больше 1.

В аддитивной группе говорят не о степенях элемента g , а о его *кратных*, которые обозначают через kg . В соответствии с этим порядок элемента g аддитивной группы G — это наименьшее из натуральных чисел m (если такие существуют), для которых

$$mg \doteq \underbrace{g + g + \dots + g}_m = 0.$$

Пример 1. Характеристика поля (см. § 1.5) есть порядок любого ненулевого элемента в его аддитивной группе.

Пример 2. Очевидно, что в конечной группе порядок любого элемента конечен. Покажем, как вычисляются порядки элементов группы S_n . Подстановка $\tau \in S_n$ называется *циклом* длины p и обозначается через $(i_1 i_2 \dots i_p)$, если она циклически переставляет i_1, i_2, \dots, i_p , т. е. $\tau(i_1) = i_2, \tau(i_2) = i_3, \dots, \tau(i_p) = i_1$, а все остальные числа оставляет на месте. Очевидно, что порядок цикла длины p равен p . Циклы τ_1 и τ_2 называются *независимыми*, если среди фактически переставляемых ими чисел нет общих; в этом случае $\tau_1 \tau_2 = \tau_2 \tau_1$. Всякая подстановка однозначно разлагается в произведение независимых циклов. Например,

$$\sigma = \left(\begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 5 & 6 & 7 & 4 & 8 & 3 & 2 & 1 \end{array} \right) = (2637)(158),$$

что наглядно показано на рис. 3, где действие подстановки σ изображено стрелками. Если подстановка σ разлагается в произведение независимых циклов длин p_1, p_2, \dots, p_s , то

$$\text{ord } \sigma = \text{НОК}\{p_1, p_2, \dots, p_s\}.$$

Например, для подстановки σ , изображенной на рис. 3, $\text{ord } \sigma = 12$.

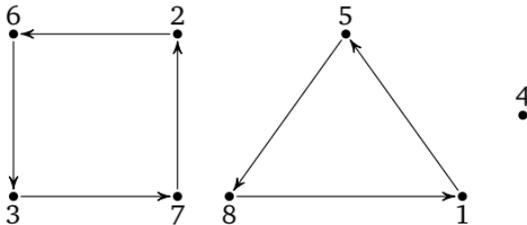


Рис. 3

Задача 1. Доказать, что порядок любого элемента группы S_n не превосходит числа

$$e^{n/e} \approx 1,44^n.$$

Пример 3. Порядок комплексного числа c в группе \mathbb{C}^* конечен тогда и только тогда, когда это число есть корень некоторой степени из единицы, что, в свою очередь, имеет место тогда и только тогда, когда $|c| = 1$, а $\arg c$ соизмерим с π , т. е. $\frac{\arg c}{\pi} \in \mathbb{Q}$.

Задача 2. Доказать, что $\operatorname{arctg} \frac{3}{4}$ несоизмерим с π .

Пример 4. Найдем элементы конечного порядка в группе $\operatorname{Isom} E^2$ движений плоскости. Пусть $\varphi \in \operatorname{Isom} E^2$, $\varphi^n = \operatorname{id}$. Для любой точки $p \in E^2$ точки

$$p, \varphi p, \varphi^2 p, \dots, \varphi^{n-1} p$$

циклически переставляются движением φ , так что их центр тяжести o неподвижен относительно φ . Следовательно, φ — либо поворот на угол вида $\frac{2\pi k}{n}$ вокруг точки o , либо отражение относительно некоторой прямой, проходящей через o .

Пример 5. Найдем порядок матрицы

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 1 \end{pmatrix}$$

как элемента группы $\operatorname{GL}_2(\mathbb{R})$. Имеем

$$A^2 = \begin{pmatrix} -1 & -1 \\ 1 & 0 \end{pmatrix}, \quad A^3 = -E,$$

откуда

$$A^4 = -A, \quad A^5 = -A^2, \quad A^6 = -A^3 = E,$$

так что $\operatorname{ord} A = 6$. Конечно, этот пример специально подобран: вероятность того, что порядок наудачу выбранной матрицы $A \in \operatorname{GL}_2(\mathbb{R})$ будет конечен, равна нулю.

Предложение 2. Если $\operatorname{ord} g = n$, то

$$\operatorname{ord} g^k = \frac{n}{(n, k)}. \quad (9)$$

Доказательство. Пусть

$$(n, k) = d, \quad n = n_1 d, \quad k = k_1 d,$$

так что $(n_1, k_1) = 1$. Имеем

$$(g^k)^m = e \Leftrightarrow n | km \Leftrightarrow n_1 | k_1 m \Leftrightarrow n_1 | m.$$

Следовательно, $\operatorname{ord} g^k = n_1$. □

Определение 1. Группа G называется *циклической*, если существует такой элемент $g \in G$, что $G = \langle g \rangle$. Всякий такой элемент называется *порождающим элементом* группы G .

Пример 6. Аддитивная группа \mathbb{Z} целых чисел является циклической, так как порождается элементом 1.

Пример 7. Аддитивная группа \mathbb{Z}_n вычетов по модулю n является циклической, так как порождается элементом [1].

Пример 8. Мультипликативная группа C_n комплексных корней n -й степени из 1 является циклической. В самом деле, эти корни суть числа

$$\varepsilon_k = \cos \frac{2\pi k}{n} + i \sin \frac{2\pi k}{n} \quad (k=0, 1, \dots, n-1).$$

Ясно, что $\varepsilon_k = \varepsilon_1^k$. Следовательно, группа C_n порождается элементом ε_1 .

Задача 3. Доказать, что группа \mathbb{Z}_n^* обратимых элементов кольца \mathbb{Z}_n (см. задачу 1.5.1) является циклической при $n \leq 7$ и $n=9$ и не является циклической при $n=8$.

Легко видеть, что в бесконечной циклической группе $G = \langle g \rangle$ порождающими элементами являются только g и g^{-1} . Так, в группе \mathbb{Z} порождающими элементами являются только 1 и -1 .

Число элементов конечной группы G называется ее *порядком* и обозначается через $|G|$. Порядок конечной циклической группы равен порядку ее порождающего элемента. Поэтому из предложения 2 следует

Предложение 3. Элемент g^k циклической группы $G = \langle g \rangle$ порядка n является порождающим тогда и только тогда, когда $(n, k) = 1$.

Пример 9. Порождающие элементы группы C_n (см. пример 8) называются *первообразными корнями* n -й степени из 1. Это корни вида ε_k , где $(n, k) = 1$. Например, первообразные корни 12-й степени из 1 — это $\varepsilon_1, \varepsilon_5, \varepsilon_7, \varepsilon_{11}$.

Циклические группы — это наиболее простые группы, которые можно себе представить. (В частности, они абелевы.) Следующая теорема дает их полное описание.

Теорема 1. Всякая бесконечная циклическая группа изоморфна группе \mathbb{Z} . Всякая конечная циклическая группа порядка n изоморфна группе \mathbb{Z}_n .

Доказательство. Если $G = \langle g \rangle$ — бесконечная циклическая группа, то в силу формулы (7) отображение $f: \mathbb{Z} \rightarrow G, k \mapsto g^k$, есть изоморфизм.

Пусть $G = \langle g \rangle$ — конечная циклическая группа порядка n . Рассмотрим отображение

$$f: \mathbb{Z}_n \rightarrow G, \quad [k] \mapsto g^k \quad (k \in \mathbb{Z}).$$

Так как

$$[k] = [l] \Leftrightarrow k \equiv l \pmod{n} \Leftrightarrow g^k = g^l,$$

отображение f корректно определено и биективно. Свойство

$$f(k+l) = f(k)f(l)$$

вытекает из той же формулы (7). Таким образом, f — изоморфизм. \square

Для понимания строения какой-либо группы важную роль играет знание ее подгрупп. Все подгруппы циклической группы могут быть легко описаны.

Теорема 2. 1) *Всякая подгруппа циклической группы является циклической.*

2) *В циклической группе порядка n порядок любой подгруппы делит n и для любого делителя q числа n существует ровно одна подгруппа порядка q .*

Доказательство. 1) Пусть $G = \langle g \rangle$ — циклическая группа и H — ее подгруппа, отличная от $\{e\}$. (Единичная подгруппа, очевидно, является циклической.) Заметим, что если $g^{-m} \in H$ для какого-либо $m \in \mathbb{N}$, то и $g^m \in H$. Пусть m — наименьшее из натуральных чисел, для которых $g^m \in H$. Докажем, что $H = \langle g^m \rangle$. Пусть $g^k \in H$. Разделим k на m с остатком:

$$k = qt + r, \quad 0 \leq r < m.$$

Имеем

$$g^r = g^k (g^m)^{-q} \in H,$$

откуда в силу определения числа m следует, что $r = 0$ и, значит, $g^k = (g^m)^q$.

2) Если $|G| = n$, то предыдущее рассуждение, примененное к $k = n$ (в этом случае $g^k = e \in H$), показывает, что $n = qt$. При этом

$$H = \{e, g^m, g^{2m}, \dots, g^{(q-1)m}\}, \quad (10)$$

и H является единственной подгруппой порядка q в группе G . Обратное, если q — любой делитель числа n и $n = qt$, то подмножество H , определяемое равенством (10), является подгруппой порядка q . \square

Следствие. В циклической группе простого порядка любая неединичная подгруппа совпадает со всей группой.

Пример 10. В группе \mathbb{Z} всякая подгруппа имеет вид $m\mathbb{Z}$, где $m \geq 0$.

Пример 11. В группе C_n корней n -й степени из 1 любая подгруппа есть группа C_q корней q -й степени из 1, где $q \mid n$.

§ 4. Системы порождающих

Пусть S — какое-либо подмножество группы G . Обозначим через $\langle S \rangle$ совокупность всевозможных произведений вида

$$g_1^{\varepsilon_1} g_2^{\varepsilon_2} \cdots g_k^{\varepsilon_k} \quad (g_1, g_2, \dots, g_k \in S; \varepsilon_1, \varepsilon_2, \dots, \varepsilon_k = \pm 1). \quad (11)$$

Это наименьшая подгруппа группы G , содержащая S . В самом деле, если какая-либо подгруппа содержит S , то она содержит и все указанные произведения. С другой стороны, само множество $\langle S \rangle$ является подгруппой, как показывают следующие равенства:

$$\begin{aligned} (g_1^{\varepsilon_1} g_2^{\varepsilon_2} \cdots g_k^{\varepsilon_k})(g_{k+1}^{\varepsilon_{k+1}} g_{k+2}^{\varepsilon_{k+2}} \cdots g_{k+l}^{\varepsilon_{k+l}}) &= g_1^{\varepsilon_1} g_2^{\varepsilon_2} \cdots g_{k+l}^{\varepsilon_{k+l}}, \\ (g_1^{\varepsilon_1} g_2^{\varepsilon_2} \cdots g_k^{\varepsilon_k})^{-1} &= g_k^{-\varepsilon_k} \cdots g_2^{-\varepsilon_2} g_1^{-\varepsilon_1}. \end{aligned}$$

Говорят, что $\langle S \rangle$ — подгруппа, порожденная подмножеством S . В частности, если S состоит из одного элемента g , то $\langle S \rangle = \langle g \rangle$ есть циклическая подгруппа, порожденная элементом g в том смысле, как это было определено в предыдущем параграфе.

Замечание 1. Удобно считать, что в число произведений (11) входит пустое произведение ($k=0$), которое по определению равно e .

Определение 1. Говорят, что группа G порождается своим подмножеством S или что S — система порождающих (элементов) группы G , если $G = \langle S \rangle$.

Конечно, любая группа G порождается подмножеством $S = G$, однако представляет интерес найти возможно меньшую систему порождающих.

Пример 1. Группа диэдра D_n (см. пример 1.11) порождается поворотом φ на угол $\frac{2\pi}{n}$ и (любым) отражением $\psi \in D_n$. В самом деле, φ порождает циклическую подгруппу C_n всех поворотов, содержащихся в группе D_n ; умножая элементы этой подгруппы на ψ , мы получим все отражения, входящие в группу D_n .

Два важных примера систем порождающих содержатся в приводимых ниже теоремах.

Подстановка, являющаяся циклом длины 2 (см. пример 3.2), называется *транспозицией*.

Теорема 1. *Группа S_n порождается транспозициями.*

Доказательство. Отметим, что каждая транспозиция обратна сама себе. Поэтому утверждение теоремы означает, что любая подстановка разлагается в произведение транспозиций.

Умножение подстановки

$$\sigma = \begin{pmatrix} 1 & 2 & \dots & n \\ k_1 & k_2 & \dots & k_n \end{pmatrix} \quad (12)$$

слева на транспозицию (ij) вызывает перестановку i и j в нижней строке. Такая операция также называется транспозицией. Очевидно, что путем последовательных транспозиций любую перестановку (k_1, k_2, \dots, k_n) можно привести к тривиальной: сначала, если $k_1 \neq 1$, меняем местами k_1 и 1, ставя тем самым 1 на первое место, затем ставим 2 на второе место и т. д. Таким образом, существуют такие транспозиции $\tau_1, \tau_2, \dots, \tau_s$, что

$$\tau_s \dots \tau_2 \tau_1 \sigma = \text{id}$$

и, значит,

$$\sigma = \tau_1 \tau_2 \dots \tau_s. \quad \square$$

Задача 1. Доказать, что группа S_n порождается *смежными транспозициями* (12), (23), ..., $(n-1, n)$, причем минимальное число смежных транспозиций, в произведение которых может быть разложена подстановка $\sigma \in S_n$, равно числу инверсий в нижней строке ее стандартной записи (12).

Теорема 2. *Группа $GL_n(K)$ порождается элементарными матрицами.*

(Определение элементарных матриц см. в § 2.1.)

Доказательство. Отметим, что матрица, обратная к элементарной, также элементарна (см. § 2.1). Поэтому утверждение теоремы означает, что любая невырожденная матрица разлагается в произведение элементарных матриц.

Умножение матрицы $A \in GL_n(K)$ слева на элементарную матрицу вызывает соответствующее элементарное преобразование ее строк. Мы знаем, что с помощью элементарных преобразований строк любую невырожденную матрицу можно привести к единичной

матрице. Таким образом, существуют такие элементарные матрицы U_1, U_2, \dots, U_s , что

$$U_s \dots U_2 U_1 A = E$$

и, значит,

$$A = U_1^{-1} U_2^{-1} \dots U_s^{-1}. \quad \square$$

Задача 2. Доказать, что группа $SL_2(\mathbb{Z})$ (см. пример 1.13) порождается матрицами

$$R = \begin{pmatrix} 0 & -1 \\ 1 & 1 \end{pmatrix}, \quad S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Задача 3. Доказать, что группа движений плоскости порождается отражениями относительно прямых. (Указание: доказать вначале, что каждый поворот и каждый параллельный перенос являются произведениями двух отражений.)

§ 5. Разбиение на смежные классы

Пусть G — группа и H — ее подгруппа. Будем говорить, что элементы $g_1, g_2 \in G$ *сравнимы по модулю H* , и писать

$$g_1 \equiv g_2 \pmod{H},$$

если

$$g_1^{-1} g_2 \in H, \quad (13)$$

т. е. $g_2 = g_1 h$, где $h \in H$. Это определение обобщает определение сравнимости целых чисел по модулю n , которое получается в случае $G = \mathbb{Z}$, $H = n\mathbb{Z}$.

Докажем, что определенное таким образом отношение сравнимости по модулю H является отношением эквивалентности:

- 1) $g \equiv g \pmod{H}$, так как $g^{-1} g = e \in H$;
- 2) если $g_1 \equiv g_2 \pmod{H}$, т. е. $g_1^{-1} g_2 \in H$, то $g_2 \equiv g_1 \pmod{H}$, так как

$$g_2^{-1} g_1 = (g_1^{-1} g_2)^{-1} \in H;$$

- 3) если $g_1 \equiv g_2 \pmod{H}$ и $g_2 \equiv g_3 \pmod{H}$, т. е. $g_1^{-1} g_2, g_2^{-1} g_3 \in H$, то $g_1 \equiv g_3 \pmod{H}$, так как

$$g_1^{-1} g_3 = (g_1^{-1} g_2)(g_2^{-1} g_3) \in H.$$

Классы этой эквивалентности называются (*левыми*) *смежными классами* группы G по подгруппе H . Ясно, что смежный класс, содержащий элемент g , имеет вид

$$gH = \{gh : h \in H\}.$$

Одним из смежных классов является сама подгруппа H .

Поскольку умножение в группе не обязано быть коммутативным, мы получим, вообще говоря, другое отношение эквивалентности, взяв вместо условия (13) аналогичное ему условие

$$g_2 g_1^{-1} \in H. \quad (14)$$

Классы этой эквивалентности называются (*правыми*) *смежными классами* группы G по подгруппе H . Они имеют вид

$$Hg = \{hg : h \in H\}.$$

Заметим, что инверсия $g \mapsto g^{-1}$ устанавливает взаимно однозначное соответствие между множествами левых и правых смежных классов. А именно,

$$(gH)^{-1} = Hg^{-1}.$$

Пример 1. Смежные классы аддитивной группы \mathbb{C} по подгруппе \mathbb{R} изображаются на комплексной плоскости прямыми, параллельными вещественной оси (рис. 4, а).

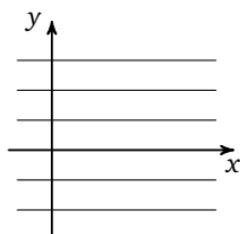


Рис. 4, а)

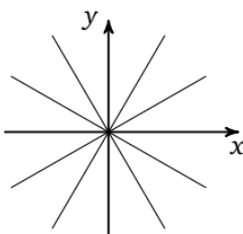


Рис. 4, б)

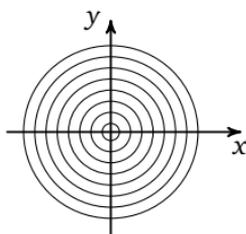


Рис. 4, в)

Пример 2. Смежные классы мультипликативной группы \mathbb{C}^* по подгруппе \mathbb{R}_+^* положительных чисел — это лучи, исходящие из начала координат (рис. 4, б).

Пример 3. Смежные классы группы \mathbb{C}^* по подгруппе

$$\mathbb{T} = \{z \in \mathbb{C}^* : |z| = 1\}$$

— это окружности с центром в начале координат (рис. 4, в).

Пример 4. В случае $G = GL_n(K)$, $H = SL_n(K)$ (см. пример 1.12) условие (13), равно как и (14), означает, что $\det g_1 = \det g_2$. Поэтому левые смежные классы в данном случае совпадают с правыми (хотя группа $GL_n(K)$ не абелева); каждый из них представляет собой совокупность всех матриц с определителем, равным какому-либо фиксированному числу.

Пример 5. В группе $G = S_n$ рассмотрим подгруппу H , состоящую из подстановок, оставляющих на месте число n . Подстановки $\sigma_1, \sigma_2 \in S_n$ принадлежат одному левому смежному классу по H , если $\sigma_1^{-1}\sigma_2(n) = n$, т. е. если

$$\sigma_1(n) = \sigma_2(n).$$

Следовательно, имеется n левых смежных классов P_1, P_2, \dots, P_n , где

$$P_k = \{\sigma \in S_n : \sigma(n) = k\}.$$

В то же время подстановки $\sigma_1, \sigma_2 \in S_n$ принадлежат одному правому смежному классу, если $\sigma_2\sigma_1^{-1}(n) = n$, т. е. если

$$\sigma_1^{-1}(n) = \sigma_2^{-1}(n).$$

Следовательно, имеется n правых смежных классов Q_1, Q_2, \dots, Q_n , где

$$Q_k = \{\sigma \in S_n : \sigma(k) = n\}.$$

Мы видим, что при $n > 2$ правые смежные классы отличны от левых, за исключением класса $Q_n = P_n = H$.

Множество левых смежных классов группы G по подгруппе H обозначается через G/H . Число смежных классов группы G по H (левых или правых, безразлично), если оно конечно, называется *индексом подгруппы H* и обозначается через $|G:H|$.

Теорема 1 (теорема Лагранжа). *Если G — конечная группа и H — любая ее подгруппа, то*

$$|G| = |G:H||H|.$$

Доказательство. Все смежные классы gH содержат одно и то же число элементов, равное $|H|$. Поскольку они образуют разбиение группы G (как классы эквивалентности), порядок группы G равен произведению их числа на $|H|$. \square

Следствие 1. *Порядок любой подгруппы конечной группы делит порядок группы.*

Мы уже видели это в случае циклических групп (теорема 3.2).

Следствие 2. *Порядок любого элемента конечной группы делит порядок группы.*

Доказательство. Это вытекает из следствия 1 и того, что порядок элемента равен порядку порождаемой им циклической подгруппы. \square

Следствие 3. *Всякая конечная группа простого порядка является циклической.*

Доказательство. В силу следствия 1 такая группа должна совпадать с циклической подгруппой, порожденной любым элементом, отличным от единицы. \square

Следствие 4. *Если $|G| = n$, то $g^n = e$ для любого $g \in G$.*

Доказательство. Пусть $\text{ord } g = m$. В силу следствия 2 имеем $m \mid n$. Значит, $g^n = e$. \square

Пример 6. Если p — простое число, то мультипликативная группа \mathbb{Z}_p^* поля \mathbb{Z}_p есть (абелева) группа порядка $p - 1$. Следовательно, $g^{p-1} = 1$ для любого элемента $g \in \mathbb{Z}_p^*$. Это означает, что

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

для любого целого числа a , не делящегося на p . Последнее утверждение есть так называемая *малая теорема Ферма*. (Другой способ ее доказательства см. в задаче 1.5.2.)

Для любого n порядок группы \mathbb{Z}_n^* обратимых элементов кольца \mathbb{Z}_n , равный количеству чисел в ряде $1, 2, \dots, n$, взаимно простых с n (см. задачу 1.5.1), обозначается через $\varphi(n)$. Функция φ , определенная таким образом на множестве натуральных чисел, называется *функцией Эйлера*. Применение следствия 4 к группе \mathbb{Z}_n^* дает

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

для любого целого числа a , взаимно простого с n . Это обобщение малой теоремы Ферма называется *теоремой Эйлера*.

Например, легко видеть, что $\varphi(125) = 125 - 25 = 100$. Отсюда следует, что $2^{100} \equiv 1 \pmod{125}$ — результат, полученный нами в примере 1.5.7 прямым вычислением.

Разбиение на смежные классы естественно возникает при изучении групп преобразований.

Пусть G — группа преобразований множества X . Будем говорить, что точки $x, y \in X$ эквивалентны относительно G , и писать $x \sim_G y$, если существует такой элемент $g \in G$, что $y = gx$. Это частный случай

эквивалентности фигур, определенной в § 2, и, следовательно, — отношение эквивалентности. Класс эквивалентности точки $x \in X$ называется ее *орбитой*. Иначе говоря, орбита точки x есть множество

$$Gx = \{gx : g \in G\}.$$

В частности, транзитивные группы преобразований (см. определение 2.1) — это группы преобразований, имеющие единственную орбиту.

Подгруппа

$$G_x = \{g \in G : gx = x\}$$

называется *стабилизатором* точки x .

Пример 7. Группа движений евклидовой плоскости транзитивна. Стабилизатором начала отсчета является ортогональная группа O_2 (см. пример 1.10).

Пример 8. Орбиты группы O_2 суть окружности с центром в начале отсчета o и сама точка o . Стабилизатор точки $p \neq o$ состоит из тождественного преобразования и отражения относительно прямой op , а стабилизатор точки o — это вся группа O_2 .

Пример 9. Группа S_n транзитивна на множестве $\{1, 2, \dots, n\}$. Стабилизатор числа n есть подгруппа $H \simeq S_{n-1}$, рассмотренная в примере 5.

Следующая теорема является обобщением (первой части) примера 5.

Теорема 2. *Имеется взаимно однозначное соответствие между орбитой Gx и множеством смежных классов G/G_x , при котором точке $y = gx \in Gx$ соответствует смежный класс gG_x .*

Доказательство. При $g_1, g_2 \in G$ имеем

$$g_1 \equiv g_2 \pmod{G_x} \Leftrightarrow g_1^{-1}g_2 \in G_x \Leftrightarrow g_1^{-1}g_2x = x \Leftrightarrow g_1x = g_2x.$$

Таким образом, элементы одного смежного класса группы G по G_x характеризуются тем, что они переводят точку x в одну и ту же точку. Более точно, все элементы смежного класса gG_x , и только они, переводят точку x в точку $y = gx$. Тем самым и установлено искомое соответствие. \square

Число элементов орбиты Gx , если оно конечно, называется ее *длиной* и обозначается через $|Gx|$.

Следствие. *Если G — конечная группа, то*

$$|G| = |Gx||G_x|. \quad (15)$$

Из этой формулы следует, что порядки стабилизаторов всех точек орбиты одинаковы. На самом деле имеется точная связь между стабилизаторами точек одной орбиты, не зависящая от конечности группы G . Мы сформулируем ее в виде задачи.

Задача 1. Доказать, что

$$G_{gx} = gG_xg^{-1}.$$

Пример 10. Пусть $K \subset E^3$ — куб. Рассмотрим группу его симметрии

$$G = \text{Sym } K = \{\varphi \in \text{Isom } E^3 : \varphi(K) = K\}.$$

Очевидно, что это конечная группа. Более того, симметрия куба полностью определяется тем, как она переставляет его вершины. Поэтому мы можем рассматривать группу G как группу преобразований множества V вершин куба K . Ввиду того что куб является правильным многогранником, любую вершину куба можно перевести в любую другую с помощью преобразования из группы G . Иначе говоря, группа G транзитивна на множестве V . Следовательно,

группа G транзитивна на множестве V . Следовательно,

$$|G| = 8|G_v|,$$

где v — какая-либо вершина. Аналогичным образом, рассматривая группу G_v как группу преобразований множества ребер, выходящих из v , можно показать, что

$$|G_v| = 3|G_{v,e}|,$$

где $G_{v,e}$ — стабилизатор в группе G_v какого-либо ребра e , выходящего из v . Группа $G_{v,e}$ состоит из тождественного преобразования и отражения относительно плоскости, проходящей через центр куба и ребро e (см. рис. 5). Таким образом,

$$|\text{Sym } K| = 8 \cdot 3 \cdot 2 = 48.$$

Задача 2. Получить тот же результат еще двумя способами, рассмотрев группу $\text{Sym } K$ как группу преобразований множества граней и множества ребер куба соответственно.

Аналогичным образом можно найти порядки групп симметрии других правильных многогранников (см. рис. 6). (По поводу определения правильных многогранников см. § 7.4.)

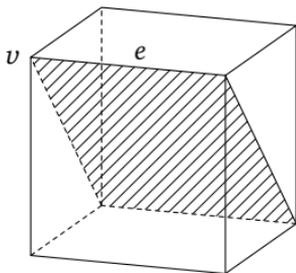


Рис. 5

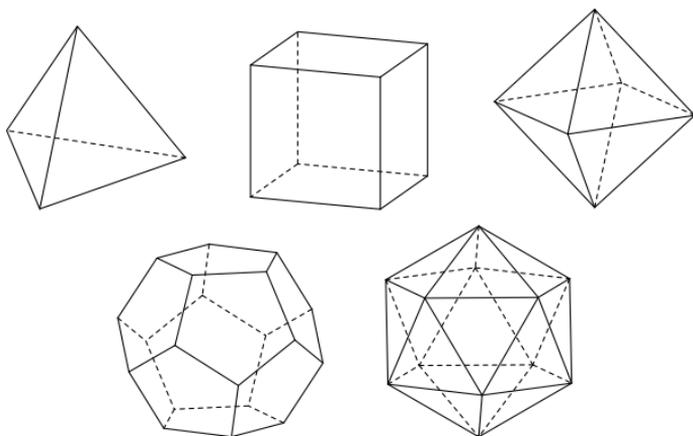


Рис. 6

Пример 11. Пусть G — группа преобразований алгебры многочленов $K[x_1, x_2, x_3, x_4]$, состоящая из всевозможных перестановок переменных x_1, x_2, x_3, x_4 . Группа G изоморфна S_4 и, следовательно, $|G| = 4! = 24$. Рассмотрим многочлен $f = x_1x_2 + x_3x_4$. Перестановка переменных из него можно получить 3 многочлена

$$x_1x_2 + x_3x_4, \quad x_1x_3 + x_2x_4, \quad x_1x_4 + x_2x_3.$$

Это означает, что $|Gf| = 3$. По формуле (15) находим

$$|G_f| = \frac{|G|}{|Gf|} = \frac{24}{3} = 8.$$

Заметим, что, если отождествить группу G с группой S_4 , то G_f будет не чем иным, как подгруппой, обозначенной в примере 1.8 через $\text{Sym } f$.

Отношение сравнимости по модулю n в аддитивной группе целых чисел согласовано с операцией сложения, что позволяет определить операцию сложения в фактормножестве. Аналогичным образом можно определить операцию в множестве смежных классов группы по подгруппе и в других случаях, но не всегда.

Определение 1. Подгруппа H группы G называется *нормальной*, если

$$gH = Hg \quad \forall g \in G \quad (16)$$

или, что эквивалентно,

$$gHg^{-1} = H \quad \forall g \in G. \quad (17)$$

В этом случае пишут $H \triangleleft G$ (или $G \triangleright H$).

Для того чтобы подгруппа H была нормальной, достаточно (но не необходимо), чтобы каждый элемент группы G был перестановочен с каждым элементом из H . В частности, в абелевой группе любая подгруппа нормальна.

Теорема 3. *Отношение сравнимости по модулю подгруппы H согласовано с операцией умножения в группе G тогда и только тогда, когда подгруппа H нормальна.*

Доказательство. Согласованность отношения сравнимости по модулю H с операцией умножения означает следующее:

$$g_1 \equiv g'_1 \pmod{H}, \quad g_2 \equiv g'_2 \pmod{H} \Rightarrow g_1 g_2 \equiv g'_1 g'_2 \pmod{H}$$

или, что эквивалентно, для любых $g_1, g_2 \in G$ и $h_1, h_2 \in H$

$$(g_1 h_1)(g_2 h_2) \equiv g_1 g_2 \pmod{H}.$$

Последнее условие, согласно определению, переписывается в виде

$$g_2^{-1} h_1 g_2 \in H.$$

Так как g_2 может быть любым элементом группы G , а h_1 — любым элементом подгруппы H , то это равносильно условию нормальности (17). \square

Задача 3. Доказать, что всякое отношение эквивалентности в группе, согласованное с операцией, есть отношение сравнимости по модулю некоторой (нормальной) подгруппы.

Таким образом, если $H \triangleleft G$, то операция умножения в группе G определяет операцию умножения в множестве G/H по правилу

$$(g_1 H)(g_2 H) = g_1 g_2 H.$$

Эта операция наследует ассоциативность операции в группе G . Для нее имеется единица — смежный класс eH . Каждый смежный класс gH имеет обратный, а именно $g^{-1}H$. Следовательно, G/H — группа. Эта группа называется *факторгруппой* группы G по H .

Очевидно, что если группа абелева, то любая ее факторгруппа также абелева.

Пример 12. Факторгруппа $\mathbb{Z}/n\mathbb{Z}$ есть группа вычетов \mathbb{Z}_n .

Пример 13. Смежные классы группы \mathbb{C} по \mathbb{R} (см. пример 1) суть прямые $L_a = \{z : \operatorname{Im} z = a\}$ ($a \in \mathbb{R}$). Операция сложения в \mathbb{C}/\mathbb{R} задается формулой $L_a + L_b = L_{a+b}$, так что факторгруппа \mathbb{C}/\mathbb{R} изоморфна группе \mathbb{R} .

Пример 14. Смежные классы группы \mathbb{C}^* по \mathbb{T} (см. пример 3) суть окружности $C_r = \{z \in \mathbb{C}^* : |z| = r\}$ ($r > 0$). Операция умножения в \mathbb{C}^*/\mathbb{T} задается формулой $C_r C_s = C_{rs}$, так что факторгруппа \mathbb{C}^*/\mathbb{T} изоморфна группе \mathbb{R}_+^* .

Пример 15. Как мы видели выше (см. пример 4), левые смежные классы группы $GL_n(K)$ по $SL_n(K)$ совпадают с правыми и имеют вид

$$M_a = \{A \in GL_n(K) : \det A = a\} \quad (a \in K^*).$$

Следовательно, $SL_n(K)$ — нормальная подгруппа. Операция умножения в факторгруппе задается формулой $M_a M_b = M_{ab}$, так что факторгруппа $GL_n(K)/SL_n(K)$ изоморфна K^* .

Пример 16. Подгруппа H (изоморфная S_{n-1}) группы S_n , рассмотренная в примере 5, не является нормальной при $n \geq 3$.

Задача 4. Доказать, что всякая факторгруппа циклической группы является циклической.

Задача 5. Доказать, что группа диагональных матриц не является нормальной подгруппой группы $GL_n(K)$ при $n \geq 2$ и $|K| \geq 3$.

§ 6. Гомоморфизмы

В любой алгебраической теории наряду с изоморфизмами рассматривают более общие отображения, называемые гомоморфизмами. Они отличаются от изоморфизмов тем, что не обязаны быть биективными. Тем не менее они позволяют установить полезные связи между алгебраическими структурами одного типа.

Дадим точное определение гомоморфизма групп.

Определение 1. Гомоморфизмом группы G в группу H называется отображение $f: G \rightarrow H$, удовлетворяющее условию

$$f(ab) = f(a)f(b) \quad \forall a, b \in G.$$

Установим некоторые общие свойства гомоморфизмов групп.

1) $f(e) = e$. В самом деле, пусть $f(e) = h \in H$; тогда

$$h^2 = f(e)^2 = f(e^2) = f(e) = h,$$

откуда $h = e$.

2) $f(a^{-1}) = f(a)^{-1}$, ибо

$$f(a)f(a^{-1}) = f(aa^{-1}) = f(e) = e.$$

3) $\text{Im } f = \{f(a) : a \in G\}$ есть подгруппа группы H (называемая образом гомоморфизма f). Это следует из определения гомоморфизма и предыдущих свойств.

4) $\text{Ker } f = \{a \in G : f(a) = e\}$ есть нормальная подгруппа группы G (называемая ядром гомоморфизма f). Действительно,

$$\begin{aligned} a, b \in \text{Ker } f &\Rightarrow f(ab) = f(a)f(b) = e^2 = e \Rightarrow ab \in \text{Ker } f, \\ a \in \text{Ker } f &\Rightarrow f(a^{-1}) = f(a)^{-1} = e^{-1} = e \Rightarrow a^{-1} \in \text{Ker } f, \\ &e \in \text{Ker } f, \end{aligned}$$

$$a \in \text{Ker } f, g \in G \Rightarrow$$

$$\begin{aligned} \Rightarrow f(gag^{-1}) &= f(g)f(a)f(g)^{-1} = f(g)ef(g)^{-1} = f(g)f(g)^{-1} = e \Rightarrow \\ &\Rightarrow gag^{-1} \in \text{Ker } f. \end{aligned}$$

5) $f(g_1) = f(g_2) \Leftrightarrow g_1 \equiv g_2 \pmod{\text{Ker } f}$; в частности, гомоморфизм f инъективен тогда и только тогда, когда $\text{Ker } f = \{e\}$. Действительно,

$$\begin{aligned} f(g_1) = f(g_2) &\Leftrightarrow f(g_1^{-1}g_2) = e \Leftrightarrow \\ &\Leftrightarrow g_1^{-1}g_2 \in \text{Ker } f \Leftrightarrow g_1 \equiv g_2 \pmod{\text{Ker } f}. \end{aligned}$$

Таким образом, гомоморфизм $f : G \rightarrow H$ является изоморфизмом (т. е. биективен) тогда и только тогда, когда $\text{Im } f = H$ и $\text{Ker } f = \{e\}$. В этом случае иногда пишут $f : G \xrightarrow{\sim} H$. Если группы G и H изоморфны (т. е. существует изоморфизм $f : G \xrightarrow{\sim} H$), то пишут $G \simeq H$.

Гомоморфизм группы в себя называется ее *эндоморфизмом*. Изоморфизм группы на себя называется ее *автоморфизмом*.

Пример 1. Пусть K — произвольное кольцо. Свойство дистрибутивности $a(b+c) = ab+ac$ означает, что отображение $x \mapsto ax$ (умножение слева на a) является эндоморфизмом аддитивной группы кольца K . (Аналогичное утверждение справедливо и для умножения справа.)

Пример 2. Пусть G — произвольная аддитивная (соответственно мультипликативная) абелева группа. Тогда для любого $n \in \mathbb{Z}$ отображение $x \mapsto nx$ (соответственно $x \mapsto x^n$) является эндоморфизмом группы G . (Для неабелевой группы это, вообще говоря, неверно.) В случае $G = C^*$ ядром этого гомоморфизма является группа C_n корней n -й степени из 1.

Пример 3. Согласно основному свойству экспоненты, отображение $x \mapsto e^x$ является гомоморфизмом аддитивной группы \mathbb{R} в мульти-

пликативную группу \mathbb{R}^* . Его образ — это подгруппа \mathbb{R}_+^* положительных чисел, а ядро тривиально.

Пример 4. Отображение $x \mapsto \cos x + i \sin x$ является гомоморфизмом группы \mathbb{R} в группу \mathbb{C}^* . Его образ есть \mathbb{T} , а ядро — $2\pi\mathbb{Z}$.

Пример 5. Формула умножения определителей означает, что отображение

$$\det: \mathrm{GL}_n(K) \rightarrow K^*, \quad A \mapsto \det A,$$

является гомоморфизмом. Его ядро — это унимодулярная группа $\mathrm{SL}_n(K)$.

Пример 6. Назовем *знаком* подстановки $\sigma \in S_n$ и обозначим через $\mathrm{sgn} \sigma$ произведение знаков верхней и нижней перестановки в ее записи (см. пример 1.1):

$$\mathrm{sgn} \begin{pmatrix} i_1 & i_2 & \dots & i_n \\ j_1 & j_2 & \dots & j_n \end{pmatrix} = \mathrm{sgn}(i_1, i_2, \dots, i_n) \cdot \mathrm{sgn}(j_1, j_2, \dots, j_n).$$

Это произведение не зависит от способа записи подстановки σ , так как от любого способа записи можно перейти к любому другому последовательными транспозициями столбиков, а при каждой такой транспозиции одновременно меняются знаки верхней и нижней перестановок, так что их произведение сохраняется. Основное свойство знака состоит в том, что отображение

$$\mathrm{sgn}: S_n \rightarrow C_2 = \{\pm 1\}, \quad \sigma \mapsto \mathrm{sgn} \sigma,$$

является гомоморфизмом. В самом деле, перемножая подстановки σ и τ , мы можем считать, что верхняя перестановка в записи σ совпадает с нижней перестановкой в записи τ :

$$\sigma = \begin{pmatrix} j_1 & j_2 & \dots & j_n \\ k_1 & k_2 & \dots & k_n \end{pmatrix}, \quad \tau = \begin{pmatrix} i_1 & i_2 & \dots & i_n \\ j_1 & j_2 & \dots & j_n \end{pmatrix}.$$

Тогда

$$\sigma\tau = \begin{pmatrix} i_1 & i_2 & \dots & i_n \\ k_1 & k_2 & \dots & k_n \end{pmatrix},$$

так что

$$\begin{aligned} \mathrm{sgn} \sigma\tau &= \mathrm{sgn}(i_1, i_2, \dots, i_n) \cdot \mathrm{sgn}(k_1, k_2, \dots, k_n) = \\ &= [\mathrm{sgn}(i_1, i_2, \dots, i_n) \mathrm{sgn}(j_1, j_2, \dots, j_n)] \times \\ &\quad \times [\mathrm{sgn}(j_1, j_2, \dots, j_n) \mathrm{sgn}(k_1, k_2, \dots, k_n)] = \\ &= \mathrm{sgn} \tau \cdot \mathrm{sgn} \sigma = \mathrm{sgn} \sigma \cdot \mathrm{sgn} \tau. \end{aligned}$$

Ядро гомоморфизма sgn называется *знакопеременной группой* и обозначается через A_n . Употребляется также следующая терминология: подстановки σ , для которых $\text{sgn } \sigma = 1$ (соответственно $\text{sgn } \sigma = -1$), называются *четными* (соответственно *нечетными*). Таким образом, A_n — это подгруппа четных подстановок.

Задача 1. Вывести следующую формулу для знака циклической подстановки:

$$\text{sgn}(i_1 i_2 \dots i_p) = (-1)^{p-1}.$$

Используя это, доказать, что знак любой подстановки равен $(-1)^{m-s}$, где m — число фактически переставляемых ею (т. е. не оставляемых на месте) символов, а s — число независимых циклов, в произведении которых она разлагается.

Теорема 1 (о гомоморфизме групп). Пусть $f: G \rightarrow H$ — гомоморфизм групп. Тогда

$$\text{Im } f \simeq G/\text{Ker } f.$$

Более точно, имеется изоморфизм

$$\varphi: \text{Im } f \xrightarrow{\simeq} G/\text{Ker } f,$$

ставящий в соответствие каждому элементу $h = f(g) \in \text{Im } f$ смежный класс $g \text{ Ker } f$.

Доказательство. Доказательство этой теоремы аналогично доказательству теоремы 5.2. Из доказанного выше свойства 5) следует, что все элементы смежного класса $g \text{ Ker } f$, и только они, переходят при гомоморфизме f в элемент $h = f(g) \in \text{Im } f$. Тем самым показано, что отображение φ , о котором идет речь в теореме, корректно определено и биективно. Остается проверить, что φ — гомоморфизм.

Пусть $g_1, g_2 \in G$, $f(g_1) = h_1$, $f(g_2) = h_2$. Тогда $f(g_1 g_2) = h_1 h_2$ и

$$\varphi(h_1 h_2) = g_1 g_2 \text{ Ker } f = (g_1 \text{ Ker } f)(g_2 \text{ Ker } f) = \varphi(h_1) \varphi(h_2),$$

что и требовалось доказать. □

Следствие. Если группа G конечна, то

$$|G| = |\text{Im } f| |\text{Ker } f|.$$

(Интересно сравнить эту формулу с формулой (15).)

Пример 7. Рассмотрим гомоморфизм

$$f: \mathbb{C} \rightarrow \mathbb{R}, \quad z \mapsto \text{Im } z.$$

Имеем $\text{Im } f = \mathbb{R}$, $\text{Ker } f = \mathbb{R}$, так что

$$\mathbb{C}/\mathbb{R} \simeq \mathbb{R}$$

— результат, уже полученный нами в примере 5.13.

Пример 8. Рассмотрим гомоморфизм

$$f: \mathbb{C}^* \rightarrow \mathbb{R}_+^*, \quad z \mapsto |z|.$$

Имеем $\text{Im } f = \mathbb{R}_+^*$, $\text{Ker } f = \mathbb{T} = \{z \in \mathbb{C}^* : |z| = 1\}$, так что

$$\mathbb{C}^*/\mathbb{T} \simeq \mathbb{R}_+^*$$

— результат, уже полученный нами в примере 5.14.

Пример 9. Отображение

$$f: \mathbb{C}^* \rightarrow \mathbb{T}, \quad z \mapsto \frac{z}{|z|},$$

также является гомоморфизмом, причем $\text{Im } f = \mathbb{T}$, $\text{Ker } f = \mathbb{R}_+^*$. Следовательно,

$$\mathbb{C}^*/\mathbb{R}_+^* \simeq \mathbb{T}.$$

(Соответствующее разбиение на смежные классы было описано в примере 5.2.)

Пример 10. Рассмотрим гомоморфизм

$$f: \mathbb{R} \rightarrow \mathbb{T}, \quad x \mapsto \cos 2\pi x + i \sin 2\pi x$$

(ср. пример 4). Так как $\text{Ker } f = \mathbb{Z}$, то мы получаем, что

$$\mathbb{R}/\mathbb{Z} \simeq \mathbb{T}.$$

Пример 11. Аналогичным образом рассмотрение гомоморфизма \det из примера 5 приводит к тому, что

$$\text{GL}_n(K)/\text{SL}_n(K) \simeq K^*$$

— результат, уже полученный в примере 5.15.

Пример 12. Рассмотрение гомоморфизма sgn из примера 6 приводит к тому, что при $n > 1$

$$S_n/A_n \simeq C_2.$$

В частности, отсюда следует, что

$$|A_n| = \frac{1}{2}n!.$$

Пример 13. Согласно определению (см. § 2), всякое аффинное преобразование f есть произведение параллельного переноса и линейного преобразования φ . Последнее называется *линейной частью* или *дифференциалом* преобразования f и обозначается через df . Формула

$$(t_a \varphi)(t_b \psi) = t_{a+\varphi(b)} \varphi \psi,$$

полученная при доказательстве предложения 2.1, показывает, что отображение

$$d: GA_n(K) \rightarrow GL_n(K), \quad f \mapsto df,$$

является гомоморфизмом. Очевидно, что

$$\text{Im } d = GL_n(K), \quad \text{Ker } d = \text{Trans}(K^n),$$

так что

$$GA_n(K)/\text{Trans}(K^n) \simeq GL_n(K).$$

Пример 14. Пусть $\Delta = A_1 A_2 A_3$ — правильный треугольник. Составив каждому движению $\varphi \in \text{Sym } \Delta$ подстановку $\sigma \in S_3$ по правилу

$$\varphi(A_i) = A_{\sigma(i)},$$

мы получим гомоморфизм

$$f: \text{Sym } \Delta \rightarrow S_3.$$

Так как всякое движение плоскости, оставляющее на месте 3 точки, не лежащие на одной прямой, тождественно, то $\text{Ker } f = \{\text{id}\}$. Докажем, что $\text{Im } f = S_3$. Так как $\text{Im } f$ — подгруппа группы S_3 и группа S_3 порождается транспозициями, то достаточно проверить, что любая транспозиция принадлежит $\text{Im } f$, т. е. может быть осуществлена некоторым движением $\varphi \in \text{Sym } \Delta$. Но это действительно так: например, транспозиция (12) осуществляется отражением относительно прямой l , показанной на рис. 7. Таким образом,

$$\text{Sym } \Delta \simeq S_3.$$

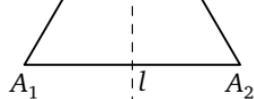


Рис. 7

Аналогично доказывается, что группа симметрии правильного тетраэдра изоморфна S_4 (проделайте это!).

Пример 15. При перестановках переменных x_1, x_2, x_3, x_4 многочлены

$$x_1x_2 + x_3x_4, \quad x_1x_3 + x_2x_4, \quad x_1x_4 + x_2x_3 \quad (18)$$

переставляются между собой. Занумеровав их каким-либо образом, мы получим гомоморфизм

$$f: S_4 \rightarrow S_3.$$

Докажем, что $\text{Im } f = S_3$. Для этого достаточно проверить, что любая транспозиция многочленов (18) может быть осуществлена некоторой перестановкой переменных x_1, x_2, x_3, x_4 . Но это действительно так: например, транспозиция первых двух многочленов (18) может быть осуществлена транспозицией переменных x_2 и x_3 . Далее, $\text{Ker } f$ — это так называемая *четверная группа Клейна*

$$V_4 = \{e, (12)(34), (13)(24), (14)(23)\}.$$

По теореме о гомоморфизме $V_4 \triangleleft S_4$ и $S_4/V_4 \simeq S_3$. Легко видеть, что группа V_4 изоморфна группе из примера 1.7.

Задача 2. Доказать, что для любого $n \in \mathbb{N}$ имеет место следующий «парадоксальный» изоморфизм:

$$\mathbb{C}^*/C_n \simeq \mathbb{C}^*.$$

Задача 3. Пусть p — простое число. Найти порядки групп $\text{GL}_2(\mathbb{Z}_p)$ и $\text{SL}_2(\mathbb{Z}_p)$.

Очевидно, что композиция гомоморфизмов $F \rightarrow G$ и $G \rightarrow H$ есть гомоморфизм $F \rightarrow H$.

Пример 16. Рассмотрим композицию гомоморфизмов

$$\det: \text{GL}_n(\mathbb{R}) \rightarrow \mathbb{R}^* \quad \text{и} \quad \text{sgn}: \mathbb{R}^* \rightarrow C_2 = \{\pm 1\},$$

где sgn обозначает знак вещественного числа. Мы получим таким образом гомоморфизм

$$\varepsilon: \text{GL}_n(\mathbb{R}) \rightarrow C_2.$$

При $n = 2$ он имеет следующий геометрический смысл: если $\varepsilon(A) = 1$ (соответственно $\varepsilon(A) = -1$), то линейное преобразование пространства E^2 , определяемое матрицей A , сохраняет (соответственно меняет) ориентацию в том смысле, что любой положительно ориентированный базис оно переводит в положительно (соответственно отрицательно) ориентированный базис. Аналогичная интерпретация возможна и при $n = 3$.

Пример 17. Композиция гомоморфизмов

$$d: \text{GA}_n(\mathbb{R}) \rightarrow \text{GL}_n(\mathbb{R}) \quad \text{и} \quad \varepsilon: \text{GL}_n(\mathbb{R}) \rightarrow C_2$$

есть гомоморфизм

$$\text{GA}_n(\mathbb{R}) \rightarrow C_2. \quad (19)$$

При $n = 2$ и 3 это позволяет распространить на аффинные преобразования евклидовой плоскости и евклидова пространства понятие сохранения или изменения ориентации. А именно, аффинное преобразование сохраняет (соответственно меняет) ориентацию, если его дифференциал сохраняет (соответственно меняет) ориентацию. В частности, можно говорить о движениях, сохраняющих или меняющих ориентацию (что мы уже делали раньше, не давая точного определения).

Пример 18. Пусть $G \subset \text{Isom } E^n$ ($n = 2$ или 3) — какая-либо подгруппа, содержащая движения, меняющие ориентацию. Рассматривая ограничение на G гомоморфизма (19), мы приходим к выводу, что подмножество движений из G , сохраняющих ориентацию, есть подгруппа индекса 2. Будем обозначать эту подгруппу через G_+ .

Пример 19. В частности, подгруппу $\text{Sym}_+ K \subset \text{Sym } K$ будем называть *группой вращений куба* K . Так как $|\text{Sym } K| = 48$ (см. пример 5.10), а $\text{Sym}_+ K$ есть подгруппа индекса 2, то

$$|\text{Sym}_+ K| = 24.$$

Докажем, что

$$\text{Sym}_+ K \simeq S_4.$$

Для этого занумеруем каким-либо образом 4 диагонали куба K и поставим в соответствие каждому движению $\varphi \in \text{Sym}_+ K$ подстановку, осуществляемую им на множестве диагоналей. Мы получим гомоморфизм

$$f: \text{Sym}_+ K \rightarrow S_4.$$

Докажем, что $\text{Im } f = S_4$, откуда уже будет следовать, что f — изоморфизм, поскольку $|\text{Sym}_+ K| = |S_4|$. Для этого достаточно проверить, что любая транспозиция принадлежит $\text{Im } f$. Но это действительно так: например, транспозиция (12) осуществляется поворотом на π вокруг прямой l , изображенной на рис. 8.

Задача 4. Доказать, что группа D_4 (группа симметрии квадрата) изоморфна группе $\text{Sym}(x_1x_2 + x_3x_4)$ (см. примеры 1.8 и 5.11).

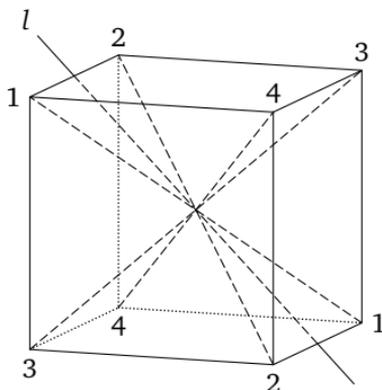


Рис. 8

Задача 5. Доказать, что $SL_2(\mathbb{Z}_2) \simeq S_3$.

Согласно определению операции в факторгруппе G/N , отображение

$$a: G \rightarrow G/N, \quad g \mapsto gN,$$

является гомоморфизмом. Оно называется *каноническим гомоморфизмом* группы G на факторгруппу G/N . Его ядром, очевидно, является подгруппа N .

Пусть $f: G \rightarrow H$ — любой сюръективный гомоморфизм. Положим $\text{Ker } f = N$. Согласно теореме 1, $H \simeq G/N$ и, если отождествить H с G/N при помощи указанного там изоморфизма, гомоморфизм f совпадет с каноническим гомоморфизмом группы G на G/N . Поэтому теорему 1 можно понимать таким образом, что никаких сюръективных гомоморфизмов групп, кроме канонических гомоморфизмов на факторгруппы, в сущности, не существует.

Глава 5

Векторные пространства

Эта и последующие две главы будут посвящены линейной алгебре, начала которой были изложены в гл. 2, и связанной с ней геометрии. Линейная алгебра является наиболее прикладным разделом алгебры. Ее аппарат так же необходим любому математику, как аппарат математического анализа.

Следует, однако, предостеречь читателя от взгляда на линейную алгебру как на манипулирование с матрицами — взгляда, игнорирующего ее идеологию, в частности геометрические образы, скрывающиеся за ее понятиями. Читатель, пошедший по этому легкому пути, много потеряет. Он будет испещрять формулами десятки страниц или перегружать компьютер в ситуациях, очевидных для того, кто действительно владеет линейной алгеброй.

За исключением общих определений, некоторых примеров и тех случаев, когда будет оговорено противное, все векторные пространства в главах, относящихся к линейной алгебре, предполагаются конечномерными. Основное поле, если это не есть какое-то конкретное поле, обозначается буквой K .

§ 1. Подпространства

Всякий базис подпространства U векторного пространства V можно дополнить до базиса всего пространства. Таким образом мы получим базис, в котором данное подпространство U выглядит весьма просто, а именно, натянута на несколько первых базисных векторов.

Определение 1. Базис пространства V называется *согласованным* с подпространством U , если U является линейной оболочкой какой-то части базисных векторов (т. е. одним из «координатных подпространств» относительно этого базиса).

Так, на рис. 1, а) базис $\{e_1, e_2\}$ согласован с подпространством U , а на рис. 1, б) — нет.

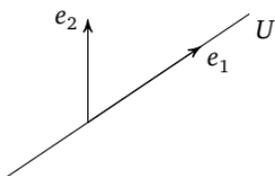


Рис. 1, а)

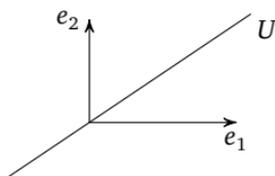


Рис. 1, б)

Согласно сказанному выше для всякого подпространства существует согласованный с ним базис.

Пусть теперь имеются два подпространства $U, W \subset V$. Очевидно, что их пересечение $U \cap W$ также является подпространством. Это наибольшее подпространство, содержащееся как в U , так и в W .

Определение 2. Суммой $U + W$ подпространств U и W называется совокупность векторов вида $u + w$, где $u \in U, w \in W$.

Это наименьшее подпространство, содержащее как U , так и W . Иными словами, это линейная оболочка объединения $U \cup W$.

Теорема 1. Для всякой пары подпространств $U, W \subset V$ существует базис пространства V , согласованный с каждым из подпространств U, W .

Доказательство. Пусть $\{e_1, \dots, e_p\}$ — базис подпространства $U \cap W$. Дополним его какими-то векторами e_{p+1}, \dots, e_k до базиса подпространства U и, с другой стороны, векторами $e_{k+1}, \dots, e_{k+l-p}$ — до базиса подпространства W . (Здесь $p = \dim U \cap W, k = \dim U, l = \dim W$.) Докажем, что векторы e_1, \dots, e_{k+l-p} линейно независимы. Дополнив их затем до базиса подпространства V , мы и получим базис, согласованный как с U , так и с W .

Предположим, что

$$\sum_{i=1}^{k+l-p} \lambda_i e_i = 0.$$

Рассмотрим вектор

$$x = \sum_{i=1}^k \lambda_i e_i = - \sum_{i=k+1}^{k+l-p} \lambda_i e_i.$$

Из первого представления вектора x следует, что он лежит в U , а из второго — что он лежит в W . Таким образом, $x \in U \cap W$ и, значит,

$$x = \sum_{i=1}^p \mu_i e_i = - \sum_{i=k+1}^{k+l-p} \lambda_i e_i.$$

Так как векторы $e_1, \dots, e_p, e_{k+1}, \dots, e_{k+l-p}$ линейно независимы, то отсюда следует, что $x = 0$ и $\lambda_i = 0$ при $i = k+1, \dots, k+l-p$. Далее, так как векторы e_1, \dots, e_k линейно независимы, то из равенства

$$\sum_{i=1}^k \lambda_i e_i = 0$$

следует, что $\lambda_i = 0$ при $i = 1, \dots, k$. □

Рисунок 2 иллюстрирует это доказательство в случае $p = 1, k = l = 2$.

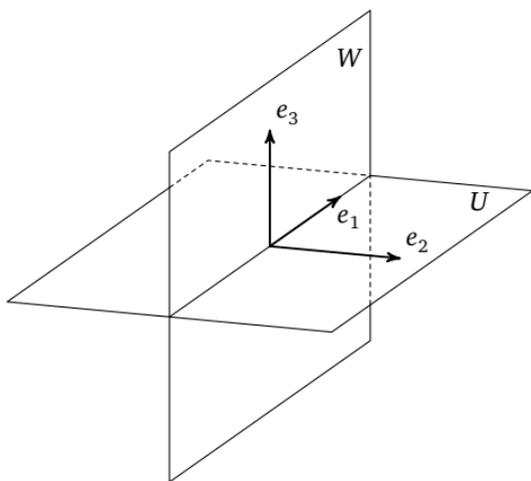


Рис. 2

Следствие. $\dim(U + W) = \dim U + \dim W - \dim(U \cap W)$.

Доказательство. В обозначениях доказательства теоремы 1 векторы e_1, \dots, e_{k+l-p} составляют базис подпространства $U + W$, так что

$$\dim(U + W) = k + l - p. \quad \square$$

Для трех подпространств теорема, аналогичная теореме 1, неверна.

Задача 1. Привести пример, подтверждающий это высказывание.

Теорема 1 описывает взаимное расположение пары подпространств. Взаимное расположение произвольного (конечного) числа подпространств описать, вообще говоря, сложно (и в некотором смысле даже невозможно). Однако нас будет в первую очередь

интересовать один важный частный случай, когда это сделать просто.

Определение 3. Подпространства U_1, \dots, U_k называются *линейно независимыми*, если из равенства $u_1 + \dots + u_k = 0$ ($u_i \in U_i$) следует, что $u_1 = \dots = u_k = 0$.

Для двух подпространств U, W линейная независимость равносильна тому, что $U \cap W = 0$. Напрашивающееся обобщение для любого числа подпространств неверно.

Задача 2. Привести пример трех линейно зависимых подпространств, все попарные пересечения которых равны нулю.

Определение 4. Суммой $U_1 + \dots + U_k$ подпространств $U_1, \dots, U_k \subset V$ называется совокупность векторов вида $u_1 + \dots + u_k$, где $u_i \in U_i$.

Это наименьшее подпространство, содержащее все подпространства U_1, \dots, U_k .

Предложение 1. Следующие свойства системы подпространств $U_1, \dots, U_k \subset V$ равносильны:

- 1) U_1, \dots, U_k линейно независимы;
- 2) объединение базисов подпространств U_1, \dots, U_k линейно независимо;
- 3) $\dim(U_1 + \dots + U_k) = \dim U_1 + \dots + \dim U_k$.

Доказательство. 1) \Leftrightarrow 2). Пусть $\{e_{i1}, \dots, e_{in_i}\}$ — базис подпространства U_i ($i = 1, \dots, k$). Предположим, что между векторами e_{ij} ($i = 1, \dots, k, j = 1, \dots, n_i$) имеется нетривиальная линейная зависимость: $\sum_{i,j} \lambda_{ij} e_{ij} = 0$. Тогда сумма векторов

$$x_i = \sum_j \lambda_{ij} e_{ij} \in U_i \quad (i = 1, \dots, k)$$

равна нулю, причем не все они равны нулю. Следовательно, подпространства U_1, \dots, U_k линейно зависимы.

Обратно, если подпространства U_1, \dots, U_k линейно зависимы, то существуют векторы $x_i \in U_i$ ($i = 1, \dots, k$), не все равные нулю, сумма которых равна нулю. Разложив каждый из них по базису своего подпространства, мы получим нетривиальную линейную зависимость между векторами e_{ij} .

2) \Leftrightarrow 3). Так как объединение базисов подпространств U_1, \dots, U_k порождает сумму $U_1 + \dots + U_k$, то каждое из свойств 2) и 3) равносильно тому, что это объединение является базисом пространства $U_1 + \dots + U_k$. Следовательно, эти свойства равносильны между собой. \square

Определение 5. Говорят, что векторное пространство V разлагается в прямую сумму подпространств U_1, \dots, U_k , если

- 1) подпространства U_1, \dots, U_k линейно независимы;
- 2) $U_1 + \dots + U_k = V$.

В этом случае пишут

$$V = U_1 \oplus \dots \oplus U_k.$$

Каждый вектор $v \in V$ однозначно представляется в виде $v = u_1 + \dots + u_k$, где $u_i \in U_i$; вектор u_i называется *проекцией* вектора v на подпространство U_i . Подчеркнем, что проекция вектора на подпространство U_i зависит не только от этого подпространства, но и от остальных слагаемых разложения.

Пример 1. Квадратная матрица A называется *симметричной*, если $A^T = A$, и *кососимметричной*, если $A^T = -A$. Симметричные (соответственно кососимметричные) матрицы образуют подпространство $L_n^+(K)$ (соответственно $L_n^-(K)$) в пространстве $L_n(K)$ всех матриц. При условии, что $\text{char } K \neq 2$, всякая матрица A может быть представлена в виде суммы симметричной и кососимметричной матриц:

$$A = \frac{1}{2}(A + A^T) + \frac{1}{2}(A - A^T).$$

С другой стороны, при том же условии очевидно, что матрица, симметричная и кососимметричная одновременно, равна нулю. Это означает, что

$$L_n(K) = L_n^+(K) \oplus L_n^-(K).$$

Пример 2. Аналогично доказывается, что пространство всех функций на вещественной прямой является прямой суммой подпространств четных и нечетных функций. (В этом примере векторное пространство и оба подпространства бесконечномерны.)

Пример 3. Пусть $\{e_1, \dots, e_n\}$ — базис векторного пространства V . Тогда

$$V = \langle e_1 \rangle \oplus \dots \oplus \langle e_n \rangle.$$

Проекция вектора $x \in V$ на $\langle e_i \rangle$ равна $x_i e_i$, где x_i есть i -я координата вектора x в базисе $\{e_1, \dots, e_n\}$, и зависит не только от e_i , но и от остальных векторов базиса.

Определения 3, 4 и 5 могут быть распространены на бесконечное число подпространств, если рассматривать только такие суммы

векторов, в которых лишь конечное число слагаемых отлично от нуля.

Пример 4. Пусть $A = K[x_1, \dots, x_n]$ — алгебра многочленов от n переменных. Обозначим через A_d подпространство однородных многочленов степени d . Так как всякий многочлен однозначно представляется в виде суммы однородных многочленов различных степеней, то

$$A = A_0 \oplus A_1 \oplus A_2 \oplus \dots = \bigoplus_{d=0}^{\infty} A_d.$$

При этом

$$A_d A_e \subset A_{d+e}. \quad (1)$$

Разложение какой-либо алгебры A в прямую сумму подпространств A_d ($d \in \mathbb{Z}$), удовлетворяющее условию (1), называется ее *градуировкой*. Алгебра, снабженная градуировкой, называется *градуированной алгеброй*. (Некоторые из подпространств A_d могут быть нулевыми. Так, в приведенном выше примере $A_d = 0$ при $d < 0$.)

Задача 3. Пусть $A = L_n(K)$ — алгебра матриц. Обозначим через A_d линейную оболочку матричных единиц E_{ij} с $j - i = d$. Доказать, что подпространства A_d задают градуировку алгебры A . (Здесь $A_d = 0$ при $|d| \geq n$.)

§ 2. Линейные отображения

Подобно гомоморфизмам групп естественно рассматривать гомоморфизмы векторных пространств. Их принято называть линейными отображениями.

Определение 1. Пусть V и U — векторные пространства над полем K . Отображение

$$\varphi: V \rightarrow U$$

называется *линейным*, если

- 1) $\varphi(a + b) = \varphi(a) + \varphi(b)$ для любых $a, b \in V$;
- 2) $\varphi(\lambda a) = \lambda \varphi(a)$ для любых $\lambda \in K, a \in V$.

Это определение отличается от определения изоморфизма векторных пространств тем, что в нем не требуется биективности. Условие 1) означает, что отображение φ должно быть гомоморфизмом аддитивных групп. Из общих свойств гомоморфизмов групп следует, что

$$\varphi(0) = 0, \quad \varphi(-a) = -\varphi(a), \quad \varphi(a - b) = \varphi(a) - \varphi(b).$$

Пример 1. Поворот есть линейное отображение (и даже изоморфизм) пространства E^2 в себя (рис. 3).

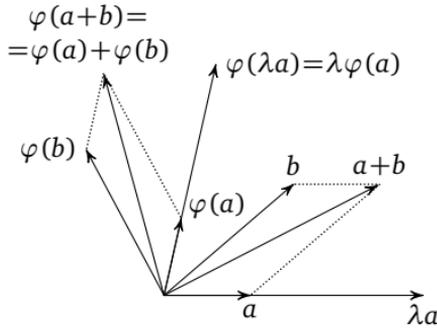


Рис. 3

Пример 2. Ортогональное проектирование на плоскость определяет линейное отображение (но не изоморфизм) пространства E^3 в пространство геометрических векторов этой плоскости.

Пример 3. Дифференцирование является линейным отображением пространства непрерывно дифференцируемых функций на заданном промежутке числовой прямой в пространство непрерывных функций на этом промежутке.

Линейное отображение $\varphi: V \rightarrow U$ однозначно определяется образами базисных векторов пространства V . В самом деле, пусть $\{e_i: i \in I\}$ — базис пространства V ; тогда для любого вектора $x = \sum_i x_i e_i$ имеем

$$\varphi(x) = \sum_i x_i \varphi(e_i).$$

С другой стороны, если $u_i \in U$ ($i \in I$) — произвольные векторы, то отображение $\varphi: V \rightarrow U$, определяемое по формуле

$$\varphi(x) = \sum_i x_i u_i,$$

как легко видеть, является линейным и $\varphi(e_i) = u_i$.

Эти соображения позволяют получить аналитическое описание линейных отображений. Пусть $\varphi: V \rightarrow U$ — линейное отображение конечномерных векторных пространств. Выберем в пространствах V и U какие-нибудь базисы $\{e_1, \dots, e_n\}$ и $\{f_1, \dots, f_m\}$. Применим отображение φ к векторам e_1, \dots, e_n и разложим полученные векторы

пространства U по базису $\{f_1, \dots, f_m\}$:

$$\varphi(e_j) = a_{1j}f_1 + a_{2j}f_2 + \dots + a_{mj}f_m = \sum_i a_{ij}f_i.$$

Числа a_{ij} ($i = 1, 2, \dots, m, j = 1, 2, \dots, n$) образуют матрицу A размера $m \times n$, которая называется *матрицей линейного отображения* φ в выбранных базисах пространств V и U . Согласно этому определению в j -м столбце матрицы A стоят координаты образа j -го базисного вектора пространства V в выбранном базисе пространства U . В матричных обозначениях это можно записать следующим образом:

$$(\varphi(e_1), \dots, \varphi(e_n)) = (f_1, \dots, f_m)A$$

(ср. формулу (8) перехода к другому базису в § 2.2).

Найдем теперь выражение координат y_1, \dots, y_m образа $y = \varphi(x)$ вектора $x \in V$ через координаты x_1, \dots, x_n вектора x . В силу линейности отображения φ имеем

$$y = \varphi\left(\sum_j x_j e_j\right) = \sum_j x_j \varphi(e_j) = \sum_{i,j} a_{ij} x_j f_i,$$

откуда

$$y_i = \sum_{j=1}^n a_{ij} x_j \quad (i = 1, 2, \dots, m). \quad (2)$$

Если обозначить через X и Y столбцы координат векторов x и y соответственно, то равенства (2) можно переписать в следующей матричной форме:

$$Y = AX. \quad (3)$$

(Ср. формулу (9) преобразования координат в § 2.2.)

Пример 4. В пространстве E^2 выберем ортонормированный базис $\{e_1, e_2\}$. Пусть φ — поворот на угол α . Тогда (см. рис. 4)

$$\varphi(e_1) = e_1 \cos \alpha + e_2 \sin \alpha,$$

$$\varphi(e_2) = -e_1 \sin \alpha + e_2 \cos \alpha.$$

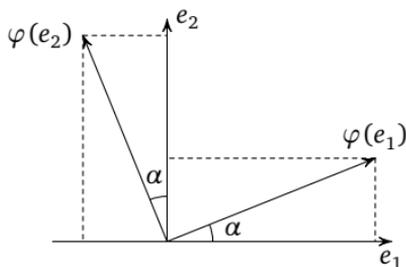


Рис. 4

Это означает, что матрица отображения φ есть

$$\begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix}. \quad (4)$$

Заметим, что в данном случае $V = U$ и мы использовали один и тот же базис $\{e_1, e_2\}$ в двух качествах: как базис пространства V и как базис пространства U , хотя, согласно определению, не обязаны были это делать.

Пример 5. Найдем матрицу проектирования из примера 2. В плоскости проектирования выберем любой базис $\{e_1, e_2\}$ и дополним его ортогональным вектором e_3 до базиса пространства. Так как при проектировании векторы e_1 и e_2 переходят сами в себя, а вектор e_3 — в нуль, то искомая матрица (относительно выбранных базисов) имеет вид

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

Как и для любого гомоморфизма групп, для линейного отображения $\varphi: V \rightarrow U$ определяются его образ

$$\text{Im } \varphi = \{\varphi(a) : a \in V\} \subset U,$$

и ядро

$$\text{Ker } \varphi = \{a \in V : \varphi(a) = 0\} \subset V.$$

Они являются не только подгруппами, но и подпространствами соответствующих векторных пространств, т. е. замкнуты относительно умножений на любые числа. Покажем, например, что $\text{Ker } \varphi$ — подпространство в V . Если $a \in \text{Ker } \varphi$, т. е. $\varphi(a) = 0$, то для любого $\lambda \in K$

$$\varphi(\lambda a) = \lambda \varphi(a) = \lambda 0 = 0,$$

т. е. $\lambda a \in \text{Ker } \varphi$.

Пример 6. Ядром отображения проектирования из примера 2 является совокупность векторов, ортогональных плоскости проектирования.

Пример 7. Ядром отображения дифференцирования из примера 3 является совокупность постоянных функций, а образом — пространство всех непрерывных функций. Последнее следует из существования первообразной у каждой непрерывной функции, доказываемого в анализе.

Согласно общим свойствам гомоморфизмов групп, установленным в § 4.6, линейное отображение $\varphi: V \rightarrow U$ инъективно тогда и только тогда, когда $\text{Ker } \varphi = 0$, а в общем случае полный прообраз любого вектора $b = \varphi(a) \in \text{Im } \varphi$ есть смежный класс $a + \text{Ker } \varphi$.

Последнее утверждение есть «бескоординатная» форма утверждения 2) теоремы 2.1.3 (а утверждение о том, что $\text{Ker } \varphi$ — подпространство, есть бескоординатная форма утверждения 1) этой теоремы).

Что можно сказать о размерностях образа и ядра линейного отображения конечномерных векторных пространств?

Теорема 1. Если в каких-то базисах пространств V и U линейное отображение $\varphi: V \rightarrow U$ имеет матрицу A , то

$$\dim \text{Im } \varphi = \text{rk } A.$$

Доказательство. Очевидно, что $\text{Im } \varphi$ есть линейная оболочка образов базисных векторов e_1, \dots, e_n пространства V и, значит, $\dim \text{Im } \varphi$ есть ранг системы векторов $\varphi(e_1), \dots, \varphi(e_n)$. Но в столбцах матрицы A как раз и записаны координаты этих векторов в каком-то базисе пространства U . Следовательно, ранг этой системы векторов равен рангу матрицы A . \square

Теорема 2. $\dim \text{Im } \varphi + \dim \text{Ker } \varphi = \dim V$.

Доказательство. Выберем базис пространства V специальным образом: сначала возьмем базис $\{e_1, \dots, e_k\}$ подпространства $\text{Ker } \varphi$, а затем дополним его какими-то векторами e_{k+1}, \dots, e_n до базиса пространства V . Так как по построению $\varphi(e_1) = \dots = \varphi(e_k) = 0$, то

$$\text{Im } \varphi = \langle \varphi(e_{k+1}), \dots, \varphi(e_n) \rangle.$$

Докажем, что векторы $\varphi(e_{k+1}), \dots, \varphi(e_n)$ линейно независимы, откуда и будет следовать утверждение теоремы.

Пусть

$$\lambda_1 \varphi(e_{k+1}) + \dots + \lambda_{n-k} \varphi(e_n) = 0.$$

Рассмотрим вектор

$$a = \lambda_1 e_{k+1} + \dots + \lambda_{n-k} e_n.$$

Предыдущее равенство означает, что $\varphi(a) = 0$, т. е.

$$a \in \text{Ker } \varphi = \langle e_1, \dots, e_k \rangle.$$

Так как $e_1, \dots, e_k, e_{k+1}, \dots, e_n$ линейно независимы, то это возможно только при $\lambda_1 = \dots = \lambda_{n-k} = 0$, что и требовалось доказать. \square

Записывая отображение φ в координатах, мы можем из теорем 1 и 2 вывести теорему 2.3.3 о размерности пространства решений системы однородных линейных уравнений. Таким образом, понятие линейного отображения позволяет получить более простое доказательство теоремы 2.3.3.

Пример 8. Пусть X — множество ребер тетраэдра и Y — множество его граней. Каждой функции f на X со значениями в поле K поставим в соответствие функцию g на Y , определяемую следующим образом:

$$g(y) = \sum_{x \subset y} f(x),$$

т. е. значение функции g на какой-либо грани равно сумме значений функции f на сторонах этой грани. Этим определяется линейное отображение

$$\varphi: F(X, K) \rightarrow F(Y, K)$$

(см. пример 1.6.2). Докажем, что если $\text{char } K \neq 2$, то оно сюръективно. Для этого достаточно показать, что $\text{Im } \varphi$ содержит δ -функции всех граней (см. пример 2.2.7). Функция f , для которой $\varphi(f)$ есть δ -функция нижней грани, изображена на рис. 5, а) (ее значения на непомеченных ребрах равны нулю). Так как

$$\dim F(X, K) = 6, \quad \dim F(Y, K) = 4,$$

то по теореме 2

$$\dim \text{Ker } \varphi = 6 - 4 = 2.$$

Функции, составляющие базис подпространства $\text{Ker } \varphi$, изображены на рис. 5, б).

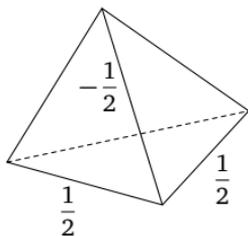


Рис. 5, а)

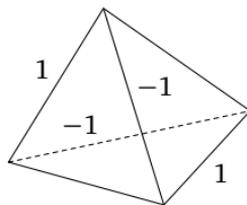
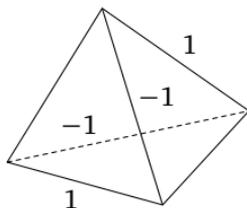


Рис. 5, б)

Задача 1. Для отображения φ из предыдущего примера найти $\dim \text{Ker } \varphi$ в случае, когда $\text{char } K = 2$.

Обратимся теперь к операциям над линейными отображениями. Линейные отображения $V \rightarrow U$ можно складывать и умножать на числа, как обычные функции:

$$(\varphi + \psi)(a) = \varphi(a) + \psi(a),$$

$$(\lambda\varphi)(a) = \lambda\varphi(a).$$

Относительно этих операций они образуют векторное пространство.

Далее, если

$$\varphi: V \rightarrow U, \quad \psi: W \rightarrow V$$

— линейные отображения, то их произведение (композиция)

$$\varphi\psi: W \rightarrow U$$

есть также линейное отображение. В самом деле,

$$\begin{aligned} (\varphi\psi)(a+b) &= \varphi(\psi(a+b)) = \varphi(\psi(a) + \psi(b)) = \\ &= \varphi(\psi(a)) + \varphi(\psi(b)) = (\varphi\psi)(a) + (\varphi\psi)(b), \\ (\varphi\psi)(\lambda a) &= \varphi(\psi(\lambda a)) = \varphi(\lambda\psi(a)) = \lambda\varphi(\psi(a)) = \lambda(\varphi\psi)(a). \end{aligned}$$

Умножение линейных отображений связано с линейными операциями свойствами

$$\begin{aligned} \varphi(\psi + \omega) &= \varphi\psi + \varphi\omega, \quad (\varphi + \psi)\omega = \varphi\omega + \psi\omega, \\ (\lambda\varphi)\psi &= \varphi(\lambda\psi) = \lambda(\varphi\psi) \quad \forall \lambda \in K. \end{aligned}$$

Докажем, например, первое из свойств дистрибутивности. Пусть

$$\varphi: V \rightarrow U, \quad \psi: W \rightarrow V, \quad \omega: W \rightarrow V$$

— линейные отображения. Для любого $a \in W$ имеем

$$\begin{aligned} (\varphi(\psi + \omega))(a) &= \varphi((\psi + \omega)(a)) = \varphi(\psi(a) + \omega(a)) = \\ &= \varphi(\psi(a)) + \varphi(\omega(a)) = (\varphi\psi)(a) + (\varphi\omega)(a) = (\varphi\psi + \varphi\omega)(a). \end{aligned}$$

Умножение линейных отображений ассоциативно, как и вообще умножение любых отображений. В самом деле, пусть M, N, P, Q — какие-то множества и

$$\varphi: N \rightarrow M, \quad \psi: P \rightarrow N, \quad \omega: Q \rightarrow P$$

— какие-то отображения. Тогда для любого $a \in Q$ имеем

$$\begin{aligned} ((\varphi\psi)\omega)(a) &= (\varphi\psi)(\omega(a)) = \varphi(\psi(\omega(a))), \\ (\varphi(\psi\omega))(a) &= \varphi((\psi\omega)(a)) = \varphi(\psi(\omega(a))), \end{aligned}$$

откуда

$$(\varphi\psi)\omega = \varphi(\psi\omega).$$

Операции над линейными отображениями соответствуют таким же операциям над их матрицами. Для линейных операций это очевидно. Докажем это для умножения. Пусть

$$\varphi: V \rightarrow U, \quad \psi: W \rightarrow V$$

— линейные отображения с матрицами $A = (a_{ij})$ и $B = (b_{jk})$ в базисах $\{e_1, \dots, e_n\}$, $\{f_1, \dots, f_m\}$, $\{g_1, \dots, g_p\}$ пространств V , U , W соответственно. Тогда

$$(\varphi\psi)(g_k) = \varphi(\psi(g_k)) = \varphi\left(\sum_j b_{jk}e_j\right) = \sum_j b_{jk}\varphi(e_j) = \sum_{i,j} a_{ij}b_{jk}f_i,$$

Следовательно, матрица отображения $\varphi\psi$ есть $C = (c_{ik})$, где

$$c_{ik} = \sum_j a_{ij}b_{jk}.$$

Это означает, что $C = AB$, что и требовалось доказать.

Пример 9. Матричное равенство, доказанное в примере 1.8.2, на языке линейных отображений означает, что произведение поворотов плоскости на углы α и β есть поворот на угол $\alpha + \beta$ (см. пример 4). Поскольку последнее утверждение геометрически очевидно, это дает доказательство формул для косинуса и синуса суммы двух углов.

Свойства операций над матрицами, полученные нами в § 1.8 прямыми вычислениями, могут быть теперь выведены из соответствующих свойств операций над линейными отображениями.

Очевидно, что тождественное отображение

$$\text{id}: V \rightarrow V$$

линейно. Матрица тождественного отображения $\text{id}: V \rightarrow V$ в двух одинаковых базисах пространства V есть единичная матрица E . Поэтому свойства единичной матрицы (см. формулу (10) § 1.8) есть просто перевод на матричный язык очевидных равенств

$$\varphi \cdot \text{id} = \varphi, \quad \text{id} \cdot \varphi = \varphi,$$

где $\varphi: V \rightarrow V$ — линейное отображение, задаваемое матрицей A , а id в первом случае обозначает тождественное отображение пространства V , а во втором — тождественное отображение пространства U .

Напомним, что отображение множеств обратимо тогда и только тогда, когда оно биективно. Если $\varphi: V \rightarrow U$ — биективное линейное

отображение, то обратное отображение $\varphi^{-1}: U \rightarrow V$ также линейно. В самом деле, для любых $a, b \in U$ пусть $c, d \in V$ — такие векторы, что $\varphi(c) = a$, $\varphi(d) = b$; тогда $\varphi(c + d) = a + b$ и, следовательно,

$$\varphi^{-1}(a + b) = c + d = \varphi^{-1}(a) + \varphi^{-1}(b).$$

Аналогично проверяется и второе свойство линейности.

Очевидно, что линейное отображение $\varphi: V \rightarrow U$ биективно тогда и только тогда, когда $\text{Im } \varphi = U$ и $\text{Ker } \varphi = 0$. В этом случае оно является изоморфизмом, а матрицей обратного отображения (в тех же базисах) служит матрица, обратная матрице φ .

Задача 2. Используя линейные отображения, доказать, что ранг произведения двух матриц (не обязательно квадратных) не превосходит ранга каждой из них (теорема 2.3.5), а если одна из этих матриц невырождена, то ранг произведения равен рангу другой матрицы.

§ 3. Сопряженное пространство

Векторные пространства представляют мир, в котором живут персонажи линейной алгебры. Простейшими из них, помимо векторов, являются линейные функции, которые, как мы увидим, в некотором смысле двойственны векторам.

Напомним определение линейной функции, введенное нами в § 2.4.

Определение 1. *Линейной функцией (или линейной формой)* на векторном пространстве V называется всякая функция $\alpha: V \rightarrow K$, обладающая свойствами

- 1) $\alpha(x + y) = \alpha(x) + \alpha(y)$;
- 2) $\alpha(\lambda x) = \lambda \alpha(x)$.

Иными словами, линейная функция — это линейное отображение пространства V в поле K , рассматриваемое как (одномерное) векторное пространство.

Пример 1. Как доказывалось в курсе элементарной геометрии, функция $\alpha(x) = (a, x)$ ($a \in E^3$) является линейной функцией на пространстве E^3 .

Пример 2. Функция $\alpha(f) = f(x_0)$ ($x_0 \in X$) является линейной функцией на пространстве $F(X, K)$ функций на множестве X со значениями в K (см. пример 1.6.2).

Пример 3. Функция $\alpha(f) = f'(x_0)$ ($x_0 \in \mathbb{R}$) является линейной функцией на пространстве $C^1(\mathbb{R})$ дифференцируемых функций на вещественной прямой.

Пример 4. Функция $\alpha(f) = \int_a^b f(x) dx$ является линейной функцией на пространстве $C[a, b]$ непрерывных функций на отрезке $[a, b]$.

Пример 5. Следом квадратной матрицы называется сумма ее диагональных элементов. След матрицы X обозначается через $\text{tr } X$. Функция $\alpha(X) = \text{tr } X$ является линейной функцией на пространстве $L_n(K)$ квадратных матриц.

Если x_1, \dots, x_n — координаты вектора x в базисе $\{e_1, \dots, e_n\}$, то

$$\alpha(x) = a_1 x_1 + \dots + a_n x_n, \quad (5)$$

где $a_i = \alpha(e_i)$. Таким образом, линейная функция однозначно определяется своими значениями на базисных векторах, называемыми ее коэффициентами в данном базисе. Коэффициенты могут быть произвольными: для любых $a_1, \dots, a_n \in K$ функция α , определяемая формулой (5), является линейной.

Линейные функции образуют подпространство в пространстве $F(V, K)$ всех функций на V со значениями в K .

Определение 2. Пространство линейных функций на V называется сопряженным пространством по отношению к V и обозначается через V^* .

Пусть $\{e_1, \dots, e_n\}$ — базис пространства V . Линейные функции $\varepsilon_1, \dots, \varepsilon_n \in V^*$, определяемые равенствами $\varepsilon_i(x) = x_i$, называются координатными функциями относительно базиса $\{e_1, \dots, e_n\}$. Они составляют базис пространства V^* , который называется сопряженным базисом по отношению к $\{e_1, \dots, e_n\}$. Из его определения следует, что для любого вектора $x \in V$

$$x = \sum_i \varepsilon_i(x) e_i. \quad (6)$$

Сопряженный базис может быть также определен условиями

$$\varepsilon_i(e_j) = \delta_{ij} \doteq \begin{cases} 1 & \text{при } i = j, \\ 0 & \text{при } i \neq j \end{cases} \quad (\text{символ Кронекера}).$$

Из предыдущего следует, что $\dim V^* = \dim V$, так что пространства V и V^* изоморфны, хотя между ними и не существует никакого естественного (выделенного) изоморфизма. Однако второе сопряжен-

ное пространство $V^{**} = (V^*)^*$ оказывается естественно изоморфным пространству V .

Из определения операций в пространстве V^* следует, что для любого вектора $x \in V$ функция f_x на V^* , определенная по формуле

$$f_x(\alpha) = \alpha(x),$$

является линейной.

Теорема 1. *Отображение $x \mapsto f_x$ является изоморфизмом пространства V на пространство V^{**} .*

Доказательство. Из определения линейных функций следует, что $f_{x+y} = f_x + f_y$ и $f_{\lambda x} = \lambda f_x$. Остается проверить, что отображение $x \mapsto f_x$ биективно. Пусть $\{e_1, \dots, e_n\}$ — базис пространства V и $\{\varepsilon_1, \dots, \varepsilon_n\}$ — сопряженный базис пространства V^* . Тогда

$$f_{e_i}(\varepsilon_j) = \varepsilon_j(e_i) = \delta_{ij},$$

так что $\{f_{e_1}, \dots, f_{e_n}\}$ — базис пространства V^{**} , сопряженный базису $\{\varepsilon_1, \dots, \varepsilon_n\}$. Отображение $x \mapsto f_x$ переводит вектор с координатами x_1, \dots, x_n в базисе $\{e_1, \dots, e_n\}$ пространства V в вектор с такими же координатами в базисе $\{f_{e_1}, \dots, f_{e_n}\}$ пространства V^{**} . Следовательно, оно биективно. \square

В дальнейшем мы будем отождествлять пространства V и V^{**} посредством указанного изоморфизма, т. е. рассматривать каждый вектор $x \in V$ одновременно и как линейную функцию на V^* (и писать $x(\alpha)$ вместо $f_x(\alpha)$). При таком соглашении пространства V и V^* будут играть совершенно симметричную роль.

Следствие. *Всякий базис пространства V^* сопряжен некоторому базису пространства V .*

Задача 1. Показать, что линейные функции $\varepsilon_1, \dots, \varepsilon_n$ (где $n = \dim V$) составляют базис пространства V^* тогда и только тогда, когда не существует ненулевого вектора $x \in V$, для которого $\varepsilon_1(x) = \dots = \varepsilon_n(x) = 0$.

Задача 2. Пусть V — пространство многочленов степени $\leq n$ над полем K . Показать, что линейные функции $\varepsilon_0, \varepsilon_1, \dots, \varepsilon_n$, определяемые равенствами

$$\varepsilon_i(f) = f(x_i),$$

где x_0, x_1, \dots, x_n — различные элементы поля K , составляют базис пространства V^* , и найти сопряженный базис пространства V . Показать, что формула (6) в этом случае превращается в интерполяционную формулу Лагранжа.

Задача 3. Пусть V то же, что и в задаче 2, причем $\text{char } K = 0$. Показать, что линейные функции $\varepsilon_0, \varepsilon_1, \dots, \varepsilon_n$, определяемые равенствами

$$\varepsilon_i(f) = f^{(i)}(x_0),$$

где $x_0 \in K$, составляют базис пространства V^* , и найти сопряженный базис пространства V . Показать, что формула (6) в этом случае превращается в формулу Тейлора.

Замечание 1. Теорема 1 неверна для бесконечномерных пространств. Если пространство V бесконечномерно, то пространство V^* и, тем более, V^{**} имеют бóльшую размерность. Например, пусть $V = K^\infty$ — пространство финитных последовательностей (см. пример 2.2.9). Это пространство счетномерно. Линейные функции на нем имеют вид

$$\alpha(x_1, x_2, \dots) = a_1 x_1 + a_2 x_2 + \dots$$

(ввиду финитности последовательности (x_1, x_2, \dots) сумма фактически конечна). Здесь a_1, a_2, \dots могут быть произвольными элементами поля K . Поэтому пространство V^* изоморфно пространству в с е х последовательностей, которое, как можно показать (попробуйте это сделать!), несчетномерно.

Имеется естественное взаимно однозначное соответствие между подпространствами пространств V и V^* , при котором каждому k -мерному подпространству пространства V соответствует $(n - k)$ -мерное подпространство пространства V^* (где $n = \dim V$).

Определение 3. *Аннулятором* подпространства $U \subset V$ называется подпространство

$$U^0 = \{\alpha \in V^* : \alpha(x) = 0 \ \forall x \in U\}.$$

Теорема 2. $\dim U^0 = \dim V - \dim U$.

Доказательство. Пусть $\{e_1, \dots, e_n\}$ — такой базис пространства V , что $U = \langle e_1, \dots, e_k \rangle$, и $\{\varepsilon_1, \dots, \varepsilon_n\}$ — сопряженный базис пространства V^* . Тогда $U^0 = \langle \varepsilon_{k+1}, \dots, \varepsilon_n \rangle$. \square

В соответствии с нашим отождествлением пространств V^{**} и V мы можем говорить об аннуляторе подпространства $W \subset V^*$ как о подпространстве пространства V . По определению

$$W^0 = \{x \in V : \alpha(x) = 0 \ \forall \alpha \in W\}.$$

Теорема 3. $(U^0)^0 = U$ для любого подпространства $U \subset V$.

Доказательство. В обозначениях доказательства теоремы 2, ясно, что $(U^0)^0 = \langle e_1, \dots, e_k \rangle = U$. \square

Следствие. Любое подпространство в V является аннулятором некоторого подпространства в V^* .

Пусть имеется система однородных линейных уравнений

$$\sum_{j=1}^n a_{ij}x_j = 0 \quad (i = 1, \dots, m). \quad (7)$$

Будем интерпретировать x_1, \dots, x_n как координаты вектора x n -мерного векторного пространства V в некотором базисе $\{e_1, \dots, e_n\}$. Тогда система (7) может быть записана в виде

$$\alpha_i(x) = 0 \quad (i = 1, \dots, m),$$

где $\alpha_1, \dots, \alpha_m \in V^*$ — линейные функции, стоящие в левых частях уравнений (7). Множество решений этой системы представляет собой аннулятор подпространства $\langle \alpha_1, \dots, \alpha_m \rangle \subset V^*$. Заметим, что размерность этого подпространства равна рангу матрицы коэффициентов системы (7). Поэтому теорема 2.3.3 о размерности пространства решений системы однородных линейных уравнений является непосредственным следствием теоремы 2.

Следствие теоремы 3 в этом контексте может быть сформулировано так:

Теорема 4. Всякое подпространство является множеством решений некоторой системы однородных линейных уравнений.

§ 4. Билинейные и квадратичные функции

Аксиоматика векторного пространства не охватывает еще всей элементарной геометрии векторов евклидова пространства, поскольку в этой аксиоматике отсутствуют такие понятия, как длина вектора и угол между векторами. Длина и угол могут быть выражены через скалярное произведение векторов. Одним из основных свойств скалярного умножения геометрических векторов является его линейность по каждому множителю. В этом параграфе мы рассмотрим функции двух векторных аргументов, являющиеся обобщением скалярного умножения.

Определение 1. Билинейной функцией (или билинейной формой) на векторном пространстве V называется функция $\alpha: V \times V \rightarrow K$, линейная по каждому аргументу.

Пример 1. Как доказывается в курсе элементарной геометрии, скалярное умножение геометрических векторов является билинейной функцией на пространстве E^3 .

Пример 2. Функция

$$\alpha(f, g) = \int_a^b f(x)g(x) dx$$

является билинейной функцией на пространстве $C[a, b]$.

Пример 3. Функция

$$\alpha(X, Y) = \text{tr } XY$$

является билинейной функцией на пространстве $L_n(K)$.

Пример 4. Определитель матрицы второго порядка как функция ее строк есть билинейная функция на пространстве K^2 .

Пусть $\{e_1, \dots, e_n\}$ — базис пространства V . Тогда для векторов $x = \sum_i x_i e_i$, $y = \sum_j y_j e_j$ получаем

$$\alpha(x, y) = \sum_{i,j} a_{ij} x_i y_j, \quad \text{где } a_{ij} = \alpha(e_i, e_j). \quad (8)$$

Матрица $A = (a_{ij})$ называется *матрицей билинейной функции* α в базисе $\{e_1, \dots, e_n\}$. Как видно из предыдущей формулы, билинейная функция однозначно определяется своей матрицей.

Формула (8) может быть переписана в матричных обозначениях:

$$\alpha(x, y) = X^T A Y, \quad (9)$$

где X и Y — столбцы координат векторов x и y соответственно.

При переходе к другому базису

$$(e'_1, \dots, e'_n) = (e_1, \dots, e_n)C$$

координаты векторов x и y преобразуются по формулам

$$X = CX', \quad Y = CY'.$$

Подставляя эти выражения в (9), получаем

$$\alpha(x, y) = (X')^T C^T A C Y',$$

откуда следует, что в базисе $\{e'_1, \dots, e'_n\}$ матрица функции α равна

$$A' = C^T A C. \quad (10)$$

Основная задача теории билинейных функций — это приведение матрицы билинейной функции к возможно более простому виду за счет выбора подходящего базиса. В связи с этим важно знать свойства матрицы билинейной функции, которые не зависят от выбора базиса.

Определение 2. Ядром билинейной функции α называется подпространство

$$\text{Ker } \alpha = \{y \in V : \alpha(x, y) = 0 \ \forall x \in V\}.$$

Функция α называется невырожденной, если $\text{Ker } \alpha = 0$.

Все билинейные функции, рассмотренные в примерах 1—4, невырожденны. Так, невырожденность скалярного умножения следует из того, что $(y, y) > 0$ при $y \neq 0$. Аналогично доказывается невырожденность билинейной функции в примере 2.

Задача 1. Доказать невырожденность билинейных функций в примерах 3 и 4.

Очевидно, что если $\{e_1, \dots, e_n\}$ — базис пространства V , то

$$\text{Ker } \alpha = \{y \in V : \alpha(e_i, y) = 0, \ i = 1, \dots, n\}.$$

Записывая эти условия в координатах, получаем систему однородных линейных уравнений, матрицей коэффициентов которой служит матрица A функции α . Следовательно,

$$\dim \text{Ker } \alpha = n - \text{rk } A. \quad (11)$$

В частности, $\text{Ker } \alpha = 0$ тогда и только тогда, когда $\text{rk } A = n$, т. е. когда матрица A невырожденна.

Из формулы (11) следует, что ранг матрицы билинейной функции α не зависит от базиса. Он называется рангом билинейной функции α и обозначается через $\text{rk } \alpha$.

Определение 3. Билинейная функция α называется симметрической (соответственно кососимметрической), если $\alpha(x, y) = \alpha(y, x)$ (соответственно $\alpha(x, y) = -\alpha(y, x)$) для любых $x, y \in V$.

Так, билинейные функции в примерах 1 и 2 симметричны.

Билинейная функция в примере 3 также симметрична. В самом деле, если $X = (x_{ij}), Y = (y_{ij})$, то

$$\text{tr } XY = \sum_{i,j} x_{ij}y_{ji} = \sum_{i,j} y_{ji}x_{ij} = \sum_{i,j} y_{ij}x_{ji} = \text{tr } YX.$$

Билинейная функция в примере 4 кососимметрична. Но, конечно, существуют билинейные функции, которые не являются ни симметрическими, ни кососимметрическими.

Билинейная функция является симметрической (соответственно кососимметрической) тогда и только тогда, когда ее матрица A симметрична (соответственно кососимметрична), т. е. $A^T = A$ (соответственно $A^T = -A$).

Определение 4. Пусть α — симметрическая билинейная функция над полем K характеристики $\neq 2$. Функция $q: V \rightarrow K$, определяемая по формуле

$$q(x) = \alpha(x, x),$$

называется *квадратичной функцией* (или *квадратичной формой*), ассоциированной с функцией α .

В координатах квадратичная функция записывается в виде

$$q(x) = \sum_{i,j} a_{ij}x_i x_j, \quad (12)$$

т. е. является однородным многочленом второй степени.

Симметрическая билинейная функция α может быть восстановлена по соответствующей квадратичной функции q по формуле

$$\alpha(x, y) = \frac{1}{2}[q(x+y) - q(x) - q(y)]. \quad (13)$$

Билинейная функция α называется *поляризацией* квадратичной функции q .

Таким образом, имеется взаимно однозначное соответствие между симметрическими билинейными и квадратичными функциями. Имея в виду это соответствие, все понятия, относящиеся к симметрическим билинейным функциям (матрица, ранг, невырожденность и т. д.), переносят на квадратичные функции. В дальнейшем изложении мы будем из соображений удобства иногда говорить о симметрических билинейных, иногда — о квадратичных функциях.

Геометрические ассоциации, связанные со скалярным умножением векторов, могут быть полезны при изучении произвольных билинейных функций. Этим объясняется вводимая ниже терминология.

Пусть α — симметрическая или кососимметрическая билинейная функция над полем K характеристики $\neq 2$. Векторы $x, y \in V$ называются *ортогональными* (относительно α), если $\alpha(x, y) = 0$;

в этом случае пишут $x \perp y$. Ясно, что это отношение симметрично: если $x \perp y$, то и $y \perp x$. Отметим, что в случае кососимметрической функции α каждый вектор ортогонален самому себе.

Определение 5. Ортогональным дополнением к подпространству U (относительно α) называется подпространство

$$U^\perp = \{y \in V : \alpha(x, y) = 0 \ \forall x \in U\}.$$

В частности, $V^\perp = \text{Ker } \alpha$.

Предложение 1. Если функция α невырождена, то

$$\dim U^\perp = \dim V - \dim U \quad \text{и} \quad (U^\perp)^\perp = U.$$

Доказательство. Пусть $\{e_1, \dots, e_k\}$ — базис в U . Тогда

$$U^\perp = \{y \in V : \alpha(e_i, y) = 0, \ i = 1, \dots, k\}. \quad (14)$$

Записывая эти условия в координатах, мы получаем систему однородных линейных уравнений. Эти уравнения линейно независимы, так как для любых $\lambda_1, \dots, \lambda_k$, не равных нулю одновременно, линейная функция

$$\sum_{i=1}^k \lambda_i \alpha(e_i, y) = \alpha\left(\sum_{i=1}^k \lambda_i e_i, y\right)$$

в силу невырожденности функции α не является нулевой. Следовательно,

$$\dim U^\perp = n - k,$$

где $n = \dim V$.

По той же формуле

$$\dim(U^\perp)^\perp = n - (n - k) = k = \dim U.$$

Однако ясно, что $(U^\perp)^\perp \supset U$. Следовательно, $(U^\perp)^\perp = U$. □

Определение 6. Подпространство U называется невырожденным относительно билинейной функции α , если ее ограничение на U невырожденно.

Предложение 2. $V = U \oplus U^\perp$ тогда и только тогда, когда подпространство U невырожденно.

Доказательство. Из (14) ясно, что в любом случае

$$\dim U^\perp \geq \dim V - \dim U.$$

С другой стороны,

$$U \cap U^\perp = \text{Ker } \alpha|_U,$$

так что если $U \cap U^\perp = 0$, то подпространство U невырожденно. Обратно, если подпространство U невырожденно, то $U \cap U^\perp = 0$ и

$$\dim(U + U^\perp) = \dim U + \dim U^\perp \geq \dim V,$$

откуда следует, что $U + U^\perp = V$. □

Пусть α — симметрическая билинейная функция.

Базис $\{e_1, \dots, e_n\}$ пространства V называется *ортогональным* (относительно α), если его векторы попарно ортогональны. В ортогональном базисе матрица функции α диагональна, а сама функция α и соответствующая ей квадратичная функция q записываются в виде

$$\alpha(x, y) = a_1 x_1 y_1 + \dots + a_n x_n y_n, \quad (15)$$

$$q(x) = a_1 x_1^2 + \dots + a_n x_n^2. \quad (16)$$

Теорема 1. *Для любой симметрической билинейной функции существует ортогональный базис.*

Доказательство. Докажем это утверждение индукцией по $n = \dim V$. При $n = 1$ доказывать нечего. Пусть $n > 1$. Если $\alpha \equiv 0$, то доказывать опять-таки нечего. Если $\alpha \neq 0$, то в силу формулы (13) $q \neq 0$, т. е. существует такой вектор e_1 , что

$$\alpha(e_1, e_1) = q(e_1) \neq 0.$$

Согласно предложению 2,

$$V = \langle e_1 \rangle \oplus \langle e_1 \rangle^\perp.$$

По предположению индукции существует ортогональный базис $\{e_2, \dots, e_n\}$ пространства $\langle e_1 \rangle^\perp$. Добавляя к нему вектор e_1 , мы получаем ортогональный базис $\{e_1, e_2, \dots, e_n\}$ пространства V . □

Следующая теорема дает более явный способ построения ортогонального базиса (при ограничениях, указанных в ее формулировке).

Пусть $\{e_1, \dots, e_n\}$ — некоторый базис пространства V и A — матрица функции α в этом базисе. Обозначим через A_k матрицу ограничения функции α на подпространство $V_k = \langle e_1, \dots, e_k \rangle$ в базисе $\{e_1, \dots, e_k\}$ этого подпространства, т. е. левый верхний угол порядка k матрицы A . Число $\delta_k = \det A_k$ будем называть *угловым минором* порядка k матрицы A . Положим также $V_0 = 0$, $\delta_0 = 1$.

Теорема 2. *Если все угловые миноры $\delta_1, \dots, \delta_n$ матрицы A отличны от нуля, то существует единственный ортогональный базис*

$\{f_1, \dots, f_n\}$ пространства V , удовлетворяющий условиям

$$f_k \in e_k + V_{k-1} \quad (k = 1, \dots, n). \quad (17)$$

При этом

$$q(f_k) = \alpha(f_k, f_k) = \frac{\delta_k}{\delta_{k-1}} \quad (k = 1, \dots, n). \quad (18)$$

Доказательство. Докажем теорему индукцией по n . При $n = 1$ имеем

$$f_1 = e_1, \quad q(f_1) = \delta_1 \left(= \frac{\delta_1}{\delta_0} \right).$$

При $n > 1$ применим предположение индукции к базису $\{e_1, \dots, e_{n-1}\}$ пространства V_{n-1} . Пусть $\{f_1, \dots, f_{n-1}\}$ — ортогональный базис пространства V_{n-1} , удовлетворяющий условиям теоремы. Будем искать вектор f_n в виде

$$f_n = e_n + \sum_{i=1}^{n-1} \lambda_i f_i \in e_n + V_{n-1}.$$

Заметим, что

$$q(f_i) = \frac{\delta_i}{\delta_{i-1}} \neq 0 \quad \text{при } i = 1, \dots, n-1,$$

и поэтому условия ортогональности

$$0 = \alpha(f_n, f_i) = \alpha(e_n, f_i) + \lambda_i q(f_i) \quad (i = 1, \dots, n-1)$$

удовлетворяются при подходящем выборе чисел $\lambda_1, \dots, \lambda_{n-1}$, причем эти числа определяются однозначно. Так как $f_n \notin V_{n-1}$, то $\{f_1, \dots, f_n\}$ — базис пространства V .

Остается проверить равенство (18) при $k = n$. Так как матрица перехода от базиса $\{e_1, \dots, e_n\}$ к базису $\{f_1, \dots, f_n\}$ является (верхней) унитреугольной (т. е. треугольной с единицами на диагонали), то ее определитель равен 1 и формула (10) показывает, что определитель матрицы функции α не меняется при переходе к базису $\{f_1, \dots, f_n\}$. Однако в этом базисе матрица функции α диагональна, причем ее диагональные элементы равны

$$q(f_1), \dots, q(f_{n-1}), q(f_n).$$

Следовательно,

$$\delta_n = q(f_1) \dots q(f_{n-1}) q(f_n).$$

Такое же рассуждение, примененное к ограничению функции α на подпространство V_{n-1} (или, если угодно, предположение индукции), показывает, что

$$\delta_{n-1} = q(f_1) \dots q(f_{n-1}).$$

Отсюда следует, что

$$q(f_n) = \frac{\delta_n}{\delta_{n-1}}. \quad \square$$

Процесс построения ортогонального базиса, описанный в доказательстве теоремы, называется *процессом ортогонализации Грама—Шмидта*. Рисунок 6 иллюстрирует его в случае, когда α — скалярное умножение в E^3 .

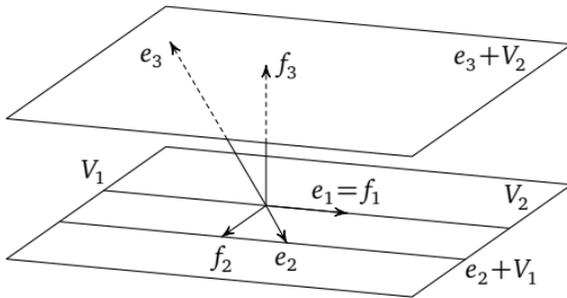


Рис. 6

Пусть $\{e_1, \dots, e_n\}$ — базис пространства V , ортогональный относительно функции α . За счет нормировки векторов e_i числа $a_i = q(e_i)$ можно умножать на квадраты любых ненулевых элементов поля K . Кроме того, переставляя базисные векторы, можно переставлять и эти числа. Однако, как видно из доказательства теоремы 1, в выборе ортогонального базиса имеется гораздо больший произвол. Как можно изменять числа a_i , пользуясь этим произволом? Ответ на этот вопрос существенно зависит от поля K .

Пусть $K = \mathbb{C}$. Тогда путем нормировки базисных векторов числа a_i могут быть сделаны равными 1 или 0, и после подходящей перестановки базисных векторов функция q приводится к так называемому *нормальному виду*:

$$q(x) = x_1^2 + \dots + x_r^2.$$

Число r является инвариантом, так как $r = \text{rk } q$.

Пусть теперь $K = \mathbb{R}$. Тогда путем нормировки базисных векторов числа a_i могут быть сделаны равными ± 1 или 0, и после подходящей перестановки базисных векторов функция q приводится к *нормальному виду*:

$$q(x) = x_1^2 + \dots + x_k^2 - x_{k+1}^2 - \dots - x_{k+l}^2. \quad (19)$$

Сумма $k + l = \text{rk } q$ является инвариантом, но являются ли инвариантами k и l по отдельности? Для ответа на этот вопрос введем одно важное понятие.

Определение 7. Вещественная квадратичная функция q называется *положительно определенной*, если $q(x) > 0$ при $x \neq 0$. Вещественная симметрическая билинейная функция называется *положительно определенной*, если соответствующая ей квадратичная функция положительно определена.

Так, скалярное умножение геометрических векторов является положительно определенной симметрической билинейной функцией.

Очевидно, что нормальный вид положительно определенной квадратичной функции есть $x_1^2 + \dots + x_n^2$, т. е. в некотором базисе ее матрица единичная и, в частности, имеет определитель 1. Из формулы (10) следует, что при переходе к другому базису определитель ее матрицы умножается на квадрат определителя матрицы перехода и, значит, остается положительным. Таким образом, определитель матрицы положительно определенной квадратичной функции в любом базисе положителен.

Теорема 3. Число k в нормальном виде (19) произвольной вещественной квадратичной функции q есть максимальная размерность подпространства, на котором функция q положительно определена.

Доказательство. Очевидно, что функция q положительно определена на k -мерном подпространстве $\langle e_1, \dots, e_k \rangle$. Пусть теперь U — произвольное подпространство, на котором функция q положительно определена, и $W = \langle e_{k+1}, \dots, e_n \rangle$. Так как $q(x) \leq 0$ при $x \in W$, то $U \cap W = 0$. Отсюда следует, что $\dim U \leq k$. \square

Аналогично определяются *отрицательно определенные* квадратичные и симметрические билинейные функции. Так же, как и выше, доказывается, что число l в нормальном виде квадратичной функции q есть максимальная размерность подпространства, на котором функция q отрицательно определена. Впрочем, это и пря-

мо следует из теоремы 3, если ее применить к квадратичной функции $-q$.

Следствие (закон инерции). Числа k и l в нормальном виде (19) вещественной квадратичной функции q не зависят от выбора базиса, в котором эта функция имеет нормальный вид.

Эти числа называются соответственно *положительным* и *отрицательным индексами инерции* квадратичной функции q (а также соответствующей симметрической билинейной функции α). Пара (k, l) называется *сигнатурой* функции q (или функции α).

Пример 5. Квадратичная функция $q(x) = x_1x_2$ путем (невыврожденного) преобразования координат

$$x_1 = x'_1 + x'_2, \quad x_2 = x'_1 - x'_2$$

приводится к виду $q(x) = x'^2_1 - x'^2_2$. Поэтому ее сигнатура равна $(1, 1)$.

Задача 2. Найти сигнатуру симметрической билинейной функции из примера 3 (в случае $K = \mathbb{R}$).

Теорема 2 позволяет (при указанных в ней ограничениях) определить индексы инерции вещественной квадратичной функции по угловым минорам $\delta_1, \dots, \delta_n$ ее матрицы в каком-либо базисе.

Теорема 4 (метод Якоби). Если все угловые миноры δ_k матрицы вещественной квадратичной функции q отличны от нуля, то отрицательный индекс инерции функции q равен числу перемен знака в последовательности

$$1, \delta_1, \delta_2, \dots, \delta_n. \quad (20)$$

(Определение числа перемен знаков в последовательности вещественных чисел см. в § 3.4.)

Доказательство непосредственно следует из теоремы 2. \square

Заметим, что в условиях теоремы функция q невырождена, так что сумма ее индексов инерции равна n .

Следствие (критерий Сильвестра). Вещественная квадратичная функция является положительно определенной тогда и только тогда, когда все угловые миноры ее матрицы положительны.

Доказательство. Если все угловые миноры положительны, то, в частности, они отличны от нуля, и применение метода Якоби доказывает, что функция является положительно определенной. Обратно, если функция положительно определена, то ее ограничение

на любое подпространство V_k (в обозначениях теоремы 2) также положительно определено и, согласно замечанию перед теоремой 3, определитель δ_k его матрицы положителен. \square

Замечание 1. Модифицируя процесс ортогонализации, можно показать (попробуйте это сделать), что метод Якоби годится и в том случае, когда некоторые из угловых миноров равны нулю, лишь бы в последовательности $\delta_1, \delta_2, \dots, \delta_n$ не было двух нулей подряд (в частности, может быть $\delta_n = 0$, но тогда должно быть $\delta_{n-1} \neq 0$). При этом если $\delta_k = 0$ при каком-то $k < n$, то автоматически $\delta_{k-1}\delta_{k+1} < 0$.

Как мы видели, в случаях $K = \mathbb{C}$ или \mathbb{R} никакие изменения диагонального вида матрицы квадратичной функции, кроме тех, которые достигаются уже путем перестановки базисных векторов и их умножения на числа, невозможны, но так обстоит дело не всегда.

Пусть $K = \mathbb{Z}_p$ — поле вычетов по простому модулю $p \neq 2$. Известно (см. теорему 9.1.7), что \mathbb{Z}_p^* — циклическая группа. Следовательно, $(\mathbb{Z}_p^*)^2 = \{a^2 : a \in \mathbb{Z}_p^*\}$ — подгруппа индекса 2. Ее элементы называются *квадратичными вычетами*, а элементы второго смежного класса — *квадратичными невычетами*. Пусть $\varepsilon \in \mathbb{Z}_p^*$ — фиксированный квадратичный невычет.

Теорема 5. *Всякая невырожденная квадратичная функция над полем \mathbb{Z}_p ($p \neq 2$) может быть приведена к одному из двух видов*

$$\begin{aligned} x_1^2 + \dots + x_{n-1}^2 + x_n^2, \\ x_1^2 + \dots + x_{n-1}^2 + \varepsilon x_n^2 \end{aligned}$$

в зависимости от того, является определитель ее матрицы квадратичным вычетом или невычетом.

Лемма 1. *Всякая невырожденная квадратичная функция q в векторном пространстве размерности $n \geq 2$ над полем \mathbb{Z}_p представляет единицу, т. е. уравнение $q(x) = 1$ имеет решение.*

Доказательство. Достаточно рассмотреть случай $n = 2$. Можно считать, что

$$q(x) = a_1 x_1^2 + a_2 x_2^2,$$

где $a_1, a_2 \neq 0$. Уравнение $q(x) = 1$ может быть представлено в виде

$$a_1 x_1^2 = 1 - a_2 x_2^2.$$

Когда x_1 пробегает поле \mathbb{Z}_p , левая часть последнего уравнения принимает $\frac{p+1}{2}$ различных значений. Аналогично, когда x_2 пробегает поле \mathbb{Z}_p , правая

где число диагональных клеток равно m . Очевидно, что при этом $\text{rk } \alpha = 2m$.

Теорема 6. Для любой кососимметрической билинейной функции существует симплектический базис.

Доказательство. Докажем это утверждение индукцией по n . При $n = 1$ доказывать нечего. Пусть $n > 1$. Если $\alpha \equiv 0$, то доказывать опять-таки нечего. Если $\alpha \neq 0$, то существуют такие векторы e_1 и e_2 , что $\alpha(e_1, e_2) \neq 0$. Домножив один из этих векторов на подходящее число, можно добиться того, чтобы

$$\alpha(e_1, e_2) = -\alpha(e_2, e_1) = 1.$$

Матрица ограничения функции α на $\langle e_1, e_2 \rangle$ в базисе $\{e_1, e_2\}$ имеет вид $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ и, в частности, невырожденна. Согласно предложению 2,

$$V = \langle e_1, e_2 \rangle \oplus \langle e_1, e_2 \rangle^\perp.$$

По предположению индукции в пространстве $\langle e_1, e_2 \rangle^\perp$ существует симплектический базис $\{e_3, e_4, \dots, e_n\}$. Добавляя к нему векторы e_1 и e_2 , мы получаем симплектический базис $\{e_1, e_2, e_3, e_4, \dots, e_n\}$ пространства V . \square

Следствие. Ранг кососимметрической билинейной функции всегда является четным числом.

§ 5. Евклидово пространство

Свойства операций над геометрическими векторами, включая скалярное умножение, находят наиболее полное отражение в понятии евклидова векторного пространства.

Определение 1. Евклидовым (векторным) пространством называется вещественное векторное пространство с фиксированной положительно определенной симметрической билинейной функцией.

Обычно эта фиксированная билинейная функция называется скалярным умножением и обозначается $(,)$.

Пример 1. Пространство геометрических векторов с обычным скалярным умножением.

Пример 2. Пространство \mathbb{R}^n со скалярным умножением

$$(x, y) = x_1 y_1 + \dots + x_n y_n,$$

где $x = (x_1, \dots, x_n)$, $y = (y_1, \dots, y_n)$.

Пример 3. Пространство $C_2[0, 1]$ непрерывных функций на отрезке $[0, 1]$ со скалярным умножением

$$(f, g) = \int_0^1 f(x)g(x) dx. \quad (21)$$

В евклидовом пространстве определяются *длина* вектора и *угол* между векторами таким образом, что в случае геометрических векторов они совпадают с обычной длиной и обычным углом. А именно, длина $|x|$ вектора x определяется по формуле

$$|x| = \sqrt{(x, x)}.$$

Для определения угла необходимо сначала доказать

Предложение 1. Для любых векторов x, y евклидова пространства

$$|(x, y)| \leq |x||y|, \quad (22)$$

причем равенство имеет место тогда и только тогда, когда векторы x и y пропорциональны.

Неравенство (22) называется *неравенством Коши—Буняковского*.

Доказательство. Если $y = \lambda x$, то

$$|(x, y)| = |\lambda|(x, x) = |\lambda||x|^2 = |x||y|.$$

Если векторы x и y не пропорциональны, то они составляют базис двумерного подпространства. Матрица скалярного умножения на этом подпространстве в базисе $\{x, y\}$ имеет вид

$$\begin{pmatrix} (x, x) & (x, y) \\ (x, y) & (y, y) \end{pmatrix}.$$

Ввиду положительной определенности скалярного умножения ее определитель должен быть положителен; но это и означает, что

$$|(x, y)| < |x||y|. \quad \square$$

Пример 4. Для евклидова пространства $C_2[0, 1]$ из примера 3 неравенство Коши—Буняковского означает, что

$$\left(\int_0^1 f(x)g(x) dx \right)^2 < \left(\int_0^1 (f(x))^2 dx \right) \left(\int_0^1 (g(x))^2 dx \right) \quad (23)$$

для любых непропорциональных непрерывных функций f и g на отрезке $[0, 1]$. В частности, при $g = 1$ мы получаем, что

$$\left(\int_0^1 f(x) dx \right)^2 < \int_0^1 (f(x))^2 dx \quad (24)$$

для любой непостоянной непрерывной функции f .

Угол \widehat{xy} между ненулевыми векторами x и y евклидова пространства определяется по формуле

$$\cos \widehat{xy} = \frac{(x, y)}{|x||y|}.$$

В частности, угол \widehat{xy} равен 0 или π тогда и только тогда, когда векторы x и y пропорциональны; $\widehat{xy} = \pi/2$ тогда и только тогда, когда векторы x и y ортогональны.

Неравенство Коши—Буняковского является частным случаем более общего неравенства, относящегося к произвольной конечной системе векторов $\{a_1, \dots, a_k\}$ евклидова пространства.

Определение 2. Матрица

$$G(a_1, \dots, a_k) = \begin{pmatrix} (a_1, a_1) & (a_1, a_2) & \dots & (a_1, a_k) \\ (a_2, a_1) & (a_2, a_2) & \dots & (a_2, a_k) \\ \dots & \dots & \dots & \dots \\ (a_k, a_1) & (a_k, a_2) & \dots & (a_k, a_k) \end{pmatrix}$$

называется *матрицей Грама* системы векторов $\{a_1, \dots, a_k\}$.

Теорема 1. Для любых векторов a_1, \dots, a_k евклидова пространства справедливо неравенство

$$\det G(a_1, \dots, a_k) \geq 0,$$

причем равенство имеет место тогда и только тогда, когда векторы a_1, \dots, a_k линейно зависимы.

Доказательство. Если $\sum_i \lambda_i a_i = 0$, то $\sum_i \lambda_i (a_i, a_j) = 0$ при всех j , а это означает, что линейная комбинация строк матрицы $G(a_1, \dots, a_k)$ с коэффициентами $\lambda_1, \dots, \lambda_k$ равна нулю. Поэтому если векторы a_1, \dots, a_k линейно зависимы, то $\det G(a_1, \dots, a_k) = 0$. Если же они линейно независимы, то так же, как в случае $k = 2$, доказывается, что $\det G(a_1, \dots, a_k) > 0$. \square

Задача 1. Получить соотношение между двугранными углами тетраэдра, рассмотрев матрицу Грама системы единичных векторов,

ортогональных его граням. С помощью этого соотношения найти двугранный угол правильного тетраэдра.

Определение 3. Базис евклидова пространства, в котором скалярное умножение имеет нормальный вид (см. § 4), называется *ортонормированным*.

Ортонормированность базиса $\{e_1, \dots, e_n\}$ означает, что матрица скалярного умножения в этом базисе (т.е. матрица $G(e_1, \dots, e_n)$) является единичной матрицей или, иными словами, что базисные векторы попарно ортогональны и имеют длину 1.

Согласно общей теории, в любом конечномерном евклидовом пространстве существует ортонормированный базис. Такой базис, конечно, не единствен. Дадим описание всех ортонормированных базисов, исходя из какого-либо одного ортонормированного базиса $\{e_1, \dots, e_n\}$.

Пусть

$$(e'_1, \dots, e'_n) = (e_1, \dots, e_n)C.$$

Тогда матрица скалярного умножения в базисе $\{e'_1, \dots, e'_n\}$ имеет вид

$$C^T E C = C^T C.$$

(см. формулу (10)). Следовательно, базис $\{e'_1, \dots, e'_n\}$ является ортонормированным тогда и только тогда, когда

$$C^T C = E, \quad (25)$$

т.е. когда матрица C^T обратна матрице C .

Определение 4. Матрица C , для которой $C^{-1} = C^T$, называется *ортогональной*.

Равенство (25) означает следующие соотношения между элементами матрицы C :

$$\sum_k c_{ki} c_{kj} = \delta_{ij} \quad \text{при всех } i, j.$$

Условие ортогональности также может быть записано в виде

$$C C^T = E, \quad (26)$$

что означает следующие соотношения между матричными элементами:

$$\sum_k c_{ik} c_{jk} = \delta_{ij} \quad \text{при всех } i, j.$$

Заметим, что из равенства (25) следует соотношение $\det C = \pm 1$ (но не наоборот!).

Ограничение скалярного умножения на любое подпространство U евклидова пространства V также является положительно определенной и потому невырожденной симметрической билинейной функцией, и предложение 4.2 показывает, что

$$V = U \oplus U^\perp.$$

Это означает, что для каждого вектора $x \in V$ имеется единственное представление в виде

$$x = y + z, \quad y \in U, \quad z \in U^\perp. \quad (27)$$

Вектор y называется (*ортогональной*) *проекцией* вектора x на подпространство U и обозначается через $\text{pr}_U x$; вектор z называется *ортогональной составляющей* вектора x относительно подпространства U и обозначается через $\text{ort}_U x$.

Если $\{e_1, \dots, e_k\}$ — ортонормированный базис подпространства U , то проекция $\text{pr}_U x$ может быть найдена по формуле

$$\text{pr}_U x = \sum_{i=1}^k (x, e_i) e_i. \quad (28)$$

Более общо, если $\{e_1, \dots, e_k\}$ — ортогональный (но не обязательно ортонормированный) базис подпространства U , то

$$\text{pr}_U x = \sum_{i=1}^k \frac{(x, e_i)}{(e_i, e_i)} e_i. \quad (29)$$

Для построения ортогонального базиса евклидова пространства V может быть применен *процесс ортогонализации*, описанный в доказательстве теоремы 4.2. В предыдущих обозначениях, если $\{e_1, \dots, e_n\}$ — какой-либо базис пространства V , то базис $\{f_1, \dots, f_n\}$, получаемый в результате ортогонализации, задается формулами

$$f_k = \text{ort}_{V_{k-1}} e_k \quad (k = 1, \dots, n). \quad (30)$$

Пользуясь тем, что $\{f_1, \dots, f_{k-1}\}$ — ортогональный базис пространства V_{k-1} , проекцию $\text{pr}_{V_{k-1}} e_k$ и, тем самым, вектор f_k можно найти по формуле (29).

Пример 5. Пусть V — пространство многочленов степени ≤ 3 со скалярным умножением (21). Применим процесс ортогонализации к базису

$$e_1 = 1, \quad e_2 = x, \quad e_3 = x^2, \quad e_4 = x^3.$$

Заметим, что $(e_i, e_j) = \frac{1}{i+j-1}$. Имеем

$$f_1 = e_1 = 1, \quad (f_1, f_1) = 1,$$

$$f_2 = e_2 - \frac{(e_2, f_1)}{(f_1, f_1)} f_1 = x - \frac{1}{2}, \quad (f_2, f_2) = (f_2, e_2) = \frac{1}{12},$$

$$f_3 = e_3 - \frac{(e_3, f_2)}{(f_2, f_2)} f_2 - \frac{(e_3, f_1)}{(f_1, f_1)} f_1 = x^2 - x + \frac{1}{6}, \quad (f_3, f_3) = (f_3, e_3) = \frac{1}{180},$$

$$f_4 = e_4 - \frac{(e_4, f_3)}{(f_3, f_3)} f_3 - \frac{(e_4, f_2)}{(f_2, f_2)} f_2 - \frac{(e_4, f_1)}{(f_1, f_1)} f_1 = x^3 - \frac{3}{2}x^2 + \frac{3}{5}x - \frac{1}{20},$$

$$(f_4, f_4) = (f_4, e_4) = \frac{1}{2800}.$$

Задача 2. Применяя процесс ортогонализации к столбцам матрицы, доказать, что каждая матрица $A \in \text{GL}_n(\mathbb{R})$ может быть единственным образом представлена в виде $A = OB$, где O — ортогональная матрица, а B — треугольная матрица с положительными элементами на диагонали.

Определим расстояние ρ между векторами евклидова пространства по формуле

$$\rho(x, y) = |x - y|.$$

Это расстояние удовлетворяет аксиомам метрического пространства, в частности аксиоме треугольника

$$\rho(x, z) \leq \rho(x, y) + \rho(y, z). \quad (31)$$

Неравенство (31) следует из неравенства

$$|x + y| \leq |x| + |y|, \quad (32)$$

которое, в свою очередь, легко выводится из неравенства Коши—Буняковского (проделайте это!)

Расстояние между подмножествами X и Y метрического пространства определяется по формуле

$$\rho(X, Y) = \inf_{x \in X, y \in Y} \rho(x, y).$$

Теорема 2. Расстояние от вектора x евклидова пространства V до подпространства $U \subset V$ равно $|\text{ort}_U x|$, причем единственным ближайшим к x вектором подпространства U является $\text{pr}_U x$.

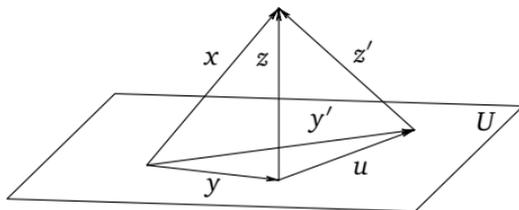


Рис. 7

Доказательство. См. рис. 7, где $y = \text{pr}_U x$, $z = \text{ort}_U x$. Для любого $y' \in U$, $y' \neq y$, имеем:

$$\rho(x, y') = |z'| = \sqrt{|z|^2 + |u|^2} > |z| = \rho(x, y). \quad \square$$

Пример 6. В силу вычислений, сделанных в предыдущем примере, квадратным трехчленом, ближайшим к x^3 в смысле метрики пространства $C_2[0, 1]$, является $\frac{3}{2}x^2 - \frac{3}{5}x + \frac{1}{20}$, причем расстояние от x^3 до этого трехчлена равно $\frac{1}{20\sqrt{7}}$.

Следующая теорема дает явную формулу для расстояния от вектора x до подпространства U , заданного произвольным базисом $\{e_1, \dots, e_k\}$.

Теорема 3.
$$(\rho(x, U))^2 = \frac{\det G(e_1, \dots, e_k, x)}{\det G(e_1, \dots, e_k)}$$

Доказательство. Если $x \in U$, то $\rho(x, U) = 0$ и $\det G(e_1, \dots, e_k, x) = 0$, так что доказываемая формула верна.

Пусть $x \notin U$ и $z = \text{ort}_U x$. Применяя теорему 4.2 к базису $\{e_1, \dots, e_k, x\}$ пространства $U \oplus \langle x \rangle$, получаем

$$|z|^2 = (z, z) = \frac{\delta_{k+1}}{\delta_k} = \frac{\det G(e_1, \dots, e_k, x)}{\det G(e_1, \dots, e_k)},$$

что и требовалось доказать. □

Полученная формула может быть применена к вычислению объема параллелепипеда в евклидовом пространстве.

Параллелепипедом, натянутым на векторы a_1, \dots, a_n евклидова пространства, называется множество

$$P(a_1, \dots, a_n) = \left\{ \sum_i x_i a_i : 0 \leq x_i \leq 1 \right\}.$$

Основанием этого n -мерного параллелепипеда называют $(n-1)$ -мерный параллелепипед $P(a_1, \dots, a_{n-1})$, а его высотой — длину вектора $\text{ort}_{\langle a_1, \dots, a_{n-1} \rangle} a_n$. При $n = 2, 3$ это согласуется с терминологией

элементарной геометрии. Руководствуясь известными формулами для площади параллелограмма и объема трехмерного параллелепипеда, примем следующее индуктивное

Определение 5. Объемом n -мерного ($n > 1$) параллелепипеда называется произведение объема его основания на высоту. Объемом одномерного параллелепипеда $P(a)$ называется длина вектора a .

Объем параллелепипеда P обозначается через $\text{vol } P$.

Теорема 4. $(\text{vol } P(a_1, \dots, a_n))^2 = \det G(a_1, \dots, a_n)$.

Доказательство. Докажем эту формулу индукцией по n . При $n = 1$ она верна по определению. При $n > 1$ имеем, согласно определению,

$$\text{vol } P(a_1, \dots, a_n) = \text{vol } P(a_1, \dots, a_{n-1}) \cdot h,$$

где h — длина вектора $\text{ort}_{\langle a_1, \dots, a_{n-1} \rangle} a_n$, т. е. расстояние от вектора a_n до подпространства $\langle a_1, \dots, a_{n-1} \rangle$. Используя предположение индукции и теорему 3, получаем

$$\begin{aligned} (\text{vol } P(a_1, \dots, a_n))^2 &= \det G(a_1, \dots, a_{n-1}) \cdot \frac{\det G(a_1, \dots, a_{n-1}, a_n)}{\det G(a_1, \dots, a_{n-1})} = \\ &= \det G(a_1, \dots, a_n). \quad \square \end{aligned}$$

В частности, мы видим, что, хотя основание параллелепипеда и зависит от того, какой из заданных векторов мы считаем «последним», объем параллелепипеда в смысле данного выше определения зависит лишь от самого параллелепипеда. Наряду с формулами для площади параллелограмма и объема трехмерного параллелепипеда все это служит неплохим обоснованием приведенного определения, однако по-настоящему убедительное обоснование может быть получено лишь в рамках теории меры, объясняющей, что вообще следует называть объемом множества.

Пусть векторы a_1, \dots, a_n выражаются через векторы какого-нибудь ортонормированного базиса $\{e_1, \dots, e_n\}$ при помощи матрицы A :

$$(a_1, \dots, a_n) = (e_1, \dots, e_n)A.$$

Теорема 5. $\text{vol } P(a_1, \dots, a_n) = |\det A|$.

Доказательство следует из того, что

$$G(a_1, \dots, a_n) = A^T E A = A^T A$$

и, значит,

$$\det G(a_1, \dots, a_n) = (\det A)^2. \quad \square$$

Доказанное равенство можно понимать как «геометрический смысл» числа $|\det A|$. Что касается знака числа $\det A$, то он может быть истолкован как ориентация системы векторов $\{a_1, \dots, a_n\}$ (по отношению к базису $\{e_1, \dots, e_n\}$). Напомним, что при введении определителей порядка n в § 2.4 мы как раз руководствовались тем, что определители порядков 2 и 3 задают ориентированную площадь параллелограмма и ориентированный объем параллелепипеда соответственно.

В § 2.2 было показано, что строение векторного пространства (над данным полем) зависит лишь от его размерности. Верно ли то же самое для евклидовых векторных пространств? Для того чтобы ответить на этот вопрос, надо прежде всего понять, какие евклидовы пространства следует считать «одинаково устроенными» или, точнее, изоморфными. Естественно принять следующее

Определение 6. Евклидовы векторные пространства V и U называются *изоморфными*, если существует биективное отображение $f: V \rightarrow U$, являющееся изоморфизмом векторных пространств и удовлетворяющее условию

$$(f(a), f(b)) = (a, b) \quad \forall a, b \in V.$$

Само отображение f называется при этом *изоморфизмом* пространств V и U .

Ясно, что изоморфными могут быть только евклидовы пространства одинаковой размерности. Оказывается, верно и обратное.

Теорема 6. Любые два евклидовых векторных пространства одинаковой (конечной) размерности изоморфны.

Доказательство. Пусть V и U — какие-то n -мерные евклидовы пространства. Выберем в них ортонормированные базисы $\{v_1, \dots, v_n\}$ и $\{u_1, \dots, u_n\}$ соответственно, и пусть $f: V \rightarrow U$ — изоморфизм векторных пространств, переводящий v_i в u_i ($i = 1, \dots, n$). Тогда

$$(f(v_i), f(v_j)) = (u_i, u_j) = \delta_{ij} = (v_i, v_j),$$

откуда по линейности вытекает, что

$$(f(a), f(b)) = (a, b)$$

для любых $a, b \in V$. □

В частности, любое двумерное (соответственно трехмерное) евклидово пространство устроено совершенно так же, как E^2 (соответственно E^3). Пользуясь этим, в тех случаях, когда рассматрива-

емые векторы лежат в двумерном или трехмерном подпространстве, для доказательства каких-либо утверждений о них можно привлекать теоремы элементарной геометрии. Например, таким способом можно доказать неравенство Коши—Буняковского (22), неравенство треугольника (31) и теорему 2.

§ 6. Эрмитовы пространства

При желании ввести метрику в комплексном векторном пространстве подобно тому, как это делается в вещественном пространстве, мы наталкиваемся на ту трудность, что в комплексном пространстве не существует положительно определенных квадратичных функций. Эту трудность можно обойти, введя в рассмотрение так называемые полуторалинейные функции (не очень удачный термин, но лучшего не придумано).

Определение 1. Пусть V — комплексное векторное пространство. Функция $\alpha: V \times V \rightarrow \mathbb{C}$ называется *полуторалинейной*, если она линейна по второму аргументу и антилинейна по первому. Последнее означает, что

$$\begin{aligned}\alpha(x_1 + x_2, y) &= \alpha(x_1, y) + \alpha(x_2, y), \\ \alpha(\lambda x, y) &= \bar{\lambda} \alpha(x, y).\end{aligned}$$

Замечание 1. Иногда требуют, чтобы полуторалинейная функция была, наоборот, линейна по первому аргументу и антилинейна по второму.

Теория полуторалинейных функций аналогична теории билинейных функций. Поэтому мы изложим ее кратко, останавливаясь более подробно лишь в тех местах, где имеется существенное различие.

Пусть $\{e_1, \dots, e_n\}$ — базис пространства V . Полуторалинейная функция α определяется числами $a_{ij} = \alpha(e_i, e_j)$. А именно,

$$\alpha(x, y) = \sum_{i,j} a_{ij} \bar{x}_i y_j. \quad (33)$$

Матрица $A = (a_{ij})$ называется *матрицей функции* α в базисе $\{e_1, \dots, e_n\}$. При переходе к другому базису

$$(e'_1, \dots, e'_n) = (e_1, \dots, e_n)C$$

она преобразуется по правилу

$$A' = C^* A C, \quad (34)$$

где $C^* = \bar{C}^T$. (Черта обозначает комплексное сопряжение, примененное ко всем элементам матрицы C .) Функция α называется *невырожденной*, если

$$\text{Ker } \alpha \doteq \{y \in V : \alpha(x, y) = 0 \ \forall x \in V\} = 0.$$

Это равносильно невырожденности матрицы A .

Полуторалинейная функция α называется эрмитовой (соответственно косоэрмитовой), если $\alpha(y, x) = \alpha(x, y)$ (соответственно $\alpha(y, x) = -\alpha(x, y)$). При умножении эрмитовой функции на i получается косоэрмитова функция, и наоборот.

Функция α является эрмитовой (соответственно косоэрмитовой) тогда и только тогда, когда ее матрица A удовлетворяет условию $A^* = A$ (соответственно $A^* = -A$). Такие матрицы называются *эрмитовыми* (соответственно *косоэрмитовыми*). Заметим, что диагональные элементы эрмитовой матрицы вещественны, а косоэрмитовой — чисто мнимы.

Каждой эрмитовой полуторалинейной функции α соответствует *эрмитова квадратичная функция*

$$q(x) = \alpha(x, x).$$

Легко видеть, что все ее значения вещественны. Соотношения

$$\begin{aligned} q(x+y) &= q(x) + q(y) + \alpha(x, y) + \alpha(y, x), \\ q(x+iy) &= q(x) + q(y) + i\alpha(x, y) - i\alpha(y, x) \end{aligned}$$

позволяют восстановить α по q . В частности, если $q \equiv 0$, то и $\alpha \equiv 0$.

Пусть α — эрмитова полуторалинейная функция. Так же, как в случае симметрических билинейных функций, определяется *ортogonalность* векторов и *ортogonalное дополнение* к подпространству относительно α . Имеет место аналог предложения 4.2. С его помощью доказывается, что всякая эрмитова полуторалинейная функция и одновременно соответствующая ей квадратичная функция приводятся к *нормальному виду*

$$\begin{aligned} \alpha(x, y) &= \bar{x}_1 y_1 + \dots + \bar{x}_k y_k - \bar{x}_{k+1} y_{k+1} - \dots - \bar{x}_{k+l} y_{k+l}, \\ q(x) &= |x_1|^2 + \dots + |x_k|^2 - |x_{k+1}|^2 - \dots - |x_{k+l}|^2. \end{aligned} \quad (35)$$

Эрмитова квадратичная функция q (и соответствующая ей эрмитова полуторалинейная функция) называется *положительно определенной*, если $q(x) > 0$ при $x \neq 0$. Это имеет место тогда и только тогда, когда в нормальном виде (35) $k = n$, $l = 0$.

В общем случае имеет место *закон инерции*, утверждающий, что числа k и l определены однозначно. Они называются *положительным* и *отрицательным индексами инерции* функции q .

Так как для всякой комплексной матрицы

$$\det A^* = \overline{\det A},$$

то определитель эрмитовой матрицы всегда веществен. Если все угловые миноры матрицы эрмитовой полуторалинейной функции отличны от нуля, то можно так же, как в случае билинейной функции, провести ортогонализацию базисных векторов и получить отсюда метод Якоби для определения индексов инерции по знакам угловых миноров. В частности, имеет место аналог критерия Сильвестра: эрмитова квадратичная функция положительно определена тогда и только тогда, когда все угловые миноры ее матрицы положительны.

Комплексным аналогом евклидовых пространств являются эрмитовы пространства. *Эрмитовым пространством* называется комплексное векторное пространство, в котором фиксирована некоторая положительно определенная эрмитова полуторалинейная функция, называемая *скалярным умножением* и обозначаемая (\cdot, \cdot) .

Пример 1. Пространство \mathbb{C}^n со скалярным умножением

$$(x, y) = \bar{x}_1 y_1 + \dots + \bar{x}_n y_n.$$

Пример 2. Пространство непрерывных комплекснозначных функций на отрезке $[0, 1]$ со скалярным умножением

$$(f, g) = \int_0^1 \overline{f(x)} g(x) dx.$$

В эрмитовом пространстве определяется *длина* вектора по формуле

$$|x| = \sqrt{(x, x)}.$$

В нем выполняются неравенство Коши—Буняковского

$$|(x, y)| \leq |x| |y|$$

и неравенство треугольника

$$|x + y| \leq |x| + |y|$$

(докажите их).

Базис $\{e_1, \dots, e_n\}$ эрмитова пространства называется *ортонормированным*, если в этом базисе скалярное умножение имеет нормальный вид, т. е. если

$$(e_i, e_j) = \delta_{ij}.$$

Матрица перехода от одного ортонормированного базиса к другому удовлетворяет условию $C^* = C^{-1}$. Такие комплексные матрицы называются *унитарными*.

Задача 1. Записать условие унитарности матрицы через матричные элементы двумя способами.

Заметим, что определитель унитарной матрицы C по модулю равен 1. В самом деле, беря определитель от обеих частей равенства $C^*C = E$, получаем

$$\overline{\det C} \cdot \det C = 1,$$

а это и означает, что $|\det C| = 1$.

Так же, как и в случае евклидова пространства, для любого подпространства U эрмитова пространства V получаем разложение

$$V = U \oplus U^\perp.$$

Если $\{e_1, \dots, e_k\}$ — ортогональный базис подпространства U , то ортогональная проекция вектора $x \in V$ на U может быть найдена по формуле

$$\text{pr}_U x = \sum_{i=1}^k \frac{(e_i, x)}{(e_i, e_i)} e_i.$$

(Обратите внимание на отличие этой формулы от формулы (29).)

В эрмитовом пространстве также справедливы аналоги теорем 5.2 и 5.3.

С математической точки зрения эрмитовы пространства полезны по той же причине, что и комплексные числа. Это станет ясным в следующей главе. С физической точки зрения эрмитовы пространства необходимы для построения адекватной квантово-механической картины мира.

Линейные операторы

Теория линейных операторов — это ядро линейной алгебры и главный источник ее многочисленных приложений. Как и билинейная функция, линейный оператор в конечномерном векторном пространстве задается квадратной матрицей, так что в каком-то смысле это объекты одинаковой сложности (но, конечно, симметрическая или кососимметрическая билинейная функция проще, чем произвольный линейный оператор).

Мы сохраняем соглашения, принятые во введении к предыдущей главе.

§1. Матрица линейного оператора

Определение 1. *Линейным оператором (или линейным преобразованием) в векторном пространстве V называется линейное отображение пространства V в себя.*

Более подробно, линейный оператор — это отображение $\mathcal{A} : V \rightarrow V$, удовлетворяющее условиям:

- 1) $\mathcal{A}(x + y) = \mathcal{A}x + \mathcal{A}y$ для любых $x, y \in V$;
- 2) $\mathcal{A}(\lambda x) = \lambda \mathcal{A}x$ для любых $x \in V, \lambda \in K$.

(Мы обычно будем обозначать линейные операторы рукописными буквами.)

Если в пространстве V выбран базис $\{e_1, \dots, e_n\}$, то линейный оператор может быть задан матрицей.

Определение 2. *Матрицей линейного оператора \mathcal{A} в базисе $\{e_1, \dots, e_n\}$ называется матрица $A = (a_{ij})$, определяемая из равенств*

$$\mathcal{A}e_j = \sum_i a_{ij}e_i. \quad (1)$$

Иначе говоря, в j -м столбце матрицы A стоят координаты вектора $\mathcal{A}e_j$ в базисе $\{e_1, \dots, e_n\}$. (Обратите внимание, что, в отличие от определения матрицы линейного отображения, в этом определении фигурирует только один базис!)

Равенства (1) можно переписать в следующей матричной форме:

$$(\mathcal{A}e_1, \dots, \mathcal{A}e_n) = (e_1, \dots, e_n)A \quad (2)$$

(Ср. определение матрицы перехода в § 2.2, формула (8).)

Очевидно, что для любых векторов $f_1, \dots, f_n \in V$ существует единственный линейный оператор \mathcal{A} , переводящий базисные векторы e_1, \dots, e_n в f_1, \dots, f_n соответственно. Это оператор, переводящий каждый вектор $x = \sum_i x_i e_i$ в вектор $\sum_i x_i f_i$. Следовательно, линейный оператор однозначно определяется своей матрицей, и любая квадратная матрица порядка n является матрицей некоторого линейного оператора (в данном базисе).

Найдем явное выражение координат образа $y = \mathcal{A}x$ вектора x . При $x = \sum_j x_j e_j$ имеем

$$y = \sum_j x_j \mathcal{A}e_j = \sum_{i,j} a_{ij} x_j e_i = \sum_i y_i e_i,$$

где

$$y_i = \sum_j a_{ij} x_j. \quad (3)$$

Если обозначить через X и Y столбцы координат векторов x и y соответственно, то равенства (3) можно переписать в следующей матричной форме:

$$Y = AX. \quad (4)$$

(Ср. формулу (9) преобразования координат в § 2.2.)

Замечание. Как видно из этих формул, линейные преобразования пространства K^n , определенные в примере 4.1.3, — это то же, что линейные преобразования в смысле определения 1. При этом матрица, определяющая линейное преобразование пространства K^n в смысле примера 4.1.3, — это его матрица в базисе из единичных строк в смысле определения 2.

Выясним, как преобразуется матрица линейного оператора при переходе к другому базису

$$(e'_1, \dots, e'_n) = (e_1, \dots, e_n)C.$$

В силу линейности оператора \mathcal{A} имеем

$$(\mathcal{A}e'_1, \dots, \mathcal{A}e'_n) = (\mathcal{A}e_1, \dots, \mathcal{A}e_n)C = (e_1, \dots, e_n)AC = (e'_1, \dots, e'_n)C^{-1}AC.$$

Таким образом, если обозначить через A' матрицу оператора \mathcal{A} в базисе $\{e'_1, \dots, e'_n\}$, то

$$A' = C^{-1}AC. \quad (5)$$

Перейдя к другому базису, матрицу линейного оператора часто можно привести к более простому виду. В частности, такая возможность открывается, если известно какое-либо инвариантное подпространство.

Определение 3. Подпространство $U \subset V$ называется *инвариантным* относительно оператора \mathcal{A} , если

$$\mathcal{A}U \subset U$$

(т. е. $\mathcal{A}u \in U$ для любого $u \in U$).

Ограничение $\mathcal{A}|_U$ линейного оператора \mathcal{A} на инвариантное подпространство U является линейным оператором в U .

Если базис $\{e_1, \dots, e_n\}$ пространства V выбран таким образом, что $U = \langle e_1, \dots, e_k \rangle$ (а это всегда можно сделать), то матрица оператора \mathcal{A} в этом базисе имеет вид

$$A = \begin{pmatrix} B & D \\ 0 & C \end{pmatrix}, \quad (6)$$

где B — матрица оператора $\mathcal{A}|_U$ в базисе $\{e_1, \dots, e_k\}$, C — квадратная матрица порядка $n - k$ и D — какая-то матрица размера $k \times (n - k)$. Обратное, если матрица оператора \mathcal{A} в базисе $\{e_1, \dots, e_n\}$ имеет вид (6), где B — квадратная матрица порядка k , то $U = \langle e_1, \dots, e_k \rangle$ — инвариантное подпространство.

Еще лучше обстоит дело, когда пространство V удается разложить в прямую сумму двух инвариантных подпространств U и W :

$$V = U \oplus W.$$

Если $\{e_1, \dots, e_k\}$ — базис подпространства U , а $\{e_{k+1}, \dots, e_n\}$ — базис подпространства W , то $\{e_1, \dots, e_n\}$ — базис пространства V и в этом базисе матрица оператора \mathcal{A} имеет вид

$$A = \begin{pmatrix} B & 0 \\ 0 & C \end{pmatrix}, \quad (7)$$

где B — матрица оператора $\mathcal{A}|_U$ в базисе $\{e_1, \dots, e_k\}$, а C — матрица оператора $\mathcal{A}|_W$ в базисе $\{e_{k+1}, \dots, e_n\}$.

Более общо, если пространство V разложено в прямую сумму k инвариантных подпространств V_1, V_2, \dots, V_k , то в базисе пространств

ва V , составленном из базисов этих подпространств, матрица оператора \mathcal{A} имеет вид

$$A = \begin{pmatrix} A_1 & & & 0 \\ & A_2 & & \\ & & \ddots & \\ 0 & & & A_k \end{pmatrix}, \quad (8)$$

где A_i — матрица оператора $\mathcal{A}|_{V_i}$.

Пример 1. Поворот на угол α является линейным оператором в E^2 (см. пример 5.2.1). В примере 5.2.4 мы доказали, что его матрица в ортонормированном базисе $\{e_1, e_2\}$ есть матрица

$$\Pi(\alpha) = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix}. \quad (9)$$

В частности, поворот на $\frac{\pi}{2}$ имеет в таком базисе матрицу

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Найдем его матрицу A' в базисе

$$e'_1 = 2e_2, \quad e'_2 = e_1 - e_2. \quad (10)$$

Как видно из рис. 1,

$$\mathcal{A}e'_1 = -e'_1 - 2e'_2, \quad \mathcal{A}e'_2 = e'_1 + e'_2.$$

Это означает, что

$$A' = \begin{pmatrix} -1 & 1 \\ -2 & 1 \end{pmatrix}.$$

Матрица A' может быть, конечно, найдена и по формуле (5). Из формулы (10) получаем

$$C = \begin{pmatrix} 0 & 1 \\ 2 & -1 \end{pmatrix}, \quad C^{-1} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 1 & 0 \end{pmatrix}.$$

Следовательно,

$$A' = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 2 & -1 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 1 & 0 \end{pmatrix} \begin{pmatrix} -2 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} -1 & 1 \\ -2 & 1 \end{pmatrix}.$$

Пример 2. Аналогично, поворот вокруг какой-либо оси на угол α является линейным оператором в E^3 . В ортонормированном базисе $\{e_1, e_2, e_3\}$, при условии, что вектор e_3 направлен по оси поворота,

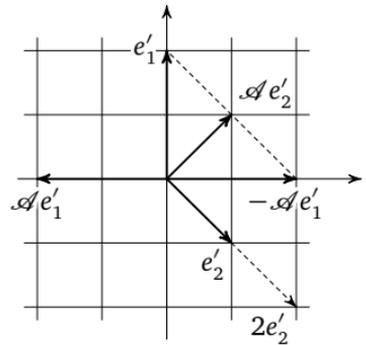


Рис. 1

матрица этого оператора имеет вид

$$A = \begin{pmatrix} \cos \alpha & -\sin \alpha & 0 \\ \sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \Pi(\alpha) & 0 \\ 0 & 1 \end{pmatrix}.$$

Этот вид согласуется с разложением пространства E^3 в прямую сумму двух инвариантных подпространств:

$$E^3 = \langle e_1, e_2 \rangle \oplus \langle e_3 \rangle. \quad (11)$$

Пример 3. В примере 5.2.2 мы рассматривали ортогональное проектирование на плоскость как линейное отображение пространства E^3 в пространство векторов этой плоскости. Однако его можно рассматривать и как линейный оператор в пространстве E^3 . В ортонормированном базисе, первые два вектора которого лежат в плоскости проектирования, его матрица имеет вид

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

Разложение (11) и в этом случае является разложением в прямую сумму инвариантных подпространств.

Пример 4. Дифференцирование — линейный оператор в пространстве многочленов. Это пространство бесконечномерно, но оно является объединением конечномерных инвариантных подпространств, состоящих из многочленов не выше заданной степени. В базисе $\{1, x, x^2, \dots, x^n\}$ пространства многочленов степени не выше n оператор дифференцирования имеет матрицу

$$A = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 2 & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 & n \\ 0 & 0 & 0 & \dots & 0 & 0 \end{pmatrix}.$$

В базисе $\{1, \frac{x}{1!}, \frac{x^2}{2!}, \dots, \frac{x^n}{n!}\}$ этот же оператор имеет более простую матрицу

$$\begin{pmatrix} 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & 0 & \dots & 0 & 0 \end{pmatrix}. \quad (12)$$

Пример 5. Пусть φ — какое-либо биективное преобразование множества X . Тогда отображение φ_* , определяемое формулой

$$(\varphi_*f)(x) = f(\varphi^{-1}(x)), \quad (13)$$

является линейным оператором в пространстве $F(X, K)$ функций на X со значениями в K . (Можно было бы действовать на аргумент функции самим преобразованием φ , а не его обратным, но последнее удобнее по причине, которая будет объяснена в гл. 10.) Например, пусть $X = \mathbb{R}$, $K = \mathbb{R}$, $\varphi(x) = x + a$ ($a \in \mathbb{R}$). Тогда

$$(\varphi_*f)(x) = f(x - a).$$

(График функции φ_*f получается из графика функции f сдвигом вправо на a .) Так как

$$\begin{aligned} \cos(x - a) &= \cos a \cdot \cos x + \sin a \cdot \sin x, \\ \sin(x - a) &= -\sin a \cdot \cos x + \cos a \cdot \sin x, \end{aligned}$$

то подпространство $\langle \cos x, \sin x \rangle$ инвариантно относительно φ_* , причем матрица ограничения преобразования φ_* на это подпространство в базисе $\{\cos x, \sin x\}$ имеет вид

$$\begin{pmatrix} \cos a & -\sin a \\ \sin a & \cos a \end{pmatrix} = \Pi(a).$$

Пример 6. В любой алгебре преобразование

$$L_a : x \mapsto ax \quad (a \in A),$$

называемое *левым умножением* на элемент a , является линейным оператором. Рассмотрим, например, поле \mathbb{C} как алгебру над \mathbb{R} . Равенства

$$(a + bi) \cdot 1 = a + bi, \quad (a + bi) \cdot i = -b + ai$$

показывают, что матрица оператора L_{a+bi} в базисе $\{1, i\}$ есть

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}.$$

Задача 1. Найти матрицу левого умножения на $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ в алгебре $L_2(K)$ в базисе, составленном из матричных единиц. Доказать инвариантность подпространств $\langle E_{11}, E_{21} \rangle$ и $\langle E_{12}, E_{22} \rangle$.

Линейные операторы в одном векторном пространстве можно складывать, умножать друг на друга и умножать на числа. Эти операции определяются так же, как для общих линейных отображений (см. § 5.2). Им соответствуют такие же операции над матрицами, т. е., например, матрица произведения двух линейных операторов в каком-либо базисе равна произведению их матриц в том же базисе.

Из свойств операций над линейными отображениями, доказанных в § 5.2, следует что совокупность всех линейных операторов в векторном пространстве V является ассоциативной алгеброй. Мы будем обозначать эту алгебру через $L(V)$. Отметим, что если $\dim V = n$, то $\dim L(V) = \dim L_n(K) = n^2$.

Алгебра $L(V)$ обладает единицей. Ею является *тождественный оператор*, который мы будем обозначать буквой \mathcal{E} . Матрица оператора \mathcal{E} в любом базисе есть единичная матрица E .

Линейный оператор $\mathcal{A} \in L(V)$ обратим тогда и только тогда, когда $\text{Ker } \mathcal{A} = 0$ и $\text{Im } \mathcal{A} = V$. Из теоремы 5.2.2 следует, что в конечномерном случае, если $\text{Ker } \mathcal{A} = 0$, то автоматически $\text{Im } \mathcal{A} = V$, и обратно. С другой стороны, ясно, что линейный оператор обратим тогда и только тогда, когда его матрица обратима, т. е. невырождена.

Невырожденные линейные операторы в пространстве V образуют группу, обозначаемую $GL(V)$ и называемую *полной линейной группой* пространства V .

В общем случае размерность подпространства $\text{Im } \mathcal{A}$ называется *рангом линейного оператора* \mathcal{A} и обозначается через $\text{rk } \mathcal{A}$. В силу теоремы 5.2.1 она равна рангу матрицы оператора \mathcal{A} (в любом базисе).

Из формулы (5) следует, что определитель матрицы оператора \mathcal{A} не зависит от выбора базиса. Он называется *определителем линейного оператора* \mathcal{A} и обозначается через $\det \mathcal{A}$.

§ 2. Собственные векторы

Основная задача теории линейных операторов состоит в приведении матрицы линейного оператора к возможно более простому виду за счет выбора подходящего базиса.

Как мы уже отмечали, для этого полезно знать инвариантные подпространства. Особую роль играют одномерные инвариантные

подпространства. Их рассмотрение приводит к понятию собственного вектора.

Определение 1. Ненулевой вектор $e \in V$ называется *собственным вектором* оператора \mathcal{A} , если $\mathcal{A}e = \lambda e$. Число $\lambda \in K$ называется при этом *собственным значением* оператора \mathcal{A} , отвечающим собственному вектору e .

Очевидно, что ненулевой вектор e является собственным тогда и только тогда, когда одномерное подпространство $\langle e \rangle$ инвариантно. В базисе, составленном из собственных векторов (если таковой существует), матрица оператора диагональна, что является пределом мечтаний.

Пример 1. Для оператора дифференцирования в пространстве многочленов единственным с точностью до пропорциональности собственным вектором является многочлен 1 (причем соответствующее собственное значение равно 0). Таким образом, в этом случае из собственных векторов нельзя составить базиса.

Пример 2. Собственные векторы поворота на угол $\alpha \neq k\pi$ в трехмерном пространстве — это векторы, лежащие на оси поворота, причем соответствующее им собственное значение равно 1. При $\alpha = k\pi$ собственными (с собственным значением $(-1)^k$) являются также векторы, ортогональные оси поворота. Таким образом, базис из собственных векторов в этом примере существует только тогда, когда $\alpha = 0$ или π (если считать, что $0 \leq \alpha < 2\pi$).

Для существования собственного вектора с собственным значением λ необходимо и достаточно, чтобы оператор $\mathcal{A} - \lambda \mathcal{E}$ был вырожден, т. е. чтобы $\det(\mathcal{A} - \lambda \mathcal{E}) = 0$. Если $A = (a_{ij})$ — матрица оператора \mathcal{A} в каком-либо базисе, то

$$\det(\mathcal{A} - t\mathcal{E}) = \begin{vmatrix} a_{11} - t & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} - t & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} - t \end{vmatrix},$$

откуда видно, что $\det(\mathcal{A} - t\mathcal{E})$ представляет собой многочлен степени n от t .

Определение 2. Многочлен

$$f_{\mathcal{A}}(t) = (-1)^n \det(\mathcal{A} - t\mathcal{E}) = \det(t\mathcal{E} - \mathcal{A})$$

называется *характеристическим многочленом* оператора \mathcal{A} .

Легко видеть, что коэффициент при t^n многочлена $f_{\mathcal{A}}(t)$ равен 1, а коэффициент при t^{n-1} равен $-\text{tr } A$, где $\text{tr } A$ — след матрицы A

(сумма ее диагональных элементов). Свободный член многочлена $f_{\mathcal{A}}(t)$ равен $f_{\mathcal{A}}(0) = (-1)^n \det A$.

Задача 1. Доказать, что коэффициент при t^{n-k} многочлена $f_{\mathcal{A}}(t)$ равен $(-1)^k \times$ (сумма главных миноров порядка k матрицы A). (Главным минором квадратной матрицы называется определитель ее подматрицы, расположенной симметрично относительно главной диагонали.)

Отметим, что характеристический многочлен линейного оператора в силу своего определения не зависит от выбора базиса. Отсюда, в частности, вытекает, что след матрицы линейного оператора не зависит от базиса.

Выше была фактически доказана

Теорема 1. Собственные значения линейного оператора — это в точности корни его характеристического многочлена.

Следствие. Любой линейный оператор в комплексном векторном пространстве имеет собственный вектор.

Линейный оператор в вещественном векторном пространстве может не иметь собственных векторов, как показывает пример поворота плоскости на угол $\alpha \neq 0, \pi$. Однако использование комплексных чисел позволяет получить полезную информацию и о линейных операторах над полем вещественных чисел. Это достигается с помощью так называемой *комплексификации*.

Пусть V — вещественное векторное пространство. Построим из него комплексное векторное пространство $V(\mathbb{C})$ аналогично тому, как из поля \mathbb{R} строится поле \mathbb{C} . А именно, элементами пространства $V(\mathbb{C})$ будем считать пары (x, y) , где $x, y \in V$. Определим сложение таких пар и умножение на комплексные числа по правилам

$$\begin{aligned}(x_1, y_1) + (x_2, y_2) &= (x_1 + x_2, y_1 + y_2), \\ (\lambda + i\mu)(x, y) &= (\lambda x - \mu y, \mu x + \lambda y).\end{aligned}$$

Легко проверить, что при этом получится векторное пространство над \mathbb{C} . Согласно данному определению, сложение пар вида $(x, 0)$ и их умножение на вещественные числа сводится к соответствующим операциям над их первыми компонентами. отождествим каждую пару вида $(x, 0)$ с вектором $x \in V$; тогда пространство V окажется вложенным в $V(\mathbb{C})$ в виде вещественного подпространства. При этом окажется, что

$$(x, y) = x + iy.$$

Любой базис пространства V (над \mathbb{R}) является в то же время базисом пространства $V(\mathbb{C})$ (над \mathbb{C}). Однако в пространстве $V(\mathbb{C})$ существуют и другие базисы.

Любой линейный оператор \mathcal{A} в пространстве V однозначно продолжается до линейного оператора $\mathcal{A}_{\mathbb{C}}$ в пространстве $V(\mathbb{C})$. При этом в базисе, составленном из вещественных векторов, оператор $\mathcal{A}_{\mathbb{C}}$ имеет такую же матрицу, как и оператор \mathcal{A} .

Оператор $\mathcal{A}_{\mathbb{C}}$ может иметь мнимые собственные значения и соответствующие им мнимые собственные векторы. Какой смысл они имеют в вещественных терминах?

Предложение 1. Вектор $x + iy$ ($x, y \in V$) является собственным вектором оператора $\mathcal{A}_{\mathbb{C}}$ с мнимым собственным значением $\lambda + i\mu$ ($\lambda, \mu \in \mathbb{R}$, $\mu \neq 0$) тогда и только тогда, когда $U = \langle x, y \rangle \subset V$ — двумерное инвариантное подпространство для оператора \mathcal{A} , причем

$$\begin{aligned}\mathcal{A}x &= \lambda x - \mu y, \\ \mathcal{A}y &= \mu x + \lambda y.\end{aligned}\tag{14}$$

Доказательство проводится непосредственным вычислением. Работы (14) означают, что в базисе $\{x, y\}$ пространства U оператор $\mathcal{A}|_U$ имеет матрицу

$$\begin{pmatrix} \lambda & \mu \\ -\mu & \lambda \end{pmatrix}.\tag{15}$$

Из них также следует, что вектор $x - iy$ является собственным вектором оператора $\mathcal{A}_{\mathbb{C}}$ с собственным значением $\lambda - i\mu$.

Пример 3. Оператор \mathcal{A} поворота евклидовой плоскости на угол α в ортонормированном базисе $\{e_1, e_2\}$ имеет матрицу $\Pi(\alpha)$ (см. (9)). Следовательно, вектор $e_1 + ie_2$ является собственным вектором оператора $\mathcal{A}_{\mathbb{C}}$ с собственным значением $\cos \alpha - i \sin \alpha$, а вектор $e_1 - ie_2$ — собственным вектором с собственным значением $\cos \alpha + i \sin \alpha$. Таким образом, матрица поворота может быть приведена к диагональному виду в комплексном пространстве.

В качестве следствия предыдущего предложения получается важная

Теорема 2. Для любого линейного оператора над полем вещественных чисел существует одномерное или двумерное инвариантное подпространство.

При заданном собственном значении λ собственные векторы находятся из системы однородных линейных уравнений

$$(A - \lambda E)X = 0,$$

где X обозначает столбец координат неизвестного вектора. Вместе с нулевым вектором они составляют подпространство

$$V_\lambda(\mathcal{A}) = \text{Ker}(\mathcal{A} - \lambda \mathcal{E}),$$

называемое *собственным подпространством* оператора \mathcal{A} , отвечающим собственному значению λ . Размерность этого подпространства равна $n - \text{rk}(\mathcal{A} - \lambda \mathcal{E})$, где $n = \dim V$.

Теорема 3. *Собственные подпространства, отвечающие различным собственным значениям $\lambda_1, \dots, \lambda_k$ оператора \mathcal{A} , линейно независимы.*

Доказательство. Докажем утверждение теоремы индукцией по k . При $k = 1$ доказывать нечего. Пусть $k > 1$ и

$$e_1 + \dots + e_{k-1} + e_k = 0 \quad (e_i \in V_{\lambda_i}(\mathcal{A})).$$

Применяя оператор \mathcal{A} , получаем

$$\lambda_1 e_1 + \dots + \lambda_{k-1} e_{k-1} + \lambda_k e_k = 0.$$

Вычитая отсюда исходное равенство, умноженное на λ_k , получаем

$$(\lambda_1 - \lambda_k) e_1 + \dots + (\lambda_{k-1} - \lambda_k) e_{k-1} = 0,$$

откуда в силу предположения индукции следует, что $e_1 = \dots = e_{k-1} = 0$. Но тогда и $e_k = 0$. \square

Следствие. *Если характеристический многочлен $f_{\mathcal{A}}(t)$ имеет n различных корней, то существует базис из собственных векторов оператора \mathcal{A} .*

Указанное условие не является необходимым для существования базиса из собственных векторов. Так, для тождественного оператора \mathcal{E} все векторы являются собственными, и, стало быть, любой базис состоит из собственных векторов, однако его характеристический многочлен $f_{\mathcal{E}}(t) = (t - 1)^n$ имеет единственный (но n -кратный) корень 1.

Рассмотрим два более интересных (и важных) примера.

Пример 4. Пусть $V = U \oplus W$. Линейный оператор \mathcal{P} , определяемый формулой

$$\mathcal{P}(y + z) = y \quad (y \in U, z \in W),$$

называется *проектором* на U параллельно W . Очевидно, что

$$U = V_1(\mathcal{P}), \quad W = V_0(\mathcal{P}).$$

В базисе пространства V , составленном из базисов подпространств U и W , оператор \mathcal{P} записывается диагональной матрицей с числами 1 и 0 по диагонали.

Задача 2. Доказать, что линейный оператор \mathcal{P} является проектором (для каких-то U и W) тогда и только тогда, когда $\mathcal{P}^2 = \mathcal{P}$.

Пример 5. В тех же обозначениях линейный оператор \mathcal{R} над полем характеристики $\neq 2$, определяемый формулой

$$\mathcal{R}(y + z) = y - z \quad (y \in U, z \in W),$$

называется *отражением* относительно U параллельно W . Очевидно, что

$$U = V_1(\mathcal{R}), \quad W = V_{-1}(\mathcal{R}).$$

В базисе пространства V , составленном из базисов U и W , оператор \mathcal{R} записывается диагональной матрицей с числами 1 и -1 по диагонали.

Задача 3. Доказать, что линейный оператор \mathcal{R} является отражением (для каких-то U и W) тогда и только тогда, когда $\mathcal{R}^2 = \mathcal{E}$.

Для получения необходимого и достаточного условия существования базиса из собственных векторов докажем сначала

Предложение 2. *Характеристический многочлен ограничения линейного оператора на инвариантное подпространство делит характеристический многочлен самого оператора.*

Доказательство. Пусть \mathcal{B} — ограничение оператора \mathcal{A} на инвариантное подпространство $U \subset V$. В базисе пространства V , первые векторы которого составляют базис подпространства U , матрица A оператора \mathcal{A} имеет вид (6), где B — матрица оператора \mathcal{B} . Следовательно,

$$f_{\mathcal{A}}(t) = f_{\mathcal{B}}(t) \cdot \det(tE - C). \quad \square \quad (16)$$

Следствие. *Размерность собственного подпространства линейного оператора не превосходит кратности соответствующего корня характеристического многочлена.*

Доказательство. Пусть $\dim V_{\lambda}(\mathcal{A}) = k$. Тогда характеристический многочлен ограничения оператора \mathcal{A} на $V_{\lambda}(\mathcal{A})$ равен $(t - \lambda)^k$. Применяя предложение 2 к подпространству $V_{\lambda}(\mathcal{A})$, мы получаем, что $(t - \lambda)^k$ делит характеристический многочлен оператора \mathcal{A} . \square

Пример 6. Рассмотрим оператор дифференцирования в пространстве многочленов степени не выше n . Из вида его матрицы, найденной в примере 1.4, следует, что его характеристический много-

член равен t^{n+1} . Он имеет корень 0 кратности $n + 1$, однако размерность соответствующего собственного подпространства равна 1 (см. пример 1). Этот пример показывает, что размерность собственного подпространства может быть строго меньше кратности соответствующего корня характеристического многочлена.

Теорема 4. Для существования базиса из собственных векторов линейного оператора \mathcal{A} необходимо и достаточно, чтобы выполнялись следующие условия:

1) характеристический многочлен $f_{\mathcal{A}}(t)$ разлагается на линейные множители;

2) размерность каждого собственного подпространства равна кратности соответствующего корня многочлена $f_{\mathcal{A}}(t)$.

Доказательство. Пусть $\lambda_1, \dots, \lambda_s$ — все корни многочлена $f_{\mathcal{A}}(t)$ и k_1, \dots, k_s — их кратности. Собственное подпространство, отвечающее λ_i , обозначим через V_i . Согласно следствию предложения 2, $\dim V_i \leq k_i$ и, значит,

$$\sum_i \dim V_i \leq \sum_i k_i \leq n. \quad (17)$$

Однако единственный способ получить базис из собственных векторов — это взять объединение базисов собственных подпространств. Для того чтобы при этом действительно получился базис пространства V , необходимо и достаточно, чтобы

$$\sum_i \dim V_i = n.$$

Ввиду (17) это равносильно тому, что $\sum_i k_i = n$ и $\dim V_i = k_i$ для всех i .

Первое из этих условий означает, что $f_{\mathcal{A}}(t)$ разлагается на линейные множители, а второе — это как раз условие 2) теоремы. \square

§ 3. Линейные операторы и билинейные функции в евклидовом пространстве

Пусть V — евклидово пространство и $\{e_1, \dots, e_n\}$ — его ортонормированный базис.

Каждому вектору $a \in V$ соответствует линейная функция

$$\varphi_a(x) = (x, a). \quad (18)$$

При этом коэффициенты $\varphi_a(e_i) = (e_i, a)$ линейной функции φ_a в базисе $\{e_1, \dots, e_n\}$ равны координатам вектора a в этом базисе. Отсюда следует, что отображение $a \mapsto \varphi_a$ есть изоморфизм пространства V на пространство V^* . Отметим, что определение этого изоморфизма не зависит от выбора базиса. Таким образом, в случае конечномерного евклидова пространства как бы исчезает разница между пространством и его сопряженным. Часто говорят «отождествим при помощи канонического изоморфизма евклидово пространство V с его сопряженным пространством», имея в виду указанный выше изоморфизм.

Аналогично, каждому линейному оператору \mathcal{A} в пространстве V соответствует билинейная функция

$$\varphi_{\mathcal{A}}(x, y) = (x, \mathcal{A}y). \quad (19)$$

При этом матрица билинейной функции $\varphi_{\mathcal{A}}(x, y)$ в базисе $\{e_1, \dots, e_n\}$ совпадает с матрицей оператора \mathcal{A} в этом базисе. В самом деле, $\varphi_{\mathcal{A}}(e_i, e_j) = (e_i, \mathcal{A}e_j)$ есть не что иное, как i -я координата вектора $\mathcal{A}e_j$. Отсюда следует, что отображение $\mathcal{A} \mapsto \varphi_{\mathcal{A}}$ есть изоморфизм пространства линейных операторов на пространство билинейных функций в пространстве V . Определение этого изоморфизма не зависит от выбора базиса. Однако в неортонормированном базисе матрица функции $\varphi_{\mathcal{A}}$ не обязана совпадать с матрицей оператора \mathcal{A} .

Для каждой билинейной функции φ можно определить «транспонированную» функцию

$$\varphi^T(x, y) = \varphi(y, x),$$

матрицей которой в любом базисе является транспонированная матрица функции φ . Линейный оператор \mathcal{A}^* , соответствующий функции $\varphi_{\mathcal{A}}^T$, называется *сопряженным оператором* по отношению к \mathcal{A} . Иначе говоря, сопряженный оператор определяется тождеством

$$(\mathcal{A}^*x, y) = (x, \mathcal{A}y). \quad (20)$$

Матрицей оператора \mathcal{A}^* в ортонормированном базисе является транспонированная матрица оператора \mathcal{A} .

Симметрическим (соответственно кососимметрическим) билинейным функциям соответствуют так называемые *симметрические* (соответственно *кососимметрические*) линейные операторы. Они

характеризуются тем, что $\mathcal{A}^* = \mathcal{A}$ (соответственно $\mathcal{A}^* = -\mathcal{A}$), а в матричных терминах — тем, что их матрица в ортонормированном базисе симметрична (соответственно кососимметрична). Симметрические операторы называют также *самосопряженными*.

Пример 1. *Ортогональный проектор* на подпространство является симметрическим оператором (проверьте это).

Линейные операторы, для которых $\mathcal{A}^* = \mathcal{A}^{-1}$, называются *ортогональными*. Иначе говоря, оператор \mathcal{A} ортогонален, если

$$(\mathcal{A}x, \mathcal{A}y) = (x, y), \quad (21)$$

т. е. если \mathcal{A} сохраняет скалярное произведение векторов. Из тождества

$$(x, y) = \frac{1}{2}(|x + y|^2 - |x|^2 - |y|^2)$$

следует, что оператор \mathcal{A} ортогонален тогда и только тогда, когда он сохраняет длины векторов.

Замечание. Мы видим, таким образом, что ортогональные линейные преобразования пространства \mathbb{R}^n , определенные в примере 4.1.9, — это то же, что ортогональные операторы в смысле данного выше определения, если считать, что скалярное умножение в \mathbb{R}^n определено как в примере 5.5.2.

Пример 2. Линейный оператор, индуцированный в пространстве геометрических векторов любым движением, ортогонален.

Пример 3. *Ортогональное отражение* относительно подпространства (т. е. отражение параллельно ортогональному подпространству) является ортогональным оператором.

В матричных терминах ортогональные операторы характеризуются тем, что их матрица в ортонормированном базисе ортогональна (см. определение 5.5.4).

Предложение 1. *Линейный оператор любого из рассмотренных выше трех типов, т. е. симметрический, кососимметрический или ортогональный, обладает следующим свойством: если подпространство U инвариантно, то и его ортогональное дополнение U^\perp инвариантно.*

Доказательство. Рассмотрим наиболее сложный случай ортогонального оператора \mathcal{A} . Заметим, прежде всего, что оператор $\mathcal{A}|_U$ также ортогонален и, следовательно, невырожден. Поэтому для лю-

бого вектора $x \in U$ найдется такой вектор $z \in U$, что $x = \mathcal{A}z$. Возьмем теперь любой вектор $y \in U^\perp$. Тогда, используя предыдущие обозначения, получаем для любого $x \in U$

$$(x, \mathcal{A}y) = (\mathcal{A}z, \mathcal{A}y) = (z, y) = 0,$$

откуда следует, что $\mathcal{A}y \in U^\perp$. □

С помощью этого предложения и теоремы 2.2 мы можем, рассуждая индукцией по размерности, получить канонический вид для матриц линейных операторов рассматриваемых трех типов.

Теорема 1. *Для любого симметрического оператора \mathcal{A} существует ортонормированный базис из собственных векторов.*

Доказательство. Достаточно доказать существование хотя бы одного собственного вектора. В силу теоремы 2.2 достаточно сделать это для двумерного пространства. Матрица симметрического оператора в ортонормированном базисе в этом случае имеет вид

$\begin{pmatrix} a & b \\ b & c \end{pmatrix}$, и характеристический многочлен равен

$$f_{\mathcal{A}}(t) = t^2 - (a+c)t + (ac - b^2).$$

Дискриминант этого квадратного трехчлена

$$D = (a+c)^2 - 4(ac - b^2) = (a-c)^2 + 4b^2$$

всегда неотрицателен, так что $f_{\mathcal{A}}(t)$ имеет вещественные корни и, значит, \mathcal{A} имеет собственные векторы. □

Следствие 1. *Характеристический многочлен симметрического оператора разлагается на линейные множители (над \mathbb{R}); размерность каждого собственного подпространства равна кратности соответствующего корня; собственные подпространства, отвечающие различным корням, ортогональны друг другу.*

Доказательство. Для доказательства последнего утверждения надо заметить, что если $\{e_1, \dots, e_n\}$ — базис из собственных векторов оператора \mathcal{A} , причем $\mathcal{A}e_i = \lambda_i e_i$, то $V_\lambda(\mathcal{A})$ есть линейная оболочка тех e_i , для которых $\lambda_i = \lambda$. Впрочем, его легко можно доказать и непосредственно. В самом деле, пусть $x \in V_\lambda(\mathcal{A})$, $y \in V_\mu(\mathcal{A})$, $\lambda \neq \mu$. Тогда

$$\lambda(x, y) = (\mathcal{A}x, y) = (x, \mathcal{A}y) = \mu(x, y),$$

откуда $(x, y) = 0$. □

Используя описанное выше соответствие между симметрическими операторами и симметрическими билинейными функциями, получаем

Следствие 2. Для любой квадратичной функции q в евклидовом пространстве существует ортонормированный базис, в котором ее матрица диагональна, т. е.

$$q(x) = \lambda_1 x_1^2 + \dots + \lambda_n x_n^2. \quad (22)$$

Нужно понимать, что в формулировке этого следствия речь идет об ортонормированности в смысле скалярного умножения, а не в смысле симметрической билинейной функции φ , соответствующей q . Однако, поскольку матрица функции φ в указанном базисе диагональна, этот базис является также ортогональным (но, вообще говоря, не ортонормированным) в смысле функции φ .

Отметим, что числа $\lambda_1, \dots, \lambda_n$ — это собственные значения соответствующего симметрического оператора и, следовательно, определены однозначно с точностью до перестановки.

Выражение (22) называют *каноническим видом* квадратичной функции q , а нахождение ортонормированного базиса, в котором функция q имеет такой вид, часто называют *приведением к главным осям*.

Используя соответствие между симметрическими операторами и квадратичными функциями в евклидовом пространстве в обратном направлении, можно получить другое доказательство существования собственного вектора у симметрического оператора.

А именно, пусть q — квадратичная функция, соответствующая данному симметрическому оператору \mathcal{A} , т. е.

$$q(x) = (\mathcal{A}x, x).$$

Заметим, что функция q , будучи непрерывной, должна иметь максимум на единичной сфере S пространства V , задаваемой уравнением

$$(x, x) = 1.$$

Предложение 2. Всякая точка максимума функции q на сфере S является собственным вектором оператора \mathcal{A} , а сам максимум равен соответствующему собственному значению.

Доказательство. Касательное пространство сферы S в точке x задается уравнением

$$(x, dx) = 0,$$

Следовательно, конечный результат сколь угодно сложного реально-го движения твердого тела с закрепленной точкой — такой же, как при простом повороте вокруг подходящей оси на подходящий угол. Эта совершенно не тривиальная теорема называется *теоремой Эйлера*.

Ортогональные операторы в евклидовом пространстве V образуют подгруппу группы $GL(V)$, называемую *ортогональной группой* и обозначаемую $O(V)$. Соответственно этому ортогональные матрицы образуют подгруппу группы $GL_n(\mathbb{R})$, обозначаемую O_n (это согласуется с обозначением, введенным в примере 4.1.9).

Как мы уже заметили в § 5.5, определитель ортогональной матрицы равен ± 1 . Ортогональные матрицы с определителем 1 образуют подгруппу индекса 2 в O_n , обозначаемую SO_n . Соответственно этому ортогональные операторы с определителем 1 образуют подгруппу индекса 2 в $O(V)$, называемую *специальной ортогональной группой* и обозначаемую $SO(V)$. Операторы из $SO(V)$ геометрически истолковываются как ортогональные операторы, сохраняющие ориентацию пространства (см. пример 4.6.16).

Пример 4. Группа $O_2 = O(E^2)$ состоит из поворотов, составляющих подгруппу $SO_2 = SO(E^2)$, и отражений относительно прямых. Обозначим через s_α поворот на угол α и через r_α — отражение относительно прямой, образующей угол α с какой-либо фиксированной прямой l . Ясно, что $s_\alpha s_\beta = s_{\alpha+\beta}$. Далее, произведение поворота и отражения меняет ориентацию и, следовательно, является отражением. Проследив за какой-нибудь одной точкой (см. рис. 2, а), б)), легко установить, что

$$s_\alpha r_\beta = r_{\beta+\frac{\alpha}{2}}, \quad r_\beta s_\alpha = r_{\beta-\frac{\alpha}{2}}.$$

Наконец, произведение двух отражений сохраняет ориентацию и, следовательно, является поворотом. Проследив за одной точкой

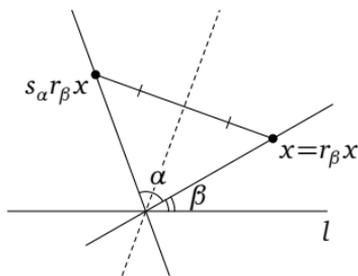


Рис. 2, а)

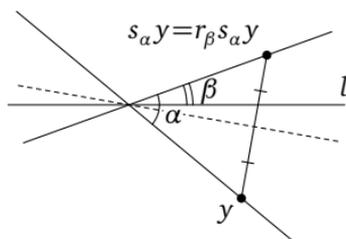


Рис. 2, б)

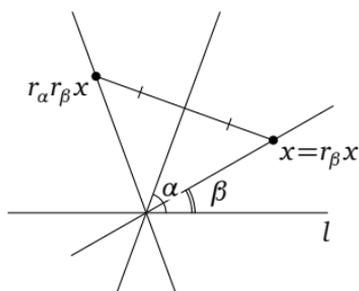


Рис. 3

(рис. 3), легко установить, что

$$r_\alpha r_\beta = s_{2(\alpha-\beta)},$$

т. е. произведение двух отражений есть поворот на удвоенный угол между их осями. В частности, отсюда следует, что группа O_2 порождается отражениями.

Задача 1. Доказать, что группа $O(V)$ порождается отражениями относительно $(n-1)$ -мерных подпространств (где $n = \dim V$).

Всякий линейный оператор в евклидовом пространстве единственным образом представляется в виде суммы симметрического и кососимметрического операторов (ср. пример 5.1.1). Имеется мультипликативный аналог этого разложения, в котором кососимметрический оператор заменяется ортогональным (почему так происходит, станет ясно в гл. 12).

Теорема 4. *Всякий невырожденный линейный оператор в евклидовом пространстве единственным образом представляется в виде произведения положительно определенного симметрического и ортогонального операторов.*

Такое представление линейного оператора называется его *полярным разложением*.

Перед тем как доказывать эту теорему, докажем следующее

Предложение 3. *Всякий положительно определенный симметрический оператор \mathcal{B} единственным образом представляется в виде $\mathcal{B} = \mathcal{C}^2$, где \mathcal{C} — также положительно определенный симметрический оператор.*

Доказательство. Пусть $\lambda_1, \dots, \lambda_s$ — (различные) собственные значения оператора \mathcal{B} и V_1, \dots, V_s — соответствующие собственные подпространства. По условию λ_i положительны. Положим $\mu_i = \sqrt{\lambda_i}$ (арифметическое значение корня). Тогда линейный оператор \mathcal{C} , действующий в V_i как умножение на μ_i , удовлетворяет условиям предложения. (В частности, он симметричен, поскольку его матрица в ортонормированном базисе, составленном из собственных векторов оператора \mathcal{B} , диагональна.)

Обратно, пусть оператор \mathcal{C} удовлетворяет условиям предложения. Пусть μ_1, \dots, μ_s — его (различные) собственные значения и W_1, \dots, W_s — соответствующие собственные подпространства. Тогда

оператор $\mathcal{C}^2 = \mathcal{B}$ действует на W_i как умножение на μ_i^2 . Следовательно, при подходящей нумерации $\mu_i^2 = \lambda_i$ и $W_i = V_i$. Это показывает, что оператор \mathcal{C} определен однозначно. \square

Доказательство теоремы 4. Пусть \mathcal{A} — невырожденный линейный оператор. Предположим, что $\mathcal{A} = \mathcal{C}\mathcal{O}$, где \mathcal{C} — положительно определенный симметрический, а \mathcal{O} — ортогональный операторы. Тогда

$$\mathcal{A}\mathcal{A}^* = \mathcal{C}\mathcal{O}\mathcal{O}^*\mathcal{C}^* = \mathcal{C}^2.$$

Ввиду предложения 3 этим однозначно определяется оператор \mathcal{C} , а тем самым и \mathcal{O} .

Обратно, из равенства

$$(x, \mathcal{A}\mathcal{A}^*y) = (\mathcal{A}^*x, \mathcal{A}^*y)$$

и невырожденности оператора \mathcal{A} (и, значит, \mathcal{A}^*) следует, что $\mathcal{A}\mathcal{A}^*$ — положительно определенный симметрический оператор. Пользуясь предложением 3, найдем такой положительно определенный симметрический оператор \mathcal{C} , что $\mathcal{A}\mathcal{A}^* = \mathcal{C}^2$, и положим $\mathcal{O} = \mathcal{C}^{-1}\mathcal{A}$. Тогда $\mathcal{A} = \mathcal{C}\mathcal{O}$ и

$$\mathcal{A}\mathcal{A}^* = \mathcal{C}\mathcal{O}\mathcal{O}^*\mathcal{C} = \mathcal{C}^2,$$

откуда после сокращения на \mathcal{C} получаем, что $\mathcal{O}\mathcal{O}^* = \mathcal{E}$, т. е. \mathcal{O} — ортогональный оператор. \square

Пример 5. Всякую деформацию твердого тела с закрепленной точкой в первом приближении можно рассматривать как невырожденный линейный оператор. Пусть $\mathcal{A} = \mathcal{C}\mathcal{O}$ — полярное разложение этого оператора. Тогда \mathcal{O} — это поворот вокруг некоторой оси, который не является истинной деформацией в том смысле, что он не приводит к возникновению каких-либо напряжений в теле. С другой стороны, оператор \mathcal{C} по теореме 1 есть комбинация растяжений (или сжатий) в трех взаимно перпендикулярных направлениях и тем самым представляет собой «чистую деформацию». Именно этот оператор, называемый тензором деформации, участвует в формулировке закона Гука.

Задача 2. Доказать, что всякую матрицу $A \in GL_n(\mathbb{R})$ можно представить в виде $A = O_1 D O_2$, где O_1, O_2 — ортогональные матрицы, а D — диагональная матрица с положительными элементами. Насколько однозначно такое представление?

Аналогичная теория имеется для линейных операторов в эрмитовом пространстве, причем она даже проще, так как в эрмитовом пространстве всякий линейный оператор имеет собственный век-

тор. Изложим ее вкратце, опуская доказательства, аналогичные приведенным выше в евклидовом случае.

Для любого линейного оператора \mathcal{A} в эрмитовом пространстве определяется сопряженный оператор \mathcal{A}^* по формуле (20). Если оператор \mathcal{A} в некотором ортонормированном базисе имеет матрицу A , то оператор \mathcal{A}^* в том же базисе имеет матрицу A^* . (Напомним, что $A^* = \bar{A}^T$.)

Линейный оператор \mathcal{A} называется эрмитовым (соответственно косоэрмитовым, унитарным), если $\mathcal{A}^* = \mathcal{A}$ (соответственно $\mathcal{A}^* = -\mathcal{A}$, $\mathcal{A}^* = \mathcal{A}^{-1}$). Это эквивалентно тому, что его матрица в ортонормированном базисе эрмитова (соответственно косоэрмитова, унитарна). Эрмитовы операторы называют также самосопряженными.

Для любого из этих типов линейных операторов доказывается существование ортонормированного базиса из собственных векторов. При этом собственные значения эрмитова оператора вещественны, косоэрмитова — чисто мнимы, а унитарного — по модулю равны единице.

Докажем, например, что собственные значения эрмитова оператора \mathcal{A} вещественны. Пусть e — собственный вектор оператора \mathcal{A} с собственным значением λ . Тогда

$$\bar{\lambda}(e, e) = (\mathcal{A}e, e) = (e, \mathcal{A}e) = \lambda(e, e),$$

откуда $\bar{\lambda} = \lambda$.

Формула (19) устанавливает биекцию между множествами эрмитовых операторов и эрмитовых полуторалинейных функций. При этом в любом ортонормированном базисе матрицы эрмитова оператора и соответствующей ему эрмитовой функции совпадают.

Применяя теорему о существовании ортонормированного базиса из собственных векторов эрмитова оператора, мы получаем, что для любой эрмитовой квадратичной функции q в эрмитовом пространстве существует ортонормированный базис, в котором ее матрица диагональна, т. е.

$$q(x) = \lambda_1|x_1|^2 + \dots + \lambda_n|x_n|^2. \quad (23)$$

Числа $\lambda_1, \dots, \lambda_n$ определены однозначно с точностью до перестановки, так как это собственные значения соответствующего эрмитова оператора. Выражение (23) называют каноническим видом эрмитовой квадратичной функции q .

Эрмитов оператор называется *положительно определенным*, если соответствующая ему эрмитова квадратичная функция положительно определена или, что равносильно, если все его собственные значения положительны.

Унитарные операторы в эрмитовом пространстве V образуют подгруппу группы $GL(V)$, называемую *унитарной группой* и обозначаемую $U(V)$. Соответственно этому унитарные матрицы образуют подгруппу группы $GL_n(\mathbb{C})$, обозначаемую U_n .

Унитарные операторы (соответственно матрицы) с определителем 1 образуют подгруппу в $U(V)$ (соответственно в U_n), называемую *специальной унитарной группой* и обозначаемую $SU(V)$ (соответственно SU_n).

Всякий невырожденный линейный оператор в эрмитовом пространстве единственным образом представляется в виде произведения положительно определенного эрмитова и унитарного операторов. Такое представление линейного оператора называется его *полярным разложением*. В одномерном случае линейный оператор есть просто комплексное число, а его полярное разложение — тригонометрическая форма этого числа. Поскольку тригонометрическая форма комплексного числа связана с полярными координатами на плоскости, это объясняет термин «полярное разложение» в общем случае.

Комплексификация $V(\mathbb{C})$ евклидова пространства V каноническим образом превращается в эрмитово пространство, если определить скалярное умножение по формуле

$$(x_1 + iy_1, x_2 + iy_2) = [(x_1, x_2) + (y_1, y_2)] + i[(x_1, y_2) - (y_1, x_2)].$$

При этом комплексное продолжение $\mathcal{A}_{\mathbb{C}}$ симметрического (соответственно кососимметрического, ортогонального) оператора \mathcal{A} будет эрмитовым (соответственно косоэрмитовым, унитарным) оператором.

Используя эти соображения, можно дать еще одно доказательство существования собственного вектора у симметрического оператора \mathcal{A} в евклидовом пространстве V . А именно, пусть $x + iy$ ($x, y \in V$) — какой-либо собственный вектор оператора $\mathcal{A}_{\mathbb{C}}$. Так как оператор $\mathcal{A}_{\mathbb{C}}$ эрмитов, то соответствующее собственное значение λ вещественно и, значит,

$$\mathcal{A}x = \lambda x, \quad \mathcal{A}y = \lambda y.$$

Хотя бы один из векторов x, y отличен от нуля; он и будет собственным вектором оператора \mathcal{A} .

§ 4. Жорданова форма

Для некоторых специальных типов линейных операторов, как, например, симметрических, эрмитовых и унитарных, рассмотренных в предыдущем параграфе, удается доказать возможность приведения их матрицы к диагональному виду. В общем случае для этого имеются препятствия, указанные в теореме 2.4.

Первое из них состоит в том, что характеристический многочлен может не разлагаться на линейные множители, т. е. иметь менее чем n корней. Его не существует для линейных операторов над полем комплексных чисел. В случае линейного оператора над полем вещественных чисел можно работать с его комплексификацией, что в какой-то мере снимает проблему: выбор удачного базиса из комплексных векторов позволяет понять и действие исходного оператора в вещественном пространстве. Так, в § 2 мы видели, что всякому мнимому собственному вектору отвечает двумерное инвариантное подпространство в вещественном пространстве. Как будет показано в § 9.5, аналогичное расширение основного поля возможно и в общем случае.

Второе препятствие состоит в том, что размерность какого-либо собственного подпространства может оказаться меньше кратности соответствующего корня характеристического многочлена. Тогда приходится расстаться с мечтой привести матрицу оператора к диагональной форме, но, если характеристический многочлен разлагается на линейные множители, ее можно привести к так называемой жордановой форме, минимально отличающейся от диагональной. Этому и посвящен настоящий параграф.

Коль скоро собственных векторов может оказаться недостаточно, естественно рассмотреть какие-то более общие векторы.

Определение 1. Вектор $e \in V$ называется *корневым вектором* линейного оператора \mathcal{A} , отвечающим числу $\lambda \in K$, если

$$(\mathcal{A} - \lambda \mathcal{E})^m e = 0$$

для некоторого $m \in \mathbb{Z}_+$. Наименьшее из таких m называется *высокой* корневого вектора e .

В частности, собственные векторы — это корневые векторы высоты 1. Удобно считать нулевой вектор корневым вектором высоты 0 (отвечающим любому λ).

Пример 1. Для оператора дифференцирования в пространстве $C^\infty(\mathbb{R})$ бесконечно дифференцируемых функций собственные векторы, отвечающие числу λ — это функции, пропорциональные $e^{\lambda x}$, а корневые векторы — это функции вида $p(x)e^{\lambda x}$, где $p(x)$ — многочлен; при этом высота такого корневого вектора равна $\deg p + 1$. В частности, корневые векторы, отвечающие числу 0, — это многочлены.

Если e — корневой вектор высоты $m > 0$, то вектор

$$f = (\mathcal{A} - \lambda \mathcal{E})^{m-1} e$$

собственный с собственным значением λ . Следовательно, λ — корень характеристического многочлена.

Легко видеть, что корневые векторы, отвечающие корню λ , образуют подпространство. Оно называется *корневым подпространством* и обозначается $V^\lambda(\mathcal{A})$. Ясно, что

$$V^\lambda(\mathcal{A}) \supset V_\lambda(\mathcal{A}).$$

Если e — корневой вектор высоты $m > 0$, то $(\mathcal{A} - \lambda \mathcal{E})e$ — корневой вектор высоты $m - 1$. Отсюда следует, что корневое подпространство $V^\lambda(\mathcal{A})$ инвариантно относительно $\mathcal{A} - \lambda \mathcal{E}$, а значит, и относительно \mathcal{A} .

Множество корневых векторов высоты $\leq m$ — это не что иное, как ядро оператора $(\mathcal{A} - \lambda \mathcal{E})^m$. Таким образом, корневое подпространство $V^\lambda(\mathcal{A})$ — это объединение возрастающей цепочки подпространств

$$\text{Ker}(\mathcal{A} - \lambda \mathcal{E}) \subset \text{Ker}(\mathcal{A} - \lambda \mathcal{E})^2 \subset \dots$$

В конечномерной ситуации эта цепочка, начиная с некоторого места, стабилизируется, и, значит, $V^\lambda(\mathcal{A}) = \text{Ker}(\mathcal{A} - \lambda \mathcal{E})^m$ для некоторого m . В базисе пространства $V^\lambda(\mathcal{A})$, согласованном с этой цепочкой подпространств, оператор $\mathcal{A} - \lambda \mathcal{E}$ записывается нильтреугольной матрицей (т. е. треугольной матрицей с нулями на диагонали), а оператор \mathcal{A} соответственно этому — треугольной матрицей с числом λ на диагонали. Отсюда мы получаем два следствия:

- 1) характеристический многочлен ограничения оператора \mathcal{A} на $V^\lambda(\mathcal{A})$ равен $(t - \lambda)^k$, где $k = \dim V^\lambda(\mathcal{A})$;
- 2) при $\mu \neq \lambda$ оператор $\mathcal{A} - \mu \mathcal{E}$ невырожден на $V^\lambda(\mathcal{A})$.

Задача 1. Доказать, что высота любого корневого вектора, отвечающего корню λ , не превосходит $\dim V^\lambda(\mathcal{A})$.

Докажем теперь ключевое утверждение, оправдывающее понятие корневого вектора.

Предложение 1. *Размерность корневого подпространства равна кратности соответствующего корня характеристического многочлена.*

Доказательство. В базисе $\{e_1, \dots, e_n\}$ пространства V , первые k векторов которого составляют базис подпространства $V^\lambda(\mathcal{A})$, матрица A оператора \mathcal{A} имеет вид (6), где B — матрица оператора $\mathcal{B} = \mathcal{A}|_{V^\lambda(\mathcal{A})}$. Следовательно,

$$f_{\mathcal{A}}(t) = f_{\mathcal{B}}(t) \cdot \det(tE - C) = (t - \lambda)^k \det(tE - C).$$

Пусть \mathcal{C} — линейный оператор в пространстве $W = \langle e_{k+1}, \dots, e_n \rangle$, задаваемый матрицей C . Нам нужно доказать, что λ не является корнем многочлена $\det(tE - C)$, т. е. собственным значением оператора \mathcal{C} .

Предположим противное. Тогда существует такой ненулевой вектор $e \in W$, что $\mathcal{C}e = \lambda e$. Это означает, что

$$\mathcal{A}e = \lambda e + u, \quad u \in V^\lambda(\mathcal{A}),$$

и, следовательно, $(\mathcal{A} - \lambda \mathcal{E})e = u$ — корневой вектор, но тогда и e — корневой вектор, что противоречит определению $V^\lambda(\mathcal{A})$. \square

Предложение 2. *Корневые подпространства, отвечающие различным корням $\lambda_1, \dots, \lambda_k$, линейно независимы.*

Доказательство. Доказательство аналогично доказательству теоремы 2.3 о линейной независимости собственных подпространств. Пусть

$$e_1 + \dots + e_{k-1} + e_k = 0 \quad (e_i \in V^{\lambda_i}(\mathcal{A})).$$

Применим к этому равенству оператор $(\mathcal{A} - \lambda_k \mathcal{E})^m$, где m — высота вектора e_k . Мы получим

$$(\mathcal{A} - \lambda_k \mathcal{E})^m e_1 + \dots + (\mathcal{A} - \lambda_k \mathcal{E})^m e_{k-1} = 0.$$

Если доказывать предложение индукцией по k , то предположение индукции даст

$$(\mathcal{A} - \lambda_k \mathcal{E})^m e_1 = \dots = (\mathcal{A} - \lambda_k \mathcal{E})^m e_{k-1} = 0.$$

Так как оператор $\mathcal{A} - \lambda_k \mathcal{E}$ невырожден на каждом из подпространств $V^{\lambda_1}(\mathcal{A}), \dots, V^{\lambda_{k-1}}(\mathcal{A})$, то отсюда следует, что

$$e_1 = \dots = e_{k-1} = 0;$$

но тогда и $e_k = 0$. □

Предложения 1 и 2 в совокупности позволяют сделать следующий вывод.

Теорема 1. Если характеристический многочлен $f_{\mathcal{A}}(t)$ разлагается на линейные множители, то

$$V = \bigoplus_{i=1}^s V^{\lambda_i}(\mathcal{A}),$$

где $\lambda_1, \dots, \lambda_s$ — (различные) корни многочлена $f_{\mathcal{A}}(t)$.

Исследуем теперь более подробно действие оператора \mathcal{A} на каждом из корневых подпространств.

Определение 2. Линейный оператор \mathcal{N} называют *нильпотентным*, если существует такое $m \in \mathbb{Z}_+$, что $\mathcal{N}^m = 0$. Наименьшее из таких m называют *высотой* нильпотентного оператора \mathcal{N} .

Пример 2. Оператор дифференцирования в пространстве многочленов степени не выше n является нильпотентным оператором высоты $n + 1$.

Так как $V^\lambda(\mathcal{A}) = \text{Ker}(\mathcal{A} - \lambda \mathcal{E})^m$ для некоторого m , то оператор

$$\mathcal{N} = (\mathcal{A} - \lambda \mathcal{E})|_{V^\lambda(\mathcal{A})}$$

нильпотентен. Поэтому наша задача сводится к исследованию нильпотентных операторов.

Пусть \mathcal{N} — нильпотентный оператор в векторном пространстве V .

Высотой вектора $e \in V$ относительно \mathcal{N} называется наименьшее m , для которого $\mathcal{N}^m e = 0$, т. е. высота вектора e как корневого вектора оператора \mathcal{N} (отвечающего корню 0). Очевидно, что высота любого вектора не превосходит высоты самого оператора \mathcal{N} , причем существуют векторы, высота которых равна высоте оператора \mathcal{N} . Мы будем обозначать высоту вектора e через $\text{ht } e$.

Предложение 3. Если $e \in V$ — вектор высоты m , то векторы

$$e, \mathcal{N}e, \mathcal{N}^2e, \dots, \mathcal{N}^{m-1}e$$

линейно независимы. Более точно, всякая нетривиальная линейная комбинация

$$u = \lambda_0 e + \lambda_1 \mathcal{N}e + \lambda_2 \mathcal{N}^2e + \dots + \lambda_{m-1} \mathcal{N}^{m-1}e$$

является ненулевым вектором высоты $m - k$, где k — номер первого ненулевого коэффициента. В частности, если $\lambda_0 \neq 0$, то u — вектор высоты m .

Доказательство. Так как $\mathcal{N}^{m-1}e \neq 0$, но $\mathcal{N}^m e = 0$, то

$$\mathcal{N}^{m-k-1}u = \lambda_k \mathcal{N}^{m-1}e \neq 0, \quad \mathcal{N}^{m-k}u = 0. \quad \square$$

Определение 3. Подпространство $\langle e, \mathcal{N}e, \mathcal{N}^2e, \dots, \mathcal{N}^{m-1}e \rangle$ ($m = \text{ht } e$) называется *циклическим подпространством* нильпотентного оператора \mathcal{N} , порожденным вектором e .

Очевидно, что циклическое подпространство инвариантно относительно \mathcal{N} . Ограничение оператора \mathcal{N} на циклическое подпространство $\langle e, \mathcal{N}e, \mathcal{N}^2e, \dots, \mathcal{N}^{m-1}e \rangle$ имеет высоту m и в базисе $\{\mathcal{N}^{m-1}e, \mathcal{N}^{m-2}e, \dots, \mathcal{N}e, e\}$ задается матрицей

$$J(0) = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & 0 & \dots & 0 & 0 \end{pmatrix},$$

называемой *нильпотентной жордановой клеткой* (порядка m) (ср. пример 1.4).

Предложение 4. Пусть $e \in V$ — вектор максимальной высоты m (равной высоте оператора \mathcal{N}) и

$$U = \langle e, \mathcal{N}e, \mathcal{N}^2e, \dots, \mathcal{N}^{m-1}e \rangle$$

— порожденное им циклическое подпространство. Тогда существует инвариантное подпространство $W \subset V$, дополнительное к U (т. е. такое, что $V = U \oplus W$).

Доказательство. Нам нужно доказать существование такого инвариантного подпространства $W \subset V$, что $U \cap W = 0$ и $U + W = V$. Заведомо существуют инвариантные подпространства, обладающие первым из этих свойств: таковым является, например, нулевое подпространство. Выберем из них максимальное (т. е. такое, которое нельзя увеличить с сохранением указанного свойства). Обозначим его через W и докажем, что $U + W = V$.

Предположим, что это не так, и пусть $v \notin U + W$. Так как $\mathcal{N}^m v = 0$, то существует такое k , что $\mathcal{N}^{k-1}v \notin U + W$, но $\mathcal{N}^k v \in U + W$. Заме-

нив v на $\mathcal{N}^{k-1}v$, мы можем считать, что

$$\mathcal{N}v \in U + W.$$

Пусть $\mathcal{N}v = u + w$ ($u \in U$, $w \in W$). Из равенства $\mathcal{N}^m v = 0$ получаем

$$\mathcal{N}^{m-1}u + \mathcal{N}^{m-1}w = 0,$$

откуда $\mathcal{N}^{m-1}u = 0$. (Напомним, что подпространства U и W инвариантны и линейно независимы.) Это означает, что $\text{ht } u < m$. Согласно предложению 3, отсюда следует, что u содержится в подпространстве

$$\mathcal{N}U = \langle \mathcal{N}e, \mathcal{N}^2e, \dots, \mathcal{N}^{m-1}e \rangle.$$

Пусть $u = \mathcal{N}u'$ ($u' \in U$). Тогда $\mathcal{N}(v - u') \in W$. Заменив v на $v - u'$, мы можем считать, что

$$\mathcal{N}v \in W.$$

Рассмотрим теперь подпространство

$$W' = W + \langle v \rangle$$

Из нашего условия следует, что оно инвариантно. Докажем, что $U \cap W' = 0$.

Предположим противное. Пусть $y \in U \cap W'$ — ненулевой вектор. Так как $U \cap W = 0$, то $y = z + \lambda v$, где $z \in W$ и $\lambda \neq 0$. Разделив y на λ , можно считать, что $\lambda = 1$, т. е. $y = z + v$. Но тогда $v = y - z \in U + W$, что противоречит нашему выбору вектора v .

Таким образом, $U \cap W' = 0$, что противоречит нашему выбору подпространства W . Следовательно, $U + W = V$, что и требовалось доказать. \square

Предложения 3 и 4 позволяют доказать следующую теорему.

Теорема 2. *Пространство V может быть разложено в прямую сумму циклических подпространств оператора \mathcal{N} . Количество слагаемых в таком разложении равно $\dim \text{Ker } \mathcal{N}$.*

Доказательство. Будем доказывать теорему индукцией по $n = \dim V$. При $n = 1$ утверждение теоремы очевидно. При $n > 1$ пусть $U \subset V$ — циклическое подпространство, порожденное каким-либо вектором максимальной высоты. Согласно предложению 4, существует такое инвариантное подпространство $W \subset V$, что $V = U \oplus W$. По предположению индукции пространство W может быть разложено в прямую сумму циклических подпространств. Вместе с подпространством U это дает нужное разложение всего пространства V .

Докажем второе утверждение теоремы. Пусть

$$V = V_1 \oplus \dots \oplus V_k$$

— разложение пространства V в прямую сумму циклических подпространств оператора \mathcal{N} . Очевидно, что

$$\text{Ker } \mathcal{N} = \text{Ker } \mathcal{N}|_{V_1} \oplus \dots \oplus \text{Ker } \mathcal{N}|_{V_k}.$$

Так как

$$\dim \text{Ker } \mathcal{N}|_{V_i} = 1$$

при любом i , то $\dim \text{Ker } \mathcal{N} = k$. □

Возвращаясь к произвольному линейному оператору \mathcal{A} , заметим, что в циклическом подпространстве нильпотентного оператора $\mathcal{N} = (\mathcal{A} - \lambda \mathcal{E})|_{V^\lambda(\mathcal{A})}$ оператор \mathcal{A} задается матрицей вида

$$J(\lambda) = J(0) + \lambda E = \begin{pmatrix} \lambda & 1 & 0 & \dots & 0 & 0 \\ 0 & \lambda & 1 & \dots & 0 & 0 \\ 0 & 0 & \lambda & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \lambda & 1 \\ 0 & 0 & 0 & \dots & 0 & \lambda \end{pmatrix}.$$

Такая матрица называется *жордановой клеткой* с собственным значением λ .

Определение 4. *Жордановой матрицей* называется клеточно-диагональная матрица

$$J = \begin{pmatrix} J_1 & & & 0 \\ & J_2 & & \\ & & \ddots & \\ 0 & & & J_n \end{pmatrix}$$

в которой J_1, J_2, \dots, J_k — какие-то жордановы клетки.

Комбинируя теоремы 1 и 2, мы приходим к следующему результату.

Теорема 3. *Если характеристический многочлен $f_{\mathcal{A}}(t)$ разлагается на линейные множители, то существует базис, в котором матрица оператора \mathcal{A} жорданова.*

Следствие. *Матрица любого линейного оператора над полем комплексных чисел приводится к жордановой форме.*

Базис, в котором оператор \mathcal{A} имеет жорданову матрицу, называется *жордановым*. Как видно из доказательства теоремы 2, в его

выборе, вообще говоря, имеется большой произвол. Однако сама жорданова форма матрицы линейного оператора определена однозначно с точностью до перестановки клеток. Это будет доказано в § 9.3.

Очевидно, что в жордановой форме матрицы оператора \mathcal{A} суммарный порядок жордановых клеток с собственным значением λ равен $\dim V^\lambda(\mathcal{A})$, т. е. кратности λ как корня характеристического многочлена. Из второй части теоремы 2 следует, что число жордановых клеток с собственным значением λ равно $\dim V_\lambda(\mathcal{A})$.

Задача 2. Доказать, что максимальный порядок жордановых клеток с собственным значением λ в жордановой форме матрицы оператора \mathcal{A} равен высоте нильпотентного оператора $\mathcal{N} = (\mathcal{A} - \lambda \mathcal{E})|_{V^\lambda(\mathcal{A})}$.

Матрицы A и B называются *подобными*, если существует такая невырожденная матрица C , что $B = C^{-1}AC$. Подобные матрицы можно рассматривать как матрицы одного линейного оператора в разных базисах. Следствие теоремы 3 можно сформулировать таким образом, что всякая комплексная матрица подобна жордановой.

Задача 3. Доказать, что всякая комплексная матрица подобна своей транспонированной матрице.

§ 5. Функции от линейного оператора

Пусть \mathcal{A} — линейный оператор в n -мерном векторном пространстве V над полем K .

Для любого многочлена

$$f(t) = a_0 t^m + a_1 t^{m-1} + \dots + a_{m-1} t + a_m \in K[t]$$

можно определить его значение от оператора \mathcal{A} по формуле

$$f(\mathcal{A}) = a_0 \mathcal{A}^m + a_1 \mathcal{A}^{m-1} + \dots + a_{m-1} \mathcal{A} + a_m \mathcal{E}.$$

Ясно, что

$$(f + g)(\mathcal{A}) = f(\mathcal{A}) + g(\mathcal{A}), \quad (fg)(\mathcal{A}) = f(\mathcal{A})g(\mathcal{A}). \quad (24)$$

Аналогичным образом можно определить многочлен от матрицы. При этом, если оператор \mathcal{A} имеет в некотором базисе матрицу A , то оператор $f(\mathcal{A})$ будет иметь в том же базисе матрицу $f(A)$.

Так как пространство всех линейных операторов конечномерно (при нашем молчаливом предположении, что пространство V конечномерно), то среди степеней оператора \mathcal{A} может быть лишь конечное число линейно независимых. Следовательно, существуют такие ненулевые многочлены f , что $f(\mathcal{A}) = 0$. Они называются *аннулирующими многочленами* оператора \mathcal{A} . Аннулирующий многочлен наименьшей степени называется *минимальным (аннулирующим) многочленом* оператора \mathcal{A} . Мы будем обозначать его через $m_{\mathcal{A}}$.

Всякий аннулирующий многочлен f делится на минимальный. В самом деле, если остаток от деления f на $m_{\mathcal{A}}$ отличен от нуля, то он является аннулирующим многочленом меньшей степени, чем $m_{\mathcal{A}}$, что противоречит определению минимального многочлена. Отсюда, кстати, следует, что минимальный многочлен определен однозначно с точностью до постоянного множителя. Для того чтобы определить его вполне однозначно, будем считать, что его старший коэффициент равен единице.

Задача 1. Найти минимальные многочлены нулевого и тождественного операторов.

Аналогично определяются аннулирующие и минимальный многочлены матрицы. Минимальный многочлен линейного оператора равен минимальному многочлену его матрицы в любом базисе.

Если пространство V разложено в прямую сумму инвариантных подпространств оператора \mathcal{A} , то минимальный многочлен оператора \mathcal{A} равен наименьшему общему кратному минимальных многочленов его ограничений на эти подпространства. Пользуясь этим, легко найти минимальный многочлен линейного оператора по жордановой форме его матрицы (если, конечно, она приводится к жордановой форме). Для этого надо прежде всего найти минимальный многочлен жордановой клетки.

Лемма 1. Минимальный многочлен жордановой клетки порядка m с собственным значением λ равен $(t - \lambda)^m$.

Доказательство. Пусть \mathcal{A} — линейный оператор, задаваемый такой жордановой клеткой. Тогда $\mathcal{N} = \mathcal{A} - \lambda\mathcal{E}$ — нильпотентный оператор высоты m , т. е.

$$(\mathcal{A} - \lambda\mathcal{E})^m = 0, \quad (\mathcal{A} - \lambda\mathcal{E})^{m-1} \neq 0.$$

Это означает, что $(t - \lambda)^m$ — аннулирующий многочлен, но никакой его собственный делитель не является аннулирующим многочленом. Следовательно, $(t - \lambda)^m$ — минимальный многочлен. \square

Пусть теперь \mathcal{A} — произвольный линейный оператор, характеристический многочлен $f_{\mathcal{A}}$ которого разлагается на линейные множители. Пусть $\lambda_1, \dots, \lambda_s$ — все (различные) корни многочлена $f_{\mathcal{A}}$. Из леммы 1 и предшествующего ей замечания следует

Теорема 1. Минимальный многочлен оператора \mathcal{A} равен

$$m_{\mathcal{A}}(t) = \prod_{i=1}^s (t - \lambda_i)^{m_i},$$

где m_i — максимальный порядок жордановых клеток с собственным значением λ_i в жордановой форме матрицы оператора \mathcal{A} .

Следствие 1. Жорданова форма матрицы оператора \mathcal{A} диагональна тогда и только тогда, когда его минимальный многочлен не имеет кратных корней.

Пример 1. Пусть \mathcal{A} — линейный оператор в комплексном векторном пространстве, удовлетворяющий условию $\mathcal{A}^m = \mathcal{E}$ для некоторого натурального m . Тогда многочлен $t^m - 1$ является аннулирующим для оператора \mathcal{A} . Так как он не имеет кратных корней, то минимальный многочлен оператора \mathcal{A} тем более не имеет кратных корней. Следовательно, жорданова форма матрицы оператора \mathcal{A} диагональна. Ясно, что ее диагональные элементы (собственные значения оператора \mathcal{A}) суть какие-то корни m -й степени из 1.

Пример 2. Найдем все линейные операторы \mathcal{A} , удовлетворяющие условию $\mathcal{A}^3 = \mathcal{A}^2$. Это условие означает, что $t^3 - t^2$ является аннулирующим многочленом оператора \mathcal{A} или, что равносильно, минимальный многочлен оператора \mathcal{A} делит $t^3 - t^2 = t^2(t - 1)$. Ввиду теоремы 1 оно выполняется тогда и только тогда, когда жорданова форма матрицы оператора \mathcal{A} состоит только из клеток вида

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad (0), \quad (1).$$

Число клеток каждого вида может быть произвольным (в том числе равным нулю), лишь бы сумма их порядков равнялась n .

Следствие 2. Если жорданова форма оператора \mathcal{A} диагональна, то и жорданова форма его ограничения $\mathcal{A}|_U$ на любое инвариантное подпространство $U \subset V$ диагональна.

Доказательство. Очевидно, что минимальный многочлен $m_{\mathcal{A}}$ оператора \mathcal{A} является, во всяком случае, аннулирующим многочленом для оператора $\mathcal{A}|_U$. Следовательно, минимальный многочлен оператора $\mathcal{A}|_U$ делит многочлен $m_{\mathcal{A}}$, и если последний из этих

многочленов не имеет кратных корней, то и первый обладает этим свойством. \square

Следствие 3 (теорема Гамильтона—Кэли). $f_{\mathcal{A}}(\mathcal{A}) = 0$.

В частности, для линейного оператора \mathcal{A} в двумерном векторном пространстве получаем:

$$\mathcal{A}^2 - (\operatorname{tr} \mathcal{A})\mathcal{A} + (\det \mathcal{A})\mathcal{E} = 0.$$

Конечно, это легко проверить и прямым вычислением (проделайте это!).

Замечание 1. Теорема Гамильтона—Кэли верна и без предположения о том, что характеристический многочлен $f_{\mathcal{A}}$ разлагается на линейные множители. Это можно доказать следующим образом. Как будет показано в § 9.5, существует расширение L поля K , в котором $f_{\mathcal{A}}$ разлагается на линейные множители. Рассматривая матрицу A оператора \mathcal{A} как матрицу с элементами из L , мы можем утверждать в силу предыдущего следствия, что она аннулируется своим характеристическим многочленом; но очевидно, что характеристический многочлен матрицы A не зависит от того, рассматриваем мы ее как матрицу с элементами из K или как матрицу с элементами из L . Этим же способом доказывается, что если минимальный многочлен оператора \mathcal{A} разлагается на линейные множители над K , то и его характеристический многочлен разлагается на линейные множители над K .

Пользуясь теоремой Гамильтона—Кэли, можно свести вычисление любого многочлена f от линейного оператора \mathcal{A} к вычислению многочлена степени $< n$ от этого оператора. А именно, разделим f на $f_{\mathcal{A}}$ с остатком:

$$f = qf_{\mathcal{A}} + p, \quad \deg p < n. \quad (25)$$

Тогда

$$f(\mathcal{A}) = p(\mathcal{A}).$$

Предположим, что $K = \mathbb{R}$ или \mathbb{C} и многочлен $f_{\mathcal{A}}$ разлагается на линейные множители (что всегда имеет место, если $K = \mathbb{C}$). Пусть $\lambda_1, \dots, \lambda_s$ — все его (различные) корни и k_1, \dots, k_s — их кратности, так что

$$k_1 + \dots + k_s = n. \quad (26)$$

Тогда из (25) следует, что

$$f^{(j)}(\lambda_i) = p^{(j)}(\lambda_i) \quad \text{при } i = 1, \dots, s; j = 0, 1, \dots, k_i - 1. \quad (27)$$

(Мы считаем здесь, что $f^{(0)} = f$ для любой функции f .) Равенства (27) однозначно определяют многочлен p , как показывает следующее

Предложение 1. Пусть $\lambda_1, \dots, \lambda_s \in K$ — различные числа и k_1, \dots, k_s — натуральные числа, удовлетворяющие условию (26). Обозначим через P_n пространство многочленов степени $< n$. Тогда отображение $\varphi: P_n \rightarrow K^n$, ставящее в соответствие каждому многочлену $p \in P_n$ набор чисел

$$(p^{(j)}(\lambda_i): i = 1, \dots, s; j = 0, 1, \dots, k_i - 1),$$

является изоморфизмом векторных пространств.

Доказательство. Очевидно, что φ — линейное отображение. Так как $\dim P_n = \dim K^n = n$, то достаточно доказать, что $\text{Ker } \varphi = 0$. Но $\text{Ker } \varphi$ состоит из многочленов, для которых каждое из чисел λ_i является корнем кратности $\geq k_i$, а ненулевой многочлен степени $< n$ не может иметь так много корней (с учетом кратностей). \square

Задача нахождения многочлена p степени $< n$, для которого числа $p^{(j)}(\lambda_i)$ ($i = 1, \dots, s; j = 0, 1, \dots, k_i - 1$) равны каким-то заданным числам, называется задачей интерполяции (с кратными узлами). В случае простых узлов, т. е. когда $k_1 = \dots = k_s = 1$, ответ может быть дан в виде интерполяционной формулы Лагранжа.

Пример 3. Вычислим A^m , где

$$A = \begin{pmatrix} 1 & 0 & -3 \\ 1 & -1 & -6 \\ -1 & 2 & 5 \end{pmatrix}.$$

Имеем

$$f_{\mathcal{A}}(t) = \begin{vmatrix} t-1 & 0 & 3 \\ -1 & t+1 & 6 \\ 1 & -2 & t-5 \end{vmatrix} = t^3 - 5t^2 + 8t - 4 = (t-1)(t-2)^2.$$

Интерполяционный многочлен

$$p(t) = at^2 + bt + c$$

определяется условиями

$$\begin{aligned} p(1) &= a + b + c = 1, \\ p(2) &= 4a + 2b + c = 2^m, \\ p'(2) &= 4a + b = m \cdot 2^{m-1}, \end{aligned}$$

откуда

$$a = (m - 2) \cdot 2^{m-1} + 1,$$

$$b = -(3m - 8) \cdot 2^{m-1} - 4,$$

$$c = (2m - 6) \cdot 2^{m-1} + 4.$$

Следовательно,

$$\begin{aligned} A^m &= 2^{m-1} [(m - 2)A^2 - (3m - 8)A + (2m - 6)E] + A^2 - 4A + 4E = \\ &= 2^{m-1} \begin{pmatrix} 3m - 6 & -6m + 12 & -9m + 12 \\ 3m - 4 & -6m + 8 & -9m + 6 \\ -m & 2m & 3m + 2 \end{pmatrix} + \begin{pmatrix} 4 & -6 & -6 \\ 2 & -3 & -3 \\ 0 & 0 & 0 \end{pmatrix}. \end{aligned}$$

Изложенная теория может быть обобщена с многочленов на произвольные аналитические функции, но для этого мы должны исследовать топологические свойства алгебры линейных операторов.

Пусть V — векторное пространство над полем $K = \mathbb{R}$ или \mathbb{C} .

Определение 1. *Нормой* в пространстве V называется всякая функция $\| \cdot \| : V \rightarrow \mathbb{R}$, обладающая свойствами

- 1) $\|x\| > 0$ при $x \neq 0$;
- 2) $\|\lambda x\| = |\lambda| \|x\|$;
- 3) $\|x + y\| \leq \|x\| + \|y\|$.

Приведем примеры норм в K^n .

Пример 4. $\|x\| = \max_i |x_i|$.

Пример 5. Евклидова (эрмитова) норма $\|x\| = \sqrt{\sum_i |x_i|^2}$.

Пример 6. $\|x\| = \sum_i |x_i|$.

Определение 2. Последовательность векторов x_m называется *сходящейся по норме* к вектору $x \in V$, если $\lim_{m \rightarrow \infty} \|x_m - x\| = 0$.

Легко видеть, что сходимость по любой из приведенных выше норм означает просто покоординатную сходимость. На самом деле это справедливо вообще для всех норм в конечномерном пространстве, как показывает следующее

Предложение 2. Для любых двух норм $\| \cdot \|_1$ и $\| \cdot \|_2$ в конечномерном векторном пространстве V существуют такие положительные константы a и b , что

$$a \leq \frac{\|x\|_2}{\|x\|_1} \leq b \quad \text{при всех } x \in V, x \neq 0.$$

Доказательство. Достаточно сравнить произвольную норму с какой-либо фиксированной. Пусть $\|x\|_1 = \sum_i |x_i|$, где x_1, \dots, x_n — координаты вектора x в базисе $\{e_1, \dots, e_n\}$. Тогда

$$\|x\|_2 \leq \sum_i |x_i| \|e_i\|_2 \leq b \|x\|_1,$$

где $b = \max_i \|e_i\|_2$. Неравенства

$$\left| \|x + \Delta x\|_2 - \|x\|_2 \right| \leq \|\Delta x\|_2 \leq b \|\Delta x\|_1$$

показывают, что $\|\cdot\|_2$ — непрерывная функция в топологии по координатной сходимости. Пусть a — ее минимум на «единичной сфере» $\|x\|_1 = 1$ в смысле первой нормы. Тогда $\|x\|_2 \geq a \|x\|_1$ при всех $x \in V$. \square

Замечание 2. В бесконечномерном пространстве различные нормы, вообще говоря, определяют различные топологии. Проверьте это, например, для норм

$$\|f\|_1 = \int_0^1 |f(x)| dx, \quad \|f\|_2 = \max_{0 \leq x \leq 1} |f(x)|$$

в пространстве непрерывных функций на отрезке $[0, 1]$.

Пусть V — конечномерное векторное пространство с фиксированной нормой $\|\cdot\|$.

Определение 3. Ряд $\sum_{m=1}^{\infty} x_m$ ($x_m \in V$) называется *абсолютно сходящимся*, если числовой ряд $\sum_{m=1}^{\infty} \|x_m\|$ сходится.

Точно так же, как для числовых рядов, доказываются следующие утверждения.

Предложение 3. *Всякий абсолютно сходящийся ряд $\sum_{m=1}^{\infty} x_m$ ($x_m \in V$) сходится, причем*

$$\left\| \sum_{m=1}^{\infty} x_m \right\| \leq \sum_{m=1}^{\infty} \|x_m\|.$$

Предложение 4. *Сумма абсолютно сходящегося ряда не изменяется ни при какой перестановке его членов.*

Определим теперь норму в пространстве линейных операторов на V .

Определение 4. *Нормой линейного оператора \mathcal{A} называется число*

$$\|\mathcal{A}\| = \max_{\|x\|=1} \|\mathcal{A}x\| = \max_{x \neq 0} \frac{\|\mathcal{A}x\|}{\|x\|}.$$

Предложение 5. *Определенная таким образом функция в пространстве линейных операторов действительно является нормой. Кроме того, она обладает свойством*

$$\|\mathcal{A}\mathcal{B}\| \leq \|\mathcal{A}\| \|\mathcal{B}\|.$$

Доказательство. Имеем

$$\begin{aligned} \|\mathcal{A} + \mathcal{B}\| &= \max_{\|x\|=1} \|(\mathcal{A} + \mathcal{B})x\| = \max_{\|x\|=1} \|\mathcal{A}x + \mathcal{B}x\| \leq \\ &\leq \max_{\|x\|=1} (\|\mathcal{A}x\| + \|\mathcal{B}x\|) \leq \max_{\|x\|=1} \|\mathcal{A}x\| + \max_{\|x\|=1} \|\mathcal{B}x\| = \|\mathcal{A}\| + \|\mathcal{B}\|. \end{aligned}$$

Остальные свойства нормы очевидны. Далее,

$$\begin{aligned} \|\mathcal{A}\mathcal{B}\| &= \max_{x \neq 0} \frac{\|\mathcal{A}\mathcal{B}x\|}{\|x\|} = \max_{\mathcal{B}x \neq 0} \frac{\|\mathcal{A}\mathcal{B}x\|}{\|\mathcal{B}x\|} \cdot \frac{\|\mathcal{B}x\|}{\|x\|} \leq \\ &\leq \max_{\mathcal{B}x \neq 0} \frac{\|\mathcal{A}\mathcal{B}x\|}{\|\mathcal{B}x\|} \cdot \max_{\mathcal{B}x \neq 0} \frac{\|\mathcal{B}x\|}{\|x\|} \leq \max_{y \neq 0} \frac{\|\mathcal{A}y\|}{\|y\|} \cdot \max_{x \neq 0} \frac{\|\mathcal{B}x\|}{\|x\|} = \|\mathcal{A}\| \|\mathcal{B}\|. \quad \square \end{aligned}$$

Задача 2. Найти явный вид нормы линейного оператора для каждой из трех приведенных выше норм в пространстве K^n .

Очевидно, что норма линейного оператора не меньше, чем модуль любого его собственного значения.

Теорема 2. Пусть ряд $f(t) = \sum_{m=0}^{\infty} a_m t^m$ ($a_m \in K$) сходится при $|t| < R$. Тогда ряд

$$f(\mathcal{A}) = \sum_{m=0}^{\infty} a_m \mathcal{A}^m \tag{28}$$

абсолютно сходится для любого линейного оператора \mathcal{A} , удовлетворяющего условию $\|\mathcal{A}\| < R$.

Доказательство. Как известно, из сходимости степенного ряда $f(t)$ при $|t| < R$ следует его абсолютная сходимость в том же интервале (круге). Так как

$$\|a_m \mathcal{A}^m\| \leq |a_m| \|\mathcal{A}\|^m,$$

то ряд $f(\mathcal{A})$ абсолютно сходится при $\|\mathcal{A}\| < R$. □

Равенство (28) считается определением функции f от линейного оператора \mathcal{A} . При этом сохраняются свойства (24). Аналогичным образом определяется функция от матрицы. Как и в случае многочленов, если A — матрица оператора \mathcal{A} в каком-либо базисе, то $f(A)$ — матрица оператора $f(\mathcal{A})$ в том же базисе.

Предположим теперь, как и выше, что характеристический многочлен $f_{\mathcal{A}}$ имеет корни $\lambda_1, \dots, \lambda_s$ кратностей k_1, \dots, k_s , причем $k_1 + \dots + k_s = n$. Если $\|\mathcal{A}\| < R$, то $|\lambda_i| < R$ при $i = 1, \dots, s$.

Теорема 3. В условиях теоремы 2 найдем многочлен p степени $< n$, удовлетворяющий условиям (27). Тогда $f(\mathcal{A}) = p(\mathcal{A})$.

Доказательство. Для любого m положим

$$f_m(t) = \sum_{k=0}^m a_k t^k$$

и обозначим через p_m многочлен степени $< n$, удовлетворяющий условиям (27) для многочлена f_m вместо f . Согласно предыдущему, $f_m(\mathcal{A}) = p_m(\mathcal{A})$. Из предложения 1 следует, что $\lim_{m \rightarrow \infty} p_m = p$. Имеем теперь

$$f(\mathcal{A}) = \lim_{m \rightarrow \infty} f_m(\mathcal{A}) = \lim_{m \rightarrow \infty} p_m(\mathcal{A}) = p(\mathcal{A}). \quad \square$$

Согласно сформулированному выше общему принципу, для любого линейного оператора \mathcal{A} можно определить его экспоненту $e^{\mathcal{A}}$ ($= \exp \mathcal{A}$) по формуле

$$e^{\mathcal{A}} = \mathcal{E} + \frac{\mathcal{A}}{1!} + \frac{\mathcal{A}^2}{2!} + \frac{\mathcal{A}^3}{3!} + \dots \quad (29)$$

Как и для чисел, путем перемножения рядов с использованием предложения 4 устанавливается

Теорема 4. $e^{\mathcal{A}+\mathcal{B}} = e^{\mathcal{A}} e^{\mathcal{B}}$ при $\mathcal{A}\mathcal{B} = \mathcal{B}\mathcal{A}$.

(При $\mathcal{A}\mathcal{B} \neq \mathcal{B}\mathcal{A}$ это свойство, как правило, не имеет места, и можно сказать, что лишь благодаря этому обстоятельству существует теория групп Ли.)

При фиксированном \mathcal{A} положим

$$\mathcal{G}(t) = e^{t\mathcal{A}} \quad (t \in K). \quad (30)$$

Очевидно, что $\mathcal{G}(0) = \mathcal{E}$. Из теоремы 4 следует, что

$$\mathcal{G}(t+s) = \mathcal{G}(t)\mathcal{G}(s), \quad \mathcal{G}(-t) = \mathcal{G}(t)^{-1}.$$

Таким образом, операторы $\mathcal{G}(t)$ образуют группу. Она называется *однопараметрической группой*, порожденной оператором \mathcal{A} .

Пример 7. Пусть \mathcal{D} — оператор дифференцирования в пространстве многочленов степени $\leq n$. Тогда

$$(e^{t\mathcal{D}} f)(x) = f(x) + \frac{f'(x)}{1!}t + \frac{f''(x)}{2!}t^2 + \dots = f(x+t).$$

Пример 8. $e^{t \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}} = \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix}$ (проверьте).

Для операторной функции вещественной или комплексной переменной можно обычным образом определить производную. При этом очевидно, что дифференцирование операторной функции сводится к дифференцированию матричных элементов.

Теорема 5. $\mathcal{G}'(t) = \mathcal{G}(t)\mathcal{A} = \mathcal{A}\mathcal{G}(t)$.

Доказательство. Так как

$$\mathcal{G}(t + \Delta t) = \mathcal{G}(t)\mathcal{G}(\Delta t) = \mathcal{G}(\Delta t)\mathcal{G}(t),$$

то

$$\begin{aligned} \mathcal{G}'(t) &= \lim_{\Delta t \rightarrow 0} \frac{\mathcal{G}(t + \Delta t) - \mathcal{G}(t)}{\Delta t} = \mathcal{G}(t) \lim_{\Delta t \rightarrow 0} \frac{\mathcal{G}(\Delta t) - \mathcal{E}}{\Delta t} = \\ &= \left(\lim_{\Delta t \rightarrow 0} \frac{\mathcal{G}(\Delta t) - \mathcal{E}}{\Delta t} \right) \mathcal{G}(t), \end{aligned}$$

и доказательство, как и в случае числовой экспоненты, сводится к выводу «замечательного предела»

$$\lim_{t \rightarrow 0} \frac{e^{t\mathcal{A}} - \mathcal{E}}{t} = \mathcal{A}. \quad (31)$$

Имеем

$$\frac{e^{t\mathcal{A}} - \mathcal{E}}{t} = \mathcal{A} \left[\mathcal{E} + t \left(\frac{\mathcal{A}}{2!} + t \frac{\mathcal{A}^2}{3!} + \dots \right) \right].$$

Ряд, заключенный в круглые скобки, при $|t| < 1$ мажорируется сходящимся числовым рядом

$$\frac{\|\mathcal{A}\|}{2!} + \frac{\|\mathcal{A}\|^2}{3!} + \frac{\|\mathcal{A}\|^3}{4!} + \dots$$

и потому абсолютно сходится, причем его сумма по норме не превосходит суммы указанного числового ряда. Отсюда и следует (31). \square

Теорема 5 позволяет найти в общем виде решение системы однородных линейных дифференциальных уравнений с постоянными коэффициентами

$$x'_i(t) = \sum_{j=1}^n a_{ij}x_j(t) \quad (i = 1, \dots, n). \quad (32)$$

(Здесь $x_1(t), \dots, x_n(t)$ — неизвестные функции переменной t .) Согласно общей теории, система (32) имеет единственное решение, удовлетворяющее начальным условиям вида

$$x_i(0) = x_{i0} \quad (i = 1, \dots, n). \quad (33)$$

Перепишем систему (32) в векторной форме:

$$x'(t) = Ax(t), \quad (34)$$

где $x(t)$ — вектор-столбец с координатами $x_i(t)$, а A — матрица с элементами a_{ij} . Начальное условие (33) можно записать в форме

$$x(0) = x_0, \quad (35)$$

где x_0 — вектор-столбец с координатами x_{i0} . Тогда решением будет

$$x(t) = e^{tA}x_0. \quad (36)$$

Доказательство этого получается непосредственной проверкой с помощью теоремы 5.

Пример 9. Найдем решение системы дифференциальных уравнений

$$\begin{cases} x'_1(t) = x_1(t) - 3x_3(t), \\ x'_2(t) = x_1(t) - x_2(t) - 6x_3(t), \\ x'_3(t) = -x_1(t) + 2x_2(t) + 5x_3(t), \end{cases}$$

удовлетворяющее начальным условиям

$$x_1(0) = 1, \quad x_2(0) = 1, \quad x_3(0) = 0.$$

Матрица A этой системы совпадает с матрицей примера 3. Мы должны вычислить $f(A)$, где $f(u) = e^{tu}$ (здесь t выступает как константа). Интерполяционный многочлен $p(u) = au^2 + bu + c$ определяется условиями

$$\begin{aligned} p(1) &= a + b + c = e^t, \\ p(2) &= 4a + 2b + c = e^{2t}, \\ p'(2) &= 4a + b = te^{2t}, \end{aligned}$$

откуда

$$\begin{aligned} a &= (t-1)e^{2t} + e^t, \\ b &= -(3t-4)e^{2t} - 4e^t, \\ c &= (2t-3)e^{2t} + 4e^t. \end{aligned}$$

Следовательно,

$$\begin{aligned} e^{tA} &= e^{2t}[(t-1)A^2 - (3t-4)A + (2t-3)E] + e^t(A^2 - 4A + 4E) = \\ &= e^{2t} \begin{pmatrix} 3t-3 & -6t+6 & -9t+6 \\ 3t-2 & -6t+4 & -9t+3 \\ -t & 2t & 3t+1 \end{pmatrix} + e^t \begin{pmatrix} 4 & -6 & -6 \\ 2 & -3 & -3 \\ 0 & 0 & 0 \end{pmatrix}. \end{aligned}$$

Решение, удовлетворяющее заданным начальным условиям, получается умножением матрицы e^{tA} на столбец $\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$. Таким образом находим

$$x_1(t) = (-3t+3)e^{2t} - 2e^t,$$

$$x_2(t) = (-3t+2)e^{2t} - e^t,$$

$$x_3(t) = te^{2t}.$$

Аффинные и проективные пространства

§ 1. Аффинные пространства

В элементарной геометрии мы имеем дело не только с векторами, но и с точками (и даже главным образом с точками). Подобно тому как аксиоматика векторного пространства отражает в обобщенном виде свойства векторов элементарной геометрии, аксиоматика аффинного пространства отражает свойства точек и векторов элементарной геометрии в их взаимосвязи.

В «обычном» евклидовом пространстве элементарной геометрии можно определить операцию сложения точки и вектора. А именно, суммой точки p и вектора x называется точка, являющаяся концом вектора, равного x , отложенного от точки p . Свойства этой операции и лежат в основе следующего определения.

Пусть V — векторное пространство над полем K .

Определение 1. *Аффинным пространством*, ассоциированным с векторным пространством V , называется множество S вместе с операцией сложения $S \times V \rightarrow S$, удовлетворяющей следующим условиям:

1) $p + (x + y) = (p + x) + y$ ($p \in S, x, y \in V$);

2) $p + 0 = p$ ($p \in S, 0$ — нулевой вектор);

3) для любых $p, q \in S$ существует единственный вектор x , такой, что $p + x = q$.

Элементы множества S называются *точками*. Вектор x из условия 3) называется *вектором, соединяющим точки p и q* , и обозначается через \overline{pq} . Из условия 1) следует, что

$$\overline{pq} + \overline{qr} = \overline{pr} \quad \forall p, q, r \in S.$$

Всякое векторное пространство V можно рассматривать как аффинное, считая, что точки — это те же векторы, и определив операцию сложения точки и вектора как сложение векторов. При этом вектор \overline{pq} будет разностью векторов q и p .

С другой стороны, если в аффинном пространстве S фиксировать некоторую точку o — «начало отсчета», то можно отождествить

каждую точку p с ее радиус-вектором \overline{op} . При этом сложение точки и вектора превратится просто в сложение векторов. Такое отождествление точек с векторами называется *векторизацией* аффинного пространства. (Конечно, оно зависит от начала отсчета.)

Размерностью аффинного пространства по определению считается размерность соответствующего векторного пространства.

Точка o (начало отсчета) вместе с базисом $\{e_1, \dots, e_n\}$ пространства V называется *репером* аффинного пространства S . С каждым репером связана *аффинная система координат* в пространстве S . А именно, каждой точке $p \in S$ приписываются координаты, равные координатам вектора \overline{op} в базисе $\{e_1, \dots, e_n\}$. Легко видеть, что

- 1) координаты точки $p + x$ равны суммам соответствующих координат точки p и вектора x ;
- 2) координаты вектора \overline{pq} равны разностям соответствующих координат точек q и p .

Основными объектами элементарной геометрии являются прямые и плоскости. Следующее определение вводит соответствующие понятия в геометрию аффинных пространств.

Определение 2. *Плоскостью* в аффинном пространстве S называется подмножество вида

$$P = p_0 + U, \quad (1)$$

где p_0 — некоторая точка, а U — подпространство пространства V .

Подпространство U однозначно определяется как совокупность всех векторов, соединяющих точки плоскости P , и называется *направляющим подпространством* плоскости P . Сумма точки из P и вектора из U принадлежит P . Относительно этой операции плоскость P является аффинным пространством, ассоциированным с векторным пространством U .

По определению $\dim P = \dim U$. Нульмерная плоскость есть точка. Одномерная плоскость называется *прямой*. Плоскость размерности $n - 1$ называется *гиперплоскостью*.

В качестве точки p_0 в равенстве (1), определяющем плоскость P , может быть взята любая точка этой плоскости.

Очевидно, что пересечение плоскостей, если оно не пусто, также является плоскостью.

Для любого подмножества $M \subset S$ и любой точки $p_0 \in M$ плоскость

$$p_0 + \langle \overline{p_0 p} : p \in M \rangle$$

является наименьшей плоскостью, содержащей M . Эта плоскость называется *аффинной оболочкой* множества M и обозначается через $\text{aff } M$.

Теорема 1. *Через любые $k + 1$ точек аффинного пространства проходит плоскость размерности $\leq k$; при этом, если эти точки не содержатся в плоскости размерности $< k$, через них проходит единственная плоскость размерности k .*

Доказательство. Пусть $p_0, p_1, \dots, p_k \in S$. Тогда

$$P = p_0 + \langle \overline{p_0 p_1}, \dots, \overline{p_0 p_k} \rangle$$

есть плоскость размерности $\leq k$, проходящая через p_0, p_1, \dots, p_k . Если $\dim P = k$, то векторы $\overline{p_0 p_1}, \dots, \overline{p_0 p_k}$ линейно независимы и P является единственной k -мерной плоскостью, проходящей через p_0, p_1, \dots, p_k . \square

Точки $p_0, p_1, \dots, p_k \in S$ называются *аффинно зависимыми*, если они лежат в плоскости размерности $< k$, и *аффинно независимыми* в противном случае. Из доказательства теоремы 1 видно, что точки p_0, p_1, \dots, p_k аффинно зависимы тогда и только тогда, когда векторы $\overline{p_0 p_1}, \dots, \overline{p_0 p_k}$ линейно зависимы. В то же время из определения ясно, что свойство точек быть аффинно зависимыми или независимыми не зависит от их нумерации (в частности, от того, какую из них мы возьмем за p_0).

Другая точка зрения на плоскости состоит в том, что это множества решений систем линейных уравнений.

Пусть дана система линейных уравнений

$$\sum_{j=1}^n a_{ij} x_j = b_i \quad (i = 1, \dots, m). \quad (2)$$

Будем интерпретировать x_1, \dots, x_n как координаты точек n -мерного аффинного пространства S относительно некоторого репера $(o; e_1, \dots, e_n)$. Тогда решения системы (2) можно понимать как точки пространства S . Предположим, что эта система совместна и $p_0 \in S$ — одно из ее решений. Тогда, согласно теореме 2.1.3, множество всех решений системы (2) имеет вид $p_0 + U$, где $U \subset V$ — подпространство решений соответствующей системы однородных линейных уравнений, и, стало быть, является плоскостью пространства S .

Обратно, пусть $P = p_0 + U$ — некоторая плоскость. Согласно теореме 5.3.4, подпространство U может быть задано системой однородных линейных уравнений. Заменив свободные члены этих урав-

нений значениями, принимаемыми левыми частями в точке p_0 , мы получим систему линейных уравнений, задающую плоскость P . Тем самым доказана

Теорема 2. *Всякая плоскость есть множество решений некоторой системы линейных уравнений.*

Обсудим взаимное расположение двух плоскостей

$$P_1 = p_1 + U_1, \quad P_2 = p_2 + U_2.$$

Очевидно, что если они пересекаются и p_0 — одна из точек пересечения, то

$$P_1 \cap P_2 = p_0 + (U_1 \cap U_2).$$

Теорема 3. *Плоскости P_1 и P_2 пересекаются тогда и только тогда, когда*

$$\overline{p_1 p_2} \in U_1 + U_2.$$

Доказательство. Плоскости P_1 и P_2 пересекаются тогда и только тогда, когда существуют такие векторы $u_1 \in U_1$, $u_2 \in U_2$, что

$$p_1 + u_1 = p_2 + u_2.$$

Это равенство может быть переписано в виде

$$\overline{p_1 p_2} = u_1 - u_2.$$

Поэтому существование таких векторов u_1, u_2 как раз и означает, что $\overline{p_1 p_2} \in U_1 + U_2$. \square

Плоскости P_1 и P_2 называются *параллельными*, если $U_1 \subset U_2$ или $U_2 \subset U_1$, и *скрещивающимися*, если $P_1 \cap P_2 = \emptyset$ и $U_1 \cap U_2 = 0$.

Задача 1. Какова наименьшая размерность пространства, в котором существуют скрещивающиеся двумерные плоскости?

Задача 2. Определить $\dim \text{aff}(P_1 \cup P_2)$.

Линейные комбинации точек аффинного пространства, вообще говоря, не определены. Однако некоторым из них можно придать смысл. А именно, назовем *барицентрической линейной комбинацией* точек $p_1, \dots, p_k \in S$ линейную комбинацию вида $\sum_i \lambda_i p_i$, где $\sum_i \lambda_i = 1$, и будем считать ее равной точке p , определяемой равенством

$$\overline{op} = \sum_i \lambda_i \overline{op_i},$$

где $o \in S$. Благодаря условию $\sum_i \lambda_i = 1$ это определение не зависит от выбора точки o . Действительно, пусть o' — любая другая точка. Тогда

$$\overline{o'p} = \overline{o'o} + \overline{op} = \sum_i \lambda_i (\overline{o'o} + \overline{op}_i) = \sum_i \lambda_i \overline{o'p}_i.$$

В частности, *центр тяжести* системы точек $\{p_1, \dots, p_k\}$ можно определить как

$$\text{center}(p_1, \dots, p_k) = \frac{1}{k}(p_1 + \dots + p_k).$$

Задача 3. Показать, что на обычной евклидовой плоскости

- а) центр тяжести точек p, q есть середина отрезка pq ;
- б) центр тяжести множества вершин треугольника есть точка пересечения его медиан.

Барицентрическая комбинация $\lambda p + \mu q$ двух точек p и q есть точка r , лежащая на прямой pq и обладающая тем свойством, что

$$\overline{pr} = \frac{\mu}{\lambda} \overline{rq} \quad (3)$$

(если $\lambda = 0, \mu = 1$, то $r = q$). В самом деле, приняв точку r за точку o в данном выше определении барицентрической линейной комбинации, мы получаем:

$$0 = \lambda \overline{rp} + \mu \overline{rq},$$

откуда и следует (3).

В случае обычной евклидовой плоскости точка $r = \lambda p + \mu q$ делит отрезок pq в отношении $\mu : \lambda$ (она лежит на самом отрезке, если $\lambda, \mu \geq 0$, и на его продолжении, если $\lambda < 0$ или $\mu < 0$). Допуская вольность речи, мы будем употреблять это выражение и в общем случае, когда понятие отрезка не имеет смысла.

Задача 4. Аффинная оболочка множества $M \subset S$ есть совокупность всех барицентрических линейных комбинаций точек из M .

Предположим, что поле K содержит более двух элементов.

Теорема 4. *Непустое подмножество $P \subset S$ является плоскостью тогда и только тогда, когда вместе с любыми двумя различными точками оно содержит проходящую через них прямую.*

Доказательство. Утверждение «только тогда» очевидно. Пусть теперь $P \subset S$ — непустое подмножество, обладающее указанным свойством. Фиксируем произвольную точку $p_0 \in P$ и рассмотрим подмножество $U = \{u \in V : p_0 + u \in P\} \subset V$. Нам нужно доказать, что U — подпространство. Ясно, что

оно содержит 0. Далее, если $u \in U$ — любой ненулевой вектор и $\lambda \in K$, то точка $p_0 + \lambda u$ лежит на прямой, проходящей через p_0 и $p_0 + u$ и, следовательно, $\lambda u \in U$. Докажем, наконец, что если $u_1, u_2 \in U$ — непропорциональные векторы, то $u_1 + u_2 \in U$. Пусть λ — любой элемент поля K , отличный от 0 и 1. Легко видеть, что точка $p = p_0 + u_1 + u_2$ лежит на прямой, проходящей через точки $p_1 = p_0 + \lambda u_1 \in P$ и $p_2 = p_0 + \frac{\lambda}{\lambda-1} u_2 \in P$, а именно

$$p = \frac{1}{\lambda} p_1 + \frac{\lambda-1}{\lambda} p_2.$$

Следовательно, $p \in P$. □

Замечание 1. Если поле K состоит из двух элементов, то любая прямая пространства S состоит всего из двух точек и, таким образом, любое непустое подмножество $P \subset S$ удовлетворяет условию теоремы (но не любое непустое подмножество является плоскостью).

Пусть p_0, p_1, \dots, p_n — такие точки n -мерного аффинного пространства S , что векторы $\overline{p_0 p_1}, \dots, \overline{p_0 p_n}$ линейно независимы. Тогда каждая точка $p \in S$ единственным образом представляется в виде

$$p = \sum_{i=0}^n x_i p_i, \quad \text{где} \quad \sum_{i=0}^n x_i = 1.$$

В самом деле, это равенство можно переписать в виде

$$\overline{p_0 p} = \sum_{i=1}^n x_i \overline{p_0 p_i},$$

откуда следует, что в качестве x_1, \dots, x_n можно (и должно) взять координаты вектора $\overline{p_0 p}$ в базисе $\{\overline{p_0 p_1}, \dots, \overline{p_0 p_n}\}$; после этого x_0 определяется равенством $x_0 = 1 - \sum_{i=1}^n x_i$.

Числа x_0, x_1, \dots, x_n называются *барицентрическими координатами* точки p относительно p_0, p_1, \dots, p_n .

Пример 1. Пусть точки u, v, w , лежащие на сторонах qr, rp, pq треугольника pqr (см. рис. 1), делят эти стороны в отношениях $\lambda : 1, \mu : 1, \nu : 1$ соответственно. Докажем *теорему Чевы*: прямые pu, qv, rw пересекаются в одной точке тогда и только тогда, когда

$$\lambda \mu \nu = 1.$$

Для этого рассмотрим барицентрические координаты относительно точек p, q, r . Барицентрические координаты точек u, v, w суть

$$\left(0, \frac{1}{1+\lambda}, \frac{\lambda}{1+\lambda}\right), \left(\frac{\mu}{1+\mu}, 0, \frac{1}{1+\mu}\right), \left(\frac{1}{1+\nu}, \frac{\nu}{1+\nu}, 0\right)$$

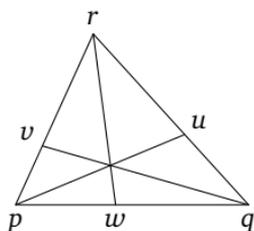


Рис. 1

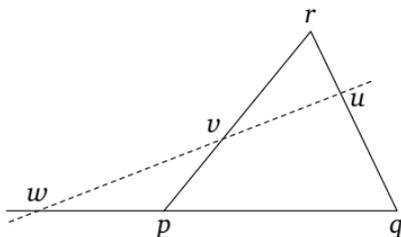


Рис. 2

соответственно. Точки прямой ru являются барицентрическими комбинациями точек p и u . Их барицентрические координаты x, y, z характеризуются тем, что $z : y = \lambda$. Аналогично, барицентрические координаты точек прямой qv характеризуются тем, что $x : z = \mu$, а барицентрические координаты точек прямой rw — тем, что $y : x = \nu$. Точка, барицентрические координаты которой удовлетворяют всем этим трем условиям, существует тогда и только тогда, когда $\lambda\mu\nu = 1$.

Задача 5. Точки p_0, p_1, \dots, p_k аффинно независимы тогда и только тогда, когда ранг матрицы, составленной из их барицентрических координат, равен $k + 1$.

Задача 6. Пусть точки u, v, w , лежащие на сторонах qr, rp, pq треугольника pqr или их продолжениях (см. рис. 2), делят эти стороны в отношениях $\lambda : 1, \mu : 1, \nu : 1$ соответственно. Используя предыдущую задачу, доказать *теорему Менелая*: точки u, v, w лежат на одной прямой тогда и только тогда, когда

$$\lambda\mu\nu = -1.$$

§ 2. Аффинные отображения

Пусть S и S' — аффинные пространства, ассоциированные с векторными пространствами V и V' соответственно (над одним и тем же полем).

Определение 1. *Аффинным отображением* пространства S в пространство S' называется всякое отображение $f : S \rightarrow S'$, обладающее свойством

$$f(p + x) = f(p) + \varphi(x) \quad (p \in S, x \in V), \quad (4)$$

где φ — некоторое линейное отображение пространства V в пространство V' .

Из (4) вытекает, что

$$\varphi(\overline{pq}) = \overline{f(p)f(q)} \quad (p, q \in S). \quad (5)$$

Тем самым линейное отображение φ однозначно определяется по f . Оно называется *дифференциалом* отображения f и обозначается через df .

Векторизуем пространства S и S' , приняв за начала отсчета какие-то точки o и o' соответственно. Полагая в (4) $p = o$, мы получаем следующее представление аффинного отображения f в векторизованной форме:

$$f(x) = \varphi(x) + b \quad (b \in V'), \quad (6)$$

где $b = \overline{o'f(o)}$. Отсюда, в свою очередь, получается запись отображения f в координатах:

$$y_i = \sum_{j=1}^n a_{ij}x_j + b_i \quad (i = 1, \dots, m), \quad (7)$$

где x_1, \dots, x_n — координаты точки x , а y_1, \dots, y_m — координаты точки $y = f(x)$.

Обратно, как легко проверить, для любого линейного отображения $\varphi: V \rightarrow V'$ и любого вектора $b \in V'$ отображение, определяемое формулой (6), аффинно и его дифференциал равен φ .

Пусть S'' — еще одно аффинное пространство и $g: S' \rightarrow S''$ — аффинное отображение.

Предложение 1. *Отображение $gf: S \rightarrow S''$ является аффинным, причем*

$$d(gf) = dg \cdot df. \quad (8)$$

Доказательство. При $p \in S$, $x \in V$ имеем

$$\begin{aligned} (gf)(p+x) &= g(f(p+x)) = g(f(p) + df(x)) = \\ &= g(f(p)) + dg(df(x)) = (gf)(p) + (dg \cdot df)(x). \quad \square \end{aligned}$$

При $K = \mathbb{R}$ дифференциал аффинного отображения есть частный случай дифференциала произвольного гладкого отображения, рассматриваемого в анализе, а формула (8) есть частный случай формулы для дифференциала произведения гладких отображений (или «сложной функции»).

Предложение 2. *Аффинное отображение биективно тогда и только тогда, когда его дифференциал биективен.*

Доказательство. Выберем начала отсчета o и o' в пространствах S и S' таким образом, чтобы $f(o) = o'$. Тогда отображение f в векторизованной форме будет совпадать со своим дифференциалом, откуда и следует доказываемое утверждение. \square

Биективное аффинное отображение называется *изоморфизмом* аффинных пространств. Аффинные пространства называются *изоморфными*, если между ними существует изоморфизм.

Следствие. *Конечномерные аффинные пространства (над одним и тем же полем) изоморфны тогда и только тогда, когда они имеют одинаковую размерность.*

Очевидно, что при аффинном отображении $f: S \rightarrow S'$ всякая плоскость $P = p + U$ пространства S переходит в плоскость $f(P) = f(p) + df(U)$ пространства S' . Если f биективно, то $\dim f(P) = \dim P$.

Предложение 3. Пусть $f: S \rightarrow S'$ — аффинное отображение. Тогда

$$f\left(\sum_i \lambda_i p_i\right) = \sum_i \lambda_i f(p_i)$$

для любой барицентрической линейной комбинации $\sum_i \lambda_i p_i$ точек p_1, \dots, p_k .

Доказательство. Векторизуем пространство S . Тогда f запишется в виде (6), и мы получим

$$f\left(\sum_i \lambda_i p_i\right) = \varphi\left(\sum_i \lambda_i p_i\right) + b = \sum_i \lambda_i (\varphi(p_i) + b) = \sum_i \lambda_i f(p_i). \quad \square$$

В частности, центр тяжести системы точек при аффинном отображении переходит в центр тяжести системы их образов.

Частным случаем аффинных отображений являются аффинно-линейные функции.

Определение 2. *Аффинно-линейной функцией* на аффинном пространстве S называется всякая функция $f: S \rightarrow K$, обладающая свойством

$$f(p + x) = f(p) + \alpha(x) \quad (p \in S, x \in V), \quad (9)$$

где α — некоторая линейная функция на векторном пространстве V , называемая *дифференциалом* функции f и обозначаемая через df .

Иначе говоря, аффинно-линейная функция — это аффинное отображение пространства S в поле K , рассматриваемое как аффинная прямая.

В векторизованной форме с началом в точке o аффинно-линейная функция f записывается в виде

$$f(x) = \alpha(x) + b \quad (b \in K), \quad (10)$$

где $b = f(o)$. Отсюда, в свою очередь, получается запись функции f в координатах:

$$f(x) = \sum_i a_i x_i + b. \quad (11)$$

Частным случаем аффинно-линейных функций являются постоянные функции. Они характеризуются тем, что их дифференциал равен нулю. Если f — непостоянная аффинно-линейная функция, то ее многообразия уровня $f(p) = c$ суть параллельные гиперплоскости с направляющим подпространством, задаваемым уравнением $df(x) = 0$.

Аффинно-линейные функции образуют $(n + 1)$ -мерное подпространство (где $n = \dim S$) в пространстве всех функций на S . Это ясно хотя бы из их координатной записи (11).

Докажем одно утверждение, которое нам понадобится в следующем параграфе.

Предложение 4. *Барицентрические координаты суть аффинно-линейные функции.*

Доказательство. Пусть x_0, x_1, \dots, x_n — барицентрические координаты относительно точек p_0, p_1, \dots, p_n . Если векторизовать пространство S , приняв точку p_0 за начало отсчета, то x_1, \dots, x_n будут обычными координатами относительно базиса $(\overline{p_0 p_1}, \dots, \overline{p_0 p_n})$. Следовательно, x_1, \dots, x_n — аффинно-линейные функции. Так как

$x_0 = 1 - \sum_{i=1}^n x_i$, то x_0 — также аффинно-линейная функция. (Это мож-

но было бы также доказать, приняв за начало отсчета какую-нибудь другую из точек p_i .) \square

Аффинное отображение аффинного пространства S в себя называется *аффинным преобразованием*. Биективные аффинные преобразования образуют группу, называемую *полной аффинной группой* пространства S и обозначаемую через $GA(S)$. (Это согласуется с тем определением, которое было дано в § 4.2 в векторной форме.)

В силу предложения 1 отображение

$$d: GA(S) \rightarrow GL(V)$$

является гомоморфизмом групп. Его ядро есть группа параллельных переносов

$$t_a: p \mapsto p + a \quad (a \in V).$$

Обозначим ее через $\text{Trans}(S)$.

Предложение 5. Для любых $f \in \text{GA}(S)$ и $a \in V$ имеем

$$ft_a f^{-1} = t_{df(a)}. \quad (12)$$

Доказательство. Применяя преобразование $ft_a f^{-1}$ к точке $q = f(p)$, получаем

$$ft_a f^{-1}(q) = ft_a(p) = f(p + a) = f(p) + df(a) = q + df(a). \quad \square$$

Тот факт, что преобразование $ft_a f^{-1}$ является параллельным переносом, можно было бы доказать, вычислив его дифференциал при помощи предложения 1.

Если фиксировано начало отсчета $o \in S$ и тем самым аффинное пространство S отождествлено с векторным пространством V , то группа $\text{GL}(V)$ становится подгруппой группы $\text{GA}(S)$. Это не что иное, как стабилизатор точки o в группе $\text{GA}(S)$. Из записи аффинных преобразований в векторизованной форме (6) следует, что всякое аффинное преобразование $f \in \text{GA}(S)$ единственным образом представляется в виде

$$f = t_b \varphi \quad (\varphi \in \text{GL}(V), b \in V). \quad (13)$$

Ясно, что $\varphi = df$ не зависит от выбора начала отсчета, но вектор $b = of(o)$ от этого, вообще говоря, зависит.

Задача 1. Доказать, что при переходе к началу отсчета $o' = o + a$ ($a \in V$) вектор b заменяется на вектор

$$b' = b + \varphi(a) - a. \quad (14)$$

Пример 1. Согласно предложению 4.2.2, всякое движение евклидовой плоскости E^2 является (биективным) аффинным преобразованием. То же верно для евклидова пространства E^3 .

Пример 2. Гомотетия с центром в точке o и коэффициентом λ есть аффинное преобразование, задаваемое формулой

$$f(o + x) = o + \lambda x.$$

Ясно, что $df = \lambda \mathcal{E}$. Докажем, что всякое аффинное преобразование f , для которого $df = \lambda \mathcal{E}$ ($\lambda \neq 1$), есть гомотетия с центром в некоторой

точке. Для этого достаточно доказать, что f имеет неподвижную точку. Запишем f в векторизованной форме:

$$f(x) = \lambda x + b \quad (b \in V).$$

Уравнение $f(x) = x$ приводится к виду $(1 - \lambda)x = b$ и, следовательно, имеет (единственное) решение.

Гомотетия с коэффициентом -1 называется *центральной симметрией*.

Задача 2. Доказать, что произведение гомотетий с центрами в разных точках с коэффициентами λ и μ при $\lambda\mu \neq 1$ есть гомотетия, а при $\lambda\mu = 1$ — нетривиальный параллельный перенос.

Группа аффинных преобразований определяет *аффинную геометрию* в том смысле, что задачей аффинной геометрии является изучение свойств фигур, инвариантных при (биективных) аффинных преобразованиях. Так как при таких преобразованиях любая плоскость переходит в плоскость той же размерности, а любая барицентрическая линейная комбинация точек — в барицентрическую линейную комбинацию их образов с теми же коэффициентами, то понятия плоскости и барицентрической комбинации точек, а следовательно, понятия параллельных прямых, параллелограмма, отрезка, середины отрезка, центра тяжести системы точек, выпуклого множества, симплекса и т. д. относятся к числу понятий аффинной геометрии. Но, например, понятия квадрата и окружности к числу таковых не относятся, так как при аффинном преобразовании квадрат может перейти в параллелограмм, не являющийся квадратом, а окружность — в эллипс, не являющийся окружностью.

Следующая теорема показывает, что в аффинной геометрии все симплексы равны (например, на аффинной плоскости все треугольники равны).

Теорема 1. Пусть $\{p_0, p_1, \dots, p_n\}$ и $\{q_0, q_1, \dots, q_n\}$ — две системы аффинно независимых точек в n -мерном аффинном пространстве S . Тогда существует единственное аффинное преобразование f , переводящее p_i в q_i при $i = 0, 1, \dots, n$.

Доказательство. Существует единственное линейное преобразование φ пространства V , переводящее базис $\{\overline{p_0p_1}, \dots, \overline{p_0p_n}\}$ в базис $\{\overline{q_0q_1}, \dots, \overline{q_0q_n}\}$. Векторизуем пространство S , приняв за начало отсчета точку p_0 . Тогда искомое аффинное преобразование f записывается в виде

$$f(x) = \varphi(x) + \overline{p_0q_0}.$$

□

Задача 3. Доказать, что в вещественной аффинной геометрии все параллелепипеды равны.

Задача 4. Пусть $P_1, P_2, P'_1, P'_2 \subset S$ — плоскости с направляющими подпространствами U_1, U_2, U'_1, U'_2 соответственно. Предположим, что $\dim P_1 = \dim P'_1$, $\dim P_2 = \dim P'_2$, $\dim U_1 \cap U_2 = \dim U'_1 \cap U'_2$ и пересечения $P_1 \cap P_2$ и $P'_1 \cap P'_2$ пусты или непусты одновременно. Доказать, что тогда существует преобразование $f \in GA(S)$, переводящее P_1 в P'_1 и P_2 в P'_2 .

В аффинной геометрии не существует понятия расстояния между точками, так как любую пару различных точек с помощью аффинного преобразования можно перевести в любую другую такую пару. Однако при аффинных преобразованиях сохраняется так называемое отношение тройки точек, лежащих на одной прямой.

Пусть точки p_1, p_2, p_3 лежат на одной прямой l . Тогда если $p_2 \neq p_3$, то $\overline{p_1 p_3} = c \overline{p_3 p_2}$ ($c \in K$). Число c и называется (*простым*) *отношением* тройки точек p_1, p_2, p_3 и обозначается через (p_1, p_2, p_3) . Если $p_1 \neq p_2 = p_3$, то полагают $(p_1, p_2, p_3) = \infty$. Если $p_1 = p_2 = p_3$, то (p_1, p_2, p_3) не определено. Если $c = \frac{\lambda}{\mu}$, то говорят также, что точка p_3 делит отрезок $p_1 p_2$ в отношении $\lambda : \mu$ (хотя само понятие отрезка определено только в вещественной геометрии). При $\lambda + \mu = 1$ это означает, что

$$p_3 = \mu p_1 + \lambda p_2.$$

Ясно, что отношение точек p_1, p_2, p_3 сохраняется при любом аффинном преобразовании, не стягивающем прямую l в точку (в частности, при любом биективном аффинном преобразовании).

Задача 5. Выяснить, как изменяется отношение тройки точек при перестановках этих точек. Какое наибольшее и какое наименьшее число различных значений оно может принимать?

Задача 6. Построить треугольник pqr по точкам u, v, w на его сторонах qr, rp, pq (или их продолжениях), делящих их в отношениях $\lambda : 1, \mu : 1, \nu : 1$ соответственно (рис. 3). (Указание: рассмотреть произведение гомотетий с центрами в точках u, v, w , переводящих r в q, p в r, q в p соответственно.)

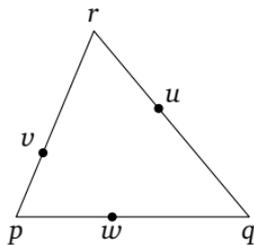


Рис. 3

§ 3. Выпуклые множества

Пусть S — аффинное пространство, ассоциированное с векторным пространством V над полем вещественных чисел.

Определение 1. Отрезком, соединяющим точки $p, q \in S$, называется множество

$$pq = \{\lambda p + (1 - \lambda)q : 0 \leq \lambda \leq 1\} \subset S.$$

Определение 2. Множество $M \subset S$ называется *выпуклым*, если вместе с любыми двумя точками оно содержит соединяющий их отрезок.

Очевидно, что пересечение выпуклых множеств выпукло. Любая плоскость является выпуклым множеством.

Определение 3. *Выпуклой линейной комбинацией* точек пространства S называется их барицентрическая линейная комбинация с неотрицательными коэффициентами.

Предложение 1. *Выпуклое множество $M \subset S$ вместе с любыми точками p_0, p_1, \dots, p_k содержит любую их выпуклую линейную комбинацию $p = \sum_i \lambda_i p_i$.*

Доказательство. Докажем, что $p \in M$, индукцией по числу l коэффициентов $\lambda_0, \lambda_1, \dots, \lambda_k$, отличных от нуля. При $l = 1$ точка p совпадает с одной из точек p_0, p_1, \dots, p_k , так что доказывать нечего. Пусть $l > 1$. Будем считать для определенности, что $\lambda_k \neq 0$. Заметим, что $\lambda_k \neq 1$, так как иначе все остальные коэффициенты были бы равны нулю. Имеем:

$$p = (1 - \lambda_k) \left(\sum_{i=0}^{k-1} \frac{\lambda_i}{1 - \lambda_k} p_i \right) + \lambda_k p_k,$$

т. е. p лежит на отрезке $p'p_k$, где

$$p' = \sum_{i=0}^{k-1} \frac{\lambda_i}{1 - \lambda_k} p_i.$$

По предположению индукции $p' \in M$. Следовательно, и $p \in M$. \square

Предложение 2. *Для любого множества $M \subset S$ множество $\text{conv } M$ всевозможных выпуклых линейных комбинаций точек из M является выпуклым множеством.*

Доказательство. Пусть $p = \sum_i \lambda_i p_i$ и $q = \sum_i \mu_i q_i$ — выпуклые линейные комбинации точек из M . Тогда при $0 \leq \lambda \leq 1$

$$\lambda p + (1 - \lambda)q = \sum_i \lambda \lambda_i p_i + \sum_i (1 - \lambda) \mu_i q_i$$

есть также выпуклая линейная комбинация точек из M . □

Множество $\text{conv } M$ является наименьшим выпуклым множеством, содержащим M ; оно называется *выпуклой оболочкой* множества M .

Выпуклая оболочка системы $\{p_0, p_1, \dots, p_n\}$ аффинно независимых точек n -мерного пространства называется *n -мерным симплексом* с вершинами в точках p_0, p_1, \dots, p_n . Иными словами, симплекс состоит из точек, барицентрические координаты которых относительно p_0, p_1, \dots, p_n неотрицательны. Нульмерный симплекс — это точка, одномерный симплекс — отрезок, двумерный — треугольник, трехмерный — тетраэдр (треугольная пирамида).

Точка множества $M \subset S$ называется *внутренней*, если некоторая ее окрестность целиком содержится в M , и *граничной* в противном случае. Очевидно, что точки симплекса, барицентрические координаты которых относительно вершин симплекса положительны (и только они), являются внутренними.

Предложение 3. *Выпуклое множество M содержит внутренние точки тогда и только тогда, когда $\text{aff } M = S$.*

Доказательство. Если $\text{aff } M = S$, то M содержит систему из $n + 1$ аффинно независимых точек. Но тогда M содержит симплекс с вершинами в этих точках и, значит, содержит внутренние точки. Обратное очевидно. □

Задача 1. Доказать, что замыкание выпуклого множества выпукло, причем всякая его внутренняя точка является внутренней точкой самого множества.

Выпуклое множество, содержащее внутренние точки, называется *выпуклым телом*.

Предложение 4. *Пусть p — внутренняя точка выпуклого тела M и q — любая его точка. Тогда все точки отрезка pq , за исключением, быть может, точки q , являются внутренними точками тела M .*

Доказательство. Рассмотрим точку

$$r = \lambda p + (1 - \lambda)q \quad (0 < \lambda \leq 1).$$

Имеем

$$p = \frac{1}{\lambda}r + \frac{\lambda-1}{\lambda}q.$$

Если точка r' достаточно близка к r , то точка

$$p' = \frac{1}{\lambda}r' + \frac{\lambda-1}{\lambda}q$$

близка к точке p и, следовательно, лежит в M (см. рис. 4). Так как

$$r' = \lambda p' + (1-\lambda)q,$$

то отсюда следует, что $r' \in M$. □

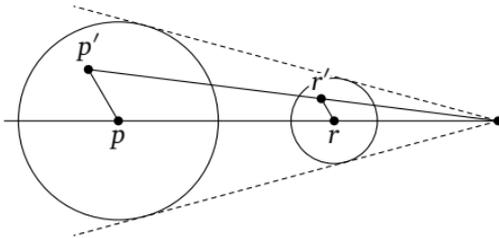


Рис. 4

Следствие 1. *Внутренние точки выпуклого тела образуют выпуклое множество.*

Следствие 2. *Всякая точка выпуклого тела является пределом его внутренних точек.*

Множество внутренних точек выпуклого тела M обозначим через M° . Это открытое выпуклое тело.

Согласно предложению 3, всякое выпуклое множество $M \subset S$ является выпуклым телом в $\text{aff } M$. Допуская вольность речи, часто говорят о внутренних точках произвольного выпуклого множества M , имея в виду его внутренние точки в пространстве $\text{aff } M$.

Для любой непостоянной аффинно-линейной функции f на пространстве S (см. § 2) положим

$$H_f = \{p \in S : f(p) = 0\},$$

$$H_f^+ = \{p \in S : f(p) \geq 0\}, \quad H_f^- = \{p \in S : f(p) \leq 0\} (= H_{-f}^+).$$

Множество H_f является гиперплоскостью. Множества H_f^+ и H_f^- называются (замкнутыми) полупространствами, ограничиваемыми гиперплоскостью H_f . Из предложения 2.3, примененного к аффинно-линейной функции f , следует, что всякое полупространство яв-

ляется выпуклым множеством. С другой стороны, всякий отрезок, соединяющий точку из H_f^+ с точкой из H_f^- , пересекает гиперплоскость H_f .

Определение 4. Гиперплоскость H_f называется *опорной гиперплоскостью* замкнутого выпуклого тела M , если $M \subset H_f^+$ и H_f содержит некоторую (граничную) точку тела M . Полупространство H_f^+ называется при этом *опорным полупространством* тела M .

Предложение 5. Гиперплоскость H , проходящая через граничную точку замкнутого выпуклого тела M , является опорной тогда и только тогда, когда $H \cap M^\circ = \emptyset$.

Доказательство. Если $H \cap M^\circ \neq \emptyset$, то точки множества M° (и тем самым точки тела M) имеются по обе стороны от H . Обратно, если точки тела M имеются по обе стороны от H , то, поскольку каждая точка тела M является пределом точек множества M° , по обе стороны от H имеются даже точки этого множества. Отрезок, соединяющий две такие точки, целиком лежит в M° и пересекает H , так что $H \cap M^\circ \neq \emptyset$. \square

Ключевой теоремой теории выпуклых множеств является следующая *теорема отделимости*.

Теорема 1. Через любую граничную точку замкнутого выпуклого тела проходит опорная гиперплоскость.

Доказательство. Пусть p — граничная точка замкнутого выпуклого тела M в n -мерном аффинном пространстве. Докажем индукцией по k , что при $k \leq n - 1$ через точку p проходит k -мерная плоскость, не пересекающая M° . При $k = 0$ такой плоскостью является сама точка p . Предположим, что уже удалось найти $(k - 1)$ -мерную плоскость P с нужными свойствами. Выберем любую $(k + 1)$ -мерную плоскость S' , содержащую P и какую-нибудь внутреннюю точку p_0 тела M , и попытаемся найти нужную нам k -мерную плоскость среди плоскостей, содержащих P и содержащихся в S' .

Рассмотрим выпуклое тело $M' = M \cap S'$ в пространстве S' . Ясно, что $M^\circ \cap S' \subset (M')^\circ$. Обратно, всякая точка $r \in (M')^\circ$ является внутренней точкой отрезка, соединяющего точку p_0 с некоторой точкой $q \in M' \subset M$ (см. рис. 5), и потому принадлежит M° . Таким образом,

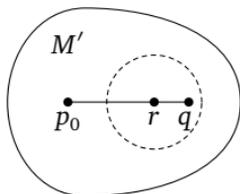


Рис. 5

$$(M')^\circ = M^\circ \cap S'.$$

В частности, отсюда следует, что $P \cap (M')^\circ = \emptyset$, и нам достаточно доказать, что в S' существует опорная гиперплоскость тела M' , содержащая P . Изменив обозначения, будем считать, что $S' = S$, $M' = M$ и $k + 1 = n$.

Итак, пусть P — это $(n - 2)$ -мерная плоскость, проходящая через точку p и не пересекающая M° . Докажем, что существует опорная гиперплоскость тела M , содержащая P .

Каждая гиперплоскость H , содержащая плоскость P , разбивается ею на две полуплоскости (или полугиперплоскости, если угодно), скажем, H^+ и H^- (не путать с предыдущими обозначениями полупространств). Если ни одна из полуплоскостей H^+ и H^- не пересекает M° , то все доказано. Если они обе пересекают M° , то и P пересекает M° , так что этот случай невозможен.

Пусть теперь H^+ пересекает M° , а H^- не пересекает. Начнем поворачивать гиперплоскость H вокруг P , условно говоря, по часовой стрелке. Ясно, что при небольшом повороте полуплоскость H^+ по-прежнему будет пересекать M° . Однако при повороте на π она перейдет в полуплоскость H^- , которая M° не пересекает. Поэтому существует некий минимальный поворот, при котором H^+ перестает пересекать M° . Повернутую таким образом гиперплоскость H обозначим через H_0 .

Согласно построению, полуплоскость H_0^+ не пересекает множество M° , но при малейшем повороте против часовой стрелки начинает его пересекать (см. рис. 6). С другой стороны, если бы полуплоскость H_0^- пересекала M° , то она сохранила бы это пересечение

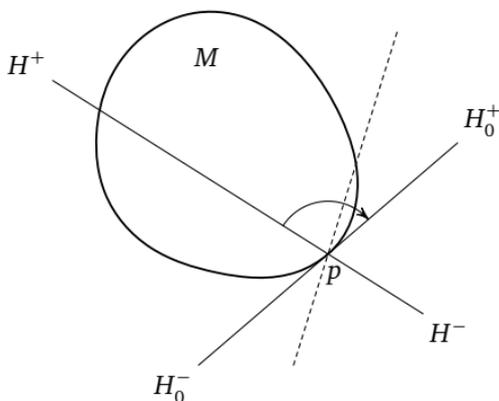


Рис. 6

при любом небольшом повороте. Но, как мы уже отмечали, обе половины гиперплоскости, содержащей P , не могут пересекать M° . Следовательно, H_0^- не пересекает M° и, значит, H_0 — опорная гиперплоскость. \square

Замечание 1. Фактически мы доказали более сильное утверждение, а именно, что любая плоскость, проходящая через точку p и не пересекающая M° , содержится в некоторой опорной гиперплоскости.

Замечание 2. Через данную граничную точку p тела M может проходить либо единственная опорная гиперплоскость, как на рис. 6, либо бесконечно много таких гиперплоскостей, как на рис. 7. Опорная гиперплоскость может содержать и другие точки тела M , кроме точки p , как на рис. 8.

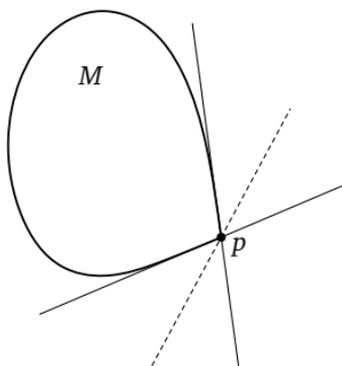


Рис. 7

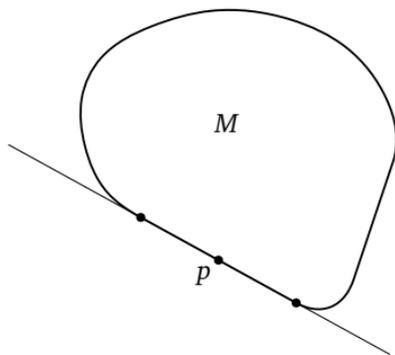


Рис. 8

Теорема 2. Всякое замкнутое выпуклое множество M является пересечением некоторого (быть может, бесконечного) числа полупространств.

Доказательство. Заметим, что всякая гиперплоскость H_f является пересечением полупространств H_f^+ и H_f^- . Отсюда следует, что и плоскость любой размерности является пересечением полупространств. Поэтому доказательство теоремы сводится к случаю, когда M — тело.

Докажем, что замкнутое выпуклое тело M является пересечением своих опорных полупространств. Пусть $q \notin M$ и p — какая-либо внутренняя точка тела M . Отрезок pq пересекает границу тела M в некоторой точке $r \neq q$. Проведем через эту точку опорную гипер-

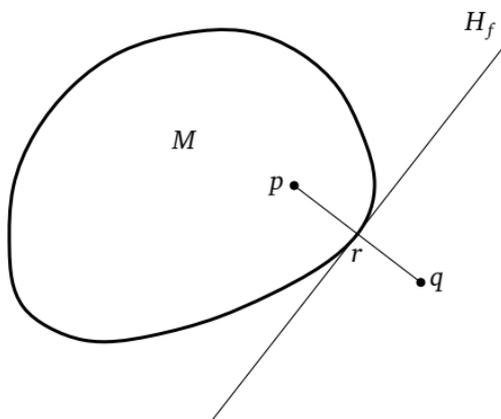


Рис. 9

плоскость H_f (см. рис. 9). Так как $f(p) > 0$, а $f(r) = 0$, то $f(q) < 0$, т. е. $q \notin H_f^+$. \square

Определение 5. Пересечение конечного числа полупространств называется *выпуклым многогранником*. Выпуклый многогранник, являющийся телом, называется *телесным*.

Иными словами, выпуклый многогранник есть множество решений конечной системы линейных неравенств. Отметим, что выпуклый многогранник не обязан быть ограниченным. Так, например, все пространство S является выпуклым многогранником (пересечение пустого множества полупространств). Выпуклый многогранник не обязан быть телесным (хотя иногда это и требуют).

Очевидно, что пересечение конечного числа выпуклых многогранников является выпуклым многогранником. Любая плоскость является выпуклым многогранником.

Пример 1. Симплекс с вершинами p_0, p_1, \dots, p_n является выпуклым многогранником, так как он может быть задан линейными неравенствами $x_i \geq 0$ ($i = 0, 1, \dots, n$), где x_0, x_1, \dots, x_n — барицентрические координаты относительно p_0, p_1, \dots, p_n .

Пример 2. Выпуклый многогранник, задаваемый линейными неравенствами $0 \leq x_i \leq 1$ ($i = 1, \dots, n$), где x_1, \dots, x_n — аффинные координаты относительно некоторого репера, называется *n -мерным параллелепипедом*.

Определение 6. Точка p выпуклого множества M называется *крайней*, если она не является внутренней точкой никакого отрезка, целиком лежащего в M .

Теорема 3. *Всякое ограниченное замкнутое выпуклое множество M является выпуклой оболочкой множества $E(M)$ своих крайних точек.*

Доказательство. Положим $\tilde{M} = \text{conv } E(M)$. Очевидно, что $\tilde{M} \subset M$. Докажем индукцией по $\dim S$, что $M \subset \tilde{M}$. При $\dim S = 0$ доказывать нечего. Пусть $\dim S > 0$ и $p \in M$. Докажем, что $p \in \tilde{M}$. Будем считать, что M — тело, так как иначе можно применить предположение индукции. Рассмотрим два случая.

1-й случай. Пусть p — граничная точка. Проведем через p опорную гиперплоскость H . Тогда $M \cap H$ — ограниченное замкнутое выпуклое множество, и всякая его крайняя точка является в то же время крайней точкой множества M . По предположению индукции $M \cap H$ является выпуклой оболочкой множества своих крайних точек. Следовательно, $p \in \tilde{M}$.

2-й случай. Пусть p — внутренняя точка. Проведем через p любую прямую. В силу ограниченности множества M она пересекает его по некоторому отрезку qr , содержащему точку p . Точки q и r являются граничными точками тела M и по доказанному принадлежат \tilde{M} . Следовательно, $p \in \tilde{M}$. \square

Теорема 4 (теорема Минковского—Вейля). *Следующие свойства ограниченного множества $M \subset S$ равносильны:*

- 1) M — выпуклый многогранник;
- 2) M — выпуклая оболочка конечного числа точек.

Доказательство. 1) Пусть

$$M = \bigcap_{i=1}^m H_{f_i}^+ \quad (15)$$

— выпуклый многогранник. Докажем, что всякая его крайняя точка есть единственная точка пересечения некоторых из гиперплоскостей H_{f_1}, \dots, H_{f_m} . Отсюда будет следовать, что M имеет лишь конечное число крайних точек. С другой стороны, по теореме 3 он является их выпуклой оболочкой.

Пусть $p \in M$ — крайняя точка. Положим

$$J = \{j : f_j(p) = 0\} \subset \{1, \dots, m\},$$

$$P = \{x \in S : f_j(x) = 0 \text{ при } j \in J\}.$$

Так как $f_i(p) > 0$ при $i \notin J$, то p является внутренней точкой выпуклого многогранника $M \cap P$ в пространстве P . Но p — крайняя точка

множества M и, следовательно, — крайняя точка множества $M \cap P$. Это означает, что $\dim P = 0$, т. е. $P = \{p\}$.

2) Пусть $M = \text{conv}\{p_1, \dots, p_k\}$. Будем считать, что $\text{aff } M = S$, и рассмотрим выпуклый многогранник

$$M^* = \left\{ f : f(p_i) \geq 0 \text{ при } i = 1, \dots, k, \sum_{i=1}^k f(p_i) = 1 \right\}$$

в пространстве аффинно-линейных функций на S . Так как аффинно-линейная функция на S однозначно определяется своими значениями в точках p_1, \dots, p_k , а для функций из M^* эти значения принадлежат отрезку $[0, 1]$, то M^* — ограниченный многогранник. По доказанному он является выпуклой оболочкой конечного числа точек, скажем, f_1, \dots, f_m .

По теореме 2 множество M (очевидно, что оно замкнуто) может быть задано линейными неравенствами. Следовательно,

$$M = \{p \in S : f(p) \geq 0 \forall f \in M^*\} = \{p \in S : f_i(p) \geq 0 \text{ при } i = 1, \dots, m\}.$$

Таким образом, M — выпуклый многогранник. \square

Определение 7. Гранью выпуклого многогранника M называется всякое непустое пересечение этого многогранника с некоторым числом его опорных гиперплоскостей. (Сам многогранник M также считается своей гранью как пересечение с пустым множеством опорных гиперплоскостей.)

Нульмерная грань называется *вершиной*, одномерная — *ребром*, $(n - 1)$ -мерная (где $n = \dim \text{aff } M$) — *гипергранью*. Пусть многогранник M задан формулой (15). Следующая теорема показывает, что для нахождения его граней можно ограничиться рассмотрением гиперплоскостей H_{f_1}, \dots, H_{f_m} .

Теорема 5. Всякая грань Γ многогранника M имеет вид

$$\Gamma = M \cap \left(\bigcap_{j \in J} H_{f_j} \right), \quad (16)$$

где $J \subset \{1, \dots, m\}$.

Доказательство. Пусть Γ' — грань многогранника M . Положим

$$J = \{j : \Gamma' \subset H_{f_j}\} \subset \{1, \dots, m\}.$$

Для каждого $i \notin J$ существует такая точка $p_i \in \Gamma'$, что $f_i(p_i) > 0$. Пусть p — центр тяжести системы этих точек. Тогда $f_i(p) > 0$ при всех $i \notin J$.

Определим теперь грань Γ по формуле (16) и докажем, что $\Gamma' = \Gamma$. Ясно, что $\Gamma' \subset \Gamma$ и что точка p является внутренней точкой гра-

ни Γ . Следовательно, всякая опорная гиперплоскость, проходящая через p , содержит Γ . Значит, $\Gamma' = \Gamma$. \square

Таким образом, если выпуклый многогранник задан системой линейных неравенств, то его грани получаются заменой части этих неравенств равенствами (но так, чтобы при этом получилось непустое множество). Нужно, однако, иметь в виду, что на определенной таким образом грани некоторые другие неравенства могут автоматически обращаться в равенства.

Пример 3. Грани n -мерного параллелепипеда, задаваемого неравенствами $0 \leq x_i \leq 1$ ($i = 1, \dots, n$), выделяются тем, что некоторые координаты равны 0 или 1. В частности, вершины — это точки, все координаты которых равны 0 или 1.

Задача 2. Найти грани сечения n -мерного параллелепипеда $0 \leq x_i \leq 1$ ($i = 1, \dots, n$) гиперплоскостью $x_1 + \dots + x_n = n/2$.

Задача 3. Найти грани n -мерного симплекса.

Задача 4. Доказать, что все грани многогранника $\text{conv}\{p_1, \dots, p_k\}$ являются выпуклыми оболочками некоторых из точек p_1, \dots, p_k .

Изучение комбинаторного строения выпуклых многогранников — это увлекательная и важная область математики. Вот два примера результатов из этой области.

1. Назовем f -вектором n -мерного ограниченного выпуклого многогранника последовательность $(a_0, a_1, \dots, a_{n-1})$, где a_k — число k -мерных граней этого многогранника. Каковы необходимые и достаточные условия для того, чтобы данная последовательность n натуральных чисел была f -вектором некоторого n -мерного многогранника? При $n = 3$ это следующие условия (*теорема Штейница*):

$$a_0 - a_1 + a_2 = 2, \quad 4 \leq a_0, a_2 \leq \frac{2a_1}{3}.$$

В общем случае ответ неизвестен.

2. Назовем вершины p и q выпуклого многогранника *смежными*, если отрезок pq является ребром этого многогранника. Легко видеть, что единственным трехмерным выпуклым многогранником, у которого любые две вершины смежны, является тетраэдр. Совершенно иная ситуация в 4-мерном пространстве. Как показал Д. Гейл, там существуют выпуклые многогранники с любым числом вершин, у которых любые две вершины смежные. Например, пусть M — выпуклая оболочка точек

$$p_i = (t_i, t_i^2, t_i^3, t_i^4), \quad i = 1, \dots, N,$$

где t_1, \dots, t_N — различные вещественные числа. Тогда

1) каждая из точек p_i является вершиной многогранника M (и это все его вершины: см. задачу 4);

2) каждый из отрезков $p_i p_j$ ($i \neq j$) является ребром многогранника M . Докажите это самостоятельно.

Предложение 6. *Крайние точки выпуклого многогранника M — это в точности его вершины.*

Доказательство. Если точка p является внутренней точкой отрезка, целиком лежащего в M , то любая опорная гиперплоскость, проходящая через p , содержит этот отрезок и, следовательно, p не может быть вершиной. Обратно, если точка p не является вершиной, то она является внутренней точкой некоторой грани положительной размерности и, значит, не может быть крайней точкой. \square

Важнейшие применения выпуклых многогранников вне математики связаны с *линейным программированием*. Основная задача линейного программирования формулируется таким образом: найти максимум (минимум) заданной аффинно-линейной функции на заданном выпуклом многограннике. Очевидно, что задача о минимуме функции f равносильна задаче о максимуме функции $-f$; поэтому можно говорить только об одной из этих задач.

В основе линейного программирования лежит следующая

Теорема 6. *Максимум аффинно-линейной функции f на ограниченном выпуклом многограннике M достигается в одной из его вершин.*

Доказательство. Согласно теореме 3 и предложению 6, каждая точка p многогранника M представляется в виде выпуклой линейной комбинации его вершин p_1, \dots, p_k :

$$p = \sum_{i=1}^k \lambda_i p_i, \quad \sum_{i=1}^k \lambda_i = 1, \quad \lambda_i \geq 0 \quad (i = 1, \dots, k).$$

В силу предложения 2.3

$$f(p) = \sum_{i=1}^k \lambda_i f(p_i) \leq \max_i f(p_i),$$

откуда и следует утверждение теоремы. \square

Приведем два примера ситуаций, в которых возникает задача линейного программирования.

Пример 4 (задача о получении максимальной прибыли). Некоторое предприятие располагает ресурсами P_1, \dots, P_m в количестве b_1, \dots, b_m соответственно и планирует произвести продукцию типов

Π_1, \dots, Π_n в количестве x_1, \dots, x_n соответственно. Пусть a_{ij} — количество ресурса P_i , нужное для производства единицы продукции Π_j , и c_j — цена единицы продукции Π_j . Очевидно, что должны выполняться неравенства

$$\sum_{j=1}^n a_{ij}x_j \leq b_i \quad (i=1, \dots, m), \quad x_j \geq 0 \quad (j=1, \dots, n).$$

Они задают некоторый выпуклый многогранник M в n -мерном пространстве с координатами x_1, \dots, x_n . Для получения максимальной прибыли нужно выбрать точку $(x_1, \dots, x_n) \in M$, в которой линейная функция $\sum_{j=1}^n c_j x_j$ (цена произведенной продукции) максимальна.

Пример 5 (транспортная задача). Имеются поставщики A_1, \dots, A_m , располагающие неким продуктом в количестве a_1, \dots, a_m соответственно, и потребители B_1, \dots, B_n , которые должны получить этот продукт в количестве b_1, \dots, b_n соответственно, причем $\sum_{i=1}^m a_i = \sum_{j=1}^n b_j$. Пусть x_{ij} — количество продукта, которое предполагается доставить от A_i к B_j , и c_{ij} — стоимость доставки единицы продукта от A_i к B_j . Должны выполняться условия

$$\sum_{j=1}^n x_{ij} = a_i, \quad \sum_{i=1}^m x_{ij} = b_j, \quad x_{ij} \geq 0.$$

Они задают некоторый выпуклый многогранник в mn -мерном пространстве с координатами x_{ij} ($i=1, \dots, m, j=1, \dots, n$). Задача состоит в минимизации линейной функции $\sum_{i,j} c_{ij}x_{ij}$ (общей стоимости перевозки) на этом многограннике.

Основной метод решения задачи линейного программирования, называемый *симплекс-методом*, состоит в движении по ребрам многогранника M в направлении возрастания функции f до тех пор, пока это возможно. Движение заканчивается в одной из вершин, в которых достигается максимум функции f (см. рис. 10).

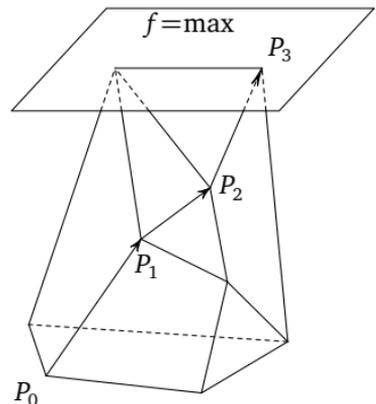


Рис. 10

§ 4. Евклидовы аффинные пространства

Объединяя аксиоматику евклидова векторного пространства с аксиоматикой аффинного пространства, мы можем теперь ввести понятие, охватывающее всю элементарную геометрию.

Определение 1. Аффинное пространство S , ассоциированное с евклидовым векторным пространством V , называется *евклидовым аффинным пространством* (или просто *евклидовым пространством*, если ясно, о чем идет речь).

Среди всех аффинных систем координат в евклидовом пространстве выделяются системы координат, определяемые ортонормированными реперами. Они называются *прямоугольными системами координат*.

Расстояние ρ между точками евклидова пространства определяется по формуле

$$\rho(p, q) = |\overline{pq}|.$$

Оно удовлетворяет аксиомам метрического пространства. В частности, неравенство треугольника следует из неравенства (32) § 5.5 для длины суммы векторов.

Нахождение расстояния от точки $p \in S$ до плоскости $P = p_0 + U$ с помощью векторизации сводится к нахождению расстояния от вектора $x = \overline{p_0 p} \in V$ до подпространства U . А именно, пусть $x = y + z$, где $y \in U$, $z \in U^\perp$. Приняв точку p_0 за начало отсчета, мы получаем по теореме 5.5.2:

$$\rho(p, P) = \rho(x, U) = |z|.$$

Точка $q = p_0 + y$ является «основанием перпендикуляра, опущенного из точки p на плоскость P ».

Задача 1. Доказать, что расстояние между плоскостями $P_1 = p_1 + U_1$ и $P_2 = p_2 + U_2$ евклидова пространства может быть найдено по формуле

$$\rho(P_1, P_2) = |\text{ort}_{U_1+U_2} \overline{p_1 p_2}|.$$

Определение 2. *Движением* евклидова аффинного пространства S называется всякое его аффинное преобразование, дифференциал которого является ортогональным оператором. (В частности, отсюда следует, что всякое движение биективно.)

Очевидно, что движение сохраняет расстояние между точками и, наоборот, всякое аффинное преобразование, сохраняющее расстояние между точками, является движением.

Замечание 1. На самом деле можно показать, что всякое биективное преобразование пространства S , сохраняющее расстояние между точками, автоматически является аффинным преобразованием и, следовательно, — движением.

Движения евклидова пространства S образуют группу, обозначаемую $\text{Isom } S$. Движение называется *собственным* (или сохраняющим ориентацию), если его дифференциал принадлежит $\text{SO}(V)$, и *несобственным* (или меняющим ориентацию) в противном случае. Собственные движения образуют подгруппу индекса 2 в $\text{Isom } S$, обозначаемую $\text{Isom}_+ S$ (ср. пример 4.6.18).

Пример 1. Важным примером несобственного движения является (ортогональное) *отражение* r_H относительно гиперплоскости H . Пусть e — единичный вектор, ортогональный H . Всякую точку $p \in S$ можно единственным образом представить в виде

$$p = q + \lambda e \quad (q \in H).$$

По определению

$$r_H p = q - \lambda e$$

(см. рис. 11). Дифференциал отражения r_H есть (ортогональное) отражение относительно направляющего подпространства гиперплос-

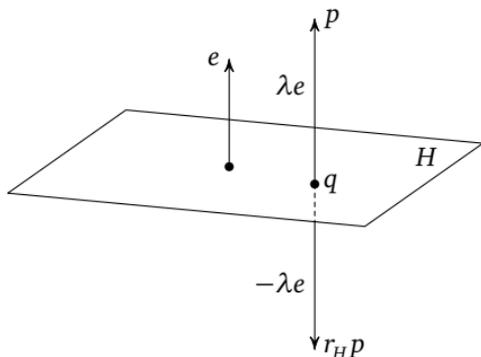


Рис. 11

кости H в пространстве V . Пусть H_1 и H_2 — две гиперплоскости. Если они параллельны, то $dr_{H_1} = dr_{H_2}$ и, следовательно,

$$d(r_{H_1} r_{H_2}) = dr_{H_1} \cdot dr_{H_2} = \mathcal{E}.$$

В этом случае $r_{H_1} r_{H_2}$ — параллельный перенос на удвоенный общий перпендикуляр гиперплоскостей H_1 и H_2 (см. рис. 12). Если же H_1

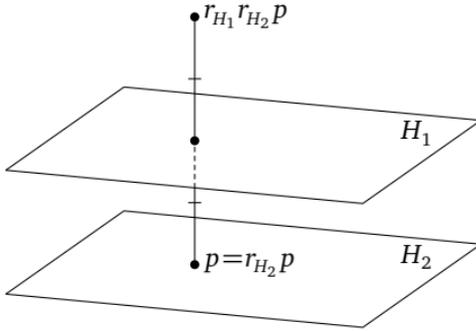


Рис. 12

и H_2 пересекаются по $(n - 2)$ -мерной плоскости P , то $r_{H_1} r_{H_2}$ — поворот вокруг P на удвоенный угол между H_1 и H_2 , т. е. движение, оставляющее на месте все точки плоскости P и осуществляющее поворот на указанный угол в любой двумерной плоскости, ортогональной P (ср. пример 6.3.4).

Задача 2. Доказать, что группа $\text{Isom } S$ порождается отражениями относительно гиперплоскостей.

Если в пространстве S выбрано начало отсчета, то всякое движение однозначно представляется в виде (13), где $\varphi \in O(V)$. Однако вектор b , вообще говоря, зависит от начала отсчета. Следующая теорема даст некое каноническое представление любого движения.

Теорема 1. Для всякого движения f существует однозначно определенная плоскость $P = p_0 + U$ со следующими свойствами:

- 1) плоскость P инвариантна относительно f , причем $f|_P$ — параллельный перенос на некоторый вектор $b \in U$ (быть может, нулевой);
- 2) оператор $\mathcal{A} = df$ не имеет ненулевых неподвижных векторов в U^\perp .

Доказательство. Если искомая плоскость P существует, то оператор \mathcal{A} действует на ее направляющем подпространстве U тождественно и, значит, U совпадает с подпространством неподвижных векторов этого оператора. Пусть, далее, q — любая точка. Представим ее в виде $q = p + c$, где $p \in P$, а $c \in U^\perp$. Имеем:

$$f(q) = f(p) + \mathcal{A}c = p + b + \mathcal{A}c = q + b + (\mathcal{A}c - c).$$

Так как $b \in U$, а $\mathcal{A}c - c \in U^\perp$, то вектор $\mathcal{A}c - c$ совпадает с ортогональной составляющей вектора $qf(q)$ относительно подпространства U . С другой стороны, так как оператор $\mathcal{A} - \mathcal{E}$ невырожден на U^\perp , то вектор c однозначно определяется по вектору $\mathcal{A}c - c = (\mathcal{A} - \mathcal{E})c$. Но, зная вектор c , мы можем найти точку $p = q - c$ плоскости P . Таким образом, если искомая плоскость существует, то она единственна.

Проведенный анализ также показывает, как построить нужную плоскость. А именно, в качестве ее направляющего подпространства возьмем подпространство U неподвижных векторов оператора \mathcal{A} . Тогда будет выполнено условие 2). Далее, возьмем любую точку q и рассмотрим вектор $a = qf(q)$. Пусть $a = b + d$, где $b \in U$, $d \in U^\perp$. Найдем такой вектор $c \in U^\perp$, что $\mathcal{A}c - c = d$. Положим $p = q - c$ и рассмотрим плоскость $P = p + U$. Проверим, что для нее выполнено условие 1). Имеем:

$$f(p) = f(q) - \mathcal{A}c = q + a - \mathcal{A}c = p + c + b + d - \mathcal{A}c = p + b \in P$$

и для любого $u \in U$

$$f(p + u) = f(p) + \mathcal{A}u = p + b + u = (p + u) + b,$$

что и требовалось доказать. \square

Задача 3. Доказать, что плоскость P есть совокупность точек, которые при движении f перемещаются на наименьшее расстояние (равное длине вектора b).

Плоскость P называется *осью движения* f . Движение f определяется своей осью $P = p_0 + U$, вектором $b \in U$ и ортогональным преобразованием $\mathcal{B} = \mathcal{A}|_{U^\perp}$ пространства U^\perp , не имеющим неподвижных векторов. Как следует из описания ортогональных преобразований, для собственных движений размерность $\dim U^\perp$ четна, а для несобственных — нечетна.

Пользуясь этой теоремой, опишем движения евклидовой прямой, плоскости и трехмерного пространства в терминах элементарной геометрии. Через P будем обозначать ось движения f .

Пусть f — движение евклидовой прямой. Возможны 2 случая.

1) $\dim P = 1$. В этом случае f — параллельный перенос.

2) $\dim P = 0$, т. е. P — точка. В этом случае $\mathcal{B} = -\mathcal{E}$ и f — отражение (симметрия) относительно точки P .

Пусть f — движение евклидовой плоскости. Возможны 3 случая.

1) $\dim P = 2$. В этом случае f — параллельный перенос.

2) $\dim P = 1$, т. е. P — прямая. В этом случае $\mathcal{R} = -\mathcal{E}$ и f — отражение относительно прямой P или *скользящее отражение*, т. е. композиция отражения относительно P и параллельного переноса вдоль P .

3) $\dim P = 0$, т. е. P — точка. В этом случае f — (нетривиальный) поворот вокруг точки P .

Пусть, наконец, f — движение трехмерного евклидова пространства. Возможны 4 случая.

1) $\dim P = 3$. В этом случае f — параллельный перенос.

2) $\dim P = 2$. В этом случае f — отражение относительно плоскости P или *скользящее отражение*, т. е. композиция отражения относительно P и параллельного переноса на вектор, параллельный P .

3) $\dim P = 1$. В этом случае f — (нетривиальный) поворот вокруг прямой P или *винтовое движение*, т. е. композиция поворота вокруг P и параллельного переноса вдоль P .

4) $\dim P = 0$. В этом случае f — *зеркальный поворот*, т. е. композиция (нетривиального) поворота вокруг некоторой прямой и отражения относительно плоскости, перпендикулярной этой прямой; при этом указанные прямая и плоскость пересекаются в точке P .

Пример 2. Обозначим через $s_{p,\alpha}$ поворот в евклидовой плоскости на угол α вокруг точки p . Его дифференциал есть, очевидно, поворот на тот же угол в соответствующем векторном пространстве. Рассмотрим произведение $s_{p,\alpha} s_{q,\beta}$ поворотов вокруг разных точек. Вычисляя его дифференциал, мы находим, что оно представляет собой поворот на угол $\alpha + \beta$ вокруг некоторой третьей точки, если только этот угол не кратен 2π (и параллельный перенос — в противном случае). Для того чтобы найти центр поворота, можно воспользоваться следующим предложением.

Предложение 1. Пусть pqr — треугольник с углами α, β, γ (см. рис. 13). Тогда

$$s_{p,2\alpha} s_{q,2\beta} s_{r,2\gamma} = \text{id}.$$

Доказательство. Обозначим через l, m, n прямые, содержащие стороны qr, rp, pq треугольника, и через r_l, r_m, r_n — отражения относительно этих прямых. Тогда (см. пример 6.3.4)

$$s_{p,2\alpha} = r_m r_n, \quad s_{q,2\beta} = r_n r_l, \quad s_{r,2\gamma} = r_l r_m,$$

откуда перемножением получается доказываемое равенство. \square

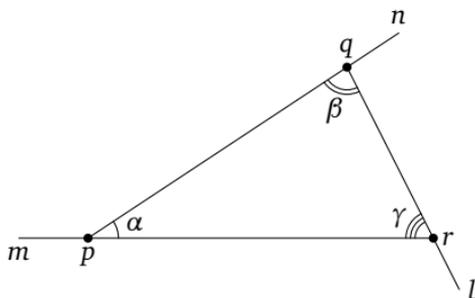


Рис. 13

Задача 4. Пользуясь доказанным предложением, указать способ построения центра поворота $s_{p, \alpha}$, $s_{q, \beta}$ из примера 2.

Для любой фигуры M евклидова пространства S можно определить ее *группу симметрии*

$$\text{Sym } M = \{f \in \text{Isom } S : f(M) = M\}.$$

Таким образом возникают, например, кристаллографические группы как группы симметрии кристаллических структур.

Отметим, что если группа $\text{Sym } M$ содержит несобственные движения, то группа

$$\text{Sym}_+ M = \{f \in \text{Isom}_+ S : f(M) = M\}$$

является ее подгруппой индекса 2 как ядро гомоморфизма

$$\text{Sym } M \rightarrow \{\pm 1\}, \quad f \mapsto \det df.$$

Если M — ограниченный выпуклый многогранник, то группа $\text{Sym } M$ конечна, так как движение, отображающее многогранник M на себя, однозначно определяется тем, как оно переставляет его вершины, а таких перестановок может быть лишь конечное число. Кроме того, группа $\text{Sym } M$ сохраняет центр тяжести множества вершин многогранника M и потому фактически представляет собой подгруппу ортогональной группы.

Наиболее симметричны так называемые правильные многогранники.

Пусть M — телесный выпуклый многогранник в n -мерном евклидовом пространстве. Назовем *флагом* многогранника M набор его граней $\{F_0, F_1, \dots, F_{n-1}\}$, где $\dim F_k = k$ и $F_0 \subset F_1 \subset \dots \subset F_{n-1}$.

Определение 3. Выпуклый многогранник M называется *правильным*, если для любых двух его флагов существует движение $f \in \text{Sym } M$, переводящее первый из этих флагов во второй.

Так как движение $f \in \text{Sym } M$, очевидно, определяется тем, куда оно переводит какой-нибудь один флаг, то порядок группы симметрии правильного многогранника равен числу его флагов.

Двумерные правильные многогранники — это обычные *правильные многоугольники*. Их группы симметрии были описаны в примере 4.1.11.

Трёхмерные правильные многогранники — это *платоновы тела*, т. е. правильный тетраэдр T , куб K , октаэдр O , додекаэдр D и икосаэдр I . (См. рис. 6 §4.5.) Куб и октаэдр, а также додекаэдр и икосаэдр — это так называемые двойственные правильные многогранники, имеющие одинаковые группы симметрии, так как центры граней одного из двойственных многогранников являются вершинами другого. (Тетраэдр двойствен сам себе.)

Согласно предыдущему, порядок группы симметрии $\text{Sym } P$ трёхмерного правильного многогранника P равен числу его флагов, т. е.

$$|\text{Sym } P| = (\text{число вершин}) \times \\ \times (\text{число ребер, выходящих из каждой вершины}) \times 2.$$

Следовательно,

$$|\text{Sym } T| = 24, \quad |\text{Sym } K| = |\text{Sym } O| = 48, \quad |\text{Sym } D| = |\text{Sym } I| = 120.$$

Группа $\text{Sym}_+ P$ имеет вдвое меньший порядок. Она состоит из поворотов вокруг прямых, проходящих через центр многогранника P и через его граничную точку, которая является либо вершиной, либо серединой ребра, либо центром грани.

Задача 5. Перечислить все элементы группы симметрии куба.

В рамках группового подхода аналогично евклидовой геометрии столь же просто определяется *псевдоевклидова геометрия*.

Вещественное векторное пространство, в котором фиксирована симметрическая билинейная функция α сигнатуры (k, l) , где $k, l > 0$, $k + l = n = \dim V$, называется *псевдоевклидовым векторным пространством* сигнатуры (k, l) . Группа линейных преобразований пространства V , сохраняющих функцию α , называется *псевдоортогональной группой* и обозначается $O(V, \alpha)$. Соответствующая ей группа матриц в базисе, в котором α имеет нормальный вид, обозначается $O_{k,l}$.

Аффинное пространство S , ассоциированное с псевдоевклидовым векторным пространством V , называется *псевдоевклидовым аффинным пространством* соответствующей сигнатуры, а группа $\text{Isom } S = d^{-1}(\text{O}(V, \alpha))$ — группой его движений. Геометрия, определяемая этой группой, и есть псевдоевклидова геометрия.

Пространство-время специальной теории относительности — это псевдоевклидово аффинное пространство сигнатуры $(3, 1)$. Оно называется *пространством Минковского*, а группа его движений — *группой Пуанкаре*. (Соответствующая группа псевдоортогональных преобразований называется *группой Лоренца*.)

Задача 6. Описать группу $O_{1,1}$. (Указание: использовать систему координат, в которой соответствующая квадратичная функция имеет вид $q(x) = x_1x_2$.)

Задача 7. Сформулировать и доказать «третий признак равенства треугольников» на псевдоевклидовой плоскости.

§ 5. Квадрики

Простейшими объектами аффинной и евклидовой геометрий являются плоскости, которые, как мы знаем, задаются системами линейных уравнений. Естественным обобщением плоскостей (называемых также *линейными многообразиями*) являются так называемые *алгебраические многообразия* — подмножества аффинного пространства, задаваемые произвольными системами алгебраических уравнений. Их изучением занимается алгебраическая геометрия. Это обширный раздел математики, который не может быть представлен в настоящем курсе. Мы лишь слегка коснемся некоторых общих вопросов алгебраической геометрии в гл. 9, а в этом параграфе рассмотрим простейший после плоскостей тип алгебраических многообразий — квадрики, задаваемые одним алгебраическим уравнением второй степени. К их числу относятся такие объекты элементарной геометрии, как окружности и сферы.

Будем считать, что $\text{char } K \neq 2$.

Определение 1. *Аффинно-квадратичной функцией* на аффинном пространстве S называется всякая функция $Q: S \rightarrow K$, имеющая в векторизованной форме вид

$$Q(x) = q(x) + l(x) + c, \quad (17)$$

где q — квадратичная функция, l — линейная функция, а c — константа.

Пусть \dot{q} — поляризация квадратичной функции q , т. е. соответствующая ей симметрическая билинейная функция.

Лемма 1. При переносе начала отсчета o в точку $o' = o + a$ ($a \in V$) слагаемые выражения (17) преобразуются следующим образом:

$$q'(x) = q(x), \quad l'(x) = 2\dot{q}(a, x) + l(x), \quad c' = q(a) + l(a) + c. \quad (18)$$

Доказательство. Имеем

$$\begin{aligned} Q(o' + x) &= Q(o + a + x) = q(a + x) + l(a + x) + c = \\ &= q(a) + 2\dot{q}(a, x) + q(x) + l(a) + l(x) + c = \\ &= q(x) + (2\dot{q}(a, x) + l(x)) + (q(a) + l(a) + c). \quad \square \end{aligned}$$

В частности, квадратичная функция q не зависит от выбора начала отсчета.

В координатах выражение (17) принимает вид

$$Q(x) = \sum_{i,j} a_{ij}x_i x_j + \sum_i b_i x_i + c \quad (a_{ij} = a_{ji}). \quad (19)$$

Коэффициентам b_i и c можно придать следующий смысл:

$$c = Q(o), \quad b_i = \frac{\partial Q}{\partial x_i}(o). \quad (20)$$

Линейная функция

$$l(x) = \sum_i b_i x_i$$

называется *дифференциалом* функции Q в точке o и обозначается $d_o Q$. В случае $K = \mathbb{R}$ это согласуется с обычным определением дифференциала.

Определение 2. Точка o называется *центром* аффинно-квадратичной функции Q , если

$$Q(o + x) = Q(o - x) \quad \forall x \in V. \quad (21)$$

Ясно, что это имеет место тогда и только тогда, когда $d_o Q = 0$. Поэтому множество всех центров функции Q задается системой линейных уравнений

$$\frac{\partial Q}{\partial x_1} = \dots = \frac{\partial Q}{\partial x_n} = 0. \quad (22)$$

Оно либо является плоскостью некоторого числа измерений, либо пусто. Легко видеть, что матрица коэффициентов системы (22) — это удвоенная матрица (a_{ij}) квадратичной функции q . Следовательно, если q невырождена, то Q имеет единственный центр.

Положим

$$X(Q) = \{p \in S : Q(p) = 0\}.$$

Определение 3. Множество вида $X(Q)$, где Q — аффинно-квадратичная функция, если только оно не пусто и не является плоскостью, называется *квадрикой* или *гиперповерхностью второго порядка*.

Квадрика на плоскости называется *коникой* или *кривой второго порядка*. Квадрика в трехмерном пространстве называется также *поверхностью второго порядка*.

Определение 4. Точка o называется *центром* квадрики, если эта квадрика симметрична относительно o , т. е. вместе со всякой точкой $o + x$ ($x \in V$) содержит точку $o - x$. Центр квадрики, лежащий на ней самой, называется ее *вершиной*.

Квадрика называется *центральной*, если она имеет (хотя бы один) центр.

Очевидно, что всякий центр аффинно-квадратичной функции Q является центром квадрики $X(Q)$. Как будет показано ниже, верно и обратное.

Докажем некоторые простые геометрические свойства квадрик.

Предложение 1. *Любая прямая либо целиком лежит на квадрике, либо пересекается с ней не более чем в двух точках.*

Доказательство. Так как начало отсчета o может быть выбрано в любой точке, то без ограничения общности можно считать, что прямая проходит через o . Пусть функция Q в векторизованной форме имеет вид (17). Тогда пересечение прямой $L = o + \langle x \rangle = \{o + tx : t \in K\}$ ($x \in V$) с квадрикой $X(Q)$ определяется условием

$$Q(tx) = t^2q(x) + tl(x) + c = 0, \quad (23)$$

представляющим собой квадратное уравнение относительно t . Если все коэффициенты этого уравнения равны нулю, то $L \subset X(Q)$; в противном случае оно имеет не более двух корней, а это означает, что пересечение $L \cap X(Q)$ содержит не более двух точек. \square

Предложение 2. *Если o — вершина квадрики X , то вместе с любой точкой $p \neq o$ квадрика X содержит всю прямую op .*

Доказательство. Пусть $p = o + x$ ($x \in V$); тогда X содержит три различные точки o , $o + x$, $o - x$ прямой op и, следовательно, — всю прямую. \square

Всякое подмножество аффинного пространства, содержащее точку o и вместе с любой точкой $p \neq o$ всю прямую op , называется конусом с вершиной в точке o . Квадрика называется конической, если она имеет (хотя бы одну) вершину.

Предложение 3. *Всякая квадрика содержит точку, не являющуюся ее вершиной.*

Доказательство. Если бы все точки квадрики были ее вершинами, то в силу предложения 2 вместе с любыми двумя точками она содержала бы проходящую через них прямую и, согласно теореме 1.4, была бы плоскостью, а это противоречит определению квадрики. \square

Очевидно, что пропорциональные аффинно-квадратичные функции определяют одну и ту же квадрику. Обратное утверждение не столь очевидно; оно составляет содержание следующей теоремы.

Теорема 1. *Пусть X — квадрика в аффинном пространстве над бесконечным полем K . Если $X = X(Q_1) = X(Q_2)$ для каких-то аффинно-квадратичных функций Q_1, Q_2 , то эти функции пропорциональны.*

Доказательство. Возьмем в качестве начала отсчета какую-нибудь точку o квадрики X , не являющуюся ее вершиной. Тогда в векторизованной форме

$$Q_1(x) = q_1(x) + l_1(x), \quad Q_2(x) = q_2(x) + l_2(x),$$

где $l_1, l_2 \neq 0$. Точки пересечения прямой $\{o + tx : t \in K\}$ с квадрикой X определяются любым из уравнений

$$t^2 q_1(x) + t l_1(x) = 0, \quad t^2 q_2(x) + t l_2(x) = 0.$$

Так как эти уравнения должны иметь одинаковые решения (относительно t), то при $l_1(x), l_2(x) \neq 0$ мы получаем

$$\frac{q_1(x)}{l_1(x)} = \frac{q_2(x)}{l_2(x)},$$

откуда

$$q_1(x)l_2(x) = q_2(x)l_1(x). \quad (24)$$

Применяя следствие теоремы 3.7.1, получаем, что это равенство верно при всех x .

Предположим, что линейные функции l_1 и l_2 не пропорциональны. Тогда в подходящем базисе $l_1(x) = x_1$, $l_2(x) = x_2$ и равенство (24) записывается в виде

$$q_1(x)x_2 = q_2(x)x_1.$$

Рассматривая члены в левой и правой частях этого равенства, мы видим, что должно быть

$$q_1(x) = l(x)x_1, \quad q_2(x) = l(x)x_2,$$

где $l(x)$ — какая-то линейная функция, и, значит,

$$Q_1(x) = (l(x) + 1)x_1, \quad Q_2(x) = (l(x) + 1)x_2.$$

Так как $X = X(Q_1)$, то X содержит гиперплоскость $x_1 = 0$. Так как в то же время $X = X(Q_2)$, то функция Q_2 должна тождественно обращаться в нуль на этой гиперплоскости. Однако ни один из ее множителей $l(x) + 1$ и x_2 не обращается на ней тождественно в нуль (первый из них не обращается в нуль уже в точке o). Поскольку в алгебре многочленов нет делителей нуля, мы тем самым приходим к противоречию.

Итак, $l_2 = \lambda l_1$ ($\lambda \in K^*$). Из (24) получаем тогда, что и $q_2 = \lambda q_1$, и, значит, $Q_2 = \lambda Q_1$. \square

Следствие 1. *Всякий центр квадрики $X(Q)$ является также центром функции Q .*

Доказательство. Если o — центр квадрики $X(Q)$, то $X(Q) = X(\bar{Q})$, где

$$\bar{Q}(o + x) = Q(o - x).$$

Следовательно, $\bar{Q} = \lambda Q$ ($\lambda \in K^*$). Сравнивая члены второй степени в выражениях Q и \bar{Q} , мы видим, что должно быть $\lambda = 1$, т. е. $\bar{Q} = Q$, а это и означает, что o — центр функции Q . \square

Следствие 2. *Если квадрика $X(Q)$ инвариантна относительно некоторого параллельного переноса, то и функция Q инвариантна относительно этого переноса.*

Доказательство. Если квадрика $X(Q)$ переходит в себя при параллельном переносе на вектор a , то $X(Q) = X(\bar{Q})$, где

$$\bar{Q}(p) = Q(p + a).$$

Далее рассуждаем так же, как в доказательстве следствия 1. \square

Замечание 1. Анализ приведенного выше доказательства теоремы 1 с учетом замечания 3.7.2 показывает, что она верна и для конечных полей, исключая только поле \mathbb{Z}_3 . (Напомним, что мы считаем, что $\text{char } K \neq 2$.) Над полем \mathbb{Z}_3 можно привести следующий контрпример: уравнения $x_1^2 + x_1x_2 + 1 = 0$ и $x_2^2 + x_1x_2 + 1 = 0$ задают одну и ту же конику в \mathbb{Z}_3^2 , состоящую из точек $(1, 1)$ и $(-1, -1)$. Однако оба следствия теоремы верны и для поля \mathbb{Z}_3 .

Пусть аффинно-квадратичная функция Q представлена в векторизованной форме выражением (17). Положим

$$\text{Ker } Q = \text{Ker } q \cap \text{Ker } l \quad (25)$$

(где $\text{Ker } q \doteq \text{Ker } \dot{q}$).

Предложение 4. *Функция Q инвариантна относительно параллельного переноса на вектор a тогда и только тогда, когда $a \in \text{Ker } Q$.*

В частности, отсюда следует, что $\text{Ker } q \cap \text{Ker } l$ не зависит от выбора начала отсчета.

Доказательство. Инвариантность функции Q относительно параллельного переноса на вектор a равносильна тому, что она сохраняет свой вид при переносе начала отсчета из точки o в точку $o' = o + a$. Ввиду леммы 1 это происходит тогда и только тогда, когда $a \in \text{Ker } Q$. \square

Таким образом, если $U = \text{Ker } Q \neq 0$, то квадрака $X = X(Q)$ вместе с каждой точкой p содержит целую плоскость $p + U$. Такая квадрака называется *цилиндрической* с направляющим подпространством U . Выберем базис пространства V таким образом, чтобы последние d его векторов составляли базис подпространства U . Тогда выражение Q не будет содержать последних d координат. Пусть $S_0 \subset S$ — любая плоскость, направляющее подпространство которой натянуто на первые $n - d$ базисных векторов, и X_0 — квадрака, задаваемая в S_0 уравнением $Q = 0$ (т. е. $X_0 = X \cap S_0$). Тогда $X = X_0 + U$ (см. рис. 14).

Квадрика, не являющаяся цилиндрической, называется *невырожденной*. Ввиду сказанного выше описание всех квадрик сводится к описанию невырожденных квадрик.

Предложение 5. *Невырожденная квадрака имеет не более одного центра.*

Доказательство. Пусть o и o' — два центра квадраки X . Обозначим через s и s' центральные симметрии относительно o и o' соответственно. Тогда $sX = s'X = X$ и, следовательно, $ss'X = X$. Так

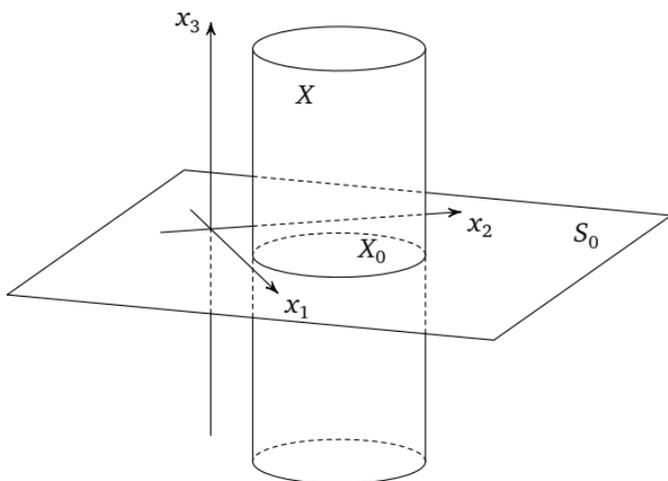


Рис. 14

как

$$d(ss') = ds \cdot ds' = (-\mathcal{E})^2 = \mathcal{E},$$

то ss' — (нетривиальный) параллельный перенос и, значит, квадрика X цилиндрическая. \square

Нецилиндрические квадрики можно разбить на три типа.

I. Неконические центральные квадрики. Выбрав начало отсчета в центре квадрики и умножив ее уравнение на подходящее число, мы приведем его к виду

$$q(x_1, \dots, x_n) = 1, \quad (26)$$

где q — невырожденная квадратичная функция.

II. Конические квадрики. Выбрав начало отсчета в вершине квадрики, мы приведем ее уравнение к виду

$$q(x_1, \dots, x_n) = 0, \quad (27)$$

где q — невырожденная квадратичная функция. При этом у нас еще остается возможность умножить уравнение на любое число $\lambda \neq 0$.

III. Нецентральные квадрики. Так как $\text{Ker } q \cap \text{Ker } l = 0$, но $\text{Ker } q \neq 0$ (иначе квадрика была бы центральной), то $\dim \text{Ker } q = 1$ и

$$V = \text{Ker } l \oplus \text{Ker } q. \quad (28)$$

Выбрав начало отсчета на квадрике и базис пространства V , согласованный с разложением (28), мы приведем уравнение квадрики

к виду

$$u(x_1, \dots, x_{n-1}) = x_n, \quad (29)$$

где $u = q|_{\text{Ker } l}$ — невырожденная квадратичная функция от $n - 1$ переменных. При этом остается возможность умножить уравнение на любое число $\lambda \neq 0$, одновременно разделив на λ последний базисный вектор.

Возможности дальнейшего упрощения уравнения квадрики за счет выбора подходящего базиса в пространстве V зависят от поля K (см. § 5.4). При $K = \mathbb{C}$ или \mathbb{R} мы можем привести квадратичную функцию q к нормальному виду.

Рассмотрим более подробно случай $K = \mathbb{R}$. В этом случае уравнение невырожденной квадрики может быть приведено к одному и только одному из следующих видов:

I. Неконические центральные квадрики:

$$x_1^2 + \dots + x_k^2 - x_{k+1}^2 - \dots - x_n^2 = 1 \quad (0 < k \leq n). \quad (30)$$

II. Конические квадрики:

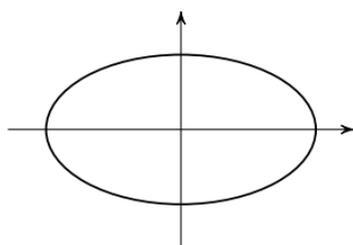
$$x_1^2 + \dots + x_k^2 - x_{k+1}^2 - \dots - x_n^2 = 0 \quad \left(\frac{n}{2} \leq k < n\right). \quad (31)$$

(Неравенство $k \geq \frac{n}{2}$ достигается за счет возможного умножения уравнения на -1 .)

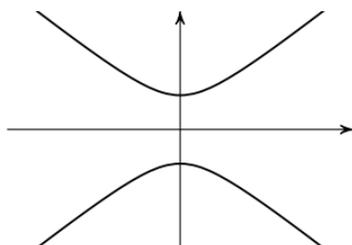
III. Нецентральные квадрики:

$$x_1^2 + \dots + x_k^2 - x_{k+1}^2 - \dots - x_{n-1}^2 = x_n \quad \left(\frac{n-1}{2} \leq k < n\right). \quad (32)$$

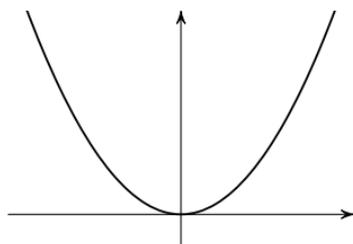
Полученный результат можно интерпретировать как классификацию вещественных квадратик с точностью до аффинных преобразований. В самом деле, если квадрики X_1 и X_2 задаются одним и тем же уравнением в аффинных системах координат, связанных с реперами $\{o; e_1, \dots, e_n\}$ и $\{o'; e'_1, \dots, e'_n\}$ соответственно, то X_1 переводится в X_2 аффинным преобразованием, переводящим репер $\{o; e_1, \dots, e_n\}$ в репер $\{o'; e'_1, \dots, e'_n\}$. Обратно, если квадратика X_1 переводится в квадратик X_2 аффинным преобразованием f , то X_1 и X_2 задаются одним и тем же уравнением в аффинных системах координат, связанных с реперами $\{o; e_1, \dots, e_n\}$ и $\{f(o); df(e_1), \dots, df(e_n)\}$ соответственно.



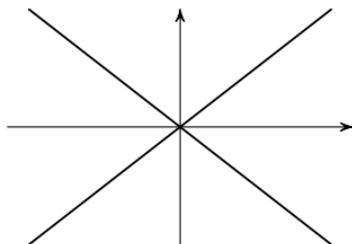
Эллипс



Гипербола



Парабола



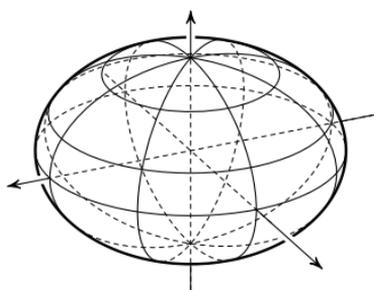
Пара пересекающихся прямых

Рис. 15

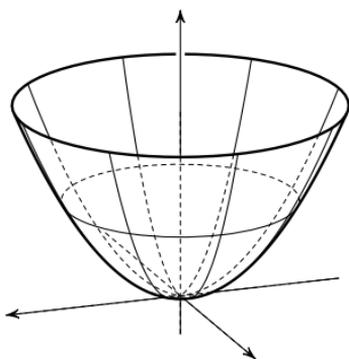
В частности, при $n = 2$ и 3 получаются хорошо известные классы вещественных кривых и поверхностей второго порядка, перечисленные в табл. 1 и представленные на рис. 15 и 16 соответственно.

Таблица 1

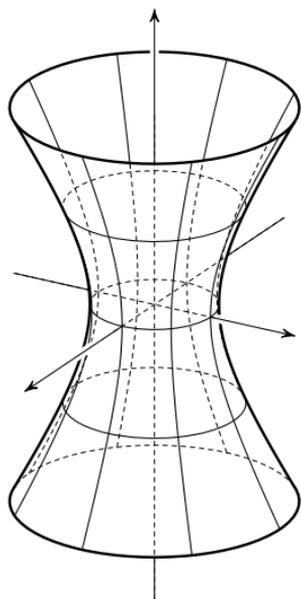
n	Тип	k	Название
2	I	2	эллипс
		1	гипербола
	II	1	пара пересекающихся прямых
	III	1	парабола
3	I	3	эллипсоид
		2	однополостный гиперболоид
		1	двуполостный гиперболоид
	II	2	квадратичный конус
	III	2	эллиптический параболоид
		1	гиперболический параболоид



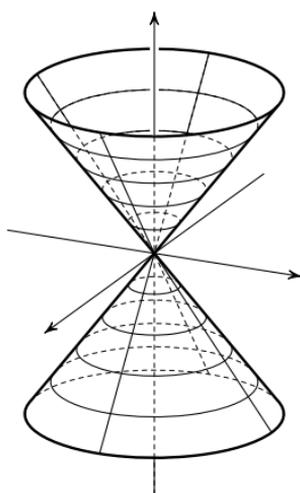
Эллипсоид



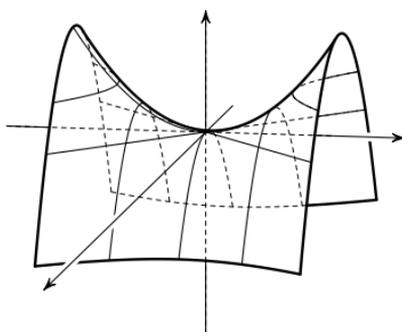
Эллиптический параболоид



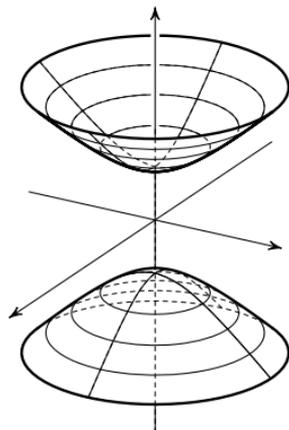
Однополостный гиперboloид



Конус



Гиперbолический параболоид



Двуполостный гиперboloид

Рис. 16

В произвольной размерности квадрики типа I при $k = n$ называются *эллипсоидами*, а при $k < n$ — *гиперболоидами*; квадрики типа II называются *квадратичными конусами*; квадрики типа III при $k = n - 1$ называются *эллиптическими параболоидами*, а при $k < n - 1$ — *гиперболическими параболоидами*.

Вещественная квадрика X является гладкой гиперповерхностью (см. определение в § 12.1) в окрестности точки $p \in X$ тогда и только тогда, когда $d_p Q \neq 0$, т.е. когда p не является вершиной; при этом уравнение $d_p Q(x - p) = 0$ задает касательную гиперплоскость квадрики X в точке p . В частности, неконические квадрики гладки всюду.

Замечательным свойством вещественных (и комплексных) квадрик, которым, вообще говоря, не обладают гиперповерхности бóльших порядков, является высокая степень их аффинной симметрии.

Пусть X — вещественная квадрика. Обозначим через $G(X)$ группу всех аффинных преобразований, отображающих X на себя.

Теорема 2. *Если X — неконическая квадрика, то группа $G(X)$ транзитивно действует на X ; если X — коническая квадрика, то группа $G(X)$ транзитивно действует на дополнении к множеству вершин в X .*

Доказательство. Если X — цилиндрическая квадрика с направляющим подпространством U , то группа $G(X)$ содержит группу параллельных переносов на векторы из U , которая транзитивно действует на любой плоскости вида $p + U$. Поэтому доказательство теоремы в этом случае сводится к ее доказательству для невырожденной квадрики X_0 в пространстве меньшей размерности (см. обозначение выше).

Пусть X — эллипсоид, задаваемый в векторизованной форме уравнением $q(x) = 1$, где q — положительно определенная квадратичная функция. Превратим пространство V в евклидово, приняв за скалярное умножение поляризацию квадратичной функции q . Тогда X будет единичной сферой в этом пространстве, а группа $G(X)$ будет, во всяком случае, содержать ортогональную группу $O(V)$. (На самом деле она будет с ней совпадать, но нам это не нужно.) Пусть x, x' — любые векторы из X ; тогда

$$V = \langle x \rangle \oplus \langle x \rangle^\perp = \langle x' \rangle \oplus \langle x' \rangle^\perp.$$

Рассмотрим линейное преобразование $\varphi \in GL(V)$, переводящее x в x' и отображающее подпространство $\langle x \rangle^\perp$ на $\langle x' \rangle^\perp$ таким образом,

чтобы это был изоморфизм евклидовых пространств. Очевидно, что $\varphi \in O(V)$, и по построению $\varphi(x) = x'$.

Случай, когда X — гиперboloид, разбирается аналогично, с той разницей, что, взяв за скалярное умножение поляризацию квадратичной функции q , мы превратим V не в евклидово, а в псевдоевклидово пространство некоторой сигнатуры (k, l) (где $k + l = n$). Подпространства $\langle x \rangle^\perp$ и $\langle x' \rangle^\perp$ в этом случае будут псевдоевклидовыми пространствами сигнатуры $(k - 1, l)$ и, следовательно, будут изоморфны.

Пусть теперь X — квадратичный конус, задаваемый в векторизованной форме уравнением $q(x) = 0$, где q — квадратичная функция сигнатуры (k, l) (где $k + l = n$). Превратим, как и выше, пространство V в псевдоевклидово. Для любого ненулевого вектора $x \in X$ существует такой вектор $y \in V$, что $(x, y) \neq 0$. Нормировав вектор y , можно считать, что $(x, y) = 1$. Далее, не нарушая этого равенства, можно заменить y на $y - \frac{1}{2}(y, y)x$ и тем самым добиться, чтобы $(y, y) = 0$. Тогда в двумерном подпространстве $\langle x, y \rangle$ скалярное умножение будет иметь матрицу $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ и, значит, будет невырожденным сигнатуры $(1, 1)$. Отсюда следует, что

$$V = \langle x, y \rangle \oplus \langle x, y \rangle^\perp,$$

где $\langle x, y \rangle^\perp$ — псевдоевклидово (или евклидово) пространство сигнатуры $(k - 1, l - 1)$. Прделав такие же построения для другого ненулевого вектора $x' \in X$, мы получим аналогичное разложение

$$V = \langle x', y' \rangle \oplus \langle x', y' \rangle^\perp.$$

Рассмотрим линейное преобразование $\varphi \in GL(V)$, переводящее x в x' , y в y' и отображающее подпространство $\langle x, y \rangle^\perp$ на $\langle x', y' \rangle^\perp$ таким образом, чтобы это был изоморфизм псевдоевклидовых пространств. Тогда $\varphi \in O(V, q) \subset G(X)$, и по построению $\varphi(x) = x'$.

Пусть, наконец, X — параболоид, задаваемый в векторизованной форме уравнением (29). Всякий вектор $x \in V$ будем представлять в виде $x = y + te$, где $y \in \text{Ker } l$, $t \in \mathbb{R}$, а e — базисный вектор подпространства $\text{Ker } q$, так что $x \in X$ тогда и только тогда, когда $u(y) = t$. Для любого $a \in \text{Ker } l$ рассмотрим аффинное преобразование

$$f_a: y + te \mapsto y + a + (t + 2\dot{u}(a, y) + u(a))e.$$

Если $u(y) = t$, то

$$u(y+a) = t + 2\dot{u}(a, y) + u(a),$$

и обратно. Это означает, что $f_a \in G(X)$. Очевидно, что преобразования f_a ($a \in \text{Ker } l$) образуют группу, транзитивно действующую на X . \square

Задача 1. Доказать, что если X — параболоид, задаваемый уравнением (29), то группа $G(X)$ транзитивно действует в области $u(x_1, \dots, x_{n-1}) < x_n$.

С каждым параболоидом $X = X(Q)$ каноническим образом связано одномерное подпространство $\text{Ker } q \subset V$, называемое *особым направлением* параболоида X . Так как $\text{Ker } q \not\subset \text{Ker } l$ при любом выборе начала отсчета, то при $x \in \text{Ker } q$ уравнение (23) имеет ровно одно решение. Следовательно, любая прямая особого направления пересекает параболоид ровно в одной точке; более того, это пересечение по той же причине трансверсально (см. рис. 17).

Задача 2. Доказать, что для любого неособого направления параболоида X существует прямая этого направления, которая не пересекает X .

Посмотрим теперь, к какому виду можно привести уравнение квадрики в евклидовом пространстве, если ограничиться прямоугольными системами координат. Как и в аффинной геометрии, задача сводится к случаю невырожденных квадрик. Рассмотрим, как и выше, три типа таких квадрик.

I. **Неконические центральные квадрики.** Из теоремы о приведении квадратичной функции к главным осям (следствие 2 теоремы 6.3.1) следует, что уравнение такой квадрики в прямоугольной системе координат может быть приведено к виду

$$\lambda_1 x_1^2 + \dots + \lambda_n x_n^2 = 1 \quad (\lambda_1, \dots, \lambda_n \neq 0). \quad (33)$$

Числа $\lambda_1, \dots, \lambda_n$ определены однозначно с точностью до перестановки.

II. **Конические квадрики.** Уравнение такой квадрики в прямоугольной системе координат может быть приведено к виду

$$\lambda_1 x_1^2 + \dots + \lambda_n x_n^2 = 0 \quad (\lambda_1, \dots, \lambda_n \neq 0). \quad (34)$$

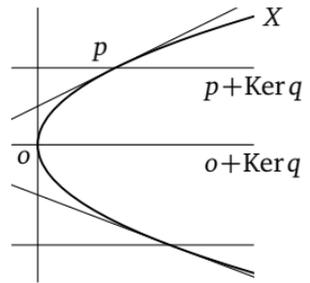


Рис. 17

Числа $\lambda_1, \dots, \lambda_n$ определены однозначно с точностью до перестановки и одновременного умножения на число $\lambda \neq 0$.

III. Нецентральные квадрики (параболоиды). Выбрав начало отсчета произвольно и приведя квадратичную функцию q к главным осям, мы получим прямоугольную систему координат, в которой уравнение параболоида будет иметь вид

$$\lambda_1 x_1^2 + \dots + \lambda_{n-1} x_{n-1}^2 + b_1 x_1 + \dots + b_{n-1} x_{n-1} + b_n x_n + c = 0$$

$$(\lambda_1, \dots, \lambda_{n-1}, b_n \neq 0).$$

За счет переноса начала отсчета по координатам x_1, \dots, x_{n-1} можно убрать линейные члены, содержащие эти координаты. (При этом, вообще говоря, изменится свободный член.) После этого за счет переноса начала отсчета по координате x_n можно убрать свободный член. Наконец, умножив уравнение на подходящее число, можно привести его к виду

$$\lambda_1 x_1^2 + \dots + \lambda_{n-1} x_{n-1}^2 = x_n \quad (\lambda_1, \dots, \lambda_{n-1} \neq 0). \quad (35)$$

Покажем, что начало отсчета, при котором уравнение параболоида приводится к виду (35), определено однозначно. Для этого охарактеризуем его в инвариантных терминах.

Пусть $\{o; e_1, \dots, e_n\}$ — репер, в котором уравнение параболоида имеет вид (35). Тогда особое направление этого параболоида есть $\langle e_n \rangle$, а его касательная гиперплоскость в точке o задается уравнением $x_n = 0$. Следовательно, если базис $\{e_1, \dots, e_n\}$ ортонормированный, то касательная гиперплоскость параболоида в точке o ортогональна особому направлению. Такая точка называется *вершиной* параболоида (хотя это и не согласуется с определением 4), а проходящая через нее прямая особого направления — *осью* параболоида. Подчеркнем, что эти определения имеют смысл лишь применительно к параболоидам в евклидовом пространстве.

Предложение 6. *Всякий параболоид в евклидовом пространстве имеет единственную вершину.*

(См. рис. 17, где вершиной изображенной там параболы является точка o .)

Доказательство. Пусть p — точка параболоида с координатами x_1, \dots, x_n . Дифференцируя уравнение (35), находим, что координаты нормального вектора параболоида в точке p суть

$$2\lambda_1 x_1, \dots, 2\lambda_{n-1} x_{n-1}, -1.$$

Для того чтобы точка p была вершиной параболоида, необходимо и достаточно, чтобы этот вектор был пропорционален e_n , а это имеет место тогда и только тогда, когда $x_1 = \dots = x_{n-1} = 0$, т. е. $p = o$. \square

Следствие. Коэффициенты $\lambda_1, \dots, \lambda_{n-1}$ в уравнении (35) определены однозначно с точностью до перестановки и одновременного умножения на -1 .

Доказательство. Как мы показали, начало отсчета, при котором уравнение параболоида приводится к виду (35), определено однозначно. Вектор e_n как единичный вектор особого направления определен однозначно с точностью до умножения на -1 , приводящего к умножению на -1 левой части уравнения (35). Если вектор e_n фиксирован, то мы уже не можем умножить уравнение на число $\lambda \neq 1$, не изменив его правой части; но тогда числа $\lambda_1, \dots, \lambda_{n-1}$ определены однозначно с точностью до перестановки как собственные значения симметрического оператора, соответствующего квадратичной функции q . \square

Аналогично тому, как это было сделано выше применительно к аффинной классификации квадрик, полученные результаты можно интерпретировать как классификацию квадрик в евклидовом пространстве с точностью до движений.

§ 6. Проективные пространства

На фотографическом снимке или (реалистическом) рисунке плоской местности изображения параллельных прямых, вообще говоря, пересекаются, а изображения равных отрезков одной прямой, вообще говоря, не равны (см. рис. 18 на следующей странице). Это говорит о том, что отображение местности на плоскость снимка или рисунка не является аффинным. То же самое можно сказать и об изображении на сетчатке нашего глаза. Во всех этих случаях мы имеем дело с центральным проектированием.

Еще одним житейским примером центрального проектирования может служить световое пятно на полу от лампы с круглым абажуром. Когда абажур направлен вертикально вниз, то граница этого пятна имеет форму окружности, как и край самого абажура. Но когда мы начинаем поворачивать абажур вокруг горизонтальной оси, эта окружность превращается в эллипс, который, вытягиваясь все больше и больше, в какой-то момент, когда его дальний край

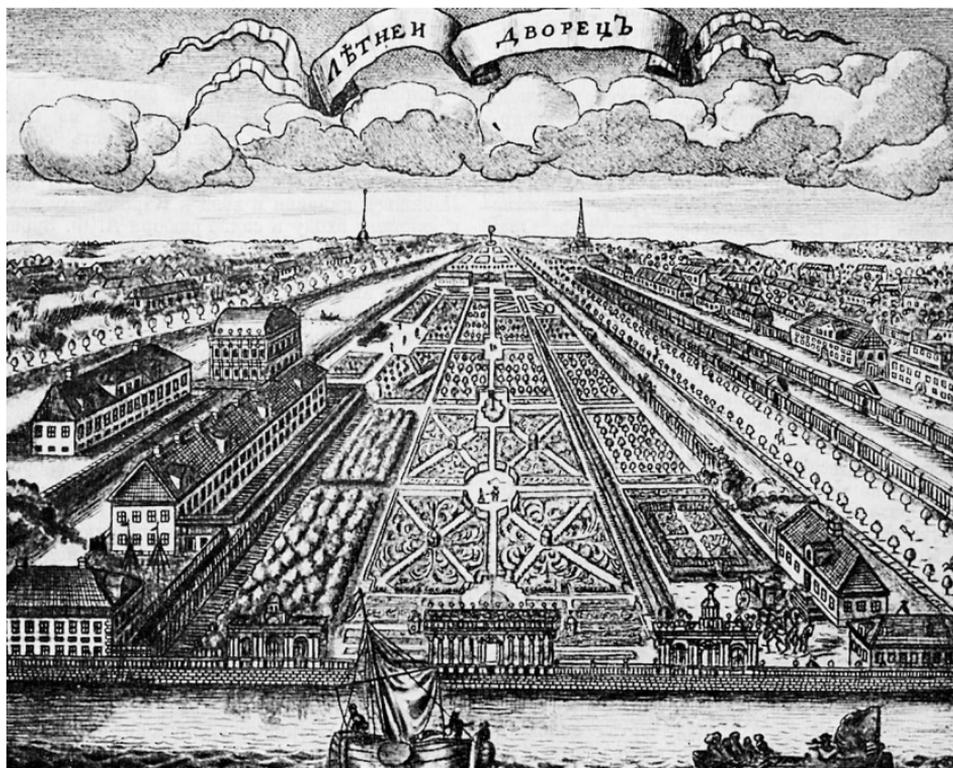


Рис. 18. Гравюра Летнего сада А. Зубова. (Воспроизводится по книге: Вергунов А. П., Горюхов В. А. Русские сады и парки. М.: Наука, 1988.)

уходит в бесконечность, превращается в параболу. Когда мы продолжаем поворачивать абакюр, парабола «раскрывается», превращаясь в ветвь гиперболы, и если бы мы приставили точно такой же абакюр с противоположной стороны лампы, мы увидели бы другую ветвь этой гиперболы. Таким образом, край абажура проектируется на пол то в виде эллипса, то в виде параболы, то в виде гиперболы.

Отметим еще одно обстоятельство. На рисунке плоской местности изображения параллельных прямых пересекаются в точке, которая не имеет прообраза на местности (иначе прямые не были бы параллельны). С другой стороны, в тот момент, когда граница светового пятна от лампы с абакюром превращается в параболу, изображение самой высокой точки края абажура исчезает, уходя в бесконечность. Таким образом, мало того, что центральное проектирование не является аффинным отображением, оно вообще не сюръективно и не всюду определено.

Для изучения центрального проектирования удобно рассмотреть множество, называемое проективной плоскостью, «точками» которого являются прямые, проходящие через центр проектирования, а пересечения этих прямых с плоскостью проектирования считать изображениями соответствующих «точек». При этом «точки», соответствующие прямым, параллельным выбранной плоскости проектирования, не получают никакого изображения. (Но они будут иметь изображение при другом выборе плоскости проектирования). Они называются «бесконечно удаленными точками» по отношению к данной плоскости проектирования.

Далее, множество «точек», соответствующих прямым, лежащим в какой-либо плоскости, проходящей через центр проектирования, естественно называть «прямой» проективной плоскости. На плоскости проектирования такая «прямая», за вычетом ее «бесконечно удаленной точки», изображается в виде обычной прямой. Единственным исключением является «прямая», соответствующая плоскости, параллельной плоскости проектирования. Она целиком состоит из «бесконечно удаленных точек» и не получает никакого изображения. Эта «прямая» называется «бесконечно удаленной прямой» по отношению к данной плоскости проектирования.

Описанную конструкцию можно интерпретировать как добавление к аффинной плоскости «бесконечно удаленных точек», составляющих «бесконечно удаленную прямую». При этом к каждой прямой из пучка параллельных прямых аффинной плоскости добавляется одна и та же «бесконечно удаленная точка». В построенной таким образом «плоскости» любые две прямые пересекаются.

Важно, однако, отметить, что все «точки» и «прямые» проективной плоскости равноправны. Понятие бесконечной удаленности относительно: оно зависит от выбора плоскости проектирования.

Обобщая эти идеи на произвольную размерность и произвольное поле, мы приходим к следующим определениям.

Определение 1. Множество одномерных подпространств $(n+1)$ -мерного векторного пространства V над полем K называется n -мерным проективным пространством над K и обозначается PV . Для всякого $(k+1)$ -мерного подпространства $U \subset V$ подмножество $PU \subset PV$ называется k -мерной плоскостью пространства PV .

В частности, нульмерные плоскости — это точки пространства PV ; одномерные плоскости называются *прямыми*, $(n-1)$ -мерные — *гиперплоскостями*.

Очевидно, что пересечение плоскостей, если только оно не пусто, также является плоскостью.

Пространство PK^{n+1} , построенное указанным образом по пространству строк K^{n+1} , обозначается также KP^n .

Для всякого ненулевого вектора $x \in V$ мы будем обозначать через \hat{x} одномерное подпространство $\langle x \rangle$, рассматриваемое как точка пространства PV .

Пусть S — какая-либо гиперплоскость пространства V , не проходящая через нуль, и V_S — ее направляющее подпространство. Определим отображение

$$\varphi_S: PV \setminus PV_S \rightarrow S,$$

ставящее в соответствие каждой точке $\hat{x} \in PV \setminus PV_S$ ($x \in V \setminus V_S$) точку пересечения прямой $\langle x \rangle$ с S (см. рис. 19).

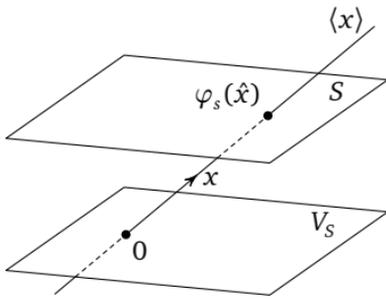


Рис. 19

Определение 2. Гиперплоскость S вместе с отображением φ_S называется *аффинной картой* пространства PV . Точки гиперплоскости PV_S пространства PV называются *бесконечно удаленными* по отношению к аффинной карте S .

Замечание 1. Термин «аффинная карта» вполне согласуется с обычным употреблением слова «карта». Точно так же как географическая карта

представляет собой отображение части земной поверхности на лист бумаги, аффинная карта представляет собой отображение части проективного пространства на аффинное пространство.

Замечание 2. отождествляя точки проективного пространства с их изображениями на аффинной карте, мы иногда будем говорить об аффинной карте как о части проективного пространства. Имея это в виду, можно сказать, что проективное пространство получается из аффинного пространства добавлением бесконечно удаленных точек.

Каждая k -мерная плоскость пространства PV , не лежащая целиком в PV_S , за вычетом ее бесконечно удаленных точек изображается k -мерной плоскостью на аффинной карте S . Плоскости, целиком лежащие в PV_S , называются *бесконечно удаленными* по отношению к S .

Однородными координатами точки $\hat{x} \in PV$ называются координаты вектора x в каком-либо выбранном базисе пространства V . Однородные координаты точки определены лишь с точностью до одновременного умножения на число $\lambda \neq 0$. Этим они отличаются от координат в привычном смысле слова. Кроме того, они не могут быть равны нулю одновременно. Точка с однородными координатами x_0, x_1, \dots, x_n обозначается $(x_0 : x_1 : \dots : x_n)$.

Неоднородными координатами точки пространства PV называются аффинные координаты ее изображения на какой-либо аффинной карте. В отличие от однородных координат, неоднородные координаты точки определены однозначно, но они могут быть вообще не определены, а именно, они не определены для точек, бесконечно удаленных по отношению к выбранной аффинной карте.

Установим связь между однородными и неоднородными координатами. Пусть $\{e_0, e_1, \dots, e_n\}$ — базис пространства V . Рассмотрим аффинную карту

$$S_0 = e_0 + \langle e_1, \dots, e_n \rangle \quad (36)$$

(см. рис. 20). Изображением точки $\hat{x} = (x_0 : x_1 : \dots : x_n)$ на S_0 служит точка

$$e_0 + \frac{x_1}{x_0}e_1 + \dots + \frac{x_n}{x_0}e_n,$$

аффинные координаты которой относительно репера $\{e_0; e_1, \dots, e_n\}$

суть $\frac{x_1}{x_0}, \dots, \frac{x_n}{x_0}$. Таким образом, при указанном выборе аффинной карты и репера неоднородными координатами точки $(x_0 : x_1 : \dots : x_n)$ служат отношения $\frac{x_1}{x_0}, \dots, \frac{x_n}{x_0}$. Точки с $x_0 = 0$ являются бесконечно удаленными по отношению к S_0 .

Аналогично, неоднородными координатами точки \hat{x} на аффинной карте

$$S_i = e_i + \langle e_0, e_1, \dots, e_{i-1}, e_{i+1}, \dots, e_n \rangle \quad (37)$$

служат отношения $\frac{x_0}{x_i}, \frac{x_1}{x_i}, \dots, \frac{x_{i-1}}{x_i}, \frac{x_{i+1}}{x_i}, \dots, \frac{x_n}{x_i}$. Точки с $x_i = 0$ являются бесконечно удаленными по отношению к S_i .

Отметим, что карты S_0, S_1, \dots, S_n составляют «атлас» в том смысле, что они покрывают все пространство PV .

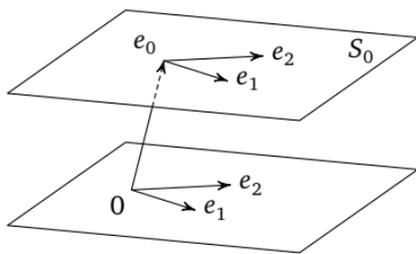


Рис. 20

Задача 1. Доказать, что не существует атласа пространства PV из меньшего числа карт.

Задача 2. Пусть y_1, \dots, y_n — неоднородные координаты изображения точки $\hat{x} \in PV$ на карте S_0 . Найти ее неоднородные координаты на карте S_1 .

Теорема 1. Через любые $k + 1$ точек проективного пространства проходит плоскость размерности $\leq k$, причем если эти точки не содержатся в плоскости размерности $< k$, то через них проходит единственная плоскость размерности k .

Доказательство. Перевод утверждения теоремы на язык векторных пространств есть следующее очевидное утверждение: любые $k + 1$ векторов содержатся в подпространстве размерности $\leq k + 1$, и если они не содержатся в подпространстве размерности $< k + 1$, то они содержатся в единственном подпространстве размерности $k + 1$. \square

Теорема 2. Пусть Π_1 и Π_2 — плоскости n -мерного проективного пространства. Если $\dim \Pi_1 + \dim \Pi_2 \geq n$, то $\Pi_1 \cap \Pi_2 \neq \emptyset$, причем

$$\dim(\Pi_1 \cap \Pi_2) \geq \dim \Pi_1 + \dim \Pi_2 - n. \quad (38)$$

Например, любые две прямые проективной плоскости пересекаются.

Доказательство. Если $\Pi_1 = PU_1$, $\Pi_2 = PU_2$, то

$$\dim U_1 + \dim U_2 = \dim \Pi_1 + \dim \Pi_2 + 2 \geq n + 2 > \dim V.$$

Следовательно, $U_1 \cap U_2 \neq 0$ и, значит, $\Pi_1 \cap \Pi_2 = P(U_1 \cap U_2) \neq \emptyset$. Более точно,

$$\dim(U_1 \cap U_2) \geq \dim U_1 + \dim U_2 - \dim V,$$

откуда следует (38). \square

Всякий невырожденный линейный оператор $\mathcal{A} \in GL(V)$ переводит одномерные подпространства в одномерные подпространства и тем самым определяет некоторое биективное преобразование $\hat{\mathcal{A}}$ пространства PV .

Определение 3. Преобразования вида $\hat{\mathcal{A}}$, где $\mathcal{A} \in GL(V)$, называются *проективными преобразованиями* пространства PV .

Очевидно, что проективное преобразование переводит любую плоскость пространства PV в плоскость той же размерности.

Отображение $\mathcal{A} \mapsto \hat{\mathcal{A}}$ является гомоморфизмом группы $GL(V)$ в группу преобразований пространства PV . Его образ есть группа

всех проективных преобразований пространства PV , называемая также *полной проективной группой* пространства PV и обозначаемая через $\text{PGL}(V)$.

Лемма 1. Ядро гомоморфизма $\mathcal{A} \mapsto \widehat{\mathcal{A}}$ есть группа скалярных операторов $\lambda \mathcal{E}$ ($\lambda \in K^*$).

Доказательство. Если оператор \mathcal{A} переводит каждое одномерное подпространство в себя, то все ненулевые векторы являются его собственными векторами. Но очевидно, что сумма собственных векторов с различными собственными значениями не может быть собственным вектором. Следовательно, все собственные значения оператора \mathcal{A} одинаковы, а это и означает, что он скалярен. \square

Таким образом,

$$\text{PGL}(V) \simeq \text{GL}(V) / \{\lambda \mathcal{E} : \lambda \in K^*\}.$$

Посмотрим, как представляется проективное преобразование $\widehat{\mathcal{A}}$ на аффинной карте S . Оператор \mathcal{A} осуществляет аффинное отображение гиперплоскости S на гиперплоскость $\mathcal{A}S$. Изображение точки $\widehat{\mathcal{A}}\hat{x} = \widehat{\mathcal{A}x}$ ($x \in S$) на карте S есть центральная проекция (с центром в нуле) точки $\mathcal{A}x \in \mathcal{A}S$ на S (см. рис. 21). Таким образом, можно

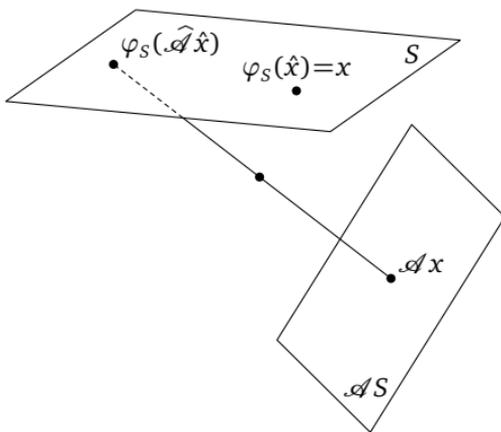


Рис. 21

сказать, что, с точки зрения аффинной карты, проективное преобразование есть композиция аффинного отображения и центрального проектирования.

В координатах это выглядит следующим образом. Пусть матрица оператора \mathcal{A} в базисе $\{e_0, e_1, \dots, e_n\}$ имеет вид $A = (a_{ij})_{i,j=0}^n$. Рассмотрим

рим неоднородные координаты в пространстве PV , определяемые репером $\{e_0; e_1, \dots, e_n\}$ аффинной карты S_0 (см. (36)). Пусть

$$x = e_0 + x_1 e_1 + \dots + x_n e_n,$$

так что точка $\hat{x} \in PV$ имеет неоднородные координаты x_1, \dots, x_n . Обозначим через y_1, \dots, y_n неоднородные координаты ее образа. Тогда

$$y_i = \frac{a_{i0} + \sum_{j=1}^n a_{ij} x_j}{a_{00} + \sum_{j=1}^n a_{0j} x_j} \quad (i = 1, \dots, n). \quad (39)$$

Например, проективные преобразования прямой суть дробно-линейные преобразования

$$y = \frac{ax + b}{cx + d} \quad (ad - bc \neq 0). \quad (40)$$

(При $c \neq 0$ точка $-d/c$ переходит в бесконечно удаленную точку, а бесконечно удаленная точка переходит в точку a/c .)

Если $\mathcal{A}S = S$, то преобразование $\widehat{\mathcal{A}}$ представляется на карте S как аффинное преобразование. Следующая лемма показывает, что всякое аффинное преобразование пространства S получается таким образом.

Лемма 2. *Всякое аффинное преобразование гиперплоскости $S \subset V$, не проходящей через нуль, единственным образом продолжается до линейного преобразования пространства V .*

Доказательство. Репер $\{e_0; e_1, \dots, e_n\}$ гиперплоскости S есть в то же время базис пространства V (см. рис. 20). Продолжением аффинного преобразования f гиперплоскости S является линейное преобразование пространства V , переводящее базис $\{e_0, e_1, \dots, e_n\}$ в базис $\{f(e_0), df(e_1), \dots, df(e_n)\}$. \square

Рассматривая аффинное пространство S как часть проективного пространства PV , можно сказать, что группа $GA(S)$ есть подгруппа группы $PGL(V)$.

Задача 3. Доказать, что для всякого проективного преобразования комплексного проективного пространства существует аффинная карта, на которой оно представляется как аффинное преобразование.

Геометрия, определяемая группой проективных преобразований, называется *проективной геометрией*. Следующая теорема при

сравнении с теоремой 2.1 показывает, насколько группа проективных преобразований богаче группы аффинных преобразований.

Назовем систему $n + 2$ точек n -мерного проективного пространства *системой точек общего положения*, если никакие $n + 1$ из них не лежат в одной гиперплоскости.

Теорема 3. Пусть $\{p_0, p_1, \dots, p_{n+1}\}$ и $\{q_0, q_1, \dots, q_{n+1}\}$ — две системы точек общего положения n -мерного проективного пространства PV . Тогда существует единственное проективное преобразование, переводящее p_i в q_i при $i = 0, 1, \dots, n + 1$.

Доказательство. Пусть $p_i = \hat{e}_i$, $q_i = \hat{f}_i$, где e_i, f_i ($i = 0, 1, \dots, n + 1$) — ненулевые векторы пространства V . Условие теоремы означает, что $\{e_0, e_1, \dots, e_n\}$ (соответственно $\{f_0, f_1, \dots, f_n\}$) — базис пространства V и все координаты вектора e_{n+1} (соответственно f_{n+1}) в этом базисе отличны от нуля. Нормировав векторы e_0, e_1, \dots, e_n (соответственно f_0, f_1, \dots, f_n) некоторым вполне определенным образом, можно добиться того, чтобы $e_{n+1} = e_0 + e_1 + \dots + e_n$ (соответственно $f_{n+1} = f_0 + f_1 + \dots + f_n$). При этих условиях пусть \mathcal{A} — линейный оператор, переводящий базис $\{e_0, e_1, \dots, e_n\}$ в $\{f_0, f_1, \dots, f_n\}$. Тогда $\mathcal{A}e_{n+1} = f_{n+1}$ и $\hat{\mathcal{A}}$ есть единственное проективное преобразование, удовлетворяющее требованию теоремы. \square

В частности, любые 3 различные точки проективной прямой проективным преобразованием можно перевести в любые 3 различные точки. Из-за этого в проективной геометрии не существует не только понятия расстояния между точками, но и понятия отношения тройки точек прямой, имеющегося в аффинной геометрии. Однако существует некий инвариант четверки точек прямой.

А именно, пусть p_1, p_2, p_3, p_4 — точки прямой $PU \subset PV$. Выберем в пространстве U какой-либо базис $\{e_1, e_2\}$ и для любых векторов $u, v \in U$ обозначим через $\det(u, v)$ определитель матрицы, составленной из их координат в этом базисе. Пусть $p_i = \hat{u}_i$ ($i = 1, 2, 3, 4$). Легко видеть, что выражение

$$(p_1, p_2; p_3, p_4) = \frac{\det(u_1, u_3)}{\det(u_3, u_2)} : \frac{\det(u_1, u_4)}{\det(u_4, u_2)} \quad (41)$$

не зависит ни от нормировки векторов u_i , ни от выбора базиса $\{e_1, e_2\}$ в U . Оно называется *двойным отношением* четверки точек p_1, p_2, p_3, p_4 .

Пусть L — аффинная карта прямой PU . Выберем базис $\{e_1, e_2\}$ так, чтобы $L = e_2 + \langle e_1 \rangle$, и пусть $u_i = e_2 + x_i e_1$. Тогда x_i — неоднород-

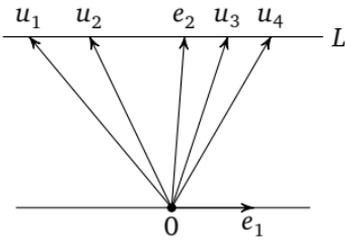


Рис. 22

ная координата точки p_i на карте L (см. рис. 22) и

$$\det(u_i, u_j) = \begin{vmatrix} x_i & 1 \\ x_j & 1 \end{vmatrix} = x_i - x_j.$$

Следовательно,

$$(p_1, p_2; p_3, p_4) = \frac{x_1 - x_3}{x_3 - x_2} : \frac{x_1 - x_4}{x_4 - x_2}. \quad (42)$$

Подчеркнем, что в силу наличия инвариантного определения (41) выражение (42) не зависит от выбора аффинной карты и координаты на ней.

Замечание 3. Двойное отношение считается определенным, если среди точек p_1, p_2, p_3, p_4 нет трех одинаковых. При этом если $p_2 = p_3$ или $p_1 = p_4$, его значение считается равным ∞ .

Задача 4. Выяснить что происходит с двойным отношением $(p_1, p_2; p_3, p_4) = \delta$ при перестановках точек p_1, p_2, p_3, p_4 . Доказать, что выражение $\frac{(\delta^2 - \delta + 1)^3}{\delta^2(\delta - 1)^2}$ не меняется ни при каких перестановках.

Задача 5. Изучив изображение четырех симметрично расположенных вдоль центральной аллеи квадратных цветников на рис. 18, показать, что гравер существенно искажил перспективу. (Указание: сравнить двойное отношение трех равноотстоящих точек центральной аллеи, определяемых этими цветниками, и ее бесконечно удаленной точки с двойным отношением изображений этих точек на гравюре.)

Так как двойное отношение определялось в терминах, инвариантных относительно линейных преобразований пространства V , то оно сохраняется при любых проективных преобразованиях.

Перейдем теперь к проективной теории квадрик. Как мы сейчас увидим, она проще аффинной. Это одно из проявлений совершенства проективной геометрии, завораживавшего еще математиков XIX в., которые считали, что все геометрии следует выводить из проективной.

Подмножество векторного пространства V будем называть *конусом*, если оно инвариантно относительно умножений на числа, т. е. вместе со всяким вектором содержит все пропорциональные ему векторы. (Это то же самое, что конус с вершиной в нуле в смысле определения, данного в § 5.) В частности, квадрика $X \subset V$ явля-

ется конусом в этом смысле тогда и только тогда, когда $X = X(Q)$, где Q — квадратичная функция в пространстве V . Такие квадратики будем называть *квадратичными конусами* (что слегка расходится с терминологией § 5).

Для любого конуса $X \subset V$ назовем его *проективизацией* и обозначим через PX подмножество пространства PV , образованное всеми одномерными подпространствами, содержащимися в X . Ясно, что изображением подмножества PX на аффинной карте S является пересечение $X \cap S$.

Определение 4. *Квадрикой* в пространстве PV называется проективизация квадратичного конуса в пространстве V .

Иными словами, это подмножество вида $PX(Q)$, где Q — квадратичная функция в пространстве V , при условии, что оно не пусто и не является плоскостью. Изображение проективной квадратики на аффинной карте, если оно не пусто и не является плоскостью, представляет собой аффинную квадратику. Однако тип этой квадратики зависит от аффинной карты. (Вспомните световое пятно от лампы с абажуром.)

Проективная квадрatica $PX(Q)$ называется *невырожденной*, если квадратичная функция Q невырожденна.

Замечание 4. Используя идеи доказательства теоремы 5.1, нетрудно показать, что, если только поле K содержит более пяти элементов, пересечение $X(Q) \cap S$ никогда не пусто и может быть (гипер)плоскостью только тогда, когда Q есть произведение двух линейных функций (и, следовательно, $PX(Q)$ есть объединение двух гиперплоскостей).

В однородных координатах проективная квадрatica $PX(Q)$ задается уравнением

$$Q(x_0, x_1, \dots, x_n) = \sum_{i,j=0}^n a_{ij}x_i x_j = 0 \quad (a_{ij} = a_{ji}). \quad (43)$$

Ее изображение на аффинной карте S_0 задается в аффинных координатах уравнением

$$Q(1, x_1, \dots, x_n) = 0, \quad (44)$$

а ее пересечение с бесконечно удаленной по отношению к S_0 гиперплоскостью задается в однородных координатах на этой гиперповерхности уравнением

$$Q(0, x_1, \dots, x_n) = 0. \quad (45)$$

Отметим, что всякая квадратика X на карте S_0 (а следовательно, и на любой аффинной карте) является изображением некоторой проективной квадратки \bar{X} . Уравнение квадратки \bar{X} в однородных координатах получается из уравнения квадратки X , если вставить x_0 во все линейные члены и x_0^2 — в свободный член. Из теоремы 5.1 (в тех случаях, когда она верна) следует, что квадратика \bar{X} однозначно определяется квадратикой X .

Пример 1. Рассмотрим конику $C \subset \mathbb{R}P^2$, задаваемую в однородных координатах уравнением

$$x_0^2 - x_1^2 - x_2^2 = 0.$$

Ее изображением на аффинной карте S_0 является эллипс

$$x_1^2 + x_2^2 = 1;$$

бесконечно удаленных по отношению к S_0 точек на C нет. На аффинной карте $x_0 - x_2 = 1$ та же коника изображается параболой

$$y = x_1^2,$$

где $y = x_0 + x_2$; при этом имеется одна бесконечно удаленная точка $(1 : 0 : 1)$. Наконец, на аффинной карте S_2 коника C изображается гиперболой

$$x_0^2 - x_1^2 = 1;$$

при этом имеются две бесконечно удаленные точки $(1 : 1 : 0)$ и $(1 : -1 : 0)$. Все это хорошо видно на карте S_0 , где изображение

прямой $x_0 - x_2 = 0$, бесконечно удаленной по отношению к карте $x_0 - x_2 = 1$, задается уравнением $x_2 = 1$ и касается изображения коники, а изображение прямой $x_2 = 0$, бесконечно удаленной по отношению к S_2 , пересекает изображение коники в двух точках (см. рис. 23). Таким образом, можно сказать, что парабола касается бесконечно удаленной прямой, а гипербола пересекает ее в двух точках. Нетрудно видеть, что бесконечно удаленная точка параболы соответствует ее особому направлению (см. § 5), а бесконечно удаленные точки гиперболы — ее асимптотам.

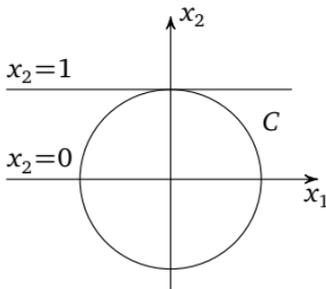


Рис. 23

Нетрудно видеть, что бесконечно удаленная точка параболы соответствует ее особому направлению (см. § 5), а бесконечно удаленные точки гиперболы — ее асимптотам.

Задача 6. Доказать, что всякий параболоид в вещественном аффинном пространстве касается бесконечно удаленной гиперплоскости.

Если квадратичная функция Q вырождена и одномерное подпространство $\langle x_0 \rangle$ содержится в ее ядре, то конус $X(Q)$ вместе со всяким одномерным подпространством $\langle x \rangle \neq \langle x_0 \rangle$ содержит двумерное подпространство $\langle x, x_0 \rangle$. Это означает, что квадратика $PX(Q)$ вместе со всякой точкой $\hat{x} \neq \hat{x}_0$ содержит прямую $\hat{x}\hat{x}_0$, т. е. является конусом с вершиной в \hat{x}_0 . Ее изображением на аффинной карте будет конус или цилиндр в зависимости от того, принадлежит точка \hat{x}_0 этой карте или нет. (Таким образом, в проективной геометрии исчезает разница между конусами и цилиндрами.)

Замечание 5. Отметим, что согласно нашей терминологии аффинное изображение вырожденной проективной квадратки может быть невырожденной аффинной квадратикой (конусом).

В случаях $K = \mathbb{C}$ или \mathbb{R} в пространстве V можно выбрать базис, в котором квадратичная функция Q имеет нормальный вид. Отсюда следует, что уравнение всякой невырожденной квадратки в комплексном проективном пространстве может быть приведено к виду

$$x_0^2 + x_1^2 + \dots + x_n^2 = 0, \quad (46)$$

а в вещественном — к виду

$$x_0^2 + x_1^2 + \dots + x_k^2 - x_{k+1}^2 - \dots - x_n^2 = 0 \quad \left(\frac{n-1}{2} \leq k < n \right). \quad (47)$$

(Неравенство $k \geq \frac{n-1}{2}$ достигается за счет возможного умножения уравнения на -1 .)

Мы видим, таким образом, что все невырожденные комплексные квадратки проективно эквивалентны, а невырожденные вещественные квадратки распадаются на $\left[\frac{n-1}{2} \right] + 1$ классов проективной эквивалентности.

Тот факт, что квадратки, задаваемые уравнением (47), при различных k проективно не эквивалентны, вытекает из теоремы 5.1 и закона инерции. Однако он проявляется и в различиях геометрического строения этих квадратик. Следующая теорема указывает одно из таких различий.

Теорема 4. Максимальная размерность плоскостей, содержащихся в вещественной проективной квадратике (47), равна $n - k - 1$.

Доказательство. Очевидно, что k -мерная плоскость Π_0 , задаваемая уравнениями

$$x_{k+1} = \dots = x_n = 0,$$

не пересекается с квадрикой (47). Так как всякая плоскость размерности $\geq n - k$ пересекается с Π_0 , то она не может целиком содержаться в квадрике.

С другой стороны, перейдя к другому базису, уравнение квадрики (47) можно записать в виде

$$y_0 y_{k+1} + y_1 y_{k+2} + \dots + y_{n-k-1} y_n + y_{n-k}^2 + \dots + y_k^2 = 0,$$

откуда видно, что $(n - k - 1)$ -мерная плоскость

$$y_0 = y_1 = \dots = y_k = 0$$

содержится в квадрике. □

В частности, квадрика (47) не содержит прямых линий тогда и только тогда, когда $k = n - 1$. Такая квадрика называется *овальной*. Одним из ее аффинных изображений является эллипсоид. При $k < n - 1$ квадрика называется *линейчатой*.

В табл. 2 перечислены невырожденные квадрики в $\mathbb{R}P^2$ и $\mathbb{R}P^3$ и их аффинные изображения. В каждом случае указывается также

Таблица 2

n	k	Название	Аффинное изображение	Бесконечно удаленная часть
2	1	коника	эллипс	\emptyset
			парабола	точка
			гипербола	пара точек
3	2	овальная квадрика	эллипсоид	\emptyset
			эллиптический параболоид	точка
			двуполостный гиперболоид	коника
	1	линейчатая квадрика	однополостный гиперболоид	коника
			гиперболический параболоид	пара прямых

бесконечно удаленная часть квадрики по отношению к соответствующей аффинной карте.

Отметим, что линейчатая квадратика в $\mathbb{R}P^3$ «соткана» из двух семейств прямых: см. рис. 24, где показаны ее аффинные изображения.

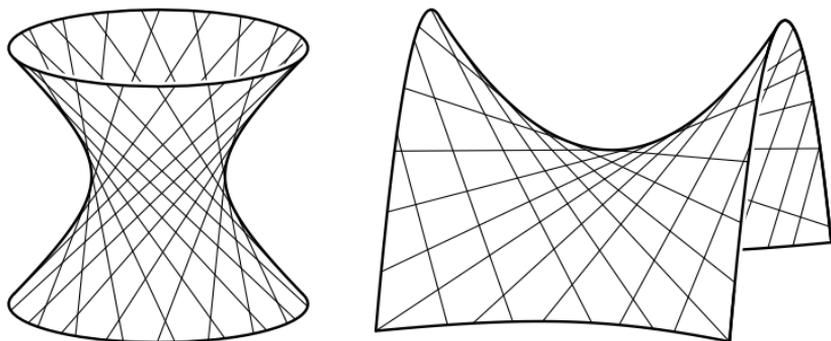


Рис. 24

Следующая теорема тесно связана с теоремой 5.2 и является ее проективным аналогом.

Теорема 5. Для любой невырожденной вещественной проективной квадрики PX группа $G(PX)$ проективных преобразований, отображающих X на себя, действует на PX транзитивно.

Доказательство. Из условия невырожденности следует, что нулевой вектор является единственной вершиной конуса X . Поэтому теорема 5.2 в применении к X означает, что группа линейных преобразований, отображающих конус X на себя, транзитивно действует на множестве его ненулевых векторов. Доказываемое утверждение получается отсюда проективизацией. \square

Задача 7. Доказать, что если PX — овальная квадратика, то группа $G(PX)$ транзитивно действует на множестве (упорядоченных) троек различных точек из PX .

Группа $G(PX)$ в случае овальной квадрики PX может служить базой для построения конформной геометрии и геометрии Лобачевского. А именно, конформная геометрия реализуется на самой квадратике PX , а геометрия Лобачевского — на ее внутренности. В обоих случаях группой, определяющей геометрию в смысле § 4.2, является группа $G(PX)$ (но действующая на разных множествах).

Задача 8. Доказать, что группа $G(PX)$ действует транзитивно на внутренности овальной квадрики PX . (Ср. задачу 5.1.)

Тензорная алгебра

Тензорная алгебра — это скорее язык, чем содержательная теория, но язык очень полезный и, более того, совершенно необходимый. Он позволяет, в частности, охватить единым взглядом и даже организовать в одну алгебру все объекты, рассматриваемые в линейной алгебре.

§ 1. Тензорное произведение векторных пространств

Начнем со следующего весьма общего понятия, охватывающего многие объекты, рассматривавшиеся нами ранее.

Пусть V_1, \dots, V_p и U — векторные пространства над полем K .
Отображение

$$\varphi: V_1 \times \dots \times V_p \rightarrow U \quad (1)$$

называется *полилинейным* (или, точнее, *p -линейным*), если оно линейно по каждому из p аргументов при фиксированных значениях других аргументов. Такие отображения образуют векторное пространство — подпространство в векторном пространстве всех отображений из $V_1 \times \dots \times V_p$ в U . Обозначим это векторное пространство через $\text{Hom}(V_1, \dots, V_p; U)$.

Если пространства V_1, \dots, V_p и U конечномерны, то и пространство $\text{Hom}(V_1, \dots, V_p; U)$ конечномерно. Более точно,

$$\dim \text{Hom}(V_1, \dots, V_p; U) = \dim V_1 \cdot \dots \cdot \dim V_p \cdot \dim U,$$

так как полилинейное отображение (1) определяется своими значениями на наборах базисных векторов пространств V_1, \dots, V_p , которые, в свою очередь, определяются своими координатами в базисе пространства U .

При $U = K$ мы получаем пространство $\text{Hom}(V_1, \dots, V_p; K)$ *полилинейных функций* на $V_1 \times \dots \times V_p$. В частности, $\text{Hom}(V; K)$ есть пространство V^* , сопряженное V .

Тензорное произведение векторных пространств V и W естественным образом возникает при рассмотрении всевозможных билинейных отображений $\varphi: V \times W \rightarrow U$. Как мы увидим, среди этих отображений имеется одно «универсальное», через которое могут быть описаны все остальные. Соответствующее пространство U и называется тензорным произведением пространств V и W .

Предложение 1. Пусть V и W — векторные пространства с базисами $\{e_i: i \in I\}$ и $\{f_j: j \in J\}$ соответственно. Следующие свойства билинейного отображения $\varphi: V \times W \rightarrow U$ эквивалентны:

1) векторы $\varphi(e_i, f_j)$ ($i \in I, j \in J$) составляют базис пространства U ;

2) каждый вектор $z \in U$ единственным образом представляется в виде $z = \sum_i \varphi(e_i, y_i)$ ($y_i \in W$);

3) каждый вектор $z \in U$ единственным образом представляется в виде $z = \sum_j \varphi(x_j, f_j)$ ($x_j \in V$).

(В случае бесконечномерных пространств предполагается, что лишь конечное число слагаемых в суммах отлично от нуля.)

Доказательство. Если $z = \sum_{i,j} z_{ij} \varphi(e_i, f_j)$, то $z = \sum_i \varphi(e_i, y_i)$, где $y_i = \sum_j z_{ij} f_j$, и наоборот. Отсюда вытекает эквивалентность свойств 1) и 2). Аналогично доказывается эквивалентность свойств 1) и 3). \square

Следствие. Выполнение условия 1) не зависит от выбора базисов в пространствах V и W .

Определение 1. Тензорным произведением двух векторных пространств V и W называется векторное пространство T вместе с билинейным отображением

$$\otimes: V \times W \rightarrow T, \quad (x, y) \mapsto x \otimes y,$$

удовлетворяющим следующему условию: если $\{e_i: i \in I\}$ и $\{f_j: j \in J\}$ — базисы пространств V и W соответственно, то $\{e_i \otimes f_j: i \in I, j \in J\}$ — базис пространства T .

Согласно доказанному выше, выполнение последнего условия не зависит от выбора базисов в V и W .

Очевидно, что тензорное произведение существует для любых векторных пространств V и W : достаточно взять векторное пространство T с базисом $\{t_{ij}: i \in I, j \in J\}$ и определить билинейное отображение $\otimes: V \times W \rightarrow T$, задав его на парах базисных векторов по формуле $e_i \otimes f_j = t_{ij}$.

Тензорное произведение единственно в следующем смысле: если (T_1, \otimes_1) и (T_2, \otimes_2) — два тензорных произведения пространств V и W , то имеется (единственный) изоморфизм $\psi: T_1 \rightarrow T_2$, удовлетворяющий условию

$$\psi(x \otimes_1 y) = x \otimes_2 y \quad (2)$$

для любых $x \in V$, $y \in W$. В самом деле, искомый изоморфизм можно построить, задав его на базисных векторах по формуле

$$\psi(e_i \otimes_1 f_j) = e_i \otimes_2 f_j.$$

По соображениям линейности условие (2) будет тогда выполняться для любых $x \in V$, $y \in W$.

Тензорное произведение векторных пространств V и W обозначается через $V \otimes W$, а при необходимости указать основное поле — через $V \otimes_K W$. Из определения следует, что в конечномерном случае

$$\dim(V \otimes W) = \dim V \cdot \dim W. \quad (3)$$

Пример 1. Рассмотрим билинейное отображение

$$\otimes: K[x] \times K[y] \rightarrow K[x, y],$$

определяемое по формуле

$$(f \otimes g)(x, y) = f(x)g(y).$$

Так как произведения $x^i \otimes y^j = x^i y^j$ ($i, j = 0, 1, 2, \dots$) составляют базис пространства $K[x, y]$, то $K[x, y] = K[x] \otimes K[y]$. Аналогично,

$$K[x_1, \dots, x_m, y_1, \dots, y_n] = K[x_1, \dots, x_m] \otimes K[y_1, \dots, y_n]. \quad (4)$$

В следующих двух примерах V и W — конечномерные векторные пространства с базисами $\{e_1, \dots, e_n\}$ и $\{f_1, \dots, f_m\}$. Через $\{\varepsilon_1, \dots, \varepsilon_n\}$ и $\{\theta_1, \dots, \theta_m\}$ обозначаются сопряженные базисы пространств V^* и W^* соответственно.

Пример 2. Для любых $\alpha \in V^*$ и $y \in W$ определим линейное отображение $\alpha \otimes y$ из V в W по формуле

$$(\alpha \otimes y)(x) = \alpha(x)y. \quad (5)$$

Мы получим билинейное отображение

$$\otimes: V^* \times W \rightarrow \text{Hom}(V; W).$$

Легко видеть, что $\varepsilon_i \otimes f_j$ — это линейное отображение, задаваемое матрицей E_{ji} . Так как такие матрицы составляют базис пространства всех матриц размера $m \times n$, то

$$\text{Hom}(V; W) = V^* \otimes W. \quad (6)$$

Пример 3. Для любых $\alpha \in V^*$ и $\beta \in W^*$ определим билинейную функцию $\alpha \otimes \beta$ на $V \times W$ по формуле

$$(\alpha \otimes \beta)(x, y) = \alpha(x)\beta(y). \quad (7)$$

Мы получим билинейное отображение

$$\otimes: V^* \times W^* \rightarrow \text{Hom}(V, W; K).$$

При этом $(\varepsilon_i \otimes \theta_j)(x, y) = x_i y_j$, где x_1, \dots, x_n и y_1, \dots, y_m — координаты векторов x и y соответственно. Так как всякая билинейная функция γ на $V \times W$ однозначно представляется в виде $\gamma(x, y) = \sum_{i,j} c_{ij} x_i y_j$, то функции $\varepsilon_i \otimes \theta_j$ составляют базис пространства $\text{Hom}(V, W; K)$. Следовательно,

$$\text{Hom}(V, W; K) = V^* \otimes W^*. \quad (8)$$

Свойство универсальности тензорного произведения, о котором говорилось в начале этого параграфа, выражается в следующем.

Предложение 2. Для произвольного билинейного отображения $\varphi: V \times W \rightarrow U$ существует единственное линейное отображение $\psi: V \otimes W \rightarrow U$, такое, что

$$\varphi(x, y) = \psi(x \otimes y) \quad (9)$$

для любых $x \in V, y \in W$.

Доказательство. Искомое линейное отображение задается на базисных векторах пространства $V \otimes W$ по формуле

$$\psi(e_i \otimes f_j) = \varphi(e_i, f_j). \quad \square$$

Каждый элемент $z \in V \otimes W$ единственным образом представляет в виде

$$z = \sum_{i,j} z_{ij} e_i \otimes f_j \quad (z_{ij} \in K). \quad (10)$$

Числа z_{ij} называются *координатами* элемента z относительно заданных базисов пространств V и W . В частности, в конечномерном случае элемент z задается матрицей (z_{ij}) размера $m \times n$, где $m = \dim V, n = \dim W$.

Элемент $z \in V \otimes W$ называется *разложимым*, если он представляется в виде

$$z = x \otimes y \quad (x \in V, y \in W). \quad (11)$$

Ясно, что если $x = \sum_i x_i e_i$, $y = \sum_j y_j f_j$, то $z_{ij} = x_i y_j$. В конечномерном случае это означает, что $\text{rk}(z_{ij}) \leq 1$. Таким образом, разложимые элементы, хотя они и порождают пространство $V \otimes W$, составляют весьма малую его часть (за исключением случаев, когда V или W одномерны).

Задача 1. Доказать, что представление ненулевого разложимого элемента $z \in V \otimes W$ в виде (11) единственно с точностью до замены $x \mapsto \lambda x$, $y \mapsto \lambda^{-1} y$ ($\lambda \in K^*$).

Часто бывают полезны и другие представления элемента тензорного произведения, вытекающие из предложения 1. А именно, всякий элемент $z \in V \otimes W$ единственным образом представляется в виде

$$z = \sum_i e_i \otimes y_i \quad (y_i \in W), \quad (12)$$

а также в виде

$$z = \sum_j x_j \otimes f_j \quad (x_j \in V). \quad (13)$$

Задача 2. Доказать, что всякий элемент $z \in V \otimes W$ представляется в виде

$$z = \sum_{k=1}^r v_k \otimes w_k, \quad (14)$$

где векторы $v_1, \dots, v_r \in V$, а также векторы $w_1, \dots, w_r \in W$ линейно независимы. Такое представление единственно с точностью до замены

$$v_k \mapsto \sum_l a_{kl} v_l, \quad w_k \mapsto \sum_l b_{kl} w_l,$$

где $A = (a_{kl})$ и $B = (b_{kl})$ — невырожденные квадратные матрицы порядка r , связанные соотношением $A^T B = E$. Число r равно рангу матрицы координат элемента z .

Важным применением тензорного умножения является операция расширения основного поля, с простейшим частным случаем которой — комплексификацией вещественного векторного пространства — мы уже встречались в § 6.2.

Пусть V — векторное пространство над полем K и L — какое-либо расширение поля K , т. е. поле, содержащее K в качестве подполя. Рассматривая L как векторное пространство над K , мы можем образовать тензорное произведение

$$V(L) = L \otimes V.$$

Согласно определению, это векторное пространство над K . Однако его можно превратить в векторное пространство над L , определив умножение на элементы поля L по правилу

$$\lambda(\mu \otimes v) = \lambda\mu \otimes v \quad (\lambda, \mu \in L, v \in V).$$

Исходное пространство V можно считать вложенным в $V(L)$, отождествив каждый вектор $v \in V$ с вектором $1 \otimes v \in V(L)$. При таком соглашении $\lambda \otimes v = \lambda v$. Рассматривая разложение элемента $V(L)$ по базису второго множителя, мы получаем, что всякий базис пространства V над K является базисом пространства $V(L)$ над L . Однако смысл расширения основного поля заключается в том, что в пространстве $V(L)$ существуют и другие базисы, в которых изучаемые объекты (например, линейные операторы) могут иметь более простой вид.

С другой стороны, если $\{\theta_i : i \in I\}$ — базис L над K , то всякий вектор пространства $V(L)$ однозначно представляется в виде $\sum_i \theta_i v_i$, где v_i ($i \in I$) — какие-то векторы пространства V (лишь конечное число которых отлично от нуля). Например, всякий вектор комплексификации $V(\mathbb{C})$ вещественного векторного пространства V однозначно представляется в виде $x + iy$, где $x, y \in V$.

Операция тензорного умножения векторных пространств в определенном смысле коммутативна и ассоциативна. А именно, для любых векторных пространств V и W имеется изоморфизм

$$V \otimes W \xrightarrow{\sim} W \otimes V, \quad (15)$$

при котором $x \otimes y$ ($x \in V, y \in W$) переходит в $y \otimes x$. В самом деле, искомым изоморфизм определяется условием, что базисные векторы $e_i \otimes f_j$ пространства $V \otimes W$ переходят в соответствующие базисные векторы $f_j \otimes e_i$ пространства $W \otimes V$. Аналогично, для любых векторных пространств U, V, W имеется изоморфизм

$$(U \otimes V) \otimes W \xrightarrow{\sim} U \otimes (V \otimes W), \quad (16)$$

при котором $(x \otimes y) \otimes z$ ($x \in U, y \in V, z \in W$) переходит в $x \otimes (y \otimes z)$.

Отождествляя пространства $(U \otimes V) \otimes W$ и $U \otimes (V \otimes W)$ при помощи изоморфизма (16), мы можем говорить о тензорном произведении любого конечного числа векторных пространств V_1, V_2, \dots, V_p , не указывая расстановки скобок. Индукция по p показывает, что всевозможные тензорные произведения базисных векторов пространств V_1, \dots, V_p составляют базис пространства $V_1 \otimes \dots \otimes V_p$. С другой стороны, это свойство можно принять за определение $V_1 \otimes \dots \otimes V_p$, т. е. можно сразу определить тензорное произведение нескольких векторных пространств так же, как это было сделано для двух пространств (заменив билинейное отображение p -линейным).

В силу предложения 2 имеется изоморфизм

$$\text{Hom}(V \otimes W; U) \cong \text{Hom}(V, W; U), \quad (17)$$

переводящий линейное отображение $\psi: V \otimes W \rightarrow U$ в билинейное отображение $\varphi: V \times W \rightarrow U$, определяемое равенством (9). В частности, при $U = K$ получаем изоморфизм

$$(V \otimes W)^* \cong \text{Hom}(V, W; K). \quad (18)$$

Предложение 2 тривиально обобщается на любое конечное число векторных пространств (вместо двух пространств V и W). Это дает изоморфизм

$$\text{Hom}(V_1 \otimes \dots \otimes V_p; U) \cong \text{Hom}(V_1, \dots, V_p; U), \quad (19)$$

переводящий линейное отображение $\psi: V_1 \otimes \dots \otimes V_p \rightarrow U$ в p -линейное отображение $\varphi: V_1 \times \dots \times V_p \rightarrow U$, определяемое равенством

$$\varphi(x_1, \dots, x_p) = \psi(x_1 \otimes \dots \otimes x_p). \quad (20)$$

В частности, при $U = K$ получаем изоморфизм

$$(V_1 \otimes \dots \otimes V_p)^* \cong \text{Hom}(V_1, \dots, V_p; K). \quad (21)$$

Элементы вида $x_1 \otimes \dots \otimes x_p$ называются *разложимыми элементами* тензорного произведения $V_1 \otimes \dots \otimes V_p$. Наличие канонического изоморфизма (19) эквивалентно следующему основному принципу тензорной алгебры, позволяющему определять линейные отображения тензорного произведения, задавая их на разложимых элементах: для любого p -линейного отображения $\varphi: V_1 \times \dots \times V_p \rightarrow U$ существует единственное линейное отображение $\psi: V_1 \otimes \dots \otimes V_p \rightarrow U$, удовлетворяющее условию (20).

Для тензорных произведений конечномерных пространств имеются и другие естественные изоморфизмы, играющие важную роль в тензорной алгебре.

Прежде всего, можно обобщить пример 3 на любое число конечномерных векторных пространств V_1, \dots, V_p . Полагая

$$(\alpha_1 \otimes \dots \otimes \alpha_p)(x_1, \dots, x_p) = \alpha_1(x_1) \dots \alpha_p(x_p) \quad (22)$$

при $\alpha_1 \in V_1^*, \dots, \alpha_p \in V_p^*$, мы получаем, что

$$\text{Hom}(V_1, \dots, V_p; K) = V_1^* \otimes \dots \otimes V_p^*. \quad (23)$$

В сочетании с изоморфизмом (21) это дает изоморфизм

$$V_1^* \otimes \dots \otimes V_p^* \xrightarrow{\sim} (V_1 \otimes \dots \otimes V_p)^*. \quad (24)$$

Комбинируя равенство (6) с изоморфизмами (19) и (24), мы получаем для любых конечномерных векторных пространств V_1, \dots, V_p и U изоморфизм

$$V_1^* \otimes \dots \otimes V_p^* \otimes U \xrightarrow{\sim} \text{Hom}(V_1, \dots, V_p; U). \quad (25)$$

Имея в виду построенные выше естественные изоморфизмы, обычно отождествляют соответствующие векторные пространства, т. е. считают, что $V \otimes W = W \otimes V$, $(U \otimes V) \otimes W = U \otimes (V \otimes W)$, $V^* \otimes W^* \otimes U = \text{Hom}(V, W; U)$ (для конечномерных пространств) и т. д.

Для любых линейных операторов $\mathcal{A} \in L(V)$ и $\mathcal{B} \in L(W)$ можно определить линейный оператор $\mathcal{A} \otimes \mathcal{B} \in L(V \otimes W)$, задав его на разложимых элементах по формуле

$$(\mathcal{A} \otimes \mathcal{B})(x \otimes y) = \mathcal{A}x \otimes \mathcal{B}y. \quad (26)$$

Оператор $\mathcal{A} \otimes \mathcal{B}$ называется *тензорным произведением* операторов \mathcal{A} и \mathcal{B} .

Пусть $A = (a_{ij})$ — матрица оператора \mathcal{A} в базисе $\{e_1, \dots, e_n\}$ пространства V и $B = (b_{kl})$ — матрица оператора \mathcal{B} в базисе $\{f_1, \dots, f_m\}$ пространства W . Тогда матрица оператора $\mathcal{A} \otimes \mathcal{B}$ в базисе $\{e_1 \otimes f_1, e_1 \otimes f_2, \dots, e_1 \otimes f_m, e_2 \otimes f_1, e_2 \otimes f_2, \dots, e_2 \otimes f_m, \dots, e_n \otimes f_1, e_n \otimes f_2, \dots, e_n \otimes f_m\}$ пространства $V \otimes W$ имеет вид

$$\begin{pmatrix} a_{11}B & a_{12}B & \dots & a_{1n}B \\ a_{21}B & a_{22}B & \dots & a_{2n}B \\ \dots & \dots & \dots & \dots \\ a_{n1}B & a_{n2}B & \dots & a_{nn}B \end{pmatrix}. \quad (27)$$

Она называется *тензорным произведением* матриц A и B и обозначается через $A \otimes B$.

Легко видеть, что $\text{tr } A \otimes B = \text{tr } A \cdot \text{tr } B$ и, значит,

$$\text{tr } \mathcal{A} \otimes \mathcal{B} = \text{tr } \mathcal{A} \cdot \text{tr } \mathcal{B}. \quad (28)$$

Задача 3. Доказать, что

$$\det \mathcal{A} \otimes \mathcal{B} = (\det \mathcal{A})^m (\det \mathcal{B})^n. \quad (29)$$

Задача 4. Предположим, что характеристический многочлен оператора \mathcal{A} имеет (с учетом кратностей) n корней $\lambda_1, \dots, \lambda_n$, а характеристический многочлен оператора \mathcal{B} имеет m корней μ_1, \dots, μ_m . Доказать, что тогда характеристический многочлен оператора $\mathcal{A} \otimes \mathcal{B}$ имеет nm корней $\lambda_i \mu_j$ ($i = 1, \dots, n; j = 1, \dots, m$). Вывести отсюда (при сделанных предположениях) формулы (28) и (29).

Задача 5. Доказать, что (для конечномерных пространств V и W) пространство $L(V \otimes W)$ является тензорным произведением пространств $L(V)$ и $L(W)$ относительно определенного выше тензорного умножения линейных операторов.

Аналогичным образом определяется тензорное произведение нескольких линейных операторов.

§ 2. Тензорная алгебра векторного пространства

В этом параграфе V есть n -мерное векторное пространство.

Пространство

$$T_q^p(V) = \underbrace{V \otimes \dots \otimes V}_p \otimes \underbrace{V^* \otimes \dots \otimes V^*}_q$$

называется пространством *тензоров типа* (p, q) на V . (Пространство $T_0^0(V)$ полагают равным K .) Очевидно, что $\dim T_q^p(V) = n^{p+q}$. Имеем $T_0^1(V) = V$, $T_1^0(V) = V^*$. Более общо,

$$T_q^0(V) = \text{Hom}(\underbrace{V, \dots, V}_q; K), \quad (30)$$

$$T_q^1(V) = \text{Hom}(\underbrace{V, \dots, V}_q; V). \quad (31)$$

В частности, тензоры типа $(0, 2)$ — это билинейные функции, тензоры типа $(1, 1)$ — это линейные операторы, а тензоры типа $(1, 2)$ — это билинейные операции (структуры алгебр) на V .

Тензорное умножение определяет билинейную операцию

$$\otimes: T_q^p(V) \times T_s^r(V) \rightarrow T_{q+s}^{p+r}(V)$$

таким образом, что

$$\begin{aligned} (x_1 \otimes \dots \otimes x_p \otimes \alpha_1 \otimes \dots \otimes \alpha_q) \otimes (x_{p+1} \otimes \dots \otimes x_{p+r} \otimes \alpha_{q+1} \otimes \dots \otimes \alpha_{q+s}) = \\ = x_1 \otimes \dots \otimes x_{p+r} \otimes \alpha_1 \otimes \dots \otimes \alpha_{q+s}. \end{aligned}$$

Пример 1. Пространство

$$T_2^2(V) = V \otimes V \otimes V^* \otimes V^* = (V \otimes V) \otimes (V \otimes V)^*$$

можно трактовать как $L(V \otimes V)$. При этом тензорное умножение

$$T_1^1(V) \times T_1^1(V) \rightarrow T_2^2(V)$$

совпадает с тензорным умножением линейных операторов в смысле § 1. В самом деле, в силу билинейности обоих умножений достаточно проверить это для разложимых линейных операторов. Пусть $\mathcal{A} = u \otimes \alpha$, $\mathcal{B} = v \otimes \beta$ ($u, v \in V$, $\alpha, \beta \in V^*$), и пусть $\mathcal{A} \otimes \mathcal{B}$ — тензорное произведение операторов \mathcal{A} и \mathcal{B} в смысле § 1. Учитывая, что $(\alpha \otimes \beta)(x \otimes y) = \alpha(x)\beta(y)$ (см. пример 1.3 и определение изоморфизма (18)), получаем:

$$\begin{aligned} (\mathcal{A} \otimes \mathcal{B})(x \otimes y) &= \mathcal{A}x \otimes \mathcal{B}y = \alpha(x)\beta(y)u \otimes v = \\ &= ((\alpha \otimes \beta)(x \otimes y))u \otimes v = ((u \otimes v) \otimes (\alpha \otimes \beta))(x \otimes y). \end{aligned}$$

Следовательно,

$$\mathcal{A} \otimes \mathcal{B} = u \otimes v \otimes \alpha \otimes \beta,$$

что и требовалось доказать.

Другая важная операция над тензорами — это *свертка*, представляющая собой линейное отображение

$$T_q^p(V) \rightarrow T_{q-1}^{p-1}(V) \quad (p, q > 0),$$

определяемое следующим образом. Рассмотрим отображение

$$\underbrace{V \times \dots \times V}_p \times \underbrace{V^* \times \dots \times V^*}_q \rightarrow T_{q-1}^{p-1}(V),$$

$$(x_1, \dots, x_p, \alpha_1, \dots, \alpha_q) \mapsto \alpha_1(x_1)(x_2 \otimes \dots \otimes x_p \otimes \alpha_2 \otimes \dots \otimes \alpha_q).$$

Очевидно, что оно полилинейно. Следовательно, существует линейное отображение $T_q^p(V) \rightarrow T_{q-1}^{p-1}(V)$, при котором

$$x_1 \otimes \dots \otimes x_p \otimes \alpha_1 \otimes \dots \otimes \alpha_q \mapsto \alpha_1(x_1)(x_2 \otimes \dots \otimes x_p \otimes \alpha_2 \otimes \dots \otimes \alpha_q).$$

Это и есть свертка.

В данном определении свертка производилась «по первым множителям» в произведениях $\underbrace{V \otimes \dots \otimes V}_p$ и $\underbrace{V^* \otimes \dots \otimes V^*}_q$, тензорным произведением которых является $T_q^p(V)$. Совершенно так же определяется свертка по любой паре множителей.

Пример 2. Свертка линейного оператора (как тензора типа $(1, 1)$) — это его след. Действительно, в силу линейности достаточно проверить это утверждение для разложимых операторов, т. е. операторов вида $x \otimes \alpha$ ($x \in V$, $\alpha \in V^*$). Оператор такого вида равен нулю на $(n-1)$ -мерном подпространстве $\text{Ker } \alpha$ и действует как умножение на $\alpha(x)$ на факторпространстве $V/\text{Ker } \alpha$. Следовательно, его след равен $\alpha(x)$, что совпадает со сверткой.

Пример 3. Свертка тензорного произведения линейного оператора \mathcal{A} и вектора x по второму множителю V и первому (единственному) множителю V^* равна вектору $\mathcal{A}x$. Действительно, для разложимого оператора $\mathcal{A} = u \otimes \alpha$ результат указанной свертки есть $\alpha(x)u$, что совпадает с $\mathcal{A}x$.

Пример 4. Свертка тензорного произведения линейных операторов \mathcal{A} и \mathcal{B} по второму множителю V и первому множителю V^* равна обычному произведению $\mathcal{A}\mathcal{B}$ операторов \mathcal{A} и \mathcal{B} . Действительно, для разложимых операторов $\mathcal{A} = u \otimes \alpha$ и $\mathcal{B} = v \otimes \beta$ результат указанной свертки есть оператор $\alpha(v)u \otimes \beta$, переводящий каждый вектор $x \in V$ в вектор $\alpha(v)\beta(x)u$; с другой стороны,

$$\mathcal{A}\mathcal{B}x = \beta(x)\mathcal{A}v = \alpha(v)\beta(x)u.$$

Пример 5. Как следует из формулы (7), свертка тензорного произведения билинейной функции α и двух векторов x и y по двум парам множителей равна $\alpha(x, y)$ или $\alpha(y, x)$ в зависимости от того, как комбинируются сворачиваемые множители.

Свертку тензорного произведения тензоров T и U часто называют сверткой тензора T с тензором U (по указанным индексам).

Пусть $\{e_i\}$ — базис пространства V и $\{\varepsilon_j\}$ — сопряженный базис пространства V^* . Тогда $\{e_{i_1} \otimes \dots \otimes e_{i_p} \otimes \varepsilon_{j_1} \otimes \dots \otimes \varepsilon_{j_q}\}$ — базис пространства $T_q^p(V)$. Любой тензор T типа (p, q) может быть выражен

через этот базис:

$$T = \sum_{i_1, \dots, i_p, j_1, \dots, j_q} T_{i_1 \dots i_p j_1 \dots j_q} e_{i_1} \otimes \dots \otimes e_{i_p} \otimes \varepsilon_{j_1} \otimes \dots \otimes \varepsilon_{j_q}.$$

Числа $T_{i_1 \dots i_p j_1 \dots j_q}$ называются *координатами тензора* T в базисе $\{e_i\}$ пространства V .

Пример 6. Координаты линейного оператора как тензора типа $(1, 1)$ — это в точности элементы матрицы этого оператора. В самом деле, если $\mathcal{A} = \sum_{i,j} \mathcal{A}_{ij} e_i \otimes \varepsilon_j$, то $\mathcal{A} e_j = \sum_i \mathcal{A}_{ij} e_i$.

Пример 7. Аналогично, координаты билинейной функции как тензора типа $(0, 2)$ — это элементы матрицы этой функции.

В символике, изобретенной Эйнштейном, используются как нижние, так и верхние индексы, причем базисные векторы пространства V нумеруются нижними индексами, а базисные векторы пространства V^* — верхними. Соответствующие индексы у координат тензора пишутся, напротив, вверху и внизу соответственно. Если в каком-либо произведении какой-либо индекс встречается дважды, причем один раз вверху, а другой — внизу (другие повторения не допускаются), то предполагается суммирование по этому индексу. Так, предыдущая формула записывается в виде

$$T = T_{j_1 \dots j_q}^{i_1 \dots i_p} e_{i_1} \otimes \dots \otimes e_{i_p} \otimes \varepsilon^{j_1} \otimes \dots \otimes \varepsilon^{j_q}. \quad (32)$$

Пример 8. Координаты образа вектора x при действии линейного оператора \mathcal{A} записываются в этой символике формулой

$$(\mathcal{A}x)^i = \mathcal{A}^i_j x^j.$$

Пример 9. Произведение линейных операторов \mathcal{A} и \mathcal{B} записывается формулой

$$(\mathcal{A}\mathcal{B})^i_k = \mathcal{A}^i_j \mathcal{B}^j_k.$$

Координаты тензорного произведения $T \otimes U$ тензоров $T \in T_q^p(V)$ и $U \in T_s^r(V)$ суть произведения координат множителей:

$$(T \otimes U)_{j_1 \dots j_{q+s}}^{i_1 \dots i_{p+r}} = T_{j_1 \dots j_q}^{i_1 \dots i_p} U_{j_{q+1} \dots j_{q+s}}^{i_{p+1} \dots i_{p+r}}.$$

Координаты свертки S тензора $T \in T_q^p(V)$ по первой паре множителей (или, как говорят, по первой паре индексов) находятся по формуле

$$S_{j_2 \dots j_q}^{i_2 \dots i_p} = T_{kj_2 \dots j_q}^{ki_2 \dots i_p}.$$

Это следует из равенства (32), если учесть, что свертка произведения $e_{i_1} \otimes \dots \otimes e_{i_r} \otimes \varepsilon^{j_1} \otimes \dots \otimes \varepsilon^{j_q}$ равна $\delta_{i_1}^{j_1} e_{i_2} \otimes \dots \otimes e_{i_r} \otimes \varepsilon^{j_2} \otimes \dots \otimes \varepsilon^{j_q}$ (где δ_i^j — символ Кронекера). Аналогично находятся координаты свертки тензора T по любой паре индексов.

В евклидовом векторном пространстве V имеется выделенный тензор $g \in T_2^0(V)$, определяющий скалярное умножение. Он называется *метрическим тензором* пространства V . Свертка метрического тензора с любым тензором $T \in T_q^p(V)$ по любому индексу тензора g и первому верхнему индексу тензора T есть тензор $\tilde{T} \in T_{q+1}^{p-1}(V)$, координаты которого находятся по формуле

$$\tilde{T}_{j_1 \dots j_q}^{i_2 \dots i_p} = g_{jk} T_{j_1 \dots j_q}^{ki_2 \dots i_p}.$$

Переход от тензора T к тензору \tilde{T} называется *спуском* первого верхнего индекса тензора T . Аналогично определяется спуск любого верхнего индекса.

В ортонормированном базисе пространства V имеем $g_{jk} = \delta_{jk}$, откуда

$$\tilde{T}_{j_1 \dots j_q}^{i_2 \dots i_p} = T_{j_1 \dots j_q}^{ji_2 \dots i_p}.$$

Из этого можно сделать два вывода. Во-первых, операция спуска индекса обратима. Обратная операция называется *подъемом индекса*. Во-вторых, если ограничиться ортонормированными базисами, в евклидовом векторном пространстве исчезает разница между верхними и нижними индексами тензоров.

Пример 10. При спуске индекса вектора $u \in V$ получается линейная функция

$$\alpha(x) = g_{jk} x^j u^k = (x, u).$$

Тем самым устанавливается уже известный нам канонический изоморфизм между евклидовым пространством V и его сопряженным пространством V^* .

Пример 11. При спуске индекса линейного оператора \mathcal{A} получается билинейная функция

$$\alpha(x, y) = g_{jk} x^j \mathcal{A}_l^k y^l = (x, \mathcal{A}y).$$

Тем самым устанавливается уже известный нам канонический изоморфизм между пространствами линейных операторов и билинейных функций в евклидовом пространстве (см. § 6.3).

Тензоры типа $(p, 0)$ называются *контравариантными тензорами* степени p . Положим

$$T^p(V) = T_0^p(V).$$

Пространства $T^0(V) = K$, $T^1(V) = V$, $T^2(V)$, ... можно организовать в алгебру. Для этого нам понадобится понятие внешней прямой суммы векторных пространств.

С разложением векторного пространства в прямую сумму подпространств мы уже встречались в § 5.1. Соответствующее определение может быть дано следующим образом.

Определение 1. Говорят, что векторное пространство V разлагается в *прямую сумму* подпространств V_1, \dots, V_k , если каждый элемент $x \in V$ единственным образом представляется в виде $x = x_1 + \dots + x_k$, где $x_i \in V_i$. При этом пишут

$$V = V_1 \oplus \dots \oplus V_k.$$

В случае двух подпространств V_1, V_2 единственность представления любого элемента $x \in V$ в виде $x = x_1 + x_2$ ($x_1 \in V_1, x_2 \in V_2$) равносильно тому, что $V_1 \cap V_2 = 0$.

Имеется другой подход к понятию прямой суммы, при котором пространства V_1, \dots, V_k заранее не предполагаются вложенными в какое-то одно пространство.

Определение 2. *Прямой суммой* векторных пространств V_1, \dots, V_k называется векторное пространство $V_1 \oplus \dots \oplus V_k$, составленное из всех последовательностей (x_1, \dots, x_k) , где $x_i \in V_i$, с покомпонентными операциями сложения и умножения на элементы основного поля K .

Более подробно, операции в $V_1 \oplus \dots \oplus V_k$ определяются следующим образом:

$$\begin{aligned} (x_1, \dots, x_k) + (y_1, \dots, y_k) &= (x_1 + y_1, \dots, x_k + y_k), \\ \lambda(x_1, \dots, x_k) &= (\lambda x_1, \dots, \lambda x_k). \end{aligned}$$

Тот факт, что при этом действительно получается векторное пространство, очевиден. В частности, его нулем является последовательность $(0, \dots, 0)$.

Прямую сумму в смысле определения 1 называют *внутренней*, а в смысле определения 2 — *внешней*. Однако эти два понятия тесно связаны друг с другом.

А именно, рассматривая последовательности вида $(0, \dots, x, \dots, 0)$, где $x \in V_i$ стоит на i -м месте, мы видим, что операции над ними сводятся к соответствующим операциям над i -ми компонентами. отождествляя элемент $x \in V_i$ с такой последовательностью, мы получаем вложение пространства V_i в качестве подпространства в пространство $V_1 \oplus \dots \oplus V_k$. При этом каждый элемент из $V_1 \oplus \dots \oplus V_k$ единственным образом представляется в виде суммы элементов из этих подпространств. Это означает, что пространство $V_1 \oplus \dots \oplus V_k$ есть прямая сумма подпространств V_1, \dots, V_k . Имея в виду указанное отождествление, элемент (x_1, \dots, x_k) внешней прямой суммы $V_1 \oplus \dots \oplus V_k$ часто записывают как $x_1 + \dots + x_k$.

Обратно, пусть какое-то векторное пространство V разложено в прямую сумму своих подпространств V_1, \dots, V_k . образуем внешнюю прямую сумму $V_1 \oplus \dots \oplus V_k$. Тогда отображение

$$V_1 \oplus \dots \oplus V_k \rightarrow V, \quad (x_1, \dots, x_k) \mapsto x_1 + \dots + x_k,$$

является изоморфизмом векторных пространств.

Все предыдущее можно распространить на случай бесконечного числа слагаемых V_1, V_2, \dots , если рассматривать не все последовательности (x_1, x_2, \dots) , где $x_i \in V_i$, а только *финитные*, т. е. такие, в которых лишь конечное число членов отлично от нуля.

Вернемся к построению тензорной алгебры. Рассмотрим бесконечную прямую сумму

$$T(V) = \bigoplus_{p=0}^{\infty} T^p(V). \quad (33)$$

Так как

$$T^p(V) \otimes T^q(V) \subset T^{p+q}(V),$$

то тензорное умножение определяет в $T(V)$ структуру градуированной алгебры. Эта алгебра называется *тензорной алгеброй* пространства V . Отметим, что она ассоциативна (но не коммутативна) и обладает единицей, каковой является единица поля $K = T^0(V)$.

Аналогично, тензоры типа $(0, p)$ называются *ковариантными тензорами* степени p . Положим $T_p(V) = T_p^0(V)$. Алгебра

$$T_*(V) = \bigoplus_{p=0}^{\infty} T_p(V)$$

называется *алгеброй полилинейных функций* на V . Тензорное умножение полилинейных функций имеет простую интерпретацию.

А именно, значения $(p+q)$ -линейной функции $\alpha \otimes \beta$ ($\alpha \in T_p(V)$, $\beta \in T_q(V)$) находятся по формуле

$$(\alpha \otimes \beta)(x_1, \dots, x_{p+q}) = \alpha(x_1, \dots, x_p) \beta(x_{p+1}, \dots, x_{p+q}). \quad (34)$$

В самом деле, по соображениям линейности достаточно проверить справедливость этой формулы для $\alpha = \alpha_1 \otimes \dots \otimes \alpha_p$ ($\alpha_1, \dots, \alpha_p \in V^*$) и $\beta = \beta_1 \otimes \dots \otimes \beta_q$ ($\beta_1, \dots, \beta_q \in V^*$); но в этом случае она легко следует из формулы (22).

С другой стороны, так как

$$T_p(V) = T^p(V^*),$$

то алгебру ковариантных тензоров можно понимать как тензорную алгебру пространства V^* .

Согласно основному принципу тензорной алгебры (см. §1) всякое p -линейное отображение

$$\varphi: \underbrace{V \times \dots \times V}_p \rightarrow U \quad (35)$$

«пропускается» через $T^p(V)$ в том смысле, что существует такое (однозначно определенное) линейное отображение

$$\psi: T^p(V) \rightarrow U, \quad (36)$$

что

$$\varphi(x_1, \dots, x_p) = \psi(x_1 \otimes \dots \otimes x_p) \quad (37)$$

для любых $x_1, \dots, x_p \in V$. При $U = K$ это дает изоморфизм

$$T_p(V) \xrightarrow{\sim} (T^p(V))^* \quad (38)$$

(частный случай изоморфизма (21)).

Если рассматривать только симметрические или же только кососимметрические полилинейные отображения, то мы придем к понятию симметрической или, соответственно, внешней степени пространства V . Этому посвящены следующие два параграфа.

§ 3. Симметрическая алгебра

Определение 1. Полилинейное отображение (35) называется *симметрическим*, если

$$\varphi(x_{i_1}, \dots, x_{i_p}) = \varphi(x_1, \dots, x_p)$$

для любой перестановки (i_1, \dots, i_p) чисел $1, \dots, p$.

Очевидно, что в этом определении можно ограничиться перестановками двух аргументов.

При $U = K$ оно превращается в определение *симметрической полилинейной функции*.

Пусть $\{e_1, \dots, e_n\}$ — базис пространства V .

Определение 2. Векторное пространство S вместе с симметрическим p -линейным отображением

$$\underbrace{V \times \dots \times V}_p \rightarrow S, \quad (x_1, \dots, x_p) \mapsto x_1 \vee \dots \vee x_p, \quad (39)$$

называется p -й *симметрической степенью* пространства V , если векторы $e_{i_1} \vee \dots \vee e_{i_p}$ с $i_1 \leq \dots \leq i_p$ составляют базис пространства S .

Подчеркнем, что выражение $x_1 \vee \dots \vee x_p$ в (39) следует понимать как единое целое. Это просто способ обозначения образа элемента (x_1, \dots, x_p) .

Данное определение не зависит от выбора базиса пространства V . В самом деле, если $\{e'_1, \dots, e'_n\}$ — другой базис, то векторы $e'_{j_1} \vee \dots \vee e'_{j_p}$ с $j_1 \leq \dots \leq j_p$ также составляют базис пространства S , так как их число равно числу векторов $e_{i_1} \vee \dots \vee e_{i_p}$ с $i_1 \leq \dots \leq i_p$ и последние через них линейно выражаются.

Симметрическая степень существует: достаточно взять векторное пространство S с базисом $\{s_{i_1 \dots i_p} : i_1 \leq \dots \leq i_p\}$ и определить p -линейное отображение (39), задав его на наборах базисных векторов пространства V по формуле $e_{i_1} \vee \dots \vee e_{i_p} = s_{j_1 \dots j_p}$, где j_1, \dots, j_p — числа i_1, \dots, i_p , расположенные в порядке неубывания.

Симметрическая степень единственна в следующем смысле: если (S_1, \vee_1) и (S_2, \vee_2) — две p -е симметрические степени пространства V , то имеется (единственный) изоморфизм $\psi : S_1 \rightarrow S_2$, удовлетворяющий условию

$$\psi(x_1 \vee_1 \dots \vee_1 x_p) = x_1 \vee_2 \dots \vee_2 x_p$$

для любых $x_1, \dots, x_p \in V$. Искомый изоморфизм можно построить, задав его на базисных векторах по формуле

$$\psi(e_{i_1} \vee_1 \dots \vee_1 e_{i_p}) = e_{i_1} \vee_2 \dots \vee_2 e_{i_p} \quad (i_1 \leq \dots \leq i_p).$$

Симметрическая степень пространства V обозначается через $S^p(V)$.

Следующее предложение выражает свойство универсальности симметрической степени, аналогичное свойству универсальности тензорного произведения (см. предложение 1.2).

Предложение 1. Для произвольного симметрического p -линейного отображения (35) существует единственное линейное отображение $\psi: S^p(V) \rightarrow U$, такое, что

$$\varphi(x_1, \dots, x_p) = \psi(x_1 \vee \dots \vee x_p) \quad (40)$$

для любых $x_1, \dots, x_p \in V$.

Доказательство. Искомое линейное отображение задается на базисных векторах пространства $S^p(V)$ по формуле

$$\psi(e_{i_1} \vee \dots \vee e_{i_p}) = \varphi(e_{i_1}, \dots, e_{i_p}) \quad (i_1 \leq \dots \leq i_p).$$

Ввиду симметричности отображения φ эта формула автоматически будет выполняться при любых i_1, \dots, i_p , а уже отсюда по линейности вытекает (40). \square

Элементы вида $x_1 \vee \dots \vee x_p$ ($x_1, \dots, x_p \in V$) симметрической степени $S^p(V)$ называются *разложимыми*. Предложение 1 позволяет определять линейные отображения пространства $S^p(V)$, задавая их на разложимых элементах так, чтобы выполнялись условия полилинейности и симметрии относительно множителей x_1, \dots, x_p .

В частности, имеется билинейное отображение

$$\vee: S^p(V) \times S^q(V) \rightarrow S^{p+q}(V),$$

задаваемое на разложимых элементах по формуле

$$(x_1 \vee \dots \vee x_p) \vee (x_{p+1} \vee \dots \vee x_{p+q}) = x_1 \vee \dots \vee x_{p+q}. \quad (41)$$

Рассмотрим прямую сумму

$$S(V) = \bigoplus_{p=0}^{\infty} S^p(V).$$

Определенная выше операция \vee превращает $S(V)$ в градуированную алгебру. Эта алгебра называется *симметрической алгеброй* пространства V . Очевидно, что она ассоциативна, коммутативна и обладает единицей (каковой является единица поля $K = S^0V$). Из определения (41) следует, что всякий разложимый элемент $x_1 \vee \dots \vee x_p \in S^p(V)$ совпадает с произведением элементов x_1, \dots, x_p в алгебре $S(V)$.

Симметрическая алгебра векторного пространства на самом деле не является новым для нас объектом. В ней можно узнать алгебру многочленов. А именно, поставив в соответствие каждому произведению $e_{i_1} \vee \dots \vee e_{i_p}$ ($i_1 \leq \dots \leq i_p$) одночлен $u_{i_1} \dots u_{i_p}$ от переменных u_1, \dots, u_n , мы получим изоморфизм алгебры $S(V)$ на алгебру $K[u_1, \dots, u_n]$.

Более того, если рассматривать e_1, \dots, e_n как координатные функции на сопряженном пространстве V^* , то любой элемент алгебры $S(V)$ как многочлен от e_1, \dots, e_n будет определять некоторую функцию на V^* . В этом смысле можно сказать, что алгебра $S(V)$ есть алгебра многочленов на V^* (хотя в случае конечного поля ее элементы нельзя отождествлять с определяемыми ими функциями). Соответственно этому алгебра $S(V^*)$ есть алгебра многочленов на V .

Если $\text{char } K = 0$, то пространство $S^p(V)$ можно отождествить с подпространством симметрических тензоров в $T^p(V)$.

А именно, для любой перестановки $\sigma \in S_p$ определим линейное преобразование $T \mapsto T^\sigma$ пространства $T^p(V)$, задав его на разложимых тензорах формулой

$$(x_1 \otimes \dots \otimes x_p)^\sigma = x_{\sigma(1)} \otimes \dots \otimes x_{\sigma(p)}. \quad (42)$$

Отметим, что

$$\begin{aligned} ((x_1 \otimes \dots \otimes x_p)^\sigma)^\tau &= (x_{\sigma(1)} \otimes \dots \otimes x_{\sigma(p)})^\tau = \\ &= x_{\sigma\tau(1)} \otimes \dots \otimes x_{\sigma\tau(p)} = (x_1 \otimes \dots \otimes x_p)^{\sigma\tau} \end{aligned}$$

и, следовательно,

$$(T^\sigma)^\tau = T^{\sigma\tau} \quad (43)$$

для любого тензора $T \in T^p(V)$.

Тензор $T \in T^p(V)$ называется *симметрическим*, если $T^\sigma = T$ для любой подстановки $\sigma \in S_p$. Симметрические тензоры образуют подпространство в $T^p(V)$; обозначим его $ST^p(V)$.

Предположим, что $\text{char } K = 0$. Тогда можно определить оператор *симметрирования* Sym в пространстве $T^p(V)$ по формуле

$$\text{Sym } T = \frac{1}{p!} \sum_{\sigma \in S_p} T^\sigma. \quad (44)$$

Ясно, что $\text{Sym } T \in ST^p(V)$ при любом T и $\text{Sym } T = T$ при $T \in ST^p(V)$. Это означает, что Sym — проектор на $ST^p(V)$.

Предложение 2. При условии $\text{char } K = 0$ имеется изоморфизм $\mu: S^p(V) \rightarrow ST^p(V)$, для которого

$$\mu(x_1 \vee \dots \vee x_p) = \text{Sym}(x_1 \otimes \dots \otimes x_p). \quad (45)$$

Доказательство. Так как правая часть (45) полилинейна и симметрична относительно x_1, \dots, x_p , то имеется линейное отображение $\mu: S^p(V) \rightarrow ST^p(V)$, удовлетворяющее условию (45). Оно переводит базисные векторы $e_{i_1} \vee \dots \vee e_{i_p}$ ($i_1 \leq \dots \leq i_p$) пространства $S^p(V)$ в тензоры $\text{Sym}(e_{i_1} \otimes \dots \otimes e_{i_p})$ ($i_1 \leq \dots \leq i_p$).

Рассматривая разложение симметрических тензоров по базисным векторам $e_{i_1} \otimes \dots \otimes e_{i_p}$ пространства $T^p(V)$, мы видим, что коэффициенты при базисных векторах, отличающихся только порядком индексов, одинаковы. Поэтому тензоры $\text{Sym}(e_{i_1} \otimes \dots \otimes e_{i_p})$ с $i_1 \leq \dots \leq i_p$ составляют базис пространства $ST^p(V)$. Следовательно, μ — изоморфизм. \square

Пользуясь этим изоморфизмом, пространство $S^p(V)$ часто отождествляют с $ST^p(V)$.

Заметим, что подпространство

$$ST(V) = \bigoplus_{p=0}^{\infty} ST^p(V) \subset T(V)$$

относительно не является подалгеброй в $T(V)$, но его отождествление с $S(V)$ позволяет ввести в нем структуру алгебры. Умножение в этой алгебре выглядит следующим образом:

$$T \vee U = \text{Sym}(T \otimes U). \quad (46)$$

Применим вышеизложенное к сопряженному пространству V^* . Положим

$$S_p(V) = S^p(V^*), \quad ST_p(V) = ST^p(V^*).$$

Пространство $ST_p(V)$ есть не что иное, как пространство симметрических p -линейных функций на V . Операция симметрирования выглядит следующим образом:

$$(\text{Sym } \alpha)(x_1, \dots, x_p) = \frac{1}{p!} \sum_{\sigma \in S_p} \alpha(x_{\sigma(1)}, \dots, x_{\sigma(p)}). \quad (47)$$

Каждой симметрической p -линейной функции $\alpha \in ST_p(V)$ поставим в соответствие многочлен $f_\alpha \in S_p(V)$ по формуле

$$f_\alpha(x) = \alpha(x, \dots, x) \quad (48)$$

подобно тому, как § 5.4 каждой симметрической билинейной функции была поставлена в соответствие квадратичная функция.

Предложение 3. Если $\text{char } K = 0$, то отображение

$$ST_p(V) \rightarrow S_p(V), \quad \alpha \mapsto f_\alpha, \quad (49)$$

есть изоморфизм векторных пространств, обратный изоморфизму $\mu: S_p(V) \rightarrow ST_p(V)$, определяемому как в предложении 2.

Доказательство. Достаточно рассмотреть симметрические p -линейные функции вида

$$\alpha = \text{Sym}(\alpha_1 \otimes \dots \otimes \alpha_p) = \mu(\alpha_1 \vee \dots \vee \alpha_p),$$

где $\alpha_1, \dots, \alpha_p \in V^*$. Для такой функции

$$f_\alpha(x) = \alpha_1(x) \dots \alpha_p(x) = (\alpha_1 \vee \dots \vee \alpha_p)(x).$$

Это означает, что

$$f_\alpha = \alpha_1 \vee \dots \vee \alpha_p = \mu^{-1}(\alpha),$$

что и требовалось доказать. \square

Симметрическая полилинейная функция α называется *поляризацией* однородного многочлена f_α .

Пример 1. Поляризацией многочлена

$$f(x) = x_1^3 + x_2^2 x_3$$

является симметрическая трилинейная функция

$$\alpha(x, y, z) = x_1 y_1 z_1 + \frac{1}{3}(x_3 y_2 z_2 + x_2 y_3 z_2 + x_2 y_2 z_3).$$

(Здесь x, y, z — векторы трехмерного пространства, x_i, y_i, z_i ($i = 1, 2, 3$) — их координаты.)

Замечание 1. В случае поля положительной характеристики отображение (49), вообще говоря, не является изоморфизмом. Так, над полем характеристики 2 симметрической билинейной функции $\alpha(x, y) = x_1 y_2 + x_2 y_1$ соответствует нулевая квадратичная функция, а квадратичная функция $f(x) = x_1 x_2$ не соответствует никакой симметрической билинейной функции.

Замечание 2. Формула (48) позволяет каждой (а не только симметрической) p -линейной функции поставить в соответствие однородный многочлен степени p . Однако определенное таким образом линейное отображение $T_p(V) \rightarrow S_p(V)$ при $p > 1$ не будет изоморфизмом.

Умножение в алгебре симметрических полилинейных функций

$$ST_*(V) = \bigoplus_{p=0}^{\infty} ST_p(V),$$

соответствующее умножению в алгебре

$$S_*(V) = \bigoplus_{p=0}^{\infty} S_p(V),$$

выглядит следующим образом:

$$\begin{aligned} (\alpha \vee \beta)(x_1, \dots, x_{p+q}) &= \\ &= \frac{p! q!}{(p+q)!} \sum_{(i_1, \dots, i_p | i_{p+1}, \dots, i_{p+q})} \alpha(x_{i_1}, \dots, x_{i_p}) \beta(x_{i_{p+1}}, \dots, x_{i_{p+q}}), \end{aligned} \quad (50)$$

где суммирование происходит по всем разбиениям $(i_1, \dots, i_p | i_{p+1}, \dots, i_{p+q})$ множества $\{1, \dots, p+q\}$ на две группы из p и q элементов соответственно (порядок чисел в каждой группе безразличен). Это следует из формул (46), (34) и (47), если учесть симметричность функций α и β . Произведение $\alpha \vee \beta$ называется *симметрическим произведением* функций α и β .

Симметрическое произведение p линейных функций $\alpha_1, \dots, \alpha_p \in V^*$ задается формулой

$$(\alpha_1 \vee \dots \vee \alpha_p)(x_1, \dots, x_p) = \frac{1}{p!} \text{per}(\alpha_i(x_j)), \quad (51)$$

где $\text{per} A$ — *перманент* квадратной матрицы A , определяемый аналогично определителю с той разницей, что все члены, независимо от четности перестановки, берутся со знаком плюс.

Замечание 3. В случае поля положительной характеристики формула (50) для симметрического произведения не имеет смысла. Ситуацию можно исправить, убрав коэффициент перед суммой. Определенная таким образом операция в $ST_*(V)$ будет по-прежнему ассоциативной и коммутативной, но полученная алгебра не будет изоморфна алгебре $S_*(V)$.

Аналогично тензорному произведению линейных операторов можно определить *симметрическую степень* $S^p \mathcal{A}$ линейного оператора как линейный оператор в пространстве $S^p(V)$, действующий на разложимые элементы по формуле

$$(S^p \mathcal{A})(x_1 \vee \dots \vee x_p) = \mathcal{A} x_1 \vee \dots \vee \mathcal{A} x_p. \quad (52)$$

Если отождествить пространство $S^p(V)$ с пространством $ST^p(V)$ симметрических тензоров (в случае $\text{char } K = 0$), то оператор $S^p \mathcal{A}$ будет не чем иным, как ограничением p -й тензорной степени оператора \mathcal{A} на инвариантное подпространство $ST^p(V) \subset T^p(V)$.

Пример 2. Имея в виду приложение к теории линейных представлений групп в § 11.4, вычислим след симметрического квадрата $S^2 \mathcal{A}$ линейного оператора \mathcal{A} . Пусть $\{e_1, \dots, e_n\}$ — базис пространства V . Тогда векторы $e_i \vee e_j$ с $i \leq j$ составляют базис пространства $S^2(V)$. Имеем (в символике Эйнштейна)

$$\begin{aligned} (S^2 \mathcal{A})(e_i \vee e_j) &= \mathcal{A} e_i \vee \mathcal{A} e_j = \mathcal{A}_i^k e_k \vee \mathcal{A}_j^l e_l = \\ &= \mathcal{A}_i^k \mathcal{A}_j^l e_k \vee e_l = \frac{1}{2} (\mathcal{A}_i^k \mathcal{A}_j^l + \mathcal{A}_i^l \mathcal{A}_j^k) e_k \vee e_l. \end{aligned}$$

Следовательно,

$$\text{tr } S^2 \mathcal{A} = \frac{1}{2} (\mathcal{A}_i^i \mathcal{A}_j^j + \mathcal{A}_i^j \mathcal{A}_j^i) = \frac{1}{2} ((\text{tr } \mathcal{A})^2 + \text{tr } \mathcal{A}^2). \quad (53)$$

Задача 1. Предположим, что характеристический многочлен оператора \mathcal{A} имеет (с учетом кратностей) n корней $\lambda_1, \dots, \lambda_n$. Доказать, что тогда характеристический многочлен оператора $S^2 \mathcal{A}$ имеет $n(n+1)/2$ корней $\lambda_i \lambda_j$ ($1 \leq i \leq j \leq n$). Вывести отсюда формулу (53).

§ 4. Алгебра Грассмана

Алгебра Грассмана, или внешняя алгебра, строится аналогично симметрической с той разницей, что условие симметричности заменяется условием кососимметричности. При этом следует предполагать, что $\text{char } K \neq 2$. (Случай $\text{char } K = 2$ может быть включен в общую схему, но он требует особой заботы.)

Определение 1. Полилинейное отображение (35) называется *кососимметрическим*, если

$$\varphi(x_{i_1}, \dots, x_{i_p}) = \text{sgn}(i_1, \dots, i_p) \varphi(x_1, \dots, x_p)$$

для любой перестановки (i_1, \dots, i_p) чисел $1, \dots, p$.

В этом определении можно ограничиться перестановками двух аргументов (потребовав, чтобы при этом образ умножался на -1). Ясно также, что если φ — кососимметрическое p -линейное отображение, то $\varphi(x_1, \dots, x_p) = 0$ всякий раз, когда среди векторов x_1, \dots, x_p есть одинаковые.

При $U = K$ данное определение превращается в определение *кососимметрической полилинейной функции*.

Пусть $\{e_1, \dots, e_n\}$ — базис пространства V .

Определение 2. Векторное пространство Λ вместе с кососимметрическим p -линейным отображением

$$\underbrace{V \times \dots \times V}_p \rightarrow \Lambda, \quad (x_1, \dots, x_p) \mapsto x_1 \wedge \dots \wedge x_p, \quad (54)$$

называется p -й *внешней степенью* пространства V , если векторы $e_{i_1} \wedge \dots \wedge e_{i_p}$ с $i_1 < \dots < i_p$ составляют базис пространства Λ .

По тем же причинам, что и определение симметрической степени, это определение не зависит от выбора базиса пространства V .

Так же как и симметрическая степень, внешняя степень существует и единственна. Она обозначается через $\Lambda^p(V)$.

Из определения следует, что

$$\dim \Lambda^p(V) = C_n^p = \frac{n(n-1)\dots(n-p+1)}{p!}.$$

В частности, $\Lambda^p(V) = 0$ при $p > n$.

Элементы пространства $\Lambda^p(V)$ называются *поливекторами* или, точнее, *p -векторами*. В частности, 1-векторы — это просто векторы пространства V ; 2-векторы называют *бивекторами*, 3-векторы — *тривекторами*.

Свойство универсальности внешней степени выражается следующим предложением, доказательство которого аналогично доказательству предложения 3.1.

Предложение 1. Для любого кососимметрического p -линейного отображения (35) существует единственное линейное отображение $\psi: \Lambda^p(V) \rightarrow U$, такое, что

$$\varphi(x_1, \dots, x_p) = \psi(x_1 \wedge \dots \wedge x_p) \quad (55)$$

для любых $x_1, \dots, x_p \in V$.

Поливекторы вида $x_1 \wedge \dots \wedge x_p$ ($x_1, \dots, x_p \in V$) называются *разложимыми*.

Имеется билинейное отображение

$$\wedge: \Lambda^p(V) \times \Lambda^q(V) \rightarrow \Lambda^{p+q}(V).$$

задаваемое на разложимых поливекторах по формуле

$$(x_1 \wedge \dots \wedge x_p) \wedge (x_{p+1} \wedge \dots \wedge x_{p+q}) = x_1 \wedge \dots \wedge x_{p+q}. \quad (56)$$

Рассмотрим прямую сумму

$$\Lambda(V) = \bigoplus_{p=0}^{\infty} \Lambda^p(V).$$

Операция \wedge превращает $\Lambda(V)$ в градуированную алгебру, которая называется *внешней алгеброй*, или *алгеброй Грассмана*, пространства V . Она ассоциативна и обладает единицей, но не коммутативна. Однако в ней выполняется следующее свойство, заменяющее коммутативность:

$$u \wedge v = (-1)^{pq} v \wedge u \quad \text{при } u \in \Lambda^p(V), v \in \Lambda^q(V).$$

Градуированные алгебры, обладающие этим свойством, называют *суперкоммутативными*. (Суперкоммутативность лежит в основе так называемой суперматематики.)

Всякий разложимый поливектор $x_1 \wedge \dots \wedge x_p \in \Lambda^p(V)$ совпадает с произведением векторов x_1, \dots, x_p в алгебре $\Lambda(V)$.

В отличие от симметрической алгебры, внешняя алгебра конечномерна. Более точно, поскольку ее базисные векторы $e_{i_1} \wedge \dots \wedge e_{i_p}$ ($i_1 < \dots < i_p$) находятся во взаимно однозначном соответствии с подмножествами множества $\{1, \dots, n\}$, то

$$\dim \Lambda(V) = 2^n.$$

Пространство $\Lambda^p(V)$ можно отождествить с подпространством кососимметрических тензоров в $T^p(V)$.

А именно, тензор $T \in T^p(V)$ называется *кососимметрическим*, если $T^\sigma = (\text{sgn } \sigma)T$ для любой подстановки $\sigma \in S_p$. Кососимметрические тензоры образуют подпространство в $T^p(V)$; обозначим его $\Lambda T^p(V)$.

Предположим, что $\text{char } K = 0$. Тогда можно определить оператор *альтернирования* Alt в пространстве $T^p(V)$ по формуле

$$\text{Alt } T = \frac{1}{p!} \sum_{\sigma \in S_p} (\text{sgn } \sigma) T^\sigma. \quad (57)$$

Это проектор на $\Lambda T^p(V)$.

Предложение 2. При условии $\text{char } K = 0$ имеется изоморфизм $\mu: \Lambda^p(V) \rightarrow \Lambda T^p(V)$, при котором

$$\mu(x_1 \wedge \dots \wedge x_p) = \text{Alt}(x_1 \otimes \dots \otimes x_p). \quad (58)$$

Доказательство аналогично доказательству предложения 3.2. Единственное дополнительное соображение состоит в том, что в

разложении кососимметрического тензора по базисным векторам $e_{i_1} \otimes \dots \otimes e_{i_p}$ пространства $T^p(V)$ коэффициенты при базисных векторах, среди индексов которых есть одинаковые, равны нулю. \square

Пользуясь этим изоморфизмом, пространство $\Lambda^p(V)$ часто отождествляют с $\Lambda T^p(V)$.

Задача 1. Доказать, что

$$T^2(V) = ST^2(V) \oplus \Lambda T^2(V),$$

но, если только $\dim V > 1$, $T^p(V) \neq ST^p(V) + \Lambda T^p(V)$ при $p > 2$.

Подпространство

$$\Lambda T(V) = \bigoplus_{p=0}^n \Lambda T^p(V) \subset T(V)$$

не является подалгеброй в $T(V)$, но его отождествление с $\Lambda(V)$ позволяет ввести в нем структуру алгебры, умножение в которой выглядит следующим образом:

$$T \wedge U = \text{Alt}(T \otimes U).$$

Применим вышеизложенное к сопряженному пространству. Положим

$$\Lambda_p(V) = \Lambda^p(V^*), \quad \Lambda T_p(V) = \Lambda T^p(V^*).$$

Пространство $\Lambda T_p(V)$ есть не что иное, как пространство кососимметрических p -линейных функций на V . Операция альтернирования выглядит следующим образом:

$$(\text{Alt } \alpha)(x_1, \dots, x_p) = \frac{1}{p!} \sum_{\sigma \in S_p} (\text{sgn } \sigma) \alpha(x_{\sigma(1)}, \dots, x_{\sigma(p)}). \quad (59)$$

Умножение в алгебре кососимметрических полилинейных функций

$$\Lambda T_*(V) = \bigoplus_{p=0}^{\infty} \Lambda T_p(V),$$

соответствующее умножению в алгебре

$$\Lambda_*(V) = \bigoplus_{p=0}^{\infty} \Lambda_p(V),$$

выглядит следующим образом:

$$\begin{aligned}
 (\alpha \wedge \beta)(x_1, \dots, x_{p+q}) &= \\
 &= \frac{p! q!}{(p+q)!} \sum_{(i_1, \dots, i_p | i_{p+1}, \dots, i_{p+q})} \operatorname{sgn}(i_1, \dots, i_{p+q}) \alpha(x_{i_1}, \dots, x_{i_p})(x_{i_{p+1}}, \dots, x_{i_{p+q}}),
 \end{aligned} \tag{60}$$

где суммирование, как и в формуле (50), происходит по всем разбиениям $(i_1, \dots, i_p | i_{p+1}, \dots, i_{p+q})$ множества $\{1, \dots, p+q\}$ на две группы из p и q элементов соответственно. Произведение $\alpha \wedge \beta$ называется *внешним произведением* функций α и β .

Внешнее произведение p линейных функций $\alpha_1, \dots, \alpha_p \in V^*$ задается формулой

$$(\alpha_1 \wedge \dots \wedge \alpha_p)(x_1, \dots, x_p) = \frac{1}{p!} \det(\alpha_i(x_j)). \tag{61}$$

Замечание 1. В случае поля положительной характеристики формула (60) не имеет смысла. Однако, если убрать коэффициент перед суммой, полученная алгебра будет по-прежнему изоморфна алгебре $\Lambda_*(V)$. Иногда такое определение внешнего умножения принимают и в случае поля нулевой характеристики.

Аналогично симметрической степени линейного оператора определяется *внешняя степень* $\Lambda^p \mathcal{A}$ линейного оператора \mathcal{A} .

Задача 2. Доказать, что

$$\operatorname{tr} \Lambda^2 \mathcal{A} = \frac{1}{2} ((\operatorname{tr} \mathcal{A})^2 - \operatorname{tr} \mathcal{A}^2). \tag{62}$$

В то время как понятие симметрической алгебры есть лишь новый взгляд на алгебру многочленов, понятие алгебры Грассмана является действительно новым для нашего курса, хотя неявно мы соприкоснулись с ним в теории определителей. Приложения алгебры Грассмана, о которых будет рассказано ниже, можно рассматривать как развитие теории определителей.

Пусть V есть n -мерное векторное пространство над полем K характеристики $\neq 2$.

Теорема 1. 1) Система векторов $\{a_1, \dots, a_p\}$ пространства V линейно зависима тогда и только тогда, когда $a_1 \wedge \dots \wedge a_p = 0$.

2) Если системы векторов $\{a_1, \dots, a_p\}$ и $\{b_1, \dots, b_p\}$ линейно независимы, то $\langle a_1, \dots, a_p \rangle = \langle b_1, \dots, b_p \rangle$ тогда и только тогда, когда p -векторы $a_1 \wedge \dots \wedge a_p$ и $b_1 \wedge \dots \wedge b_p$ пропорциональны.

Доказательство. 1) Если векторы a_1, \dots, a_p линейно зависимы, то один из них линейно выражается через остальные. Пусть, например,

$$a_p = \sum_{i=1}^{p-1} \lambda_i a_i.$$

Тогда

$$a_1 \wedge \dots \wedge a_{p-1} \wedge a_p = \sum_{i=1}^{p-1} \lambda_i a_1 \wedge \dots \wedge a_{p-1} \wedge a_i = 0.$$

Если векторы a_1, \dots, a_p линейно независимы, то их можно включить в базис пространства V . Тогда p -вектор $a_1 \wedge \dots \wedge a_p$ будет одним из векторов базиса пространства $\Lambda^p(V)$, построенного согласно определению внешней степени. Следовательно, он отличен от нуля.

2) Если $\langle a_1, \dots, a_p \rangle = \langle b_1, \dots, b_p \rangle$, то векторы b_1, \dots, b_p линейно выражаются через векторы a_1, \dots, a_p и, значит, p -вектор $b_1 \wedge \dots \wedge b_p$ линейно выражается через p -векторы вида $a_{i_1} \wedge \dots \wedge a_{i_p}$. Однако

$$a_{i_1} \wedge \dots \wedge a_{i_p} = \begin{cases} \pm a_1 \wedge \dots \wedge a_p, & \text{если } i_1, \dots, i_p \text{ различны,} \\ 0 & \text{в противном случае.} \end{cases}$$

Следовательно, $b_1 \wedge \dots \wedge b_p = \lambda a_1 \wedge \dots \wedge a_p$.

Если $\langle a_1, \dots, a_p \rangle \neq \langle b_1, \dots, b_p \rangle$, то существует такой базис $\{e_1, \dots, e_n\}$ пространства V , что

$$\langle a_1, \dots, a_p \rangle = \langle e_1, \dots, e_p \rangle, \quad \langle b_1, \dots, b_p \rangle = \langle e_{d+1}, \dots, e_{d+p} \rangle \quad (0 < d \leq p).$$

По уже доказанному p -вектор $a_1 \wedge \dots \wedge a_p$ пропорционален $e_1 \wedge \dots \wedge e_p$, а p -вектор $b_1 \wedge \dots \wedge b_p$ пропорционален $e_{d+1} \wedge \dots \wedge e_{d+p}$. Но p -векторы $e_1 \wedge \dots \wedge e_p$ и $e_{d+1} \wedge \dots \wedge e_{d+p}$ не пропорциональны, так как они суть различные векторы базиса пространства $\Lambda^p(V)$, построенного согласно определению внешней степени. Следовательно, и p -векторы $a_1 \wedge \dots \wedge a_p$ и $b_1 \wedge \dots \wedge b_p$ не пропорциональны. \square

Совокупность разложимых p -векторов называется *грассмановым конусом*. Проективизация этого конуса называется *грассмановым многообразием* и обозначается $\text{Gr}_p(V)$. Согласно доказанной теореме, точки многообразия $\text{Gr}_p(V)$ находятся во взаимно однозначном соответствии с p -мерными подпространствами пространства V .

Пусть $\{e_1, \dots, e_n\}$ — фиксированный базис пространства V и $\{a_1, \dots, a_p\}$ — базис подпространства U . Найдем координаты p -вектора $a_1 \wedge \dots \wedge a_p$ в базисе пространства $\Lambda^p(V)$, образованном произведениями $e_{i_1} \wedge \dots \wedge e_{i_p}$ с $i_1 < \dots < i_p$.

Пусть $A = (a_{ij})$ — матрица размера $p \times n$, образованная координатами векторов a_1, \dots, a_p в базисе $\{e_1, \dots, e_n\}$, т. е.

$$a_i = \sum_j a_{ij} e_j \quad (i = 1, \dots, p).$$

Имеем тогда

$$a_1 \wedge \dots \wedge a_p = \sum_{i_1, \dots, i_p} a_{1i_1} \dots a_{pi_p} e_{i_1} \wedge \dots \wedge e_{i_p}.$$

Если среди индексов i_1, \dots, i_p есть одинаковые, то $e_{i_1} \wedge \dots \wedge e_{i_p} = 0$. Если же они все различны, то мы можем переставить множители в $e_{i_1} \wedge \dots \wedge e_{i_p}$ так, чтобы их индексы шли в порядке возрастания; при этом все произведение умножится на $(-1)^s$, где s — число инверсий в последовательности (i_1, \dots, i_p) . Отсюда следует, что

$$a_1 \wedge \dots \wedge a_p = \sum_{i_1 < \dots < i_p} M_{i_1 \dots i_p} e_{i_1} \wedge \dots \wedge e_{i_p}, \quad (63)$$

где $M_{i_1 \dots i_p}$ — минор порядка p матрицы A , образованный столбцами с номерами i_1, \dots, i_p .

Согласно теореме 1, числа $M_{i_1 \dots i_k}$ однозначно определяют подпространство U . Они называются его *плюккеровыми координатами*. Это не что иное, как однородные координаты соответствующей точки проективного пространства $P\Lambda^p(V)$. Они определены с точностью до одновременного умножения на число $c \neq 0$. Кроме того, так как разложимые p -векторы составляют лишь часть пространства $\Lambda^p(V)$, плюккеровы координаты подпространства не могут быть произвольными. Они связаны соотношениями, описываемыми следующей ниже теоремой.

Для того чтобы было удобнее написать эти соотношения, примем следующее соглашение: если заданы какие-то числа $\mu_{i_1 \dots i_p}$ для $i_1 < \dots < i_p$, то будем считать, что числа $\mu_{i_1 \dots i_p}$ автоматически определены для любых i_1, \dots, i_p таким образом, что при перестановке любых двух индексов число $\mu_{i_1 \dots i_p}$ умножается на -1 (и, тем самым, оно равно нулю, если какие-либо два индекса совпадают). В частности, $M_{i_1 \dots i_p}$ для любых i_1, \dots, i_p будет тогда равно определителю матрицы порядка p , составленной из столбцов матрицы A с номерами i_1, \dots, i_p (в указанном порядке).

Теорема 2. Числа $\mu_{i_1 \dots i_p}$ являются плюккеровыми координатами некоторого p -мерного подпространства $U \subset V$ тогда и только тогда, когда они не равны одновременно нулю и для любых $i_1, \dots, i_{p+1}, j_1, \dots, j_{p-1}$ выполнено

соотношение

$$\sum_{k=1}^{p+1} (-1)^k \mu_{i_1 \dots \hat{i}_k \dots i_{p+1}} \mu_{i_k j_1 \dots j_{p-1}} = 0, \quad (64)$$

где крышка обозначает пропуск соответствующего индекса.

Соотношения (64) называются соотношениями Плюккера.

Замечание 2. Так как левая часть соотношения (64) кососимметрична по i_1, \dots, i_{p+1} и по j_1, \dots, j_{p-1} , то можно считать, что $i_1 < \dots < i_{p+1}$ и $j_1 < \dots < j_{p-1}$.

Доказательство. Докажем, что соотношения (64) выполняются для плюккеревых координат $M_{i_1 \dots i_p}$ p -мерного подпространства $U \subset V$. Разлагая определитель $M_{i_k j_1 \dots j_{p-1}}$ по первому столбцу, получаем

$$M_{i_k j_1 \dots j_{p-1}} = \sum_s a_{s i_k} N_s,$$

где N_s не зависит от k . Поэтому достаточно доказать, что

$$\sum_{k=1}^{p+1} (-1)^k M_{i_1 \dots \hat{i}_k \dots i_{p+1}} a_{s i_k} = 0 \quad (65)$$

для всех s . Припишем к матрице A ее s -ю строку. Полученную матрицу размера $(p+1) \times n$ обозначим через A_s . Тогда левая часть равенства (65) есть с точностью до знака результат разложения по последней строке определителя матрицы порядка $p+1$, составленной из столбцов матрицы A_s с номерами i_1, \dots, i_{p+1} . Так как матрица A_s имеет две одинаковые строки, то этот определитель равен нулю.

Обратно, пусть числа $\mu_{i_1 \dots i_p}$ не равны одновременно нулю и удовлетворяют соотношениям (64). Докажем, что существует такая матрица A размера $p \times n$, что

$$\mu_{i_1 \dots i_p} = M_{i_1 \dots i_p} \quad (66)$$

(где $M_{i_1 \dots i_p}$ имеет тот же смысл, что и выше) для всех i_1, \dots, i_p .

Считая для определенности, что $\mu_{1 \dots p} = 1$, будем искать матрицу A в виде

$$A = \begin{pmatrix} 1 & 0 & \dots & 0 & a_{1,p+1} & \dots & a_{1n} \\ 0 & 1 & \dots & 0 & a_{2,p+1} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 & a_{p,p+1} & \dots & a_{pn} \end{pmatrix}.$$

При $j > p$ имеем тогда

$$M_{1 \dots \hat{i} \dots p j} = (-1)^{p-i} a_{ij}.$$

Поэтому мы должны взять

$$a_{ij} = (-1)^{p-i} \mu_{1 \dots \hat{i} \dots p j};$$

тогда равенство (66) будет выполнено во всех случаях, когда множество $\{i_1, \dots, i_p\}$ не более чем одним элементом отличается от множества $\{1, \dots, p\}$.

Докажем теперь индукцией по m , что равенство (66) выполнено, если множество $\{i_1, \dots, i_p\}$ отличается от множества $\{1, \dots, p\}$ m элементами. Пусть, скажем, $i_1 \notin \{1, \dots, p\}$. Согласно условию, выполнено соотношение

$$\mu_{i_1 i_2 \dots i_p} = \mu_{1 \dots p} \mu_{i_1 i_2 \dots i_p} = \sum_{k=1}^p (-1)^{k+1} \mu_{i_1 1 \dots \hat{k} \dots p} \mu_{k i_2 \dots i_p}. \quad (67)$$

С другой стороны, по уже доказанной части теоремы аналогичное соотношение выполнено для миноров матрицы A :

$$M_{i_1 i_2 \dots i_p} = \sum_{k=1}^p (-1)^{k+1} M_{i_1 1 \dots \hat{k} \dots p} M_{k i_2 \dots i_p}. \quad (68)$$

По предположению индукции правые части равенств (67) и (68) совпадают. Следовательно, $\mu_{i_1 \dots i_p} = M_{i_1 \dots i_p}$. \square

Пример 1. При $n = 4$, $p = 2$ соотношения Плюккера сводятся к одному соотношению

$$\mu_{12} \mu_{34} + \mu_{23} \mu_{14} + \mu_{31} \mu_{24} = 0. \quad (69)$$

Пример 2. При $p = n - 1$ нетривиальных соотношений Плюккера нет. Следовательно, всякий $(n - 1)$ -вектор разложим.

Задача 3. Доказать, что при $p \leq q$ существует билинейное отображение

$$\varphi: \Lambda^p(V) \times \Lambda^q(V^*) \rightarrow \Lambda^{q-p}(V^*),$$

задаваемое на разложимых элементах формулой

$$\begin{aligned} \varphi(x_1 \wedge \dots \wedge x_p, \alpha_1 \wedge \dots \wedge \alpha_q) &= \\ &= \sum_{i_1, \dots, i_p} \operatorname{sgn}(i_1, \dots, i_p, j_1, \dots, j_{q-p}) \alpha_{i_1}(x_1) \dots \alpha_{i_p}(x_p) \alpha_{j_1} \wedge \dots \wedge \alpha_{j_{q-p}}, \end{aligned}$$

где суммирование происходит по всем различным i_1, \dots, i_p , а $\{j_1, \dots, j_{q-p}\}$ есть дополнение к $\{i_1, \dots, i_p\}$ в $\{1, \dots, q\}$, упорядоченное произвольным образом.

Задача 4. Доказать, что если δ — ненулевой элемент из $\Lambda^n(V^*)$, то отображение

$$\Lambda^p(V) \rightarrow \Lambda^{n-p}(V^*), \quad u \mapsto \varphi(u, \delta),$$

где φ — билинейное отображение из задачи 3, является изоморфизмом, переводящим разложимые элементы в разложимые. Вывести отсюда другим способом (по сравнению с примером 2), что всякий $(n - 1)$ -вектор разложим.

В качестве другого приложения алгебры Грассмана выведем так называемый пфаффиан кососимметричной матрицы четного порядка.

Пусть $n = 2m$ и $A = (a_{ij})$ — кососимметричная матрица порядка n . Рассмотрим бивектор

$$a = \sum_{i < j} a_{ij} (e_i \wedge e_j) = \frac{1}{2} \sum_{i, j} a_{ij} (e_i \wedge e_j),$$

где $\{e_1, \dots, e_n\}$ — фиксированный базис пространства V . Вычислим его m -ю степень в алгебре $\Lambda(V)$:

$$\begin{aligned} a^m &= \underbrace{a \wedge \dots \wedge a}_m = \frac{1}{2^m} \sum_{i_1, \dots, i_n} a_{i_1 i_2} \dots a_{i_{n-1} i_n} e_{i_1} \wedge \dots \wedge e_{i_n} = \\ &= \frac{1}{2^m} \left(\sum_{(i_1, \dots, i_n)} \operatorname{sgn}(i_1, \dots, i_n) a_{i_1 i_2} \dots a_{i_{n-1} i_n} \right) e_1 \wedge \dots \wedge e_n, \end{aligned}$$

где последняя сумма берется по всем перестановкам (i_1, \dots, i_n) чисел $1, \dots, n$. Слагаемые этой суммы, отличающиеся лишь порядком пар $(i_1, i_2), \dots, (i_{n-1}, i_n)$ и порядком элементов в каждой паре, равны между собой. Следовательно,

$$a^m = m! \left(\sum_{(i_1 i_2 | \dots | i_{n-1} i_n)} \operatorname{sgn}(i_1, \dots, i_n) a_{i_1 i_2} \dots a_{i_{n-1} i_n} \right) e_1 \wedge \dots \wedge e_n, \quad (70)$$

где суммирование происходит по всем разбиениям множества $\{1, \dots, n\}$ на пары $(i_1, i_2), \dots, (i_{n-1}, i_n)$ (порядок пар и порядок элементов в каждой паре выбираются произвольно).

Выражение

$$\operatorname{pf} A = \sum_{(i_1 i_2 | \dots | i_{n-1} i_n)} \operatorname{sgn}(i_1, \dots, i_n) a_{i_1 i_2} \dots a_{i_{n-1} i_n} \quad (71)$$

называется *пфафффианом* матрицы A . Формула (70) переписывается в виде

$$a^m = m! (\operatorname{pf} A) e_1 \wedge \dots \wedge e_n. \quad (72)$$

Она справедлива и в том случае, когда векторы e_1, \dots, e_n линейно зависимы; но тогда она означает просто, что $a^m = 0$.

Теорема 3. 1) $\operatorname{pf} C A C^T = \det C \cdot \operatorname{pf} A$ для любой матрицы C порядка n .

2) $(\operatorname{pf} A)^2 = \det A$.

Доказательство. 1) Вначале докажем эту формулу в предположении, что $\operatorname{char} K = 0$. Пусть $\{e'_1, \dots, e'_n\}$ — базис пространства V и

$$(e_1, \dots, e_n) = (e'_1, \dots, e'_n) C.$$

Выразим бивектор $a = \frac{1}{2} \sum_{i, j} a_{ij} e_i \wedge e_j$ через e'_1, \dots, e'_n . Пусть $C = (c_{ij})$; тогда

$$a = \frac{1}{2} \sum_{i, j, k, l} a_{ij} c_{ki} c_{lj} e'_k \wedge e'_l = \frac{1}{2} \sum_{k, l} a'_{kl} e'_k \wedge e'_l,$$

Пример 3. Для кососимметрической матрицы A порядка 4 имеем

$$\text{pf } A = a_{12}a_{34} + a_{23}a_{14} + a_{31}a_{24}.$$

Сравнивая эту формулу с формулой (69), мы видим, что условие разложимости бивектора $a = \sum_{i < j} a_{ij} e_i \wedge e_j$ может быть записано в виде $\text{pf } A = 0$, что ввиду теоремы 3 равносильно условию $\det A = 0$. Так как ранг кососимметрической матрицы всегда четен, то это условие, в свою очередь, равносильно условию $\text{rk } A \leq 2$. С другой стороны, легко доказать и непосредственно, что бивектор a разложим тогда и только тогда, когда $\text{rk } A \leq 2$.

Коммутативная алгебра

Наиболее важными типами алгебраических структур, для которых имеется содержательная теория, являются кольца (в частности, поля) и группы. В этой главе мы разовьем темы абелевых групп и коммутативных ассоциативных колец, начатые в гл. 1 и 3. Впрочем, некоторые общие определения и простейшие факты, приведенные в § 2, 3, относятся к более общим кольцам.

§ 1. Конечно порожденные абелевы группы

Абелевы группы до некоторой степени схожи с векторными пространствами, с которыми читатель уже хорошо знаком. Во всяком случае, понятие линейной зависимости в теории абелевых групп также играет важную роль.

Напомним, что элементы аддитивной абелевой группы можно умножать на целые числа (что соответствует возведению в степень в мультипликативной группе). Эта операция обладает такими же свойствами, как умножение векторов на элементы основного поля.

А именно, пусть A — аддитивная абелева группа. Тогда легко проверить, что

$$k(a + b) = ka + kb, \quad (1)$$

$$(k + l)a = ka + la, \quad (2)$$

$$(kl)a = k(la) \quad (3)$$

для любых $a, b \in A$, $k, l \in \mathbb{Z}$. (Свойство (2) в мультипликативном варианте было доказано в § 4.3.) Из (1) и (2) выводятся аналогичные свойства для вычитания:

$$k(a - b) = ka - kb, \quad (k - l)a = ka - la.$$

Для любого подмножества $S \subset A$ совокупность всех линейных комбинаций

$$k_1 a_1 + \dots + k_n a_n \quad (a_i \in S, k_i \in \mathbb{Z})$$

есть наименьшая подгруппа группы A , содержащая S . Она называется *подгруппой, порожденной подмножеством S* , и обозначается через $\langle S \rangle$. Если $\langle S \rangle = A$, то говорят, что группа A *порождается* подмножеством S или что S — *система порождающих* группы A . (Это согласуется с понятиями, введенными в § 4.4 для произвольных групп.) Абелева группа, допускающая конечную систему порождающих, называется *конечно порожденной*. Конечно порожденные абелевы группы аналогичны конечномерным векторным пространствам.

Система $\{a_1, \dots, a_n\}$ элементов группы A называется *линейно независимой*, если $k_1 a_1 + \dots + k_n a_n = 0$ только при $k_1 = \dots = k_n = 0$. Линейно независимая система порождающих называется *базисом*.

В отличие от векторных пространств, не всякая конечно порожденная абелева группа обладает базисом. Так, группа \mathbb{Z}_n порождается одним элементом, но она не обладает базисом, так как всякий ее элемент a удовлетворяет нетривиальному соотношению $na = 0$.

Определение 1. Конечно порожденная абелева группа, обладающая базисом, называется *свободной*.

Для свободных абелевых групп справедливы аналоги некоторых теорем о векторных пространствах, доказанных в § 2.2.

Теорема 1. *Все базисы свободной абелевой группы L содержат одно и то же число элементов.*

Доказательство. Пусть $\{e_1, \dots, e_n\}$ и $\{e'_1, \dots, e'_m\}$ — базисы группы L . Предположим, что $m > n$. Имеем:

$$(e'_1, \dots, e'_m) = (e_1, \dots, e_n)C,$$

где C — некоторая целочисленная матрица размера $n \times m$. По основной лемме о линейной зависимости столбцы матрицы C линейно зависимы как элементы пространства \mathbb{Q}^n . Отсюда следует, что между ними имеется нетривиальная линейная зависимость с целыми коэффициентами; но тогда такая же линейная зависимость имеется между элементами e'_1, \dots, e'_m группы L , что невозможно. \square

Число элементов базиса свободной абелевой группы L называется ее *рангом* и обозначается через $\text{rk } L$. Очевидно, что всякая свободная абелева группа ранга n изоморфна группе \mathbb{Z}^n строк длины n , составленных из целых чисел.

Замечание 1. Нулевая группа считается свободной абелевой группой ранга 0.

Опишем все базисы свободной абелевой группы L . Пусть $\{e_1, \dots, e_n\}$ — какой-либо один базис и e'_1, \dots, e'_n — какие-то элементы группы L . Имеем:

$$(e'_1, \dots, e'_n) = (e_1, \dots, e_n)C, \quad (4)$$

где C — целочисленная квадратная матрица порядка n .

Теорема 2. *Элементы e'_1, \dots, e'_n составляют базис группы L тогда и только тогда, когда $\det C = \pm 1$.*

Доказательство. Если $\det C = \pm 1$, то матрица C^{-1} является целочисленной и

$$(e_1, \dots, e_n) = (e'_1, \dots, e'_n)C^{-1}.$$

Отсюда следует, что элементы e'_1, \dots, e'_n порождают группу L , а из невырожденности матрицы C — что они линейно независимы.

Обратно, пусть $\{e'_1, \dots, e'_n\}$ — базис группы L . Тогда

$$(e_1, \dots, e_n) = (e'_1, \dots, e'_n)D \quad (5)$$

для некоторой целочисленной матрицы D . Из (4) и (5) следует, что $CD = E$ и, значит, $\det C \cdot \det D = 1$. Так как $\det C$ и $\det D$ — целые числа, то $\det C = \pm 1$. \square

Если уподоблять свободные абелевы группы векторным пространствам, то роль подпространств следует отвести подгруппам. Это частично оправдывается следующей теоремой.

Теорема 3. *Всякая подгруппа N свободной абелевой группы L ранга n является свободной абелевой группой ранга $\leq n$.*

Доказательство проведем индукцией по n . При $n = 0$ доказывать нечего.

При $n > 0$ пусть $\{e_1, \dots, e_n\}$ — базис группы L . Рассмотрим подгруппу $L_1 = \langle e_1, \dots, e_{n-1} \rangle \subset L$. Это свободная абелева группа ранга $n - 1$. По предположению индукции подгруппа $N_1 = N \cap L_1$ является свободной абелевой группой ранга $m \leq n - 1$. Пусть $\{f_1, \dots, f_m\}$ — ее базис.

Рассмотрим последние координаты всех элементов из N в базисе $\{e_1, \dots, e_n\}$ группы L . Они образуют подгруппу в группе \mathbb{Z} , которая по теореме 4.3.2 имеет вид $k\mathbb{Z}$ для некоторого $k \in \mathbb{Z}_+$. Если $k = 0$, то $N = N_1$ и все доказано. Если $k > 0$, то пусть f_{m+1} — какой-либо элемент из N , последняя координата которого равна k ; тогда $\{f_1, \dots, f_m, f_{m+1}\}$ — базис группы N и также все доказано. \square

Аналогия между подгруппами свободной абелевой группы и подпространствами векторного пространства все же неполная. В отли-

чие от векторных пространств, в свободной абелевой группе ранга $n > 0$ существуют подгруппы того же ранга, не совпадающие со всей группой. Так, подгруппа $m\mathbb{Z} \subset \mathbb{Z}$ при $m > 0$ имеет ранг 1, как и вся группа \mathbb{Z} .

Однако связь между свободными абелевыми группами и векторными пространствами не исчерпывается аналогией между ними. Свободная абелева группа ранга n может быть вложена в виде подгруппы в n -мерное евклидово векторное пространство E^n . А именно, пусть $\{e_1, \dots, e_n\}$ — какой-либо базис пространства E^n . Тогда подгруппа L , порожденная векторами e_1, \dots, e_n (т. е. совокупность векторов с целыми координатами в базисе $\{e_1, \dots, e_n\}$) является свободной абелевой группой ранга n . Этот геометрический образ (см. рис. 1) очень помогает восприятию свободных абелевых групп.

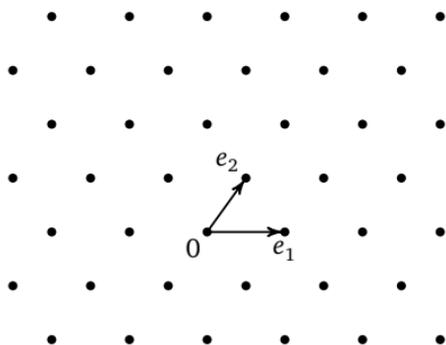


Рис. 1

Подгруппы $L \subset E^n$, получаемые указанным выше способом, называются *решетками* в E^n .

Задача 1. Параллелепипед $P(e_1, \dots, e_n)$, натянутый на какой-либо базис $\{e_1, \dots, e_n\}$ решетки $L \subset E^n$, называется *фундаментальным параллелепипедом* этой решетки. Доказать, что его объем не зависит от выбора базиса решетки L .

Для решеток в E^n имеется также аксиоматическое описание, использующее топологическое понятие дискретности.

Определение 2. Подмножество $M \subset E^n$ называется *дискретным*, если в каждом ограниченном подмножестве пространства E^n имеется лишь конечное число его элементов.

Очевидно, что всякая решетка дискретна. Более общо, подгруппа, порожденная любой линейно независимой системой векторов

(т. е. решетка в подпространстве пространства E^n), также дискретна.

Задача 2. Доказать, что подгруппа $L \subset E^n$ дискретна тогда и только тогда, когда ее пересечение с некоторой окрестностью нуля состоит только из нуля.

Теорема 4. *Всякая дискретная подгруппа $L \subset E^n$ порождена некоторой линейно независимой системой векторов пространства E^n .*

Доказательство. Пусть $U \subset E^n$ — линейная оболочка подгруппы L . Очевидно, что L — дискретная подгруппа в U . Поэтому, заменив пространство E^n на пространство U , мы можем свести доказательство к случаю, когда линейная оболочка подгруппы L совпадает со всем пространством.

В этом случае подгруппа L содержит некоторый базис $\{e_1, \dots, e_n\}$ пространства E^n . Рассмотрим решетку L_0 в E^n , порожденную этим базисом. В любом смежном классе группы L по L_0 имеется вектор, принадлежащий параллелепипеду $P(e_1, \dots, e_n)$. Так как пересечение $L \cap P(e_1, \dots, e_n)$ конечно, то индекс $|L : L_0|$ конечен. Если он равен d , то $dx \in L_0$ для любого $x \in L$ (см. следствие 4 теоремы 4.5.1). Таким образом,

$$L_0 \subset L \subset d^{-1}L_0. \quad (6)$$

Заметим, что $d^{-1}L_0$ — это решетка в E^n , порожденная базисом $\{d^{-1}e_1, \dots, d^{-1}e_n\}$. По теореме 3 из (6) следует, что L — свободная абелева группа, причем

$$n = \text{rk } L_0 \leq \text{rk } L \leq \text{rk } d^{-1}L_0 = n,$$

т. е. $\text{rk } L = n$. Базис группы L является в то же время базисом пространства E^n . Это означает, что L — решетка в E^n . \square

Следствие. *Всякая дискретная подгруппа $L \subset E^n$, линейная оболочка которой совпадает с E^n , является решеткой в E^n .*

Пример 1. Решетки в E^3 играют важную роль в кристаллографии. Кристаллические структуры характеризуются тем, что расположение атомов в них периодически повторяется во всех трех измерениях (см. рис. 2 § 4.2). Более точно, пусть Γ — группа симметрии некоторой кристаллической структуры (мысленно продолженной на все пространство). Обозначим через L группу всех таких векторов a , что параллельный перенос t_a принадлежит Γ . Сказанное выше означает, что L порождает пространство E^3 (как векторное пространство). С другой стороны, так как в любой ограниченной

части пространства имеется лишь конечное число атомов кристаллической структуры, группа L является дискретной подгруппой пространства E^3 . Следовательно, L — решетка в E^3 .

Как правило, группа Γ , кроме параллельных переносов, содержит и другие движения. Именно они определяют симметрию реальных кристаллов, которую мы можем наблюдать, точнее, группа G симметрии любого кристалла, симметрия структуры которого описывается группой Γ , совпадает с группой $d\Gamma$ линейных частей движений из Γ .

Пользуясь полученным выше описанием группы параллельных переносов, содержащихся в Γ , можно получить информацию о группе G . А именно, для любого $\gamma \in \Gamma$ и любого $a \in L$ имеем:

$$t_{d\gamma(a)} = \gamma t_a \gamma^{-1} \in \Gamma$$

(см. формулу (5) § 4.2). Следовательно, любое преобразование $g \in G$ сохраняет решетку L и, значит, в базисе этой решетки записывается целочисленной матрицей. Отсюда получаем, что $\text{tr } g \in \mathbb{Z}$. Но, с другой стороны, если g — поворот на угол α вокруг какой-либо оси, то $\text{tr } g = 2 \cos \alpha + 1$. Следовательно, $2 \cos \alpha \in \mathbb{Z}$. Это означает, что

$$\alpha \in \left\{ 0, \frac{\pi}{3}, \frac{\pi}{2}, \frac{2\pi}{3}, \pi \right\}.$$

В частности, кристаллы, в отличие от цветов и низших животных, не могут иметь поворотной симметрии 5-го порядка.

Дадим теперь более точное описание подгрупп свободных абелевых групп. Ключевую роль в этом описании будет играть одно вспомогательное утверждение о целочисленных матрицах.

Определение 3. Целочисленными элементарными преобразованиями строк матрицы называются преобразования следующих трех типов:

- 1) прибавление к одной строке другой, умноженной на целое число;
- 2) перестановка двух строк;
- 3) умножение одной строки на -1 .

Аналогично определяются целочисленные элементарные преобразования столбцов.

Прямоугольную матрицу $C = (c_{ij})$ размера $n \times m$ назовем диагональной и обозначим через $\text{diag}(u_1, \dots, u_p)$, если $c_{ij} = 0$ при $i \neq j$ и $c_{ii} = u_i$ при $i = 1, \dots, p$, где $p = \min\{n, m\}$.

Предложение 1. *Всякую целочисленную прямоугольную матрицу $C = (c_{ij})$ с помощью целочисленных элементарных преобразований строк и столбцов можно привести к виду $\text{diag}(u_1, \dots, u_p)$, где $u_1, \dots, u_p \geq 0$ и $u_i \mid u_{i+1}$ при $i = 1, \dots, p - 1$.*

Доказательство. Если $C = 0$, то доказывать нечего. Если $C \neq 0$, но $c_{11} = 0$, то путем перестановки строк и столбцов добьемся того, чтобы $c_{11} \neq 0$. Далее, умножив, если нужно, первую строку на -1 , добьемся того, чтобы $c_{11} > 0$. После этого с помощью целочисленных элементарных преобразований строк и столбцов будем стремиться уменьшить c_{11} .

Если какой-то элемент c_{i1} ($i \geq 2$) не делится на c_{11} , то разделим его на c_{11} с остатком:

$$c_{i1} = qc_{11} + r \quad (0 < r < c_{11}).$$

Вычтя из i -й строки 1-ю строку, умноженную на q и переставив 1-ю и i -ю строки полученной матрицы, мы уменьшим c_{11} . Аналогично, если какой-то элемент c_{1j} ($j \geq 2$) не делится на c_{11} , мы можем уменьшить c_{11} , работая со столбцами.

Если все элементы первого столбца и первой строки делятся на c_{11} , но какой-то элемент c_{ij} с $i, j \geq 2$ не делится на c_{11} , то поступим следующим образом. Вычтя из i -й строки 1-ю строку с подходящим коэффициентом, добьемся того, чтобы $c_{i1} = 0$ (при этом c_{ij} по-прежнему не будет делиться на c_{11}). После этого прибавим к 1-й строке i -ю строку. Элемент c_{11} при этом не изменится, но элемент c_{1j} перестанет делиться на c_{11} , и мы сможем применить описанную выше процедуру для уменьшения c_{11} .

Поступая таким образом, мы в конце концов придем к ситуации, когда все элементы матрицы делятся на c_{11} . Вычитая подходящие кратные 1-й строки из всех остальных строк и подходящие кратные 1-го столбца из всех остальных столбцов, мы получаем матрицу вида

$$\begin{pmatrix} u_1 & 0 & \dots & 0 \\ 0 & \boxed{C_1} & & \\ \vdots & & & \\ 0 & & & \end{pmatrix},$$

где все элементы матрицы C_1 делятся на u_1 . Последнее свойство будет сохраняться при любых целочисленных элементарных преобразованиях строк и столбцов матрицы C_1 .

Поступая таким же образом с матрицей C_1 и т. д., мы в конце концов приведем матрицу C к требуемому виду. \square

Для матриц размера 2×1 или 1×2 описанная в этом доказательстве процедура есть не что иное, как алгоритм Евклида для нахождения наибольшего общего делителя двух целых чисел.

Пример 2. Проиллюстрируем процедуру, описанную в этом доказательстве, на конкретном примере:

$$\begin{aligned} \begin{pmatrix} 2 & 6 & 2 \\ 2 & 3 & 4 \\ 4 & 2 & 4 \end{pmatrix} &\rightarrow \begin{pmatrix} 2 & 6 & 2 \\ 0 & -3 & 2 \\ 4 & 2 & 4 \end{pmatrix} \rightarrow \begin{pmatrix} 2 & 3 & 4 \\ 0 & -3 & 2 \\ 4 & 2 & 4 \end{pmatrix} \rightarrow \\ &\rightarrow \begin{pmatrix} 2 & 1 & 4 \\ 0 & -3 & 2 \\ 4 & -2 & 4 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 2 & 4 \\ -3 & 0 & 2 \\ -2 & 4 & 4 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 6 & 14 \\ 0 & 8 & 12 \end{pmatrix} \rightarrow \\ &\rightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 6 & 14 \\ 0 & 2 & -2 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & -2 \\ 0 & 6 & 14 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 20 \end{pmatrix}. \end{aligned}$$

Здесь с самого начала все элементы первого столбца и первой строки делились на $c_{11} = 2$, но элемент $c_{22} = 3$ не делился на c_{11} ; поэтому мы вычли из второй строки первую и прибавили к первой строке полученную вторую. Конечно, используя особенности конкретной матрицы, тот же результат можно было бы получить более коротким способом.

Задача 3. Доказать, что $u_i = d_i/d_{i-1}$, где d_i — наибольший общий делитель миноров i -го порядка исходной матрицы C (d_0 считается равным 1 и $0/0$ считается равным 0).

Замечание 2. Как следует из предыдущей задачи, числа u_1, \dots, u_p однозначно определяются матрицей C . Если отказаться от требования $u_i | u_{i+1}$, то процедура приведения целочисленной матрицы к диагональному виду несколько облегчится, но диагональный вид уже не будет, вообще говоря, определен однозначно.

Теорема 5. Для всякой подгруппы N свободной абелевой группы L ранга n существует такой базис $\{e_1, \dots, e_n\}$ группы L и такие натуральные числа u_1, \dots, u_m ($m \leq n$), что $\{u_1 e_1, \dots, u_m e_m\}$ — базис группы N и $u_i | u_{i+1}$ при $i = 1, \dots, m-1$.

Доказательство. Согласно теореме 3 группа N является свободной абелевой группой ранга $m \leq n$. Пусть $\{e_1, \dots, e_n\}$ — какой-нибудь базис группы L и $\{f_1, \dots, f_m\}$ — базис группы N . Тогда

$$(f_1, \dots, f_m) = (e_1, \dots, e_n)C,$$

где C — целочисленная матрица размера $n \times m$ и ранга m . У нас есть возможность делать следующие «элементарные» преобразования базисов групп L и N :

- 1) прибавление к одному базисному элементу другого, умноженного на целое число;
- 2) перестановка двух базисных элементов;
- 3) умножение одного базисного элемента на -1 .

Элементарные преобразования базиса группы L приводят к целочисленным элементарным преобразованиям строк матрицы C , а элементарные преобразования базиса группы N — к целочисленным элементарным преобразованиям столбцов этой матрицы. Согласно предложению 1, с помощью таких преобразований мы можем добиться того, чтобы

$$C = \text{diag}(u_1, \dots, u_m),$$

где $u_1, \dots, u_m > 0$ и $u_i \mid u_{i+1}$ при $i = 1, \dots, m - 1$. (Среди чисел u_1, \dots, u_m не может быть нулей, так как $\text{rk } C = m$.) Но это как раз и означает, что полученные в результате преобразований базисы $\{e_1, \dots, e_n\}$ и $\{f_1, \dots, f_m\}$ групп L и N связаны соотношениями

$$f_i = u_i e_i \quad (i = 1, \dots, m). \quad \square$$

Базис $\{e_1, \dots, e_n\}$ группы L , удовлетворяющий требованиям теоремы, не единствен. Однако числа u_1, \dots, u_m , как мы увидим ниже, определены однозначно. Они называются *инвариантными множителями* подгруппы $N \subset L$.

Задача 4. Доказать, что при $m = n$ индекс $|L : N|$ конечен и равен произведению инвариантных множителей.

Задача 5. Пусть L — решетка в E^n и N — ее подрешетка. Доказать, что индекс $|L : N|$ равен отношению объемов фундаментальных параллелепипедов решеток N и L .

На рисунке 2, иллюстрирующем теорему 5, точками изображены элементы решетки $L \subset E^2$, а кружками — элементы подрешетки N . Векторы e_1 и e_2 составляют базис решетки L , удовлетворяющий требованиям теоремы; при этом $u_1 = 1$, $u_2 = 4$.

Изучим теперь строение произвольных конечно порожденных абелевых групп. Для этого нам понадобится понятие прямой суммы абелевых групп.

Определение 4. Говорят, что (аддитивная) абелева группа A разлагается в *прямую сумму* подгрупп A_1, \dots, A_k , если каждый эле-

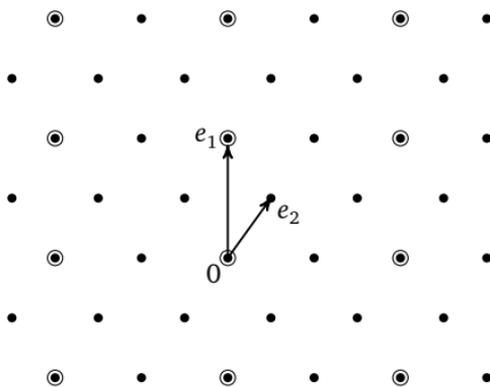


Рис. 2

мент $a \in A$ единственным образом представляется в виде $a = a_1 + \dots + a_k$, где $a_i \in A_i$. При этом пишут

$$A = A_1 \oplus \dots \oplus A_k.$$

В случае двух подгрупп A_1, A_2 единственность представления любого элемента $a \in A$ в виде $a = a_1 + a_2$ ($a_1 \in A_1, a_2 \in A_2$) равносильна тому, что $A_1 \cap A_2 = 0$.

Определение 5. *Прямой суммой* (аддитивных) абелевых групп A_1, \dots, A_k называется абелева группа $A_1 \oplus \dots \oplus A_k$, составленная из всех последовательностей (a_1, \dots, a_k) , где $a_i \in A_i$, с покомпонентной операцией сложения.

Так, например, $\underbrace{\mathbb{Z} \oplus \dots \oplus \mathbb{Z}}_n = \mathbb{Z}^n$. Отметим, что если группы A_1, \dots, A_k конечны, то

$$|A_1 \oplus \dots \oplus A_k| = |A_1| \dots |A_k|.$$

Прямую сумму в смысле определения 4 называют *внутренней*, а в смысле определения 5 — *внешней*. Эти два понятия связаны между собой так же, как в случае векторных пространств (см. § 8.2).

В случае мультипликативных абелевых групп G_1, \dots, G_k обычно говорят не о прямой сумме, а о *прямом произведении* и пишут $G_1 \times \dots \times G_k$. Это согласуется с общим определением прямого произведения групп, которое будет дано в § 10.1.

Рассмотрим вначале разложения циклических групп в прямую сумму (циклических) подгрупп.

Напомним, что всякая бесконечная циклическая группа изоморфна аддитивной группе \mathbb{Z} , а всякая конечная циклическая

группа порядка n изоморфна аддитивной группе \mathbb{Z}_n вычетов по модулю n .

Задача 6. Доказать, что группа \mathbb{Z} не может быть разложена в прямую сумму двух ненулевых подгрупп.

Предложение 2. Пусть $n = n_1 \dots n_s$, где числа n_1, \dots, n_s попарно взаимно просты. Тогда отображение

$$\varphi: \mathbb{Z}_n \rightarrow \mathbb{Z}_{n_1} \oplus \dots \oplus \mathbb{Z}_{n_s}, \quad [a]_n \mapsto ([a]_{n_1}, \dots, [a]_{n_s}),$$

является изоморфизмом групп.

Доказательство. Из определения сложения в группах вычетов и в прямой сумме следует, что φ — гомоморфизм. Найдем его ядро. Ясно, что $\varphi([a]_n) = 0$ тогда и только тогда, когда a делится на каждое из чисел n_1, \dots, n_s . Так как эти числа попарно взаимно просты, то это имеет место, только если a делится на $n_1 \dots n_s = n$, т. е. если $[a]_n = 0$. Таким образом, $\text{Ker } \varphi = 0$ и, значит, φ — инъективный гомоморфизм. Так как порядки групп \mathbb{Z}_n и $\mathbb{Z}_{n_1} \oplus \dots \oplus \mathbb{Z}_{n_s}$ равны, то отсюда следует, что φ — изоморфизм. \square

Замечание 3. Доказанное утверждение носит название *китайской теоремы об остатках* и обычно формулируется следующим образом: если числа n_1, \dots, n_s попарно взаимно просты, то система сравнений $x \equiv m_i \pmod{n_i}$ ($i = 1, \dots, s$) имеет решение при любых m_1, \dots, m_s .

Задача 7. Найти прообразы элементов $[1]_3 \in \mathbb{Z}_3$ и $[1]_5 \in \mathbb{Z}_5$ при изоморфизме $\varphi: \mathbb{Z}_{15} \rightarrow \mathbb{Z}_3 \oplus \mathbb{Z}_5$, построенном, как в предложении 2.

Следствие. Если $n = p_1^{k_1} \dots p_s^{k_s}$ — разложение числа n на простые множители, то

$$\mathbb{Z}_n \simeq \mathbb{Z}_{p_1^{k_1}} \oplus \dots \oplus \mathbb{Z}_{p_s^{k_s}}. \quad (7)$$

Задача 8. Доказать более сильную версию китайской теоремы об остатках, не предполагающую попарной взаимной простоты чисел n_1, \dots, n_s : система сравнений $x \equiv m_i \pmod{n_i}$ ($i = 1, \dots, s$) имеет решение тогда и только тогда, когда $m_i \equiv m_j \pmod{(n_i, n_j)}$ при любых i, j .

Определение 6. Конечная группа, порядок которой есть степень простого числа p , называется *примарной*, или *p -группой*.

Таким образом, всякая конечная циклическая группа разлагается в прямую сумму примарных циклических подгрупп.

Задача 9. Доказать, что примарная циклическая группа не может быть разложена в прямую сумму двух ненулевых подгрупп.

Теорема 6. *Всякая конечно порожденная абелева группа A разлагается в прямую сумму примарных и бесконечных циклических подгрупп, причем набор порядков этих подгрупп определен однозначно.*

Доказательство. Пусть $\{a_1, \dots, a_n\}$ — система порождающих группы A . Рассмотрим гомоморфизм

$$\varphi: \mathbb{Z}^n \rightarrow A, \quad (k_1, \dots, k_n) \mapsto k_1 a_1 + \dots + k_n a_n.$$

По теореме о гомоморфизме $A \simeq \mathbb{Z}^n / N$, где $N = \text{Ker } \varphi$. По теореме 5 существует такой базис $\{e_1, \dots, e_n\}$ группы \mathbb{Z}^n и такие натуральные числа u_1, \dots, u_m ($m \leq n$), что $\{u_1 e_1, \dots, u_m e_m\}$ — базис группы N и $u_i \mid u_{i+1}$ при $i = 1, \dots, m-1$. Рассмотрим гомоморфизм

$$\psi: \mathbb{Z}^n \rightarrow \mathbb{Z}_{u_1} \oplus \dots \oplus \mathbb{Z}_{u_m} \oplus \underbrace{\mathbb{Z} \oplus \dots \oplus \mathbb{Z}}_{n-m},$$

$$l_1 e_1 + \dots + l_n e_n \mapsto ([l_1]_{u_1}, \dots, [l_m]_{u_m}, l_{m+1}, \dots, l_n).$$

Очевидно, что $\text{Ker } \psi = N$. Отсюда следует, что

$$A \simeq \mathbb{Z}_{u_1} \oplus \dots \oplus \mathbb{Z}_{u_m} \oplus \underbrace{\mathbb{Z} \oplus \dots \oplus \mathbb{Z}}_{n-m}. \quad (8)$$

(Если $u_1 = \dots = u_q = 1$, то первые q слагаемых имеют вид $\mathbb{Z}_1 = \mathbb{Z}/\mathbb{Z} = 0$ и их следует отбросить.)

Таким образом, группа A разлагается в прямую сумму циклических подгрупп. Каждое конечное слагаемое этого разложения, в свою очередь, может быть разложено в прямую сумму примарных циклических подгрупп. Тем самым получится требуемое разложение группы A .

Перейдем к доказательству единственности. Пусть $\langle c \rangle_q$ обозначает циклическую группу порядка q с порождающим элементом c . Предположим, что группа A каким-то образом разложена в прямую сумму примарных и бесконечных циклических подгрупп:

$$A = \langle c_1 \rangle_{p_1^{k_1}} \oplus \dots \oplus \langle c_s \rangle_{p_s^{k_s}} \oplus \langle c_{s+1} \rangle_\infty \oplus \dots \oplus \langle c_{s+t} \rangle_\infty \quad (9)$$

(среди простых чисел p_1, \dots, p_s могут быть одинаковые). Рассмотрим так называемую подгруппу кручения

$$\text{Tor } A \doteq \{a \in A: ma = 0 \text{ для некоторого } m \in \mathbb{Z}, m \neq 0\}. \quad (10)$$

Очевидно, что $\text{Tor } A$ есть сумма первых s слагаемых разложения (9). Следовательно, $A/\text{Tor } A \simeq \mathbb{Z}^t$. Так как определение подгруппы $\text{Tor } A$

не зависит от разложения (9), то тем самым показано, что число t не зависит от этого разложения.

Далее, для каждого простого числа p можно рассмотреть *подгруппу p -кручения*

$$\text{Tor}_p A \doteq \{a \in A : p^k a = 0 \text{ для некоторого } k \in \mathbb{Z}_+\}. \quad (11)$$

Очевидно, что $\text{Tor}_p A$ есть сумма тех конечных слагаемых разложения (9), порядки которых суть степени p . Поэтому сумма этих слагаемых не зависит от разложения (9). Тем самым доказательство единственности сводится к случаю, когда A — примарная группа.

Пусть $|A| = p^k$ и группа A каким-то образом разложена в прямую сумму циклических подгрупп:

$$A = \langle c_1 \rangle_{p^{k_1}} \oplus \dots \oplus \langle c_r \rangle_{p^{k_r}} \quad (k_1 + \dots + k_r = k). \quad (12)$$

Докажем индукцией по k , что набор чисел $\{k_1, \dots, k_r\}$ не зависит от разложения (12).

При $k = 1$ утверждение очевидно. При $k > 1$ рассмотрим подгруппу

$$pA \doteq \{pa : a \in A\} \subset A.$$

Очевидно, что

$$pA = \langle pc_1 \rangle_{p^{k_1-1}} \oplus \dots \oplus \langle pc_r \rangle_{p^{k_r-1}};$$

в частности, при $k_i = 1$ соответствующее слагаемое просто исчезает. Так как определение подгруппы pA не зависит от разложения (12), то по предположению индукции набор отличных от единицы чисел k_1, \dots, k_r не зависит от этого разложения. Что касается единиц, то их количество может быть определено из соотношения $k_1 + \dots + k_r = k$ и потому также не зависит от разложения (12). \square

Замечание 4. Сами подгруппы разложения (12) при $r > 1$ определены не однозначно. Например, при $k_1 = \dots = k_r = 1$ группу A можно рассматривать как r -мерное векторное пространство над полем \mathbb{Z}_p , и ее разложение в прямую сумму циклических подгрупп — это то же самое, что разложение векторного пространства в прямую сумму одномерных подпространств, которое, очевидно, не единственно.

Замечание 5. Если группа A конечна, то в ее разложении не может быть бесконечных слагаемых и, значит, она разлагается в прямую сумму примарных циклических подгрупп.

Условие $u_i \mid u_{i+1}$ не было использовано в доказательстве теоремы. Однако оно позволяет восстановить числа u_1, \dots, u_m , т. е. инвариант-

ные множители подгруппы $N \subset \mathbb{Z}^n$, по числу n и порядкам слагаемых разложения (9) и тем самым доказать, что инвариантные множители подгруппы свободной абелевой группы L не зависят от выбора базиса группы L , удовлетворяющего требованиям теоремы 5. А именно, анализируя доказательство теоремы 6, мы видим, что для каждого простого числа p степень, в которой оно входит в разложение u_m , равна максимальной степени p среди чисел $p_1^{k_1}, \dots, p_s^{k_s}$; степень, в которой число p входит в разложение u_{m-1} , равна его максимальной степени среди оставшихся чисел $p_1^{k_1}, \dots, p_s^{k_s}$ и т. д. Тем самым числа u_1, \dots, u_m определяются однозначно.

В частности, всякая конечная абелева группа A допускает разложение

$$A = \langle a_1 \rangle_{u_1} \oplus \dots \oplus \langle a_r \rangle_{u_m}, \quad (13)$$

в котором $u_i \mid u_{i+1}$ при $i = 1, \dots, m - 1$. Можно считать, что $u_1 \neq 1$; иначе какое-то число первых слагаемых можно было бы отбросить. При этом условия числа u_1, \dots, u_m определены однозначно. Они называются *инвариантными множителями* группы A . Их произведение равно $|A|$.

Последний инвариантный множитель имеет простой смысл.

Определение 7. Наименьшее общее кратное порядков элементов конечной группы называется ее *экспонентой*.

Следствие 4 теоремы 4.5.1 показывает, что экспонента любой конечной группы делит ее порядок.

Предложение 3. *Экспонента конечной абелевой группы A равна ее последнему инвариантному множителю u_m .*

Доказательство. Ясно, что $u_m a = 0$ для любого $a \in A$. Это означает, что экспонента группы A делит u_m ; но так как в A имеется циклическая подгруппа порядка u_m , то экспонента равна u_m . \square

Следствие. *Конечная абелева группа A является циклической тогда и только тогда, когда ее экспонента равна ее порядку.*

Доказательство. Группа A является циклической тогда и только тогда, когда в разложении (13) имеется только одно слагаемое, но это как раз и означает, что $u_m = |A|$. \square

Этот критерий цикличности конечной абелевой группы имеет интересное приложение.

Теорема 7. *Любая конечная подгруппа мультипликативной группы поля (в частности, мультипликативная группа всякого конечно го поля) является циклической.*

Доказательство. Пусть G — конечная подгруппа мультипликативной группы поля K . Предположим, что ее экспонента равна m . Тогда $g^m = 1$ для всех $g \in G$. Так как уравнение $x^m = 1$ имеет в поле K не более m решений, то $|G| \leq m$ и, значит, $|G| = m$. \square

Задача 10. Найти какие-нибудь порождающие элементы групп \mathbb{Z}_7^* и \mathbb{Z}_{41}^* .

Задача 11. Доказать, что для любого простого $p \neq 2$ группа \mathbb{Z}_p^* обратимых элементов кольца \mathbb{Z}_{p^k} является циклической. (Указание: доказать, что элемент $[p+1]_{p^k}$ этой группы имеет порядок p^{k-1} .)

Задача 12. Доказать, что группа $\mathbb{Z}_{2^k}^*$ обратимых элементов кольца \mathbb{Z}_{2^k} не является циклической при $k > 2$; более точно, $\mathbb{Z}_{2^k}^* = \langle 3 \rangle \times \langle -1 \rangle \cong \mathbb{Z}_{2^{k-2}} \oplus \mathbb{Z}_2$.

Пример 3. При нечетном простом p группа \mathbb{Z}_p^* является циклической группой четного порядка и, следовательно, квадраты ее элементов образуют подгруппу индекса 2. Поэтому отображение, ставящее в соответствие каждому квадратичному вычету по модулю p число 1, а каждому квадратичному невычету — число -1 , является гомоморфизмом группы \mathbb{Z}_p^* в (мультипликативную) группу $\{\pm 1\}$. Образ вычета $[k]_p$ при этом отображении обозначается через $\left(\frac{k}{p}\right)$ и называется *символом Лежандра*.

Вычет $[-1]_p$ является единственным элементом порядка 2 в группе \mathbb{Z}_p^* . Он является квадратом тогда и только тогда, когда в этой группе имеется элемент порядка 4, т. е. когда $|\mathbb{Z}_p^*| = p - 1$ делится на 4. Таким образом,

$$\left(\frac{-1}{p}\right) = (-1)^{\frac{p-1}{2}}.$$

Задача 13. Доказать, что многочлен $x^4 + 1$ приводим над любым конечным полем. (Указание: доказать вначале, что хотя бы один из элементов $-1, 2, -2$ является квадратом в этом поле.)

§ 2. Идеалы и факторкольца

Обобщая конструкцию кольца вычетов \mathbb{Z}_n , изложенную в §1.5, можно рассматривать отношения эквивалентности, согласованные с операциями, в произвольных кольцах. Так как кольцо — это прежде всего абелева группа по сложению, то такое отношение должно быть отношением сравнимости по модулю некоторой аддитивной

подгруппы (см. § 4.5, в частности, задачу 4.5.3). Выясним, какой должна быть эта подгруппа для того, чтобы отношение эквивалентности было согласовано с умножением.

Пусть A — кольцо и $I \subset A$ — его подгруппа по сложению.

Предложение 1. *Отношение сравнимости по модулю I согласовано с умножением тогда и только тогда, когда подгруппа I инвариантна относительно умножений слева и справа на любые элементы из A .*

Последнее означает, что для любых $x \in I$ и $a \in A$ должны иметь место включения $ax \in I$ и $xa \in I$. Аддитивная подгруппа I , удовлетворяющая этим условиям, называется (двусторонним) идеалом кольца A . Подгруппа, удовлетворяющая первому (соответственно второму) из этих условий, называется левым (соответственно правым) идеалом. Понятно, что в коммутативном кольце нет разницы между левыми, правыми и двусторонними идеалами.

Доказательство. Пусть отношение сравнимости по модулю I согласовано с операцией умножения. Тогда для любого $a \in A$

$$x \equiv 0 \pmod{I} \Rightarrow ax \equiv a \cdot 0 = 0 \pmod{I}.$$

Это означает, что I — левый идеал. Аналогично доказывается, что I — правый идеал.

Обратно, пусть I — идеал, и пусть

$$a \equiv a' \pmod{I}, \quad b \equiv b' \pmod{I},$$

т. е.

$$a' = a + x, \quad b' = b + y \quad (x, y \in I).$$

Тогда

$$a'b' = ab + ay + xb + xy \equiv ab \pmod{I}. \quad \square$$

Итак, если I — идеал кольца A , то на факторгруппе A/I можно определить операцию умножения по правилу

$$(a + I)(b + I) = ab + I.$$

Легко видеть, что эта операция дистрибутивна относительно сложения. Построенное таким образом кольцо называется факторкольцом кольца A по идеалу I и обозначается A/I . Если кольцо A коммутативно, ассоциативно или обладает единицей, то и факторкольцо обладает соответствующим свойством.

Пример 1. В поле нет нетривиальных (т.е. отличных от нуля и всего поля) идеалов. В самом деле, если x — ненулевой элемент поля K , то всякий элемент поля K может быть представлен в виде ax , где $a \in K$, и поэтому всякий идеал, содержащий x , совпадает с K .

Пример 2. Всякая аддитивная подгруппа кольца \mathbb{Z} имеет вид $n\mathbb{Z}$, где $n \in \mathbb{Z}_+$ (см. пример 4.3.10), и является идеалом. Факторкольцо $\mathbb{Z}/n\mathbb{Z}$ при $n \neq 0$ — это не что иное, как кольцо вычетов \mathbb{Z}_n .

Примеры идеалов в других кольцах мы рассмотрим несколько позже, а сейчас покажем, каким образом идеалы и факторкольца возникают при рассмотрении гомоморфизмов колец.

Отображение f кольца A в кольцо B называется *гомоморфизмом*, если оно сохраняет операции, т.е. если

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

для любых $x, y \in A$. Образ $\text{Im } f$ гомоморфизма f является подкольцом кольца B , а его ядро

$$\text{Ker } f = \{x \in A : f(x) = 0\}$$

— идеалом кольца A .

Согласно определению факторкольца A/I , отображение

$$\pi : A \rightarrow A/I, \quad a \mapsto a + I,$$

является гомоморфизмом. Оно называется *каноническим гомоморфизмом* кольца A на факторкольцо A/I . Его ядром, очевидно, является идеал I .

Имеет место следующая теорема о гомоморфизме колец, аналогичная теореме о гомоморфизме групп (теореме 4.6.1).

Теорема 1. Пусть $f : A \rightarrow B$ — гомоморфизм колец. Тогда

$$\text{Im } f \simeq A/\text{Ker } f.$$

Более точно, имеется изоморфизм

$$\varphi : \text{Im } f \xrightarrow{\sim} A/\text{Ker } f,$$

ставящий в соответствие каждому элементу $b = f(a) \in \text{Im } f$ смежный класс $\pi(a) = a + \text{Ker } f$.

Доказательство. Благодаря теореме 4.6.1 мы уже знаем, что отображение φ является изоморфизмом аддитивных групп. Оста-

ется только проверить, что оно сохраняет операцию умножения. Пусть $f(x) = u$ и $f(y) = v$. Тогда $f(xy) = uv$ и

$$\varphi(uv) = \pi(xy) = \pi(x)\pi(y) = \varphi(u)\varphi(v). \quad \square$$

Пример 3. Рассмотренная нами в § 3.6 редукция по модулю p является гомоморфизмом кольца $\mathbb{Z}[t]$ на кольцо $\mathbb{Z}_p[t]$. Его ядро есть идеал $p\mathbb{Z}[t]$, образованный многочленами, все коэффициенты которых делятся на p . Следовательно,

$$\mathbb{Z}[t]/p\mathbb{Z}[t] \simeq \mathbb{Z}_p[t].$$

Пример 4. Пусть K — поле и $c \in K$ — его фиксированный элемент. Как мы фактически доказали в § 3.1, отображение

$$K[t] \rightarrow K, \quad f \mapsto f(c),$$

является гомоморфизмом колец. В силу теоремы Безу его ядро состоит из всех многочленов, делящихся на $t - c$. Следовательно,

$$K[t]/(t - c)K[t] \simeq K.$$

Пример 5. Пусть $t^2 + pt + q \in \mathbb{R}[t]$ — квадратный трехчлен с отрицательным дискриминантом и $c \in \mathbb{C}$ — один из его мнимых корней. Отображение

$$\mathbb{R}[t] \rightarrow \mathbb{C}, \quad f \mapsto f(c),$$

является гомоморфизмом колец. Его образ совпадает с \mathbb{C} , а ядро состоит из всех многочленов с вещественными коэффициентами, делящихся на $(t - c)(t - \bar{c}) = t^2 + pt + q$. Следовательно,

$$\mathbb{R}[t]/(t^2 + pt + q)\mathbb{R}[t] \simeq \mathbb{C}.$$

В случае, когда A — алгебра над полем K , в определении (левого, правого или двустороннего) идеала I требуется, чтобы он выдерживал также умножение на элементы поля K , т. е. был подпространством. Если I — (двусторонний) идеал алгебры A , то в факторкольце A/I определяется операция умножения на элементы поля K по правилу

$$\lambda(a + I) = \lambda a + I$$

и тем самым оно превращается в алгебру над K , называемую факторалгеброй алгебры A по идеалу I .

Замечание 1. Если A — алгебра с единицей 1 , то идеалы алгебры A — это то же, что идеалы кольца A . В самом деле, пусть I — левый идеал кольца A . Тогда для любых $x \in I$ и $\lambda \in K$ имеем

$$\lambda x = (\lambda 1)x \in I.$$

Это означает, что I — подпространство и, следовательно, — левый идеал алгебры A . Аналогично обстоит дело с правыми идеалами.

Пример 6. Непосредственно проверяется, что матрицы, у которых все столбцы, кроме первого, равны нулю, образуют левый идеал в алгебре $L_n(K)$ матриц порядка n . Аналогично, матрицы, у которых все строки, кроме первой, равны нулю, образуют правый идеал. Однако нетривиальных двусторонних идеалов в алгебре $L_n(K)$ нет. В самом деле, пусть $I \subset L_n(K)$ — ненулевой двусторонний идеал и $A = (a_{ij})$ — ненулевая матрица из этого идеала. Предположим, что $a_{kl} \neq 0$. Для любых i, j имеем

$$E_{ik} A E_{lj} = a_{kl} E_{ij} \in I,$$

и, значит, $E_{ij} \in I$. Следовательно, $I = L_n(K)$.

Пример 7. Нильтреугольные матрицы образуют идеал в алгебре всех треугольных матриц.

Пример 8. Функции, обращающиеся в нуль в заданной точке $x_0 \in X$, образуют идеал в алгебре $F(X, K)$ всех функций на множестве X со значениями в поле K .

Отображение f алгебры A в алгебру B называется *гомоморфизмом*, если оно линейно и сохраняет операцию умножения, т. е.

$$f(xy) = f(x)f(y)$$

для любых $x, y \in A$. Образ $\text{Im } f$ гомоморфизма f является подалгеброй алгебры B , а его ядро $\text{Ker } f$ — идеалом алгебры A .

Для любого идеала I алгебры A определяется *канонический гомоморфизм*

$$\pi: A \rightarrow A/I, \quad a \mapsto a + I,$$

ядром которого является I .

Имеет место *теорема о гомоморфизме алгебр*, формулируемая точно так же, как теорема о гомоморфизме колец.

Пример 9. Отображение, ставящее в соответствие каждой треугольной матрице ее диагональную часть, является гомоморфизмом алгебры треугольных матриц на алгебру диагональных матриц. Ядром этого гомоморфизма служит идеал нильтреугольных матриц.

Следовательно, факторалгебра алгебры треугольных матриц по идеалу нильтреугольных матриц изоморфна алгебре диагональных матриц.

Пример 10. Отображение, ставящее в соответствие каждой функции $f \in F(X, K)$ ее значение в точке $x_0 \in X$, является гомоморфизмом алгебры $F(X, K)$ на поле K , рассматриваемое как (одномерная) алгебра над самим собой. Его ядром служит идеал $I(x_0)$ функций, обращающихся в нуль в точке x_0 . Следовательно,

$$F(X, K)/I(x_0) \simeq K.$$

Используя понятие прямой суммы абелевых групп (см. § 1) и векторных пространств (см. § 8.2), определим прямые суммы колец и алгебр.

Определение 1. Говорят, что кольцо (соответственно алгебра) A разлагается в *прямую сумму* своих подколец (соответственно подалгебр) A_1, \dots, A_k , если

- 1) оно разлагается в прямую сумму A_1, \dots, A_k как аддитивная группа (соответственно как векторное пространство);
- 2) $A_i A_j = 0$ при $i \neq j$.

Последнее условие (при наличии условия 1) равносильно тому, что A_1, \dots, A_k — идеалы. Оно обеспечивает следующее «покомпонентное» правило умножения:

$$(x_1 + \dots + x_k)(y_1 + \dots + y_k) = x_1 y_1 + \dots + x_k y_k \quad (x_i, y_i \in A_i).$$

Пусть теперь A_1, \dots, A_k — какие-то кольца или алгебры.

Определение 2. *Прямой суммой* колец (соответственно алгебр) A_1, \dots, A_k называется их прямая сумма $A_1 \oplus \dots \oplus A_k$ как аддитивных групп (соответственно как векторных пространств) с покомпонентной операцией умножения:

$$(x_1, \dots, x_k)(y_1, \dots, y_k) = (x_1 y_1, \dots, x_k y_k) \quad (x_i, y_i \in A_i).$$

Очевидно, что определенная таким образом операция умножения в $A_1 \oplus \dots \oplus A_k$ дистрибутивна по отношению к сложению (соответственно билинейна), так что $A_1 \oplus \dots \oplus A_k$ действительно является кольцом (соответственно алгеброй). Если все кольца A_1, \dots, A_k коммутативны, ассоциативны или обладают единицей, то и их прямая сумма обладает соответствующим свойством.

Прямая сумма колец или алгебр в смысле определения 1 называется *внутренней*, а в смысле определения 2 — *внешней*. Между

этим двумя понятиями имеется такая же связь, как и в случае векторных пространств.

Пример 11. Пусть $n = n_1 \dots n_s$, где числа n_1, \dots, n_s попарно взаимно просты. Тогда отображение

$$\mathbb{Z}_n \xrightarrow{\sim} \mathbb{Z}_{n_1} \oplus \dots \oplus \mathbb{Z}_{n_s}, \quad (14)$$

рассмотренное в предложении 1.2, является не только изоморфизмом групп, но и изоморфизмом колец. Оно индуцирует изоморфизм мультипликативных групп обратимых элементов этих колец:

$$\mathbb{Z}_n^* \xrightarrow{\sim} \mathbb{Z}_{n_1}^* \times \dots \times \mathbb{Z}_{n_s}^*. \quad (15)$$

Задача 1. Используя результат примера 11, получить следующую формулу для функции Эйлера (см. пример 4.5.6):

$$\varphi(n) = n \left(1 - \frac{1}{p_1}\right) \dots \left(1 - \frac{1}{p_s}\right),$$

где p_1, \dots, p_s — все (различные) простые делители числа n .

Задача 2. Доказать, что группа \mathbb{Z}_n^* является циклической тогда и только тогда, когда $n = 2, 4, p^k$ или $2p^k$, где p — нечетное простое число. (Указание: использовать задачи 1.11 и 1.12.)

Задача 3. Доказать, что для составного n утверждение « $a^{n-1} \equiv 1 \pmod{n}$ для любого целого a , взаимно простого с n » (наивный аналог малой теоремы Ферма: см. пример 4.5.6) может быть верно, только если n является произведением не менее трех различных нечетных простых чисел, и что наименьшее составное n , для которого это утверждение верно — это 561.

Пример 12. Отображение, ставящее в соответствие каждой диагональной матрице последовательность ее диагональных элементов, является изоморфизмом алгебры диагональных матриц порядка n над полем K на прямую сумму n экземпляров поля K .

Начиная с этого момента, будем предполагать, что A — коммутативное ассоциативное кольцо с единицей.

Для любого подмножества $S \subset A$ совокупность всех «линейных комбинаций»

$$a_1 x_1 + \dots + a_m x_m \quad (x_1, \dots, x_m \in S; a_1, \dots, a_m \in A)$$

является наименьшим идеалом, содержащим S . Оно называется идеалом, порожденным подмножеством S , и обозначается через (S) . В частности, идеал (u) , порожденный одним элементом u , называется главным.

Определение 3. Целостное кольцо, в котором всякий идеал является главным, называется *кольцом главных идеалов*.

В частности, всякое поле по тривиальным причинам является кольцом главных идеалов.

Теорема 2. *Всякое евклидово кольцо является кольцом главных идеалов.*

Доказательство. Очевидно, что нулевой идеал является главным. Пусть I — ненулевой идеал кольца A , и пусть u — наименьший по норме ненулевой элемент идеала I . Остаток при делении на u любого элемента идеала I принадлежит I и, следовательно, может быть только нулем. Это означает, что $I = (u)$. \square

Таким образом, кольца \mathbb{Z} и $K[t]$ (где K — поле) являются кольцами главных идеалов.

Замечание 2. Легко видеть, что свойство «всякий идеал является главным» сохраняется при переходе к факторкольцу. Однако факторкольцо кольца главных идеалов не обязано быть кольцом главных идеалов в смысле нашего определения, так как в нем могут появиться делители нуля. Так, в кольце $\mathbb{Z}_n = \mathbb{Z}/n\mathbb{Z}$ все идеалы главные, но оно является кольцом главных идеалов только при простом n (когда оно является полем).

Замечание 3. Существуют кольца главных идеалов, не являющиеся ни полями, ни евклидовыми кольцами, например, кольцо чисел вида $a + b\sqrt{-19}$, где $a, b \in \mathbb{Z}$ или $a, b \in \mathbb{Z} + \frac{1}{2}$.

Свойства делимости, доказанные в § 3.5 для евклидовых колец, обобщаются на произвольные кольца главных идеалов.

Теорема 3. *В кольце главных идеалов A для любых элементов x, y существует наибольший общий делитель d , и он может быть представлен в виде $d = ax + by$, где $a, b \in A$.*

Доказательство. Рассмотрим идеал

$$(x, y) = \{ax + by : a, b \in A\},$$

порожденный элементами x и y . Существует такой элемент $d \in A$, что $(x, y) = (d)$. Это и будет наибольший общий делитель элементов x и y . По самому построению он представляется в виде $d = ax + by$. \square

Замечание 4. Обозначение (x, y) для идеала, порожденного элементами x и y , хорошо согласуется с обозначением (x, y) для их наибольшего общего делителя.

Теорема о существовании и единственности разложения на простые множители также справедлива в любых кольцах главных идеалов. В самом деле, доказательство единственности, данное в § 3.5 для евклидовых колец, с учетом теоремы 3 дословно переносится на кольца главных идеалов. Что касается существования, то оно будет позже доказано для гораздо более широкого класса колец (см. теорему 7.1).

Следующая теорема является обобщением теоремы 1.5.1.

Теорема 4. Пусть u — ненулевой необратимый элемент кольца главных идеалов A . Факторкольцо $A/(u)$ является полем тогда и только тогда, когда элемент u прост.

Доказательство. Для всякого $a \in A$ будем обозначать через $[a]$ смежный класс $a + (u) \in A/(u)$. Если $u = vw$, где v и w необратимы, то $[v][w] = 0$, но $[v], [w] \neq 0$, так что в кольце $A/(u)$ есть делители нуля и, стало быть, оно не является полем.

Обратно, если элемент u прост, то для всякого $x \notin (u)$ элементы x и u взаимно просты и, следовательно, существуют такие a и b , что $ax + bu = 1$. Переходя к смежным классам, получаем $[a][x] = 1$ в $A/(u)$. Таким образом, всякий ненулевой элемент кольца $A/(u)$ обратим и, значит, оно является полем. \square

Пример 13. Выясним, когда простое число p является простым элементом кольца $\mathbb{Z}[i]$ целых гауссовых чисел (см. пример 3.5.1). Так как $\mathbb{Z}[i] \simeq \mathbb{Z}[t]/(t^2 + 1)$ (ср. пример 5), то

$$\mathbb{Z}[i]/(p) \simeq \mathbb{Z}[t]/(t^2 + 1, p) \simeq \mathbb{Z}_p[t]/(t^2 + 1)$$

(см. пример 3). Согласно примеру 1.3, многочлен $t^2 + 1$ неприводим над \mathbb{Z}_p тогда и только тогда, когда $p \equiv -1 \pmod{4}$. Двукратное применение теоремы 4 показывает, что последнее условие и является необходимым и достаточным для того, чтобы элемент p был прост в $\mathbb{Z}[i]$.

Пусть $p \equiv 1 \pmod{4}$, и пусть $p = \pi_1 \dots \pi_s$ ($s \geq 2$) — разложение p на простые множители в кольце $\mathbb{Z}[i]$. Переходя к нормам, получаем

$$N(\pi_1) \dots N(\pi_s) = N(p) = p^2,$$

откуда $s = 2$ и $N(\pi_1) = N(\pi_2) = p$. Если $\pi_1 = a + bi$ ($a, b \in \mathbb{Z}$), то $a^2 + b^2 = p$ (и $\pi_2 = a - bi$). Таким образом, всякое простое число вида $4k + 1$ представимо в виде суммы квадратов двух целых чисел.

Задача 4. Доказать, что простые элементы кольца $\mathbb{Z}[i]$ — это, с точностью до ассоциированности, простые натуральные числа

вида $4k + 3$, числа вида $a + bi$ ($a, b \in \mathbb{N}$), где $a^2 + b^2$ есть простое натуральное число вида $4k + 1$ и число $1 + i$.

Задача 5. Пользуясь однозначностью разложения на простые множители в кольце $\mathbb{Z}[i]$, доказать, что натуральное число n представимо в виде суммы квадратов двух целых чисел тогда и только тогда, когда в его разложение на простые множители (в \mathbb{Z}) все множители вида $4k + 3$ входят в четной степени, и найти число таких представлений в этом случае.

Следующая теорема является обобщением примера 11.

Теорема 5. Пусть u, v — взаимно простые элементы кольца главных идеалов A . Тогда

$$A/(uv) \simeq A/(u) \oplus A/(v). \quad (16)$$

Доказательство. Отображение

$$f: A \rightarrow A/(u) \oplus A/(v), \quad a \mapsto (a + (u), a + (v)),$$

является гомоморфизмом колец. Пусть a и b — такие элементы кольца A , что $au + bv = 1$. Тогда

$$f(bv) = (1 + (u), 0 + (v)), \quad f(au) = (0 + (u), 1 + (v)),$$

откуда следует, что гомоморфизм f сюръективен. Очевидно, что $\text{Ker } f = (uv)$. Это и дает изоморфизм (16). \square

Пример 14. Если $f \in K[t]$ — неприводимый многочлен над полем K , то факторкольцо $K[t]/(f)$ является полем. Например, факторкольцо $\mathbb{R}[t]/(t^2 + 1)$ изоморфно \mathbb{C} (см. пример 5). Напротив, если $f = (t - c_1) \dots (t - c_n)$, где c_1, \dots, c_n различны, то из теоремы 5 следует, что

$$K[t]/(f) \simeq K[t]/(t - c_1) \oplus \dots \oplus K[t]/(t - c_n) \simeq \underbrace{K \oplus \dots \oplus K}_n$$

(см. пример 4).

§ 3. Модули над кольцами главных идеалов

Ввиду свойств (1)—(3) на абелевы группы можно смотреть как на «векторные пространства над \mathbb{Z} ». Аналогичным образом можно определить «векторные пространства» и над более общими кольцами. Они называются модулями.

Понятие модуля оказывается очень полезным. В частности, теория модулей над кольцами главных идеалов, которая будет изложена в настоящем параграфе, включает в себя теорию конечно порожденных абелевых групп, которой был посвящен §1, и теорему о приведении матрицы линейного оператора к жордановой форме.

Начнем, однако, с общих понятий.

Пусть A — ассоциативное кольцо с единицей.

Определение 1. (Левым) A -модулем (или модулем над A) называется аддитивная абелева группа M с операцией умножения (слева) на элементы кольца A , обладающей следующими свойствами:

- 1) $a(x + y) = ax + ay$ для любых $a \in A$, $x, y \in M$;
- 2) $(a + b)x = ax + bx$ для любых $a, b \in A$, $x \in M$;
- 3) $(ab)x = a(bx)$ для любых $a, b \in A$, $x \in M$;
- 4) $1x = x$ для любого $x \in M$.

В частности, модули над полем — это векторные пространства; модули над \mathbb{Z} — это просто аддитивные абелевы группы. Имеются, однако, и другие важные примеры модулей.

Пример 1. Модуль над кольцом многочленов $K[t]$ (K — поле) — это векторное пространство над K с линейным оператором, играющим роль умножения на t .

Пример 2. Кольцо A является модулем над самим собой (произведение элемента модуля на элемент кольца определяется как произведение этих элементов в кольце).

Пример 3. Всякое векторное пространство V тавтологическим образом является модулем над кольцом $L(V)$ всех линейных операторов в V .

Замечание 1. Аналогичным образом определяются *правые модули*. Разница состоит в том, что в этом случае элементы кольца A пишутся в произведении справа от элементов модуля и соответственно этому при умножении на произведение элементов кольца элемент модуля умножается сначала на первый множитель (а не на второй, как в случае левых модулей). Если кольцо A коммутативно, то разницы между левыми и правыми модулями нет (и элементы кольца могут писаться в произведении с любой стороны от элементов модуля).

Подмножество N модуля M называется *подмодулем*, если оно замкнуто относительно сложения и умножения на элементы кольца A . Всякий подмодуль является модулем относительно тех же операций.

Пример 4. Подмодуль абелевой группы, рассматриваемой как \mathbb{Z} -модуль — это просто подгруппа.

Пример 5. Подмодуль $K[t]$ -модуля (см. пример 1) — это подпространство, инвариантное относительно оператора умножения на t .

Пример 6. Подмодуль кольца A , рассматриваемого как (левый) модуль над самим собой — это левый идеал этого кольца.

Так же, как это было сделано для векторных пространств в § 8.2 и для абелевых групп в § 1, определяется (внутренняя и внешняя) *прямая сумма* модулей.

Определим теперь понятие фактормодуля.

В A -модуле M отношение эквивалентности R следует считать согласованным с операцией умножения на элементы кольца A , если

$$x \sim_R x' \Rightarrow ax \sim_R ax'.$$

Отношение сравнимости по модулю аддитивной подгруппы $N \subset M$ согласовано с операцией умножения на элементы кольца A тогда и только тогда, когда N — подмодуль. В этом случае на факторгруппе M/N можно определить операцию умножения на элементы кольца A по правилу

$$a(x + N) = ax + N,$$

превратив ее тем самым в A -модуль, называемый *фактормодулем* модуля M по подмодулю N и обозначаемый через M/N .

В частности, таким образом определяется *факторпространство* V/U векторного пространства V по подпространству U . Фактормодули \mathbb{Z} -модулей — это то же, что факторгруппы.

Отображение f модуля M в модуль N (над тем же кольцом) называется *гомоморфизмом*, если

$$\begin{aligned} f(x + y) &= f(x) + f(y), \\ f(ax) &= af(x). \end{aligned}$$

Обратимый гомоморфизм называется *изоморфизмом*.

Если $f: M \rightarrow N$ — какой-либо гомоморфизм модулей, то его образ

$$\text{Im } f = \{f(x) : x \in M\} \subset N$$

— подмодуль модуля N , а его ядро

$$\text{Ker } f = \{x \in M : f(x) = 0\} \subset M$$

— подмодуль модуля M .

Для любого подмодуля $N \subset M$ определяется канонический гомоморфизм

$$\pi: M \rightarrow M/N, \quad x \mapsto x + N,$$

ядром которого является N .

Теорема 1 (о гомоморфизме модулей). Пусть $f: M \rightarrow N$ — гомоморфизм A -модулей. Тогда

$$\text{Im } f \simeq M/\text{Ker } f.$$

Более точно, имеется изоморфизм

$$\varphi: \text{Im } f \xrightarrow{\sim} M/\text{Ker } f,$$

ставящий в соответствие каждому элементу $y = f(x) \in \text{Im } f$ смежный класс $\pi(x) = x + \text{Ker } f$.

Доказательство. Благодаря теореме 4.6.1 мы уже знаем, что отображение φ является изоморфизмом аддитивных групп. Остается только проверить, что оно перестановочно с умножениями на элементы кольца A . Пусть $f(x) = y$. Тогда $f(ax) = ay$ при $a \in A$ и

$$\varphi(ay) = \pi(ax) = a\pi(x) = a\varphi(y). \quad \square$$

Пусть M — некоторый A -модуль.

Для любого подмножества $S \subset M$ совокупность всех линейных комбинаций

$$a_1x_1 + \dots + a_kx_k \quad (x_i \in S, a_i \in A)$$

есть наименьший подмодуль, содержащий S . Он называется *подмодулем, порожденным подмножеством S* , и обозначается через $\langle S \rangle$. Если $\langle S \rangle = M$, то говорят, что модуль M порождается подмножеством S или что S — *система порождающих* модуля M . Модуль, допускающий конечную систему порождающих, называется *конечно порожденным*.

Модуль, порождаемый одним элементом, называется *циклическим*.

Идеал

$$\text{Ann } M = \{a \in A: aM = 0\}$$

называется *аннулятором* модуля M . Если $\text{Ann } M \neq 0$, то модуль называется *периодическим*.

Теорема 2. Всякий циклический A -модуль M изоморфен модулю вида A/I , где I — левый идеал кольца A . Если кольцо A коммутатив-

но, то идеал I совпадает с $\text{Ann } M$ и тем самым определен модулем M однозначно.

Доказательство. Пусть $M = \langle x \rangle$ — циклический A -модуль. Отображение

$$f: A \rightarrow M, \quad a \mapsto ax,$$

является гомоморфизмом модулей, причем $\text{Im } f = M$. По теореме о гомоморфизме $M \simeq A/I$, где $I = \text{Ker } f$. Второе утверждение теоремы очевидно. \square

Система $\{x_1, \dots, x_n\}$ элементов модуля M называется *линейно независимой*, если $a_1x_1 + \dots + a_nx_n = 0$ ($a_i \in A$) только при $a_1 = \dots = a_n = 0$. Линейно независимая система порождающих называется *базисом*.

Конечно порожденный модуль, обладающий базисом, называется *свободным*. Свободный циклический модуль изоморфен A (как A -модуль).

Для конечно порожденных модулей над кольцами главных идеалов можно построить теорию, вполне аналогичную теории конечно порожденных абелевых групп.

Начиная с этого момента, мы будем предполагать, что A — кольцо главных идеалов.

Теорема 3. *Все базисы свободного A -модуля L содержат одно и то же число элементов.*

Доказательство. Пусть p — какой-либо простой элемент кольца A . Тогда $A/(p)$ — поле и L/pL — векторное пространство над этим полем. Если $\{e_1, \dots, e_n\}$ — базис модуля L , то $\{[e_1], \dots, [e_n]\}$ (где $[x]$ обозначает класс $x + pL$) — базис этого векторного пространства. Следовательно, $n = \dim L/pL$. \square

Число элементов базиса свободного модуля L называется его *рангом* и обозначается через $\text{rk } L$.

Теорема 4. *Всякий подмодуль N свободного A -модуля L ранга n является свободным A -модулем ранга $t \leq n$, причем существует такой базис $\{e_1, \dots, e_n\}$ модуля L и такие (ненулевые) элементы $u_1, \dots, u_m \in A$, что $\{u_1e_1, \dots, u_me_m\}$ — базис подмодуля N и $u_i \mid u_{i+1}$ при $i = 1, \dots, t - 1$.*

Доказательство. Первое утверждение теоремы при $n = 1$ есть определение кольца главных идеалов; при $n > 1$ она доказывается точно так же, как для $A = \mathbb{Z}$ (см. теорему 1.3).

Доказательство второго утверждения, как и в случае $A = \mathbb{Z}$, основано на приведении матрицы C перехода от базиса модуля L к бази-

су модуля N к диагональному виду с помощью элементарных преобразований этих базисов.

В случае, когда A — евклидово кольцо, элементарными преобразованиями базиса A -модуля называются:

- 1) прибавление к одному элементу другого, умноженного на элемент кольца A ;
- 2) перестановка двух элементов;
- 3) умножение одного элемента на обратимый элемент кольца A .

Приведение матрицы C к диагональному виду в этом случае может быть осуществлено так же, как в доказательстве предложения 1.1, с той оговоркой, что минимизировать следует не сам элемент c_{11} (что не имеет смысла), а его норму.

В общем случае понятие элементарного преобразования следует расширить. Назовем *квазиэлементарным преобразованием* базиса $\{x_1, \dots, x_p\}$ какого-либо A -модуля замену двух элементов x_i и x_j их линейными комбинациями

$$ax_i + bx_j, \quad cx_i + dx_j,$$

где $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ — обратимая матрица с элементами из кольца A . (Обратимость матрицы равносильна обратимости ее определителя.) Ясно, что преобразование, обратное к квазиэлементарному, также является квазиэлементарным и что элементарные преобразования являются квазиэлементарными.

Любую пару элементов $\{x, y\}$ самого кольца A с помощью квазиэлементарного преобразования можно привести к виду $\{d, 0\}$, где $d = (x, y)$. В самом деле, существуют такие $a, b \in A$, что $ax + by = d$.

Рассмотрим матрицу $\begin{pmatrix} a & b \\ -y/d & x/d \end{pmatrix}$. Она обратима, так как ее определитель равен 1. Соответствующее квазиэлементарное преобразование переводит $\{x, y\}$ в $\{d, 0\}$.

Следовательно, если в каком-либо столбце или какой-либо строке матрицы C имеются элементы x, y , то с помощью квазиэлементарного преобразования строк или столбцов из них можно получить элементы $d, 0$. Такого рода преобразований достаточно, чтобы, следуя в целом доказательству предложения 1.1, привести матрицу C к диагональному виду. \square

Изучим теперь строение произвольных конечно порожденных A -модулей.

Всякий нетривиальный циклический A -модуль изоморфен либо A , либо $A/(u)$, где u — необратимый ненулевой элемент.

Если $(u, v) = 1$, то изоморфизм колец

$$A/(u, v) \xrightarrow{\sim} A/(u) \oplus A/(v),$$

построенный в доказательстве теоремы 2.5, является, как легко понять, и изоморфизмом A -модулей. Следовательно, если $u = p_1^{k_1} \dots p_s^{k_s}$ — разложение элемента u на простые множители, то имеет место изоморфизм A -модулей

$$A/(u) \simeq A/(p_1^{k_1}) \oplus \dots \oplus A/(p_s^{k_s}). \quad (17)$$

Определение 2. Конечно порожденный A -модуль M , аннулятор которого содержит степень простого элемента $p \in A$, называется *примарным* или, точнее, *p -примарным*.

Таким образом, всякий периодический циклический A -модуль разлагается в прямую сумму примарных циклических подмодулей.

Теорема 5. *Всякий конечно порожденный A -модуль M разлагается в прямую сумму примарных и свободных циклических подмодулей, причем набор аннуляторов этих подмодулей определен однозначно.*

Доказательство этой теоремы аналогично доказательству теоремы 1.6. В частности, существование требуемого разложения выводится из теоремы 4 и изоморфизма (17). Для доказательства его единственности (в указанном смысле) следует рассмотреть *подмодуль кручения*

$$\text{Tor } M \doteq \{x \in M : ax = 0 \text{ для некоторого } a \in A, a \neq 0\}$$

и, для каждого простого элемента $p \in A$, *подмодуль p -кручения*

$$\text{Tor}_p M \doteq \{x \in M : p^k x = 0 \text{ для некоторого } k \in \mathbb{Z}_+\}.$$

Единственность разложения примарного модуля в прямую сумму примарных циклических подмодулей доказывается по индукции, как и в случае абелевых групп. Однако соображение, использовавшее порядок группы, в общем случае не работает. Вместо него можно применить следующее соображение: если модуль M разложен в прямую сумму p -примарных циклических подмодулей, то число слагаемых равно размерности подмодуля $\{x \in M : px = 0\}$ как векторного пространства над полем $A/(p)$. \square

Так же, как и в случае абелевых групп, для всякого периодического A -модуля M одновременно с доказательством теоремы 5 получается, что

$$M \simeq A/(u_1) \oplus \dots \oplus A/(u_m), \quad (18)$$

где u_1, \dots, u_m — такие необратимые ненулевые элементы кольца A , что $u_i \mid u_{i+1}$ при $i = 1, \dots, m - 1$. Элементы u_1, \dots, u_m определены однозначно с точностью до умножения на обратимые элементы. Они называются *инвариантными множителями* модуля M . Очевидно, что

$$\text{Ann } M = (u_m). \quad (19)$$

В случае $A = K[t]$ (K — поле) теорема 5 описывает строение линейных операторов в векторных пространствах над полем K (см. пример 1). Условие конечной порожденности заведомо выполняется, если векторное пространство конечномерно. Более того, в этом случае отсутствуют свободные слагаемые, так как свободный циклический модуль над $K[t]$ имеет бесконечную размерность над K . Результат выглядит особенно просто, если поле K алгебраически замкнуто. В этом случае примарные циклические модули имеют вид

$$K[t]/((t - \lambda)^m) \quad (\lambda \in K).$$

Такой модуль является m -мерным векторным пространством над K с базисом

$$\{[(t - \lambda)^{m-1}], \dots, [t - \lambda], [1]\},$$

где $[f(t)]$ обозначает класс $f(t) + ((t - \lambda)^m)$. Оператор умножения на t записывается в этом базисе жордановой клеткой

$$J(\lambda) = \underbrace{\begin{pmatrix} \lambda & 1 & & & 0 \\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda & 1 \\ 0 & & & & \lambda \end{pmatrix}}_m.$$

Из предыдущего вытекает

Теорема 6. *Всякий линейный оператор в конечномерном векторном пространстве над алгебраически замкнутым полем в некотором базисе записывается жордановой матрицей, причем эта матрица определена однозначно с точностью до перестановки диагональных клеток.*

Напомним, что первое утверждение этой теоремы было доказано другим способом в § 6.4.

Из (19) следует, что последний инвариантный множитель $K[t]$ -модуля, ассоциированного с линейным оператором \mathcal{A} , — это не что иное, как минимальный многочлен оператора \mathcal{A} (ср. теорему 6.5.1).

Задача 1. Доказать, что оператор умножения на t в $K[t]$ -модуле $K[t]/(h(t))$, где $h(t) = t^n + a_1 t^{n-1} + \dots + a_{n-1} t + a_n$, имеет в базисе $\{[t^{n-1}], [t^{n-2}], \dots, [t], [1]\}$ матрицу

$$\begin{pmatrix} -a_1 & 1 & 0 & \dots & 0 & 0 \\ -a_2 & 0 & 1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ -a_{n-1} & 0 & 0 & \dots & 0 & 1 \\ -a_n & 0 & 0 & \dots & 0 & 0 \end{pmatrix},$$

а его характеристический многочлен равен $h(t)$. Вывести отсюда, что произведение инвариантных множителей $K[t]$ -модуля, ассоциированного с любым линейным оператором \mathcal{A} , равно характеристическому многочлену оператора \mathcal{A} .

Задача 2. Вывести из предыдущей задачи теорему Гамильтона—Кэли (следствие 3 теоремы 6.5.1).

Задача 3. Получить канонический вид матрицы линейного оператора над полем вещественных чисел.

Задача 4. Получить канонический вид матрицы линейного оператора в четырехмерном векторном пространстве над полем \mathbb{Z}_2 .

§ 4. Нётеровы кольца

Начиная с этого момента и до конца главы термин «кольцо» означает «коммутативное ассоциативное кольцо с единицей». Подкольца предполагаются содержащими единицу, гомоморфизмы колец — переводящими единицу в единицу.

Естественным расширением класса колец главных идеалов является класс нётеровых колец.

Определение 1. Кольцо A называется *нётеровым*, если выполняется любое из следующих эквивалентных условий:

- 1) всякий идеал порождается конечным числом элементов;
- 2) не существует бесконечной строго возрастающей цепочки идеалов $I_1 \subset I_2 \subset \dots \subset I_n \subset \dots$ ($I_n \neq I_{n+1}$).

Напомним, что идеалом, порожденным элементами $x_1, \dots, x_n \in A$, называется идеал $(x_1, \dots, x_n) = \{a_1x_1 + \dots + a_nx_n : a_1, \dots, a_n \in A\}$.

Эквивалентность условий 1) и 2) доказывается следующим образом. Пусть $I_1 \subset I_2 \subset \dots$ — возрастающая последовательность идеалов. Тогда $I = \bigcup_{n=1}^{\infty} I_n$ есть также идеал. Если он порождается конечным числом элементов, то все они принадлежат идеалу I_n для некоторого достаточно большого n и, значит, $I = I_n$, так что последовательность не является строго возрастающей.

Обратно, если некоторый идеал I не порождается конечным числом элементов, то существует такая последовательность элементов $x_1, x_2, \dots \in I$, что последовательность идеалов

$$(x_1) \subset (x_1, x_2) \subset \dots$$

является строго возрастающей.

Предложение 1. *Всякое факторкольцо A/I нётерова кольца нётерово.*

Доказательство. Пусть J — идеал кольца A/I . Тогда его полный прообраз при каноническом гомоморфизме $\pi: A \rightarrow A/I$ является идеалом кольца A , и если этот идеал порождается элементами $x_1, \dots, x_n \in A$, то идеал J порождается элементами $\pi(x_1), \dots, \pi(x_n)$. \square

Конечно порожденные модули над произвольными нётеровыми кольцами устроены не так просто, как над кольцами главных идеалов. Однако следующая теорема показывает, что и они в некотором смысле похожи на конечномерные векторные пространства.

Теорема 1. *Всякий подмодуль N конечно порожденного модуля M над нётеровым кольцом A конечно порожден.*

Замечание 1. В случае $M = A$ (т. е. когда M — свободный циклический модуль) утверждение этой теоремы совпадает с первым определением нётерова кольца.

Доказательство. Пусть $M = \langle x_1, \dots, x_n \rangle$. Докажем утверждение теоремы индукцией по n .

При $n = 1$ можно считать, что $M = A/I$, где I — идеал кольца A (см. теорему 3.2). Тогда N — идеал кольца A/I , и утверждение теоремы вытекает из предложения 1.

При $n > 1$ рассмотрим подмодуль $M_1 = \langle x_1, \dots, x_{n-1} \rangle \subset M$ и положим $N_1 = N \cap M_1$. По предположению индукции модуль N_1 конечно порожден. Пусть $N_1 = \langle y_1, \dots, y_k \rangle$. Фактормодуль N/N_1 является под-

модулем циклического модуля M/M_1 и конечно порожден по уже доказанному. Пусть $N/N_1 = \langle z_1 + N_1, \dots, z_l + N_1 \rangle$. Тогда

$$N = \langle y_1, \dots, y_k, z_1, \dots, z_l \rangle. \quad \square$$

Как можно доказать нётеровость какого-либо кольца? Одним из основных инструментов для этого является следующая теорема.

Теорема 2 (теорема Гильберта о базисе идеала). *Кольцо многочленов $A[x]$ над нётеровым кольцом A нётерово.*

Доказательство. Пусть I — идеал кольца $A[x]$. Обозначим через $A[x]_n$ совокупность многочленов степени $\leq n$. Это свободный A -модуль с базисом $\{1, x, \dots, x^n\}$. Положим $I_n = I \cap A[x]_n$. По теореме 1 I_n — конечно порожденный A -модуль. Очевидно, что $I = \bigcup_{n=0}^{\infty} I_n$.

Обозначим через J_n совокупность коэффициентов при x^n всех многочленов из I_n . Очевидно, что это идеал кольца A и что $J_n \subset J_{n+1}$. В силу нётеровости кольца A существует такое m , что $J_n = J_m$ при всех $n \geq m$. Поэтому для всякого многочлена $f \in I_n$ ($n \geq m$) найдется такой многочлен $g \in I_m$, что $f - x^{n-m}g \in I_{n-1}$. Это показывает, что идеал I кольца $A[x]$ порождается своим подмножеством I_m . Следовательно, если I_m порождается какими-то многочленами f_1, \dots, f_k как A -модуль, то I порождается этими же многочленами как $A[x]$ -модуль. \square

Следствие 1. *Кольцо многочленов от любого числа переменных над нётеровым кольцом нётерово.*

Говорят, что кольцо B порождается элементами u_1, \dots, u_n над подкольцом A , если каждый его элемент может быть представлен в виде многочлена от u_1, \dots, u_n с коэффициентами из A . В этом случае имеется гомоморфизм

$$f: A[x_1, \dots, x_n] \xrightarrow{\text{на}} B, \quad x_i \mapsto u_i$$

(где $A[x_1, \dots, x_n]$ обозначает кольцо многочленов от x_1, \dots, x_n с коэффициентами из A), и, следовательно,

$$B \simeq A[x_1, \dots, x_n]/\text{Ker } f.$$

Часто пишут $B = A[u_1, \dots, u_n]$, хотя это не означает, что B есть кольцо многочленов от n независимых переменных (между u_1, \dots, u_n могут быть алгебраические зависимости).

Следствие 2. *Всякое кольцо, конечно порожденное над нётеровым подкольцом, нётерово.*

Замечание 2. Имеется следующий «абсолютный» вариант следствия 2, не привязанный ни к какому выделенному подкольцу.

Говорят, что кольцо порождается элементами u_1, \dots, u_n , если каждый его элемент может быть представлен в виде многочлена от u_1, \dots, u_n с целыми коэффициентами. В этом случае оно изоморфно факторкольцу кольца $\mathbb{Z}[x_1, \dots, x_n]$ и, следовательно, является нётеровым. Таким образом, *всякое конечно порожденное кольцо нётерово.*

При работе с кольцами делители нуля, если они есть, часто доставляют неприятности. Существуют методы борьбы с ними. Наиболее «злостными» из делителей нуля являются нильпотентные элементы.

Элемент a кольца A называется *нильпотентным*, если $a^m = 0$ для некоторого натурального m . Легко видеть, что совокупность всех нильпотентных элементов является идеалом кольца A . Он называется (*нильпотентным*) *радикалом* кольца A и обозначается через $\text{rad } A$. Факторкольцо $A/\text{rad } A$ уже не имеет нильпотентных элементов (отличных от нуля).

Пример 1. Пусть A — кольцо главных идеалов. Найдем $\text{rad}(A/(u))$, где $u \in A$ — ненулевой необратимый элемент. Пусть $u = p_1^{k_1} \dots p_s^{k_s}$ — разложение элемента u на простые множители. Элемент $a + (u) \in A/(u)$ нильпотентен тогда и только тогда, когда $a^n \in (u)$ для некоторого натурального n ; но из единственности разложения на простые множители в кольце A следует, что это имеет место тогда и только тогда, когда a делится на $p_1 \dots p_s$. Таким образом,

$$\text{rad}(A/(u)) = (p_1 \dots p_s)/(u).$$

Задача 1. Доказать, что

$$\text{rad}(A_1 \oplus \dots \oplus A_k) = \text{rad } A_1 \oplus \dots \oplus \text{rad } A_k.$$

Определение 2. Идеал I кольца A , не равный A , называется *простым*, если факторкольцо A/I не имеет делителей нуля.

Это означает, что из $ab \in I$ следует, что $a \in I$ или $b \in I$.

Например, в кольце главных идеалов A ненулевой идеал (p) является простым тогда и только тогда, когда p — простой элемент.

Идеал I кольца A , не равный A , называется *максимальным*, если он не содержится ни в каком большем идеале, не равном A . Из второго определения нётерова кольца следует, что в нётеровом кольце имеется хотя бы один максимальный идеал.

Предложение 2. *Идеал I кольца A максимален тогда и только тогда, когда факторкольцо A/I является полем.*

Доказательство. Очевидно, что идеал I максимален тогда и только тогда, когда в факторкольце A/I нет нетривиальных идеалов. Мы видели (см. пример 2.1), что в поле нет нетривиальных идеалов. Обратное, пусть в каком-то кольце K нет нетривиальных идеалов. Тогда для любого ненулевого элемента $a \in K$ идеал (a) совпадает с K и, в частности, содержит единицу, а это как раз и означает, что элемент a обратим. Следовательно, K — поле. \square

Следствие. *Всякий максимальный идеал прост.*

Теорема 3. *Радикал нётерова кольца A совпадает с пересечением всех простых идеалов.*

Доказательство. Если элемент a не нильпотентен, то мы можем построить кольцо $A' = A[a^{-1}]$ дробей вида b/a^n ($b \in A$, $n \in \mathbb{Z}_+$) аналогично тому, как было построено поле дробей целостного кольца в § 3.10. Ввиду следствия 2 теоремы 2 кольцо A' нётерово. Поэтому в нем имеется какой-то максимальный идеал I' . Так как элемент a обратим в A' , то $a \notin I'$. Положим $I = I' \cap A$. Тогда кольцо A/I вкладывается в поле A'/I' и, следовательно, не имеет делителей нуля. Таким образом, I — простой идеал кольца A , не содержащий a . \square

Замечание 3. Используя трансфинитные средства (например, лемму Цорна), легко доказать, что в любом (не обязательно нётеровом) кольце имеется максимальный идеал. Отсюда следует, что и теорема 3 на самом деле справедлива для любых колец.

§ 5. Алгебраические расширения

Если кольцо A является подкольцом кольца B , то говорят, что B — *расширение* кольца A . В этом случае B — не просто кольцо: оно является алгеброй над A , что дает дополнительные возможности для его изучения. (Определение алгебры над кольцом такое же, как над полем.)

Введем некоторую терминологию, относящуюся к этой ситуации.

Элемент $u \in B$ называется *алгебраическим над A* , если он удовлетворяет некоторому нетривиальному алгебраическому уравнению с коэффициентами из A , и *трансцендентным* в противном случае. В частности, любой элемент $a \in A$ алгебраичен над A , так как он

удовлетворяет линейному уравнению $x - a = 0$. Кольцо B называется *алгебраическим расширением* кольца A , если всякий его элемент алгебраичен над A .

Более общо, элементы $u_1, \dots, u_n \in B$ называются *алгебраически зависимыми* над A , если они удовлетворяют некоторому нетривиальному алгебраическому уравнению (с n неизвестными) с коэффициентами из A .

Совокупность элементов кольца B , которые могут быть представлены в виде $f(u_1, \dots, u_n)$, где f — многочлен с коэффициентами из A , является его подкольцом (содержащим A). Оно называется *подкольцом, порожденным над A элементами u_1, \dots, u_n* , и обозначается через $A[u_1, \dots, u_n]$. Если u_1, \dots, u_n алгебраически независимы, то оно изоморфно кольцу многочленов от n переменных с коэффициентами из A ; в общем случае оно изоморфно факторкольцу кольца многочленов по идеалу алгебраических зависимостей между u_1, \dots, u_n . Расширение B кольца A называется *конечно порожденным*, если существуют такие элементы $u_1, \dots, u_n \in B$, что $B = A[u_1, \dots, u_n]$.

Если кольцо B (а, значит, и A) не имеет делителей нуля, то можно рассмотреть поля отношений $K = \text{Quot } A$, $L = \text{Quot } B$ и считать, что все действие разворачивается в «большом» поле L . Имеет место следующая диаграмма включений:

$$\begin{array}{ccc} A & \subset & B \\ \cap & & \cap \\ K & \subset & L \end{array} \quad (20)$$

Если элементы $u_1, \dots, u_n \in L$ алгебраически зависимы над K , то они алгебраически зависимы и над A , так как коэффициенты алгебраической зависимости можно сделать «целыми», т. е. принадлежащими A , умножив всю зависимость на их общий знаменатель.

Рассмотрим вначале алгебраические расширения полей. Ключом к их пониманию является вводимое ниже понятие конечно расширения, а единственной идеей доказательств приводимых ниже утверждений — использование теоремы о том, что всякое подпространство конечномерного векторного пространства конечномерно.

Если поле L является расширением поля K , то его можно рассматривать как векторное пространство над K . Размерность этого векторного пространства обозначается через $\dim_K L$.

Определение 1. Расширение L поля K называется *конечным*, если $\dim_K L < \infty$. Число $\dim_K L$ в этом случае называется *степенью* расширения L .

Способ получения конечных расширений полей дается следующей теоремой.

Теорема 1. Пусть $h \in K[x]$ — неприводимый многочлен степени n . Тогда $L = K[x]/(h)$ — конечное расширение поля K , причем $\dim_K L = n$.

Доказательство. Тот факт, что L — поле, вытекает из общей теоремы 2.4. Далее, из возможности и единственности деления с остатком в $K[x]$ следует, что всякий элемент из L однозначно представляется в виде

$$a_0 + a_1x + \dots + a_{n-1}x^{n-1} + (h) \quad (a_0, a_1, \dots, a_{n-1} \in K).$$

Это означает, что смежные классы

$$1 + (h), \quad x + (h), \quad \dots, \quad x^{n-1} + (h)$$

составляют базис поля L над K . □

Элемент $\alpha = x + (h) \in L$ является, очевидно, корнем многочлена h в поле L , причем $L = K[\alpha]$. Поэтому переход от поля K к полю L называется *присоединением* к полю K корня неприводимого многочлена h .

Расширения описанного типа называются *простыми*. В § 11.6 мы покажем, что всякое конечное расширение поля нулевой характеристики является простым (теорема о примитивном элементе). Однако это обстоятельство не играет существенной роли для понимания дальнейшего, и мы пока не будем доказывать (и использовать) эту теорему.

Пример 1. Если $a \in K$ — элемент, не являющийся квадратом в поле K , то поле $K[\sqrt{a}]$, получающееся присоединением к K корня многочлена $x^2 = a$, является расширением степени 2, или, как говорят, *квадратичным расширением* поля K . В частности, $\mathbb{R}[\sqrt{-1}] = \mathbb{C}$.

Пусть L — какое-то расширение поля K .

Если элемент $u \in L$ алгебраичен над K , то совокупность всех многочленов $f \in K[x]$, для которых $f(u) = 0$, является ненулевым идеалом кольца $K[x]$. Порождающий элемент этого идеала называется *минимальным многочленом* элемента u и обозначается через m_u (ср. определение минимального многочлена линейного оператора в § 6.5). Отметим, что минимальный многочлен неприводим. В самом деле, если $m_u = fg$, то либо $f(u) = 0$, либо $g(u) = 0$, так

что один из многочленов f и g должен иметь такую же степень, что и m_u . Степень многочлена m_u называется *степенью элемента* u над K .

Теорема 2. *Элемент $u \in L$ алгебраичен над K тогда и только тогда, когда $K[u]$ — конечномерное векторное пространство над K . При этом условии $K[u]$ есть поле и его размерность над K равна степени u над K .*

Доказательство. Если пространство $K[u]$ конечномерно над K , то оно порождается конечным числом степеней элемента u . Следовательно, найдется такое n , что u^n линейно выражается через предыдущие степени, а это и означает, что u алгебраичен над K .

Обратно, пусть u — алгебраический элемент степени n над K . Тогда u^n линейно выражается через предыдущие степени элемента u . Последовательно умножая это выражение на u и заменяя образующую при этом n -ю степень элемента u ее выражением через предыдущие степени, мы получаем, что и любая степень элемента u , а значит, и любой элемент пространства $K[u]$, линейно выражается через $1, u, \dots, u^{n-1}$. Следовательно, $\dim_K K[u] \leq n$.

Более точно, рассмотрим гомоморфизм

$$\varphi: K[x] \rightarrow L, \quad f \mapsto f(u).$$

Его образ есть $K[u]$, а ядро есть идеал, порожденный минимальным многочленом m_u элемента u . Следовательно,

$$K[u] \simeq K[x]/(m_u).$$

Так как многочлен m_u неприводим, то, согласно теореме 1, $K[u]$ есть поле и его размерность над K равна $\deg m_u = n$. \square

Следствие. *Всякое конечное расширение поля является алгебраическим.*

Пример 2. Пусть p — простое число. Так как число $\varepsilon_p = \cos \frac{2\pi}{p} + i \sin \frac{2\pi}{p} \in \mathbb{C}$ является корнем многочлена $x^{p-1} + \dots + x + 1$, неприводимого над \mathbb{Q} (пример 3.6.2), то $\mathbb{Q}[\varepsilon_p]$ есть расширение степени $p-1$ поля \mathbb{Q} . Оно содержит все комплексные корни p -й степени из 1 и называется *круговым полем* (или *полем деления круга*).

Теорема 3. *Если L — конечное расширение поля K , а M — конечное расширение поля L , то M — конечное расширение поля K , причем*

$$\dim_K M = \dim_K L \cdot \dim_L M.$$

Доказательство. Если $\{e_i\}$ — базис расширения L над K , а $\{f_j\}$ — базис расширения M над L , то $\{e_i f_j\}$ — базис расширения M над K . \square

Для любых элементов $u_1, \dots, u_n \in L$ совокупность элементов поля L , которые представимы в виде отношения элементов кольца $K[u_1, \dots, u_n]$, является подполем, изоморфным $\text{Quot } K[u_1, \dots, u_n]$. Оно называется *подполем, порожденным над K элементами u_1, \dots, u_n* , и обозначается через $K(u_1, \dots, u_n)$. В частности, если $u \in L$ — алгебраический над K элемент, то, согласно теореме 2, $K(u) = K[u]$ (феномен «уничтожения иррациональности в знаменателе»).

Если $K(u_1, \dots, u_n) = L$, то говорят, что поле L порождается над K элементами u_1, \dots, u_n .

Теорема 4. Если поле L порождается над K конечным числом алгебраических элементов u_1, \dots, u_n , то оно является конечным расширением поля K .

Доказательство. Рассмотрим «башню расширений»

$$K \subset K(u_1) \subset K(u_1, u_2) \subset \dots \subset K(u_1, \dots, u_n) = L.$$

Так как $K(u_1, \dots, u_m) = K(u_1, \dots, u_{m-1})(u_m)$ и элемент u_m , будучи алгебраичен над K , тем более алгебраичен над $K(u_1, \dots, u_{m-1})$, то все «этажи» башни являются конечными расширениями. По теореме 3 отсюда следует, что и L — конечное расширение поля K . \square

Теорема 5. Пусть L — какое-либо расширение поля K . Совокупность \bar{K} всех элементов поля L , алгебраических над K , является подполем, алгебраически замкнутым в L .

(Последнее означает, что всякий элемент поля L , алгебраический над \bar{K} , принадлежит \bar{K} , т. е. алгебраичен уже над K .)

Доказательство. Если $u, v \in \bar{K}$, то, согласно теореме 4, $K(u, v) \subset \bar{K}$; в частности,

$$u + v, uv, u^{-1} \in \bar{K}.$$

Это означает, что \bar{K} — подполе поля L .

Пусть $u \in L$ — элемент, алгебраический над \bar{K} , и пусть $u_1, \dots, u_n \in \bar{K}$ — коэффициенты алгебраического уравнения, корнем которого он является. По теореме 4 $K' = K(u_1, \dots, u_n)$ — конечное расширение поля K . Так как элемент u алгебраичен над K' , то $K'(u)$ — конечное расширение поля K' . Следовательно, $K'(u)$ — конечное расширение поля K и, значит, $K'(u) \subset \bar{K}$; в частности, $u \in \bar{K}$. \square

Поле \bar{K} называется *алгебраическим замыканием* поля K в L .

Например, поле всех алгебраических чисел есть алгебраическое замыкание $\bar{\mathbb{Q}}$ поля \mathbb{Q} в \mathbb{C} . Так как поле \mathbb{C} алгебраически замкнуто, то и $\bar{\mathbb{Q}}$ алгебраически замкнуто (в абсолютном смысле, а не только в \mathbb{C}). Всякое конечное расширение поля \mathbb{Q} называется *полем алгебраических чисел* (так что существует много различных полей алгебраических чисел). Нетрудно доказать, что всякое поле алгебраических чисел изоморфно подполю поля $\bar{\mathbb{Q}}$ (проделайте это!).

В простом расширении поля K , полученном присоединением корня неприводимого многочлена f , этот многочлен не обязан (хотя и может) иметь более одного корня. Если мы хотим получить поле, в котором f разлагается на линейные множители, необходимо, вообще говоря, дальнейшее расширение.

Определение 2. Расширение L поля K называется *полем разложения* многочлена $f \in K[x]$ (не обязательно неприводимого), если f разлагается в $L[x]$ на линейные множители и поле L порождается над K его корнями.

Гомоморфизмы (в частности, изоморфизмы) расширений поля K , тождественные на K , называются *гомоморфизмами (изоморфизмами) над K* .

Теорема 6. Поле разложения любого многочлена $f \in K[x]$ существует и единственно с точностью до изоморфизма над K .

Для доказательства второй части теоремы нам понадобится

Лемма 1. Пусть $P(\alpha)$ — расширение поля P , полученное присоединением корня α неприводимого многочлена $h \in P[x]$, и φ — гомоморфизм поля P в некоторое поле F . Гомоморфизм φ продолжается до гомоморфизма $\psi: P(\alpha) \rightarrow F$ ровно столькоими способами, сколько различных корней имеет в F многочлен h^φ , полученный из многочлена h применением к его коэффициентам гомоморфизма φ .

Доказательство. Искомое продолжение ψ , если оно существует, задается формулой

$$\psi(a_0 + a_1\alpha + \dots + a_m\alpha^m) = \varphi(a_0) + \varphi(a_1)\beta + \dots + \varphi(a_m)\beta^m$$

$$(a_0, a_1, \dots, a_m \in P), \quad (21)$$

где $\beta = \psi(\alpha)$ — некоторый элемент поля F . Применяя эту формулу к равенству $h(\alpha) = 0$, получаем, что $h^\varphi(\beta) = 0$. Обратно, если $\beta \in F$ — корень многочлена h^φ , то формула (21) корректно определяет гомоморфизм $\psi: P(\alpha) \rightarrow F$. \square

Доказательство теоремы 6. Рассмотрим последовательность расширений

$$K = K_0 \subset K_1 \subset K_2 \subset \dots,$$

в которой K_i получается из K_{i-1} присоединением корня какого-либо неприводимого множителя f_i степени > 1 многочлена f над K_{i-1} . Так как число неприводимых множителей многочлена f каждый раз увеличивается, то эта последовательность не может быть бесконечной. Последний ее член $K_s = L$ и является полем разложения многочлена f .

Пусть теперь \tilde{L} — другое поле разложения. Построим последовательность гомоморфизмов

$$\varphi_i: K_i \rightarrow \tilde{L} \quad (i=0, 1, \dots, s)$$

так, чтобы

$$\varphi_0 = \text{id}, \quad \varphi_i|_{K_{i-1}} = \varphi_{i-1}.$$

Согласно лемме, i -й шаг этого построения будет возможен, если многочлен $\tilde{f}_i = f_i^{\varphi_{i-1}}$ имеет корень в \tilde{L} . Так как f_i делит f в $K_{i-1}[x]$, то \tilde{f}_i делит f в $\tilde{L}[x]$. Но многочлен f разлагается в $\tilde{L}[x]$ на линейные множители и, следовательно, любой его делитель положительной степени имеет корень в \tilde{L} . Таким образом, искомые гомоморфизмы φ_i существуют. Последний из них

$$\varphi_s = \varphi: L \rightarrow \tilde{L}$$

является изоморфизмом, так как, согласно определению поля разложения, поле \tilde{L} является минимальным расширением поля K , над которым многочлен f разлагается на линейные множители. \square

Пример 3. Найдем степень поля разложения L кубического многочлена

$$f = x^3 + a_1x^2 + a_2x + a_3 \in K[x], \quad \text{char } K \neq 2.$$

Рассмотрим различные случаи.

1) f имеет 3 корня в K . Тогда $L = K$.

2) f имеет 1 корень в K . Тогда L — квадратичное расширение поля K .

3) f не имеет корней в K и, следовательно, неприводим над K . Пусть тогда $K_1 \supset K$ — кубическое расширение, полученное присоединением корня α_1 многочлена f . Могут представиться два случая:

а) f имеет 3 корня в K_1 ; тогда $L = K_1$;

б) f имеет только 1 корень в K_1 ; тогда L — квадратичное расширение поля K_1 и, следовательно, $\dim_K L = 6$.

Для различения случаев 3а) и 3б) рассмотрим дискриминант многочлена f , равный по определению

$$D = (\alpha_1 - \alpha_2)^2(\alpha_1 - \alpha_3)^2(\alpha_2 - \alpha_3)^2,$$

где $\alpha_1, \alpha_2, \alpha_3$ — корни многочлена f в L . (Выражение D через коэффициенты многочлена f см. в § 3.9.) Докажем, что если многочлен f не имеет корней в K , то $\dim_K L = 3$ тогда и только тогда, когда $D \in K^2$.

Заметим, что если $D \notin K^2$, то $D \notin K_1^2$, так как иначе $K(\sqrt{D})$ было бы квадратичным расширением поля K , содержащимся в K_1 , что невозможно в силу формулы умножения размерностей в башне расширений (теорема 3). Поэтому

$$D \in K^2 \Leftrightarrow D \in K_1^2 \Leftrightarrow (\alpha_1 - \alpha_2)(\alpha_1 - \alpha_3)(\alpha_2 - \alpha_3) \in K_1.$$

Далее, так как $\alpha_1 \in K_1$, а α_2 и α_3 суть корни квадратного трехчлена с коэффициентами из K_1 , то

$$(\alpha_1 - \alpha_2)(\alpha_1 - \alpha_3) \in K_1$$

и, значит,

$$D \in K^2 \Leftrightarrow \alpha_2 - \alpha_3 \in K_1 \Leftrightarrow \alpha_2, \alpha_3 \in K_1 \Leftrightarrow L = K_1$$

(здесь мы использовали, что $\text{char } K \neq 2$).

Воспользуемся теперь теоремой 6 для описания всех конечных полей.

Всякое конечное поле F имеет характеристику $p > 0$, являющуюся простым числом, и его циклическая аддитивная подгруппа, порожденная единицей, есть подполе, изоморфное полю вычетов \mathbb{Z}_p . Мы будем отождествлять это подполе с \mathbb{Z}_p . Если $\dim_{\mathbb{Z}_p} F = n$, то

$$|F| = p^n.$$

Таким образом, число элементов любого конечного поля есть степень простого числа.

Теорема 7. Для любого простого p и натурального n существует поле из p^n элементов и все такие поля изоморфны.

Доказательство этой теоремы потребует некоторой подготовки.

Пусть F — любое (может быть, бесконечное) поле характеристики $p > 0$. Рассмотрим отображение

$$\varphi: F \rightarrow F, \quad x \mapsto x^p.$$

Очевидно, что $\varphi(xy) = \varphi(x)\varphi(y)$. Кроме того, как это ни странно, $\varphi(x+y) = \varphi(x) + \varphi(y)$. Действительно, как мы видели в § 1.5,

$$(x+y)^p = \sum_{k=0}^p C_p^k x^{p-k} y^k = x^p + y^p.$$

Таким образом, φ — эндоморфизм (гомоморфизм в себя) поля F . Он называется *эндоморфизмом Фробениуса*.

Так как $\text{Ker } \varphi = 0$, то $\text{Im } \varphi = F^p \simeq F$. Очевидно, что для конечного поля $F^p = F$, так что эндоморфизм Фробениуса в этом случае является автоморфизмом.

Доказательство теоремы 7. Пусть F — конечное поле из $q = p^n$ элементов. Так как мультипликативная группа F^* имеет порядок $q - 1$, то $a^{q-1} = 1$ для любого $a \in F^*$ и, значит,

$$a^q = a \quad \forall a \in F.$$

Иначе говоря, все элементы поля F являются корнями многочлена $x^q - x$. Следовательно, F — поле разложения этого многочлена над \mathbb{Z}_p . В силу теоремы 6 это показывает, что все поля из q элементов изоморфны.

С другой стороны, пусть F — поле разложения многочлена $f = x^q - x$ над \mathbb{Z}_p . Так как $f' = -1$, то многочлен f не имеет кратных корней. Его корни — это неподвижные точки автоморфизма φ^n поля F , где φ — автоморфизм Фробениуса. Легко видеть, что неподвижные точки любого автоморфизма поля образуют подполе. Таким образом, совокупность корней многочлена f есть подполе из q элементов в F (и, следовательно, совпадает с F). Тем самым доказано существование поля из q элементов. \square

Следствие. Для любого простого p и любого натурального n существует неприводимый многочлен степени n над \mathbb{Z}_p .

Доказательство. Пусть F — поле из $q = p^n$ элементов и α — порождающий элемент его мультипликативной группы (которая, как известно, циклическая). Тогда $F = \mathbb{Z}_p(\alpha)$, и, значит, минимальный многочлен элемента α над \mathbb{Z}_p имеет степень n . \square

Поле из q элементов обозначается через \mathbb{F}_q . (В частности, $\mathbb{F}_p = \mathbb{Z}_p$ при простом p .)

Пример 4. Единственным неприводимым многочленом второй степени над полем \mathbb{Z}_2 является многочлен $x^2 + x + 1$. Присоединяя к \mathbb{Z}_2 корень этого многочлена, мы получаем поле \mathbb{F}_4 .

Задача 1. Составить таблицы сложения и умножения в поле \mathbb{F}_4 .

Во всяком конечном расширении L поля K , рассматриваемом как векторное пространство над K , можно естественным образом ввести «скалярное умножение».

А именно, для любого $u \in L$ определим линейный оператор $T(u)$ в пространстве L по формуле

$$T(u)x = ux \quad (x \in L).$$

След этого оператора назовем *следом* элемента u и обозначим через $\text{tr } u$. Очевидно, что след — линейная функция на L . Определим скалярное умножение в L по формуле

$$(u, v) = \text{tr } uv. \quad (22)$$

Это симметрическая билинейная функция на L . Если $\text{char } K = 0$, то она невырождена, так как

$$(u, u^{-1}) = \text{tr } 1 = \dim_K L \neq 0$$

для любого ненулевого элемента $u \in L$.

Пример 5. Опишем скалярное умножение в поле деления круга $\mathbb{Q}(\varepsilon_p) = \mathbb{Q}[\varepsilon_p]$ (см. пример 2). Как векторное пространство над \mathbb{Q} поле $\mathbb{Q}(\varepsilon_p)$ порождается элементами $1, \varepsilon_p, \varepsilon_p^2, \dots, \varepsilon_p^{p-1}$, сумма которых равна нулю. Базис этого пространства составляют, например, элементы $1, \varepsilon_p, \varepsilon_p^2, \dots, \varepsilon_p^{p-2}$. Записав операторы $T(\varepsilon_p^k)$ в этом базисе, легко получить, что

$$\text{tr } 1 = p - 1, \quad \text{tr } \varepsilon_p^k = -1 \quad (k = 1, \dots, p - 1).$$

Следовательно,

$$(\varepsilon_p^k, \varepsilon_p^l) = \begin{cases} p - 1 & \text{при } k + l \equiv 0 \pmod{p}, \\ -1 & \text{во всех остальных случаях.} \end{cases}$$

Скалярное произведение двух элементов поля $\mathbb{Q}(\varepsilon_p)$ вычисляется особенно просто, если один из них представлен в виде рациональной линейной комбинации элементов $1, \varepsilon_p, \varepsilon_p^2, \dots, \varepsilon_p^{p-1}$ с суммой коэффициентов, равной нулю (что всегда можно сделать). А именно,

если $\sum_{k=0}^{p-1} x_k = 0$, то

$$\left(\sum_{k=0}^{p-1} x_k \varepsilon_p^k, \sum_{k=0}^{p-1} y_k \varepsilon_p^k \right) = p \left(x_0 y_0 + \sum_{k=1}^{p-1} x_k y_{p-k} \right).$$

Часть изложенных выше результатов о расширениях полей может быть обобщена на расширения нётеровых колец, если надлежащим образом видоизменить понятия алгебраического элемента и алгебраического расширения.

Пусть кольцо B является расширением кольца A . Элемент $u \in B$ называется *целым алгебраическим* или просто *целым* над A , если он удовлетворяет нетривиальному алгебраическому уравнению с коэффициентами из A и со старшим коэффициентом, равным 1. В частности, элементы самого кольца A являются целыми над A . Всякий элемент $u \in B$, алгебраический над A , становится целым после умножения на подходящий ненулевой элемент кольца A (а именно, на старший коэффициент алгебраического уравнения с коэффициентами из A , которому удовлетворяет u).

Кольцо B называется *целым расширением* кольца A или просто *целым над A* , если всякий его элемент цел над A .

В случае когда A — поле, эти определения эквивалентны определениям алгебраического элемента и алгебраического расширения.

Следующее определение является ключевым.

Определение 3. Расширение B кольца A называется *конечным*, если B является конечно порожденным A -модулем.

Ниже формулируются частичные аналоги теорем 2—5 для расширений колец. Их доказательства практически не отличаются от доказательств соответствующих утверждений теорем 2—5. Нужно только слово «базис» всюду заменить на «систему порождающих» и теорему о конечномерности любого подпространства конечномерного векторного пространства — на теорему 4.1.

Пусть B — какое-то расширение кольца A .

Теорема 8. Элемент $u \in B$ цел над A тогда и только тогда, когда $A[u]$ — конечно порожденный A -модуль.

Следствие. Всякое конечное расширение нётерова кольца является целым.

Теорема 9. Если B — конечное расширение кольца A , а C — конечное расширение кольца B , то C — конечное расширение кольца A .

Теорема 10. Если кольцо B порождается над A конечным числом целых элементов, то оно является конечным расширением кольца A .

Напомним, что конечно порожденное (и, тем более, конечное) расширение нётерова кольца также нётерово (следствие 2 теоремы 4.2).

Теорема 11. Пусть B — какое-то расширение нётерова кольца A . Совокупность \bar{A} всех элементов кольца B , целых над A , является подкольцом, целозамкнутым в B .

(Последнее означает, что всякий элемент кольца B , целый над \bar{A} , принадлежит \bar{A} , т. е. цел уже над A .)

Кольцо \bar{A} называется *целым замыканием* кольца A в B .

Например, все алгебраические числа, целые над \mathbb{Z} , — они называются *целыми алгебраическими числами* — образуют подкольцо $\bar{\mathbb{Z}}$ в поле $\bar{\mathbb{Q}}$ всех алгебраических чисел. Поле отношений кольца $\bar{\mathbb{Z}}$ совпадает с $\bar{\mathbb{Q}}$.

Замечание 1. На самом деле следствие теоремы 8 и, тем самым, теорема 11 верны для произвольных (не обязательно нётеровых) колец. Это можно установить при помощи следующего рассуждения. Пусть $B = Ae_1 + \dots + Ae_n$. Тогда $e_i e_j = \sum_k c_{ijk} e_k$ для каких-то $c_{ijk} \in A$. Если A' — какое-либо подкольцо кольца A , содержащее все c_{ijk} , то $B' = A'e_1 + \dots + A'e_n$ — подкольцо кольца B , являющееся конечным расширением кольца A' . Для любого $u = a_1 e_1 + \dots + a_n e_n$ ($a_i \in A$) возьмем в качестве A' подкольцо, порожденное всеми c_{ijk} и a_i . Так как оно нётерово (см. замечание 4.2), то согласно следствию теоремы 8 элемент $u \in B'$ цел над A' ; но тогда он тем более цел над A .

Следующая теорема устанавливает связь между конечными расширениями полей и конечными расширениями колец.

Целостное кольцо называется *нормальным* (или *целозамкнутым*), если оно целозамкнуто в своем поле отношений. Например, кольцо \mathbb{Z} нормально в силу следствия теоремы 3.6.1. Примерами не нормальных колец могут служить кольцо многочленов без линейного члена и кольцо чисел вида $a + b\sqrt{-3}$, где $a, b \in \mathbb{Z}$ (докажите, что они не нормальны!).

Теорема 12. Пусть A — нормальное нётерово кольцо, K — его поле отношений, L — конечное расширение поля K и B — целое замыкание кольца A в L . Предположим, что $\text{char } K = 0$. Тогда B — конечное расширение кольца A .

(См. диаграмму (20).)

Доказательство. Докажем вначале, что $\operatorname{tr} u \in A$ для любого $u \in B$. Пусть $a_1, \dots, a_m \in A$ таковы, что

$$u^m + a_1 u^{m-1} + \dots + a_{m-1} u + a_m = 0.$$

Тогда

$$T(u)^m + a_1 T(u)^{m-1} + \dots + a_{m-1} T(u) + a_m E = 0. \quad (23)$$

Пусть $P \supset K$ — поле разложения характеристического многочлена оператора $T(u)$. Из (23) следует, что все корни этого многочлена в поле P целы над A ; но след $\operatorname{tr} u = \operatorname{tr} T(u)$ равен их сумме, и, значит, также цел над A . С другой стороны, $\operatorname{tr} u \in K$. Ввиду нормальности кольца A отсюда вытекает, что $\operatorname{tr} u \in A$.

Пусть $\{e_1, \dots, e_n\}$ — базис L над K . Умножив e_1, \dots, e_n на подходящие элементы кольца A , можно добиться того, чтобы $e_1, \dots, e_n \in B$. Тогда $c_{ij} \doteq (e_i, e_j) \in A$ при всех i, j и

$$\Delta \doteq \det(c_{ij}) \neq 0.$$

Выясним, когда элемент

$$u = x_1 e_1 + \dots + x_n e_n \quad (x_1, \dots, x_n \in K)$$

цел над A . Это заведомо так, если $x_1, \dots, x_n \in A$. Это условие, вообще говоря, не является необходимым, но мы сейчас покажем, что коэффициенты x_1, \dots, x_n все же не могут быть «слишком дробными».

Составляя скалярные произведения элемента u с базисными векторами, находим:

$$\sum_j c_{ij} x_j = (e_i, u) \in A \quad (i = 1, \dots, n). \quad (24)$$

Рассматривая (24) как систему линейных уравнений относительно x_1, \dots, x_n , по формулам Крамера получаем, что $x_1, \dots, x_n \in \Delta^{-1} A$.

Таким образом, кольцо B содержится в A -подмодуле, порожденном элементами $\Delta^{-1} e_1, \dots, \Delta^{-1} e_n$. Так как кольцо A нётерово, то отсюда вытекает, что B — конечно порожденный A -модуль. \square

Замечание 2. Теорема 12 и ее доказательство верны и в случае, когда $\operatorname{char} K = p > 0$, при условии, что скалярное умножение в L невырожденно. Конечное расширение L поля K , удовлетворяющее этому условию, называется *сепарабельным* (см. по этому поводу замечание 11.3.2).

Пусть K — какое-либо поле алгебраических чисел (т. е. конечное расширение поля \mathbb{Q}). Целое замыкание кольца \mathbb{Z} в K называется

кольцом целых (чисел) поля K . Обозначим его через \mathbb{Z}_K . Из теоремы 12 следует, что \mathbb{Z}_K — конечно порожденная абелева группа (по сложению). Так как группа \mathbb{Z}_K не имеет кручения, то она свободна. Более того, поскольку всякий элемент поля K становится целым после умножения на подходящее целое рациональное число, базис группы \mathbb{Z}_K является в то же время базисом поля K как векторного пространства над \mathbb{Q} , так что

$$\text{rk } \mathbb{Z}_K = \dim_{\mathbb{Q}} K.$$

Задача 2. Доказать, что в поле $\mathbb{Q}(\sqrt{d})$, где d — целое число, свободное от квадратов, целыми являются числа вида $a + b\sqrt{d}$, где $a, b \in \mathbb{Z}$ либо, если $d \equiv 1 \pmod{4}$, $a, b \in \mathbb{Z} + \frac{1}{2}$.

Задача 3. Доказать, что в поле деления круга $\mathbb{Q}(\varepsilon_p)$ (см. примеры 2 и 5) целыми являются числа

$$a_0 + a_1\varepsilon_p + \dots + a_{p-2}\varepsilon_p^{p-2}, \quad a_0, a_1, \dots, a_{p-2} \in \mathbb{Z}.$$

(Указания: 1) следуя доказательству теоремы 12, доказать вначале, что знаменателями рациональных чисел a_0, a_1, \dots, a_{p-2} могут быть только степени числа p ;

2) вместо разложения по степеням числа ε_p рассматривать разложение по степеням числа $1 - \varepsilon_p$;

3) доказать, что в кольце целых поля $\mathbb{Q}(\varepsilon_p)$ имеют место следующие ассоциированности:

$$1 - \varepsilon_p \sim 1 - \varepsilon_p^k \quad (k = 1, 2, \dots, p-1), \quad p \sim (1 - \varepsilon_p)^{p-1};$$

4) доказать, что если какое-либо целое рациональное число делится на $1 - \varepsilon_p$, то оно делится на p .)

§ 6. Конечно порожденные алгебры и аффинные алгебраические многообразия

В этом параграфе мы будем рассматривать алгебры над полем K , причем под словом «алгебра» будет пониматься коммутативная ассоциативная алгебра с единицей. Алгебра A будет называться *конечно порожденной*, если она конечно порождена над K . Отметим, что в силу следствия 2 теоремы 4.2 всякая конечно порожденная алгебра является нётеровым кольцом.

Теория конечно порожденных алгебр лежит в основе исследования систем алгебраических уравнений, составляющего предмет алгебраической геометрии.

Пусть A — алгебра без делителей нуля. Элементы u_1, \dots, u_n алгебры A будут называться *алгебраически зависимыми*, если они алгебраически зависимы над K .

Определение 1. Алгебраически независимая система элементов $\{u_1, \dots, u_d\}$ называется *базисом трансцендентности* алгебры A , если для любого $u \in A$ система $\{u_1, \dots, u_d, u\}$ алгебраически зависима или, что равносильно, если алгебра A является алгебраическим расширением подалгебры $K[u_1, \dots, u_d]$, порожденной элементами u_1, \dots, u_d . (Ср. определение базиса векторного пространства.)

Например, $\{x_1, \dots, x_n\}$ — базис трансцендентности алгебры многочленов $K[x_1, \dots, x_n]$.

Предложение 1. *Всякий базис трансцендентности алгебры A является в то же время базисом трансцендентности ее поля отношения $\text{Quot } A$ (рассматриваемого как алгебра над K).*

Доказательство. Пусть $\{u_1, \dots, u_d\}$ — базис трансцендентности алгебры A . Элементы поля $\text{Quot } A$, алгебраические над подалгеброй $K[u_1, \dots, u_d]$ — это то же, что элементы, алгебраические над подполем

$$K(u_1, \dots, u_d) = \text{Quot } K[u_1, \dots, u_d].$$

Все такие элементы образуют подполе в $\text{Quot } A$ (теорема 3.5). Так как это подполе содержит A , то оно совпадает с $\text{Quot } A$. \square

Предложение 2. *Пусть $A = K[u_1, \dots, u_n]$. Тогда всякая максимальная алгебраически независимая подсистема системы $\{u_1, \dots, u_n\}$ является базисом трансцендентности алгебры A .*

Доказательство. Пусть $\{u_1, \dots, u_d\}$ — максимальная алгебраически независимая подсистема системы $\{u_1, \dots, u_n\}$. Рассмотрим алгебраическое замыкание подполя $K(u_1, \dots, u_d)$ в поле $\text{Quot } A$. По условию оно содержит элементы u_1, \dots, u_n , а следовательно, совпадает с $\text{Quot } A$ и, в частности, содержит A . \square

Следствие. *Во всякой конечно порожденной алгебре без делителей нуля существует базис трансцендентности.*

Предложение 3 (лемма о замене). *Пусть $\{u_1, u_2, \dots, u_d\}$ — базис трансцендентности алгебры A и $v \in A$ — элемент, трансцендентный над $K[u_2, \dots, u_d]$. Тогда $\{v, u_2, \dots, u_d\}$ — также базис трансцендентности алгебры A .*

Доказательство. Ясно, что элементы v, u_2, \dots, u_d алгебраически независимы. С другой стороны, элементы v, u_1, u_2, \dots, u_d алгебраически зависимы. Рассмотрим нетривиальную алгебраическую зависимость между ними. Она должна нетривиальным образом содержать u_1 . Следовательно, элемент u_1 алгебраичен над подалгеброй $K[v, u_2, \dots, u_d]$. Таким образом, алгебраическое замыкание подполя $K(v_1, u_2, \dots, u_d)$ в $\text{Quot } A$ содержит $K(u_1, u_2, \dots, u_d)$ и, значит, совпадает с $\text{Quot } A$. \square

Теорема 1. *Все базисы трансцендентности алгебры A (если они существуют) содержат одно и то же число элементов.*

Это число называется *степеню трансцендентности* алгебры A и обозначается через $\text{tr. deg } A$.

Доказательство. Пусть $\{u_1, \dots, u_d\}$ и $\{v_1, \dots, v_e\}$ — два базиса трансцендентности. Если все элементы v_1, \dots, v_e алгебраичны над $K[u_2, \dots, u_d]$, то уже элементы u_2, \dots, u_d составляют базис трансцендентности алгебры A , что невозможно. Следовательно, существует такой номер i_1 , что элемент v_{i_1} трансцендентен над $K[u_2, \dots, u_d]$. Согласно предложению 3, $\{v_{i_1}, u_2, \dots, u_d\}$ — базис трансцендентности алгебры A . Рассуждая таким же образом, мы можем в этом базисе заменить u_2 некоторым элементом v_{i_2} и т. д. В конце концов мы получим базис трансцендентности вида

$$\{v_{i_1}, v_{i_2}, \dots, v_{i_d}\}.$$

Ясно, что числа i_1, i_2, \dots, i_d должны быть различны и что среди них должны быть все числа $1, 2, \dots, e$. Отсюда следует, что $d = e$. \square

Теорема 2 (лемма Нётер о нормализации). *В конечно порожденной алгебре $A = K[u_1, \dots, u_n]$ без делителей нуля существует такой базис трансцендентности $\{v_1, \dots, v_d\}$, что алгебра A цела над $K[v_1, \dots, v_d]$.*

Доказательство. Мы докажем эту теорему в предположении, что поле K бесконечно. В этом случае искомым базис трансцендентности можно составить из линейных комбинаций элементов u_1, \dots, u_n .

Проведем индукцию по n . Если элементы u_1, \dots, u_n алгебраически независимы, то они и составляют искомым базис трансцендентности. В противном случае рассмотрим нетривиальную алгебраическую зависимость между ними:

$$f(u_1, \dots, u_n) = 0, \quad f \in K[x_1, \dots, x_n].$$

Пусть $\deg f = m$. Если f содержит x_n^m с ненулевым коэффициентом, то элемент u_n цел над подалгеброй $B = K[u_1, \dots, u_{n-1}]$. По предположению индукции в B существует такой базис трансцендентности v_1, \dots, v_d , что алгебра B цела над $K[v_1, \dots, v_d]$. Он же будет искомым базисом трансцендентности алгебры A .

Общий случай сводится к рассмотренному с помощью подходящей замены вида

$$x_i = y_i + a_i y_n \quad (i = 1, \dots, n-1), \quad x_n = y_n \quad (a_1, \dots, a_{n-1} \in k).$$

Многочлен

$$g(y_1, \dots, y_{n-1}, y_n) = f(y_1 + a_1 y_n, \dots, y_{n-1} + a_{n-1} y_n, y_n)$$

также имеет степень m и содержит y_n^m с коэффициентом, равным

$$g_0(0, \dots, 0, 1) = f_0(a_1, \dots, a_{n-1}, 1),$$

где f_0 и g_0 — старшие однородные компоненты многочленов f и g соответственно. Так как f_0 — ненулевой однородный многочлен, то он не может быть тождественно равен нулю на гиперплоскости $x_n = 1$. Следовательно, при подходящем выборе a_1, \dots, a_{n-1} многочлен g будет содержать y_n^m с ненулевым коэффициентом. Положим

$$u_i = v_i + a_i v_n \quad (i = 1, \dots, n-1), \quad u_n = v_n;$$

тогда

$$g(v_1, \dots, v_n) = f(u_1, \dots, u_n) = 0,$$

и доказательство сводится к уже рассмотренному случаю. \square

Теорема 3. Если конечно порожденная алгебра A является полем, то это (конечное) алгебраическое расширение поля K .

Доказательство. Согласно теореме 2, существует такой базис трансцендентности $\{v_1, \dots, v_d\}$ алгебры A , что алгебра A цела над подалгеброй

$$B = K[v_1, \dots, v_d].$$

Докажем, что B — также поле. Для любого $u \in B$ существует $u^{-1} \in A$. Элемент u^{-1} цел над B , т. е.

$$u^{-m} + b_1 u^{-m+1} + \dots + b_{m-1} u^{-1} + b_m = 0$$

для некоторых $b_1, \dots, b_m \in B$. Отсюда

$$u^{-1} = -b_1 - b_2 u - \dots - b_m u^{m-1} \in B.$$

Так как алгебра B изоморфна алгебре многочленов от d переменных, то при $d > 0$ она не является полем. Следовательно, $d = 0$, т. е. A — алгебраическое расширение поля K . \square

Следствие. Если конечно порожденная алгебра A над алгебраически замкнутым полем K является полем, то $A = K$.

Теорема 4. Пусть A — конечно порожденная алгебра над алгебраически замкнутым полем K . Тогда для любого не нильпотентного элемента $a \in A$ существует такой гомоморфизм $\varphi: A \rightarrow K$, что $\varphi(a) \neq 0$

Доказательство. Следуя доказательству теоремы 4.3, возьмем какой-нибудь максимальный идеал I' алгебры $A' = A[a^{-1}]$. Поле A'/I' , будучи конечно порожденной алгеброй над K , совпадает с K . В качестве искомого гомоморфизма φ можно взять ограничение на A канонического гомоморфизма $A' \rightarrow A'/I' = K$. \square

Применим эту теорему к исследованию систем алгебраических уравнений.

Пусть $M \subset K^n$ — множество решений системы алгебраических уравнений

$$f_i(x_1, \dots, x_n) = 0 \quad (i = 1, \dots, m). \quad (25)$$

Рассмотрим алгебру

$$A = K[x_1, \dots, x_n]/(f_1, \dots, f_m) \quad (26)$$

и канонический гомоморфизм

$$\pi: K[x_1, \dots, x_n] \rightarrow A \quad (27)$$

Положим $\pi(x_i) = u_i$; тогда

$$A = K[u_1, \dots, u_n]. \quad (28)$$

Каждой точке $x \in K^n$ соответствует гомоморфизм

$$\psi_x: K[x_1, \dots, x_n] \rightarrow K, \quad f \mapsto f(x), \quad (29)$$

и, наоборот, каждый гомоморфизм

$$\psi: K[x_1, \dots, x_n] \rightarrow K$$

имеет вид ψ_x , где x — точка с координатами $(\psi(x_1), \dots, \psi(x_n))$.

Если $x \in M$, то ψ_x переводит идеал (f_1, \dots, f_m) в нуль и поэтому может быть пропущен через гомоморфизм π :

$$\begin{array}{ccc}
 K[x_1, \dots, x_n] & \xrightarrow{\psi_x} & K \\
 \searrow \pi & & \nearrow \varphi_x \\
 & & A
 \end{array} \tag{30}$$

Возникающий при этом гомоморфизм $\varphi_x: A \rightarrow K$ переводит элементы u_1, \dots, u_n , порождающие алгебру A , в координаты точки x . Обратное, каждый гомоморфизм $\varphi: A \rightarrow K$ включается в коммутативную диаграмму (30) и поэтому имеет вид φ_x , где $x \in M$.

Итак, точки множества M взаимно однозначно соответствуют гомоморфизмам алгебры A в K . Это по существу тривиальное соображение перекидывает мост между коммутативной алгеброй и алгебраической геометрией и, в частности, позволяет придать следующую форму теореме 4.

Теорема 5 (теорема Гильберта о нулях). Пусть M — множество решений системы алгебраических уравнений (25) над алгебраически замкнутым полем K , и пусть многочлен $f \in K[x_1, \dots, x_n]$ тождественно обращается в нуль на M . Тогда

$$f^k \in (f_1, \dots, f_m) \tag{31}$$

для некоторого натурального k .

Доказательство. Определим алгебру A , как выше, и положим $a = \pi(f) \in A$. Условие (31) означает, что элемент a нильпотентен. Если это не так, то по теореме 4 существует такой гомоморфизм $\varphi: A \rightarrow K$, что $\varphi(a) \neq 0$. Этот гомоморфизм определяет точку множества M , в которой многочлен f не обращается в нуль. \square

Заметим, что и, обратно, всякий многочлен $f \in K[x_1, \dots, x_n]$, удовлетворяющий условию (31), тождественно обращается в нуль на M .

Следствие. Система алгебраических уравнений (25) над алгебраически замкнутым полем K несовместна тогда и только тогда, когда

$$(f_1, \dots, f_m) \ni 1, \tag{32}$$

т. е. когда существуют такие многочлены $g_1, \dots, g_m \in K[x_1, \dots, x_n]$, что

$$f_1 g_1 + \dots + f_m g_m = 1. \tag{33}$$

Доказательство. Применим теорему Гильберта о нулях к многочлену $f = 1$. \square

Определение 2. *Аффинным алгебраическим многообразием над полем K или алгебраическим многообразием в K^n называется множество решений системы алгебраических уравнений.*

Пусть $M \subset K^n$ — алгебраическое многообразие. Функции на M со значениями в K , являющиеся ограничениями многочленов на пространстве K^n , называются *многочленами на M* . Они образуют алгебру, называемую *алгеброй многочленов на M* и обозначаемую через $K[M]$. Ядром гомоморфизма ограничения

$$\rho: K[x_1, \dots, x_n] \rightarrow K[M]$$

является идеал $I(M)$, состоящий из всех многочленов на K^n , тождественно обращающихся в нуль на M . Имеем

$$K[M] \simeq K[x_1, \dots, x_n]/I(M).$$

По теореме Гильберта о базисе идеала идеал $I(M)$ имеет конечную систему порождающих:

$$I(M) = (f_1, \dots, f_m).$$

Очевидно, что многообразие M может быть задано уравнениями (25). Идеал $I(M)$ называется *идеалом многообразия M* .

Каждая точка $x \in M$ определяет гомоморфизм

$$\varphi_x: K[M] \rightarrow K, \quad f \mapsto f(x). \quad (34)$$

Проведенное выше рассуждение показывает, что тем самым устанавливается взаимно однозначное соответствие между точками многообразия M и гомоморфизмами алгебры $K[M]$ в K . Отметим, что алгебра $K[M]$, будучи алгеброй функций на M , не имеет нильпотентных элементов.

Обратно, пусть $A = K[u_1, \dots, u_n]$ — конечно порожденная алгебра. Рассмотрим гомоморфизм

$$\pi: K[x_1, \dots, x_n] \rightarrow A, \quad x_i \mapsto u_i.$$

Его ядро есть некоторый идеал I алгебры многочленов $K[x_1, \dots, x_n]$. Пусть

$$I = (f_1, \dots, f_m)$$

и $M \subset K^n$ — алгебраическое многообразие, определяемое системой уравнений (25). Тогда точки многообразия M взаимно однозначно

соответствуют гомоморфизмам алгебры A в K . Однако идеал $I(M)$ может быть больше, чем I , и из-за этого алгебра $K[M]$ может не совпадать с алгеброй A .

В любом случае $\text{Ker } \rho \supset \text{Ker } \pi$ и, следовательно, имеется гомоморфизм

$$\sigma : A \rightarrow K[M].$$

Его ядро состоит из тех элементов алгебры A , которые приходят (при гомоморфизме π) из многочленов, тождественно равных нулю на M . Согласно теореме Гильберта о нулях, если поле K алгебраически замкнуто, то

$$\text{Ker } \sigma = \text{rad } A.$$

В частности, если алгебра A не имеет нильпотентных элементов (и поле K алгебраически замкнуто), то $A = K[M]$.

Итак, в случае алгебраически замкнутого поля K мы установили взаимно однозначное соответствие между алгебраическими многообразиями в K^n и алгебрами с n порождающими, не имеющими нильпотентных элементов.

Для любой конечно порожденной алгебры A над полем K множество всех ее гомоморфизмов в K назовем ее *спектром* и обозначим через $\text{Spes } A$. Если в алгебре A выбрана система порождающих из n элементов, то согласно предыдущему $\text{Spes } A$ отождествляется с алгебраическим многообразием в K^n .

В частности, если M — алгебраическое многообразие в K^n , то $\text{Spes } K[M]$ отождествляется с M при выборе в качестве порождающих элементов ограничений координатных функций пространства K^n . При других выборах систем порождающих получаются другие «модели» спектра, которые считаются алгебраическими многообразиями, изоморфными M . Соответственно этому внутренними свойствами многообразия M считаются такие свойства, которые могут быть выражены в терминах алгебры $K[M]$.

При изучении этих свойств удобно пользоваться *топологией Зарисского*. А именно, подмножество $N \subset M$ считается замкнутым в топологии Зарисского, если оно может быть задано уравнениями вида

$$f_i = 0 \quad (i = 1, \dots, m),$$

где $f_1, \dots, f_m \in K[M]$. Например, замкнутые подмножества в K^n — это в точности алгебраические многообразия. Замкнутые подмно-

жества в любом алгебраическом многообразии $M \subset K^n$ — это алгебраические многообразия в K^n , содержащиеся в M .

Нетрудно проверить, что данное определение удовлетворяет аксиомам топологии, т. е. что пересечение любого числа и объединение конечного числа замкнутых подмножеств замкнуты. Например, объединение подмножеств, задаваемых уравнениями $f_i = 0$ ($i = 1, \dots, m$) и $g_j = 0$ ($j = 1, \dots, p$) соответственно, может быть задано уравнениями $f_i g_j = 0$ ($i = 1, \dots, m; j = 1, \dots, p$).

Топология Зарисского, за исключением тривиальных случаев, не является хаусдорфовой. Например, замкнутые подмножества прямой K^1 в топологии Зарисского — это вся прямая и конечные подмножества; поэтому если поле K бесконечно, то любые два непустых открытых подмножества пересекаются.

Ввиду своей бедности топология Зарисского играет в основном вспомогательную роль как удобный язык при изучении алгебраических многообразий. Однако некоторые грубые свойства алгебраических многообразий ею все же улавливаются.

Определение 3. Топологическое пространство называется *нётеровым*, если в нем не существует бесконечной строго убывающей последовательности замкнутых подмножеств.

Каждому замкнутому подмножеству N аффинного алгебраического многообразия M соответствует идеал $I_M(N)$ алгебры $K[M]$, состоящий из всех многочленов, тождественно равных нулю на N ; при этом $N_1 \supset N_2$ тогда и только тогда, когда $I_M(N_1) \subset I_M(N_2)$. Поэтому из нётеровости алгебры $K[M]$ следует, что многообразие M является нётеровым топологическим пространством (в топологии Зарисского).

Определение 4. Топологическое пространство M называется *неприводимым*, если оно непусто и удовлетворяет любому из следующих эквивалентных условий:

- 1) его нельзя представить в виде объединения двух собственных замкнутых подмножеств;
- 2) любые два его непустых открытых подмножества пересекаются.

(Сравните это определение с определением связного топологического пространства.)

Теорема 6. *Аффинное алгебраическое многообразие M неприводимо тогда и только тогда, когда алгебра $K[M]$ не имеет делителей нуля.*

Доказательство. Пусть $f_1, f_2 \in K[M]$ — такие ненулевые многочлены, что $f_1 f_2 = 0$. Тогда $M = N_1 \cup N_2$, где N_i ($i = 1, 2$), — замкнутое подмножество, выделяемое уравнением $f_i = 0$.

Обратно, пусть $M = N_1 \cup N_2$, где N_1, N_2 — собственные замкнутые подмножества. Возьмем какие-нибудь ненулевые многочлены $f_1 \in I_M(N_1)$, $f_2 \in I_M(N_2)$; тогда $f_1 f_2 = 0$. \square

Предложение 4. *Всякое нётерово топологическое пространство M единственным образом представляется в виде*

$$M = \bigcup_{i=0}^s M_i, \quad (35)$$

где M_1, \dots, M_s — неприводимые замкнутые подмножества, ни одно из которых не содержится ни в каком другом.

Подмножества M_i называются *неприводимыми компонентами* пространства M .

Доказательство. Предположим, что существуют нётеровы топологические пространства, которые нельзя представить в виде конечного объединения неприводимых замкнутых подмножеств. Назовем их плохими. Пусть M_0 плохое. Тогда M_0 приводимо, т. е. $M_0 = M_1 \cup N_1$, где M_1 и N_1 — собственные замкнутые подмножества. Ясно, что хотя бы одно из них плохое. Пусть это M_1 . Тогда $M_1 = M_2 \cup N_2$, где M_2 и N_2 — собственные замкнутые подмножества, причем хотя бы одно из них плохое. Продолжая этот процесс, мы получаем бесконечную строго убывающую последовательность замкнутых подмножеств $M_0 \supset M_1 \supset M_2 \supset \dots$, что противоречит нётеровости пространства M_0 .

Таким образом, всякое нётерово топологическое пространство M можно представить в виде (35), где M_1, \dots, M_s — неприводимые замкнутые подмножества. Убрав те из них, которые содержатся в других, мы можем добиться того, чтобы ни одно из них не содержалось в другом. Покажем, что при этом условии разложение (35) единственно.

Пусть $M = \bigcup_{j=1}^t N_j$ — другое такое разложение. Тогда для любого j

$$N_j = \bigcup_{i=1}^s (M_i \cap N_j),$$

откуда в силу неприводимости N_j следует, что существует такое i , что $N_j \subset M_i$. Аналогично, существует такое k , что $M_i \subset N_k$; но тогда $N_j \subset N_k$, откуда следует, что $j = k$ и $N_j = M_i$. Таким образом,

$\{N_1, \dots, N_t\} \subset \{M_1, \dots, M_s\}$. Аналогично доказывается обратное включение. \square

В частности, всякое аффинное алгебраическое многообразие единственным образом разлагается на неприводимые компоненты.

Пример 1. Пусть q — многочлен второй степени от n переменных над алгебраически замкнутым полем K . Уравнение $q = 0$ задает в K^n гиперповерхность M . Возможны следующие случаи:

1) q не разлагается на линейные множители; тогда M — неприводимая квадрака;

2) q разлагается на два непропорциональных линейных множителя; тогда гиперповерхность M есть объединение двух гиперплоскостей, которые и являются ее неприводимыми компонентами;

3) q есть квадрат линейного многочлена; тогда M — гиперплоскость (но в этом случае $I(M) \neq (q)$).

Все это может быть выведено из более общей теоремы 7.5, которая будет доказана в следующем параграфе.

Важнейшей характеристикой неприводимого аффинного алгебраического многообразия является его размерность.

Определение 5. Размерностью неприводимого аффинного алгебраического многообразия M называется число

$$\dim M = \text{tr. deg } K[M].$$

В частности,

$$\dim K^n = \text{tr. deg } K[x_1, \dots, x_n] = n.$$

Размерность алгебраического многообразия обладает следующим свойством, аналогичным свойству размерности векторного пространства.

Теорема 7. Пусть N — неприводимое замкнутое подмножество неприводимого аффинного алгебраического многообразия M . Тогда $\dim N \leq \dim M$, причем равенство имеет место только тогда, когда $N = M$.

Доказательство. Пусть

$$\rho: K[M] \rightarrow K[N]$$

— гомоморфизм ограничения. Ясно, что если элементы $\rho(f_1), \dots, \rho(f_k)$ алгебраически независимы в $K[N]$, то элементы f_1, \dots, f_k алгебраически независимы в $K[M]$. Отсюда следует первое утверждение теоремы.

Предположим теперь, что $N \neq M$. Пусть $\{\rho(f_1), \dots, \rho(f_k)\}$ — базис трансцендентности алгебры $K[N]$ и $f \in I_M(N)$, $f \neq 0$. Докажем, что тогда f_1, \dots, f_k, f алгебраически независимы в $K[M]$, откуда будет следовать второе утверждение теоремы.

Предположим, что f_1, \dots, f_k, f связаны некоторой нетривиальной алгебраической зависимостью. Эту зависимость можно представить в виде

$$a_0(f_1, \dots, f_k)f^m + a_1(f_1, \dots, f_k)f^{m-1} + \dots + a_m(f_1, \dots, f_k) = 0,$$

где a_0, a_1, \dots, a_m — некоторые многочлены, не все равные нулю. Можно считать, что $a_m \neq 0$, — иначе равенство можно сократить на f . Применяя гомоморфизм ρ , получаем тогда, что

$$a_m(\rho(f_1), \dots, \rho(f_k)) = 0,$$

а это противоречит алгебраической независимости $\rho(f_1), \dots, \rho(f_k)$. \square

§ 7. Разложение на простые множители

Одной из основных задач арифметики является разложение на простые множители в кольцах целых алгебраических чисел. Аналогичная проблема для конечно порожденных алгебр возникает в алгебраической геометрии (например, в связи с описанием линейных расслоений над алгебраическими многообразиями). Несмотря на различие между этими двумя типами колец, проблема разложения на простые множители в них может трактоваться до некоторой степени единообразно. Схема такого подхода будет изложена в конце этого параграфа, а в первой части параграфа будут доказаны теоремы о существовании и единственности разложения на простые множители для некоторых типов колец.

Заметим, что если A — целостное кольцо, то для элементов $a, b \in A$ условие $b \mid a$ равносильно тому, что $(a) \subset (b)$, и соответственно условие $a \sim b$ равносильно тому, что $(a) = (b)$.

Теорема 1. *В нётеровом целостном кольце каждый необратимый ненулевой элемент может быть разложен в произведение простых элементов.*

(Имеется в виду, что это произведение может состоять только из одного множителя.)

Доказательство. Предположим, что существуют необратимые ненулевые элементы, которые нельзя разложить на простые множители. Такие элементы будем называть плохими. Пусть a_0 — плохой элемент. Тогда он, в частности, не является простым и, значит, $a_0 = a_1 b_1$, где a_1 и b_1 — необратимые элементы. Ясно, что хотя бы один из элементов a_1 и b_1 плохой. Пусть это a_1 . Тогда $a_1 = a_2 b_2$, где a_2 и b_2 — необратимые элементы, причем хотя бы один из них плохой. Продолжая этот процесс, мы получаем бесконечную строго возрастающую последовательность идеалов

$$(a_0) \subset (a_1) \subset (a_2) \subset \dots,$$

что противоречит нётеровости. □

Что касается единственности разложения на простые множители, то она, конечно, может иметь место лишь с точностью до перестановки множителей и умножения их на обратимые элементы. Поэтому в дальнейшем, говоря о единственности, мы будем понимать ее именно в этом смысле.

Анализируя доказательство единственности разложения на простые множители в евклидовом кольце, данное в § 3.5, мы видим, что оно опирается только на одно свойство кольца: если простой элемент p делит произведение ab , то он делит a или b ; иначе говоря, идеал (p) , порожденный любым простым элементом p , прост. Это приводит нас к следующей теореме.

Теорема 2. *Если в целостном кольце A главный идеал, порожденный любым простым элементом, прост, то любой элемент этого кольца не более чем одним способом может быть разложен в произведение простых элементов.*

Заметим, что главный идеал, порожденный не простым необратимым ненулевым элементом, не может быть простым ни в каком кольце.

Определение 1. Целостное кольцо A называется *факториальным*, если каждый его необратимый ненулевой элемент может быть разложен на простые множители, причем это разложение единственно в указанном выше смысле.

В частности, всякое кольцо главных идеалов факториально (см. § 2).

Очевидно, что в факториальном кольце главный идеал, порожденный любым простым элементом, прост.

В факториальном кольце для любых двух элементов a и b существует *наибольший общий делитель* $\text{НОД}\{a, b\}$ — общий делитель, делящийся на все другие общие делители. А именно, если

$$a = \prod_{i=1}^s p_i^{k_i}, \quad b = \prod_{i=1}^s p_i^{l_i} \quad (k_i, l_i \geq 0),$$

где p_1, \dots, p_s — простые элементы, то

$$\text{НОД}\{a, b\} = \prod_{i=1}^s p_i^{\min\{k_i, l_i\}}.$$

Наибольший общий делитель определен однозначно с точностью до умножения на обратимый элемент.

Элементы a и b факториального кольца называются *взаимно простыми*, если $\text{НОД}\{a, b\} = 1$, т. е. если в разложениях элементов a и b на простые множители не содержится общих (с точностью до ассоциированности) множителей.

Следующая теорема является обобщением (следствия) теоремы 3.6.1.

Теорема 3. *Всякое факториальное кольцо нормально.*

Доказательство этой теоремы ничем не отличается от доказательства теоремы 3.6.1.

Теорема 4. *Кольцо многочленов $A[x]$ над факториальным кольцом A также факториально.*

Доказательство этой теоремы требует некоторой подготовки.

Многочлен $f \in A[x]$ назовем *примитивным*, если его коэффициенты взаимно просты в совокупности.

Пусть K — поле отношений кольца A . Очевидно, что всякий многочлен $h \in K[x]$ можно представить в виде $h = \lambda h_1$, где $\lambda \in K^*$, а $h_1 \in A[x]$ — примитивный многочлен.

Лемма 1 (лемма Гаусса). *Если многочлен $f \in A[x]$ разлагается в произведение двух многочленов в кольце $K[x]$, то он разлагается в произведение двух пропорциональных им многочленов в кольце $A[x]$.*

Эта лемма доказывается так же, как теорема 3.6.2. Небольшое отличие состоит в том, что факторкольцо кольца A по идеалу, порожденному простым элементом, в общем случае может не быть полем, но оно во всяком случае не имеет делителей нуля, а этого достаточно для доказательства.

Следствие. Если многочлен $f \in A[x]$ может быть разложен в произведение двух многочленов меньшей степени в кольце $K[x]$, то он может быть разложен в произведение многочленов меньшей степени и в кольце $A[x]$.

Доказательство теоремы 4. Следствие леммы Гаусса и очевидные соображения показывают, что простыми элементами кольца $A[x]$ могут быть только элементы следующих двух типов:

- 1) простые элементы кольца A ;
- 2) примитивные многочлены $h \in A[x]$, неприводимые над полем K .

С другой стороны, ясно, что все эти элементы действительно являются простыми и что любой необратимый ненулевой элемент кольца $A[x]$ может быть разложен в произведение таких элементов. Если имеется два таких разложения многочлена $f \in A[x]$, то, рассматривая их как разложения в кольце $K[x]$, факториальность которого известна, мы заключаем, что множители второго типа в этих разложениях ассоциированы в $K[x]$; но так как они являются примитивными многочленами, то они ассоциированы и в $A[x]$. После сокращения на эти множители мы получаем два разложения на простые множители элемента кольца A и можем воспользоваться факториальностью этого кольца. \square

Рассуждая по индукции, из доказанной теоремы можно вывести

Следствие. Кольцо многочленов $K[x_1, \dots, x_n]$ от n переменных над полем K факториально при любом n .

Простые элементы кольца $K[x_1, \dots, x_n]$ называются *неприводимыми многочленами*.

Очевидно, что всякий многочлен первой степени неприводим.

Лемма 2. Если многочлен $f \in K[x_1, \dots, x_n]$ над бесконечным полем K обращается в нуль во всех точках гиперплоскости

$$l \doteq a_1 x_1 + \dots + a_n x_n + b = 0,$$

то он делится на l .

Доказательство. Перейдя к другой аффинной системе координат, мы можем считать, что $l = x_1$. Тогда условие леммы означает, что все члены многочлена f содержат x_1 и, значит, f делится на l . \square

Приведем два примера разложения многочленов на линейные множители с использованием этой леммы и факториальности кольца многочленов.

Пример 1. Вычислим другим способом определитель Вандермонда $V(x_1, \dots, x_n)$ (см. пример 2.4.5). Очевидно, что $V(x_1, \dots, x_n)$ — многочлен от x_1, \dots, x_n . При $x_i = x_j$ (i, j различны) он обращается в нуль, так как соответствующая матрица в этом случае имеет две одинаковые строки. По лемме 2 мы можем заключить, что многочлен $V(x_1, \dots, x_n)$ делится на $x_i - x_j$ в кольце $K[x_1, \dots, x_n]$ при любых различных i, j ; но тогда из факториальности этого кольца следует, что $V(x_1, \dots, x_n)$ делится на $\prod_{i>j} (x_i - x_j)$. Легко видеть, что $V(x_1, \dots, x_n)$ — однородный многочлен степени $\frac{n(n-1)}{2}$. Поэтому

$$V(x_1, \dots, x_n) = c \prod_{i>j} (x_i - x_j) \quad (c \in K).$$

Сравнивая коэффициенты при $x_2 x_3^2 \dots x_n^{n-1}$, получаем, что $c = 1$.

Задача 1. Доказать аналогичным способом, что

$$x^3 + y^3 + z^3 - 3xyz = (x + y + z)(x + \omega y + \bar{\omega}z)(x + \bar{\omega}y + \omega z),$$

где $\omega = -\frac{1}{2} + i\frac{\sqrt{3}}{2}$.

Применим результат о факториальности кольца многочленов к описанию $(n-1)$ -мерных алгебраических многообразий в K^n .

Пусть K — алгебраически замкнутое поле. Для всякого многочлена $f \in K[x_1, \dots, x_n]$ обозначим через $M(f)$ алгебраическое многообразие в K^n , задаваемое уравнением $f = 0$.

Теорема 5. *Отображение $p \mapsto M(p)$ устанавливает взаимно однозначное соответствие между неприводимыми многочленами от n переменных, рассматриваемыми с точностью до ассоциированности, и $(n-1)$ -мерными неприводимыми алгебраическими многообразиями в K^n ; при этом идеал многообразия $M(p)$ порождается многочленом p .*

Доказательство. 1) Пусть $p \in K[x_1, \dots, x_n]$ — неприводимый многочлен. Тогда идеал (p) прост и, следовательно, многообразие $M(p)$ неприводимо и $I(M(p)) = (p)$. В частности, многочлен p с точностью до ассоциированности однозначно восстанавливается по многообразию $M(p)$.

2) В предыдущих обозначениях, имеем

$$K[M(p)] = K[x_1, \dots, x_n]/(p) = K[u_1, \dots, u_n],$$

где u_1, \dots, u_n — ограничения координатных функций x_1, \dots, x_n пространства K^n на $M(p)$. Предположим, что многочлен p содержит

нетривиальным образом переменную x_n . Тогда и любой многочлен из идеала (p) содержит x_n . Это означает, что u_1, \dots, u_{n-1} алгебраически независимы. Следовательно, $\dim M(p) = n - 1$.

3) Обратно, пусть $M \subset K^n$ — $(n - 1)$ -мерное неприводимое алгебраическое многообразие. Возьмем любой ненулевой многочлен $f \in I(M)$ и разложим его на неприводимые множители. Из простоты идеала $I(M)$ следует, что хотя бы один из этих множителей принадлежит $I(M)$. Пусть это будет неприводимый многочлен p . Тогда $M \subset M(p)$, но из совпадения размерностей следует, что на самом деле $M = M(p)$. \square

Пусть $f \in K[x_1, \dots, x_n]$ — любой необратимый ненулевой многочлен. Разложим его на неприводимые множители:

$$f = p_1^{k_1} \dots p_s^{k_s}.$$

Из теоремы 5 очевидным образом следует, что

$$M(f) = M(p_1) \cup \dots \cup M(p_s)$$

есть разложение многообразия $M(f)$ на неприводимые компоненты.

Задача 2. Найти $I(M(f))$.

Такие же результаты получатся, если вместо пространства K^n взять любое неприводимое аффинное алгебраическое многообразие M , для которого алгебра $K[M]$ факториальна. (Единственным местом, где в общем случае требуются дополнительные соображения, является п. 2) доказательства теоремы.)

Если же алгебра $K[M]$ не факториальна, то в ней существуют простые элементы, которые порождают главные идеалы, не являющиеся простыми, и одновременно в M существуют $(n - 1)$ -мерные неприводимые подмногообразия, идеалы которых не являются главными.

Пример 2. Пусть $Q \subset K^3$ — квадратичный конус, задаваемый уравнением $xu = z^2$. Имеем

$$K[Q] = K[x, y, z]/(xu - z^2) = K[u, v, w],$$

где u, v, w связаны соотношением $uv = w^2$. Очевидно, что u, v, w (как и все линейные формы от них) — простые элементы алгебры $K[Q]$. Поэтому соотношение $uv = w^2$ является нарушением факториальности. Это находит свое отражение в том, что идеалы (u) , (v) , (w) не являются простыми (например, $uv \in (w)$, но $u \notin (w)$)

и $v \notin (w)$, и одновременно в том, что идеалы образующих конуса Q не являются главными (например, идеал оси x есть (v, w) , а идеал оси y есть (u, w)).

Наблюдения подобного рода приводят к мысли о том, что, может быть, правильнее рассматривать не простые элементы алгебры $K[M]$, а простые идеалы, соответствующие $(n - 1)$ -мерным неприводимым подмногообразиям. И, действительно, на этом пути может быть построена очень красивая теория, причем не только для конечных порожденных алгебр, но и для любых нётеровых колец. Изложим кратко основные идеи этой теории.

Пусть A — нормальное нётерово кольцо. Назовем *нормированием* кольца A сюръективное отображение

$$v: A \setminus \{0\} \rightarrow \mathbb{Z}_+,$$

обладающее следующими свойствами:

- 1) $v(ab) = v(a) + v(b)$;
- 2) $v(a + b) \geq \min\{v(a), v(b)\}$.

Совокупность элементов $a \in A$, для которых $v(a) > 0$, есть простой идеал кольца A . Он называется *идеалом нормирования* v и обозначается через $\mathfrak{p}(v)$.

С каждым простым элементом $p \in A$, для которого идеал (p) прост, связано нормирование v_p , определяемое следующим образом: $v_p(a)$ есть наибольшая степень элемента p , на которую делится a . Очевидно, что $\mathfrak{p}(v_p) = (p)$. Если кольцо A факториально, то для любых его ненулевых элементов верно следующее:

$$b \mid a \Leftrightarrow v_p(a) \geq v_p(b) \quad \forall p.$$

В общем же случае нормирований вида v_p недостаточно для выяснения вопроса о делимости элементов. Каким должно быть их разумное обобщение, подсказывают следующие две задачи.

Задача 3. Доказать, что всякий ненулевой простой главный идеал является минимальным среди ненулевых простых идеалов кольца A .

Задача 4. Доказать, что в факториальном кольце любой минимальный простой идеал является главным.

Минимальные простые идеалы кольца A называются его *простыми дивизорами*. В рассматривавшемся выше случае, когда $A = K[M]$, простые дивизоры — это идеалы $(n - 1)$ -мерных неприводимых подмногообразий многообразия M .

Можно доказать, что любой простой дивизор \mathfrak{p} является идеалом некоторого однозначно определенного нормирования $v_{\mathfrak{p}}$. Идея доказательства состоит в том, что идеал \mathfrak{p} становится главным при вложении кольца A

в подходящее кольцо вида $A[u^{-1}]$, где $u \in A \setminus \mathfrak{p}$. Очевидно, что если $\mathfrak{p} = (p)$, то $v_{\mathfrak{p}} = v_p$.

В случае когда $A = K[M]$ и $\mathfrak{p} = I(N)$, где N есть $(n-1)$ -мерное неприводимое подмногообразие многообразия M , число $v_{\mathfrak{p}}(f)$ при $f \in K[M]$ имеет смысл «порядка нуля» функции f на подмногообразии N .

Пример 3. В примере 2 плоскость $x = 0$ касается конуса Q по оси y , плоскость $y = 0$ касается него по оси x , а плоскость $z = 0$ трансверсально пересекает его по осям x и y . Поэтому, если обозначить через \mathfrak{p} и \mathfrak{q} идеалы осей x и y в алгебре $K[Q]$, то

$$\begin{aligned} v_{\mathfrak{p}}(u) &= 0, & v_{\mathfrak{p}}(v) &= 2, & v_{\mathfrak{p}}(w) &= 1, \\ v_{\mathfrak{q}}(u) &= 2, & v_{\mathfrak{q}}(v) &= 0, & v_{\mathfrak{q}}(w) &= 1, \end{aligned}$$

что согласуется с соотношением $uv = w^2$.

Основные свойства нормирований $v_{\mathfrak{p}}$, оправдывающие их рассмотрение, состоят в следующем:

- 1) для любого $a \in A \setminus \{0\}$ множество таких \mathfrak{p} , что $v_{\mathfrak{p}}(a) > 0$, конечно;
- 2) для любых $a, b \in A \setminus \{0\}$

$$b \mid a \iff v_{\mathfrak{p}}(a) \geq v_{\mathfrak{p}}(b) \quad \forall \mathfrak{p}.$$

Исторически эта теория была впервые построена для колец целых чисел круговых полей в работах Куммера по теореме Ферма. В отличие от общего случая, в кольцах целых алгебраических чисел все нетривиальные простые идеалы, как мы покажем ниже, являются минимальными. Кроме того, в этом случае теория может быть интерпретирована как теорема об однозначном разложении идеалов на простые множители.

А именно, определим умножение идеалов по правилу

$$a\mathfrak{b} = \left\{ \sum_{i=1}^k a_i b_i : a_1, \dots, a_k \in a, b_1, \dots, b_k \in \mathfrak{b} \right\}.$$

Очевидно, что это умножение коммутативно и ассоциативно и что $(a)(\mathfrak{b}) = (a\mathfrak{b})$. Тем самым полугруппа ненулевых элементов кольца A , рассматриваемых с точностью до ассоциированности, оказывается вложенной в полугруппу идеалов.

Можно доказать, что если A — кольцо целых чисел некоторого поля алгебраических чисел, то всякий ненулевой идеал в нем однозначно разлагается в произведение простых идеалов. Число $v_{\mathfrak{p}}(a)$ интерпретируется тогда как показатель при \mathfrak{p} в разложении идеала (a) .

Два идеала называются *эквивалентными*, если они становятся равными после умножения их на подходящие главные идеалы. Классы эквивалентных идеалов кольца A целых алгебраических чисел образуют группу, называемую *группой классов идеалов* кольца A и обозначаемую через $Cl A$. Она измеряет отклонение кольца A от факториальности.

Пусть K — поле алгебраических чисел, а \mathbb{Z}_K — кольцо его целых чисел.

Теорема 6. *Любой ненулевой идеал \mathfrak{a} кольца \mathbb{Z}_K является подгруппой конечного индекса (по сложению).*

Доказательство. Мы знаем, что существует такой базис $\{e_1, \dots, e_n\}$ пространства K над \mathbb{Q} , что

$$\mathbb{Z}_K = \mathbb{Z}e_1 \oplus \dots \oplus \mathbb{Z}e_n.$$

Пусть a — какой-либо ненулевой элемент идеала \mathfrak{a} . Отображение $x \mapsto ax$ является невырожденным линейным преобразованием пространства K над \mathbb{Q} и, следовательно, $\{ae_1, \dots, ae_n\}$ — также базис этого пространства. Так как он содержится в \mathfrak{a} , то \mathfrak{a} — подгруппа конечного индекса в \mathbb{Z}_K . \square

Следствие. *Любой нетривиальный простой идеал \mathfrak{p} кольца \mathbb{Z}_K является максимальным идеалом.*

Доказательство. Факторкольцо $\mathbb{Z}_K/\mathfrak{p}$ конечно и не имеет делителей нуля. Доказываемое утверждение вытекает из следующей леммы. \square

Лемма 3. *Всякое конечное кольцо A без делителей нуля является полем.*

Доказательство. Пусть $a \in A$ — ненулевой элемент. Отображение

$$A \rightarrow A, \quad x \mapsto ax,$$

ввиду отсутствия в кольце A делителей нуля инъективно и, значит, сюръективно. В частности, существует такой элемент b , что $ab = 1$. \square

Следствие. *Любой нетривиальный простой идеал кольца \mathbb{Z}_K является минимальным простым идеалом.*

Доказательство. Если бы существовал меньший простой идеал, то он не был бы максимальным. \square

Пример 4. Кольцо целых чисел поля $\mathbb{Q}(\sqrt{-5})$ есть $\mathbb{Z}[\sqrt{-5}]$. Определим норму $N(c)$ числа $c = a + b\sqrt{-5} \in \mathbb{Z}[\sqrt{-5}]$ ($a, b \in \mathbb{Z}$) по формуле

$$N(c) = c\bar{c} = a^2 + 5b^2 \in \mathbb{Z}.$$

Очевидно, что норма мультипликативна:

$$N(c_1c_2) = N(c_1)N(c_2).$$

Поэтому, если c — обратимый элемент кольца $\mathbb{Z}[\sqrt{-5}]$, то $N(c) = \pm 1$. Отсюда следует, что обратимы только ± 1 . Если c — не простой необратимый ненулевой элемент, то $N(c)$ представляется в виде произведения двух норм, отличных от 1. С помощью этого соображения легко показать, что все элементы, участвующие в равенстве

$$2 \cdot 3 = (1 + \sqrt{-5})(1 - \sqrt{-5}), \quad (36)$$

просты. Таким образом, кольцо $\mathbb{Z}[\sqrt{-5}]$ не факториально.

С точки зрения теории идеалов равенство (36) находит следующее объяснение:

$$(2) = \mathfrak{p}^2, \quad (3) = \mathfrak{q}_1 \mathfrak{q}_2, \quad (1 + \sqrt{-5}) = \mathfrak{p} \mathfrak{q}_1, \quad (1 - \sqrt{-5}) = \mathfrak{p} \mathfrak{q}_2,$$

где

$$\mathfrak{p} = (2, 1 + \sqrt{-5}) = (2, 1 - \sqrt{-5}), \quad \mathfrak{q}_1 = (3, 1 + \sqrt{-5}), \quad \mathfrak{q}_2 = (3, 1 - \sqrt{-5})$$

— простые идеалы. (Проверьте это!) Можно показать, что

$$\text{Cl } \mathbb{Z}[\sqrt{-5}] \simeq \mathbb{Z}_2.$$

Вообще, группа классов идеалов кольца целых любого поля алгебраических чисел конечна. Это означает, в частности, что некоторая степень любого идеала является главным идеалом.

Группы

§ 1. Прямые и полупрямые произведения

В § 9.1 были рассмотрены прямые суммы аддитивных абелевых групп. Конечно, название операции в группе неважно. Ничто не мешает сделать то же для мультипликативных групп; только в этом случае естественно говорить не о прямой сумме, а о прямом произведении. Более существенно то, что можно отказаться от коммутативности. Дадим соответствующие точные определения.

Определение 1. Говорят, что группа G разлагается в *прямое произведение* своих подгрупп G_1, \dots, G_k , если

1) каждый элемент $g \in G$ единственным образом представляется в виде $g = g_1 \dots g_k$, где $g_i \in G_i$;

2) $g_i g_j = g_j g_i$ при $g_i \in G_i, g_j \in G_j, i \neq j$.

В этом случае пишут $G = G_1 \times \dots \times G_k$. Очевидно, что если группа G конечна, то

$$|G| = |G_1| \dots |G_k|.$$

Из условия 1) следует, что $G_i \cap G_j = \{e\}$ при $i \neq j$, но, как мы видели уже на примере векторных пространств (ср. с задачей 5.1.2), при $k > 2$ последнее условие является более слабым, чем условие 1).

Из условия 2) вытекает следующее правило умножения элементов группы G :

$$(g_1 \dots g_k)(g'_1 \dots g'_k) = (g_1 g'_1) \dots (g_k g'_k) \quad (g_i, g'_i \in G_i). \quad (1)$$

В частности, легко видеть, что каждая из подгрупп G_i нормальна. Как вытекает из следующей леммы, условие 2) можно заменить требованием, чтобы подгруппы G_1, \dots, G_k были нормальны.

Лемма 1. Пусть G_1 и G_2 — нормальные подгруппы группы G , причём $G_1 \cap G_2 = \{e\}$. Тогда $g_1 g_2 = g_2 g_1$ для любых $g_1 \in G_1, g_2 \in G_2$.

Доказательство. Имеем

$$g_1 g_2 g_1^{-1} g_2^{-1} = g_1 (g_2 g_1^{-1} g_2^{-1}) = (g_1 g_2 g_1^{-1}) g_2^{-1} \in G_1 \cap G_2 = \{e\},$$

откуда $g_1 g_2 = g_2 g_1$.

□

Рассмотрим отдельно случай двух множителей.

Предложение 1. *Группа G разлагается в прямое произведение своих подгрупп G_1 и G_2 тогда и только тогда, когда*

1) *подгруппы G_1 и G_2 нормальны;*

2) $G_1 \cap G_2 = \{e\}$;

3) $G = G_1 G_2$, т. е. каждый элемент $g \in G$ представляется в виде $g = g_1 g_2$, где $g_1 \in G_1$, $g_2 \in G_2$.

Доказательство. Утверждение «только тогда» уже доказано выше. Пусть, обратно, выполнены условия 1)–3) предложения. Тогда по лемме 1 $g_1 g_2 = g_2 g_1$ при $g_1 \in G_1$, $g_2 \in G_2$. Остается проверить единственность представления элемента $g \in G$ в виде $g = g_1 g_2$, где $g_1 \in G_1$, $g_2 \in G_2$. Пусть

$$g_1 g_2 = g'_1 g'_2 \quad (g_1, g'_1 \in G_1, g_2, g'_2 \in G_2).$$

Тогда

$$g_1^{-1} g'_1 = g_2 g'_2^{-1} \in G_1 \cap G_2 = \{e\},$$

откуда

$$g_1 = g'_1, \quad g_2 = g'_2. \quad \square$$

Пример 1. Пусть $G = \{e, a, b, c\}$ — нециклическая группа порядка 4. Легко видеть, что квадрат любого из элементов a, b, c равен единице, а произведение любых двух из них (в любом порядке) равно третьему (см. пример 4.1.7, где выписана таблица умножения этой группы). Отсюда следует, что G есть прямое произведение любых двух различных циклических подгрупп второго порядка, например,

$$G = \{e, a\} \times \{e, b\}.$$

Пример 2. Возможность и единственность представления комплексного числа, отличного от нуля, в тригонометрической форме означает, что

$$\mathbb{C}^* = \mathbb{R}_+^* \times \mathbb{T}$$

(см. обозначения в примерах 4.5.2 и 4.5.3).

Пример 3. Пусть $G = \text{GL}_n^+(\mathbb{R})$ — группа матриц с положительным определителем, G_1 — подгруппа скалярных матриц λE с $\lambda > 0$ и $G_2 = \text{SL}_n(\mathbb{R})$. Тогда $G = G_1 \times G_2$. В самом деле, G_1 и G_2 — нормальные подгруппы (причем элементы подгруппы G_1 коммутируют со всеми

элементами группы), $G_1 \cap G_2 = \{e\}$ и $G = G_1 G_2$, так как каждая матрица $A \in G$ может быть представлена в виде

$$A = \lambda A_1 = (\lambda E) A_1,$$

где

$$\lambda = \sqrt[n]{\det A}, \quad A_1 = \frac{1}{\lambda} A \in \text{SL}_n(\mathbb{R}) = G_2.$$

Задача 1. Выяснить, при каких n

$$\text{GL}_n(\mathbb{R}) = \{\lambda E : \lambda \in \mathbb{R}^*\} \times \text{SL}_n(\mathbb{R}).$$

Дадим теперь определение внешнего прямого произведения групп.

Определение 2. *Прямым произведением* групп G_1, \dots, G_k называется совокупность последовательностей (g_1, \dots, g_k) , где $g_i \in G_i$, с n -компонентной операцией умножения:

$$(g_1, \dots, g_k)(g'_1, \dots, g'_k) = (g_1 g'_1, \dots, g_k g'_k).$$

Очевидно, что таким образом действительно получается группа. Она обозначается через $G_1 \times \dots \times G_k$.

Отождествляя каждый элемент $g \in G_i$ с последовательностью $(e, \dots, g, \dots, e) \in G_1 \times \dots \times G_k$, где g стоит на i -м месте, мы получаем вложение G_i в $G_1 \times \dots \times G_k$ в виде подгруппы. Группа $G_1 \times \dots \times G_k$ есть прямое произведение этих подгрупп в смысле определения 1.

Обратно, если некоторая группа G разлагается в прямое произведение своих подгрупп G_1, \dots, G_k , то отображение

$$G_1 \times \dots \times G_k \rightarrow G, \quad (g_1, \dots, g_k) \mapsto g_1 \dots g_k,$$

в силу формулы (1) является изоморфизмом групп.

Пример 4. Группа (невырожденных) диагональных матриц порядка n изоморфна группе

$$(K^*)^n = \underbrace{K^* \times \dots \times K^*}_n.$$

Гораздо чаще, чем разложение группы в прямое произведение, встречается разложение в так называемое полупрямое произведение. Перед тем как дать соответствующие определения, поговорим об автоморфизмах групп.

Определение 3. *Аutomорфизмом* группы называется ее изоморфизм на себя.

Пример 5. Отображение $x \mapsto ax$ ($a \neq 0$) является автоморфизмом аддитивной группы поля.

Пример 6. Отображение $X \mapsto (X^T)^{-1}$ является автоморфизмом группы невырожденных матриц.

Все автоморфизмы группы G образуют группу, обозначаемую через $\text{Aut } G$.

Для любого элемента $g \in G$ отображение $a(g): x \mapsto gxg^{-1}$ ($x \in G$) является автоморфизмом:

$$a(g)(xy) = gxyg^{-1} = (gxyg^{-1})(gyg^{-1}) = (a(g)x)(a(g)y).$$

Такой автоморфизм называется *внутренним автоморфизмом*, определяемым элементом g .

Отображение $g \mapsto a(g)$ является гомоморфизмом группы G в группу $\text{Aut } G$:

$$a(gh)x = ghx(gh)^{-1} = g(hxh^{-1})g^{-1} = a(g)a(h)x.$$

Его ядро есть *центр* Z группы G :

$$Z = \{z \in G : zg = gz \ \forall g \in G\}.$$

Его образ есть подгруппа группы $\text{Aut } G$, называемая *группой внутренних автоморфизмов* группы G и обозначаемая через $\text{Int } G$. По теореме о гомоморфизме

$$\text{Int } G \simeq G/Z.$$

Пример 7. Легко доказать, что при $n \geq 3$ центр группы S_n тривиален. Следовательно,

$$\text{Int } S_n \simeq S_n.$$

Пример 8. Центр группы $\text{GL}_n(K)$ (где K — поле) состоит из скалярных матриц и изоморфен мультипликативной группе K^* . Факторгруппа $\text{GL}_n(K)/\{\lambda E : \lambda \in K^*\}$ есть не что иное, как проективная группа $\text{PGL}_n(K)$ (группа проективных преобразований $(n-1)$ -мерного проективного пространства PK^n , ассоциированного с векторным пространством K^n). Таким образом,

$$\text{Int } \text{GL}_n(K) \simeq \text{PGL}_n(K).$$

Пусть $\varphi \in \text{Aut } G$ — любой автоморфизм и $g \in G$. Непосредственная проверка показывает, что

$$\varphi a(g) \varphi^{-1} = a(\varphi(g)).$$

Следовательно, $\text{Int } G$ — нормальная подгруппа группы $\text{Aut } G$.

Конечно, нетривиальные внутренние автоморфизмы имеются лишь у неабелевых групп.

Пример 9. Найдем группу $\text{Aut } S_3$. Так как при любом изоморфизме групп сохраняются порядки элементов, то всякий автоморфизм φ группы S_3 переводит транспозиции в транспозиции. Более того, так как группа S_3 порождается транспозициями, то автоморфизм φ однозначно определяется тем, как он переставляет транспозиции. Следовательно,

$$|\text{Aut } S_3| \leq |S_3| = 6.$$

Но группа $\text{Int } S_3$, как мы видели выше, содержит как раз 6 элементов. Следовательно,

$$\text{Aut } S_3 = \text{Int } S_3.$$

Пример 10. Найдем группу $\text{Aut } \mathbb{Z}_n$. Пусть $\varphi \in \text{Aut } \mathbb{Z}_n$ и $\varphi([1]) = [k]$. Тогда

$$\varphi([l]) = \varphi(\underbrace{[1] + \dots + [1]}_l) = \underbrace{[k] + \dots + [k]}_l = [kl] = [k][l],$$

где умножение в последнем выражении понимается в смысле кольца \mathbb{Z}_n . Таким образом, всякий автоморфизм группы \mathbb{Z}_n имеет вид

$$\varphi_a: x \mapsto ax$$

для некоторого $a \in \mathbb{Z}_n$. Обратно, для любого $a \in \mathbb{Z}_n$ отображение φ_a является гомоморфизмом группы \mathbb{Z}_n в себя и

$$\varphi_a \varphi_b = \varphi_{ab}.$$

Следовательно, гомоморфизм φ_a обратим, т. е. является автоморфизмом, тогда и только тогда, когда обратим элемент a в кольце \mathbb{Z}_n . Таким образом,

$$\text{Aut } \mathbb{Z}_n \simeq \mathbb{Z}_n^*,$$

где \mathbb{Z}_n^* — группа обратимых элементов кольца \mathbb{Z}_n .

Пользуясь понятием внутреннего автоморфизма, можно следующим образом переформулировать определение нормальной подгруппы: подгруппа нормальна, если она инвариантна относительно всех внутренних автоморфизмов группы G .

Пусть N — нормальная подгруппа группы G , а H — любая подгруппа. Тогда произведение

$$NH \doteq \{nh: n \in N, h \in H\}$$

является подгруппой, как показывают следующие тождества:

$$\begin{aligned}(n_1 h_1)(n_2 h_2) &= n_1 (h_1 n_2 h_1^{-1}) h_1 h_2, \\ (nh)^{-1} &= (h^{-1} n^{-1} h) h^{-1},\end{aligned}\tag{2}$$

Кроме того, $NH = HN$.

Определение 4. Говорят, что группа G разлагается в *полупрямое произведение* подгрупп N и H , если

- 1) N — нормальная подгруппа;
- 2) $N \cap H = \{e\}$;
- 3) $NH = G$.

При этом пишут $G = N \rtimes H$ (или $G = H \ltimes N$).

Свойства 2) и 3) эквивалентны тому, что каждый элемент группы G единственным образом представляется в виде произведения nh , где $n \in N$, $h \in H$. В частности, если группа G конечна, то

$$|G| = |N||H|.$$

Пример 11. $S_n = A_n \rtimes \langle (12) \rangle$.

Пример 12. $S_4 = V_4 \rtimes S_3$, где V_4 — четверная группа Клейна (см. пример 4.6.15), а S_3 вложена в S_4 в виде подгруппы, оставляющей на месте символ 4. В самом деле, легко видеть, что для каждого $k \in \{1, 2, 3, 4\}$ в V_4 имеется единственная подстановка, переводящая 4 в k . Отсюда следует, что каждая подстановка $\sigma \in S_4$ единственным образом представляется в виде $\sigma = \tau\rho$, где $\tau \in V_4$, $\rho \in S_3$.

Пример 13. $GL_n(K) = SL_n(K) \rtimes \{\text{diag}(\lambda, 1, \dots, 1) : \lambda \in K^*\}$.

Пример 14. Группа $GA(S)$ аффинных преобразований аффинного пространства S есть полупрямое произведение (нормальной) подгруппы $\text{Trans}(S)$ параллельных переносов и группы $GL(V)$ линейных преобразований ассоциированного векторного пространства V , вложенной в $GA(S)$ в виде подгруппы, оставляющей на месте какую-либо фиксированную точку.

Пример 15. Группа $\text{Isom } S$ движений евклидова аффинного пространства S есть полупрямое произведение группы параллельных переносов и группы $O(V)$ ортогональных преобразований ассоциированного евклидова векторного пространства.

Если $G = N \rtimes H$, то $G/N \simeq H$. Однако не следует думать, что для всякой нормальной подгруппы N найдется такая подгруппа H (изоморфная G/N), что $G = N \rtimes H$. Например, для (нормальной) подгруппы $2\mathbb{Z}$ в группе \mathbb{Z} не существует дополнительной подгруппы.

Пусть $G = N \rtimes H$. Для каждого $h \in H$ обозначим через $\alpha(h)$ ограничение на N внутреннего автоморфизма $a(h)$ группы G . Очевидно, что $\alpha(h) \in \text{Aut } N$ и что отображение $h \mapsto \alpha(h)$ является гомоморфизмом группы H в группу $\text{Aut } N$. Первую из формул (2) можно переписать в виде

$$(n_1 h_1)(n_2 h_2) = (n_1 \alpha(h_1) n_2)(h_1 h_2). \quad (3)$$

Пусть теперь имеются какие-то группы N и H и задан гомоморфизм

$$\alpha: H \rightarrow \text{Aut } N.$$

Определим в декартовом произведении $N \times H$ операцию умножения по формуле

$$(n_1, h_1)(n_2, h_2) = (n_1 \alpha(h_1) n_2, h_1 h_2). \quad (4)$$

подсказанной формулой (3). Непосредственно проверяется, что операция (4) удовлетворяет аксиомам группы. Полученная группа G называется (*внешним*) *полупрямым произведением* групп N и H , определяемым гомоморфизмом α , и обозначается $N \rtimes_{\alpha} H$ или просто $N \rtimes H$. Если отождествить группу N с подгруппой группы G , состоящей из пар вида (n, e) , а группу H — с подгруппой, состоящей из пар вида (e, h) , то группа G будет полупрямым произведением этих подгрупп в смысле определения 4.

Обратно, если какая-то группа G разлагается в полупрямое произведение своих подгрупп N и H и $\alpha: H \rightarrow \text{Aut } N$ — гомоморфизм, определенный, как было указано выше, то отображение

$$NH \rightarrow G, \quad (n, h) \mapsto nh,$$

есть изоморфизм групп.

Прямое произведение является частным случаем полупрямого: оно получается, если в качестве α взять тривиальный гомоморфизм.

Условимся обозначать через $\langle a \rangle_n$ циклическую группу порядка n с порождающим элементом a .

Пример 16. Опишем группы, являющиеся полупрямыми произведениями циклических групп $\langle a \rangle_n$ и $\langle b \rangle_m$ порядков n и m соответственно. Гомоморфизм

$$\alpha: \langle b \rangle_m \rightarrow \text{Aut} \langle a \rangle_n \simeq \mathbb{Z}_n^*$$

определяется образом элемента b , который представляет собой возведение в некоторую степень k в группе $\langle a \rangle_n$. (См. пример 10.) Чис-

ло k (его задание существенно лишь по модулю n) должно удовлетворять условию

$$k^m \equiv 1 \pmod{n}.$$

В частности, если число $|\mathbb{Z}_n^*| = \varphi(n)$ взаимно просто с m , то отсюда следует, что $k = 1$, а это соответствует прямому произведению. Например, всякое полупрямое произведение групп $\langle a \rangle_7$ и $\langle b \rangle_5$ является прямым произведением. Полупрямое произведение групп $\langle a \rangle_n$ и $\langle b \rangle_m$, отвечающее числу k , обозначим через $\langle a \rangle_n \lambda_k \langle b \rangle_m$. Оно определяется соотношением

$$bab^{-1} = a^k.$$

Например, $\langle a \rangle_n \lambda_{-1} \langle b \rangle_2$ есть группа диэдра D_n . Некоторые из полученных таким образом полупрямых произведений могут оказаться изоморфными. А именно, при $(s, m) = 1$ мы можем заменить элемент b элементом b^s , который также порождает группу $\langle b \rangle$; при этом k заменится на k^s . Это показывает, что существенно не само число k , а циклическая подгруппа, порожденная элементом $[k]$ в \mathbb{Z}_n^* . Например, имеются всего две неизоморфные группы, разлагающиеся в полупрямое произведение групп $\langle a \rangle_{11}$ и $\langle b \rangle_5$ (одна из них есть прямое произведение этих групп).

§ 2. Коммутант

Пусть G — какая-либо группа. *Коммутатором* элементов $x, y \in G$ называется элемент

$$(x, y) = xyx^{-1}y^{-1}.$$

Очевидны свойства:

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

Подгруппа, порожденная всеми коммутаторами, называется *коммутантом* группы G и обозначается (G, G) или G' . Ввиду свойства 2) для образования коммутанта достаточно брать произведения коммутаторов. Коммутант тривиален тогда и только тогда, когда группа G абелева.

Очевидно, что если $\varphi: G \rightarrow H$ — какой-либо гомоморфизм групп, то $\varphi(G') \subset H'$, а если $\varphi(G) = H$, то $\varphi(G') = H'$. В частности, коммутант инвариантен относительно всех внутренних автоморфизмов группы, т. е. является нормальной подгруппой.

Теорема 1. Коммутант G' группы G является наименьшей нормальной подгруппой, факторгруппа по которой абелева.

Доказательство. 1) Пусть $G/G' = A$ и $\pi: G \rightarrow A$ — канонический гомоморфизм. Тогда $A' = \pi(G') = \{e\}$ и, значит, группа A абелева.

2) Пусть $N \subset G$ — такая нормальная подгруппа, что факторгруппа $G/N = A$ абелева, и пусть $\pi: G \rightarrow A$ — канонический гомоморфизм. Тогда $\pi(G') = A' = \{e\}$ и, значит, $G' \subset N$. \square

Для анализа дальнейших примеров нам понадобятся следующие предложения, представляющие и независимый интерес.

Предложение 1. Группа A_n порождается тройными циклами, а при $n \geq 5$ — произведениями пар независимых транспозиций.

Доказательство. Так как группа S_n порождается транспозициями, то группа A_n порождается произведениями пар транспозиций. Доказываемые утверждения вытекают из следующих соотношений:

$$\begin{aligned}(ij)(jk) &= (ijk), \\(ij)(kl) &= (ijk)(jkl), \\(ij)(jk) &= [(ij)(lm)][(jk)(lm)]\end{aligned}$$

(где i, j, k, l, m различны). \square

Предложение 2. Группа $SL_n(K)$ порождается элементарными матрицами первого типа, т. е. матрицами вида $E + cE_{ij}$ ($i \neq j$).

Доказательство. Слегка модифицируя метод Гаусса, покажем, что путем элементарных преобразований только первого типа любую матрицу $A \in SL_n(K)$ можно привести к единичной матрице. Отсюда, как и при доказательстве теоремы 4.4.2, будет следовать доказываемое утверждение.

Пусть $A = (a_{ij}) \in SL_n(K)$, $n > 1$. Вначале с помощью элементарных преобразований первого типа добьемся того, чтобы $a_{11} = 1$. Если $a_{i1} \neq 0$ для некоторого $i > 1$, то этого можно добиться за один шаг, прибавив к 1-й строке i -ю строку с подходящим коэффициентом. Если $a_{i1} = 0$ для всех $i > 1$, то $a_{11} \neq 0$ и, прибавив ко 2-й строке 1-ю строку, мы придем к уже рассмотренной ситуации.

Если уже $a_{11} = 1$, то, вычитая из каждой строки 1-ю строку с подходящим коэффициентом, добьемся того, чтобы $a_{i1} = 0$ при всех $i > 1$. После этого применим описанную процедуру к матрице порядка $n - 1$, получаемой вычеркиванием из матрицы A 1-й строки и 1-го столбца. Продолжая так дальше, мы в конце концов придем к треугольной матрице, все диагональные элементы которой, за ис-

ключением, быть может, последнего, равны 1. Но так как определитель матрицы был по условию равен 1 и он не изменился при сделанных преобразованиях, то и последний диагональный элемент полученной треугольной матрицы равен 1.

Наконец, с помощью обратного хода метода Гаусса, как и обычно, приведем полученную унитреугольную матрицу к единичной матрице. \square

Пример 1. Докажем, что $S'_n = A_n$. Так как группа S_n/A_n абелева, то $S'_n \subset A_n$. Так как группа S_3 не абелева и $|A_3| = 3$, то $S'_3 = A_3$. Следовательно, при любом n группа S'_n содержит все тройные циклы и, значит, совпадает с A_n .

Пример 2. Докажем, что $A'_4 = V_4$ и $A'_n = A_n$ при $n \geq 5$. Так как группа A_4/V_4 абелева, то $A'_4 \subset V_4$, но так как группа A_4 не абелева, то $A'_4 = V_4$. Следовательно, при любом n группа A'_n содержит все произведения пар независимых транспозиций и, значит, при $n \geq 5$ совпадает с A_n .

Пример 3. Докажем, что $SL_n(K)' = GL_n(K)' = SL_n(K)$, если поле K содержит более 3 элементов. (На самом деле это верно и при $|K| \leq 3$, если только $n \geq 3$.) Так как группа $GL_n(K)/SL_n(K) \simeq K^*$ абелева, то $GL_n(K)' \subset SL_n(K)$. Непосредственное вычисление показывает, что

$$\left(\left(\begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix}, \begin{pmatrix} 1 & c \\ 0 & 1 \end{pmatrix} \right) = \begin{pmatrix} 1 & (\lambda^2 - 1)c \\ 0 & 1 \end{pmatrix} \right).$$

Следовательно, если в поле K найдется элемент $\lambda \neq 0, \pm 1$, то группа $SL_n(K)'$ содержит все элементарные матрицы первого типа и, значит, совпадает с $SL_n(K)$.

Задача 1. Найти коммутант группы $\langle a \rangle_n \lambda_k \langle b \rangle_m$ (см. пример 1.16).

Определим *кратные коммутанты* $G^{(k)}$ группы G следующим образом:

$$G^{(1)} = G', \quad G^{(k+1)} = (G^{(k)})'.$$

Определение 1. Группа G называется *разрешимой*, если существует такое натуральное m , что $G^{(m)} = \{e\}$.

Очевидно, что всякая подгруппа и всякая факторгруппа разрешимой группы разрешимы. Обратное, если нормальная подгруппа N группы G и факторгруппа G/N разрешимы, то и группа G разрешима (докажите это).

Пример 4. Как следует из примеров 1 и 2, группа S_n разрешима при $n \leq 4$ и не разрешима при $n \geq 5$.

Пример 5. Докажем индукцией по n , что группа $B_n(K)$ (невырожденных) треугольных матриц разрешима. Группа $B_1(K) \simeq K^*$ абелева. Далее, вычеркивая из каждой треугольной матрицы порядка n последнюю строку и последний столбец, мы получаем гомоморфизм

$$f: B_n(K) \rightarrow B_{n-1}(K).$$

По предположению индукции группа

$$B_{n-1}(K) \simeq B_n(K)/\text{Ker } f$$

разрешима. Группа $\text{Ker } f$ состоит из матриц вида

$$\begin{pmatrix} 1 & 0 & c_1 \\ & \ddots & \vdots \\ 0 & & 1 & c_{n-1} \\ 0 & \dots & 0 & c_n \end{pmatrix} \quad (5)$$

Поставив в соответствие каждой такой матрице число c_n , мы получаем гомоморфизм $\text{Ker } f \rightarrow K^*$, ядро которого состоит из матриц вида (5) с $c_n = 1$ и, как легко видеть, коммутативно. Следовательно, группа $\text{Ker } f$, а значит, и группа $B_n(K)$, разрешимы.

§ 3. Действия

Напомним, что через $S(X)$ мы обозначаем группу всех преобразований (взаимно однозначных отображений на себя) множества X . В частности, $S(\{1, \dots, n\}) = S_n$ — симметрическая группа, или группа подстановок.

Всякая подгруппа группы $S(X)$ называется *группой преобразований* множества X . Многочисленные примеры групп преобразований уже встречались нам в этом курсе.

Понятие действия группы близко к понятию группы преобразований, но язык действий более гибок.

Определение 1. Действием группы G на множестве X называется любой гомоморфизм

$$\alpha: G \rightarrow S(X).$$

Иначе говоря, задать действие G на X — это значит поставить в соответствие каждому $g \in G$ преобразование $\alpha(g) \in S(X)$ таким

образом, что

$$\alpha(gh) = \alpha(g)\alpha(h). \quad (6)$$

По общему свойству гомоморфизмов единице группы G соответствует тождественное преобразование id , а обратному элементу — обратное преобразование.

Результат применения преобразования $\alpha(g)$ к «точке» $x \in X$ обозначается через $\alpha(g)x$ или, если ясно, о каком действии идет речь, просто через gx . Свойство (6) переписывается в виде «ассоциативности»:

$$(gh)x = g(hx).$$

Если задано действие α группы G на X , то пишут $G : X$ или просто $G : X$.

Всякая группа преобразований $G \subset S(X)$ действует на X «тавтологически» — путем тождественного гомоморфизма $G \rightarrow S(X)$.

Обратно, для любого действия $G : X$ подгруппа $\text{Im } \alpha \subset S(X)$ есть некоторая группа преобразований множества X . По теореме о гомоморфизме

$$\text{Im } \alpha \simeq G/\text{Ker } \alpha.$$

Нормальная подгруппа $\text{Ker } \alpha \subset G$ называется *ядром неэффективности* действия α . Если $\text{Ker } \alpha = \{e\}$, то действие α называется *эффективным*.

Частным случаем действия является *линейное представление* — гомоморфизм группы G в группу $\text{GL}(V)$ (обратимых) линейных преобразований векторного пространства V .

Всякое действие $G : X$ естественным образом порождает ряд других действий: на любом инвариантном подмножестве множества X , на множестве всех (или не всех) подмножеств множества X , на фактормножестве множества X по инвариантному отношению эквивалентности и т. п. Отметим особо индуцируемое действием α линейное представление α_* группы G в пространстве всех (или не всех) функций на X со значениями в поле K , определяемое по формуле

$$(\alpha_*(g)f)(x) = f(\alpha(g)^{-1}x). \quad (7)$$

Обычно мы опускаем символ α ; тогда эта формула переписывается в виде

$$(gf)(x) = f(g^{-1}x). \quad (8)$$

Всякое действие $G : X$, ограниченное на подгруппу $H \subset G$, определяет действие $H : X$.

Для любой группы G определяются следующие три ее важных действия на самой себе:

1) действие l левыми сдвигами:

$$l(g)x = gx;$$

2) действие r правыми сдвигами:

$$r(g)x = xg^{-1};$$

3) действие a сопряжениями (внутренними автоморфизмами):

$$a(g)x = gxg^{-1}.$$

Действие l (а также r) эффективно, так что $G \simeq \text{Im } l \subset S(G)$. Отсюда, в частности, получается *теорема Кэли*: всякая конечная группа порядка n изоморфна некоторой подгруппе группы S_n . Ядро неэффективности действия a , как мы уже видели в § 1, есть центр Z группы G .

Действие $G : X$ определяет на множестве X отношение эквивалентности \sim_G по правилу:

$$x \sim_G y, \text{ если существует } g \in G, \text{ такой, что } y = gx$$

(проверьте аксиомы отношения эквивалентности!). Классы эквивалентности называются *орбитами*. Таким образом, орбита, содержащая точку x (орбита точки x), есть подмножество

$$Gx = \{gx : g \in G\} \subset X.$$

Если все элементы множества X эквивалентны (т. е. имеется всего одна орбита), то действие называется *транзитивным*.

Пример 1. Орбиты группы поворотов плоскости вокруг точки o суть окружности с центром в точке o и сама эта точка.

Пример 2. Группа всех движений и даже группа параллельных переносов действуют на плоскости транзитивно.

Пример 3. Четверная группа Клейна V_4 действует на множестве $\{1, 2, 3, 4\}$ транзитивно.

Пример 4. Если ограничить действие l (соответственно r) группы G на себе на подгруппу $H \subset G$, то орбитами будут правые (соответственно левые) смежные классы по H .

Элементы группы G , эквивалентные относительно действия a группы G на себе сопряжениями, называются *сопряженными*, а орбиты этого действия — *классами сопряженных элементов*.

Если группа G задана как группа преобразований некоторого множества X , то описание ее классов сопряженных элементов может быть получено с помощью следующего простого соображения: если элемент $g \in G$ переводит точку x в точку y , то элемент ghg^{-1} переводит точку hx в точку hy .

Пример 5. Пусть подстановка $\sigma \in S_n$ разложена в произведение независимых циклов:

$$\sigma = (i_1 i_2 \dots i_p)(j_1 j_2 \dots j_q) \dots$$

Тогда для любой подстановки $\tau \in S_n$ имеем:

$$\tau \sigma \tau^{-1} = (\tau(i_1) \tau(i_2) \dots \tau(i_p)) (\tau(j_1) \tau(j_2) \dots \tau(j_q)) \dots$$

Отсюда следует, что две подстановки сопряжены тогда и только тогда, когда в их разложениях в произведение независимых циклов наборы длин циклов совпадают. Таким образом, число классов сопряженных элементов в группе S_n равно числу (неупорядоченных) разбиений числа n в сумму натуральных чисел.

Пример 6. Группа $\text{Isom}_+ E^2$ собственных движений плоскости состоит из параллельных переносов и поворотов (см. § 7.4). Из сформулированного выше принципа следует, что движение, сопряженное при помощи движения h параллельному переносу на вектор a , есть параллельный перенос на вектор $dh(a)$ (см. рис. 1, а). (Впрочем, мы уже доказали это в § 4.2.) Аналогичным образом, движение, сопряженное при помощи собственного движения h повороту на угол α вокруг точки o , есть поворот на угол α вокруг

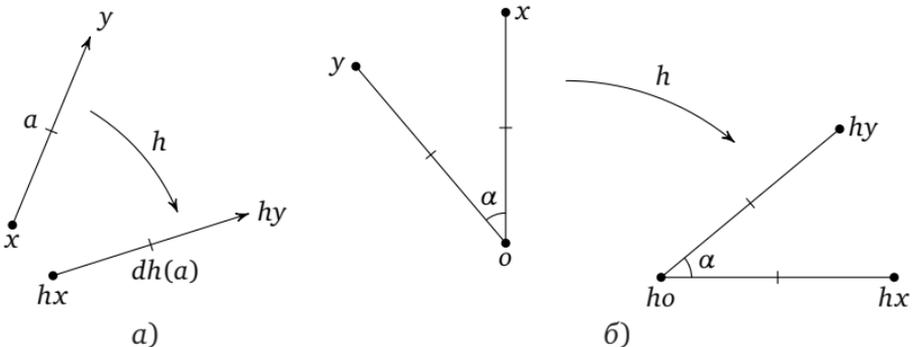


Рис. 1

точки ho (см. рис. 1, б)). Таким образом, классы сопряженных элементов группы $\text{Isom}_+ E^2$ суть подмножества следующих двух видов:

- 1) совокупность параллельных переносов на векторы заданной длины $r \geq 0$;
- 2) совокупность поворотов на заданный угол $\alpha \in (0, 2\pi)$.

Задача 1. Описать классы сопряженных элементов в группе вращений куба.

Пусть заданы действия α и β одной и той же группы G на множествах X и Y соответственно. Отображение $f: X \rightarrow Y$ называется *эквивариантным* или, точнее, *G -эквивариантным*, если для любого $g \in G$ диаграмма

$$\begin{array}{ccc} X & \xrightarrow{\alpha(g)} & X \\ f \downarrow & & \downarrow f \\ Y & \xrightarrow{\beta(g)} & Y \end{array}$$

коммутативна. Эквивариантная биекция называется *изоморфизмом действий*. (Диаграмма, состоящая из множеств и отображений, называется *коммутативной*, если композиция отображений вдоль любых двух путей, имеющих общее начало и общий конец, дает один и тот же результат.)

Подгруппа

$$G_x = \{g \in G: gx = x\}$$

называется *стабилизатором* точки x .

Для любой подгруппы $H \subset G$ определим действие группы G на множестве G/H левых смежных классов по H формулой

$$g(uH) = (gu)H.$$

Следующая теорема, показывающая, что действие группы G на орбите Gx с точностью до изоморфизма определяется стабилизатором точки x , является обобщением и уточнением теоремы 4.5.2.

Теорема 1. *Отображение*

$$f: Gx \rightarrow G/G_x, \quad y \mapsto G_x^y = \{g \in G: gx = y\}$$

является *изоморфизмом действий*.

Доказательство. 1) Покажем, что при $y = gx$ подмножество G_x^y совпадает со смежным классом gG_x . Действительно,

$$g_1x = y \Leftrightarrow g^{-1}g_1x = x \Leftrightarrow g^{-1}g_1 \in G_x \Leftrightarrow g_1 \in gG_x.$$

2) Из определения ясно, что отображение f биективно.

3) Покажем, что отображение f эквивариантно. Пусть $y = ux$ ($u \in G$). Для любого $g \in G$ имеем

$$f(gy) = f((gu)x) = (gu)G_x = g(uG_x) = gf(y). \quad \square$$

Следствие 1. *Всякое транзитивное действие группы изоморфно ее действию на множестве левых смежных классов по некоторой подгруппе.*

Следствие 2. *Если группа G конечна, то*

$$|Gx| = \frac{|G|}{|G_x|}. \quad (9)$$

(Здесь $|Gx|$ обозначает число элементов орбиты Gx .)

Легко видеть, что

$$G_{gx} = gG_xg^{-1}. \quad (10)$$

Так как в теореме 1 в качестве точки x можно взять любую точку орбиты, то отсюда следует, что для любой подгруппы $H \subset G$ и любого $g \in G$ действия группы G на G/H и на G/gHg^{-1} изоморфны.

Ядро неэффективности действия $G : G/H$ есть пересечение стабилизаторов всех точек, т. е. $\bigcap_{g \in G} gHg^{-1}$. Это наибольшая нормальная подгруппа группы G , содержащаяся в H . В частности, подгруппа H нормальна тогда и только тогда, когда она тривиально действует на G/H .

Пример 7. Пусть $K \subset E^3$ — куб. Изоморфизм $S_4 \xrightarrow{\sim} \text{Sym}_+ K$ (см. пример 4.6.19) определяет действие $S_4 : E^3$. В свою очередь, это действие индуцирует действие группы S_4 на множестве вершин куба, на множестве его диагоналей и т. д. В табл. 1 перечислены некоторые получаемые таким образом транзитивные действия $S_4 : X$. Для каждого из них указан стабилизатор H одного из элементов множества X . Во всех случаях $|X||H| = |S_4| = 24$, как и должно быть согласно следствию 2 теоремы 1.

Пример 8. Докажем, что если $|G| = n < \infty$ и p — наименьший простой делитель числа n , то всякая подгруппа $H \subset G$ индекса p нормальна. Действительно, рассмотрим действие $H : G/H$. Число элементов любой орбиты этого действия делит $|H|$ и, значит, либо равно 1, либо больше или равно p . Так как $|G/H| = p$ и имеется

Таблица 1

Элементы множества X	$ X $	$ H $	H
ребра куба	12	2	$\langle(12)\rangle$
диагонали граней куба	12	2	$\langle(12)(34)\rangle$
вершины куба	8	3	$\langle(123)\rangle$
грани куба	6	4	$\langle(1234)\rangle$
пары противоположных ребер куба	6	4	$\langle(12), (34)\rangle$
пары противоположных вершин (или диагонали) куба	4	6	S_3
пары противоположных граней куба	3	8	$\langle V_4, (1234)\rangle$

по меньшей мере одна неподвижная точка (смежный класс eH), то отсюда следует, что действие тривиально.

Задача 2. Пусть задано действие конечной группы G на конечном множестве X . Множество орбит этого действия обозначим через X/G , и для каждого элемента $g \in G$ множество его неподвижных точек в X обозначим через X^g . Доказать формулу Бернсайда:

$$|X/G| = \frac{1}{|G|} \sum_{g \in G} |X^g|.$$

(Указание: подсчитать двумя способами число элементов множества $F = \{(g, x) \in G \times X : gx = x\}$.)

Задача 3. Пользуясь формулой Бернсайда и задачей 1, найти число существенно различных раскрасок граней куба в 3 цвета. (Две раскраски считаются существенно различными, если они не могут быть совмещены путем вращения куба.)

Для действия группы G на самой себе сопряжениями стабилизатором точки x служит подгруппа

$$Z(x) = \{g \in G : gx = xg\},$$

называемая *централизатором* элемента x . Обозначим через $C(x)$ класс сопряженных элементов (орбиту этого действия), содержащий x . Для конечной группы G формула (9) дает

$$|C(x)| = \frac{|G|}{|Z(x)|}. \quad (11)$$

Действие группы G сопряжениями на самой себе порождает ее действие на множестве ее подгрупп. Подгруппы, эквивалентные относительно этого действия, называются *сопряженными*. (Так, стабилизаторы эквивалентных точек для любого действия группы G являются в силу формулы (10) сопряженными подгруппами.) Стабилизатором подгруппы $H \subset G$ для этого действия служит подгруппа

$$N(H) = \{g \in G : gHg^{-1} = H\},$$

называемая *нормализатором* подгруппы H . В случае конечной группы G формула (9) показывает, что число подгрупп, сопряженных подгруппе H , равно $[G : N(H)]$ (индексу подгруппы $N(H)$). Отметим, что $N(H) \supset H$ и потому в случае конечной группы $[G : N(H)]$ делит $[G : H]$.

§ 4. Теоремы Силова

Пусть p — простое число. Напомним, что конечная группа G называется p -группой, если $|G| = p^n$.

Теорема 1. *Центр нетривиальной p -группы нетривиален.*

Доказательство. Пусть G — нетривиальная p -группа, а Z — ее центр. Множество $G \setminus Z$ разбивается на нетривиальные классы сопряженных элементов, число элементов в каждом из которых, согласно формуле (11), делится на p . Следовательно, число элементов центра также делится на p . \square

Следствие 1. *Всякая p -группа разрешима.*

Доказательство. Пусть G — нетривиальная p -группа, а Z — ее центр. Доказывая утверждение индукцией по $n = \log_p |G|$, мы можем считать, что группа G/Z разрешима. Так как группа Z тоже разрешима (даже абелева), то отсюда следует, что и группа G разрешима. \square

Следствие 2. *Всякая группа порядка p^2 абелева.*

Доказательство. Пусть G — группа порядка p^2 и Z — ее центр. Предположим, что $Z \neq G$. Тогда $|Z| = p$ и $|G/Z| = p$, так что группа G/Z является циклической. Пусть aZ — ее порождающий элемент. Тогда каждый элемент $g \in G$ представляется в виде $g = a^k z$ ($z \in Z$). Но любые два элемента такого вида коммутируют, что противоречит нашему предположению. \square

Пусть теперь $|G| = p^n m$, где $(p, m) = 1$.

Определение 1. Силовской p -подгруппой группы G называется всякая ее подгруппа порядка p^n .

Теорема 2. Силовская p -подгруппа существует.

Доказательство. Если группа G абелева, то ее (единственной) силовской p -подгруппой является подгруппа p -кращения (см. § 9.1).

В общем случае докажем теорему индукцией по $|G|$.

Если $|G| = 1$, то доказывать нечего. Пусть $|G| > 1$. Рассмотрим разбиение группы G на классы сопряженных элементов. Возможны два случая.

1-й случай. Существует нетривиальный класс $C(x)$, число элементов которого не делится на p . Тогда $|Z(x)|$ делится на p^n и по предположению индукции в $Z(x)$ имеется подгруппа порядка p^n . Она и будет силовской p -подгруппой в G .

2-й случай. Не существует такого класса. Тогда, как и в доказательстве теоремы 1, получаем, что $|Z|$ делится на p . Пусть $|Z| = p^{n_0} m_0$, где $(p, m_0) = 1$, и пусть $Z_1 \subset Z$ — подгруппа порядка p^{n_0} . В группе G/Z_1 , порядок которой равен $p^{n-n_0} m$, по предположению индукции существует подгруппа порядка p^{n-n_0} . Ее полный прообраз при каноническом гомоморфизме $G \rightarrow G/Z_1$ и будет силовской p -подгруппой в G . \square

Теорема 3. Всякая p -подгруппа группы G содержится в некоторой силовской p -подгруппе. Все силовские p -подгруппы сопряжены.

Доказательство. Пусть $S \subset G$ — силовская p -подгруппа и S_1 — какая-либо p -подгруппа. Рассмотрим действие S_1 на G/S . Так как число элементов любой нетривиальной S_1 -орбиты делится на p , а число всех элементов множества G/S не делится на p , то S_1 имеет в G/S неподвижные точки. Если gS — такая точка, то $S_1 \subset gSg^{-1}$, что доказывает первое утверждение теоремы. Если, кроме того, S_1 — силовская p -подгруппа, то сравнение порядков дает равенство $S_1 = gSg^{-1}$. \square

Задача 1. Рассуждая аналогичным образом, доказать, что в группе порядка p^n всякая подгруппа H порядка p^k , $k < n$, имеет неподвижные точки в G/H , отличные от eH . Вывести отсюда, что $N(H) \neq H$ и что H содержится в некоторой подгруппе порядка p^{k+1} .

Теорема 4. Число силовских p -подгрупп сравнимо с 1 по модулю p .

Доказательство. Пусть S — силовская p -подгруппа и $C(S)$ — класс подгрупп, сопряженных S . По теореме 3 это и есть совокупность всех силовских p -подгрупп. При действии G на $C(S)$ со-

пряжениями стабилизатором любой подгруппы $S' \in C(S)$ служит ее нормализатор $N(S')$. Ограничим это действие на S . Тогда $C(S)$ каким-то образом разобьется на нетривиальные S -орбиты, число элементов в каждой из которых делится на p , и на неподвижные точки. Докажем, что единственной неподвижной точкой является сама подгруппа S , откуда и будет следовать что

$$|C(S)| \equiv 1 \pmod{p}.$$

Пусть $S' \in C(S)$ — неподвижная точка. Это означает, что $S \subset N(S')$. Тогда S и S' — силовские p -подгруппы группы $N(S')$ и, значит, сопряжены в ней. Но S' — нормальная подгруппа в $N(S')$. Следовательно, $S = S'$. \square

Теорема 4 в соединении с тем фактом, что число силовских p -подгрупп делит индекс (любой) силовской p -подгруппы, иногда позволяет доказать, что силовская p -подгруппа единственна и, значит, нормальна.

Пример 1. Докажем, что всякая группа G порядка pq , где p и q — различные простые числа, является полупрямым произведением циклических групп порядков p и q (см. пример 1.16). Действительно, пусть $p > q$. Тогда силовская p -подгруппа G_p нормальна в силу примера 3.8. Если G_q — силовская q -подгруппа, то $G_p \cap G_q = \{e\}$ и, значит, $|G_p G_q| = pq = |G|$. Следовательно,

$$G = G_p \rtimes G_q.$$

Пример 2. Докажем, что всякая группа G порядка 45 абелева. Действительно, пусть N_p ($p = 3, 5$) — число ее силовских p -подгрупп. Тогда

$$N_3 \equiv 1 \pmod{3}, N_3 | 5 \Rightarrow N_3 = 1,$$

$$N_5 \equiv 1 \pmod{5}, N_5 | 9 \Rightarrow N_5 = 1,$$

так что силовские подгруппы G_3 и G_5 нормальны и

$$G = G_3 \times G_5.$$

Но группа G_3 имеет порядок 9 и, значит, абелева (следствие 2 теоремы 1). Следовательно, и группа G абелева.

Задача 2. Доказать, что все группы порядка < 60 разрешимы. (Указание: доказать, что если $|G| = n < 60$, то для некоторого простого p , $p | n$, число N силовских p -подгрупп группы G не превосходит 4; если при этом $N > 1$, то, рассмотрев действие группы G на

множестве силовских p -подгрупп сопряжениями, получить нетривиальный гомоморфизм $G \rightarrow S_N$.)

§ 5. Простые группы

Определение 1. Нетривиальная группа G называется *простой*, если она не содержит нетривиальных (т. е. отличных от $\{e\}$ и G) нормальных подгрупп.

Всякая разрешимая простая группа G есть циклическая группа простого порядка. Действительно, так как $G' \neq G$, то $G' = \{e\}$, т. е. группа G абелева. Но в абелевой группе все подгруппы нормальны. Следовательно, группа G циклическая и, более того, простого порядка.

Таким образом, простые группы бывают двух типов:

1) абелевы, к которым относятся лишь циклические группы простого порядка;

2) неабелевы (и, следовательно, неразрешимые).

Примером неабелевой простой группы может служить группа A_5 (доказательство ее простоты см. ниже).

Значение простых групп может быть пояснено следующими соображениями. Пусть имеется цепочка вложенных подгрупп

$$G = G_0 \supset G_1 \supset \dots \supset G_{m-1} \supset G_m = \{e\}, \quad (12)$$

в которой $G_{i+1} \triangleleft G_i$ ($i = 0, 1, \dots, m-1$). Если факторгруппа

$$F_i = G_i / G_{i+1}$$

содержит нетривиальную нормальную подгруппу N , то между G_i и G_{i+1} можно вставить еще одну подгруппу, а именно полный образ N при каноническом гомоморфизме $G_i \rightarrow F_i$. Поэтому, если, например, группа G конечна, то можно построить цепочку (12), в которой все факторы просты. В любом случае такая цепочка, если она существует, называется *композиционным рядом* группы G .

Несложно доказывается *теорема Жордана—Гёльдера*: если группа G обладает композиционным рядом, то набор его факторов определен однозначно с точностью до перестановки. Таким образом, с каждой группой, обладающей композиционным рядом (например, с каждой конечной группой), каноническим образом связывается набор простых групп. Поэтому классификация простых групп имеет фундаментальное значение для понимания строения любых групп.

Задача 1. Доказать, что конечная группа разрешима тогда и только тогда, когда все факторы ее композиционного ряда абелевы.

Классификация неабелевых простых конечных групп чрезвычайно сложна. Она была получена в результате 30-летней работы нескольких сотен математиков всего мира, завершившейся в 1981 г. Мы ограничимся рассмотрением примеров.

Предложение 1. *Группа A_n проста при $n \geq 5$.*

Лемма 1. *Если в разложении подстановки $\sigma \in A_n$ в произведение независимых циклов присутствует цикл четной длины или два цикла одинаковой нечетной длины, то класс сопряженности подстановки σ в A_n совпадает с ее классом сопряженности в S_n .*

(Неподвижный символ следует рассматривать как цикл длины 1.)

Доказательство. В любом из рассматриваемых случаев в централизаторе подстановки σ в группе S_n имеется нечетная подстановка τ_0 . Действительно, если в разложении σ присутствует цикл четной длины, то в качестве τ_0 можно взять этот цикл, а если присутствуют циклы $(i_1 i_2 \dots i_q)$, $(j_1 j_2 \dots j_q)$ одинаковой нечетной длины q , то можно взять $\tau_0 = (i_1 j_1)(i_2 j_2) \dots (i_q j_q)$. Пусть теперь τ — любая нечетная подстановка. Тогда

$$\tau \sigma \tau^{-1} = (\tau \tau_0) \sigma (\tau \tau_0)^{-1},$$

причем подстановка $\tau \tau_0$ уже четна. □

В частности, все произведения пар независимых транспозиций, а при $n \geq 5$ и все тройные циклы сопряжены не только в S_n , но и в A_n .

Доказательство предложения 1. Пусть $N \subset A_n$ — нормальная подгруппа, содержащая какую-то подстановку $\sigma \neq e$. Возведя подстановку σ в подходящую степень, можно добиться, чтобы она имела простой порядок p . Тогда σ есть произведение какого-то числа независимых циклов длины p .

Рассмотрим отдельно три возможности.

1) Пусть $p \geq 5$. Запишем подстановку σ в виде

$$\sigma = (i_1 i_2 i_3 i_4 \dots i_p) \tau,$$

где τ — произведение остальных циклов (если их нет, то $\tau = e$), и сопряжем ее с помощью тройного цикла $(i_1 i_2 i_3)$. Мы получим (см. пример 3.5):

$$\sigma' = (i_1 i_2 i_3) \sigma (i_1 i_2 i_3)^{-1} = (i_2 i_3 i_1 i_4 \dots i_p) \tau \in N$$

и, значит, $\sigma'\sigma^{-1} = (i_1i_2i_4) \in N$. Так как все тройные циклы сопряжены в A_n и группа A_n ими порождается (см. предложение 2.1), то $N = A_n$.

2) Пусть $p = 3$. Если σ есть просто тройной цикл, то, как и выше, получаем, что $N = A_n$. В противном случае запишем подстановку σ в виде

$$\sigma = (i_1i_2i_3)(j_1j_2j_3)\tau,$$

где τ — произведение остальных циклов, и сопряжем ее с помощью подстановки $(i_1j_1)(i_2j_2)$. Мы получим:

$$\sigma' = (i_1i_2j_3)(j_1j_2i_3)\tau \in N,$$

$$\sigma'\sigma^{-1} = (i_1j_1)(i_3j_3) \in N.$$

Так как все произведения пар независимых транспозиций сопряжены в A_n и группа A_n ими порождается (см. предложение 2.1), то и в этом случае $N = A_n$.

3) Пусть, наконец, $p = 2$. Тогда подстановка σ есть произведение какого-то четного числа независимых транспозиций. Запишем ее в виде

$$\sigma = (i_1i_2)(i_3i_4)\tau,$$

где τ — произведение остальных транспозиций, и сопряжем с помощью цикла $(i_1i_2i_3)$. Мы получим:

$$\sigma' = (i_2i_3)(i_1i_4)\tau \in N,$$

$$\sigma'\sigma^{-1} = (i_1i_3)(i_2i_4) \in N,$$

откуда, как и выше, следует, что $N = A_n$. □

В частности, группа A_5 является простой группой порядка 60. В силу задачи 4.2 это наименьший порядок, который может иметь неабелева простая группа. Заметим, что при $n = 5$ предыдущее доказательство сильно упрощается, сводясь (с учетом леммы) к рассмотрению случая, когда σ — цикл длины 5.

Задача 2. Доказать, что единственной нетривиальной нормальной подгруппой группы A_4 является четверная группа Клейна V_4 .

Задача 3. Доказать, что всякая простая группа G порядка 60 изоморфна A_5 . (Указание: рассмотрев действие группы G на множестве силовских 5-подгрупп, получить вложение $G \subset A_6$; далее рассмотреть действие группы G на A_6/G .)

В качестве примера еще одной серии простых групп приведем без доказательства следующий факт: при $n \geq 2$ группа

$$\mathrm{PSL}_n(K) = \mathrm{SL}_n(K) / \{\lambda E : \lambda \in K^*, \lambda^n = 1\}$$

проста, за исключением случаев, когда $n = 2$ и K — конечное поле из 2 или 3 элементов.

Задача 4. Найти порядок группы $\mathrm{PSL}_2(\mathbb{F}_q)$, где \mathbb{F}_q — конечное поле из q элементов. Доказать, что

$$\mathrm{PSL}_2(\mathbb{F}_2) \simeq S_3, \quad \mathrm{PSL}_2(\mathbb{F}_3) \simeq A_4, \quad \mathrm{PSL}_2(\mathbb{F}_4) \simeq \mathrm{PSL}_2(\mathbb{F}_5) \simeq A_5.$$

(Указание: рассмотреть естественное действие группы $G = \mathrm{PSL}_2(\mathbb{F}_q)$ на проективной прямой $P(\mathbb{F}_q^2)$ над полем \mathbb{F}_q ; при $q = 5$ рассуждать, как в задаче 3.)

Группа $\mathrm{PSL}_2(\mathbb{F}_7)$ есть простая группа порядка 168. Это следующая по порядку неабелева простая группа после группы A_5 . Группа $\mathrm{PSL}_2(\mathbb{F}_9)$, как можно показать, изоморфна группе A_6 .

Задача 5. Доказать, что группа $\mathrm{PSL}_2(\mathbb{C})$ проста.

В качестве примера использования геометрических соображений для доказательства простоты (бесконечной) группы докажем, что группа SO_3 проста.

Всякий элемент группы SO_3 есть поворот на некоторый угол α вокруг некоторой оси. Преобразование, сопряженное с помощью элемента $g \in \mathrm{SO}_3$ повороту на угол α вокруг оси l , есть поворот на тот же угол α вокруг оси gl (ср. пример 3.6). Поэтому всякая нормальная подгруппа группы SO_3 вместе с поворотом на угол α вокруг какой-либо оси должна содержать поворот на угол α вокруг любой оси.

Легко видеть (ср. пример 6.3.4), что произведение поворотов на π вокруг осей m и m' , образующих между собой угол γ , есть поворот на угол 2γ вокруг оси, перпендикулярной плоскости осей m и m' .

Предположим теперь, что $N \subset \mathrm{SO}_3$ — нормальная подгруппа, содержащая поворот s на угол $\alpha \in (0, 2\pi)$ вокруг какой-то оси l . Пусть g — поворот на π вокруг оси m , образующей с осью l угол $\theta \in [0, \pi/2]$. Тогда

$$h = g(sgs^{-1}) = (gsg^{-1})s^{-1} \in N;$$

но так как sgs^{-1} есть поворот на π вокруг оси sm , то, согласно предыдущему замечанию, h есть поворот на угол 2γ , где γ — угол

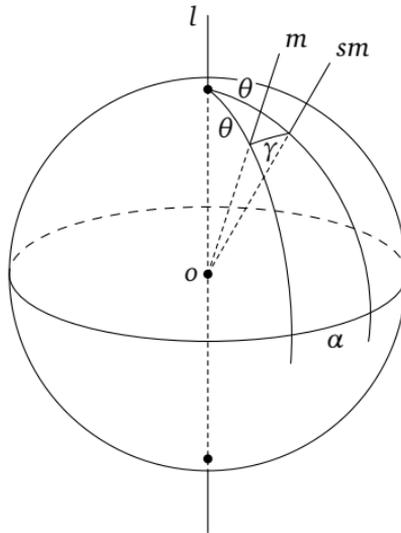


Рис. 2

между m и sm (см. рис. 2). Угол γ равен 0 при $\theta = 0$ и равен α при $\theta = \pi/2$. Из соображений непрерывности следует, что он может принимать любые значения в отрезке $[0, \alpha]$. Следовательно, группа N содержит повороты на все углы из отрезка $[0, 2\alpha]$. Возведением этих поворотов в степени можно получить поворот на любой угол. Это показывает, что $N = SO_3$.

Можно показать, что группа SO_n проста при любом $n \geq 3$, за исключением $n = 4$. Отсутствие простоты группы SO_4 является удивительным обстоятельством, которое будет обсуждено в § 12.4.

§ 6. Расширения Галуа

Расширения поля K , получаемые присоединением корней различных неприводимых многочленов, могут оказаться изоморфными или, более общо, одно из них может изоморфно вкладываться в другое. Выяснить, когда это имеет место, не так просто. Изучением гомоморфизмов (и, в частности, автоморфизмов) алгебраических расширений полей как раз и занимается теория Галуа.

В § 4.2 было рассказано о том, какую роль играют группы в геометрии и физике. Однако своим происхождением теория групп обязана теории Галуа, в которой группы появляются принципиально

иным образом. Идеи теории Галуа находят воплощение и в других разделах математики. Так, аналогом теории Галуа в топологии является теория накрытий (в частности, аналогом группы Галуа поля является фундаментальная группа топологического пространства), а в теории функций комплексной переменной — теория голоморфных отображений римановых поверхностей.

Пусть L — конечное расширение степени n поля K . Автоморфизмы поля L над K образуют группу, которую мы обозначим через $\text{Aut}_K L$.

Предложение 1. $|\text{Aut}_K L| \leq n$.

Доказательство. Поле L может быть получено из K последовательными простыми расширениями:

$$K = K_0 \subset K_1 \subset K_2 \subset \dots \subset K_s = L,$$

где K_i получается из K_{i-1} присоединением корня α_i какого-то неприводимого многочлена $f_i \in K_{i-1}[x]$. Согласно лемме 9.5.1, любой гомоморфизм $\varphi_{i-1}: K_{i-1} \rightarrow L$ продолжается до гомоморфизма $\varphi_i: K_i \rightarrow L$ не более чем

$$n_i = \deg f_i = \dim_{K_{i-1}} K_i$$

способами. Следовательно, тождественный автоморфизм поля K продолжается до автоморфизма поля L не более чем $n_1 n_2 \dots n_s = n$ способами. \square

Пусть $G \subset \text{Aut}_K L$ — какая-то (конечная) группа автоморфизмов поля L над K . Обозначим через L^G подполе G -инвариантных элементов поля L .

Теорема 1. $L^G = K$ тогда и только тогда, когда $|G| = n$. Кроме того, если $L^G = K$, то для любых полей P, Q , таких, что $K \subset P \subset Q \subset L$, всякий гомоморфизм $\varphi: P \rightarrow L$ над K продолжается до гомоморфизма $\psi: Q \rightarrow L$ ровно $\dim_P Q$ способами.

Доказательство. 1) Согласно определению, $G \subset \text{Aut}_{L^G} L$. Следовательно,

$$|G| \leq \dim_{L^G} L \leq \dim_K L = n.$$

Если же $|G| = n$, то $\dim_{L^G} L = \dim_K L$ и, значит, $L^G = K$.

2) Обратно, пусть $L^G = K$. Для любого элемента $\alpha \in L$ пусть $\{\alpha_1, \dots, \alpha_m\}$ — его G -орбита. Тогда

$$f = \prod_{i=1}^m (x - \alpha_i) \in L^G[x] = K[x] \quad (13)$$

есть минимальный многочлен элемента α над K . По построению он разлагается на различные линейные множители в $L[x]$.

Докажем теперь второе утверждение теоремы. Так как всякое конечное расширение может быть получено путем последовательных простых расширений, то достаточно доказать его в случае, когда $Q = P(\alpha)$ — простое расширение поля P . Пусть h — минимальный многочлен элемента α над P . Тогда h делит минимальный многочлен f элемента α над K в кольце $P[x]$. Следовательно, h^φ делит f в кольце $\varphi(P)[x]$ и, значит, разлагается на различные линейные множители в $L[x]$. Согласно лемме 9.5.1, гомоморфизм φ продолжается до гомоморфизма $\psi: Q \rightarrow L$ ровно $\deg h = \dim_P Q$ способами.

Применяя доказанное утверждение к случаю $P = K$, $Q = L$, получаем, что $|\text{Aut}_K L| = n$.

Остается доказать, что $G = \text{Aut}_K L$. Пусть $\varphi \in \text{Aut}_K L$. Тогда для любого $\alpha \in L$ элемент $\varphi(\alpha)$, как и α , является корнем многочлена (13), т. е. существует такой элемент $g \in G$ (быть может, зависящий от α), что $\varphi(\alpha) = g\alpha$.

Если поле L конечно, то в качестве α возьмем элемент, порождающий группу L^* , и тогда мы получим, что $\varphi = g \in G$. Если же L (и, стало быть, K) бесконечно, то для каждого $g \in G$ положим

$$L_g = \{\alpha \in L: \varphi(\alpha) = g\alpha\} \subset L.$$

Очевидно, что L_g — подпространство над K (и даже подполе). Из доказанного следует, что

$$L = \bigcup_{g \in G} L_g.$$

Отсюда мы должны получить, что на самом деле $L = L_g$ для некоторого $g \in G$. Это вытекает из следующей ниже леммы. \square

Лемма 1. *Конечномерное векторное пространство V над бесконечным полем не может быть покрыто конечным числом собственных подпространств.*

Доказательство. Пусть $V = \bigcup_{i=1}^s V_i$, где V_1, \dots, V_s — собственные подпространства. Для каждого i возьмем какую-нибудь ненулевую линейную функцию $l_i \in V^*$, обращающуюся в нуль на V_i , и рассмотрим многочлен $F = \prod_{i=1}^s l_i$. Из условия следует, что $F(v) = 0$ для любого $v \in V$; но тогда F — нулевой многочлен, что, очевидно, неверно. \square

Определение 1. Конечное расширение L поля K называется *расширением Галуа*, если

$$|\text{Aut}_K L| = \dim_K L.$$

Группа $\text{Aut}_K L$ в этом случае называется *группой Галуа* расширения L и обозначается через $\text{Gal } L/K$.

(Косая черта здесь не подразумевает никакого факторобразования.)

Из теоремы 1 следует, что если L — расширение Галуа поля K и $P \supset L$ — подполе, содержащее K , то L — расширение Галуа поля P .

Многочлен $f \in K[x]$ называется *сепарабельным*, если он не имеет кратных корней ни в каком расширении поля K .

Обозначим через f' формальную производную многочлена f .

Предложение 2. Многочлен $f \in K[x]$ сепарабелен тогда и только тогда, когда $(f, f') = 1$.

Доказательство. Заметим, прежде всего, что наибольший общий делитель любых двух многочленов $f, g \in K[x]$ может быть найден с помощью алгоритма Евклида и потому не изменяется при любом расширении поля K . С другой стороны, если над каким-либо расширением L поля K многочлен f имеет кратный корень c , то c будет также корнем производной и, значит, $(f, f') \neq 1$. Таким образом, если $(f, f') = 1$, то многочлен f сепарабелен.

Обратно, если многочлен f сепарабелен и L — какое-либо расширение поля K , над которым он разлагается на линейные множители, то ни один из его корней в L не является корнем производной (поскольку все они простые) и, значит, $(f, f') = 1$. \square

Из доказанного предложения следует, что неприводимый многочлен несепарабелен, только если его производная является нулевым многочленом. Это, конечно, невозможно, если $\text{char } K = 0$; если же $\text{char } K = p > 0$, то это имеет место только для многочленов вида

$$f = a_0 + a_1 x^p + a_2 x^{2p} + \dots + a_m x^{mp}. \quad (14)$$

Таким образом, всякий неприводимый многочлен f над полем нулевой характеристики или над полем характеристики $p \nmid \deg f$ сепарабелен.

Предложение 3. Всякий неприводимый многочлен над конечным полем сепарабелен.

Доказательство. Пусть f — несепарабельный неприводимый многочлен над конечным полем K характеристики p . Тогда он

имеет вид (14). Так как $K^p = K$ (см. § 9.5), то существуют такие $b_0, b_1, \dots, b_m \in K$, что $b_k^p = a_k$; но тогда

$$f = (b_0 + b_1x + b_2x^2 + \dots + b_mx^m)^p, \quad (15)$$

что противоречит предположению о неприводимости. \square

Примером несепарабельного неприводимого многочлена может служить многочлен

$$x^p - t = (x - \sqrt[p]{t})^p$$

над полем $\mathbb{Z}_p(t)$.

Теорема 2. Пусть $f \in K[x]$ — многочлен, все неприводимые множители которого сепарабельны. Тогда его поле разложения над K является расширением Галуа.

Доказательство получается путем анализа доказательства второй части теоремы 9.5.6 в случае $\tilde{L} = L$ с учетом того, что в условиях настоящей теоремы все многочлены \tilde{f}_i сепарабельны. \square

Заметим, что если L — поле разложения многочлена $f \in K[x]$, то любой автоморфизм φ поля L над K сохраняет множество $\{\alpha_1, \dots, \alpha_n\}$ корней многочлена f , каким-то образом их переставляя. Так как $L = K(\alpha_1, \dots, \alpha_n)$, то автоморфизм φ однозначно определяется той перестановкой, которую он осуществляет на множестве корней. Тем самым группа $\text{Aut}_K L$ изоморфно вкладывается в S_n .

Пример 1. Как следует из формулы для решения квадратного уравнения, всякое квадратичное расширение поля K характеристики $\neq 2$ имеет вид $K(\sqrt{d})$, где $d \in K \setminus K^2$. Всякое такое расширение является расширением Галуа. Его группа Галуа порождается автоморфизмом $a + b\sqrt{d} \mapsto a - b\sqrt{d}$ ($a, b \in K$).

Пример 2. Всякое конечное поле \mathbb{F}_q , $q = p^n$, является расширением Галуа поля \mathbb{Z}_p . Его группа Галуа — это циклическая группа порядка n , порождаемая автоморфизмом Фробениуса.

Пример 3. Круговое поле $K_n = \mathbb{Q}(e^{2\pi i/n})$ есть поле разложения многочлена $x^n - 1$ над \mathbb{Q} и потому является расширением Галуа поля \mathbb{Q} . Всякий автоморфизм поля K_n индуцирует автоморфизм (циклической) группы C_n корней n -й степени из 1, содержащейся в K_n . Как мы знаем, всякий автоморфизм группы C_n есть возведение в степень k , взаимно простую с n . Тем самым группа $\text{Gal } K_n/\mathbb{Q}$ вкладывается в группу \mathbb{Z}_n^* , порядок которой равен $\varphi(n)$ (где φ — функция Эйлера). На самом деле это вложение является изоморфизмом. Для того чтобы это доказать, достаточно установить, что для любого

простого p , не делящего n , существует автоморфизм поля K_n , индуцирующий на C_n возведение в p -ю степень. Последнее означает, что если f — минимальный многочлен числа $\varepsilon = e^{2\pi i/n}$ над \mathbb{Q} , то $f(\varepsilon^p) = 0$.

Можно считать, что f — примитивный многочлен с целыми коэффициентами. Тогда по лемме Гаусса $x^n - 1 = fg$, где $g \in \mathbb{Z}[x]$. Легко видеть, что многочлен $x^n - 1$ остается сепарабельным при редукции по модулю p . Следовательно, многочлены $[f]_p$ и $[g]_p$ взаимно просты.

Предположим теперь, что $f(\varepsilon^p) \neq 0$. Тогда $g(\varepsilon^p) = 0$ и, следовательно,

$$g(x^p) = f(x)h(x), \quad h \in \mathbb{Z}[x].$$

Редуцируя по модулю p , получаем

$$[g]_p^p = [f]_p [h]_p,$$

что противоречит взаимной простоте многочленов $[f]_p$ и $[g]_p$.

Таким образом, $\dim_{\mathbb{Q}} K_n = \varphi(n)$.

Пример 4. Пусть L — поле разложения неприводимого кубического многочлена f с дискриминантом D над полем K характеристики $\neq 2, 3$ (см. пример 9.5.3). Тогда L — расширение Галуа поля K , и если $D \notin K^2$, то $\dim_K L = 6$ и $\text{Gal } L/K \simeq S_3$. Если же $D \in K^2$, то $\dim_K L = 3$ и $\text{Gal } L/K \simeq A_3$. Последнее означает, что группа Галуа осуществляет только четные перестановки корней многочлена f .

Пример 5. Пусть

$$f = x^n + a_1 x^{n-1} + \dots + a_{n-1} x + a_n$$

— «общий» многочлен степени n , коэффициенты которого рассматриваются как элементы поля $K = k(a_1, \dots, a_n)$ рациональных функций от n независимых переменных над некоторым полем k . Пусть L — поле разложения многочлена f над K и $x_1, \dots, x_n \in L$ — его корни. По формулам Виета $a_k = (-1)^k \sigma_k$, где $\sigma_1, \dots, \sigma_n$ — элементарные симметрические многочлены от x_1, \dots, x_n . Следовательно, $L = k(x_1, \dots, x_n)$. Так как

$$\text{tr. deg } L = \text{tr. deg } K = n,$$

то x_1, \dots, x_n алгебраически независимы над k . В частности, они различны, так что f — сепарабельный многочлен и L — расширение

Галуа поля K . Любая перестановка корней x_1, \dots, x_n определяет автоморфизм поля L , тождественный на K . Следовательно, $\text{Gal } L/K \simeq S_n$. Одновременно мы доказали, что

$$k(x_1, \dots, x_n)^{S_n} = k(\sigma_1, \dots, \sigma_n).$$

§7. Основная теорема теории Галуа

В теории Галуа устанавливается соответствие между подполями расширения Галуа L поля K , содержащими K , и подгруппами группы $G = \text{Gal } L/K$.

А именно, каждому подполю $P \subset L$, содержащему K , ставится в соответствие подгруппа

$$G_P = \{g \in G : g|_P = \text{id}\} \subset G$$

и каждой подгруппе $H \subset G$ ставится в соответствие подполе

$$L^H = \{\alpha \in L : h\alpha = \alpha \ \forall h \in H\} \subset L.$$

Теорема 6.1 показывает, что

$$|G_P| = \dim_P L, \quad (16)$$

$$\dim_{L^H} L = |H|. \quad (17)$$

Теорема 1 (основная теорема теории Галуа). *Описанные выше отображения $P \mapsto G_P$ и $H \mapsto L^H$ взаимно обратны и, таким образом, устанавливают взаимно однозначное соответствие между множествами подполей поля L , содержащих K , и подгрупп группы G . При этом подполям P , являющимся расширениями Галуа поля K , соответствуют нормальные подгруппы H группы G , и наоборот.*

Доказательство. Очевидно, что

$$L^{G_P} \supset P.$$

В то же время из (16) и (17) вытекает, что

$$\dim_{L^{G_P}} L = |G_P| = \dim_P L.$$

Следовательно,

$$L^{G_P} = P.$$

Аналогично доказывается, что

$$G_{L^H} = H.$$

Далее, так как по теореме 6.1 все автоморфизмы подполя P над K продолжаются до автоморфизмов поля L , поле P является расширением Галуа поля K тогда и только тогда, когда преобразования из группы G , переводящие его в себя, индуцируют на нем $\dim_K P$ различных автоморфизмов. Но из формулы (16) следует, что

$$\dim_K P = |G : G_P|.$$

Поэтому P — расширение Галуа тогда и только тогда, когда все преобразования из группы G переводят его в себя.

Так как $P = L^H$, где $H = G_P$, то

$$gP = L^{gHg^{-1}}.$$

Следовательно, подполе P инвариантно относительно всех преобразований из G тогда и только тогда, когда подгруппа H нормальна. \square

Пример 1. Пусть f — неприводимый кубический многочлен над полем K характеристики $\neq 2$ с $D \notin K^2$ (см. пример 9.5.3), и пусть L — его поле разложения над K . Тогда $\text{Gal } L/K \simeq S_3$. Подгруппе $A_3 \subset S_3$ отвечает квадратичное расширение $K(\sqrt{D})$ поля K , содержащееся в L .

Пример 2. Пусть φ — автоморфизм Фробениуса конечного поля \mathbb{F}_{p^n} (p простое). Как мы знаем, $\text{Gal } \mathbb{F}_{p^n}/\mathbb{Z}_p = \langle \varphi \rangle_n$. Всякая подгруппа группы $\langle \varphi \rangle_n$ имеет вид $\langle \varphi^m \rangle_{n/m}$, где $m \mid n$. Соответствующее подполе есть подполе неподвижных точек автоморфизма φ^m . Оно имеет размерность m над \mathbb{Z}_p , т. е. изоморфно \mathbb{F}_{p^m} .

Пример 3. Пусть p — нечетное простое число. Группа Галуа G кругового поля $K_p = \mathbb{Q}(\varepsilon_p)$ (см. примеры 9.5.2 и 6.3) есть циклическая группа порядка $p - 1$. Пусть $H \subset G$ — (единственная) подгруппа индекса 2. Тогда $P = K_p^H$ — квадратичное расширение поля \mathbb{Q} . Докажем, что

$$P = \begin{cases} \mathbb{Q}(\sqrt{p}), & \text{если } p \equiv 1 \pmod{4}, \\ \mathbb{Q}(\sqrt{-p}), & \text{если } p \equiv -1 \pmod{4}. \end{cases}$$

Порождающий элемент группы G действует на группе C_p корней p -й степени из 1 как возведение в некоторую степень r (где $[r]_p$ — порождающий элемент группы \mathbb{Z}_p^*). Рассмотрим число

$$\alpha = \varepsilon_p - \varepsilon_p^r + \varepsilon_p^{r^2} - \dots - \varepsilon_p^{r^{p-2}} = \sum_{k=1}^{p-1} \binom{k}{p} \varepsilon_p^k,$$

где $\left(\frac{k}{p}\right)$ — символ Лежандра (см. пример 9.1.3). Очевидно, что

$$g(\alpha) = \begin{cases} \alpha & \text{при } g \in H, \\ -\alpha & \text{при } g \in G \setminus H. \end{cases}$$

Следовательно, $\alpha \in P$ и $\alpha^2 \in \mathbb{Q}$.

Пользуясь результатами примеров 9.5.5 и 9.1.3, получаем:

$$\alpha^2 = \frac{1}{p-1} \operatorname{tr} \alpha^2 = \frac{1}{p-1} (\alpha, \alpha) = \frac{p}{p-1} \sum_{k=1}^{p-1} \left(\frac{k}{p}\right) \left(\frac{-k}{p}\right) = p \left(\frac{-1}{p}\right) = (-1)^{\frac{p-1}{2}} p,$$

откуда и вытекает доказываемое утверждение.

Теория Галуа была создана в связи с проблемой решения алгебраических уравнений в радикалах.

Будем говорить, что элемент α какого-либо расширения поля K представляется в радикалах над K , если он выражается через элементы поля K при помощи арифметических операций и извлечения корней (любых степеней). Иначе говоря, это означает, что α принадлежит последнему полю в цепочке расширений

$$K = K_0 \subset K_1 \subset \dots \subset K_s,$$

в которой $K_i = K_{i-1}(\alpha_i)$, где $\alpha_i^{n_i} \in K_{i-1}$ для некоторого $n_i \in \mathbb{N}$.

Предложение 1. Если многочлен $f \in K[x]$ неприводим и хотя бы один его корень представляется в радикалах над K , то и все его корни обладают этим свойством.

Доказательство. Пусть α_1 и α_2 — корни многочлена f в каких-то расширениях поля K . Тогда существует изоморфизм поля $K(\alpha_1)$ на поле $K(\alpha_2)$, переводящий α_1 в α_2 . Поэтому, если α_1 представляется в радикалах над K , то и α_2 обладает этим свойством. \square

Алгебраическое уравнение $f(x) = 0$, где $f \in K[x]$, называется разрешимым в радикалах над K , если каждый его корень представляется в радикалах над K . Это равносильно тому, что поле разложения L многочлена f над K содержится в поле, получаемом из K последовательными присоединениями корней из каких-то элементов.

Основной результат Э. Галуа (1830 г.), касающийся разрешимости алгебраических уравнений в радикалах, заключается в следующей теореме.

Теорема 2. Пусть f — неприводимый многочлен над полем K нулевой характеристики и L — его поле разложения над K . Уравнение

$f(x) = 0$ разрешимо в радикалах над K тогда и только тогда, когда группа $\text{Gal } L/K$ разрешима.

Доказательство этой теоремы основано на том, что если P — поле, содержащее n различных корней n -й степени из 1, то его расширения вида $P(\alpha)$, где $\alpha^n = a \in P$ — это то же самое, что расширения Галуа с циклической группой Галуа, порядок которой делит n . Мы приведем ниже полное доказательство более простого варианта теоремы 2, относящегося к разрешимости в квадратных радикалах.

Пример 6.5 вместе с тем фактом, что группа S_n разрешима только при $n \leq 4$, показывает, что общее алгебраическое уравнение степени n над (произвольным) полем k нулевой характеристики разрешимо в радикалах только при $n \leq 4$. Разрешимость в радикалах общего уравнения степени n над k означает возможность выразить единообразно, т. е. одной и той же формулой, корни любого уравнения степени n над k через его коэффициенты и фиксированные элементы поля k при помощи арифметических операций и извлечения корней. Отсутствие такой формулы не означает невозможности решения в радикалах конкретных уравнений. Например, любое алгебраическое уравнение над \mathbb{C} разрешимо в радикалах, так как его корни лежат в \mathbb{C} .

Традиционно наибольший интерес вызывала разрешимость в радикалах алгебраических уравнений над \mathbb{Q} . Теорема 2 позволяет доказать, что для любого $n \geq 5$ существуют такие неприводимые многочлены $f \in \mathbb{Q}[x]$ степени n , что уравнение $f(x) = 0$ не разрешимо в радикалах.

Разрешимость в квадратных радикалах определяется совершенно так же, как разрешимость в радикалах, с той разницей, что разрешается извлечение только квадратных корней.

Теорема 3. Пусть f — неприводимый многочлен над полем K характеристики $\neq 2$ и L — его поле разложения над K . Уравнение $f(x) = 0$ разрешимо в квадратных радикалах над K тогда и только тогда, когда

$$\dim_K L = 2^n \quad (n \in \mathbb{N}). \quad (18)$$

Доказательство. 1) Пусть уравнение $f(x) = 0$ разрешимо в квадратных радикалах. Тогда существует такая цепочка квадратичных расширений

$$K = K_0 \subset K_1 \subset K_2 \subset \dots \subset K_s,$$

что $L \subset K_s$. Имеем

$$\dim_K L \mid \dim_K K_s = 2^s,$$

откуда и следует (18).

2) Обратное, пусть $\dim_K L = 2^n$. Тогда группа $G = \text{Gal } L/K$ есть 2-группа и, следовательно, разрешима. Рассмотрим какой-либо ее композиционный ряд

$$G = G_0 \supset G_1 \supset G_2 \supset \dots \supset G_s = \{e\}.$$

Очевидно, что все его факторы имеют порядок 2. Положив $K_i = L^{G_i}$, мы получим цепочку квадратичных расширений

$$K = K_0 \subset K_1 \subset K_2 \subset \dots \subset K_s = L,$$

которая и доказывает разрешимость уравнения $f(x) = 0$ в квадратных радикалах. \square

Замечание 1. Так как

$$\deg f = \dim_K K(\alpha),$$

где $\alpha \in L$ — какой-либо корень многочлена f , то из (18) следует, что $\deg f$ есть степень двойки. Обратное неверно.

Замечание 2. Во второй части доказательства мы пользовались тем, что L — расширение Галуа поля K . Это, конечно, верно, если $\text{char } K = 0$. Если $\text{char } K = p > 2$, то это вытекает из того, что многочлен f сепарабелен, так как его степень, будучи степенью двойки, не делится на p .

Разрешимость уравнений в квадратных радикалах вызывала интерес в связи с задачами на построение циркулем и линейкой.

Всякая задача на построение циркулем и линейкой может быть сформулирована следующим образом: дана единица измерения и даны отрезки длин a_1, \dots, a_k ; требуется построить отрезок длины α . Анализируя возможные элементарные шаги построения, можно доказать, что сформулированная выше задача может быть решена тогда и только тогда, когда число α представляется в квадратных радикалах над полем $K = \mathbb{Q}(a_1, \dots, a_k)$.

Замечание 3. Когда мы говорим о представлении вещественного числа в квадратных радикалах над полем $K \subset \mathbb{R}$, то исходное определение не исключает извлечения квадратных корней из отрицательных чисел, что выводит нас в комплексную область. Однако задание комплексного числа равносильно заданию его вещественной и мнимой частей, а арифметические

операции над комплексными числами и извлечение квадратных корней из них сводятся к арифметическим операциям над вещественными числами и извлечению квадратных корней из положительных чисел. Все эти операции выполнимы циркулем и линейкой.

В частности, если α трансцендентно над K , то задача неразрешима. Таким образом доказывается неразрешимость квадратуры круга (если радиус круга выбрать за единицу измерения, то задача равносильна построению отрезка длины π).

Если α алгебраично над K и $f \in K[x]$ — его минимальный многочлен, то в силу теоремы 3 задача разрешима тогда и только тогда, когда степень поля разложения многочлена f над K является степенью двойки. В частности, для этого необходимо, чтобы степень самого многочлена f была степенью двойки.

Пример 4. Удвоение куба сводится к построению отрезка длины $\sqrt[3]{2}$. Так как многочлен $x^3 - 2$ неприводим над \mathbb{Q} и его степень не есть степень двойки, то эта задача неразрешима.

Пример 5. Трисекция угла, равного φ , сводится к построению отрезка длины $\cos \frac{\varphi}{3}$ по отрезку длины $\cos \varphi$. По известной формуле

$$\cos \varphi = 4 \cos^3 \frac{\varphi}{3} - 3 \cos \frac{\varphi}{3},$$

так что число $\alpha = \cos \frac{\varphi}{3}$ является корнем многочлена

$$f = 4x^3 - 3x - \cos \varphi \in K[x],$$

где $K = \mathbb{Q}(\cos \varphi)$. Если речь идет об универсальном методе трисекции угла, не зависящем от величины угла φ , то мы должны рассматривать $\cos \varphi$ как независимую переменную. Тогда многочлен f неприводим над K (проверьте это!), и задача неразрешима по той же причине, что в предыдущем примере. Для конкретных углов (например, для прямого) задача, конечно, может быть разрешимой, но можно указать такие значения φ , для которых она неразрешима. Критерием разрешимости является наличие у многочлена f корня в поле K . Если, например, $\varphi = \frac{\pi}{3}$, то $K = \mathbb{Q}$ и $f = 4x^3 - 3x - \frac{1}{2}$ не имеет корней в K , так что задача неразрешима.

Пример 6. Деление окружности на n равных частей (циклотомия) сводится к построению отрезка длины $\cos \frac{2\pi}{n}$ или, что равносильно, числа $e^{2\pi i/n} = \cos \frac{2\pi}{n} + i \sin \frac{2\pi}{n}$. Поэтому циклотомия воз-

можно тогда и только тогда, когда степень кругового поля $K_n = \mathbb{Q}(e^{2\pi i/n})$ является степенью двойки. Как известно, она равна $\varphi(n)$ (см. пример 6.3). Если n — простое число, то $\varphi(n) = n - 1$, так что должно быть $n = 2^m + 1$. Легко видеть, что число $2^m + 1$ может быть простым только тогда, когда m есть степень двойки. Таким образом, число n должно иметь вид

$$n = 2^{2^k} + 1.$$

Такие числа называются *числами Ферма*. При $k = 0, 1, 2, 3, 4$ получаем простые числа

$$3, 5, 17, 257, 65537,$$

но при $k = 5$ получается уже не простое число. В настоящее время не известно ни одного простого числа Ферма, кроме перечисленных выше.

Теория Галуа позволяет дать концептуальное доказательство основной теоремы алгебры комплексных чисел, использующее только следующие два свойства полей \mathbb{R} и \mathbb{C} :

- 1) любой многочлен нечетной степени над полем \mathbb{R} имеет корень в \mathbb{R} ;
- 2) в поле \mathbb{C} возможно извлечение квадратного корня из любого элемента.

Оба эти свойства легко доказываются без привлечения основной теоремы (см. §3.4 и §1.4 соответственно).

Из свойства 1) следует, что над полем \mathbb{R} не существует неприводимых многочленов нечетной степени, отличной от единицы, и, значит, не существует нетривиальных конечных расширений нечетной степени (так как минимальный многочлен любого элемента такого расширения должен был бы иметь нечетную степень).

Пусть $f \in \mathbb{C}[x]$. Обозначим через \bar{f} многочлен, получаемый из f заменой всех коэффициентов комплексно сопряженными числами. Тогда $\overline{f\bar{f}} = \bar{f}f = f\bar{f}$ и, следовательно, $f\bar{f} \in \mathbb{R}[x]$. С другой стороны, если $c \in \mathbb{C}$ — корень многочлена $f\bar{f}$, то c или \bar{c} — корень многочлена f . Поэтому нам достаточно доказать, что всякий многочлен положительной степени с вещественными коэффициентами имеет корень в \mathbb{C} .

Пусть $f \in \mathbb{R}[x]$ — многочлен положительной степени, $L \supset \mathbb{R}$ — его поле разложения над \mathbb{R} и $G = \text{Gal } L/\mathbb{R}$. Пусть H — какая-либо силовская 2-подгруппа группы G . Рассмотрим поле $L^H = K$. Так как $\dim_{\mathbb{R}} K = |G : H|$ — нечетное число, то в силу предыдущего $G = H$, т. е. G — 2-группа. Но тогда по теореме 3 поле L содержится в поле, получаемом из \mathbb{R} последовательными присоединениями квадратных радикалов, и из сформулированного выше свойства 2) следует, что $L = \mathbb{R}$ или \mathbb{C} . Таким образом, f имеет корень в \mathbb{C} .

Линейные представления и ассоциативные алгебры

В приложениях теории групп важнейшую роль играют их линейные представления. Имеются следующие два основных источника линейных представлений групп:

1) для любой группы G дифференцируемых преобразований, оставляющих на месте некоторую точку, взятие дифференциала в этой точке есть линейное представление группы G ;

2) любое действие группы на множестве X определяет по формуле (7) § 10.3 ее линейное представление в пространстве функций на X .

С другой стороны, алгебра матриц, благодаря своему богатству и высокой эффективности производимых в ней вычислений, служит эталоном при изучении многих алгебраических структур. Сравнение с ней осуществляется при помощи линейных представлений.

§ 1. Инвариантные подпространства

Для всякого векторного пространства V над каким-то полем K мы будем обозначать через $L(V)$ (ассоциативную) алгебру всех линейных операторов на V . Если пространство V конечномерно, то линейные операторы можно задавать матрицами в каком-либо базисе, и таким образом устанавливается изоморфизм алгебры $L(V)$ и алгебры матриц $L_n(K)$ (где $n = \dim V$).

Определение 1. *Линейным представлением* множества X в векторном пространстве V называется любое отображение

$$R: X \rightarrow L(V) \tag{1}$$

Пространство V называется *пространством представления*, его размерность — *размерностью представления*, а операторы $R(x)$, $x \in X$, — *операторами представления*.

Если в множестве X определены какие-либо операции, то естественно потребовать, чтобы представление было согласовано с

ними. Так, *линейное представление группы* определяется требованиями

$$R(xy) = R(x)R(y), \quad R(e) = \mathcal{E}$$

(и тем самым может быть определено как гомоморфизм в группу $GL(V)$), а *линейное представление ассоциативной алгебры* — требованиями

$$\begin{aligned} R(x + y) &= R(x) + R(y), & R(xy) &= R(x)R(y), \\ R(\lambda x) &= \lambda R(x), & \lambda &\in K. \end{aligned}$$

Мы, однако, сначала займемся теми свойствами линейных представлений, которые имеют смысл безотносительно к каким-либо операциям в множестве X .

Определение 2. Пусть $R: X \rightarrow L(V)$ и $S: X \rightarrow L(U)$ — два линейных представления одного и того же множества X над одним и тем же полем. *Морфизмом* представления R в представление S называется любое линейное отображение $\varphi: V \rightarrow U$ со следующим свойством: для любого $x \in X$ диаграмма

$$\begin{array}{ccc} V & \xrightarrow{R(x)} & V \\ \varphi \downarrow & & \downarrow \varphi \\ U & \xrightarrow{S(x)} & U \end{array}$$

коммутативна. Обратимый морфизм называется *изоморфизмом* представлений.

Линейные представления R и S называются *изоморфными*, если существует изоморфизм представления R в представление S . В этом случае пишут $R \simeq S$. Изоморфные представления в подходящих базах пространств V и U записываются одними и теми же матрицами.

Пример 1. Линейное представление одноэлементного множества — это просто один линейный оператор в каком-то векторном пространстве. Два линейных представления одноэлементного множества над алгебраически замкнутым полем изоморфны тогда и только тогда, когда матрицы соответствующих линейных операторов приводятся к одной и той же жордановой форме. Таким образом, в этом случае классы изоморфных представлений параметризуются жордановыми матрицами.

Замечание 1. Задача описания линейных представлений двухэлементного множества, не говоря уже о более мощных множествах, считается «дикой». Это означает, что по современным представлениям она не может быть решена ни в каком разумном смысле. Однако реально интересны лишь представления множеств с теми или иными операциями (в первую очередь, групп), и в этой ситуации сложность описания представлений зависит вовсе не от количества элементов.

Всякое линейное представление $R: X \rightarrow L(V)$ над полем K можно рассматривать как линейное представление над любым расширением L поля K , продолжив операторы представления до линейных операторов в пространстве $V(L)$ (см. § 8.1). В базисе пространства $V(L)$, составленном из векторов пространства V , продолженное таким образом представление будет задаваться такими же матрицами, что и исходное представление.

Предложение 1. Пусть $R: X \rightarrow L(V)$ и $S: X \rightarrow L(W)$ — линейные представления множества X над бесконечным полем K , и пусть L — какое-либо расширение поля K . Если представления R и S изоморфны над L , то они изоморфны и над K .

Доказательство. Запишем представления R и S матрицами в каких-нибудь базисах пространств V и W . Изоморфность представлений R и S над K означает существование такой невырожденной матрицы C с элементами из K , что

$$CR(x) = S(x)C \quad \forall x \in X. \quad (2)$$

Соотношения (2) представляют собой систему однородных линейных уравнений относительно элементов матрицы C с коэффициентами из поля K . Пусть $\{C_1, \dots, C_m\}$ — фундаментальная система ее решений. Если представления R и S не изоморфны над K , то $\det(\lambda_1 C_1 + \dots + \lambda_m C_m) = 0$ при любых $\lambda_1, \dots, \lambda_m \in K$ и, следовательно, $\det(t_1 C_1 + \dots + t_m C_m)$ есть нулевой многочлен от t_1, \dots, t_m . Но тогда $\det(\lambda_1 C_1 + \dots + \lambda_m C_m) = 0$ при любых $\lambda_1, \dots, \lambda_m \in L$, а это означает, что представления R и S не изоморфны над L . \square

Для понимания структуры линейных представлений важную роль играют инвариантные подпространства.

Пусть задано представление $R: X \rightarrow L(V)$. Подпространство $U \subset V$ называется *инвариантным* относительно представления R , если оно инвариантно относительно всех операторов этого представления. Очевидно, что сумма и пересечение инвариантных подпространств являются инвариантными подпространствами.

С каждым инвариантным подпространством $U \subset V$ связаны два новых представления множества X : *подпредставление*

$$R_U: X \rightarrow L(U), \quad R_U(x) = R(x)|_U,$$

и *факторпредставление*

$$R_{V/U}: X \rightarrow L(V/U), \quad R_{V/U}(x)(v + U) = R(x)v + U.$$

Представление R_U (соответственно $R_{V/U}$) однозначно определяется тем свойством, что тождественное вложение $U \rightarrow V$ (соответственно каноническое отображение $V \rightarrow V/U$) является морфизмом представлений.

В матричной форме все это выглядит следующим образом: если базис пространства V выбран так, что какое-то число первых базисных векторов составляет базис инвариантного подпространства U , то

$$R(x) = \begin{pmatrix} R_U(x) & * \\ 0 & R_{V/U}(x) \end{pmatrix}.$$

Определение 3. Линейное представление (1) называется *неприводимым*, если $V \neq 0$ и не существует нетривиальных подпространств $U \subset V$, инвариантных относительно R .

Очевидно, что всякое одномерное представление неприводимо.

Пример 2. Представление Π аддитивной группы \mathbb{R} поворотами евклидовой плоскости E^2 , задаваемое в ортонормированном базисе $\{e_1, e_2\}$ формулой

$$\Pi(t) = \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix},$$

неприводимо, так как никакое одномерное подпространство не переходит в себя при всех поворотах. Однако если рассматривать это представление как комплексное, то оно будет приводимо. Более точно, одномерные подпространства, натянутые на векторы $e_1 - ie_2$ и $e_1 + ie_2$, будут инвариантны, и в базисе, составленном из этих векторов, представление будет иметь вид

$$\Pi(t) = \begin{pmatrix} e^{it} & 0 \\ 0 & e^{-it} \end{pmatrix}$$

(см. пример 6.2.3).

Пример 3. Изоморфизм $S_4 \xrightarrow{\sim} \text{Sym}_+ K$ (см. пример 4.6.19) определяет линейное представление группы S_4 в пространстве E^3 . Докажем, что оно неприводимо. Так как ортогональное дополнение

к инвариантному подпространству также инвариантно, то достаточно доказать, что нет одномерных инвариантных подпространств, но это очевидно. На самом деле указанное представление неприводимо не только над \mathbb{R} , но и над \mathbb{C} . Это вытекает из следующего общего предложения.

Предложение 2. Пусть $R: X \rightarrow L(V)$ — неприводимое нечетномерное вещественное линейное представление множества X . Тогда комплексификация представления R также неприводима.

Доказательство. Предположим, что $W \subset V(\mathbb{C})$ — нетривиальное инвариантное подпространство. Заметим, что если какое-либо подпространство пространства $V(\mathbb{C})$ инвариантно относительно комплексного сопряжения, то оно вместе с любым вектором содержит его вещественную и мнимую части и, значит, является комплексификацией некоторого подпространства пространства V . Отсюда следует, что инвариантные подпространства $W \cap \bar{W}$ и $W + \bar{W}$ являются комплексификациями каких-то инвариантных подпространств пространства V . В силу неприводимости представления R должно быть

$$W \cap \bar{W} = 0, \quad W + \bar{W} = V(\mathbb{C}),$$

т. е. $V(\mathbb{C}) = W \oplus \bar{W}$; но тогда $\dim V(\mathbb{C}) = 2 \dim W$, что противоречит нечетномерности пространства V . \square

Пример 4. Пусть V есть n -мерное векторное пространство с базисом $\{e_1, \dots, e_n\}$. Линейное представление M группы S_n , определяемое формулами

$$M(\sigma)e_i = e_{\sigma(i)} \quad (\sigma \in S_n),$$

называется *мономиальным представлением*. Это представление приводимо: можно указать по меньшей мере два нетривиальных инвариантных подпространства: одномерное подпространство $\langle e_1 + \dots + e_n \rangle$ и $(n-1)$ -мерное подпространство

$$V_0 = \left\{ \sum_i x_i e_i : \sum_i x_i = 0 \right\}.$$

Докажем, что если $\text{char } K = 0$, то представление $M_0 = M_{V_0}$ неприводимо. В самом деле, пусть $U \subset V_0$ — инвариантное подпространство и $x = \sum_i x_i e_i \in U$ — ненулевой вектор. Так как $\sum_i x_i = 0$, то не все числа x_1, \dots, x_n равны между собой. Пусть для определенности $x_1 \neq x_2$. Тогда

$$x - M((12))x = (x_1 - x_2)(e_1 - e_2) \in U,$$

откуда $e_1 - e_2 \in U$. Применяя к $e_1 - e_2$ операторы представления, мы получаем, что $e_i - e_j \in U$ при всех i, j ; но тогда $U = V_0$.

Пример 5. Пусть A — ассоциативная алгебра. Тогда формула

$$T(a)x = ax \quad (a, x \in A)$$

определяет линейное представление T алгебры A в своем собственном пространстве, называемое ее (*левым*) *регулярным представлением*. Подчеркнем, что это именно представление алгебры, т. е.

$$T(a+b) = T(a) + T(b), \quad T(ab) = T(a)T(b), \quad T(\lambda a) = \lambda T(a).$$

Например, второе из этих свойств эквивалентно ассоциативности умножения в A . Инвариантные подпространства для этого представления суть не что иное, как левые идеалы алгебры A .

Если $\varphi: V \rightarrow U$ — морфизм представления $R: X \rightarrow L(V)$ в представлении $S: X \rightarrow L(U)$, то $\text{Im } \varphi$ есть инвариантное подпространство в U , а $\text{Ker } \varphi$ есть инвариантное подпространство в V . Поэтому справедлива

Теорема 1. *Всякий морфизм неприводимых представлений есть либо изоморфизм, либо нулевое отображение.*

В дальнейшем мы будем без специальных оговорок предполагать, что все рассматриваемые линейные представления конечномерны.

Теорема 2 (лемма Шура). *Всякий эндоморфизм (т. е. морфизм в себя) неприводимого представления над алгебраически замкнутым полем скалярен.*

Доказательство. Пусть $R: X \rightarrow L(V)$ — данное представление. Линейный оператор $\varphi \in L(V)$ является его эндоморфизмом, если он перестановочен со всеми операторами представления. Следовательно, если φ — эндоморфизм представления R , то таковым является и $\varphi - \lambda \mathcal{E}$ при любом $\lambda \in K$. Выбрав в качестве λ какое-нибудь собственное значение оператора φ , мы получим по теореме 1, что $\varphi - \lambda \mathcal{E} = 0$. \square

Следствие 1. *Пусть $R: X \rightarrow L(V)$ и $S: X \rightarrow L(U)$ — два неприводимых представления множества X над алгебраически замкнутым полем. Тогда любые два морфизма представления R в представление S пропорциональны.*

Доказательство. Если один из морфизмов равен нулю, то доказывать нечего. Поэтому мы должны только доказать пропорциональность любых двух изоморфизмов. Но если $\varphi: V \rightarrow U$ и $\psi: V \rightarrow U$ —

два изоморфизма представления R в представление S , то $\psi^{-1}\varphi: V \rightarrow V$ есть автоморфизм представления R . По лемме Шура $\psi^{-1}\varphi = \lambda\mathcal{E}$, откуда $\varphi = \lambda\psi$. \square

Следствие 2. *Всякое неприводимое представление абелевой группы над алгебраически замкнутым полем одномерно.*

Доказательство. В случае абелевой группы все операторы представления перестановочны между собой и, значит, каждый из них является эндоморфизмом представления. По лемме Шура все они скалярны. Следовательно, любое подпространство инвариантно, и представление может быть неприводимым, только если оно одномерно. \square

Определение 4. Линейное представление $R: X \rightarrow L(V)$ называется *вполне приводимым*, если для любого инвариантного подпространства $U \subset V$ имеется инвариантное дополнительное подпространство, т. е. такое инвариантное подпространство $W \subset V$, что $V = U \oplus W$.

Отметим, что всякое неприводимое представление, как это ни странно звучит, вполне приводимо.

Пример 6. Формула

$$R(t) = \begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix} \quad (t \in \mathbb{R})$$

задает двумерное вещественное линейное представление аддитивной группы \mathbb{R} . Нетривиальные инвариантные подпространства — это только одномерные подпространства, натянутые на базисные векторы (координатные оси). Так как эти подпространства дополнители друг к другу, то представление R вполне приводимо.

Пример 7. Формула

$$S(t) = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$$

задает другое линейное представление той же группы. В этом случае единственное нетривиальное инвариантное подпространство — это одномерное подпространство, натянутое на первый базисный вектор. Поэтому представление S не является вполне приводимым.

Предложение 3. *Всякое подпредставление и всякое факторпредставление вполне приводимого представления вполне приводимы.*

Доказательство. Пусть $R: X \rightarrow L(V)$ — вполне приводимое представление и $U \subset V$ — инвариантное подпространство. Для любого инвариантного подпространства $U_1 \subset U$ в пространстве V существу-

ет инвариантное дополнительное подпространство, скажем, V_2 . Подпространство $U_2 = U \cap V_2$ будет тогда инвариантным подпространством, дополнительным к U_1 в U .

Пусть теперь $\pi: V \rightarrow V/U$ — каноническое отображение и $W_1 \subset V/U$ — инвариантное подпространство. Тогда $V_1 = \pi^{-1}(W_1)$ — инвариантное подпространство в V (содержащее U). Если $V_2 \subset V$ — дополнительное к нему инвариантное подпространство, то $W_2 = \pi(V_2)$ будет инвариантным подпространством, дополнительным к W_1 в V/U . \square

Дадим теперь другую характеристику вполне приводимых представлений.

Теорема 3. 1) Если представление $R: X \rightarrow L(V)$ вполне приводимо, то пространство V может быть разложено в прямую сумму минимальных инвариантных подпространств.

2) Обратно, если пространство V может быть разложено в сумму (не обязательно прямую) минимальных инвариантных подпространств V_1, \dots, V_m , то представление R вполне приводимо, причем для любого инвариантного подпространства $U \subset V$ в качестве инвариантного дополнительного подпространства может быть взята сумма некоторых из V_i .

(Минимальным инвариантным подпространством здесь называется подпространство, минимальное среди ненулевых инвариантных подпространств.)

Доказательство. 1) Берем любое минимальное инвариантное подпространство V_1 , находим инвариантное дополнительное подпространство, в нем берем любое минимальное инвариантное подпространство V_2 и т. д.

2) Пусть $U \subset V$ — инвариантное подпространство. Для любого подмножества $I \subset \{1, \dots, n\}$ положим $V_I = \sum_{i \in I} V_i$. Пусть I — максимальное подмножество (быть может, пустое), для которого $U \cap V_I = 0$. Тогда для любого $j \notin I$ должно быть $U \cap V_{I \cup \{j\}} \neq 0$ и, значит,

$$(U \oplus V_I) \cap V_j \neq 0.$$

Так как V_j — минимальное инвариантное подпространство, то $V_j \subset U \oplus V_I$. Следовательно,

$$V = U \oplus V_I. \quad \square$$

Пример 8. Если $\text{char } K = 0$, то мономиальное представление группы S_n , определенное в примере 4, вполне приводимо, так как

пространство V в этом случае разлагается в прямую сумму минимальных инвариантных подпространств:

$$V = \langle e_1 + \dots + e_n \rangle \oplus V_0.$$

Задача 1. Рассмотрим линейное представление Ad группы $\text{GL}_n(K)$ в пространстве $L_n(K)$, задаваемое формулой

$$\text{Ad}(A)X = AXA^{-1}.$$

Доказать, что если $\text{char } K = 0$, то $\langle E \rangle$ и $\langle X \in L_n(K) : \text{tr } X = 0 \rangle$ — минимальные инвариантные подпространства, и вывести отсюда, что представление Ad вполне приводимо.

Определение 5. Суммой линейных представлений $R_i: X \rightarrow L(V_i)$, $i = 1, \dots, m$, называется линейное представление

$$R = R_1 + \dots + R_m: X \rightarrow L(V_1 \oplus \dots \oplus V_m),$$

определяемое по формуле

$$R(x)(v_1, \dots, v_m) = (R_1(x)v_1, \dots, R_m(x)v_m).$$

В матричной записи

$$R(x) = \begin{pmatrix} R_1(x) & 0 & \dots & 0 \\ 0 & R_2(x) & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & R_m(x) \end{pmatrix}.$$

Если $R: X \rightarrow L(V)$ — какое-то линейное представление и пространство V разложено в прямую сумму инвариантных подпространств V_1, \dots, V_m , то $R \simeq R_1 + \dots + R_m$, где $R_i = R|_{V_i}$.

Теорема 3 дает следующую характеристику вполне приводимых представлений.

Следствие 1. *Линейное представление вполне приводимо тогда и только тогда, когда оно изоморфно сумме неприводимых представлений.*

Следствие 2. *Пусть $R: X \rightarrow L(V)$ — вполне приводимое представление, изоморфное сумме неприводимых представлений R_1, \dots, R_m . Тогда всякое подпредставление, а также всякое факторпредставление представления R изоморфны сумме некоторых из представлений R_i .*

Доказательство. Пусть

$$V = V_1 \oplus \dots \oplus V_m$$

— такое разложение в прямую сумму инвариантных подпространств, что $R_{V_i} \simeq R_i$, и пусть $U \subset V$ — какое-то инвариантное подпространство. По теореме 3 существует такое подмножество $I \subset \{1, \dots, m\}$, что $V = U \oplus V_I$. Ясно, что

$$R_{V/U} \simeq R_{V_I} \simeq \sum_{i \in I} R_i.$$

Далее, положим $J = \{1, \dots, m\} \setminus I$. Тогда $V = V_J \oplus V_I$ и, следовательно,

$$R_U \simeq R_{V/V_I} \simeq R_{V_J} \simeq \sum_{j \in J} R_j. \quad \square$$

Пример 9. Всякое неприводимое линейное представление одноэлементного множества X над алгебраически замкнутым полем K одномерно. Следовательно, всякое вполне приводимое представление множества X над полем K задается линейным оператором, матрица которого в некотором базисе диагональна.

Пусть $R: X \rightarrow L(V)$ — вполне приводимое представление и

$$V = V_1 \oplus \dots \oplus V_m \quad (3)$$

— какое-то разложение пространства V в прямую сумму минимальных инвариантных подпространств.

Определение 6. *Изотипной компонентой* представления R , отвечающей неприводимому представлению S множества X , называется сумма $V_{(S)}$ тех слагаемых V_i разложения (3), для которых $R_{V_i} \simeq S$, а также ограничение $R_{(S)}$ представления R на эту сумму.

Всякое (минимальное) инвариантное подпространство $U \subset V$, для которого $R_U \simeq S$, может нетривиальным образом проектироваться только на те слагаемые V_i разложения (3), для которых $R_{V_i} \simeq S$, и потому содержится в $V_{(S)}$. Это показывает, что изотипные компоненты не зависят от выбора разложения (3).

Из определения ясно, что пространство V разлагается в прямую сумму изотипных компонент, отвечающих различным неприводимым представлениям множества X .

Пример 10. Для вполне приводимого представления одноэлементного множества над алгебраически замкнутым полем изотипные компоненты — это собственные подпространства соответствующего линейного оператора.

Представление R называется *изотипным* или, точнее, *S -изотипным*, если $R = R_{(S)}$.

Изотипные представления удобно описывать следующим образом. Пусть $S: X \rightarrow L(U)$ — неприводимое представление и Z — любое векторное пространство. Определим представление

$$R: X \rightarrow L(U \otimes Z) \quad (4)$$

по формуле

$$R(x)(u \otimes z) = (S(x)u) \otimes z.$$

Если $\{z_1, \dots, z_m\}$ — базис пространства Z , то разложение

$$U \otimes Z = (U \otimes z_1) \oplus \dots \oplus (U \otimes z_m) \quad (5)$$

является разложением пространства $U \otimes Z$ в прямую сумму инвариантных подпространств, причем ограничение представления R на каждое из них изоморфно S .

Теорема 4. Если основное поле K алгебраически замкнуто, то всякое инвариантное подпространство пространства $U \otimes Z$ имеет вид $U \otimes Z_0$, где Z_0 — некоторое подпространство пространства Z .

Доказательство. Так как сумма подпространств вида $U \otimes Z_0$ есть подпространство того же вида, то достаточно доказать теорему для минимальных инвариантных подпространств.

Пусть $W \subset U \otimes Z$ — минимальное инвариантное подпространство. Для любого $w \in W$ в соответствии с разложением (5) имеем

$$w = \varphi_1(w) \otimes z_1 + \dots + \varphi_m(w) \otimes z_m,$$

где $\varphi_1, \dots, \varphi_m$ — некоторые морфизмы представления R_W в представление S . Согласно следствию 1 теоремы 2, $\varphi_i = \lambda_i \varphi$, где $\lambda_i \in K$, а φ — некоторый фиксированный изоморфизм представления R_W в представление S . Таким образом,

$$w = \varphi(w) \otimes (\lambda_1 z_1 + \dots + \lambda_m z_m)$$

и, стало быть,

$$W = U \otimes \langle \lambda_1 z_1 + \dots + \lambda_m z_m \rangle. \quad \square$$

Задача 2. Доказать, что, если поле K алгебраически замкнуто, всякий эндоморфизм представления (4) имеет вид

$$u \otimes z \mapsto u \otimes \mathcal{C}z,$$

где \mathcal{C} — некоторый линейный оператор в пространстве Z .

Теорема 5 (теорема Бернсайда). Пусть $R: X \rightarrow L(V)$ — неприводимое представление множества X над алгебраически замкнутым полем. Тогда подалгебра алгебры $L(V)$, порожденная множеством $R(X)$, совпадает с $L(V)$ за исключением тривиального случая, когда $\dim V = 1$ и $R(X) = 0$.

Заметим, что подалгебра, порожденная множеством $R(X)$, состоит из всевозможных линейных комбинаций произведений операторов $R(x)$, $x \in X$. Таким образом, теорема утверждает, что, за исключением упомянутого тривиального случая, всякий линейный оператор является линейной комбинацией произведений операторов $R(x)$, $x \in X$.

Доказательство. Как мы знаем, пространство $L(V)$ может быть отождествлено с $V \otimes V^*$ таким образом, что всякому разложимому элементу $u \otimes \alpha \in V \otimes V^*$ соответствует оператор

$$u \otimes \alpha: v \mapsto \alpha(v)u.$$

При этом соглашения произведения оператора $u \otimes \alpha$ на любой линейный оператор $\mathcal{A} \in L(V)$ выглядят следующим образом:

$$\mathcal{A}(u \otimes \alpha) = \mathcal{A}u \otimes \alpha, \quad (6)$$

$$(u \otimes \alpha).\mathcal{A} = u \otimes \mathcal{A}^*\alpha, \quad (7)$$

где \mathcal{A}^* — сопряженный оператор, определяемый по формуле

$$(\mathcal{A}^*\alpha)(v) = \alpha(\mathcal{A}v).$$

Заметим, что ввиду канонического взаимно однозначного соответствия между подпространствами пространства V и подпространствами пространства V^* , сопоставляющего каждому подпространству его аннулятор, представление

$$R^*: X \rightarrow L(V^*),$$

определяемое формулой

$$R^*(x)\alpha = R(x)^*\alpha,$$

неприводимо.

Определим представления T_l и T_r множества X в пространстве $L(V)$ по формулам

$$T_l(x).\mathcal{A} = R(x).\mathcal{A}, \quad T_r(x).\mathcal{A} = \mathcal{A}R(x).$$

Ввиду формул (6) и (7) эти представления изотипны.

Обозначим через A подалгебру алгебры $L(V)$, порожденную множеством $R(X)$. Ясно, что она является подпространством в $L(V)$, инвариантным как относительно представления T_l , так и относительно представления T_r . По теореме 4 она должна представляться в виде $A = V \otimes W_0$, где W_0 — подпространство пространства V^* , и в то же время в виде $A = V_0 \otimes V^*$, где V_0 — подпространство пространства V . Это возможно, только если A есть $L(V)$ или 0 . В последнем случае $\dim V = 1$, так как иначе представление R было бы приводимо. \square

Задача 3. Тензорным произведением представлений групп $R: G \rightarrow GL(V)$ и $S: H \rightarrow GL(W)$ называется представление

$$R \otimes S: G \times H \rightarrow GL(V \otimes W),$$

определяемое формулой

$$(R \otimes S)(g, h) = R(g) \otimes S(h).$$

(См. определение тензорного произведения линейных операторов в § 8.1.) Доказать, что тензорное произведение неприводимых представлений групп G и H над алгебраически замкнутым полем неприводимо.

Рассмотрим теперь класс вполне приводимых линейных представлений, в определенном смысле противоположный классу изотипных представлений. А именно, будем говорить, что вполне приводимое представление имеет простой спектр, если оно является суммой попарно не изоморфных неприводимых представлений или, иначе говоря, если все его (ненулевые) изотипные компоненты неприводимы.

Пример 11. Вполне приводимое представление одноэлементно-го множества над алгебраически замкнутым полем имеет простой спектр тогда и только тогда, когда все корни характеристического многочлена соответствующего линейного оператора являются простыми.

Для представлений с простым спектром инвариантные подпространства и эндоморфизмы описываются особенно просто.

Предложение 4. Пусть $R: X \rightarrow L(V)$ — вполне приводимое представление с простым спектром. Рассмотрим какое-то разложение (3) пространства V в прямую сумму минимальных инвариантных подпространств. Тогда

1) всякое инвариантное подпространство пространства V есть сумма некоторых из слагаемых разложения (3);

2) если основное поле K алгебраически замкнуто, то всякий эндоморфизм φ представления R имеет вид

$$\varphi(x) = \lambda_i x \quad \text{при } x \in V_i \quad (\lambda_1, \dots, \lambda_m \in K). \quad (8)$$

Доказательство. 1) Всякое инвариантное подпространство есть сумма минимальных инвариантных подпространств. Всякое минимальное инвариантное подпространство по определению представления с простым спектром есть изотипная компонента и, значит, совпадает с одним из слагаемых разложения (3).

2) Каждое слагаемое разложения (3), будучи изотипной компонентой, инвариантно относительно эндоморфизма φ , а по лемме Шура φ действует на нем скалярно. \square

Следствие. Для вполне приводимого представления с простым спектром разложение пространства представления в прямую сумму минимальных инвариантных подпространств единственно.

Линейное представление $R: G \rightarrow GL(V)$ группы G (над полем K характеристики $\neq 2$) называется *ортогональным* (соответственно *симплектическим*), если в пространстве V существует невырожденная симметрическая (соответственно кососимметрическая) билинейная функция, инвариантная относительно всех операторов представления.

Задача 4. Доказать, что если $R: G \rightarrow GL(V)$ — неприводимое представление группы G над алгебраически замкнутым полем, то

- а) любая ненулевая инвариантная билинейная функция в пространстве V невырожденна;
- б) любые две такие функции пропорциональны;
- в) любая такая функция либо симметрична, либо кососимметрична.

§ 2. Полная приводимость линейных представлений конечных и компактных групп

Для некоторых классов групп можно доказать априори полную приводимость всех линейных представлений.

Мы начнем с конечных групп, для которых доказательство чисто алгебраическое. Оно основано на простой идее, заключенной в следующей ниже лемме.

Пусть S — конечномерное аффинное пространство над полем K .

Лемма 1 (о неподвижной точке). *Всякая конечная группа G аффинных преобразований пространства S , порядок которой не делится на $\text{char } K$, имеет в S неподвижную точку.*

Доказательство. Неподвижной точкой является центр тяжести орбиты любой точки $p \in S$:

$$\text{center } Gp = \frac{1}{|G|} \sum_{g \in G} gp. \quad \square$$

Пусть теперь V — конечномерное векторное пространство над полем K и $G \subset \text{GL}(V)$ — некоторая группа его линейных преобразований.

Теорема 1. *Если G — конечная группа и ее порядок не делится на $\text{char } K$, то для любого G -инвариантного подпространства $U \subset V$ существует G -инвариантное дополнительное подпространство W .*

Доказательство. Задать подпространство W , дополнительное к U , — это все равно, что задать проектор \mathcal{P} на U параллельно W . Инвариантность подпространства W равносильна тому, что соответствующий проектор \mathcal{P} перестановочен со всеми преобразованиями из G .

Совокупность всех проекторов на U описывается линейными уравнениями

$$\mathcal{P}v \in U \quad \forall v \in V, \quad \mathcal{P}u = u \quad \forall u \in U$$

и, следовательно, представляет собой плоскость в пространстве $L(V)$ всех линейных операторов на V . Обозначим эту плоскость через S .

При линейном действии группы G на $L(V)$ сопряжениями плоскость S переходит в себя и на ней индуцируются какие-то аффинные преобразования. Тем самым мы получаем конечную группу аффинных преобразований плоскости S . Ее неподвижная точка и будет искомым проектором. \square

Следствие. *Всякое линейное представление конечной группы G над полем K , характеристика которого не делит $|G|$, вполне приводимо.*

Доказательство получается применением теоремы к образу группы G при рассматриваемом линейном представлении. \square

Пример 1. Докажем другим способом, что трехмерное линейное представление группы S_4 , построенное в примере 1.3, неприводимо не только над \mathbb{R} , но и над \mathbb{C} . Так как оно во всяком случае вполне приводимо, то достаточно доказать отсутствие одномерных инвари-

антных подпространств, т. е. отсутствие общих собственных векторов у операторов представления. Как мы знаем (предложение 6.2.1), всякому мнимому собственному значению вещественного линейного оператора соответствует двумерное инвариантное вещественное подпространство. Однако для рассматриваемого представления нет двумерных инвариантных вещественных подпространств.

При $K = \mathbb{R}$ или \mathbb{C} разумным обобщением конечных групп являются компактные топологические группы.

Определение 1. *Топологической группой* называется группа G , снабженная хаусдорфовой топологией таким образом, что групповые операции

$$\begin{aligned}\mu: G \times G &\rightarrow G, & (x, y) &\mapsto xy, \\ \iota: G &\rightarrow G, & x &\mapsto x^{-1},\end{aligned}$$

являются непрерывными отображениями. Гомоморфизмом топологических групп называется гомоморфизм групп, являющийся в то же время непрерывным отображением.

Примерами топологических групп могут служить аддитивные и мультипликативные группы полей \mathbb{R} и \mathbb{C} , а также группы невырожденных матриц над этими полями. Всякая группа (например, конечная) может рассматриваться как топологическая, если ее снабдить дискретной топологией.

Всякая подгруппа топологической группы, наделенная индуцированной топологией, является топологической группой. Прямое произведение топологических групп является топологической группой.

Топологическая группа называется *компактной*, если она является компактным топологическим пространством. В частности, все конечные группы компактны. Примерами бесконечных компактных групп могут служить «окружность»

$$\mathbb{T} = \{z \in \mathbb{C}^* : |z| = 1\},$$

ортогональная группа O_n и унитарная группа U_n . Докажем компактность группы O_n . Эта группа выделяется в n^2 -мерном пространстве $L_n(\mathbb{R})$ всех вещественных матриц $X = (x_{ij})$ порядка n уравнениями

$$\sum_k x_{ik}x_{jk} = \delta_{ij}$$

и, следовательно, замкнута в $L_n(\mathbb{R})$. Из выписанных уравнений следует также, что $|x_{ij}| \leq 1$, а потому группа O_n ограничена в $L_n(\mathbb{R})$. Следовательно, она компактна. Аналогично доказывается компактность группы U_n .

Всякая замкнутая подгруппа компактной группы компактна. Прямое произведение компактных групп компактно. Так, прямое произведение n экземпляров окружности \mathbb{T} есть компактная группа, называемая n -мерным тором и обозначаемая \mathbb{T}^n . Образ компактной группы при (непрерывном) гомоморфизме (в частности, линейном представлении) является компактной группой.

Имеются аналоги леммы о неподвижной точке и теоремы 1 для компактных групп. Для их доказательства мы используем понятие центра тяжести выпуклого множества.

Пусть M — непустое ограниченное выпуклое множество в вещественном аффинном пространстве S . Если $\text{aff } M = S$, то определим центр тяжести $\text{center } M$ множества M по формуле

$$\text{center } M = \mu(M)^{-1} \int_M x \mu(dx),$$

где μ — обычная мера в пространстве S , инвариантная относительно параллельных переносов. Мера μ определена с точностью до постоянного множителя, но из вида формулы ясно, что произвол в выборе μ не влияет на результат. Интеграл в правой части может определяться либо покоординатно, либо непосредственно как предел интегральных сумм, представляющих собой (с учетом множителя, стоящего перед интегралом) барицентрические линейные комбинации точек пространства S и поэтому имеющих инвариантный смысл. Первое определение показывает существование интеграла, второе — его независимость от выбора системы координат. В общем случае определим $\text{center } M$ так же, как и выше, но заменив пространство S пространством $\text{aff } M$.

Так как определение центра тяжести дается в терминах аффинной геометрии, то

$$\text{center } \alpha(M) = \alpha(\text{center } M)$$

для любого аффинного преобразования α . В частности, если множество M инвариантно относительно какого-либо аффинного преобразования, то его центр тяжести является неподвижной точкой этого преобразования.

Из определения центра тяжести ясно, что $\text{center } M \in \bar{M}$. На самом деле

$$\text{center } M \in M^\circ,$$

где M° — внутренность множества M относительно пространства $\text{aff } M$. Действительно, для любой аффинно-линейной функции f , неотрицательной на M и не равной тождественно нулю на $\text{aff } M$, имеем

$$f(\text{center } M) = \mu(M)^{-1} \int_M f(x) \mu(dx) > 0.$$

Лемма 2 (о неподвижной точке). Пусть G — компактная группа аффинных преобразований вещественного аффинного пространства S , и пусть $M \subset S$ — непустое выпуклое множество, инвариантное относительно G . Тогда группа G имеет в M неподвижную точку.

Заметим, что в качестве M можно взять все пространство S .

Доказательство. Искомой неподвижной точкой является центр тяжести выпуклой оболочки орбиты любой точки $p \in M$. \square

Теорема 2. Пусть G — компактная группа линейных преобразований векторного пространства V над полем $K = \mathbb{R}$ или \mathbb{C} . Тогда для любого G -инвариантного подпространства $U \subset V$ существует G -инвариантное дополнительное подпространство W .

Доказательство этой теоремы дословно повторяет доказательство теоремы 1. Нужно только заметить, что в случае $K = \mathbb{C}$ плоскость S , составленную проекторами на U , следует рассматривать как вещественное аффинное пространство. \square

Следствие. Всякое вещественное или комплексное линейное представление компактной топологической группы вполне приводимо.

Пример 2. Согласно теореме 2 и следствию 2 теоремы 1.2 всякое (непрерывное) комплексное линейное представление компактной абелевой группы есть сумма одномерных представлений, т. е. в некотором базисе оно записывается диагональными матрицами. В частности, это применимо к конечным абелевым группам и к группе \mathbb{T} .

Имеется другой подход к доказательству полной приводимости линейных представлений компактных групп, представляющий интерес и сам по себе.

Теорема 3. Пусть G — компактная группа линейных преобразований вещественного (соответственно комплексного) векторного пространства V . Тогда в пространстве V существует G -инвариантная положительно определенная квадратичная (соответственно эрмитова) функция.

Доказательство. Совокупность всех положительно определенных квадратичных (соответственно эрмитовых) форм есть G -инва-

риантное выпуклое множество в пространстве всех квадратичных (соответственно эрмитовых) форм. Неподвижная точка группы G в этом множестве и будет искомой формой. \square

Следствие. *Всякая компактная (в частности, конечная) подгруппа группы $GL_n(\mathbb{R})$ (соответственно $GL_n(\mathbb{C})$) сопряжена подгруппе группы O_n (соответственно U_n).*

Теорема 3 дает другой способ доказательства теоремы 2: в качестве инвариантного подпространства, дополнительного к U , можно взять ортогональное дополнение к U в смысле скалярного умножения, определяемого инвариантной квадратичной (соответственно эрмитовой) формой.

§ 3. Конечномерные ассоциативные алгебры

Техника линейных представлений позволяет, прежде всего, достаточно хорошо описать строение конечномерных ассоциативных алгебр.

Пусть A — конечномерная ассоциативная (но не обязательно коммутативная) алгебра над полем K .

Элемент $a \in A$ называется *нильпотентным*, если $a^n = 0$ для некоторого $n \in \mathbb{N}$. Алгебра A называется *нильпотентной*, если все ее элементы nilьпотентны. Всякая подалгебра и всякая факторалгебра nilьпотентной алгебры nilьпотентны. С другой стороны, если идеал I и факторалгебра A/I nilьпотентны, то и алгебра A nilьпотентна.

Пример 1. Алгебра всех nilьтреугольных (треугольных с нулями на диагонали) матриц порядка n nilьпотентна. Более того, произведение любых n элементов этой алгебры равно нулю. Как мы сейчас увидим, аналогичным свойством обладает любая nilьпотентная алгебра.

Теорема 1. *Для любой nilьпотентной алгебры A существует такое $n \in \mathbb{N}$, что произведение любых n элементов алгебры A равно нулю.*

Для любых подпространств $B, C \subset A$ условимся обозначать через BC линейную оболочку всевозможных произведений вида bc ($b \in B, c \in C$). При этом соглашении утверждение теоремы можно записать так: $A^n = 0$ для некоторого $n \in \mathbb{N}$.

Доказательство. Пусть $B \subset A$ — максимальное подпространство, для которого существует такое $n \in \mathbb{N}$, что $B^n = 0$. Предположим, что $B \neq A$, и пусть $a \in A \setminus B$. Так как $aB^n = 0$, то существует такое $k \geq 0$, что $aB^k \notin B$, но $aB^{k+1} \subset B$. Заменив a подходящим элементом из aB^k , мы можем добиться, чтобы

$$aB \subset B. \quad (9)$$

Для некоторого $m \in \mathbb{N}$ имеем:

$$a^m = 0. \quad (10)$$

Положим $C = B \oplus \langle a \rangle$. Из условий (9) и (10) следует, что $C^{nm} = 0$, но это противоречит определению подпространства B . \square

В отличие от коммутативного случая, все нильпотентные элементы произвольной ассоциативной алгебры A не обязаны образовывать идеал (и даже подпространство). Однако если I и J — два нильпотентных идеала, то их сумма

$$I + J \doteq \{x + y : x \in I, y \in J\}.$$

также является нильпотентным идеалом, так как она содержит нильпотентный идеал I , факторалгебра

$$(I + J)/I \simeq J/(I \cap J)$$

по которому также нильпотентна. Поэтому существует наибольший нильпотентный идеал. Он называется *радикалом* алгебры A и обозначается через $\text{rad } A$.

В коммутативном случае он совпадает с радикалом кольца A в смысле § 9.4.

Алгебра A называется *полупростой*, если $\text{rad } A = 0$.

Пример 2. В силу примера 9.4.1 алгебра $K[t]/(h)$ полупроста тогда и только тогда, когда многочлен h не имеет кратных неприводимых множителей.

В случае, когда $\text{char } K = 0$, полупростые алгебры могут быть охарактеризованы с другой точки зрения.

Пусть $T: A \rightarrow L(A)$ — регулярное представление алгебры A (см. пример 1.5).

Определим в A «скалярное умножение» по формуле

$$(a, b) = \text{tr } T(ab) = \text{tr } T(a)T(b). \quad (11)$$

Это симметрическая билинейная функция (может быть, вырожденная). Кроме того, она обладает свойством

$$(ab, c) = (a, bc),$$

вытекающим из ассоциативности умножения в A . (В случае, когда A — поле, это скалярное умножение совпадает с введенным в § 9.5.)

Предложение 1. *Ортогональное дополнение I^\perp к любому идеалу I алгебры A также является идеалом.*

Доказательство. Пусть $x \in I^\perp$, $a \in A$, $y \in I$. Тогда

$$(xa, y) = (x, ay) = 0, \quad (ax, y) = (y, ax) = (ya, x) = 0. \quad \square$$

Предложение 2. *Если $\text{char } K = 0$, то всякий элемент $a \in A$, ортогональный всем своим степеням, нильпотентен.*

Доказательство. Пусть

$$(a^n, a) = \text{tr } T(a)^{n+1} = 0 \quad \forall n \in \mathbb{N}.$$

Возьмем подходящее расширение поля K , в котором характеристический многочлен f оператора $T(a)$ разлагается на линейные множители:

$$f = t^{k_0} \prod_{i=1}^s (t - \lambda_i)^{k_i} \quad (\lambda_1, \dots, \lambda_s \text{ различны и отличны от } 0).$$

Тогда

$$\text{tr } T(a)^{n+1} = \sum_{i=1}^s k_i \lambda_i^{n+1} = 0.$$

Пусть n пробегает значения $1, \dots, s$. Рассмотрим предыдущие равенства как квадратную систему однородных линейных уравнений относительно k_1, \dots, k_s . Ее определитель, который лишь множителем $\lambda_1^2 \dots \lambda_s^2$ отличается от определителя Вандермонда для $\lambda_1, \dots, \lambda_s$, отличен от нуля. Следовательно, $k_1 = \dots = k_s = 0$ в поле K , что невозможно, если $\text{char } K = 0$. Поэтому $s = 0$, а это означает, что оператор $T(a)$ нильпотентен, т. е. $T(a)^m = 0$ для некоторого $m \in \mathbb{N}$. Но тогда

$$a^{m+1} = T(a)^m a = 0. \quad \square$$

Теорема 2. 1) *Если скалярное умножение (11) невырожденно, то алгебра A полупроста.*

2) Обратное, если алгебра A полупроста и $\text{char } K = 0$, то скалярное умножение (11) невырожденно.

Доказательство. 1) Пусть I — нильпотентный идеал алгебры A . Тогда для любого $x \in I$ и любого $a \in A$ элемент ax нильпотентен (так как принадлежит I) и, следовательно,

$$(a, x) = \text{tr } T(ax) = 0.$$

Таким образом, $I \subset A^\perp = 0$.

2) Обратное, если $\text{char } K = 0$, то предложения 1 и 2 показывают, что A^\perp есть нильпотентный идеал. \square

Замечание 1. На самом деле мы доказали больше: если $\text{char } K = 0$, то $\text{rad } A$ совпадает с ядром скалярного умножения (11).

Пример 3. Как следует из формулы (6), регулярное представление алгебры $L(V)$ изотипно. Более точно, оно изоморфно nR , где $n = \dim V$, а R — тавтологическое представление алгебры $L(V)$ в пространстве V (т. е. тождественное отображение $L(V) \rightarrow L(V)$). Следовательно, скалярное умножение (11) в $L(V)$ имеет вид

$$(\mathcal{A}, \mathcal{B}) = n \text{tr } \mathcal{A}\mathcal{B}. \quad (12)$$

Если \mathcal{E}_{ij} — оператор, задаваемый (в каком-либо фиксированном базисе) матричной единицей E_{ij} , то

$$\text{tr } \mathcal{E}_{ij}\mathcal{E}_{kl} = \begin{cases} 1 & \text{при } k=j, l=i, \\ 0 & \text{в остальных случаях.} \end{cases}$$

Отсюда следует, что если $\text{char } K \nmid n$, то скалярное умножение (12) невырожденно и, стало быть, алгебра $L(V)$ полупроста. (Как мы увидим (см. пример 6), она полупроста и в том случае, когда $\text{char } K \mid n$.)

Пример 4. Пусть $A = K[t]/(h)$. Мы можем рассмотреть A как $K[t]$ -модуль. Тогда h будет характеристическим многочленом оператора умножения на t (см. задачу 9.3.1). Пусть c_1, \dots, c_n — его корни (с учетом кратностей) в поле разложения. Тогда для любого многочлена $f \in K[t]$ корнями характеристического многочлена оператора умножения на $f(t)$ будут $f(c_1), \dots, f(c_n)$. Но оператор умножения на $f(t)$ в $K[t]$ -модуле A — это то же, что оператор умножения на $[f] = f + (h)$ в алгебре A , т. е. оператор $T([f])$. Следовательно,

$$\text{tr } T([f]) = \sum_i f(c_i).$$

Поэтому в базисе $\{[1], [t], [t^2], \dots, [t^{n-1}]\}$ алгебры A матрица скалярного умножения (11) имеет вид

$$\begin{pmatrix} s_0 & s_1 & s_2 & \dots & s_{n-1} \\ s_1 & s_2 & s_3 & \dots & s_n \\ s_2 & s_3 & s_4 & \dots & s_{n+1} \\ \dots & \dots & \dots & \dots & \dots \\ s_{n-1} & s_n & s_{n+1} & \dots & s_{2n-2} \end{pmatrix}, \quad (13)$$

где $s_k = c_1^k + \dots + c_n^k$. Заметим, что степенные суммы s_k можно выразить через коэффициенты многочлена h без того, чтобы находить его корни.

Согласно теореме 2, если $\text{char } K = 0$, алгебра A полупроста тогда и только тогда, когда матрица (13) невырождена. С другой стороны (см. пример 2), она полупроста тогда и только тогда, когда многочлен h не имеет кратных неприводимых множителей, что в случае $\text{char } K = 0$ равносильно тому, что c_1, \dots, c_n различны. Таким образом, мы приходим к следующему выводу: многочлен h над полем нулевой характеристики сепарабелен (т. е. не имеет кратных корней ни в каком расширении поля K) тогда и только тогда, когда матрица (13) невырождена.

Замечание 2. Последнее утверждение верно в любой характеристике и может быть доказано более непосредственно. А именно, матрица (13) может быть представлена в виде произведения

$$\begin{pmatrix} 1 & 1 & \dots & 1 \\ c_1 & c_2 & \dots & c_n \\ c_1^2 & c_2^2 & \dots & c_n^2 \\ \dots & \dots & \dots & \dots \\ c_1^{n-1} & c_2^{n-1} & \dots & c_n^{n-1} \end{pmatrix} \begin{pmatrix} 1 & c_1 & c_1^2 & \dots & c_1^{n-1} \\ 1 & c_2 & c_2^2 & \dots & c_2^{n-1} \\ 1 & c_3 & c_3^2 & \dots & c_3^{n-1} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & c_n & c_n^2 & \dots & c_n^{n-1} \end{pmatrix}.$$

Следовательно, ее определитель равен

$$\prod_{i>j} (c_i - c_j)^2,$$

т. е. дискриминанту многочлена h (см. § 3.9), откуда и следует доказываемое утверждение.

В частности, для неприводимого многочлена h над полем K положительной характеристики мы получаем, что поле $L = K[x]/(h)$ является сепарабельным расширением поля K в смысле определения, данного в замечании 9.5.2, тогда и только тогда, когда многочлен h сепарабелен.

Задача 1. Доказать, что в случае $\text{char } K = 0$ число различных корней многочлена h равно рангу матрицы (13).

Алгебра A называется *простой*, если $A \neq 0$ и A не имеет никаких нетривиальных (т. е. отличных от 0 и A) идеалов.

Пример 5. Всякое расширение L поля K является простой (коммутативной) алгеброй над K .

Задача 2. Доказать обратное утверждение: всякая простая коммутативная алгебра над K есть либо поле, содержащее K (т. е. расширение поля K), либо одномерная алгебра с нулевым умножением.

Для любой алгебры A подпространство A^2 является идеалом (как и вообще произведение двух идеалов). Если алгебра A нильпотентна, то $A^2 \neq A$. Поэтому простая алгебра A не может быть нильпотентной, за исключением тривиального случая, когда $A^2 = 0$ и $\dim A = 1$, т. е. когда A — одномерная алгебра с нулевым умножением. За исключением этого случая, всякая простая алгебра является полупростой.

Пример 6. Алгебра $L(V)$ проста (см. пример 9.2.6) и, следовательно, полупроста.

Теорема 3. *Всякая полупростая ассоциативная алгебра A разлагается в прямую сумму (нетривиальных) простых алгебр:*

$$A = A_1 \oplus \dots \oplus A_s, \quad (14)$$

причем любой идеал алгебры A является суммой некоторых слагаемых этого разложения.

Доказательство. Мы докажем эту теорему в предположении, что $\text{char } K = 0$. Если алгебра A проста, то доказывать нечего (в этом случае $s = 1$). Пусть она не проста, и пусть $A_1 \subset A$ — какой-нибудь ее минимальный идеал. Тогда либо

$$A = A_1 \oplus A_1^\perp, \quad (15)$$

либо $A_1 \subset A_1^\perp$. Но во втором случае идеал A_1 в силу предложения 2 нильпотентен, так что этот случай невозможен. В первом случае из разложения (15) следует, что всякий идеал алгебры A_1 , а также всякий идеал алгебры A_1^\perp , является идеалом алгебры A и, значит, алгебра A_1 проста, а алгебра A_1^\perp полупроста. Если алгебра A_1^\perp еще не проста, то применяем к ней такое же рассуждение и т. д.

Пусть теперь I — любой идеал алгебры A . Обозначим через π_k проекцию на k -е слагаемое разложения (14). Очевидно, что $I_k = \pi_k(I)$ — идеал алгебры A_k . Если $I_k \neq 0$, то $I_k = A_k$ и, следовательно,

$$A_k = A_k^2 = A_k I_k = A_k I \subset I.$$

Это как раз и означает, что идеал I есть сумма некоторых слагаемых разложения (14). \square

В частности, всякая (конечномерная) полупростая коммутативная ассоциативная алгебра A есть прямая сумма нескольких конечных расширений поля K (см. задачу 2), а если поле K алгебраически замкнуто — то просто нескольких копий самого поля K .

Задача 3. Доказать последнее утверждение (относящееся к алгебраически замкнутому случаю) средствами коммутативной алгебры. (Указание: рассмотреть $\text{Spec } A$ и воспользоваться теоремой Гильберта о нулях.)

Пример 7. Пусть многочлен $h \in K[x]$ не имеет кратных неприводимых множителей, и пусть $h = p_1 \dots p_s$ — его разложение на неприводимые множители над K . В силу теоремы 9.2.5 имеет место изоморфизм алгебр

$$K[t]/(h) \simeq K[t]/(p_1) \oplus \dots \oplus K[t]/(p_s). \quad (16)$$

Это и дает разложение полупростой алгебры $K[t]/(h)$ в прямую сумму простых алгебр (конечных расширений поля K). В частности, при $K = \mathbb{R}$ каждому вещественному корню многочлена h отвечает одномерное слагаемое разложения (16), изоморфное \mathbb{R} , а каждой паре сопряженных мнимых корней — двумерное слагаемое, изоморфное \mathbb{C} (см. примеры 9.2.4 и 9.2.5).

Задача 4. Вычислив двумя способами скалярное умножение (11) в алгебре $\mathbb{R}[x]/(h)$, где $h \in \mathbb{R}[x]$ — многочлен, не имеющий кратных комплексных корней, доказать, что число пар сопряженных мнимых корней многочлена h равно отрицательному индексу инерции симметричной матрицы (13). В частности, все корни многочлена h вещественны тогда и только тогда, когда матрица (13) положительно определена.

Что касается простых алгебр, то их строение в случае алгебраически замкнутого поля K описывается следующей теоремой, а общему случаю будет посвящен § 6.

Теорема 4. *Всякая нетривиальная простая ассоциативная алгебра A над алгебраически замкнутым полем K изоморфна алгебре вида $L(V)$, где V — векторное пространство над K , а всякое ее нетривиальное неприводимое представление изоморфно тавтологическому представлению алгебры $L(V)$.*

(Под тривиальным неприводимым представлением здесь понимается одномерное представление R , при котором $R(A) = 0$.)

Доказательство. Рассмотрим ограничение регулярного представления алгебры A на какое-либо минимальное инвариантное подпространство V (левый идеал) алгебры A . Полученное неприводимое представление обозначим через R . Его ядро есть идеал алгебры A . По теореме Бернсайда либо

$$A \simeq R(A) = L(V),$$

либо $\dim V = 1$ и $R(A) = 0$. Во втором случае $AV = 0$, так что

$$A_0 = \{x \in A : Ax = 0\} \neq 0;$$

но легко видеть, что A_0 — идеал алгебры A и, значит, $A_0 = A$, что противоречит нетривиальности алгебры A .

Пусть теперь $A = L(V)$, где V — какое-то векторное пространство, и R — тавтологическое представление алгебры A в пространстве V . Тогда регулярное представление T алгебры A изоморфно nR , где $n = \dim V$. Пусть $S : A \rightarrow L(U)$ — любое неприводимое представление алгебры A . Возьмем какой-нибудь ненулевой вектор $u_0 \in U$ и рассмотрим отображение

$$\varphi : A \rightarrow U, \quad a \mapsto S(a)u_0.$$

Так как

$$\varphi(T(a)x) = \varphi(ax) = S(ax)u_0 = S(a)S(x)u_0 = S(a)\varphi(x),$$

то φ — морфизм представления T в представление S . Если представление S нетривиально, то $\text{Im } \varphi = U$, так что представление S изоморфно факторпредставлению представления T . Так как S неприводимо, то $S \simeq R$ (см. следствие 2 теоремы 1.3). \square

Из теорем 3 и 4 следует, что всякая полупростая ассоциативная алгебра над алгебраически замкнутым полем изоморфна алгебре вида

$$A = L(V_1) \oplus \dots \oplus L(V_s), \quad (17)$$

где V_1, \dots, V_s — какие-то векторные пространства. Если $\dim V_i = n_i$ ($i = 1, \dots, s$), то

$$\dim A = n_1^2 + \dots + n_s^2. \quad (18)$$

Число s можно найти, если известен центр алгебры A . Вообще, центром ассоциативной алгебры A называется (коммукативная) подалгебра

$$Z(A) = \{z \in A : az = za \ \forall a \in A\}.$$

Известно (см. задачу 1.8.2), что центр алгебры $L(V)$ одномерен и состоит из скалярных преобразований. Поэтому для полупростой алгебры A , заданной разложением (17),

$$\dim Z(A) = s. \quad (19)$$

Пусть R_i ($i = 1, \dots, s$) обозначает неприводимое представление алгебры A в пространстве V_i , задаваемое проекцией на $L(V_i)$ в разложении (17). Отметим, что представления R_1, \dots, R_s попарно не изоморфны, так как их ядра различны.

Теорема 5. *Всякое нетривиальное неприводимое представление алгебры (17) изоморфно одному из представлений R_1, \dots, R_s .*

Доказательство. Пусть $S: A \rightarrow L(U)$ — нетривиальное неприводимое представление алгебры A . Так как $S(A) = L(U)$ — простая алгебра, то $\text{Ker } S$ есть сумма всех слагаемых разложения (17), кроме одного, скажем $L(V_i)$. Но тогда S кроме одного, скажем $L(V_i)$. Но тогда S фактически сводится к представлению алгебры $L(V_i)$ и по теореме 4 изоморфно R_i . \square

Имеет место также следующая теорема: *всякое линейное представление полупростой ассоциативной алгебры (над любым полем) вполне приводимо.*

§ 4. Линейные представления конечных групп

Теория ассоциативных алгебр дает важную информацию о линейных представлениях конечных групп.

Пусть G — конечная группа порядка n и K — какое-нибудь поле.

Определение 1. *Групповой алгеброй группы G над полем K называется алгебра KG , базисные элементы которой занумерованы элементами группы G , причем произведение базисных элементов с номерами $g, h \in G$ есть базисный элемент с номером gh .*

Обычно базисные элементы алгебры KG отождествляют с соответствующими элементами группы G . При таком соглашении всякий элемент алгебры KG записывается в виде

$$a = \sum_{g \in G} a_g g \quad (a_g \in K). \quad (20)$$

Из ассоциативности умножения в группе G следует ассоциативность умножения в алгебре KG .

Всякое линейное представление R группы G в векторном пространстве V над полем K однозначно продолжается до линейного представления алгебры KG в том же пространстве по формуле

$$R\left(\sum_{g \in G} a_g g\right) = \sum_{g \in G} a_g R(g).$$

Обратно, ограничение всякого линейного представления алгебры KG на G есть линейное представление группы G . Тем самым устанавливается естественное взаимно однозначное соответствие между представлениями группы и представлениями ее групповой алгебры.

Очевидно, что соответствующие друг другу представления группы G и алгебры KG имеют один и тот же набор инвариантных подпространств. В частности, неприводимым представлениям группы отвечают неприводимые представления групповой алгебры, и наоборот.

Теорема 1. Если $\text{char } K \nmid n$, то алгебра KG полупроста.

Доказательство. Воспользуемся теоремой 3.2. Для этого вычислим скалярное умножение (11) на алгебре KG . Легко видеть, что для любого $g \in G$

$$\text{tr } T(g) = \begin{cases} n, & g = e, \\ 0, & g \neq e. \end{cases}$$

Поэтому для любых $g, h \in G$

$$(g, h) = \begin{cases} n, & gh = e, \\ 0, & gh \neq e. \end{cases} \quad (21)$$

При $\text{char } K \nmid n$ это скалярное умножение невырожденно и, значит, алгебра KG полупроста. \square

В оставшейся части этого параграфа мы будем считать, что $K = \mathbb{C}$. Из теоремы 1 и результатов § 3 следует, что алгебра $\mathbb{C}G$ есть прямая сумма матричных алгебр.

Теорема 2. Группа G имеет, с точностью до изоморфизма, лишь конечное число неприводимых комплексных представлений. Их размерности n_1, \dots, n_s подчиняются соотношению

$$n_1^2 + \dots + n_s^2 = n, \quad (22)$$

а их число s равно числу классов сопряженных элементов группы G .

Доказательство. Первое утверждение теоремы вытекает из теоремы 3.5, а соотношение (22) — из формулы (18). Число s по формуле (19) равно размерности центра алгебры $\mathbb{C}G$. Найдем этот центр.

Элемент (20) принадлежит центру алгебры $\mathbb{C}G$ тогда и только тогда, когда он перестановочен со всеми элементами группы G , т. е. когда

$$hah^{-1} = \sum_{g \in G} a_g (hgh^{-1}) = \sum_{g \in G} a_{h^{-1}gh} g = a$$

для любого $h \in G$. Последнее означает, что в выражении элемента a коэффициенты при сопряженных элементах группы G равны. Следовательно, центр алгебры $\mathbb{C}G$ есть линейная оболочка элементов вида $\sum_{g \in C} g$, где C — класс сопряженных элементов, а его размерность равна числу классов сопряженных элементов. \square

Пример 1. Для абелевой группы все неприводимые представления одномерны (следствие 2 теоремы 1.2). Их число равно n , так как в этом случае каждый класс сопряженных элементов состоит из одного элемента. Это согласуется с формулой (22).

Пример 2. Так как при любом гомоморфизме группы G в абелеву группу ее коммутант переходит в единицу, то одномерные представления любой группы G сводятся к представлениям факторгруппы G/G' . В частности, группа S_n при любом n имеет ровно два одномерных представления: тривиальное и нетривиальное, ставящее в соответствие каждой подстановке ее знак.

Пример 3. Для группы S_3 заведомо имеются следующие три попарно не изоморфных неприводимых представления:

R_1 — тривиальное одномерное представление,

R'_1 — знак подстановки,

R_2 — двумерное представление, при котором S_3 изоморфно отображается на группу симметрии правильного треугольника (проверьте, что это представление неприводимо не только над \mathbb{R} , но и над \mathbb{C} !).

Так как в S_3 имеется ровно три класса сопряженных элементов (или, если угодно, так как $1^2 + 1^2 + 2^2 = 6$), этим исчерпывается список неприводимых представлений группы S_3 .

Пример 4. Аналогичным образом можно получить следующий полный список неприводимых представлений группы S_4 :

R_1 — тривиальное одномерное представление,

R'_1 — знак подстановки,

R_2 — композиция гомоморфизма $S_4 \rightarrow S_4/V_4 \simeq S_3$ и двумерного неприводимого представления группы S_3 ,

R_3 — изоморфизм на группу вращений куба,

R'_3 — изоморфизм на группу симметрии правильного тетраэдра.

Замечание 1. Как следует из примеров 3 и 4, все неприводимые комплексные представления (а, значит, и вообще все комплексные представления) групп S_3 и S_4 являются комплексификациями вещественных представлений. Можно доказать, что то же самое верно для группы S_n при любом n . Поэтому, учитывая предложение 1.1, с вещественными представлениями групп S_n можно работать так же, как с комплексными. В частности, для них справедлива вся теория, изложенная в этом параграфе.

Задача 1. Описать неприводимые представления группы диэдра D_n .

Задача 2. Доказать, что любое неприводимое представление группы $G \times H$ есть тензорное произведение неприводимых представлений групп G и H (см. определение в задаче 1.3).

Пусть

$$R_i: G \rightarrow \text{GL}(V_i) \quad (i = 1, \dots, s)$$

— все неприводимые комплексные представления группы G . Тогда, произведя подходящие отождествления, мы можем считать, что

$$\mathbb{C}G = L(V_1) \oplus \dots \oplus L(V_s), \quad (23)$$

причем R_i есть просто проекция на i -е слагаемое в этом разложении.

Подпространства $L(V_i)$ суть изотипные компоненты регулярного представления T алгебры $\mathbb{C}G$, причем ограничение T на $L(V_i)$ изоморфно $n_i R_i$, где $n_i = \dim V_i$. Поэтому для любых $a, b \in \mathbb{C}G$

$$(a, b) = \sum_{i=1}^s n_i \text{tr } R_i(a)R_i(b) \quad (24)$$

(ср. формулу (12)).

Рассмотрим теперь пространство $\mathbb{C}[G]$ всех функций на G со значениями в \mathbb{C} . Поскольку всякая функция φ на G однозначно продолжается до линейной функции на $\mathbb{C}G$ по формуле

$$\varphi \left(\sum_{g \in G} a_g g \right) = \sum_{g \in G} a_g \varphi(g),$$

то пространство $\mathbb{C}[G]$ естественно отождествляется с пространством, сопряженным к $\mathbb{C}G$.

С другой стороны, имеющееся скалярное умножение в пространстве $\mathbb{C}G$ определяет его изоморфизм с сопряженным пространством. В частности, элементу $g \in G$ при этом изоморфизме соответствует функция φ_g , определяемая согласно формуле

$$\varphi_g(h) = (g, h) = \begin{cases} n, & gh = e, \\ 0, & gh \neq e, \end{cases}$$

т. е. взятая с множителем n δ -функция $\delta_{g^{-1}}$ в точке g^{-1} .

Перенесем с помощью указанного изоморфизма скалярное умножение из пространства $\mathbb{C}G$ в пространство $\mathbb{C}[G]$. Тогда для δ -функций мы получим

$$(\delta_g, \delta_h) = \frac{1}{n^2} (g^{-1}, h^{-1}) = \begin{cases} 1/n, & gh = e, \\ 0, & gh \neq e, \end{cases}$$

а для любых функций φ и ψ —

$$(\varphi, \psi) = \frac{1}{n} \sum_{g \in G} \varphi(g) \psi(g^{-1}). \quad (25)$$

Вычислим теперь скалярные произведения матричных элементов неприводимых представлений группы G .

В каждом из пространств V_i ($i = 1, \dots, s$) выберем базис и обозначим через φ_{ijk} ($j, k = 1, \dots, n_i$) (j, k) -й матричный элемент оператора $R_i(g)$ в этом базисе. Функция $\varphi_{ijk} \in \mathbb{C}[G]$, определенная таким образом, называется (j, k) -м *матричным элементом* представления R_i .

С другой стороны, обозначим через \mathcal{E}_{ijk} линейный оператор в пространстве V_i , матрица которого в выбранном базисе есть матричная единица E_{jk} . Ввиду разложения (23) элементы \mathcal{E}_{ijk} составляют базис пространства $\mathbb{C}G$. Из формулы (24) следует, что

$$(\mathcal{E}_{ijk}, \mathcal{E}_{ikj}) = n_i, \quad (26)$$

а все остальные скалярные произведения элементов \mathcal{E}_{ijk} равны нулю.

При изоморфизме пространств $\mathbb{C}G$ и $\mathbb{C}[G]$ элементу \mathcal{E}_{ijk} в силу (26) соответствует матричный элемент φ_{ikj} с коэффициентом n_i . Следовательно,

$$(\varphi_{ijk}, \varphi_{ikj}) = \frac{1}{n_i}, \quad (27)$$

а все остальные скалярные произведения матричных элементов равны нулю.

Особенный интерес представляют суммы диагональных матричных элементов, называемые характерами представлений R_i .

Вообще, пусть $R: G \rightarrow GL(V)$ — любое линейное представление группы G .

Определение 2. Функция $\chi \in \mathbb{C}[G]$, определяемая формулой

$$\chi(g) = \text{tr } R(g),$$

называется *характером* представления R .

Очевидно, что характер суммы двух представлений равен сумме их характеров.

Так как следы сопряженных операторов равны, то

$$\chi(hgh^{-1}) = \chi(g) \quad \forall g, h \in G.$$

Функции $\chi \in \mathbb{C}[G]$, обладающие этим свойством, называются *центральными*. Они образуют в $\mathbb{C}[G]$ подпространство, которое мы обозначим через $Z\mathbb{C}[G]$. Очевидно, что $\dim Z\mathbb{C}[G] = s$.

В частности, пусть χ_i ($i = 1, \dots, s$) — характер представления R_i . Из (27) следует

Теорема 3. *Характеры χ_1, \dots, χ_s образуют ортонормированный базис пространства $Z\mathbb{C}[G]$, т. е.*

$$(\chi_i, \chi_j) = \delta_{ij}. \quad (28)$$

Пусть $R: G \rightarrow GL(V)$ — какое-то линейное представление и χ — его характер.

Следствие 1. *Кратность, с которой неприводимое представление R_i входит в представление R , равна (χ, χ_i) .*

Доказательство. Если $R \simeq \sum_{i=1}^s k_i R_i$, то $\chi = \sum_{i=1}^s k_i \chi_i$ и, следовательно, $(\chi, \chi_i) = k_i$. \square

Следствие 2. *Представление R неприводимо тогда и только тогда, когда $(\chi, \chi) = 1$.*

Доказательство. Если $R \simeq \sum_{i=1}^s k_i R_i$, то $(\chi, \chi) = \sum_{i=1}^s k_i^2 = 1$ тогда и только тогда, когда одна из кратностей k_i равна 1, а все остальные равны 0. \square

Вместо билинейного скалярного умножения (25) в пространстве $\mathbb{C}[G]$ можно рассматривать эрмитово скалярное умножение

$$(\varphi | \psi) = \frac{1}{n} \sum_{g \in G} \varphi(g) \overline{\psi(g)}, \quad (29)$$

более удобное для практических вычислений. Если в каждом из пространств V_i выбрать базис, ортонормированный относительно инвариантного эрмитова скалярного умножения (см. теорему 2.3), то операторы представления запишутся унитарными матрицами, т. е. будут выполняться соотношения

$$\varphi_{ikj}(g^{-1}) = \overline{\varphi_{ijk}(g)}.$$

Соотношения (27) и (28) в терминах эрмитовой метрики (29) означают, что матричные элементы φ_{ijk} образуют ортогональный базис пространства $\mathbb{C}[G]$, причем

$$(\varphi_{ijk} | \varphi_{ijk}) = \frac{1}{n_i},$$

а характеры χ_i образуют ортонормированный базис пространства $Z\mathbb{C}[G]$.

Пример 5. Характер одномерного представления совпадает с единственным матричным элементом и, если угодно, с самим представлением. Циклическая группа $\langle a \rangle_n$ имеет n одномерных комплексных представлений R_0, R_1, \dots, R_{n-1} , определяемых условиями

$$R_k(a) = \omega^k, \quad \omega = e^{2\pi i/n}.$$

Поэтому характеры этой группы задаются следующей таблицей:

	χ_0	χ_1	\dots	χ_{n-1}
e	1	1	\dots	1
a	1	ω	\dots	ω^{n-1}
a^2	1	ω^2	\dots	$\omega^{2(n-1)}$
\dots	\dots	\dots	\dots	\dots
a^{n-1}	1	ω^{n-1}	\dots	$\omega^{(n-1)^2}$

Соотношения ортогональности для характеров означают в данном случае, что если таблицу характеров поделить на \sqrt{n} , то получится унитарная матрица.

Пример 6. Пользуясь описанием неприводимых представлений группы S_4 , данным в примере 4, нетрудно получить следующую таблицу характеров группы S_4 :

	χ_1	χ'_1	χ_2	χ_3	χ'_3	
e	1	1	2	3	3	1
(12)	1	-1	0	-1	1	6
(12)(34)	1	1	2	-1	-1	3
(123)	1	1	-1	0	0	8
(1234)	1	-1	0	1	-1	6

В левом (входном) столбце этой таблицы указаны представители классов сопряженных элементов группы S_4 , а в крайнем правом приведены числа элементов в этих классах, необходимые для вычисления скалярных произведений. Так, например,

$$\begin{aligned}
 (\chi_2, \chi_3) &= (\chi_2 | \chi_3) = \\
 &= \frac{1}{24}(1 \cdot 2 \cdot 3 + 6 \cdot 0 \cdot (-1) + 3 \cdot 2 \cdot (-1) + 8 \cdot (-1) \cdot 0 + 6 \cdot 0 \cdot 1) = 0.
 \end{aligned}$$

Пример 7. Пусть V есть (6-мерное) пространство функций на множестве граней куба. Изоморфизм группы S_4 и группы вращений куба определяет линейное представление группы S_4 в пространстве V . Обозначим это представление через R , а его характер — через χ . Всякий элемент $g \in S_4$ каким-то образом переставляет грани куба и таким же образом $R(g)$ переставляет δ -функции этих граней. Поэтому $\chi(g) = \text{tr } R(g)$ есть число граней, которые элемент g оставляет на месте. Таким образом получается следующая таблица значений характера χ :

	e	(12)	(12)(34)	(123)	(1234)
χ	6	0	2	0	2

Вычисляя скалярные произведения этого характера и характеров неприводимых представлений группы S_4 (см. пример 6), находим:

$$(\chi | \chi_1) = 1, \quad (\chi | \chi'_1) = 0, \quad (\chi | \chi_2) = 1, \quad (\chi | \chi_3) = 1, \quad (\chi | \chi'_3) = 0.$$

Таким образом,

$$R \simeq R_1 + R_2 + R_3.$$

Задача 3. Найти минимальные инвариантные подпространства для представления R из примера 7.

Задача 4. Пусть $G \subset S_n$ — дважды транзитивная группа подстановок. (Это означает, что для любых двух упорядоченных пар различных символов найдется подстановка из G , переводящая первую пару во вторую.) Доказать, что представление группы G в пространстве функций на множестве $\{1, \dots, n\}$ разлагается в сумму ровно двух неприводимых представлений, одно из которых — тривиальное одномерное. (Указание: использовать выражение, которое дает формула Бернсайда (см. задачу 10.3.2) для числа орбит группы G в множестве $\{1, \dots, n\} \times \{1, \dots, n\}$.)

Задача 5. Составить таблицу характеров группы A_5 .

Задача 6. Пусть $R: G \rightarrow \text{GL}(V)$ — какое-то линейное представление конечной группы G . Доказать, что проектор \mathcal{P}_i пространства V на его изотипную компоненту, отвечающую неприводимому представлению R_i группы G , может быть задан формулой

$$\mathcal{P}_i = \frac{n_i}{n} \sum_{g \in G} \chi_i(g^{-1}) R(g),$$

где $n = |G|$, $n_i = \dim R_i$, а χ_i — характер представления R_i . (Указание: доказать, что элемент

$$\frac{n_i}{n} \sum_{g \in G} \chi_i(g^{-1}) g \in \mathbb{C}G$$

есть единица i -го слагаемого разложения (23), для чего вычислить его скалярные произведения с элементами группы G , пользуясь формулами (21) и (24).)

Помимо уже рассмотренной нами операции сложения представлений, имеются и другие важные операции над линейными представлениями (произвольных) групп.

Для любого линейного представления $R: G \rightarrow \text{GL}(V)$ можно определить сопряженное представление $R^*: G \rightarrow \text{GL}(V^*)$ по правилу

$$(R^*(g)\alpha)(x) = \alpha(R(g)^{-1}x) \quad (\alpha \in V^*, x \in V), \quad (30)$$

т. е. по обычному правилу действия преобразований на функции. На матричном языке это выглядит следующим образом:

$$R^*(g) = (R(g)^T)^{-1}. \quad (31)$$

Следовательно, характер сопряженного представления находится по формуле

$$\chi_{R^*}(g) = \chi_R(g^{-1}). \quad (32)$$

Определение сопряженного представления может быть переписано в следующем симметричном виде:

$$(R^*(g)\alpha)(R(g)x) = \alpha(x).$$

Отсюда следует, что $R^{**} = R$ (при каноническом отождествлении пространства V^{**} с V). Может случиться, что $R^* \simeq R$; в этом случае представление R называется *самосопряженным*.

Для комплексного линейного представления конечной группы в базисе, ортонормированном относительно инвариантного эрмитова скалярного умножения, формулы (31) и (32) принимают вид

$$R^*(g) = \overline{R(g)}, \quad \chi_{R^*}(g) = \overline{\chi_R(g)}. \quad (33)$$

Пример 8. Для неприводимых (одномерных) представлений циклической группы имеем в обозначениях примера 5:

$$R_0^* \simeq R_0, \quad R_k^* \simeq R_{n-k} \quad (k = 1, \dots, n-1).$$

Пример 9. Как следует из примера 10.3.5, в группе S_n любой элемент сопряжен своему обратному. Поэтому всякое линейное представление группы S_n самосопряженно.

Задача 7. Доказать, что если R — неприводимое представление какой-либо группы, то представление R^* также неприводимо.

Задача 8. Доказать, что все представления конечной группы G самосопряженны тогда и только тогда, когда всякий элемент группы G сопряжен своему обратному.

Определим теперь операцию умножения линейных представлений группы G .

Произведением линейных представлений $R: G \rightarrow GL(V)$ и $S: G \rightarrow GL(W)$ называется линейное представление

$$RS: G \rightarrow GL(V \otimes W), \quad g \mapsto R(g) \otimes S(g).$$

(См. определение тензорного произведения линейных операторов в § 8.1.)

Замечание 2. Иногда представление RS называют тензорным произведением представлений R и S , но мы сохраняем этот термин для представления прямого произведения двух групп, определенно в задаче 1.3.

Задача 9. Пусть в пространствах V и W выбраны какие-то базисы и представления R и S записаны в этих базисах. Будем задавать элемент пространства $V \otimes W$ матрицей Z , составленной из его координат (см. формулу (10) § 8.1). Доказать, что представление RS в этих терминах задается формулой

$$(RS)(g)Z = R(g)ZS(g)^\top. \quad (34)$$

Из формулы (28) § 8.1 следует, что

$$\chi_{RS} = \chi_R \chi_S. \quad (35)$$

Произведение неприводимых представлений, как правило, не является неприводимым. Разложение произведений неприводимых представлений на неприводимые компоненты — это одна из основных задач теории представлений. Для представлений конечных групп благодаря формуле (35) она в принципе может быть решена с помощью теории характеров.

Пример 10. Разложим в сумму неприводимых представлений квадрат представления R_3 группы S_4 (см. пример 4). Пользуясь таблицей характеров группы S_4 , приведенной в примере 6, получаем:

$$\begin{aligned} (\chi_3^2 | \chi_1) &= 1, & (\chi_3^2 | \chi'_1) &= 0, & (\chi_3^2 | \chi_2) &= 1, \\ (\chi_3^2 | \chi_3) &= 1, & (\chi_3^2 | \chi'_3) &= 1, \end{aligned}$$

Следовательно,

$$R_3^2 \simeq R_1 + R_2 + R_3 + R'_3. \quad (36)$$

Аналогично определяется произведение нескольких представлений, а также симметрическая и внешняя степени представления. Например, симметрический квадрат представления $R: G \rightarrow \text{GL}(V)$ есть представление

$$S^2R: G \rightarrow \text{GL}(S^2(V)), \quad g \mapsto S^2(R(g)).$$

(См. определение симметрического квадрата линейного оператора в § 8.3.)

Из формулы (53) § 8.3 следует, что

$$\chi_{S^2R}(g) = \frac{1}{2}(\chi_R(g)^2 + \chi_R(g^2)). \quad (37)$$

Если отождествить пространство $S^2(V)$ с пространством $ST^2(V)$ симметрических тензоров, то представление S^2R будет не чем

иным, как ограничением представления R^2 на инвариантное подпространство $ST^2(V)$. Аналогичное утверждение справедливо и для внешнего квадрата $\Lambda^2 R$ представления R . Так как $T^2(V) = ST^2(V) \oplus \Lambda T^2(V)$, то

$$R^2 \simeq S^2 R + \Lambda^2 R. \quad (38)$$

Пример 11. Закон Гука в теории упругости выражает связь между тензором деформации σ и тензором напряжений τ твердого тела в какой-либо его точке. Оба этих тензора суть симметрические операторы в пространстве E^3 . (Определение тензора деформации см. в примере 6.3.5). Подняв индексы, их можно рассматривать, если угодно, как элементы пространства $S^2(E^3)$. Закон Гука имеет вид $\sigma = \mathcal{H}\tau$, где \mathcal{H} — некоторый симметрический оператор в пространстве $S^2(E^3)$, называемый тензором упругости и характеризующий упругие свойства данного твердого тела в данной точке (при заданных температуре и давлении). Так как $\dim S^2(E^3) = 6$, то размерность пространства симметрических операторов в $S^2(E^3)$ равна $\frac{6 \cdot 7}{2} = 21$. Таким образом, в общем случае тензор упругости определяется 21 параметрами, которые должны быть найдены опытным путем.

Ситуация упрощается, если тело имеет кристаллическую структуру. А именно, пусть $G = d\Gamma$, где Γ — группа симметрии данной кристаллической структуры (см. пример 9.1.1). Тогда оператор \mathcal{H} должен быть перестановочен со всеми операторами $S^2 \mathcal{A}$, где $\mathcal{A} \in G$. Общий вид такого оператора может быть найден с помощью теории представлений. Число параметров, от которых он зависит, будет тем меньше, чем больше группа G .

Рассмотрим, например, кристалл поваренной соли (см. рис. 2 § 4.2). В этом случае группа G есть группа симметрии куба, т. е., в обозначениях примера 4, $G = R_3(S_4) \times \{\pm \mathcal{E}\}$. Второй множитель тривиально действует на $S^2(E^3)$, поэтому его можно не учитывать. Таким образом, оператор \mathcal{H} должен быть эндоморфизмом представления $S^2 R_3$ группы S_4 . По формуле (37) получаем следующую таблицу значений характера χ этого представления:

	e	(12)	(12)(34)	(123)	(1234)
χ	6	2	2	0	0

Вычисляя его скалярные произведения с характеристиками неприводимых представлений, находим, что

$$S^2 R_3 \simeq R_1 + R_2 + R'_3. \quad (39)$$

В частности, представление $S^2 R_3$ имеет простой спектр. Согласно предложению 1.4 (см. также замечание 1), общий вид его эндоморфизма зависит от 3 параметров. Итак, для определения тензора упругости кристалла поваренной соли требуется найти опытным путем значения лишь 3 параметров (вместо 21!).

Замечание 3. Ввиду изоморфизма (38) разложение (39) можно было бы найти вычитанием из разложения (36) представления $\Lambda^2 R_3$, которое, как легко показать, изоморфно R_3 .

Задача 10. Доказать, что неприводимое представление группы самосопряженно тогда и только тогда, когда оно ортогонально или симплектично (см. определение в конце § 1).

Задача 11. Доказать, что неприводимое комплексное представление конечной группы является комплексификацией вещественно-го представления тогда и только тогда, когда оно ортогонально.

Задача 12. Доказать, что сумма размерностей ортогональных неприводимых представлений конечной группы G минус сумма размерностей ее симплектических неприводимых представлений равно числу решений уравнения $x^2 = e$ в G . (Указание: подсчитать след антиавтоморфизма групповой алгебры $\mathbb{C}G$, индуцированного инверсией в группе G , в базисе из элементов группы G и в базисе, согласованном с разложением (23).)

§ 5. Инварианты

Всякое действие группы G на множестве X определяет по формуле (8) § 10.3 линейное представление этой группы в пространстве $F(X, K)$ функций на X со значениями в (любом) поле K .

Определение 1. Функция $f \in F(X, K)$ называется *инвариантом* (данного действия) группы G , если $gf = f$ для любого $g \in G$.

Иными словами, инвариант — это функция, постоянная на орбитах группы G . Знание инвариантов помогает решению важной задачи описания орбит. А именно, если какой-либо инвариант f принимает разные значения в каких-либо двух точках, то эти точки заведомо принадлежат разным орбитам. Идеальным решением за-

дачи является указание таких инвариантов f_1, \dots, f_m , что для любых двух точек, принадлежащих разным орбитам, хотя бы один из них принимает в этих точках разные значения. В этом случае говорят, что инварианты f_1, \dots, f_m *разделяют орбиты*.

Пример 1. Рассмотрим так называемое *присоединенное* линейное представление Ad группы $\text{GL}_n(\mathbb{C})$ в пространстве $L_n(\mathbb{C})$, определяемое по формуле

$$\text{Ad}(A)X = AXA^{-1}.$$

Пусть $f_X(t) = \det(tE - X)$ — характеристический многочлен матрицы X . Запишем его в виде

$$f_X(t) = t^n - f_1(X)t^{n-1} + \dots + (-1)^n f_n(X).$$

Тогда $f_k(X)$ есть сумма главных миноров порядка k матрицы X (см. задачу 6.2.1). Так как характеристические многочлены подобных матриц равны, то f_1, \dots, f_n — инварианты рассматриваемого действия группы $\text{GL}_n(\mathbb{C})$. Однако они не разделяют орбит. Действительно, две матрицы принадлежат одной орбите тогда и только тогда, когда они имеют одну и ту же жорданову форму, в то время как значения инвариантов f_1, \dots, f_n определяют лишь собственные значения матрицы.

Пример 2. Инварианты симметрической группы S_n , действующей в K^n перестановками координат, — это функции от n переменных, не меняющиеся ни при какой перестановке переменных. В частности, инвариантные многочлены этой группы — это симметрические многочлены.

Пространство $F(X, K)$ является алгеброй относительно обычной операции умножения функций, и преобразования из группы G являются автоморфизмами этой алгебры. Отсюда следует, что инварианты образуют подалгебру в $F(X, K)$.

Обычно инварианты ищут не среди всех вообще, а среди «хороших» в том или ином смысле функций. Наиболее распространенной является ситуация, когда $X = V$ — векторное пространство над полем K , а действие группы G определяется ее линейным представлением в пространстве V . В этой ситуации обычно ищут инварианты в алгебре $K[V]$ многочленов на V . (Именно так обстояло дело в примере 1.) Подалгебра инвариантов в $K[V]$ обозначается через $K[V]^G$.

Говорят, что орбиты линейной группы $G \subset \text{GL}(V)$ *разделяются инвариантами*, если для любых двух различных орбит найдется инвариант $f \in K[V]^G$, принимающий на них различные значения.

Теорема 1. Если $G \subset GL(V)$ — конечная группа и ее порядок не делится на $\text{char } K$, то ее орбиты разделяются инвариантами.

Доказательство. Пусть O_1 и O_2 — две различные орбиты. Тогда существует многочлен $f \in K[V]$, равный 1 во всех точках орбиты O_1 и 0 во всех точках орбиты O_2 . Совокупность всех многочленов степени $\leq \deg f$, обладающих этим свойством, есть некоторая плоскость S в пространстве всех многочленов степени $\leq \deg f$. Группа G сохраняет эту плоскость и действует на ней аффинными преобразованиями. Согласно лемме 2.1, в S существует неподвижная точка группы G . Это и будет искомым инвариант. \square

Задача 1. В приведенном доказательстве использовался тот факт, что для любого конечного числа точек пространства V существует многочлен, принимающий в этих точках любые наперед заданные значения. Докажите это.

Пример 3. Задание вектора $(x_1, x_2, \dots, x_n) \in K^n$ с точностью до перестановки его координат равносильно заданию многочлена

$$(x - x_1)(x - x_2) \dots (x - x_n) \in K[x].$$

Коэффициенты этого многочлена суть с точностью до знака элементарные симметрические многочлены от x_1, x_2, \dots, x_n . Следовательно, орбиты группы S_n в пространстве K^n (см. пример 2) разделяются элементарными симметрическими многочленами (являющимися инвариантами группы S_n) и, тем более, разделяются всеми инвариантами.

Если алгебра $K[V]^G$ порождается инвариантами f_1, \dots, f_m и эти инварианты принимают одинаковые значения в каких-то двух точках, то и все инварианты принимают одинаковые значения в этих точках. Поэтому, если орбиты группы G разделяются (всеми) инвариантами, то они разделяются инвариантами f_1, \dots, f_m . Так, в предыдущем примере можно было заранее предсказать (в случае, когда $\text{char } K \nmid |G|$), что орбиты группы S_n должны разделяться элементарными симметрическими многочленами, поскольку эти многочлены порождают алгебру всех симметрических многочленов.

Следующая теорема является частным случаем современной версии *теоремы Гильберта об инвариантах*. Сам Гильберт доказал эту теорему в 1891 г. для линейных представлений группы $SL_n(K)$, но основная идея его доказательства применима в гораздо более общей ситуации и в том числе для конечных групп.

Теорема 2. Если G — конечная группа и ее порядок не делится на $\text{char } K$, то алгебра $K[V]^G$ является конечно порожденной.

Утверждение теоремы означает, что существуют такие инварианты f_1, \dots, f_m , что всякий инвариант представляется (быть может, не однозначно) в виде многочлена от f_1, \dots, f_m .

Доказательство. Определим в алгебре $K[V]^G$ так называемый оператор Рейнольдса \natural (читается «бекар») по формуле

$$f^\natural = \text{center } Gf = \frac{1}{|G|} \sum_{g \in G} gf. \quad (40)$$

Это линейный оператор, обладающий следующими свойствами:

- 1) $f^\natural \in K[V]^G$ для любого $f \in K[V]$;
- 2) $f^\natural = f$ для любого $f \in K[V]^G$;
- 3) $(fh)^\natural = fh^\natural$ для любых $f \in K[V]^G$, $h \in K[V]$.

Иными словами, это проектор на алгебру инвариантов, перестановочный с умножениями на инварианты.

Заметим, что многочлен инвариантен тогда и только тогда, когда инвариантны все его однородные составляющие, и что оператор Рейнольдса переводит любой однородный многочлен в однородный многочлен той же степени.

Пусть теперь $I \subset K[V]$ — идеал, порожденный всеми однородными инвариантами положительной степени. По теореме Гильберта о базисе идеала идеал I порождается конечным числом многочленов. Ясно, что их можно выбрать среди однородных инвариантов. Пусть это будут инварианты f_1, \dots, f_m , и пусть

$$K[f_1 \dots f_m] \subset K[V]^G$$

— порожденная ими подалгебра. Мы докажем, что она совпадает с алгеброй инвариантов. Для этого докажем индукцией по n , что каждый однородный инвариант степени n принадлежит алгебре $K[f_1, \dots, f_m]$.

При $n = 0$ доказывать нечего (алгебра $K[f_1, \dots, f_m]$ содержит единицу по определению). Пусть f — произвольный однородный инвариант положительной степени. Так как $f \in I$, то существуют такие многочлены $h_1, \dots, h_m \in K[V]$, что

$$f = \sum_{i=1}^m f_i h_i.$$

Можно считать, что h_i — однородный многочлен, степень которого равна

$$\deg h_i = \deg f - \deg f_i < \deg f.$$

Применяя к предыдущему равенству оператор \natural , мы получаем тогда

$$f = \sum_{i=1}^m f_i h_i^{\natural}.$$

По предположению индукции $h_i^{\natural} \in K[f_1, \dots, f_m]$. Следовательно, и $f \in K[f_1, \dots, f_m]$. \square

Явное нахождение конечной системы порождающих алгебры инвариантов в конкретном случае может представлять собой трудную задачу.

Пример 4. Дадим новое доказательство того, что алгебра инвариантов симметрической группы S_n (см. пример 2), т. е. алгебра симметрических многочленов, порождается элементарными симметрическими многочленами $\sigma_1, \dots, \sigma_n$. В примере 10.6.5 уже было показано, что

$$K(x_1, \dots, x_n)^{S_n} = K(\sigma_1, \dots, \sigma_n),$$

причем $\sigma_1, \dots, \sigma_n$ алгебраически независимы. Так как x_1, \dots, x_n являются корнями многочлена

$$(x - x_1) \dots (x - x_n) = x^n - \sigma_1 x^{n-1} + \dots + (-1)^n \sigma_n$$

с коэффициентами из $K[\sigma_1, \dots, \sigma_n]$, то алгебра $K[x_1, \dots, x_n]$ всех многочленов и, тем более, ее подалгебра $K[x_1, \dots, x_n]^{S_n}$ являются целыми расширениями алгебры $K[\sigma_1, \dots, \sigma_n]$. В то же время

$$K[x_1, \dots, x_n]^{S_n} \subset K(x_1, \dots, x_n)^{S_n} = K(\sigma_1, \dots, \sigma_n).$$

Так как алгебра $K[\sigma_1, \dots, \sigma_n]$, будучи изоморфна алгебре многочленов от n переменных, факториальна, то она нормальна (целозамкнута в своем поле отношений). Следовательно,

$$K[x_1, \dots, x_n]^{S_n} = K[\sigma_1, \dots, \sigma_n].$$

Пример 5. Не следует думать, что алгебра инвариантов всегда порождается алгебраически независимыми элементами, как в предыдущем примере. Такая ситуация является скорее исключением, чем правилом. Рассмотрим, например, группу

$$G = \{\pm E\} \subset GL(V), \quad \text{char } K \neq 2.$$

Однородный многочлен является инвариантом этой группы тогда и только тогда, когда его степень четна. Поэтому минимальная система однородных порождающих алгебры инвариантов состоит из многочленов $f_{ij} = x_i x_j$, которые связаны соотношениями

$$f_{ij} f_{kl} = f_{ik} f_{jl}.$$

Замечание 1. Теоремы 1 и 2 справедливы и для конечных групп, порядок которых делится на $\text{char } K$, но приведенные выше доказательства в этом случае не проходят.

В случае $K = \mathbb{R}$ доказанные теоремы могут быть обобщены на произвольные компактные группы.

Теорема 3. *Орбиты компактной группы G линейных преобразований вещественного векторного пространства V разделяются инвариантами.*

Доказательство. Следуя доказательству теоремы 1, мы не можем теперь априори гарантировать существование многочлена, равного 1 на O_1 и 0 на O_2 . Однако из теоремы Вейерштрасса о равномерной аппроксимации непрерывных функций на компактных множествах многочленами следует, что существует многочлен f , положительный на O_1 и отрицательный на O_2 . Совокупность всех многочленов степени $\leq \deg f$, обладающим этим свойством, есть некоторое G -инвариантное выпуклое множество M в пространстве всех многочленов степени $\leq \deg f$. Неподвижная точка группы G в этом множестве и будет искомым инвариантом. \square

Замечание 2. Для комплексных векторных пространств аналогичная теорема неверна, как показывает пример окружности $\mathbb{T} \subset \mathbb{C}^* = \text{GL}_1(\mathbb{C})$.

Теорема 4. *Пусть G — компактная группа линейных преобразований векторного пространства V над полем $K = \mathbb{R}$ или \mathbb{C} . Тогда алгебра $K[V]^G$ является конечно порожденной.*

Эта теорема, как и теорема 2, является частным случаем теоремы Гильберта об инвариантах.

Доказательство может быть проведено так же, как и доказательство теоремы 2, если только нам удастся определить оператор Рейнольдса, обладающей свойствами 1)–3). Это можно сделать, заменив в формуле (40) суммирование по конечной группе надлежащим образом определенным интегрированием по компактной группе. (Например, в случае $G = \mathbb{T}$ это будет обычное интегрирование по окружности.) Однако мы дадим другое доказательство.

В силу полной приводимости линейных представлений компактных групп (следствие теоремы 2.2) пространство $K[V]_n$ однородных многочленов степени n на V может быть разложено в прямую сумму подпространства $K[V]_n^G$ G -инвариантных многочленов и некоторого G -инвариантного подпространства $(K[V]_n)_G$. Положим

$$K[V]_G = \bigoplus_{n=0}^{\infty} (K[V]_n)_G.$$

Очевидно, что подпространство $K[V]_G$ инвариантно относительно G и

$$K[V] = K[V]^G \oplus K[V]_G. \quad (41)$$

Определим теперь оператор \natural как проектор на $K[V]^G$ относительно разложения (41). По построению этот проектор перестановочен с действием группы G . Единственное, что нам нужно проверить — это то, что он перестановочен с умножениями на инварианты. Для этого достаточно доказать, что

$$K[V]^G K[V]_G \subset K[V]_G.$$

Умножение на инвариант $f \in K[V]^G$ перестановочно с действием группы G , т. е. является эндоморфизмом представления группы G в пространстве $K[V]$. Так как подпространство $K[V]_G$ по построению дополнительно к $K[V]^G$, то представление группы G в $K[V]_G$ разлагается в сумму нетривиальных неприводимых представлений. То же самое можно сказать и о представлении группы G в $fK[V]_G$. Значит, проекция этого подпространства на $K[V]^G$ равна нулю, т. е.

$$fK[V]_G \subset K[V]_G,$$

что и требовалось доказать. \square

Пример 6. Рассмотрим линейное представление R группы O_n в пространстве L_n^+ вещественных симметричных матриц порядка n , определяемое формулой

$$R(A)X = AXA^{-1} (= AXA^T).$$

Рассмотрим характеристический многочлен матрицы X :

$$\det(tE - X) = t^n - f_1(X)t^{n-1} + f_2(X)t^{n-2} - \dots + (-1)^n f_n(X).$$

Докажем, что

$$\mathbb{R}[L_n^+]^{R(O_n)} = \mathbb{R}[f_1, \dots, f_n]$$

и что f_1, \dots, f_n алгебраически независимы. Для этого вспомним, что каждая симметричная матрица ортогонально подобна диагональной матрице. Поэтому любой инвариант f группы $R(O_n)$ однозначно определяется своим ограничением на подпространство D диагональных матриц. Так как диагональные матрицы, отличающиеся лишь порядком диагональных элементов, ортогонально подобны, то $f|_D$ есть симметрический многочлен от диагональных элементов x_1, \dots, x_n . Непосредственно проверяется, что ограничения на D инвариантов f_1, \dots, f_n суть элементарные симметрические многочлены от x_1, \dots, x_n . Доказываемые утверждения вытекают теперь из теоремы о симметрических многочленах. Заметим, что, как и должно быть, согласно теореме 3, в этом примере орбиты разделяются инвариантами.

§ 6. Алгебры с делением

Из алгебраической замкнутости поля комплексных чисел следует, что единственные конечномерные алгебры над \mathbb{R} , являющиеся полями, — это \mathbb{R} и \mathbb{C} . Однако если отказаться от коммутативности умножения, то можно построить еще одну такую алгебру, а именно, алгебру кватернионов, которая также играет заметную роль в математике и ее приложениях. Теория становится еще более содержательной, если в качестве основного поля рассматривать вместо \mathbb{R} произвольное поле (например, \mathbb{Q}).

Определение 1. Ассоциативное кольцо с единицей, в котором каждый ненулевой элемент имеет обратный, называется *телом*. Алгебра, являющаяся телом, называется *алгеброй с делением*.

Замечание 1. Кольцо, состоящее из одного нуля, не считается телом.

Иными словами, тело — это «некоммутативное поле». Как и в поле, в теле нет делителей нуля и элементы, отличные от нуля, образуют группу по умножению (но уже не обязательно абелеву). Мультипликативная группа ненулевых элементов тела D обозначается через D^* .

Всякое тело D может рассматриваться как алгебра с делением над своим *центром*

$$Z(D) = \{z \in D : za = az \ \forall a \in D\},$$

который, очевидно, является полем.

Если D — алгебра с делением над полем K и 1 — ее единица, то элементы вида $\lambda 1$, $\lambda \in K$, образуют подкольцо, изоморфное K и содержащееся в центре $Z(D)$ алгебры D . Обычно эти элементы отождествляют с соответствующими элементами поля K . При таком соглашении $Z(D) \supset K$. Алгебра D называется *центральной*, если $Z(D) = K$.

Задача 1. Доказать, что конечномерная ассоциативная алгебра является алгеброй с делением тогда и только тогда, когда в ней нет делителей нуля.

Наиболее простые и важные примеры некоммутативных алгебр с делением — это алгебры кватернионов.

(Обобщенной) алгеброй кватернионов над полем K характеристики $\neq 2$ называется алгебра $D = D(\alpha, \beta)$, порождаемая элементами i и j , удовлетворяющими соотношениям

$$i^2 = \alpha, \quad j^2 = \beta, \quad ij = -ji \quad (\alpha, \beta \in K^*).$$

Легко видеть, что базис алгебры D над K составляют элементы $1, i, j$ и $k = ij$, причем элементы i, j, k попарно антикоммутируют и

$$k^2 = -\alpha\beta.$$

В частности, при $K = \mathbb{R}$ алгебра $D(-1, -1)$ есть обычная алгебра кватернионов \mathbb{H} , открытая Гамильтоном в 1843 г.

Алгебра $D(1, 1)$ изоморфна алгебре матриц $L_2(K)$. Изоморфизм между этими алгебрами устанавливается следующим образом:

$$1 \leftrightarrow \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad i \leftrightarrow \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad j \leftrightarrow \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad k \leftrightarrow \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Для того чтобы выяснить возможность деления в алгебре кватернионов, определим для любого кватерниона

$$q = x + yi + zj + uk \quad (x, y, z, u \in K)$$

сопряженный кватернион \bar{q} по формуле

$$\bar{q} = x - yi - zj - uk.$$

Легко видеть, что линейное отображение $q \mapsto \bar{q}$, называемое *стандартной инволюцией*, является *антиавтоморфизмом* алгебры D , т. е.

$$\overline{q_1 q_2} = \bar{q}_2 \bar{q}_1.$$

(Ввиду линейности достаточно проверить это равенство для базисных элементов.) Элемент

$$N(q) = q\bar{q} = x^2 - \alpha y^2 - \beta z^2 + \alpha\beta u^2 \in K \quad (42)$$

называется *нормой* кватерниона q . Ясно, что q обратим тогда и только тогда, когда $N(q) \neq 0$ (и в этом случае $q^{-1} = N(q)^{-1}\bar{q}$).

Алгебра $D = D(\alpha, \beta)$ является алгеброй с делением тогда и только тогда, когда уравнение

$$x^2 - \alpha y^2 - \beta z^2 + \alpha\beta u^2 = 0$$

не имеет ненулевых решений в поле K , или, как говорят, квадратичная форма (42) не представляет нуля над K . В частности, это условие выполнено, если $K = \mathbb{R}$ и $\alpha = \beta = -1$, так как в этом случае квадратичная форма (42) положительно определена.

Заметим, наконец, что для любых $a, b \in K^*$ имеем

$$(ai)^2 = a^2\alpha, \quad (bj)^2 = b^2\beta, \quad (ai)(bj) = -(bj)(ai).$$

Это показывает, что

$$D(a^2\alpha, b^2\beta) \simeq D(\alpha, \beta).$$

Задача 2. Доказать, что в алгебре $D(1, 1) \simeq L_2(K)$ норма есть определитель соответствующей матрицы. Как в матричных терминах интерпретируется стандартная инволюция этой алгебры?

Задача 3. Доказать, что $D(\alpha, 1) \simeq L_2(K)$ при любом $\alpha \in K^*$.

Если D — конечномерная алгебра с делением над полем K , то для любого $x \in D$ подалгебра $K[x]$ коммутативна и, следовательно, является полем. Поэтому всякая конечномерная алгебра с делением над алгебраически замкнутым полем K совпадает с полем K .

Когда мы имеем дело с алгебрами над алгебраически незамкнутым полем, всегда полезно посмотреть, что происходит при алгебраических расширениях этого поля. Например, для изучения вещественных алгебр полезно посмотреть, что происходит при их комплексификации. Разрешив делать алгебраические расширения, мы ставим себя в ситуацию, равнозначную той, когда основное поле алгебраически замкнуто. С другой стороны, многие свойства алгебр при таких расширениях сохраняются.

Пусть A — алгебра над полем K и P — какое-либо расширение поля K . Векторное пространство $A(P) = P \otimes_K A$ можно превратить

в алгебру над P , определив умножение его элементов правилом

$$(\lambda \otimes u)(\mu \otimes v) = \lambda\mu \otimes uv.$$

Отождествляя каждый элемент $a \in A$ с элементом $1 \otimes a \in A(P)$, мы получаем вложение алгебры A в $A(P)$. Если $\{e_1, \dots, e_n\}$ — базис алгебры A над K , то умножение в A определяется формулами вида

$$e_i e_j = \sum_k c_{ijk} e_k.$$

Элементы $c_{ijk} \in K$ называются *структурными константами* алгебры A в базисе $\{e_1, \dots, e_n\}$. Те же формулы определяют умножение в алгебре $A(P)$ в базисе $\{e_1, \dots, e_n\}$. Однако смысл расширения основного поля состоит в том, что в алгебре $A(P)$ существуют другие базисы, в которых структурные константы могут иметь более простой вид.

Для того чтобы этот метод привел к каким-то результатам, нужно, конечно, заранее доказать инвариантность каких-то свойств алгебры при расширении основного поля.

Предложение 1. *Полупростая конечномерная ассоциативная алгебра A над полем K нулевой характеристики остается полупростой при переходе к любому расширению P поля K .*

Доказательство. Воспользуемся критерием полупростоты конечномерной ассоциативной алгебры, связанным со скалярным умножением (теорема 3.2). Очевидно, что в базисе, составленном из элементов алгебры A , матрица скалярного умножения в алгебре $A(P)$ такая же, как в алгебре A . Следовательно, она невырождена, а это означает, что алгебра $A(P)$ полупроста. \square

Пусть, например, L — конечное расширение поля K . Рассмотрим его как алгебру над K . Эта алгебра полупроста (и даже проста) и, согласно доказанному, для любого расширения P поля K алгебра $L(P)$ также полупроста и, следовательно, является прямой суммой нескольких конечных расширений поля P .

Пусть $\alpha \in L \setminus K$ — какой-либо элемент и h — его минимальный многочлен над K . Тогда

$$L \supset K[\alpha] \simeq K[x]/(h)$$

и, следовательно,

$$L(P) \supset P[\alpha] \simeq P[x]/(h).$$

Если многочлен h приводим над P и, в частности, если он имеет корень в P , то $P[\alpha]$ и, тем более, $L(P)$ — не поля. Поэтому последовательными простыми алгебраическими расширениями поля K мы можем получить такое конечное расширение P , что

$$L(P) \simeq \underbrace{P \oplus \dots \oplus P}_n \quad (n = \dim_K L). \quad (43)$$

Если $P \supset K$ — такое расширение, что имеет место изоморфизм (43), то говорят что L *расщепляется* над P .

В качестве примера применения этой идеологии докажем *теорему о примитивном элементе*.

Теорема 1. *Всякое конечное расширение L поля K нулевой характеристики является простым, т. е. порождается над K одним элементом.*

Доказательство. Пусть $\dim_K L = n$. Если L не порождается как алгебра над K одним элементом, то для любого $\alpha \in L$ элементы $1, \alpha, \alpha^2, \dots, \alpha^{n-1}$ линейно зависимы. Это можно записать в виде тождественного равенства нулю определителя, составленного из координат элементов $1, \alpha, \alpha^2, \dots, \alpha^{n-1}$ в каком-либо базисе L над K . Как функция от координат элемента α этот определитель представляет собой некоторый многочлен с коэффициентами из K . Если он равен нулю при всех значениях переменных в поле K , то он является нулевым многочленом и, следовательно, равен нулю и при всех значениях переменных в любом расширении P поля K , а это, в свою очередь, означает, что алгебра $L(P)$ не порождается над P одним элементом.

Однако если L расщепляется над P , то легко доказать, что алгебра $L(P)$ порождается одним элементом. В самом деле, рассмотрим элемент $\alpha = (\alpha_1, \dots, \alpha_n) \in \underbrace{P \oplus \dots \oplus P}_n$ с различными координатами $\alpha_1, \dots, \alpha_n \in P$. Тогда определитель, составленный из координат элементов $1, \alpha, \alpha^2, \dots, \alpha^{n-1}$, есть определитель Вандермонда для $\alpha_1, \dots, \alpha_n$ и, значит, отличен от нуля. \square

Обратимся теперь к изучению некоммутативных конечномерных алгебр с делением. Очевидно, что всякая алгебра с делением является простой.

Пусть D — центральная конечномерная алгебра с делением над полем K .

Предложение 2. *Алгебра D остается простой при переходе к любому расширению P поля K .*

Доказательство. Пусть $I \subset D(P)$ — ненулевой идеал. Рассмотрим самую короткую линейную комбинацию вида

$$a = \sum_{i=1}^s \lambda_i a_i \quad (\lambda_i \in P, a_i \in D),$$

принадлежащую I и отличную от нуля. Очевидно, что a_1, \dots, a_s линейно независимы над K , так как иначе число слагаемых можно было бы сократить. Точно так же, $\lambda_1, \dots, \lambda_s$ линейно независимы над K .

Домножив элемент a на a_1^{-1} (с любой стороны), что не выведет нас за пределы идеала I , добьемся того, чтобы $a_1 = 1$. Если при этом $s > 1$, то $a_2 \notin K = Z(D)$ и, значит, существует такой элемент $c \in D^*$, что $ca_2 \neq a_2c$. Имеем

$$a - cac^{-1} = \sum_{i=2}^s \lambda_i (a_i - ca_i c^{-1}) \in I,$$

причем $a - cac^{-1} \neq 0$, так как $\lambda_2, \dots, \lambda_s$ линейно независимы над K , а $a_2 - ca_2c^{-1} \neq 0$. Это противоречит определению элемента a . Следовательно, $s = 1$; но тогда $I \ni 1$ и, значит, $I = D(P)$. \square

Теорема 2. Существует такое конечное расширение P поля K , что $D(P) \simeq L_n(P)$ для некоторого $n \in \mathbb{N}$.

Следствие. $\dim D = n^2$.

Число n называется *степенью* алгебры D и обозначается через $\deg D$. Так, степень алгебры кватернионов равна 2.

Доказательство теоремы 2. Пусть \bar{K} — максимальное алгебраическое расширение поля K . (Существование такого расширения доказывается при помощи леммы Цорна.) Это алгебраически замкнутое поле, называемое *алгебраическим замыканием* поля K . По теореме 3.4

$$D(\bar{K}) \simeq L_n(\bar{K})$$

для некоторого $n \in \mathbb{N}$. Пусть $e_{ij} \in D(\bar{K})$ — элементы, соответствующие при этом изоморфизме матричным единицам, и $P \subset \bar{K}$ — подполе, порожденное над K координатами всех этих элементов в каком-либо базисе алгебры $D(\bar{K})$, составленном из элементов алгебры D . Очевидно, что P — конечное расширение поля K и $D(P) \simeq L_n(P)$. \square

Если $P \supset K$ — такое расширение, что $D(P) \simeq L_n(P)$, то говорят, что алгебра D *расщепляется* над P .

Пример 1. Алгебра кватернионов $D(\alpha, \beta)$ расщепляется над полем $P = K(\sqrt{\alpha}, \sqrt{\beta})$.

Важную информацию об алгебре D доставляет изучение ее максимальных коммутативных подалгебр или, что то же, максимальных подполей.

Теорема 3. *Всякое максимальное подполе F алгебры D имеет размерность n над K . Всякий изоморфизм максимальных подполей продолжается до внутреннего автоморфизма алгебры D .*

(Не утверждается, что все максимальные подполя изоморфны.)

Доказательство. Заметим, прежде всего, что если F — максимальная коммутативная подалгебра в D и P — любое расширение поля K , то $F(P)$ — максимальная коммутативная подалгебра в $D(P)$. В самом деле, условие максимальнойности означает, что F совпадает со своим централизатором

$$Z_D(F) = \{x \in D : ax = xa \ \forall a \in F\},$$

а это эквивалентно тому, что

$$\dim Z_D(F) = \dim F.$$

Но в координатах определение централизатора записывается как система однородных линейных уравнений с коэффициентами из K , а размерность пространства решений любой системы однородных линейных уравнений не меняется при расширении поля.

Пусть теперь \bar{K} — алгебраическое замыкание поля K . Тогда $D(\bar{K}) \simeq L_n(\bar{K})$ и

$$F(\bar{K}) \simeq \underbrace{\bar{K} \oplus \dots \oplus \bar{K}}_m,$$

где $m = \dim F$. Последнее означает, что в $F(\bar{K})$ имеется такой базис $\{e_1, \dots, e_m\}$, что

$$e_i^2 = e_i, \quad e_i e_j = 0 \quad \text{при } i \neq j.$$

Если рассматривать алгебру $L_n(\bar{K})$ как алгебру линейных операторов, то элементам e_1, \dots, e_m будут соответствовать попарно коммутирующие проекторы. В подходящем базисе эти проекторы одновременно записываются диагональными матрицами.

Отождествим алгебру $D(\bar{K})$ с $L_n(\bar{K})$ при помощи какого-либо фиксированного изоморфизма. Тогда из сказанного выше следует, что

существует такой элемент $c \in D(\bar{K})^*$, что $cF(\bar{K})c^{-1}$ состоит из диагональных матриц. Но так как $cF(\bar{K})c^{-1}$ — максимальная коммутативная подалгебра, то она совпадает с подалгеброй всех диагональных матриц. Следовательно, $m = n$, что доказывает первое утверждение теоремы.

Пусть теперь $F_1, F_2 \subset D$ — две максимальные коммутативные подалгебры и $\varphi: F_1 \xrightarrow{\sim} F_2$ — какой-либо изоморфизм. Тогда φ продолжается до изоморфизма

$$\bar{\varphi}: F_1(\bar{K}) \xrightarrow{\sim} F_2(\bar{K}).$$

Из предыдущего следует, что существует такой элемент $c \in D(\bar{K})^*$, что $cF_1(\bar{K})c^{-1} = F_2(\bar{K})$. Более того, так как всякий автоморфизм алгебры диагональных матриц есть просто перестановка диагональных элементов и, следовательно, индуцируется внутренним автоморфизмом алгебры $L_n(\bar{K}) = D(\bar{K})$, мы можем считать, что

$$cac^{-1} = \bar{\varphi}(a) \quad \text{при } a \in F_1(\bar{K}). \quad (44)$$

Нам нужно доказать существование такого ненулевого элемента $x \in D$, что $хах^{-1} = \varphi(a)$ при $a \in F_1$ или, что равносильно,

$$xa = \varphi(a)x \quad \forall a \in F_1.$$

В координатах эти условия записываются как система однородных линейных уравнений с коэффициентами из K . Из (44) следует, что эта система имеет ненулевое решение в \bar{K} ; но тогда она имеет ненулевое решение и в K . \square

Применим развитую теорию к описанию конечномерных алгебр с делением над полем \mathbb{R} и над конечными полями.

Теорема 4 (теорема Фробениуса). *Всякая конечномерная алгебра с делением D над полем \mathbb{R} изоморфна либо \mathbb{R} , либо \mathbb{C} , либо \mathbb{H} .*

Доказательство. Центр $Z(D)$ алгебры D , будучи конечным расширением поля \mathbb{R} , изоморфен либо \mathbb{R} , либо \mathbb{C} . Во втором случае D можно рассматривать как алгебру над \mathbb{C} и, поскольку поле \mathbb{C} алгебраически замкнуто, $D \simeq \mathbb{C}$.

Пусть теперь $Z(D) = \mathbb{R}$, т. е. D — центральная алгебра. Так как всякое максимальное подполе алгебры D изоморфно \mathbb{R} или \mathbb{C} , то $\deg D = 1$ или 2 . В первом случае $D = \mathbb{R}$. Рассмотрим второй случай.

Выберем какое-либо максимальное подполе в D и отождествим его с \mathbb{C} . По теореме 3 комплексное сопряжение в \mathbb{C} продолжается до внутреннего автоморфизма алгебры D , т. е. существует такой

элемент $j \in D$, что $jzj^{-1} = \bar{z}$ для любого $z \in \mathbb{C}$. Ясно, что $j \notin \mathbb{C}$. Следовательно, $D = \mathbb{C} \oplus \mathbb{C}j$. Далее, поскольку j^2 коммутирует с j и со всеми элементами из \mathbb{C} , то $j^2 \in Z(D) = \mathbb{R}$. Умножив j на подходящее вещественное число, добьемся того, чтобы $j^2 = \pm 1$. Случай $j^2 = 1$, однако, невозможен, ибо тогда $(j+1)(j-1) = 0$. Таким образом,

$$i^2 = j^2 = -1, \quad ij = -ji,$$

откуда следует, что $D \simeq \mathbb{H}$. \square

Теорема 5 (теорема Веддербёрна). *Всякое конечное тело коммутативно, т. е. является полем.*

Доказательство. Пусть D — конечное тело и K — его центр. Тогда D есть конечномерная центральная алгебра над K . Из первого утверждения теоремы 3 следует, что все максимальные подполя тела D содержат одно и то же число элементов и, значит, изоморфны, а из второго — что все они получаются друг из друга внутренними автоморфизмами тела D . Кроме того, всякий элемент α тела D содержится в подполе $K[\alpha]$ и, следовательно, в каком-то максимальном подполе.

Пусть F — какое-либо максимальное подполе тела D . Из предыдущего следует, что группа D^* покрывается подгруппами, сопряженными F^* . Число таких подгрупп равно $[D^* : N(F^*)]$ и, во всяком случае, не превосходит $[D^* : F^*]$. Следовательно,

$$|D^*| \leq |F^*| [D^* : F^*] = |D^*|.$$

Однако равенство невозможно хотя бы потому, что все эти подгруппы содержат единицу. Единственным исключением является тривиальный случай, когда $D^* = F^*$ и, значит, $D = F (= K)$. \square

В отличие от поля \mathbb{R} и конечных полей, над многими другими полями, например над полем \mathbb{Q} , существуют центральные алгебры с делением любой степени.

В качестве примера построим центральную алгебру с делением степени 3 над полем \mathbb{Q} . Пусть $F = \mathbb{Q}(\theta)$, где θ — корень неприводимого многочлена

$$f = t^3 - 3t + 1.$$

Дискриминант многочлена f равен $81 = 9^2$; следовательно, F — расширение Галуа степени 3 поля \mathbb{Q} (см. пример 10.6.4). Пусть σ — порождающий элемент группы $\text{Gal } F/\mathbb{Q}$. Рассмотрим формальные выражения вида

$$a_0 + a_1s + a_2s^2 \quad (a_0, a_1, a_2 \in F). \quad (45)$$

Определим их умножение, руководствуясь правилами дистрибутивности и ассоциативности и соотношениями

$$s^3 = 2, \quad sa = \sigma(a)s \quad (a \in F).$$

Мы получим некоторую 9-мерную некоммутативную алгебру D над \mathbb{Q} , содержащую поле F в качестве подалгебры. Докажем, что D — алгебра с делением.

Алгебра D может быть представлена матрицами 3-го порядка над полем F . А именно, поставив в соответствие каждому элементу $a \in F$ матрицу

$$T(a) = \begin{pmatrix} a & 0 & 0 \\ 0 & \sigma(a) & 0 \\ 0 & 0 & \sigma^2(a) \end{pmatrix},$$

мы получим вложение поля F в $L_3(F)$ в виде \mathbb{Q} -подалгебры. Далее, рассмотрим матрицу

$$S = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 2 & 0 & 0 \end{pmatrix}.$$

Легко проверить, что

$$S^3 = 2E, \quad ST(a) = T(\sigma(a))S.$$

Следовательно, матрицы вида

$$T(a_0) + T(a_1)S + T(a_2)S^2 = \begin{pmatrix} a_0 & a_1 & a_2 \\ 2\sigma(a_2) & \sigma(a_0) & \sigma(a_1) \\ 2\sigma^2(a_1) & 2\sigma^2(a_2) & \sigma^2(a_0) \end{pmatrix} \quad (46)$$

$(a_0, a_1, a_2 \in F)$ образуют \mathbb{Q} -подалгебру алгебры $L_3(F)$, изоморфную D . Она состоит из всех матриц $A \in L_3(F)$, удовлетворяющих условию

$$SAS^{-1} = \sigma(A), \quad (47)$$

где $\sigma(A)$ обозначает матрицу, получаемую применением σ ко всем элементам матрицы A .

Очевидно, что если матрица A , удовлетворяющая условию (47), невырождена, то матрица A^{-1} также удовлетворяет условию (47). Поэтому для доказательства того, что D — алгебра с делением, достаточно проверить, что всякая ненулевая матрица вида (46) невырождена.

Для доказательства последнего факта мы применим редукцию по модулю 2. Пусть O — кольцо целых чисел поля F . Очевидно, что O инвариантно относительно группы $\text{Gal } F/\mathbb{Q}$, так что если $a \in O$, то $T(a) \in L_3(O)$. Далее, O содержит подкольцо $\mathbb{Z}[\theta] = \{u_0 + u_1\theta + u_2\theta^2 : u_0, u_1, u_2 \in \mathbb{Z}\}$. Так как многочлен

$$[f]_2 = t^3 + t + 1 \in \mathbb{Z}_2[t]$$

неприводим над \mathbb{Z}_2 , то факторкольцо

$$\mathbb{Z}[\theta]/2\mathbb{Z}[\theta] \simeq \mathbb{Z}_2[t]/[f]_2\mathbb{Z}_2[t]$$

есть поле из 8 элементов. Имеется естественный гомоморфизм

$$\mathbb{Z}[\theta]/2\mathbb{Z}[\theta] \rightarrow O/2O. \quad (48)$$

Так как аддитивная группа кольца O изоморфна \mathbb{Z}^3 , то $|O/2O| = 8$. Отсюда следует, что гомоморфизм (48) есть на самом деле изоморфизм, так что кольцо $O/2O$ также является полем.

Путем умножения элемента (45) алгебры D на подходящее рациональное число можно добиться того, чтобы все числа a_0, a_1, a_2 принадлежали O , но хотя бы одно из них не принадлежало $2O$. Если при этом $a_0 \in 2O$, но $a_1 \notin 2O$, то умножением на $s^{-1} = s^2/2$ мы добьемся того, чтобы $a_0 \notin 2O$. Если $a_0, a_1 \in 2O$, но $a_2 \notin 2O$, то мы получим тот же результат умножением на $s^{-2} = s/2$. Таким образом, достаточно доказать обратимость элементов (45) алгебры D , для которых

$$a_0, a_1, a_2 \in O, \quad a_0 \notin 2O.$$

При этих условиях все элементы матрицы (46) принадлежат O и ее редукция по модулю 2 есть строго треугольная матрица над полем $O/2O$. Определитель последней матрицы отличен от нуля. Отсюда следует, что определитель матрицы (46) тем более отличен от нуля, что и требовалось доказать.

Так как $\dim D = 9$, то D — центральная алгебра с делением степени 3.

Алгебры с делением естественно возникают при рассмотрении неприводимых линейных представлений над алгебраически не замкнутыми полями. А именно, пусть $R: X \rightarrow L(V)$ — нетривиальное неприводимое представление множества X над полем K . Рассмотрим совокупность D всех эндоморфизмов представления R . Очевидно, что это подалгебра в $L(V)$. В силу теоремы 1.1 всякий ненулевой эндоморфизм представления R обратим. Следовательно, D — алгебра с делением.

Пространство V можно рассматривать как D -модуль или, как говорят, векторное пространство над телом D . (Легко видеть, что всякий конечно порожденный модуль над телом обладает базисом, как и векторное пространство над полем.) Множество $R(X)$ содержится в алгебре $L_D(V)$ линейных преобразований этого векторного пространства (изоморфной алгебре матриц над D). Нетрудно получить следующее обобщение теоремы 1.5: *подалгебра алгебры $L(V)$, порожденная множеством $R(X)$, совпадает с $L_D(V)$.*

Отсюда, в свою очередь, вытекает следующее обобщение теоремы 3.4: *всякая нетривиальная простая конечномерная ассоциативная алгебра над полем K изоморфна алгебре всех линейных преобразований некоторого векторного пространства над некоторой алгеброй с делением над K .*

В частности, неприводимые вещественные представления в силу теоремы Фробениуса разбиваются на три типа, для которых $D = \mathbb{R}, \mathbb{C}$ и \mathbb{H} соответственно.

Задача 4. Показать, что представления этих трех типов характеризуются тем, что их комплексификация неприводима, разлагается в сумму двух неизоморфных неприводимых представлений и разлагается в сумму двух изоморфных неприводимых представлений соответственно.

Если не требовать ассоциативности, то определение алгебры с делением следует видоизменить. Алгебра D (не обязательно ассоциативная) называется *алгеброй с делением*, если для любых $a, b \in D, a \neq 0$, каждое из уравнений $ax = b$ и $ya = b$ имеет решение.

Задача 5. Доказать, что для ассоциативных алгебр это определение эквивалентно данному выше.

Задача 6. Доказать, что утверждение задачи 1 справедливо и для неассоциативных алгебр.

При отказе от ассоциативности возникают новые интересные примеры алгебр с делением, даже в случае поля \mathbb{R} .

Приведем конструкцию алгебры октав \mathbb{O} , являющейся наиболее интересным и важным примером неассоциативной алгебры с делением над \mathbb{R} .

Пусть V — трехмерное векторное пространство над полем \mathbb{Z}_2 . Рассмотрим 8-мерную алгебру \mathbb{O} над \mathbb{R} с базисом $\{e_a : a \in V\}$ и таблицей умножения

$$e_a e_b = \varepsilon(a, b) e_{a+b},$$

где коэффициенты $\varepsilon(a, b)$, равные ± 1 , определяются в соответствии со следующими правилами:

- 1) $\varepsilon(0, b) = \varepsilon(a, 0) = 1$, так что $e_0 = 1$ является единицей алгебры \mathbb{O} ;
- 2) $\varepsilon(a, a) = -1$ при $a \neq 0$, так что квадрат каждой «мнимой единицы» e_a ($a \neq 0$) равен -1 ;
- 3) $\varepsilon(a, b) = -\varepsilon(b, a)$ при $a, b \neq 0, a \neq b$, так что мнимые единицы антикоммутируют;
- 4) $\varepsilon(a, b) = \varepsilon(b, c) = \varepsilon(c, a)$ при $a, b, c \neq 0, a + b + c = 0$, так что любые две мнимые единицы порождают подалгебру, изоморфную алгебре кватернионов;

5) $\varepsilon(a, b)\varepsilon(b, c)\varepsilon(c, d)\varepsilon(d, a) = -1$ при различных $a, b, c, d \neq 0$, $a + b + c + d = 0$.

Ненулевые векторы пространства V можно представлять как точки проективной плоскости PV над полем \mathbb{Z}_2 . При такой интерпретации условие 4) относится к тройкам точек, лежащих на одной прямой, а условие 5) — к четверкам точек, из которых никакие три не лежат на одной прямой.

Пример выбора коэффициентов $\varepsilon(a, b)$ ($a, b \neq 0$, $a \neq b$), удовлетворяющих условиям 3)—5), приведен на рис. 1, где прямые плоскости PV условно изображены шестью прямыми и одной окружностью, а стрелка, идущая от точки a к точке b , означает, что $\varepsilon(a, b) = 1$. (Остальные коэффициенты восстанавливаются в соответствии с правилами 3) и 4).) Несложно показать, что любой другой допустимый выбор коэффициентов $\varepsilon(a, b)$ приводится к этому за счет умножения некоторых мнимых единиц на -1 .

Построенная алгебра \mathbb{O} называется *алгеброй октав* или *алгеброй Кэли*.

Как и в случае кватернионов, линейное отображение $u \mapsto \bar{u}$, оставляющее на месте единицу и умножающее все мнимые единицы на -1 , является антиавтоморфизмом алгебры \mathbb{O} . Элемент $N(u) = u\bar{u}$, называемый *нормой* октавы u , принадлежит \mathbb{R} и равен сумме квадратов ее координат. Если $u \neq 0$, то $N(u) \neq 0$ и

$$u^{-1} = N(u)^{-1}\bar{u} \quad (49)$$

есть элемент, обратный u . Для установления всех этих свойств достаточно условий 1)—3), однако ввиду отсутствия ассоциативности они еще не гарантируют того, что \mathbb{O} — алгебра с делением.

При отсутствии ассоциативности для некоторых целей может оказаться достаточным более слабое свойство альтернативности. Алгебра называется *альтернативной*, если ассоциатор

$$[uvw] = (uv)w - u(vw)$$

любых ее элементов кососимметричен по u, v, w . В частности, отсюда следует, что если какие-либо два из трех перемножаемых элементов совпадают, то имеет место ассоциативность.

Задача 7. Доказать, что в альтернативной алгебре подалгебра, порожденная любыми двумя элементами, ассоциативна.

Проверим, что алгебра \mathbb{O} альтернативна. По соображениям линейности достаточно проверить кососимметричность ассоциатора любых базисных векторов e_a, e_b, e_c ($a, b, c \in V$). Если a, b, c линейно зависимы, то e_a, e_b, e_c принадлежат ассоциативной подалгебре и проверять нечего. Пусть a, b, c

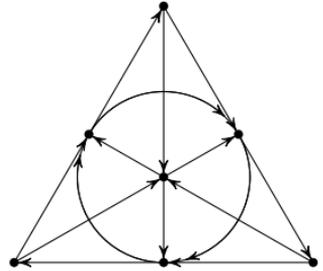


Рис. 1

линейно независимы. Докажем, что тогда не только ассоциатор, но и каждое из произведений $(e_a e_b) e_c$, $e_a (e_b e_c)$ кососимметрично по a, b, c . Рассмотрим, например, первое из них. Очевидно, что оно кососимметрично по a и b . Поэтому достаточно проверить, что оно кососимметрично по b и c . Пользуясь свойствами 3) и 4), получаем

$$(e_a e_b) e_c = \varepsilon(a, b) \varepsilon(a + b, c) e_{a+b+c} = -\varepsilon(a, b) \varepsilon(a + b + c, c) e_{a+b+c}$$

и, аналогично,

$$(e_a e_c) e_b = \varepsilon(a, c) \varepsilon(a + c, b) e_{a+b+c} = -\varepsilon(c, a) \varepsilon(b, a + b + c) e_{a+b+c};$$

но из свойства 5) следует, что

$$\varepsilon(a, b) \varepsilon(a + b + c, c) = -\varepsilon(c, a) \varepsilon(b, a + b + c),$$

откуда

$$(e_a e_b) e_c = -(e_a e_c) e_b.$$

Аналогично проверяется кососимметричность второго произведения.

Из формулы (49) для обратного элемента в алгебре \mathbb{O} вытекает, что

$$u^{-1} \in \langle 1, u \rangle,$$

и поэтому, если какие-либо два из трех перемножаемых элементов взаимно обратны, то, как и в случае их совпадения, имеет место ассоциативность. Следовательно, при $u \neq 0$ элемент $u^{-1}v$ является решением уравнения $ux = v$, а элемент vu^{-1} — решением уравнения $yu = v$, и, значит, \mathbb{O} — алгебра с делением.

Имеет место теорема: *всякая альтернативная конечномерная алгебра с делением над \mathbb{R} изоморфна либо \mathbb{R} , либо \mathbb{C} , либо \mathbb{H} , либо \mathbb{O} .*

Глава 12

Группы Ли

Определение группы Ли аналогично определению топологической группы. А именно, группой Ли называется группа G , снабженная структурой гладкого (дифференцируемого) многообразия таким образом, что групповые операции

$$\begin{aligned}\mu: G \times G &\rightarrow G, & (x, y) &\mapsto xy, \\ \iota: G &\rightarrow G, & x &\mapsto x^{-1},\end{aligned}$$

дифференцируемы. Иными словами, (локальные) координаты произведения должны быть дифференцируемыми функциями от (локальных) координат множителей и координаты обратного элемента должны быть дифференцируемыми функциями от координат самого элемента. Всякую группу Ли можно рассматривать как топологическую группу, но структура группы Ли богаче, чем структура топологической группы.

В приведенном определении под гладким многообразием можно понимать как вещественное, так и комплексное многообразие. В соответствии с этим получается определение вещественной или комплексной группы Ли. Мы будем рассматривать одновременно оба этих случая, обозначая через K поле \mathbb{R} или \mathbb{C} соответственно.

Примерами групп Ли служат аддитивная и мультипликативная группы поля K и группа невырожденных матриц $GL_n(K)$ (или, в геометрических терминах, группа $GL(V)$ обратимых линейных преобразований n -мерного векторного пространства V над полем K). В последнем случае координатами служат матричные элементы.

Теория групп Ли объединяет в себе алгебру, анализ и геометрию. Благодаря этому ее понятия и методы играют важную роль в большинстве разделов математики и теоретической физики.

§ 1. Определение и простейшие свойства групп Ли

Мы не будем пользоваться данным выше общим определением группы Ли. Ради простоты изложения мы ограничимся линейными

Более точно, пусть p_1, \dots, p_n — координаты точки p . Тогда существуют такая окрестность U точки (p_1, \dots, p_m) в пространстве K^m главных неизвестных и такая окрестность V точки (p_{m+1}, \dots, p_n) в пространстве K^d свободных неизвестных, что пересечение $M \cap \cap (U \times V)$ есть график дифференцируемого отображения $\varphi: V \rightarrow U$, задаваемого уравнениями (3), т. е. точка $(x_1, \dots, x_m, x_{m+1}, \dots, x_n) \in U \times V$ принадлежит M тогда и только тогда, когда выполнены соотношения (3).

Касательное пространство $T_p(M)$ многообразия M , заданного уравнениями (1), в точке $p \in M$ состоит из векторов $(dx_1, \dots, dx_n) \in K^n$, удовлетворяющих системе однородных линейных уравнений, получающихся дифференцированием в точке p уравнений (1):

$$df_i(p) = \sum_{j=1}^n \frac{\partial f_i}{\partial x_j}(p) dx_j = 0 \quad (i = 1, \dots, m). \quad (4)$$

Заметим, что условие, налагаемое на ранг якобиевой матрицы функций f_1, \dots, f_m , равносильно тому, что размерность пространства решений системы (4) равна $n - m$. Иногда это удобнее для проверки упомянутого условия, чем непосредственное вычисление ранга якобиевой матрицы. Если выполнено условие (2), то в качестве свободных неизвестных системы (4) могут быть взяты dx_{m+1}, \dots, dx_n .

Пространство $T_p(M)$ может быть описано как совокупность всех касательных векторов кривых на многообразии M , проходящих через точку p . Отсюда, в частности, следует, что оно не зависит от выбора системы уравнений, задающих многообразие M в окрестности точки p .

Подчеркнем, что мы понимаем касательное пространство $T_p(M)$ как подпространство векторного пространства K^n , а не как параллельную ему плоскость, проходящую через точку p .

Определение 1. *Линейной группой Ли* называется всякая подгруппа G группы $GL_n(K)$, являющаяся гладким многообразием в пространстве $L_n(K)$ всех матриц.

Так как всякая подгруппа $G \subset GL_n(K)$ инвариантна относительно умножений на свои элементы, являющихся линейными преобразованиями пространства $L_n(K)$, то условие, составляющее определение гладкого многообразия, достаточно проверить в какой-либо одной точке группы G , например, в единице e . (Единицей линейной группы является единичная матрица E .)

Задача 1. Если G — линейная группа Ли, то

$$T_g(G) = T_e(G)g$$

для любой матрицы $g \in G$.

В дальнейшем под группами Ли мы понимаем именно линейные группы Ли, а касательное пространство группы Ли G в единице обозначаем просто $T(G)$.

Пример 1. Сама группа $GL_n(K)$, будучи открытым подмножеством пространства $L_n(K)$, является n^2 -мерной группой Ли.

Пример 2. Группа $SL_n(K)$ является $(n^2 - 1)$ -мерной группой Ли. Действительно, дифференцируя в единице задающее ее уравнение $\det g = 1$, получаем линейное уравнение

$$d \det g = \operatorname{tr} dg = 0,$$

задающее $(n^2 - 1)$ -мерное пространство матриц с нулевым следом.

В § 6.5 мы дали определение производной матричной функции одной переменной. Точно так же определяются частные производные матричной функции нескольких переменных. Определим дифференциал матричной функции $\Phi = \Phi(x_1, \dots, x_n)$ по формуле

$$d\Phi = \sum_{i=1}^n \frac{\partial \Phi}{\partial x_i} dx_i.$$

Переходя к матричным элементам, легко доказать формулы

$$d(\Phi + \Psi) = d\Phi + d\Psi, \quad (5)$$

$$d(\Phi\Psi) = (d\Phi)\Psi + \Phi(d\Psi). \quad (6)$$

Используя последнюю формулу для вычисления дифференциала $d(\Phi\Phi^{-1})$, равного нулю, находим:

$$d(\Phi^{-1}) = -\Phi^{-1}(d\Phi)\Phi^{-1}.$$

Пример 3. Группа $O_n(K)$ задается матричным уравнением $gg^T = E$, которое ввиду очевидной симметрии можно рассматривать как систему $\frac{n(n+1)}{2}$ уравнений относительно матричных элементов. Дифференцируя в единице это уравнение, получаем линейное уравнение

$$d(gg^T) = dg + (dg)^T = 0,$$

задающее $\frac{n(n-1)}{2}$ -мерное пространство кососимметрических матриц. Так как

$$n^2 - \frac{n(n+1)}{2} = \frac{n(n-1)}{2},$$

то $O_n(K)$ есть $\frac{n(n-1)}{2}$ -мерная группа Ли. Заметим, что группу $O_n(\mathbb{R})$ обычно обозначают просто через O_n .

Задача 2. Доказать, что псевдоортогональная группа $O_{k,l}$, где $k+l=n$ (см. § 7.4), также является $\frac{n(n-1)}{2}$ -мерной группой Ли.

Задача 3. При четном $n=2m$ рассмотрим группу линейных преобразований пространства K^n , сохраняющих невырожденную кососимметрическую билинейную функцию

$$\alpha(x, y) = \sum_{i=1}^m (x_i y_{m+i} - x_{m+i} y_i).$$

Она называется *симплектической группой* и обозначается через $Sp_n(K)$. Доказать, что $Sp_n(K)$ есть $\frac{n(n+1)}{2}$ -мерная группа Ли.

Пример 4. Группа $B_n(K)$ невырожденных треугольных матриц, будучи открытым подмножеством в пространстве всех треугольных матриц, является $\frac{n(n+1)}{2}$ -мерной группой Ли.

Пример 5. Всякая дискретная (в частности, конечная) подгруппа группы $GL_n(K)$ является нульмерной группой Ли.

Мы будем также рассматривать вещественные группы Ли, состоящие из комплексных матриц, понимая под этим такие подгруппы группы $GL_n(\mathbb{C})$, которые являются гладкими многообразиями в пространстве $L_n(\mathbb{C})$, рассматриваемом как $2n^2$ -мерное вещественное векторное пространство.

Пример 6. Группа U_n задается в $L_n(\mathbb{C})$ матричным уравнением

$$gg^* = E \quad (7)$$

(где $g^* = \bar{g}^T$). Это уравнение можно рассматривать как систему n^2 вещественных уравнений относительно вещественных и мнимых частей элементов x_{ij} матрицы g :

$$\sum_k |x_{ik}|^2 = 1 \quad (i = 1, \dots, n),$$

$$\operatorname{Re} \sum_k x_{ik} \bar{x}_{jk} = \operatorname{Im} \sum_k x_{ik} \bar{x}_{jk} = 0 \quad (i, j = 1, \dots, n; i < j).$$

Дифференцируя в единице уравнение (7), получаем линейное уравнение

$$dg + (dg)^* = 0,$$

задающее n^2 -мерное пространство косоэрмитовых матриц. Так как $2n^2 - n^2 = n^2$, то U_n есть n^2 -мерная вещественная группа Ли.

Пример 7. Группа $SU_n = U_n \cap SL_n(\mathbb{C})$ является $(n^2 - 1)$ -мерной вещественной группой Ли (докажите это). Ее касательное пространство в единице состоит из косоэрмитовых матриц с нулевым следом.

Предложение 1. *Всякая группа Ли $G \subset GL_n(K)$ замкнута в $GL_n(K)$.*

Доказательство. Пусть \bar{G} — замыкание группы G в $GL_n(K)$. Из соображений непрерывности следует, что \bar{G} — подгруппа, а из определения гладкого многообразия — что G открыта в \bar{G} . Пусть теперь $g \in \bar{G}$. Тогда смежный класс gG открыт в \bar{G} и, следовательно, пересекается с G ; но это значит, что $gG = G$ и, в частности, $g \in G$. \square

Основной метод теории групп Ли состоит в переходе от рассмотрения группы G к рассмотрению ее касательного пространства $T_e(G) = T(G)$ (которое, как мы увидим, имеет структуру алгебры). Однако если, например, группа G дискретна, то ее касательное пространство равно нулю и не несет никакой информации о структуре группы G . В общем случае группа Ли достаточно хорошо контролируется своим касательным пространством в единице, только если она связна.

Напомним, что топологическое пространство называется *связным*, если его нельзя представить в виде объединения двух непересекающихся собственных замкнутых подмножеств. Объединение двух пересекающихся связных подмножеств топологического пространства M связно. Отсюда следует, что отношение « $x \sim y$, если x и y содержатся в некотором связном подмножестве» является отношением эквивалентности на M . Классы этой эквивалентности называются *связными компонентами* пространства M .

Если M — гладкое многообразие, то у каждой его точки есть связная окрестность (например, гомеоморфная шару). Отсюда следует, что связные компоненты многообразия M открыты в M . В то же время они замкнуты в M , так как каждая из них есть дополнение к объединению остальных.

Связную компоненту группы Ли G , содержащую единицу, обозначим через G° .

в O_n матрице вида (8). Следовательно, $O_n^\circ \supset SO_n$. С другой стороны, так как O_n есть объединение двух смежных классов по SO_n , каждый из которых, очевидно, замкнут, то $O_n^\circ = SO_n$.

Пример 11. Аналогичным образом доказывается, что группы U_n и SU_n связны.

Задача 4. Доказать, что группа $SO_{n,1}$ состоит из двух связных компонент, причем $SO_{n,1}^\circ$ есть подгруппа, образованная теми преобразованиями, которые оставляют на месте каждую из двух связных компонент гиперboloида

$$x_1^2 + \dots + x_n^2 - x_{n+1}^2 = -1.$$

(Указание: доказать, что всякое преобразование из группы $SO_{n,1}$, оставляющее на месте каждую связную компоненту указанного гиперboloида, есть произведение гиперболического поворота в двумерном подпространстве, содержащем базисный вектор e_{n+1} , и преобразования из группы SO_n (оставляющего на месте вектор e_{n+1}).)

Предложение 3. *Связная группа Ли порождается любой своей окрестностью единицы.*

Доказательство. Пусть U — окрестность единицы группы Ли G . Обозначим через \tilde{G} порожденную ею подгруппу. Для любого $g \in G$ подмножество gU , являющееся окрестностью элемента g в G , содержится в смежном классе $g\tilde{G}$. Это показывает, что все смежные классы группы G по \tilde{G} открыты в G . В то же время они замкнуты в G , так как каждый из них есть дополнение к объединению остальных. Следовательно, если группа G связна, то имеется только один смежный класс, т. е. $G = \tilde{G}$. \square

Замечание 2. Мы избрали в этой главе матричный язык. Однако ясно, что вместо матриц можно было бы говорить об определяемых ими линейных преобразованиях. Так, можно говорить о группе Ли $GL(V)$ невырожденных линейных преобразований n -мерного векторного пространства V над полем K , о группе Ли $O(V)$ ортогональных преобразований n -мерного евклидова пространства V и т. п. В некоторых случаях, когда это будет более удобно, мы будем переходить на язык линейных преобразований.

Замечание 3. Ввиду леммы 7.6.2 группа $GA(S)$ аффинных преобразований n -мерного аффинного пространства S естественным образом вкладывается в группу $GL(V)$ линейных преобразований $(n+1)$ -мерного векторного пространства V . В подходящем базисе

образ этого вложения записывается матрицами вида

$$\left(\begin{array}{ccc|c} a_{11} & \dots & a_{1n} & b_1 \\ \dots & \dots & \dots & \dots \\ a_{n1} & \dots & a_{nn} & b_n \\ \hline 0 & \dots & 0 & 1 \end{array} \right)$$

где матрица $A = (a_{ij})$ невырождена. Тем самым группу $GA(S)$, а также различные ее подгруппы, например, группу движений n -мерного евклидова пространства, можно рассматривать как линейные группы Ли.

§ 2. Экспоненциальное отображение

Связь группы Ли G с ее касательным пространством $T(G)$ осуществляется экспоненциальным отображением.

В § 6.5 была определена экспонента матрицы. Тем самым определено отображение

$$\exp: L_n(K) \rightarrow GL_n(K), \quad (9)$$

называемое *экспоненциальным отображением*.

Из определения экспоненты матрицы следует, что $\exp 0 = E$ и

$$\exp X = E + X + o(\|X\|). \quad (10)$$

Это показывает, что дифференциал экспоненциального отображения в нуле есть тождественное линейное отображение. В частности, якобиан экспоненциального отображения в нуле отличен от нуля, и по теореме о неявной функции отображение \exp осуществляет диффеоморфизм некоторой окрестности нуля в пространстве $L_n(K)$ на некоторую окрестность единицы в группе $GL_n(K)$. Обратное отображение (определенное в некоторой окрестности единицы группы $GL_n(K)$) обозначается через \log .

Замечание 1. Отображение \log задается рядом

$$\log(E + X) = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{X^n}{n},$$

абсолютно сходящимся при $\|X\| < 1$.

Замечание 2. Можно показать, что отображение \exp осуществляет диффеоморфизм открытого подмножества пространства $L_n(K)$, состоящего из

матриц, все комплексные собственные значения λ которых удовлетворяют условию $|\operatorname{Im} \lambda| < \pi$, на открытое подмножество группы $\operatorname{GL}_n(K)$, состоящее из матриц, не имеющих отрицательных собственных значений. В целом отображение (9) не является диффеоморфизмом.

Следующее предложение является обобщением известной формулы

$$e^a = \lim_{n \rightarrow \infty} \left(1 + \frac{a}{n}\right)^n.$$

Предложение 1. Пусть $g(t)$, $|t| < c$, — дифференцируемая кривая в группе $\operatorname{GL}_n(K)$, удовлетворяющая условиям

$$g(0) = E, \quad g'(0) = A. \quad (11)$$

Тогда

$$\exp A = \lim_{n \rightarrow \infty} g\left(\frac{1}{n}\right)^n. \quad (12)$$

Доказательство. Кривая $\log g(t)$ имеет при $t = 0$ тот же касательный вектор, что и $g(t)$, т. е. A . Это означает, что

$$\log g(t) = tA + o(t),$$

откуда

$$g(t) = \exp(tA + o(t))$$

и, в частности,

$$g\left(\frac{1}{n}\right) = \exp\left(\frac{A}{n} + o\left(\frac{1}{n}\right)\right).$$

Возводя последнее равенство в n -ю степень, получаем

$$g\left(\frac{1}{n}\right)^n = \exp(A + o(1)).$$

откуда и следует (12). □

Теорема 1. Пусть $G \subset \operatorname{GL}_n(K)$ — группа Ли. Тогда

$$\exp T(G) \subset G, \quad (13)$$

причем отображение \exp осуществляет диффеоморфизм некоторой окрестности нуля пространства $T(G)$ на некоторую окрестность единицы группы G .

Доказательство. Для любого $A \in T(G)$ существует кривая $g(t)$ в группе G , удовлетворяющая условиям (11). Так как группа G замкнута в $\operatorname{GL}_n(K)$ (см. предложение 1.1), то формула (12) показывает, что $\exp A \in G$.

Для того чтобы доказать второе утверждение теоремы, запишем отображение

$$\exp: T(G) \rightarrow G \quad (14)$$

во внутренних координатах окрестности единицы группы G и окрестности нуля пространства $T(G)$. В качестве таких координат можно взять свободные неизвестные системы уравнений, задающей группу G в окрестности единицы, и соответствующие свободные неизвестные системы однородных линейных уравнений, задающей пространство $T(G)$. (Напомним, что неизвестными в данном случае являются матричные элементы.) Мы получим дифференцируемое отображение f некоторой окрестности нуля пространства K^d (где $d = \dim G$) в это пространство. Из формулы (10) следует, что дифференциал отображения f в нуле есть тождественное линейное отображение, и по теореме о неявной функции отображение f осуществляет диффеоморфизм некоторой (быть может, меньшей) окрестности нуля пространства K^d на некоторую область этого пространства. Переходя к отображению \exp , мы получаем отсюда второе утверждение теоремы. \square

Пример 1. В случае $G = \mathrm{SL}_n(K)$ свойство (13) означает (см. пример 1.2), что если $\mathrm{tr} A = 0$, то $\det \exp A = 1$.

Пример 2. В случае $G = \mathrm{O}_n$ (соответственно U_n) свойство (13) означает (см. примеры 1.3 и 1.6), что экспонента кососимметричной (соответственно косоэрмитовой) матрицы является ортогональной (соответственно унитарной) матрицей. Впрочем, это легко проверить и непосредственно (попробуйте сделать это).

Теорема 2. *Связная группа Ли однозначно определяется своим касательным пространством в единице.*

Доказательство. Из предыдущей теоремы и предложения 1.3 следует, что связная группа Ли $G \subset \mathrm{GL}_n(K)$ совпадает с подгруппой, порожденной множеством $\exp T(G)$. \square

Подчеркнем, что в доказанной теореме ничего не говорится о существовании группы Ли с заданным касательным пространством. На самом деле далеко не всякое подпространство пространства матриц является касательным пространством группы Ли. Необходимое условие для этого будет указано в следующем параграфе.

Замечание 3. Вообще говоря, $G \neq \exp T(G)$, т. е. отображение (14) может не быть сюръективным (даже для связной группы Ли G). Например, для группы $G = \mathrm{SL}_2(\mathbb{R})$ пространство $T(G)$ состоит из матриц с нулевым следом.

Комплексные собственные значения такой матрицы A имеют вид $\lambda, -\lambda$, где либо $\lambda \in \mathbb{R}$, либо $\lambda \in i\mathbb{R}$. В обоих случаях

$$\operatorname{tr} \exp A = e^\lambda + e^{-\lambda} \geq -2.$$

Поэтому матрицы $g \in G$, для которых $\operatorname{tr} g < -2$ (например, $g = \begin{pmatrix} -2 & 0 \\ 0 & -1/2 \end{pmatrix}$), не принадлежат $\exp T(G)$.

Пусть $G \subset \operatorname{GL}_n(K)$, $H \subset \operatorname{GL}_m(K)$ — группы Ли. Отображение $f: G \rightarrow H$ называется *гомоморфизмом групп Ли*, если оно является гомоморфизмом групп и дифференцируемо, т. е. элементы матрицы $f(g)$ являются дифференцируемыми функциями от элементов матрицы $g \in G$. Дифференциал гомоморфизма f в единице есть линейное отображение пространства $T(G)$ в пространство $T(H)$. Мы будем обозначать его просто через df , не указывая явно, в какой точке берется дифференциал.

Теорема 3. Пусть $f: G \rightarrow H$ — гомоморфизм групп Ли. Тогда

$$f(\exp A) = \exp df(A) \quad (15)$$

для любого $A \in T(G)$.

Доказательство. Воспользуемся предложением 1. Пусть $g(t)$ — кривая в группе G , удовлетворяющая условиям (11). Тогда кривая $h(t) = f(g(t))$ в группе H удовлетворяет условиям

$$h(0) = E, \quad h'(0) = df(A).$$

Следовательно,

$$f(\exp A) = f\left(\lim_{n \rightarrow \infty} g\left(\frac{1}{n}\right)^n\right) = \lim_{n \rightarrow \infty} h\left(\frac{1}{n}\right)^n = \exp df(A). \quad \square$$

Пример 3. В применении к гомоморфизму

$$\det: \operatorname{GL}_n(\mathbb{C}) \rightarrow \mathbb{C}^* (= \operatorname{GL}_1(\mathbb{C}))$$

формула (15) означает, что

$$\det \exp A = e^{\operatorname{tr} A}$$

для любой матрицы $A \in \operatorname{L}_n(\mathbb{C})$ (см. пример 1.2).

Теорема 4. Гомоморфизм связной группы Ли в какую-либо группу Ли однозначно определяется своим дифференциалом в единице.

Доказательство. Пусть $f: G \rightarrow H$ — гомоморфизм групп Ли. Теоремы 1 и 3 показывают, что, зная df , мы можем найти $f(g)$ для элементов g из некоторой окрестности U единицы в группе G . Но если

группа G связна, то, согласно предложению 1.3, она порождается окрестностью U и, значит, мы можем найти $f(g)$ для всех $g \in G$. \square

В доказанной теореме ничего не говорится о существовании гомоморфизма групп Ли с заданным дифференциалом. На самом деле далеко не всякое линейное отображение касательных пространств является дифференциалом гомоморфизма групп Ли. Необходимое условие для этого будет указано в следующем параграфе.

Задача 1. Доказать, что ядро $\text{Ker } f$ гомоморфизма групп Ли $f: G \rightarrow H$ есть группа Ли, касательное пространство которой совпадает с $\text{Ker } df$.

Задача 2. В тех же обозначениях доказать, что если $\text{Im } df = T(H)$ и группа H связна, то $\text{Im } f = H$.

§ 3. Касательная алгебра Ли и присоединенное представление

Матрица

$$[A, B] = AB - BA$$

называется *коммутатором* матриц $A, B \in L_n(K)$. Не следует путать коммутатор в этом смысле с групповым коммутатором $(A, B) = ABA^{-1}B^{-1}$, определенным для невырожденных матриц; однако эти два понятия тесно связаны между собой.

Предложение 1. Для любых матриц $A, B \in L_n(K)$

$$[A, B] = \frac{\partial^2}{\partial t \partial s} (\exp tA, \exp sB) \Big|_{t=s=0}. \quad (16)$$

Доказательство. Дифференцируя групповой коммутатор

$$(\exp tA, \exp sB) = (\exp tA)(\exp sB)(\exp tA)^{-1}(\exp sB)^{-1}$$

по s при $s = 0$, получаем

$$\frac{\partial}{\partial s} (\exp tA, \exp sB) \Big|_{s=0} = (\exp tA)B(\exp tA)^{-1} - B.$$

Дифференцируя полученное выражение по t при $t = 0$, приходим к равенству

$$\frac{\partial^2}{\partial t \partial s} (\exp tA, \exp sB) \Big|_{t=s=0} = AB - BA = [A, B]. \quad \square$$

Теорема 1. Касательное пространство $T(G)$ группы Ли $G \subset \text{GL}_n(K)$ замкнуто относительно операции коммутирования, т. е.

$$A, B \in T(G) \Rightarrow [A, B] \in T(G).$$

Доказательство. При фиксированном t рассмотрим кривую

$$g(s) = (\exp tA, \exp sB) \in G.$$

Так как $g(0) = E$, то

$$g'(0) = \frac{\partial}{\partial s}(\exp tA, \exp sB) \Big|_{s=0} \in T(G).$$

Следовательно,

$$\frac{\partial^2}{\partial t \partial s}(\exp tA, \exp sB) \Big|_{t=s=0} = [A, B] \in T(G). \quad \square$$

Подпространство пространства матриц, замкнутое относительно операции коммутирования, называется *линейной алгеброй Ли*. Таким образом, касательное пространство $T(G)$ любой (линейной) группы Ли G является линейной алгеброй Ли. Эта алгебра Ли называется *касательной алгеброй группы G* .

Пример 1. Для группы $\text{SL}_n(K)$ доказанная теорема означает, что если $\text{tr } A = \text{tr } B = 0$, то и $\text{tr}[A, B] = 0$. На самом деле последнее равенство верно всегда. Иными словами,

$$\text{tr } AB = \text{tr } BA$$

для любых матриц A, B (см. § 5.4).

Пример 2. Для группы $\text{O}_n(K)$ доказанная теорема означает, что коммутатор двух кососимметричных матриц A, B также является кососимметричной матрицей. Это легко проверить и непосредственно:

$$[A, B]^T = (AB - BA)^T = B^T A^T - A^T B^T = BA - AB = -[A, B].$$

Операция коммутирования матриц *антикоммутативна*, т. е.

$$[A, B] + [B, A] = 0, \quad (17)$$

и удовлетворяет *тождеству Якоби*

$$[[A, B], C] + [[B, C], A] + [[C, A], B] = 0. \quad (18)$$

Последнее тождество легко проверяется непосредственным вычислением. Оно является следствием ассоциативности умножения матриц.

Любая алгебра, в которой выполняются тождества (17) и (18), называется *алгеброй Ли*. Так, пространство E^3 с операцией векторного умножения является алгеброй Ли (см. пример 1.7.5). Пространство $L_n(K)$ является алгеброй Ли относительно операции коммутирования. Теорема 1 означает, что касательное пространство $T(G)$ любой группы Ли $G \subset GL_n(K)$ является подалгеброй этой алгебры.

Теорема 2. *Дифференциал любого гомоморфизма групп Ли является гомоморфизмом их касательных алгебр.*

Доказательство. Пусть $f: G \rightarrow H$ — гомоморфизм групп Ли, и пусть $A, B \in T(G)$. Согласно теореме 2.3,

$$f((\exp tA, \exp sB)) = (f(\exp tA), f(\exp sB)) = (\exp t df(A), \exp s df(B)).$$

Как и в доказательстве теоремы 1, рассмотрим коммутатор элементов $\exp tA$ и $\exp sB$ при фиксированном t как кривую с параметром s в группе G , проходящую при $s = 0$ через единицу. Отображение df переводит касательный вектор этой кривой при $s = 0$ в касательный вектор ее образа в группе H . Таким образом,

$$df\left(\frac{\partial}{\partial s}(\exp tA, \exp sB)\Big|_{s=0}\right) = \frac{\partial}{\partial s}(\exp t df(A), \exp s df(B))\Big|_{s=0}.$$

Теперь, дифференцируя по t при $t = 0$ и учитывая, что линейное отображение df перестановочно с дифференцированием, мы получаем ввиду (16), что

$$df([A, B]) = [df(A), df(B)]. \quad \square$$

(Дифференцируемый) гомоморфизм группы Ли G в группу Ли $GL(V)$ называется *линейным представлением* группы G (как группы Ли) в пространстве V .

Для каждой группы Ли G имеется некоторое замечательное линейное представление в пространстве $T(G)$, играющее важную роль в теории групп Ли. Оно строится следующим образом.

Всякий элемент $g \in G$ определяет внутренний автоморфизм $a(g)$ группы G по формуле

$$a(g)x = gxg^{-1} \quad (x \in G). \quad (19)$$

Этот автоморфизм является ограничением на G линейного преобразования $X \mapsto gXg^{-1}$ пространства $L_n(K)$ и, в частности, дифференцируем. Его дифференциал в единице обозначается через $\text{Ad}(g)$ и называется *присоединенным оператором* элемента $g \in G$. Оператор $\text{Ad}(g)$

задается такой же формулой, что и $a(g)$:

$$\text{Ad}(g)X = gXg^{-1} \quad (X \in T(G)).$$

Так как $a(xy) = a(x)a(y)$, то

$$\text{Ad}(xy) = \text{Ad}(x) \text{Ad}(y).$$

(Впрочем, это легко проверить и непосредственно.) Далее, если $g = (g_{ij})$, $g^{-1} = (\tilde{g}_{ij})$, $X = (x_{ij})$, то $\text{Ad}(g)X = (y_{ij})$, где

$$y_{ij} = \sum_{k,l} g_{ik} x_{kl} \tilde{g}_{lj}. \quad (20)$$

В качестве координат в пространстве $T(G)$ могут быть взяты какие-то из матричных элементов; при этом остальные матричные элементы будут через них линейно выражаться. Формула (20) показывает, что матричные элементы оператора $\text{Ad}(g)$ являются рациональными и, следовательно, дифференцируемыми функциями от элементов матрицы g . Таким образом, отображение

$$\text{Ad}: G \rightarrow \text{GL}(T(G))$$

является линейным представлением группы Ли G в пространстве $T(G)$. Оно называется *присоединенным представлением* группы G .

Задача 1. Доказать, что если $f: G \rightarrow H$ — гомоморфизм групп Ли, то

$$f(\text{Ad}(g)X) = \text{Ad}(f(g))df(X) \quad (g \in G, X \in T(G)).$$

Задача 2. Доказать, что ядро присоединенного представления связанной группы Ли — это ее центр.

Гомоморфизм алгебры Ли L в алгебру Ли $L(V)$ линейных преобразований векторного пространства V (относительно операции коммутирования) называется *линейным представлением* алгебры L (как алгебры Ли) в пространстве V . Согласно теореме 2, дифференциал линейного представления группы Ли G является линейным представлением ее касательной алгебры Ли $T(G)$.

Дифференциал присоединенного представления Ad группы Ли G называется *присоединенным представлением* алгебры Ли $T(G)$ и обозначается символом ad .

Теорема 3. $\text{ad}(A)X = [A, X]$ ($A, X \in T(G)$).

Доказательство. Пусть $g(t)$ — кривая в группе G , удовлетворяющая условиям (11). Тогда

$$\text{ad}(A) = \left. \frac{d}{dt} \text{Ad}(g(t)) \right|_{t=0}.$$

Следовательно, для любого $X \in T(G)$

$$\operatorname{ad}(A)X = \left. \frac{d}{dt} \operatorname{Ad}(g(t))X \right|_{t=0} = \left. \frac{d}{dt} g(t)Xg(t)^{-1} \right|_{t=0} = AX - XA = [A, X]. \quad \square$$

Тот факт, что ad есть линейное представление алгебры $T(G)$, означает, что

$$\operatorname{ad}([A, B]) = [\operatorname{ad}(A), \operatorname{ad}(B)]$$

или, ввиду доказанной теоремы, что

$$[[A, B], C] = [A, [B, C]] - [B, [A, C]]$$

для любых $A, B, C \in T(G)$. Последнее тождество равносильно тождеству Якоби. Полученный результат можно рассматривать как концептуальное (т. е. основанное на понимании существа дела, а не на простом вычислении) доказательство тождества Якоби для операции коммутирования матриц.

С другой стороны, приведенное выше рассуждение подсказывает возможность определения присоединенного представления ad любой алгебры Ли L (не обязательно связанной с какой-либо группой Ли) по формуле

$$\operatorname{ad}(a)x = [a, x] \quad (a, x \in L).$$

Пример 3. Присоединенное представление алгебры Ли (\mathbb{E}^3, \times) определяется по формуле

$$\operatorname{ad}(a)x = a \times x.$$

Из кососимметричности смешанного произведения $(a, b, c) = (a \times b, c)$ следует, что оператор $\operatorname{ad}(a)$ кососимметричен. Тем самым определяется гомоморфизм алгебры Ли (\mathbb{E}^3, \times) в алгебру Ли $T(\operatorname{SO}_3)$ кососимметричных матриц 3-го порядка. Легко видеть, что ядро этого гомоморфизма тривиально. Так как обе алгебры трехмерны, то построенный гомоморфизм является изоморфизмом.

Задача 3. Выписать матрицы присоединенных операторов векторов ортонормированного базиса пространства \mathbb{E}^3 в этом же базисе.

Задача 4. Доказать, что для любых матриц $A, X \in L_n(K)$ справедливо равенство

$$(\exp A)X(\exp A)^{-1} = \sum_{k=0}^{\infty} \frac{1}{k!} \underbrace{[A, [A, \dots, [A, X] \dots]]}_k.$$

Задача 5. Центром алгебры Ли L называется подалгебра

$$Z(L) = \{z \in L : [z, u] = 0 \ \forall u \in L\}.$$

Доказать, что центр связной группы Ли G есть группа Ли, касательная алгебра которой совпадает с центром алгебры Ли $T(G)$. (Указание: воспользоваться задачами 2 и 2.1.)

Пример 4. Рассмотрим присоединенное представление группы Ли SU_2 . Алгебра Ли $T(SU_2)$ состоит из косоэрмитовых матриц с нулевым следом, т. е. матриц вида

$$X = \begin{pmatrix} ix_1 & x_2 + ix_3 \\ -x_2 + ix_3 & -ix_1 \end{pmatrix}. \quad (21)$$

Заметим, что

$$\det X = x_1^2 + x_2^2 + x_3^2.$$

Следовательно, $\det X$ — положительно определенная квадратичная функция в пространстве $T(SU_2)$. Приняв ее за скалярный квадрат, будем рассматривать $T(SU_2)$ как (трехмерное) евклидово пространство. Так как

$$\det \text{Ad}(g)X = \det gXg^{-1} = \det X,$$

то присоединенные операторы элементов группы SU_2 ортогональны, т. е. $\text{Ad}(SU_2) \subset O_3$. Так как группа SU_2 связна, то и ее образ связан и, значит, $\text{Ad}(SU_2) \subset SO_3$.

Далее, Ker Ad состоит из матриц, коммутирующих со всеми матрицами вида (21). Пользуясь тем, что всякая матрица, коммутирующая с матрицей $\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$, диагональна, легко доказать, что

$$\text{Ker Ad} = \{\pm E\}.$$

Аналогично доказывается, что $\text{Ker ad} = 0$. Так как $\dim SU_2 = \dim SO_3 = 3$, то $\text{Im ad} = T(SO_3)$, и из теорем 2.3 и 2.1 и предложения 1.3 следует, что

$$\text{Ad}(SU_2) = SO_3.$$

Таким образом, присоединенное представление осуществляет гомоморфизм группы SU_2 на группу SO_3 с ядром $\{\pm E\}$.

Отметим, что группа SU_2 состоит из матриц вида

$$\begin{pmatrix} a & -\bar{b} \\ b & \bar{a} \end{pmatrix}, \quad a, b \in \mathbb{C}, \quad |a|^2 + |b|^2 = 1, \quad (22)$$

и совпадает с группой кватернионов с нормой 1 в матричной модели алгебры \mathbb{H} (см. задачу 1.8.4). Таким образом, группа SU_2 представляет собой (трехмерную) сферу в 4-мерном евклидовом пространстве \mathbb{H} . Группа SO_3 получается из нее отождествлением диаметрально противоположных точек и тем самым представляет собой трехмерное вещественное проективное пространство. Пространство $T(SU_2)$ совпадает с пространством чисто мнимых кватернионов.

Задача 6. Рассуждая аналогичным образом, доказать, что присоединенное представление осуществляет гомоморфизм группы $SL_2(\mathbb{R})$ на связную компоненту группы $SO_{2,1}$ (см. задачу 1.2) с ядром $\{\pm E\}$.

Задача 7. Доказать, что присоединенное представление осуществляет гомоморфизм группы $SL_2(\mathbb{C})$ на группу $SO_3(\mathbb{C})$ с ядром $\{\pm E\}$.

§ 4. Линейные представления групп Ли

Линейные представления групп Ли весьма хорошо изучены, но мы здесь имеем возможность рассказать лишь о некоторых начальных идеях этой теории. Основная из них состоит в том, чтобы заменять изучение линейных представлений групп Ли изучением линейных представлений их касательных алгебр Ли.

Пусть G — какая-то группа Ли и

$$R: G \rightarrow GL(V)$$

— какое-то ее (конечномерное) линейное представление. Тогда

$$dR: T(G) \rightarrow L(V)$$

есть линейное представление алгебры Ли $T(G)$.

Теорема 1. *Всякое подпространство $U \subset V$, инвариантное относительно G , инвариантно относительно $T(G)$. Если группа G связна, то верно и обратное: всякое подпространство, инвариантное относительно $T(G)$, инвариантно также относительно G .*

Доказательство. 1) Пусть подпространство U инвариантно относительно G , и пусть $A \in T(G)$. Возьмем кривую $g(t)$ в G , удовлетворяющую условиям (11). Тогда

$$dR(A) = \left. \frac{d}{dt} R(g(t)) \right|_{t=0}$$

и, следовательно, для любого вектора $u \in U$

$$dR(A)u = \left. \frac{d}{dt} R(g(t))u \right|_{t=0} \in U.$$

2) Обратно, пусть U инвариантно относительно $T(G)$. По теореме 2.3 для любого $A \in T(G)$

$$R(\exp A) = \exp dR(A)$$

и, следовательно, для любого $u \in U$

$$R(\exp A)u = \sum_{k=0}^{\infty} \frac{1}{k!} dR(A)^k u \in U.$$

Таким образом, подпространство U инвариантно относительно $\exp T(G)$. Если группа G связна, то отсюда следует, что оно инвариантно относительно всей группы G . \square

Таким образом, если группа G связна, то набор инвариантных подпространств для представления R группы G и для представления dR алгебры $T(G)$ один и тот же.

Следствие. *Линейное представление R связной группы Ли G неприводимо (соответственно вполне приводимо) тогда и только тогда, когда представление dR алгебры Ли $T(G)$ неприводимо (соответственно вполне приводимо).*

Задача 1. Пусть G — связная группа Ли и H — ее связная подгруппа Ли. Доказать эквивалентность следующих условий:

- 1) H — нормальная подгруппа группы G ;
- 2) подпространство $T(H)$ инвариантно относительно присоединенного представления группы G ;
- 3) $T(H)$ — идеал алгебры $T(G)$.

Связная группа Ли называется *простой*, если она не содержит нетривиальных связных нормальных подгрупп Ли. Алгебра Ли называется *простой*, если она не содержит нетривиальных идеалов.

Задача 2. Доказать, что если касательная алгебра связной группы Ли G проста, то и группа G проста (как группа Ли). (На самом деле верно и обратное.)

Задача 3. Доказать, что группа Ли SO_3 проста. (На самом деле группа SO_3 не имеет никаких нетривиальных нормальных подгрупп. См. § 10.5.)

Классификация простых групп Ли имеет для теории групп Ли такое же значение, какое имеет классификация простых конечных групп для теории конечных групп. Она была получена в конце XIX — начале XX века В. Киллингом и Э. Картаном (сначала для комплексных групп Ли, затем для вещественных). Это одно из самых поразительных достижений математики.

Можно доказать, что группа Ли SO_n проста при любом $n \geq 5$. Однако группа Ли SO_4 , как вытекает из следующего ниже примера, простой не является.

Пример 1. Как мы видели в примере 3.4, группа SU_2 может быть интерпретирована как группа кватернионов с нормой 1. Рассмотрим линейное представление R группы Ли $G = SU_2 \times SU_2$ в пространстве \mathbb{H} , определяемое по формуле

$$R(p, q)x = pxq^{-1} \quad (x \in \mathbb{H}).$$

Так как

$$N(pxq^{-1}) = N(p)N(x)N(q)^{-1} = N(x)$$

при $p, q \in SU_2$, то $R(G) \subset O_4$. Из соображений связности следует, что $R(G) \subset SO_4$. Тем самым определен гомоморфизм

$$R: SU_2 \times SU_2 \rightarrow SO_4.$$

Если $(p, q) \in \text{Ker } R$, то, в частности, $R(p, q)1 = pq^{-1} = 1$, откуда $p = q$. Далее, как и в примере 3.4, получаем, что $p = q = \pm 1$. Так как $\dim G = \dim SO_4 = 6$, то отсюда следует, что $R(G) = SO_4$. Таким образом,

$$SO_4 \simeq (SU_2 \times SU_2) / \{(E, E), (E, -E)\}.$$

В частности, каждый из множителей произведения $SU_2 \times SU_2$ при гомоморфизме R переходит в связную нормальную подгруппу Ли группы SO_4 и, следовательно, группа SO_4 не является простой группой Ли.

Задача 4. Доказать, что группа Ли $SL_n(K)$ проста при любом $n \geq 2$.

Комплексные и вещественные группы Ли находятся в тесной связи.

Пусть G — связная комплексная группа Ли.

Определение 1. Связная вещественная подгруппа Ли $H \subset G$ называется *вещественной формой* группы G , если

$$T(G) = T(H) \oplus iT(H).$$

Замечание 1. Это определение можно распространить на несвязные группы Ли, дополнительно потребовав, чтобы каждая связная компонента группы G пересекалась с H .

Пример 2. Группа $SL_n(\mathbb{R})$ является вещественной формой группы $SL_n(\mathbb{C})$.

Пример 3. Группа SU_n также является вещественной формой группы $SL_n(\mathbb{C})$, а группа U_n — вещественной формой группы $GL_n(\mathbb{C})$. Это следует из того, что всякая комплексная матрица единственным образом представляется в виде суммы косоэрмитовой и эрмитовой матриц, а пространство эрмитовых матриц получается из пространства косоэрмитовых матриц умножением на i .

Пример 4. Группа SO_n является вещественной формой группы $SO_n(\mathbb{C})$ (которая, как можно доказать, связна).

Теорема 2. Пусть $R: G \rightarrow GL(V)$ — комплексное линейное представление связной комплексной группы Ли G , и пусть H — вещественная форма группы G . Тогда набор инвариантных подпространств для $R(G)$ и для $R(H)$ один и тот же.

Доказательство. Согласно теореме 1, подпространство $U \subset V$ инвариантно относительно G (соответственно H) тогда и только тогда, когда оно инвариантно относительно $T(G)$ (соответственно $T(H)$). Но так как

$$dR(T(G)) = dR(T(H)) + idR(T(H)),$$

то инвариантность подпространства U относительно $T(G)$ равносильна его инвариантности относительно $T(H)$. \square

Этой теоремой можно воспользоваться, чтобы доказать полную приводимость линейных представлений некоторых комплексных групп Ли.

Определение 2. Связная комплексная группа Ли называется *редуктивной*, если она обладает компактной вещественной формой.

Так, в силу приведенных выше примеров группы $GL_n(\mathbb{C})$, $SL_n(\mathbb{C})$ и $SO_n(\mathbb{C})$ редуктивны. Можно доказать (но это непросто), что всякая некоммутативная простая комплексная группа Ли редуктивна.

Замечание 2. Более естественно в определении редуктивной группы не требовать связности: тогда в число редуктивных групп войдут все конечные группы.

Из теоремы 2 и доказанной в § 11.2 полной приводимости линейных представлений компактных групп немедленно вытекает

Теорема 3. Всякое линейное представление редуктивной комплексной группы Ли вполне приводимо.

Этот прием доказательства, принадлежащий Г. Вейлю, носит название «унитарный трюк». С его помощью можно доказывать и другие теоремы. Например, он позволяет распространить на редуктив-

ные группы теорему Гильберта о конечной порожденности алгебр инвариантов, доказанную в § 11.5 для компактных групп.

Пользуясь изложенной выше теорией, найдем все неприводимые линейные представления группы Ли $SL_2(\mathbb{C})$. Этот пример играет ключевую роль в теории линейных представлений произвольных простых групп Ли.

Обозначим касательную алгебру Ли группы $SL_2(\mathbb{C})$ через $sl_2(\mathbb{C})$. Выберем в ней базис из матриц

$$H = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad E_+ = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad E_- = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}. \quad (23)$$

Непосредственно проверяется, что

$$[H, E_+] = 2E_+, \quad [H, E_-] = -2E_-, \quad [E_+, E_-] = H. \quad (24)$$

Пусть $R: SL_2(\mathbb{C}) \rightarrow GL(V)$ — какое-то линейное представление. Положим

$$dR(H) = \mathcal{H}, \quad dR(E_+) = \mathcal{E}_+, \quad dR(E_-) = \mathcal{E}_-.$$

Операторы $\mathcal{H}, \mathcal{E}_+, \mathcal{E}_-$ удовлетворяют соотношениям

$$[\mathcal{H}, \mathcal{E}_+] = 2\mathcal{E}_+, \quad [\mathcal{H}, \mathcal{E}_-] = -2\mathcal{E}_-, \quad [\mathcal{E}_+, \mathcal{E}_-] = \mathcal{H},$$

вытекающим из соотношений (24).

Лемма 1. Если v — собственный вектор оператора \mathcal{H} с собственным значением λ , то вектор \mathcal{E}_+v (соответственно \mathcal{E}_-v), если он не равен нулю, — собственный вектор оператора \mathcal{H} с собственным значением $\lambda + 2$ (соответственно $\lambda - 2$).

Доказательство. Имеем

$$\mathcal{H} \mathcal{E}_+v = \mathcal{E}_+ \mathcal{H}v + [\mathcal{H}, \mathcal{E}_+]v = \lambda \mathcal{E}_+v + 2\mathcal{E}_+v = (\lambda + 2)\mathcal{E}_+v.$$

Аналогично доказывается, что

$$\mathcal{H} \mathcal{E}_-v = (\lambda - 2)\mathcal{E}_-v. \quad \square$$

Лемма 2. Существует собственный вектор v_0 оператора \mathcal{H} , для которого $\mathcal{E}_+v_0 = 0$.

Доказательство. Так как оператор \mathcal{H} имеет лишь конечное число собственных значений, то существует такое его собственное значение λ_0 , что $\lambda_0 + 2$ не является собственным значением. Соответствующий собственный вектор v_0 и будет искомым. \square

Всякий такой вектор v_0 называется старшим вектором для представления R .

Пусть v_0 — старший вектор и λ_0 — соответствующее собственное значение оператора \mathcal{H} . Рассмотрим векторы

$$v_k = \mathcal{E}_-^k v_0 \quad (k = 0, 1, 2, \dots).$$

По лемме 1

$$\mathcal{H}v_k = (\lambda_0 - 2k)v_k.$$

Лемма 3. $\mathcal{E}_+v_k = c_k v_{k-1}$, где $c_k = k(\lambda_0 - k + 1)$.

Доказательство. Докажем эту формулу индукцией по k . Она верна при $k=0$, если считать, что $v_{-1} = 0$. Предположим, что она верна для некоторого k . Тогда

$$\begin{aligned} \mathcal{E}_+v_{k+1} &= \mathcal{E}_+\mathcal{E}_-v_k = \mathcal{E}_-\mathcal{E}_+v_k + [\mathcal{E}_+, \mathcal{E}_-]v_k = \\ &= c_k \mathcal{E}_-v_{k-1} + \mathcal{H}v_k = c_k v_k + (\lambda_0 - 2k)v_k = c_{k+1}v_k, \end{aligned}$$

где

$$c_{k+1} = c_k + \lambda_0 - 2k = (k+1)\lambda_0 - k^2 - k = (k+1)(\lambda_0 - k). \quad \square$$

Так как собственные векторы оператора \mathcal{H} , отвечающие различным собственным значениям, линейно независимы, то существует такое n , что векторы $v_0, v_1, v_2, \dots, v_n$ отличны от нуля и линейно независимы, а $v_{n+1} = 0$. Из леммы 3 следует тогда, что $c_{n+1} = 0$, т. е. $\lambda_0 = n$.

Далее, как следует из доказанных формул, линейная оболочка векторов $v_0, v_1, v_2, \dots, v_n$ инвариантна относительно операторов $\mathcal{H}, \mathcal{E}_+, \mathcal{E}_-$ и, значит, относительно всей алгебры $\mathfrak{sl}_2(\mathbb{C})$.

Если представление R неприводимо, то

$$V = \langle v_0, v_1, \dots, v_n \rangle \quad (25)$$

и операторы $\mathcal{H}, \mathcal{E}_+, \mathcal{E}_-$ задаются в базисе $\{v_0, v_1, \dots, v_n\}$ формулами

$$\mathcal{H}v_k = (n - 2k)v_k, \quad (26)$$

$$\mathcal{E}_+v_k = k(n - k + 1)v_{k-1}, \quad (27)$$

$$\mathcal{E}_-v_k = v_{k+1}, \quad (28)$$

если условиться, что $v_{-1} = v_{n+1} = 0$. Число n называется *старшим весом* представления R . Формулы (26)—(28) показывают, что неприводимое представление группы $\mathrm{SL}_2(\mathbb{C})$ полностью определяется своим старшим весом.

Обратно, если выполнено условие (25), то представление R неприводимо. В самом деле, всякое ненулевое инвариантное подпространство инвариантно, в частности, относительно оператора \mathcal{H} и, следовательно, является линейной оболочкой каких-то из его собственных векторов v_0, v_1, \dots, v_n . Но, будучи инвариантным также относительно операторов \mathcal{E}_+ и \mathcal{E}_- , оно должно содержать все эти векторы, т. е. совпадать с V .

Остается вопрос о существовании неприводимого представления с заданным старшим весом. Можно проверить прямым вычислением, что операторы $\mathcal{H}, \mathcal{E}_+, \mathcal{E}_-$, определяемые формулами (26)—(28), задают линейное представление алгебры Ли $\mathfrak{sl}_2(\mathbb{C})$, и затем, пользуясь общей теоремой, которой не было в данном курсе, доказать существование линейного представ-

ления группы $SL_2(\mathbb{C})$, имеющего своим дифференциалом это представление. Мы поступим иначе, а именно, построим нужное представление явно.

Будем считать, что группа $SL_2(\mathbb{C})$ тавтологически действует в пространстве \mathbb{C}^2 с базисом $\{x, y\}$, т. е. элемент

$$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{C})$$

действует по формулам

$$\begin{cases} gx = ax + cy, \\ gy = bx + dy. \end{cases} \quad (29)$$

Это действие индуцирует линейное действие группы $SL_2(\mathbb{C})$ в пространстве $S^n(\mathbb{C}^2)$. Если рассматривать x и y как координаты в сопряженном пространстве $(\mathbb{C}^2)^*$, а симметрическую алгебру пространства \mathbb{C}^2 отождествить с алгеброй многочленов на $(\mathbb{C}^2)^*$ (см. § 8.3), то элементы пространства $S^n(\mathbb{C}^2)$ будут однородными многочленами степени n от x и y или, как говорят, *бинарными формами* степени n , а действие группы $SL_2(\mathbb{C})$ можно будет понимать как линейную замену переменных по формулам (29). Полученное таким образом линейное представление обозначим через R_n . Согласно определению,

$$(R_n(g)f)(x, y) = f(ax + cy, bx + dy)$$

для любой бинарной формы f степени n .

Вычислим дифференциал представления R_n . Положим

$$dR_n(H) = \mathcal{H}, \quad dR(E_+) = \mathcal{E}_+, \quad dR(E_-) = \mathcal{E}_-.$$

Принимая во внимание, что

$$\exp tH = \begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix}, \quad \exp tE_+ = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}, \quad \exp tE_- = \begin{pmatrix} 1 & 0 \\ t & 1 \end{pmatrix},$$

получаем для бинарной формы $f \in S^n(\mathbb{C}^2)$

$$(\mathcal{H}f)(x, y) = \frac{d}{dt} f(e^t x, e^{-t} y) \Big|_{t=0} = x \frac{\partial f(x, y)}{\partial x} - y \frac{\partial f(x, y)}{\partial y},$$

$$(\mathcal{E}_+ f)(x, y) = \frac{d}{dt} f(x, y + tx) \Big|_{t=0} = x \frac{\partial f(x, y)}{\partial y},$$

$$(\mathcal{E}_- f)(x, y) = \frac{d}{dt} f(x + ty, y) \Big|_{t=0} = y \frac{\partial f(x, y)}{\partial x}.$$

В частности, для $f_0 = x^n$

$$\mathcal{H}f_0 = nf_0, \quad \mathcal{E}_+ f_0 = 0,$$

т. е. f_0 является старшим вектором представления R_n с собственным значением n . При этом

$$f_k = \mathcal{E}_-^k f_0 = n(n-1)\dots(n-k+1)x^{n-k}y^k,$$

так что формы f_0, f_1, \dots, f_n составляют базис пространства $S^n(\mathbb{C}^2)$. Следовательно, R_n — неприводимое представление со старшим весом n .

Одновременно мы доказали, что всякое неприводимое линейное представление алгебры Ли $\mathfrak{sl}_2(\mathbb{C})$ является дифференциалом некоторого (неприводимого) линейного представления группы $SL_2(\mathbb{C})$.

Полученные результаты дают также описание неприводимых комплексных линейных представлений групп Ли $SL_2(\mathbb{R})$ и SU_2 , являющихся вещественными формами группы $SL_2(\mathbb{C})$. В самом деле, если H — вещественная форма группы $SL_2(\mathbb{C})$ и $S: H \rightarrow GL(V)$ — ее неприводимое комплексное линейное представление, то представление dS алгебры Ли $T(H)$ однозначно продолжается до линейного представления алгебры Ли $\mathfrak{sl}_2(\mathbb{C})$, которое по доказанному выше является дифференциалом некоторого линейного представления группы $SL_2(\mathbb{C})$. Отсюда следует, что неприводимые комплексные линейные представления группы H — это в точности ограничения на H неприводимых линейных представлений группы $SL_2(\mathbb{C})$.

Задача 5. Доказать, что при $n > 0$

$$\text{Ker } R_n = \begin{cases} \{E\}, & \text{если } n \text{ нечетно,} \\ \{\pm E\}, & \text{если } n \text{ четно.} \end{cases}$$

Задача 6. Описать неприводимые линейные представления группы Ли $SO_3(\mathbb{C})$.

Задача 7. Описать неприводимые комплексные линейные представления групп Ли SO_3 и $SO_{2,1}^\circ$.

Ответы к задачам

1.8.1. $E_{ij}E_{kl} = \delta_{jk}E_{il}$ (где δ_{ij} — символ Кронекера).

2.2.1. q^n .

2.2.3. $(q^n - 1)(q^n - q)(q^n - q^2) \dots (q^n - q^{n-1})$.

2.2.4. $\frac{(q^n - 1)(q^n - q)(q^n - q^2) \dots (q^n - q^{k-1})}{(q^k - 1)(q^k - q)(q^k - q^2) \dots (q^k - q^{k-1})}$.

3.5.4. $2 = (1+i)(1-i) \sim (1+i)^2$, 3 — простой элемент, $5 = (2+i) \times (2-i)$.

3.5.5. $x, x+1, x^2+x+1, x^3+x^2+1, x^3+x+1, x^4+x^3+1, x^4+x+1, x^4+x^3+x^2+x+1$.

4.6.3. $|\mathrm{GL}_2(\mathbb{Z}_p)| = p(p+1)(p-1)^2$, $|\mathrm{SL}_2(\mathbb{Z}_p)| = p(p^2-1)$.

5.1.1. Три различных одномерных подпространства двумерного векторного пространства.

5.1.2. То же, что и в задаче 5.1.1.

5.2.1. 3.

5.4.2. $\left(\frac{n(n+1)}{2}, \frac{n(n-1)}{2} \right)$.

5.5.1.
$$\begin{vmatrix} 1 & -\cos \alpha_{12} & -\cos \alpha_{13} & -\cos \alpha_{14} \\ -\cos \alpha_{12} & 1 & -\cos \alpha_{23} & -\cos \alpha_{24} \\ -\cos \alpha_{13} & -\cos \alpha_{23} & 1 & -\cos \alpha_{34} \\ -\cos \alpha_{14} & -\cos \alpha_{24} & -\cos \alpha_{34} & 1 \end{vmatrix} = 0$$
, где α_{ij} — угол

между i -й и j -й гранями; двугранный угол правильного тетраэдра равен $\arccos 1/3$.

5.6.1. $\sum_k \bar{c}_{ki}c_{kj} = \delta_{ij}$; $\sum_k \bar{c}_{ik}c_{jk} = \delta_{ij}$.

6.1.1. $\begin{pmatrix} a & 0 & b & 0 \\ 0 & a & 0 & b \\ c & 0 & d & 0 \\ 0 & c & 0 & d \end{pmatrix}$.

6.3.2. Матрица D определена с точностью до перестановки диагональных элементов. При заданной матрице D матрицы O_1 и O_2 определены с точностью до преобразования $O_1 \mapsto O_1O$, $O_2 \mapsto O^{-1}O_2$, где O — ортогональная матрица, коммутирующая с D .

6.5.1. $t; t-1$.

6.5.2. 1) $\max_i \sum_j |a_{ij}|$, где $(a_{ij}) = A$ — матрица оператора \mathcal{A} ;

2) $\sqrt{\max_i \lambda_i}$, где $\lambda_1, \dots, \lambda_n$ — (неотрицательные) собственные значения самосопряженного оператора $\mathcal{A}^* \mathcal{A}$; 3) $\max_j \sum_i |a_{ij}|$.

7.1.1. 5.

7.1.2. $\dim(U_1 + U_2)$, если $P_1 \cap P_2 \neq \emptyset$; $\dim(U_1 + U_2) + 1$, если $P_1 \cap P_2 = \emptyset$.

7.3.2. Грани положительной размерности выделяются тем, что какие-то $k < n/2$ координат x_i равны 0 и какие-то $l < n/2$ координат равны 1 при условии, что $k + l < n - 1$ (при четном n это условие выполняется автоматически). Вершины — это точки у которых какие-то $[n/2]$ координат равны 0 и какие-то $[n/2]$ координат равны 1, а оставшаяся координата при нечетном n равна $1/2$.

7.3.3. Выпуклые оболочки всевозможных подмножеств множества вершин симплекса.

7.2.5. При перестановках точек p_1, p_2, p_3 их отношение принимает значения $c, -c - 1, \frac{1}{c}, -\frac{1}{c} - 1, -\frac{1}{c+1}, -\frac{c}{c+1}$. Наименьшее значение различных значений равно 2, если в поле K разрешимо уравнение $x^2 + x + 1 = 0$, и 3 в противном случае.

7.4.5. См. ответ к задаче 10.3.1, где перечислены все элементы группы $\text{Sym}_+ K$. Помимо этого, в группе $\text{Sym} K$ имеются шесть отражений относительно плоскостей, проходящих через ребра, три отражения относительно плоскостей, параллельных граням, восемь зеркальных поворотов на $\pi/3$ вокруг осей, проходящих через вершины, шесть зеркальных поворотов на $\pi/2$ вокруг осей, проходящих через центры граней и центральная симметрия.

7.4.6. $(x_1, x_2) \mapsto (tx_1, t^{-1}x_2); (x_1, x_2) \mapsto (tx_2, t^{-1}x_1)$, где $t \in \mathbb{R}^*$.

7.4.7. Если скалярные квадраты сторон одного треугольника (рассматриваемых как векторы) равны скалярным квадратам соответственных сторон другого треугольника, то эти треугольники равны.

7.6.2. $\frac{1}{y_1}, \frac{y_2}{y_1}, \dots, \frac{y_n}{y_1}$.

7.6.4. При перестановках из группы V_4 двойное отношение не меняется. При перестановках, оставляющих на месте точку p_4 , двойное отношение меняется так же, как взятое со знаком минус простое отношение точек p_1, p_2, p_3 (см. ответ к задаче 7.2.5), т. е. принимает значения

$$\delta, 1 - \delta, \frac{1}{\delta}, 1 - \frac{1}{\delta}, \frac{1}{1 - \delta}, \frac{\delta}{\delta - 1}.$$

9.1.7. $[10]_{15}, [6]_{15}$.

9.1.10. $[3]_7; [6]_{41}$.

9.2.5. Если $n = 2^m p_1^{k_1} \dots p_s^{k_s} q_1^{l_1} \dots q_t^{l_t}$, где p_1, \dots, p_s — различные простые числа вида $4k + 1$, а q_1, \dots, q_t — различные простые числа вида

10.3.3. 57.

10.5.4. $\frac{1}{2}q(q^2 - 1)$, если q нечетно, и $q(q^2 - 1)$, если q — степень двойки.

11.4.1. В обозначениях примера 10.1.16, $D_n = \langle a \rangle_n \rtimes \langle b \rangle_2$. При нечетном n группа D_n имеет 2 одномерных представления

$$a \mapsto 1, \quad b \mapsto \pm 1$$

и $\frac{n-1}{2}$ двумерных неприводимых представлений

$$a \mapsto \begin{pmatrix} \omega^k & 0 \\ 0 & \omega^{-k} \end{pmatrix}, \quad b \mapsto \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \left(1 \leq k < \frac{n}{2}\right),$$

где $\omega = e^{2\pi i/n}$. При четном n группа D_n имеет 4 одномерных представления

$$a \mapsto \pm 1, \quad b \mapsto \pm 1$$

и $\frac{n}{2} - 1$ двумерных неприводимых представлений, описываемых так же, как при нечетном n .

11.4.3. Одномерное подпространство констант; двумерное подпространство «четных» функций, принимающих одинаковые значения на противоположных гранях куба, с суммой значений, равной нулю; трехмерное подпространство «нечетных» функций, принимающих противоположные значения на противоположных гранях куба.

11.4.5. Таблица характеров имеет следующий вид:

	χ_1	χ_3	χ'_3	χ_4	χ_5	
e	1	3	3	4	5	1
(12)(34)	1	-1	-1	0	1	15
(123)	1	0	0	1	-1	20
(12345)	1	$\frac{1+\sqrt{5}}{2}$	$\frac{1-\sqrt{5}}{2}$	-1	0	12
(12354)	1	$\frac{1-\sqrt{5}}{2}$	$\frac{1+\sqrt{5}}{2}$	-1	0	12

11.6.2. Стандартная инволюция имеет вид

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

12.3.3. Матрицы присоединенных операторов имеют вид

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

12.4.6. Для каждого $n \in \mathbb{Z}_+$ имеется единственное $(2n + 1)$ -мерное неприводимое представление S_n группы $SO_3(\mathbb{C})$, связанное с представлением R_{2n} группы $SL_2(\mathbb{C})$ коммутативной диаграммой

$$\begin{array}{ccc} SL_2(\mathbb{C}) & \xrightarrow{R_{2n}} & GL_{2n+1}(\mathbb{C}) \\ & \searrow \text{Ad} & \nearrow S_n \\ & & SO_3(\mathbb{C}) \end{array}$$

(см. задачу 12.3.7).

12.4.7. Для каждого $n \in \mathbb{Z}_+$ имеется единственное $(2n + 1)$ -мерное неприводимое комплексное представление группы SO_3 (соответственно $SO_{2,1}^\circ$), получаемое ограничением представления S_n группы $SO_3(\mathbb{C})$ (см. ответ к задаче 12.4.6).

Словарь сокращений английских слов, употребляемых в обозначениях

Сокращение	От слова	Перевод
Ad (ad)	adjoint	присоединенный
aff	affine	аффинный
Alt	alternation	альтернирование
Ann	annihilator	аннулятор
area	area	площадь
arg	argument	аргумент
Aut	automorphism	автоморфизм
char	characteristic	характеристика
conv	convex	выпуклый
deg	degree	степень
det	determinant	определитель
diag	diagonal	диагональный
dim	dimension	размерность
exp	exponential	экспонента
Hom	homomorphism	гомоморфизм
ht	height	высота
Gr	Grassmanian	грассманиан
id	identity	тождество
Im	image	образ
Im	imaginary	мнимый
inf	infimum (лат.)	нижний
Isom	isometry	изометрия
Ker	kernel	ядро
lim	limit	предел
log	logarithm	логарифм
mod	modulo (лат.)	по модулю
ord	order	порядок
per	permanent	перманент
pf	Pfaffian	пфаффиан
Quot	quotient	частное

rad	radical	радикал
Re	real	действительный
rk	rank	ранг
sgn	signum (<i>лат.</i>)	знак
Spec	spectrum	спектр
Sym	symmetry	симметрия
Sym	symmetrization	симметрирование
Tor	torsion	кручение
tr	trace	след
tr. deg	transcendence degree	степень трансцендентности
Trans	translation	перенос
vol	volume	объем

Список литературы

1. *Атья М., Макдональд И.* Введение в коммутативную алгебру. М.: Факториал Пресс, 2003.
2. *Беклемишев Д. В.* Дополнительные главы линейной алгебры. М.: Наука, 1983.
3. *Беллман Р.* Введение в теорию матриц. М.: Наука, 1976.
4. *Берже М.* Геометрия. М.: Мир, 1984.
5. *Боревич З. И.* Определители и матрицы. М.: Наука, 1988.
6. *Ван дер Варден Б. Л.* Алгебра. М.: Наука, 1979.
7. *Винберг Э. Б.* Алгебра многочленов. М.: Просвещение, 1980.
8. *Винберг Э. Б.* Линейные представления групп. М.: Наука, 1985.
9. *Гантмахер Ф. Р.* Теория матриц. М.: Наука, 1988.
10. *Гельфанд И. М.* Лекции по линейной алгебре. М.: МЦНМО, 1998.
11. *Глазман И. М., Любич Ю. И.* Конечномерный линейный анализ в задачах. М.: Наука, 1969.
12. *Калужнин Л. А., Сущанский В. И.* Преобразования и перестановки. М.: Наука, 1985.
13. *Каргаполов М. И., Мерзляков Ю. И.* Основы теории групп. М.: Наука, 1996.
14. *Кострикин А. И.* Введение в алгебру. Часть I: Основы алгебры. М.: МЦНМО, 2009.
15. *Кострикин А. И.* Введение в алгебру. Часть II: Линейная алгебра. М.: МЦНМО, 2009.
16. *Кострикин А. И.* Введение в алгебру. Часть III: Основные структуры. М.: МЦНМО, 2009.
17. *Кострикин А. И., Манин Ю. И.* Линейная алгебра и геометрия. М.: Наука, 1986.
18. *Кострикин А. И. (ред.)* Сборник задач по алгебре. М.: МЦНМО, 2009.
19. *Куликов Л. Я., Москаленко А. И., Фомин А. А.* Сборник задач по алгебре и теории чисел. М.: Просвещение, 1993.
20. *Курош А. Г.* Курс высшей алгебры. М.: Наука, 1975.
21. *Курош А. Г.* Лекции по общей алгебре. М.: Наука, 1973.
22. *Ланкастер П.* Теория матриц. М.: Наука, 1982.
23. *Ленг С.* Алгебра. М.: Мир, 1968.

24. Мишина А. П., Проскуряков И. В. Высшая алгебра: Линейная алгебра, многочлены, общая алгебра. М.: Физматгиз, 1962. (Сер. «Справочная математическая библиотека»).
25. Постников М. М. Аналитическая геометрия. М.: Наука, 1986.
26. Постников М. М. Линейная алгебра и дифференциальная геометрия. М.: Наука, 1979.
27. Проскуряков И. В. Сборник задач по линейной алгебре. М.: Бинном. Лаборатория знаний, 2005.
28. Скорняков Л. А. Элементы алгебры. М.: Наука, 1980.
29. Скорняков Л. А. (ред.) Общая алгебра. Т. 1. М.: Наука, 1990. (Сер. «Справочная математическая библиотека»).
30. Фаддеев Д. К. Лекции по алгебре. М.: Наука, 1984.
31. Фаддеев Д. К., Соминский И. С. Сборник задач по высшей алгебре. М.: Наука, 1977.
32. Халмош П. Конечномерные векторные пространства. М.: Физматгиз, 1963; Ижевск: РХД, 2002.
33. Шафаревич И. Р. Основные понятия алгебры // Итоги науки и техники. Современные проблемы математики. Фундаментальные направления. Т. 11. М.: ВИНТИ, 1986; Ижевск: РХД, 1999.
34. Шафаревич И. Р. Основы алгебраической геометрии. М: МЦНМО, 2007.
35. Шилов Г. Е. Математический анализ. Конечномерные линейные пространства. М.: Наука, 1969.

Указатель обозначений

- \mathbb{N} 9
 \mathbb{Z} 9
 \mathbb{Z}_+ 9
 \mathbb{Q} 9
 \mathbb{R} 9
 \mathbb{R}_+ 9
 t_a 11
 \mathbb{R}^n 15
 2^X 19
 K^* 21
 \mathbb{C} 22
 i 22
 $\operatorname{Re} c$ 24
 $\operatorname{Im} c$ 24
 \bar{c} 24
 $|c|$ 25
 $\arg c$ 25
 $a \sim_R b, a \sim b$ 29
 $[a]_n$ 30
 \mathbb{Z}_n 30
 $\operatorname{char} K$ 33
 C_p^k 33
 K^n 35
 E^2, E^3 35
 $F(K, X)$ 35
 \mathbb{H} 41
 (a_{ij}) 42
 $\operatorname{diag}(a_1, \dots, a_n)$ 44
 E 44
 $L_n(K)$ 45
 $M_n(K)$ 45
 E_{ij} 45
 A^T 46
 a_{ij}^T 46
 \tilde{A} 48
 $\langle S \rangle$ 62
 $\dim V$ 63
 δ_a 63
 K^∞ 67
 $\operatorname{rk} A$ 68
 $\operatorname{area}(a_1, a_2)$ 75
 $\operatorname{vol}(a_1, a_2, a_3)$ 76
 $\det A$ 77, 80
 $|a_{ij}|$ 77
 $\operatorname{sgn}(k_1, k_2, \dots, k_n)$ 79
 $V(x_1, \dots, x_n)$ 84
 M_{ij} 85
 A_{ij} 85
 $\mathbb{R}[x]$ 92
 $K[x]$ 93
 $\deg f$ 93, 128
 $K[[x]]$ 94
 D 102
 f' 103
 $N(f)$ 113
 $L(f)$ 113
 $b|a$ 118
 $a \sim b$ 118
 $\mathbb{Z}[i]$ 119
 $\operatorname{НОД}\{a, b\}$ 119, 433
 $\operatorname{НОК}\{a, b\} ([a, b])$ 123
 $K[x_1, \dots, x_n]$ 127
 CC_n^d 128
 $\deg_{x_1} f$ 129
 $u \succ v$ 130
 s_k 132
 σ_k 132
 $D(\varphi)$ 141
 $\frac{a}{b} (a/b)$ 148
 $\operatorname{Quot} A$ 148
 $K(x)$ 149
 $S(X)$ 154
 S_n 154
 $\operatorname{Isom} E^2$ 155
 $\operatorname{Trans}(V)$ 156
 $\operatorname{GL}_n(K)$ 157
 A^* 157
 O_n 159
 $\operatorname{Sym} F$ 161
 D_n 161
 $\operatorname{SL}_n(K)$ 161
 $\operatorname{SL}_n(\mathbb{Z})$ 162
 $F_1 \underset{G}{\sim} F_2$ 162
 $\operatorname{GA}_n(K)$ 164
 $O_{3,1}$ 166
 $\langle g \rangle$ 168
 $\operatorname{ord} g$ 168
 $(i_1 i_2 \dots i_p)$ 169
 C_n 171
 ε_k 171
 $|G|$ 171
 $\langle S \rangle$ 173
 \mathbb{T} 176
 G/H 177
 $|G:H|$ 177
 $\varphi(n)$ 178
 $x \underset{G}{\sim} y$ 178
 G_x 179
 G_x 179
 $|Gx|$ 179
 $\operatorname{Sym} K$ 180
 $H \triangleleft G (G \triangleright H)$ 181
 $\operatorname{Im} f$ 184
 $\operatorname{Ker} f$ 184
 $\operatorname{sgn} \sigma$ 185
 A_n 186
 df 188, 284, 285
 V_4 189

- G_+ 190
 $U \cap V$ 193
 $U_1 + \dots + U_k$ 195
 $U_1 \oplus \dots \oplus U_k$ 196
 $L_n^+(K)$ 196
 $L_n^-(K)$ 196
 A_d 197
 $\text{Im } \varphi$ 200
 $\text{Ker } \varphi$ 200
 id 204
 $\text{tr } X$ 206
 V^* 206
 ε_k 206
 δ_{ij} 206
 U^0 208
 $\text{Ker } \alpha$ 211
 $\text{rk } \alpha$ 211
 U^\perp 213
 δ_k 214
 $(,)$ 221, 232
 $\text{pr}_U x$ 225
 $\text{ort}_U x$ 225
 $\rho(x, y)$ 226
 $\text{vol } P$ 228
 $\mathcal{A}, \mathcal{B}, \mathcal{C}, \dots$ 234
 $\Pi(\alpha)$ 237
 L_a 239
 $L(V)$ 240
 \mathcal{E} 240
 $\text{GL}(V)$ 240
 $\text{rk } \mathcal{A}$ 240
 $\det \mathcal{A}$ 240
 $f_{\mathcal{A}}(t)$ 241
 $V(\mathbb{C})$ 242
 $\mathcal{A}_{\mathbb{C}}$ 243
 $V_\lambda(\mathcal{A})$ 244
 \mathcal{P} 244
 \mathcal{R} 245
 $\varphi_{\mathcal{A}}$ 247
 \mathcal{A}^* 247, 256
 $O(V)$ 253
 SO_n 253
 $\text{SO}(V)$ 253
 $U(V)$ 257
 U_n 257
 SU_n 257
 $\text{ht } e$ 261
 $f(\mathcal{A})$ 265
 $m_{\mathcal{A}}$ 266
 P_n 269
 $\|\mathcal{A}\|$ 272
 $e^{\mathcal{A}}$ 273
 \overline{pq} 277
 $\text{aff } M$ 279
 $\text{center}(p_1, \dots, p_k)$ 281
 $\text{GA}(S)$ 286
 $\text{Trans}(S)$ 287
 $\text{conv } M$ 290
 M° 292
 H_f 292
 H_f^+, H_f^- 292
 $\rho(p, q)$ 302
 $\text{Isom } S$ 303
 $\text{Isom}_+ S$ 303
 r_H 303
 $\text{Sym } M$ 307
 $\text{Sym}_+ M$ 307
 T, K, O, D, I 308
 $O(V, \alpha)$ 308
 $d_o Q$ 310
 $X(Q)$ 311
 $G(X)$ 319
 PV 325
 KP^n 326
 \hat{x} 326
 V_S 326
 $\widehat{\mathcal{A}}$ 328
 $\text{PGL}(V)$ 329
 $\det(u, v)$ 331
 $(p_1, p_2; p_3, p_4)$ 331
 $PX(Q)$ 333
 $G(PX)$ 337
 $\text{Hom}(V_1, \dots, V_p; U)$ 338
 $\text{Hom}(V_1, \dots, V_p; K)$ 338
 $V \otimes W, V \otimes_K W$ 340
 $\alpha \otimes \beta$ 341
 $V(L)$ 343
 $T_q^P(V)$ 346
 $T_{i_1 \dots i_p j_1 \dots j_q}$ 349
 $V_1 \oplus \dots \oplus V_k$ 351
 $T(V)$ 352
 $T_*(V)$ 352
 $S^P(V)$ 354
 $S(V)$ 355
 Sym 356
 $\alpha \vee \beta$ 359
 $\text{per } A$ 359
 $\Lambda^P(V)$ 361
 $\Lambda(V)$ 362
 Alt 362
 $\alpha \wedge \beta$ 364
 $\text{Gr}_p(V)$ 365
 $P\Lambda^P(V)$ 366
 $\text{pf } A$ 369
 $\langle S \rangle$ 373
 $\text{rk } L$ 373
 $A_1 \oplus \dots \oplus A_k$ 381, 391
 $G_1 \times \dots \times G_k$ 381
 $\text{Tor } A$ 383
 $\text{Tor}_p A$ 384
 $\left(\frac{k}{p}\right)$ 386
 A/I 387
 $\text{Im } f$ 388, 397
 $\text{Ker } f$ 388, 397
 $(S), (u)$ 392
 M/N 397
 $\text{Ann } M$ 398
 $\text{Tor } M$ 401
 $\text{Tor}_p M$ 401
 $\text{rad } A$ 406, 497
 $A[u_1, \dots, u_n]$ 408
 $\dim_K L$ 408
 m_u 409
 $K(u_1, \dots, u_n)$ 411
 \bar{K} 411

\bar{A}, \bar{Z} 418	X^g 457	φ_{ijk} 508
\mathbb{Z}_K 420	$Z(x), C(x)$ 457	χ 509
tr. deg A 422	$N(h)$ 458	$Z\mathbb{C}[G]$ 509
$K[M], I(M)$ 426	$\text{PSL}_n(K)$ 464	$K[V]^G$ 517
Spec A 427	$\text{Aut}_K L$ 466	\natural 519
$\dim M$ 430	L^G 466	$Z(D)$ 523
$\rho(v)$ 437	$\text{Gal } L/K$ 468	$D(\alpha, \beta)$ 524
Cl A 438	K_n 469	$N(q)$ 525
$\text{Aut } G, \text{Int } G$ 444	G_p 471	deg D 528
$N \setminus H, H \triangleleft N$ 446	$R(x)$ 478	\mathbb{O} 534
$N \setminus_a H, N \setminus H$ 447	R_U 481	$N(u)$ 535
$(G, G), G'$ 448	$R_{V/U}$ 481	$[uvw]$ 535
$B_n(K)$ 451	$M(\sigma)$ 482	$T(G)$ 542
$G : X, G : X$ 452	$T(a)$ 483	G° 542
$l(g)$ 453	Ad 486, 551	$\text{SO}_{n,1}$ 544
$r(g), a(g)$ 453	$V_{(S)}, R_{(S)}$ 487	exp 545
$x \approx y$ 453	\mathbb{T}^n 494	$[A, B]$ 549
Gx 453	$Z(A)$ 503	ad 552
G_x 455	KG 504	$Z(L)$ 554
	$\mathbb{C}[G]$ 507	

Предметный указатель

А

- Автоморфизм 24
 - группы 184, 443
 - — внутренний 444, 551
 - Фробениуса 472
 - Алгебра 38, 39, 525
 - альтернативная 535, 536
 - ассоциативная 524
 - внешняя 362
 - градуированная 197
 - Грассмана 362
 - групповая 504, 505, 507, 508
 - инвариантов 517, 520, 521
 - касательная группы Ли 550
 - кватернионов 41, 46, 63
 - — обобщенная 524, 525, 529
 - конечно порожденная 420—424, 427
 - Кэли 535
 - Ли 551
 - — линейная 550
 - — простая 556
 - линейных операторов 240, 499, 501, 502
 - матриц 45, 197, 240, 390, 551
 - многочленов 92—94
 - — на алгебраическом многообразии 426
 - — от n переменных 127, 129, 197, 340
 - нильпотентная 496
 - нильтреугольных матриц 496
 - октав 534—536
 - полилинейных функций 352
 - полупростая 497—499, 501, 502, 504, 526
 - — простая 501, 502, 534
 - — с делением 523, 525, 527, 530, 533, 534
 - — центральная 524, 527—529, 531
 - — — расщепимая 528
 - с единицей 390
 - симметрическая 355—357
 - суперкоммутативная 362
 - тензорная 352
 - формальных степенных рядов над K 94
 - функций на множестве 39, 390, 391, 517
 - $K[t]/(h)$ 497, 499, 500, 502
 - Алгоритм Евклида 120
 - Аннулятор модуля 398
 - подпространства 208
 - Антиавтоморфизм 524
 - Антикоммутативность 19, 550
 - Аргумент комплексного числа 25
 - Ассоциативность 14, 15, 156
 - Ассоциатор 535
- ## Б
- Базис абелевой группы 373, 374
 - векторного пространства 37, 62, 63, 243
 - — —, согласованный с подпространством 192, 193
 - жорданов 264
 - модуля 399
 - ортогональный 214
 - ортонормированный 224, 233
 - симплектический 220
 - трансцендентности 421, 422

Бивектор 361

В

Вектор 35, 37

— в аффинном пространстве 277

— геометрический 35

— корневой 258, 259

— собственный 241, 243, 250

f -вектор выпуклого многогранника 299

p -вектор 361

Векторизация 278

Векторы линейно зависимые

59—61, 64, 223, 364

— — независимые 59, 60, 64, 364

—, ориентированные положительно 75, 76

— ортогональные 212, 231

Вершина выпуклого многогранника 298

— квадрики 311

— параболоида 322

Вершины смежные 299

Высота вектора 261

— корневого вектора 258, 260

— нильпотентного оператора 261, 265

— параллелепипеда 227

Вычет квадратичный 219

— числа 30

Г

Геометрия аффинная 164, 288, 289

— конформная 337

— Лобачевского 337

— проективная 330, 335

— псевдоевклидова 308

Гиперболоид 319

Гипергрань выпуклого многогранника 298

Гиперплоскость в аффинном пространстве 278

— опорная 293, 295

— проективная 325

Гиперповерхность второго порядка 311

Гомоморфизм алгебр 390

— групп 183, 186, 189

— — Ли 548, 551

— канонический 191, 388, 390, 398

— колец 388

— модулей 397

— над полем 412

Гомотетия 287, 288

Градуировка 197

Грань выпуклого многогранника 298, 299

Группа 156, 158

— абелева (коммутативная) 14, 15, 157, 182, 184, 372, 396, 484, 495, 506

— — конечно порожденная 373, 383

— — свободная 373, 374, 379

— — циклическая 385

— автоморфизмов 444

— — внутренних 444

— — конечного расширения поля 466

— аддитивная вещественных чисел 187, 481, 484

— — комплексных чисел 176, 182, 186

— — целых чисел 17, 171, 173, 182, 382

— аффинных преобразований квадрики 319

— вращений куба 190, 455, 456, 481, 511

— вычетов по модулю n 171, 382, 445

- Галилея 165
- Галуа 468
- дважды транзитивная 512
- движений евклидова аффинного пространства 303, 304, 446
- — — пространства E^3 155
- — евклидовой плоскости E^2 155, 164, 170, 175, 179, 453
- — — — собственных 454
- диагональных матриц 162, 183, 443
- диэдра 161, 190, 448, 507
- знакопеременная 186, 187, 449, 450, 462
- — порядка 4 450, 463, 464
- — порядка 5 461, 463, 464, 512
- классов идеалов 438, 440
- Клейна четверная 189, 453, 463
- кольца аддитивная 18
- конечная 177—179, 186, 456, 457, 462, 492, 504, 505, 512, 513, 516, 518, 519
- — порядка p^2 458
- — — pq 460
- Ли 537, 542, 546
- — комплексная 558
- — — редуцируемая 558
- — линейная 539, 540
- — простая 556
- — связная 544, 547, 548
- Лоренца 166, 309
- мультипликативная корней из 1 171, 173
- — поля 21, 157, 385, 537
- — — \mathbb{C} 170, 176, 183, 184, 187, 189, 442
- — — \mathbb{Z}_p 178
- невырожденных квадратных матриц см. группа полная линейная
- — треугольных матриц 162, 451, 541
- обратимых элементов кольца 157
- — — — \mathbb{Z}_n 32, 171, 178, 386, 392, 445
- однопараметрическая, порожденная оператором 273
- ортогональная 159—161, 179, 253, 254, 453, 493, 496, 522, 540, 543, 547, 550
- — специальная 253, 465, 558
- — — порядка 3 464, 553—556
- — — порядка 4 465, 557
- параллельных переносов 156, 158, 163, 287, 453
- подстановок см. группа симметрическая
- полная аффинная 164, 188, 190, 286, 287, 330, 446, 545
- — линейная 157, 163, 174, 177, 183, 187—189, 240, 287, 444, 446, 450, 486, 496, 537, 540, 541, 543, 548
- — проективная 329, 444
- , порожденная подмножеством 173
- преобразований 154, 157, 159, 166, 178, 451
- — транзитивная 163, 179
- примарная (p -группа) 382, 458
- простая 461
- псевдоортогональная 308, 541
- Пуанкаре 165, 166, 309
- разрешимая 450
- симметрии куба 180, 308
- — правильного многогранника 180, 308
- — фигуры 161, 307
- симметрическая 154, 169, 174, 177, 179, 183, 185, 187, 444, 446, 450, 454, 482, 507, 513, 517, 518, 520

Группа симметрическая поряд-
ка 3 158, 188, 191, 445, 464, 506
— — порядка 4 181, 188—190, 446,
456, 481, 492, 506, 511
— симплектическая 541
— топологическая 493
— — компактная 493, 495, 496,
521
— унимодулярная 161, 162, 175,
177, 185, 189, 191, 449, 450, 540,
543, 547, 555, 557, 559
— — проективная 464
— унитарная 257, 493, 541, 544,
547
— — специальная 257, 542, 544,
554, 558
— циклическая 171—173, 178, 183,
381, 382, 385, 447, 510, 513
— \mathbb{T} 185, 493

Д

Движение 302, 304, 306
— винтовое 306
— несобственное 303
— собственное 303

Действие 453

— группы на множестве 451—453
— левыми сдвигами 453
— правыми сдвигами 453
— сопряжениями 453
— транзитивное 453, 456
— эффе́ктивное 452

Деление окружности на n равных
частей 476

— с остатком 96, 118

Делимость элементов 118

Делитель нуля 20, 21

Детерминант 185

Диагональ главная 44

— побочная 44

Диаграмма коммутативная 455

Дивизор простой 437

Дискриминант 141, 142
— кубического многочлена
142—144

Дистрибутивность 18, 19

Дифференциал аффинно-квадра-
тической функции 310

— аффинно-линейной функции
285

— аффинного преобразования
188

— отображения 284

Длина вектора 222, 232

— орбиты 179

Дополнение алгебраическое 85

— ортогональное 213, 231

Дробь 148

— несократимая 149

— рациональная 150

— — правильная 150, 151

— — простейшая 151

Е

Единица матричная 45

— правая 158

З

Задача интерполяции 96, 269

— о получении максимальной
прибыли 300

— транспортная 301

Закон инерции 218, 232

Замыкание кольца целое 418

— поля алгебраическое 412, 528

Знак перестановки 79

— подстановки 185, 186

Значение собственное 241, 242,
250

И

Идеал алгебры 389

— главный 392, 437

— кольца двусторонний 387

- — левый 387
- — правый 387
- максимальный 406, 407
- многообразия 426, 435
- нормирования 437
- , порожденный подмножеством 392
- простой 406
- Идеалы эквивалентные 438
- Изоморфизм алгебр 41
- алгебраических структур 11
- векторных пространств 36, 205
- групп 184
- действий 455
- евклидовых пространств 229
- естественный 344, 345, 353
- канонический 350
- модулей 397
- над полем 412
- представлений 479
- Инвариант действия группы 516, 517
- Инверсия элементов 79
- Инволюция 524
- Индекс инерции 218, 232
- подгруппы 177
- К**
- Карта аффинная 326
- Квадрат симметрический представления 514
- Квадрика 311—313, 315, 316, 319, 321, 322
- коническая 312, 315, 316, 321
- невырожденная 314, 335
- нецентральная 315, 316, 322
- проективная 333, 335
- — линейчатая 336, 337
- — невырожденная 333, 337
- — овальная 336, 337
- центральная 311
- — неконическая 315, 316, 321
- цилиндрическая 314
- Кватернион 524
- сопряженный 524
- Китайская теорема об остатках 382
- Класс смежный 176—179
- сопряженных элементов 454
- эквивалентности 29
- Клетка жорданова 264, 266
- — нильпотентная 262
- Кольцо 18, 184, 387, 396, 397
- ассоциативное 18
- без делителей нуля 20, 439
- вычетов 31, 32, 47, 388, 392
- главных идеалов 393—395, 399, 406, 432
- евклидово 118, 119, 121—123, 149, 393
- коммутативное 18—20
- — ассоциативное с единицей 47, 94, 138, 403
- конечно порожденное 406
- многочленов 118, 119, 121, 389, 393, 405, 433
- — от нескольких переменных 129, 405, 434
- нётерово 403—406, 417, 418, 431, 437
- подмножеств 19, 21, 39
- с единицей 19
- факториальное 432, 433, 437
- функций 19, 20
- , целое над подкольцом 417
- целостное 118, 147, 431, 432
- — нормальное (целозамкнутое) 418
- — факториальное 432
- целых гауссовых чисел 119, 123, 394, 395
- — чисел 20, 21, 47, 118, 119, 121, 388, 393

Кольцо целых чисел поля 420, 439, 440
 Комбинация векторов линейная 37, 59, 67
 — — — нетривиальная 59
 — — — тривиальная 59
 — точек барицентрическая 280, 281
 — — выпуклая линейная 290
 Коммутант 448, 449
 — кратный 450
 Коммутативность 14, 15
 Коммутатор 448
 — матриц 549
 Комплексификация 242, 257
 Композиция отображений 10
 Компонента изотипная представления 487
 — многочлена однородная 128
 — неприводимая 429
 — связная группы Ли 543
 — — пространства 542
 Коники 311
 Константа структурная алгебры 526
 Конус 312, 332
 — грассманов 365
 — квадратичный 319, 333
 Координаты барицентрические 282, 286
 — вектора 37
 — неоднородные 327, 328
 — однородные 327
 — плюккеровы 366
 — тензора 341, 349
 Корень многочлена 99, 100
 — — кратный 100, 104, 117, 142
 — — простой 100
 — первообразный 171
 Коэффициент многочлена 93, 104
 — — старший 93

Коэффициенты линейного уравнения 48
 — линейной функции 206
 Кратность корня 100
 Кривая второго порядка 311, 317
 Критерий Сильвестра 218, 232

Л

Лемма Гаусса 124, 433
 — Даламбера 109
 — Нётер о нормализации 422
 — о замене 421
 — о линейной зависимости 61
 — о неподвижной точке 492, 495
 — Шура 483

М

Матрица 41, 50
 — билинейной функции 210
 — верхняя треугольная 52
 — — — строго 52, 54
 — вырожденная 74
 — Грама 223
 — диагональная 44, 377
 — единичная 44
 — жорданова 264
 — квадратная 44, 196
 — кососимметричная 196
 — косоэрмитова 231
 — коэффициентов 48
 — — расширенная 48
 — линейного оператора 234, 235, 237
 — — отображения 199
 — невырожденная 73, 74, 82, 226, 255
 — нижняя треугольная 52
 — нильтреугольная 259, 390
 — обратная 73, 74, 90
 — оператора 264
 — ортогональная 224, 253
 — перехода 66

- полуторалинейной функции 230
- симметричная 196
- скалярная 46
- ступенчатая 50
- транспонированная 46
- трапецеидальная 52
- унитарная 233, 257
- целочисленная 91, 162, 378
- элементарная 57, 58
- эрмитова 231
- Матрицы подобные 265
- Матричная единица 204
- Метод вращений 58
 - Гаусса 49, 54—56
 - Якоби 218, 232
- Минор 85, 91
 - главный 242
 - дополнительный 85
 - окаймляющий 91
 - угловой 214
- Многогранник выпуклый 296, 297, 300, 307
 - правильный 307, 308
 - телесный 296
- Многогранники правильные двойственные 308
- Многообразие алгебраическое 426, 427
 - — аффинное 426, 428, 430, 436
 - — неприводимое 435
 - гладкое 538
 - грассманово 365
- Многоугольник правильный 308
- Многочлен 92—96, 106, 110, 141
 - аннулирующий 266
 - деления круга 125
 - кубический 142, 470, 472
 - минимальный 266, 267, 403
 - — элемента 409
 - на алгебраическом многообразии 426
 - неполный 143
 - неприводимый 121, 395, 415, 434, 435, 468, 473
 - нормированный 102
 - от n переменных 127—130
 - — однородный 128
 - — — степени d 128
 - от матрицы (оператора) 265, 268
 - приведенный 102
 - примитивный 433
 - с вещественными коэффициентами 111—113, 142
 - с целыми коэффициентами 123—126
 - — — примитивный 124
 - сепарабельный 468
 - симметрический 132, 134, 138, 159, 517
 - — элементарный 132
 - характеристический 241, 242, 244—246, 259, 261, 264, 268, 346, 360
- Множество выпуклое 290—292, 295, 297
 - , замкнутое относительно операции 16
- Множитель инвариантный 385
 - — модуля 402
 - — подгруппы 380
- Модуль 47, 396, 397
 - комплексного числа 25, 106
 - конечно порожденный 398, 400, 401
 - периодический 398
 - правый 396
 - примарный (p -примарный) 401
 - свободный 399
 - — циклический 399
 - циклический 398
- Морфизм представлений 479, 483

Н

Наибольший общий делитель 119,
433

Наименьшее общее кратное 122

Направление особое параболои-
да 321

Начало отсчета пространства 155,
277

Невычет квадратичный 219

Неизвестные главные 53

— свободные 53

Неравенство Коши—Буняковско-
го 222, 232

Норма 118, 270

— кватерниона 525

— линейного оператора 272

— октавы 535

Нормализатор подгруппы 458

Нормирование 437

О

Область целостности 118

Оболочка аффинная 279, 281

— выпуклая 291, 297

— линейная 62, 64

Образ гомоморфизма 184, 388

— линейного отображения 200,
201

Объем параллелепипеда 228

Оператор альтернирования 362

— в евклидовом пространстве
247, 254

— в эрмитовом пространстве 255

— дифференцирования 238, 241,
245, 259, 261, 274

— кососимметрический 247, 248,
251

— косоэрмитов 256

— линейный 234, 235, 240, 242,
243, 246, 265, 350, 402, 403

— нильпотентный 261, 263

— ортогональный 248, 252

— представления 478

— присоединенный 551

— Рейнольдса 519

— самосопряженный 248, 256

— симметрирования 356

— симметрический 247—249, 257
— — положительно определен-
ный 251, 254

— сопряженный 247, 256

— тождественный 240

— унитарный 256

— эрмитов 256

— — положительно определен-
ный 257

Операция альтернирования 363

— в множестве 9

— коммутативная 12

— симметрирования 357

Определитель Вандермонда 84,
132, 435

— линейного оператора 240

— матрицы 76, 80—83, 86, 185,
210

Орбита 179, 453

Основание параллелепипеда 227

Основная теорема алгебры ком-
плексных чисел 106, 477

— — теории Галуа 471

Остаток от деления многочленов
96

Ось движения 305

— параболоида 322

Отношение двойное 331, 332

— на множестве 28

— сравнимости по модулю 30

— — — подгруппы 175

— тройки точек 289

— эквивалентности 28

— —, определяемое действием
453

— —, согласованное с операцией
29

- Отображение аффинное
283—285
— билинейное 339
— линейное 197, 198, 201—204
— полилинейное (p -линейное)
338, 340, 341, 344
— — кососимметрическое 360,
361
— — симметрическое 353, 355
— факторизации 29
— эквивариантное (G -эквивариантное) 455
— экспоненциальное 545, 547
- Отражение 245, 303
— ортогональное 248
— скользящее 306
- Отрезок 290
- П**
- Параболоид 321, 322, 335
— гиперболический 319
— эллиптический 319
- Параллелепипед 227, 296, 299
— фундаментальный 375
- Перемена знака 113
- Перенос параллельный 156
- Пересечение подпространств 193
- Перестановка элементов 78
— — тривиальная 78
— — четная (нечетная) 79
- Перманент матрицы 359
- Плоскости параллельные 280
— скрещивающиеся 280
- Плоскость 305
— бесконечно удаленная 326
— в аффинном пространстве
278—281, 295
— проективная 325, 328
- Поверхность второго порядка 311,
317
- Поворот 198, 199, 237, 241, 243,
306
— зеркальный 252, 306
- Подалгебра 41
- Подгруппа 159
— абелевой группы 16, 17
— дискретная 376
— кручения 383
— p -кручения 384
— нормальная 181, 182, 445
—, порожденная подмножеством
173, 373
— силовая 459, 460
— циклическая, порожденная
элементом 168
- Подгруппы сопряженные 458
- Подкольцо 21
—, порожденное элементами 408
- Подматрица 85
- Подмножество дискретное 375
- Подмодуль 396
— кручения 401
—, порожденный подмножеством
398
- Подполе 21, 36
— G -инвариантных элементов
466
—, порожденное элементами 411
- Подпредставление 481
- Подпространства линейно независимые 195
- Подпространство 36, 65, 208, 209
— инвариантное 236, 480, 555
— корневое 259—261
— направляющее 278
— невырожденное 213
— собственное 244, 245
— циклическое 262
- Подстановка 154
— четная (нечетная) 186
- Подход аксиоматический 27
- Подъем индекса 350
- Поле 20, 32, 39, 388
— алгебраически замкнутое 106

- Поле алгебраических чисел 412, 419
- вещественных чисел 27, 64, 67
 - вычетов 32, 34, 415
 - комплексных чисел 23, 24, 27, 38—40, 46, 63, 106, 121, 389
 - конечное 414, 415, 468, 469
 - круговое (деления круга) 410, 416, 420, 469, 472
 - отношений (дробей) кольца 148, 421
 - , порожденное элементами 411, 412, 418
 - разложения многочлена 412, 469, 470, 472
 - рациональных дробей (функций) 149
 - — чисел 21
 - числовое 32
- Поливектор 361
- разложимый 361, 368
- Полупространство 292
- опорное 293
- Поляризация 358
- квадратичной функции 212
- Порядок группы 171, 177
- элемента группы 168—170, 178
- Последовательность финитная 67
- Построение циркулем и линейкой 475
- Представление алгебры 479, 483, 552
- вполне приводимое 484—486
 - группы 452, 478, 479, 512
 - — Ли 551
 - — — неприводимое 556
 - — неприводимое 491
 - — одномерное 506
 - — ортогональное 491
 - — симплектическое 491
 - изотипное (S -изотипное) 487, 488
 - множества 478, 480
 - — одноэлементного 479, 487, 490
 - мономиальное 482, 485
 - неприводимое 481, 482, 509, 513, 516
 - присоединенное 486, 552
 - — алгебры 552
 - — — Ли 553
 - регулярное 483
 - с простым спектром 490, 491
 - самосопряженное 513
 - сопряженное 512
 - тавтологическое 499
- Представления изоморфные 479
- Преобразование аффинное 164, 165, 286, 288, 289, 330
- квазиэлементарное базиса модуля 400
 - координат 66
 - линейное 234
 - Лоренца 166
 - множества 154, 239
 - ортогональное 159
 - проективное 328—331
 - элементарное системы линейных уравнений 49
 - — строк матрицы 50, 57, 71
 - — целочисленное 377
- Приведение к главным осям 250
- Присоединение корня многочлена 409
- Программирование линейное 300
- Проективизация конуса 333
- Проектирование 198
- ортогональное 200, 238
- Проектор 244, 245
- ортогональный 248
- Проекция вектора 196
- — ортогональная 225, 233
- Произведение групп 507

- идеалов 438
- линейных операторов 349
- матриц 42
- матрицы на элемент поля 42
- отображений 10
- полупрямое 447
- — подгрупп 446
- представлений 490, 513, 514
- — тензорное 513
- прямое 381
- — групп 443
- — подгрупп 441, 442
- смешанное 553
- тензорное векторных пространств 339, 340, 344
- — матриц 346
- — операторов 345, 346
- функций внешнее 364
- — симметрическое 359
- Производная многочлена 103
- Пространство аффинное 277, 285
- — евклидово 302
- — псевдоевклидово 309
- векторное (линейное) бесконечномерное 62, 66, 208, 271
- — евклидово 221, 229, 246, 302
- — конечномерное 35, 38, 62, 63, 65, 196, 270, 271, 396, 467
- — несчетномерное 68
- — псевдоевклидово 308
- — счетномерное 67, 68
- второе сопряженное 207
- касательное 539
- — к группе Ли 540, 546, 550
- Минковского 309
- над полем из q элементов 64, 65
- непрерывных функций 64
- представления 478
- проективное 325, 328
- сопряженное 206, 207
- — к евклидову 247
- — строк длины n 35, 63, 65
- топологическое неприводимое 428
- — нётерово 428, 429
- — связное 542
- финитных последовательностей 67, 93, 208
- функций на группе 507, 508, 510
- — — центральных 509, 510
- — на множестве со значениями в поле 35, 63, 205, 239, 452
- эрмитово 232, 233
- Процесс ортогонализации Грама—Шмидта 216, 225
- Прямая в аффинном пространстве 278
- проективная 325
- Пфаффиан 369, 371
- Р**
- Радикал алгебры 497
- кольца 407
- нильпотентный 406
- Разделение орбит инвариантами 517, 521
- Разложение многочлена на множители 110, 111, 117
- полярное 254, 257
- элемента на простые множители 121, 431, 432
- Размерность аффинного алгебраического многообразия 430
- — пространства 278
- векторного пространства 63, 66
- представления 478
- Разность симметрическая 19
- элементов 14
- Разрешимость в квадратных радикалах 474

- Ранг билинейной функции 211, 221
 — квадратичной функции 212
 — линейного оператора 240
 — матрицы 68, 69, 71—73, 91, 205
 — модуля 399
 — свободной абелевой группы 373
 — системы векторов 68
 Расстояние между векторами 226
 — между плоскостями 302
 — между точками 302
 — от вектора до подпространства 226, 227
 Расширение Галуа 468
 — кольца 407, 418
 — — алгебраическое 408
 — — конечно порожденное 408
 — — конечное 417
 — — целое 417
 — основного поля 342
 — поля 411, 412, 501
 — — алгебраическое 408
 — — квадратичное 409, 420, 469
 — — конечное 409, 410, 466, 527
 — — простое 409
 — — расщепимое 527
 — — сепарабельное 419
 Ребро выпуклого многогранника 298
 Редукция многочлена по модулю p 124
 Резольвента кубическая 140
 Репер 278
 Рефлексивность 28
 Решение общее 53
 Решетка 375, 376
 Ряд абсолютно сходящийся 271
 — композиционный 461
 С
 Свертка 348
 — тензора 347
 Сигнатура квадратичной функции 218
 Символ Кронекера 206
 — Лежандра 386
 Симметричность 28
 Симметрия центральная 288
 Симплекс n -мерный 291, 299
 Симплекс-метод 301
 Система алгебраических уравнений 424, 425
 — векторов 59
 — координат аффинная 278
 — — прямоугольная 302
 — линейных уравнений 48, 89, 279, 280
 — — — неопределенная 53, 56
 — — — несовместная 49, 90
 — — — определенная 53, 54, 73
 — — — совместная 49, 55
 — — — (строго) треугольная 52
 — — — ступенчатая 52
 — однородных линейных уравнений 55, 69, 209
 — порождающих модуля 398
 — порождающих (элементов) группы 173, 373
 — решений фундаментальная 70
 — точек общего положения 331
 — элементов модуля линейно независимая 399
 Системы векторов эквивалентные 68
 — линейных уравнений эквивалентные 49
 След 348
 — матрицы 206, 242
 — элемента 416
 Соотношения Пюккера 367, 368
 Сопряжение комплексное 24
 Спектр алгебры 427
 Спуск индекса 350

- Стабилизатор 179
 — точки 455
 Степень линейного оператора
 симметрическая 359
 — многочлена 93
 — — от n переменных 128
 — — — по переменной 129
 — оператора внешняя 364
 — представления внешняя 514
 — — симметрическая 514
 — пространства внешняя 361,
 362, 368
 — — симметрическая 354, 357
 — расширения 409
 — центральной алгебры с деле-
 нием 528
 — элемента группы 166, 169
 Строка 15
 — единичная 37
 — нулевая 15
 Сумма внешняя 381
 — внутренняя 381
 — линейных представлений 486
 — матриц 42
 — подпространств 193—195
 — прямая алгебр 391
 — — внешняя 351, 391
 — — внутренняя 351, 391
 — — групп 381
 — — колец 391, 406
 — — модулей 397
 — — подгрупп 380
 — — подколец 391
 — — подпространств 196, 351
 — — пространств 351
 Схема Горнера 97

Т
 Тело 523
 — выпуклое 291—293
 — платоново 308
 Тензор 342
 — ковариантный 352
 — контравариантный 351
 — кососимметрический 362
 — метрический 350
 — разложимый 342, 344, 355
 — симметрический 356
 — типа (p, q) 346
 Теорема Безу 97
 — Бернсайда 489
 — Веддербёрна 531
 — Вильсона 102
 — Гамильтона—Кэли 268, 403
 — Гильберта о базисе идеала 405
 — — о нулях 425
 — — об инвариантах 518, 519, 559
 — Декарта 113
 — Жордана—Гёльдера 461
 — китайская об остатках 382
 — Кронекера—Капелли 69
 — Кэли 453
 — Лагранжа 177
 — Менелая 283
 — Минковского—Вейля 297
 — о гомоморфизме алгебр 390
 — — групп 186, 191
 — — колец 388
 — — модулей 398
 — о примитивном элементе 527
 — о ранге матрицы 91
 — об определителе матрицы
 с углом нулей 83
 — основная алгебры комплекс-
 ных чисел 106, 477
 — — теории Галуа 471
 — отделимости 293
 — Ферма малая 178
 — Фробениуса 530
 — Чевы 282
 — Штейница 299
 — Эйлера 178, 253
 Тождество Якоби 19, 41, 550
 Топология Зарисского 427, 428

- Топ n -мерный 494
 Точка аффинного пространства 277
 — бесконечно удаленная 326
 — внутренняя 291
 — граничная 291
 — крайняя 296, 300
 Точки аффинно зависимые 279
 — — независимые 279, 288
 Транзитивность 28
 Транспозиции смежные 174
 Транспозиция 174
 — элементов 79
 Триектор 361
 Трисекция угла 476
- У**
- Угол между векторами 222
 Удвоение куба 476
 Умножение левое на элемент 239
 — матриц 43, 45, 66, 204
 — скалярное 210, 217, 221, 232
 — тензорное векторных пространств 343
 Упорядочение лексикографическое 130, 131
 Уравнение алгебраическое 99
 — линейное 48
 — — однородное 48
 — разрешимое в радикалах 473, 474
 — — — квадратных 474
- Ф**
- Факторалгебра 389
 Факторгруппа 182, 191
 Факторкольцо 387
 Факормножество 29
 Факормодуль 397
 Факторпредставление 481
 Факторпространство 397
 Фигуры эквивалентные 162, 164
- Флаг многогранника 307
 Форма билинейная см. функция билинейная
 — вещественная группы Ли 557
 — жорданова матрицы оператора 264, 267
 — квадратичная см. функция квадратичная
 — линейная см. функция линейная
 Формула Бернсайда 457
 — Кардано 146
 — Лагранжа интерполяционная 152, 207
 — Муавра 26
 — разложения определителя 86
 — Тейлора 104, 208
 Формулы Виета 101, 133
 — Крамера 89
 Функция аффинно-квадратичная 309, 314
 — аффинно-линейная 285, 300
 — билинейная 209, 212
 — — кососимметрическая 211, 212, 220, 221
 — — невырожденная 211
 — — симметрическая 211, 212, 214
 — — — положительно определенная 217
 — дифференцируемая 538
 — квадратичная 212, 250
 — —, канонический вид 250
 — — над полем \mathbb{Z}_p 219
 — — невырожденная 212
 — —, нормальный вид 216, 217
 — — отрицательно определенная 217
 — — положительно определенная 217, 218
 — — эрмитова 231
 — — —, канонический вид 256

- — —, нормальный вид 231
- — — положительно определенная 232
- координатная 206
- линейная 75, 205, 206
- полилинейная (*m*-линейная) 77, 338
- — кососимметрическая 78, 361
- — симметрическая 354, 358
- полуторалинейная 230
- — косозэрмитова 231
- — невырожденная 231
- — эрмитова 231
- центральная 509
- Эйлера 178, 392

Х

- Характер представления 509
- Характеристика поля 32, 169

Ц

- Целая часть дроби 150
- Центр алгебры ассоциативной 503
- — Ли 554
- аффинно-квадратичной функции 310, 311
- группы 444
- — Ли 552
- квадратики 311, 313
- тела 523
- тяжести выпуклого множества 494
- — системы точек 281
- Централизатор элемента 457
- Цикл (в симметрической группе) 169
- Циклы независимые 169

Ч

- Частное неполное 96

- Частное (отношение) элементов 16
- Четность перестановки 79
- Число комплексное 24
- —, алгебраическая форма 24
- —, вещественная часть 24
- —, извлечение корня 26
- —, мнимая часть 24
- —, тригонометрическая форма 25
- мнимое 24
- сочетаний 34
- — с повторениями 128
- Ферма 477
- целое алгебраическое 418
- — гауссово 119
- чисто мнимое 24
- Член старший многочлена от n переменных 131

Э

- Экспонента группы 385
- оператора 273
- Элемент алгебраический 407, 410
- — целый 417
- ведущий 50
- единичный (в группе) 15, 156, 157
- единичный (в кольце) 18
- матричный представления 508
- нильпотентный 406, 496, 498
- нулевой (в группе) 14
- обратимый 20
- обратный (в группе) 15, 156, 157
- — правый 158
- обратный (в кольце) 20
- порождающий 171
- представимый в радикалах 473
- простой 121

Элемент противоположный
(в группе) 14
— трансцендентный 407
Элементы абелевой группы линейно независимые 373
— алгебраически зависимые 408, 421
— ассоциированные 118
— взаимно простые 120, 433
— группы, сравнимые по модулю подгруппы 175

— сопряженные 454
Эллипсоид 319
Эндоморфизм 184
— Фробениуса 415

Я

Ядро билинейной функции 211
— гомоморфизма 184, 388
— линейного отображения 200
— неэффективности 452, 456

Эрнест Борисович Винберг

КУРС АЛГЕБРЫ

Издательство Московского центра
непрерывного математического образования
119002, Москва, Большой Власьевский пер., 11. Тел. (499) 241-74-83

Подписано в печать 07.09.2012 г. Формат $60 \times 90^{1/16}$.
Бумага офсетная. Печать офсетная. Печ. л. 37. Тираж 2000. Заказ .
Отпечатано по технологии СтР в ИПК ООО «Ленинградское издательство».
194044, Санкт-Петербург, ул. Менделеевская, д. 9.
Телефон/факс: (812) 495-56-10.

Книги издательства МЦНМО можно приобрести в магазине
«Математическая книга», Большой Власьевский пер., д. 11.
Тел. (499) 241-72-85. E-mail: biblio@mccme.ru
