Э.Б.Винберг

НАЧАЛА АЛГЕБРЫ

УРСС

МОСКВА • 1998

МК НМУ
МЦНМО

Эрнест Борисович Винберг, доктор физ.-мат. наук

Винберг Э.Б.
Начала алгебры. М.: МЦНМО, МК НМУ, «УРСС», 1998. — 192 с.

Книга написана по мотивам лекций, прочитанных автором студентам 1 курса Математического колледжа НМУ в осеннем семестре 1992/93 учебного года. По сравнению с предыдущим изданием книга подверглась существенной переработке, что позволяет пользоваться ею как учебником.

Издательство «УРСС». 111672, г. Москва, ул. Новокасинская, 27/174.
Лицензия ЛР № 063377 от 23.05.94 г. Подписано к печати 20.02.98.
Формат 60×84/16. Печ. л. 12. Зак. № 25
Отпечатано в ТОО «Типография ПЭМ». 121471, г. Москва, Можайское шоссе, 25.

ISBN 5-88417-133-1

© Э.Б.Винберг, 1998
© МЦНМО, МК НМУ, 1998
© Издание: «УРСС», 1998

Оглавление

Предисловие

Настоящая книга написана по мотивам 6 лекций, прочитанных автором студентам 1 курса Математического колледжа НМУ в осеннем семестре 1992/93 учебного года. Конспективное изложение этих лекций было опубликовано тогда же. Настоящая книга существенно отличается от него, во-первых, степенью подробности, позволяющее пользоваться ею как учебником, и, во-вторых, тем, что в ней добавлены начала линейной алгебры и некоторые другие более мелкие разделы (но кое-что и выкинуто, например, факторкольца). При ее написании автор опирался на свой 35-летний опыт преподавания алгебры на механико-математическом факультете МГУ.

Нумерация теорем, предложений, лемм, примеров, задач и замечаний производится в пределах каждого параграфа. Система ссылок поясняется следующими примерами: в тексте §3.2 «теорема 1» означает теорему 1 того же параграфа, «теорема 1.4» — теорему 4 §3.1, а «теорема 1.4.2» — теорему 2 §1.4.

Глава 1. 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 этой главе состоится лишь внешнее знакомство читателя с многими из алгебраических структур, рассматриваемых в курсе. Более глубокое их понимание будет приходить в процессе дальнейшего чтения книги и решения задач.

§1.1. 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

$$M \times M \to M.$$

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,
$\mathbb{Z}_+ = \mathbb{N} \cup \{0\}$ — множество неотрицательных целых чисел,
$\mathbb{Z}$ — множество всех целых чисел,
$\mathbb{Q}$ — множество рациональных чисел,
$\mathbb{R}_+$ — множество неотрицательных действительных чисел,
$\mathbb{R}$ — множество всех действительных (= 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

$$f\colon N \to M, \quad g\colon P \to N$$

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

$$fg\colon P \to M,$$

определяемое формулой

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

т. е. результат последовательного выполнения сначала g, а потом f. 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 $\mathbb{Z}_+$, сопоставляющую 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. Пусть M — множество с операцией $\circ$, а N — множество с операцией $*$. Алгебраические структуры $(M, \circ)$ и $(N, *)$ называются *изоморфными*, если существует такое биективное отображение

$$f\colon M \to N,$$

что

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

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

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

Пример 3. Отображение

$$a \mapsto 2^a$$

является изоморфизмом множества всех действительных 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 M — множество параллельных переносов плоскости на векторы, параллельные какой-либо фиксированной прямой. Для любого действительного числа a обозначим через t_a элемент множества M, представляющий собой перенос на вектор длины $|a|$ в одном из двух возможных направлений, определяемом знаком a. (Если $a = 0$, то t_a — это тождественное преобразование.) Легко видеть, что

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

где $\circ$ обозначает умножение (композицию) параллельных переносов. Следовательно, отображение $a \mapsto t_a$ является изоморфизмом алгебраических структур $(\mathbb{R}, +)$ и $(M, \circ)$.

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 не безразличен для фактического решения какой-либо задачи. Какая-то определенная модель может предоставить для этого наибольшее удобство. Например, если какая-то модель имеет геометрический характер, то она позволяет применить геометрические методы.

§1.2. Абелевы группы

Сложение действительных чисел обладает следующими свойствами:

(С1) $a + b = b + a$ (коммутативность);
(С2) $(a + b) + c = a + (b + c)$ (ассоциативность);
(С3) $a + 0 = a$;
(С4) $a + (-a) = 0$.

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

$$x + a = b$$

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

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

Пользуясь свойствами (С2)–(С4), 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 уравнение показывает, что $b + (-a)$ действительно является решением:

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

Умножение действительных 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 (С1)–(С4), с единственной оговоркой, что в (У4) мы предполагаем, что $a \neq 0$, в то время как в (С4) никаких ограничений на a нет. Поэтому приведенный выше вывод из свойств (С1)–(С4) операции вычитания, будучи переведен на язык умножения, даст вывод из свойств (У1)–(У4) 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 действительных 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 (С1)–(С4), 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 соответ-

ствующим образом. Это позволяет использовать разработанную терминологию и систему обозначений, относящиеся к операциям над действительными числами, а также вызывает полезные ассоциации.

Приведем вначале определение абелевой группы, использующее язык сложения.

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

1) $a+b=b+a \quad \forall a,b$ (*коммутативность*);

2) $(a+b)+c=a+(b+c) \quad \forall a,b,c$ (*ассоциативность*);

3) существует такой элемент 0 (*нуль*), что $a+0=a \quad \forall a$;

4) для любого a существует такой элемент $-a$ (*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 $b-a$.

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 обозначим через $\mathbb{R}^n$. Определим сложение строк по правилу

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

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

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

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* называется множество A с операцией умножения, обладающей следующими свойствами:

1) $ab=ba \quad \forall a,b$ (*коммутативность*);

2) $(ab)c=a(bc) \quad \forall a,b,c$ (*ассоциативность*);

3) существует такой элемент e (*единица*), что $ae=a \quad \forall a$;

4) для любого a существует такой элемент a^{-1} (*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 $\frac{b}{a}$.

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.

§1.3. 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 действительными 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 $(a+b)c = ac + bc \quad \forall a, b, c$ (*дистрибутивность умножения относительно сложения*).

Выведем некоторые следствия аксиом кольца, не входящие в число следствий аксиом аддитивной абелевой группы, перечисленных в §1.2.

1) $a0 = 0a = 0 \quad \forall a$. 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) $a(-b) = (-a)b = -ab \quad \forall a, b$. В самом деле,

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

и, аналогично, $ab + (-a)b = 0$.

3) $a(b-c) = ab - ac$ и $(a-b)c = ac - bc \quad \forall a, b, c$. 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 (если такой существует) называется *единицей*, если

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

Так же, как в случае мультипликативной абелевой группы, доказывается, что в кольце не может быть двух различных единиц.

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 оставить лишь одно. Аналогичное замечание относится к определению единицы.

Пример 1. Числовые множества $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$ являются коммутативными ассоциативными кольцами с единицей относительно обычных операций сложения и умножения.

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

Пример 3. Множество всех функций, определенных на заданном подмножестве числовой прямой, является коммутативным ассоциативным кольцом с единицей относительно обычных операций сложения и умножения функций.

Пример 4. Множество векторов пространства с операциями сложения и векторного умножения является некоммутативным и неассоциативным кольцом. Однако в нем выполняются следующие тождества, которые в некотором смысле заменяют коммутативность и ассоциативность:

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

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

Антикоммутативность очевидна из определения векторного умножения. По поводу проверки тождества Якоби см. пример 8.2.

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

$$A \triangle B = (A \setminus B) \cup (B \setminus A)$$

и пересечения, взятых в качестве сложения и умножения соответственно. Доказать, что это кольцо коммутативно и ассоциативно.

Элемент a^{-1} кольца с единицей (если такой существует) 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. Элемент, имеющий обратный, называется *обратимым*.

Определение 2. *Полем* 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 действительных 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.)

В любом поле выполнено следующее важное свойство:

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

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 равенств:

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

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 подмножества X_1 и X_2 и положим

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

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^*.

§1.4. Подгруппы, подкольца и подполя

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

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

В этом случае операция $\circ$ определена в множестве N и превращает его в некоторую алгебраическую структуру. Если операция $\circ$ в M обладает некоторым свойством, имеющим характер тождественного соотношения (например, коммутативна или ассоциативна), 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, так как не содержит противоположных элементов ко всем своим элементам, кроме нуля.

Определение 1. Подмножество B аддитивной абелевой группы A называется *подгруппой*, если

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

Замечание. 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 1. В аддитивной группе $\mathbb{R}$ имеется следующая цепочка подгрупп:

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

Пример 2. 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 1′. Подмножество B мультипликативной абелевой группы A называется *подгруппой*, если

1) B замкнуто относительно умножения;

2) $a \in B \implies a^{-1} \in B$;

3) $e \in B$.

Пример 3. В группе $\mathbb{R}^*$ имеется следующая цепочка подгрупп:

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

Соображения, с которых начинается этот параграф, могут быть распространены на алгебраические структуры с несколькими операциями. Таким образом мы приходим к следующим понятиям подкольца и подполя.

Определение 2. 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 4. Цепочка подгрупп аддитивной группы $\mathbb{R}$, приведенная в примере 1, является в то же время цепочкой подколец.

Пример 5. При любом $n \in \mathbb{Z}_+$ множество $n\mathbb{Z}$ является подкольцом кольца $\mathbb{Z}$. (Ср. задачу 1.)

Задача 2. Доказать, что все конечные подмножества множества M образуют подкольцо кольца 2^M из задачи 3.1.

Определение 3. Подмножество L поля K называется *подполем*, если

1) L является подкольцом кольца K;

2) $a \in L$, $a \neq 0 \implies a^{-1} \in L$;

3) $1 \in L$.

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

Пример 6. Поле $\mathbb{Q}$ является подполем поля $\mathbb{R}$.

Задача 3. 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 4. Доказать, что поле $\mathbb{Q}$ не имеет нетривиальных подполей (т. е. отличных от него самого).

§1.5. 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 действительных чисел приводит к необходимости расширить его до большего поля, называемого полем комплексных чисел.

Для того чтобы лучше понять, что такое поле комплексных чисел, нужно прежде подумать над тем, что такое поле действительных чисел. Строгое построение поля действительных чисел обычно приводится в курсе анализа. Мы не будем входить в его детали. Однако заметим, что имеется несколько определений действительных 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 действительных 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 действительных 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 доказывается, что этими свойствами он определяется однозначно с точностью до изоморфизма, и с помощью какой-либо конструкции доказывается его существование. В случае поля действительных чисел такими аксиомами (помимо аксиом поля) могут быть аксиомы порядка, аксиома Архимеда и аксиома непрерывности.

Замечание 1. Нетрудно доказать, что любые две модели поля действительных чисел не просто изоморфны, но между ними имеется *единственный* изоморфизм. (Доказательство сводится к тому, 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. *Полем комплексных чисел* называется всякое поле $\mathbb{C}$, обладающее следующими свойствами:

1) оно содержит в качестве подполя поле $\mathbb{R}$ действительных 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.

Теорема 1. *Поле комплексных чисел существует и единственно с точностью до изоморфизма, переводящего все действительные числа в себя. Каждое комплексное число однозначно представляется в виде $a + bi$, где $a, b \in \mathbb{R}$, а i — (фиксированный) элемент, квадрат которого равен -1.*

Доказательство. 1) Пусть $\mathbb{C}$ — какое-то поле комплексных чисел (если оно существует). Рассмотрим его подмножество

$$K = \{\, a + bi : a, b \in \mathbb{R} \}.$$

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

$$(a_1 + b_1 i) + (a_2 + b_2 i) = (a_1 + a_2) + (b_1 + b_2)i, \tag{1}$$

$$(a_1 + b_1 i)(a_2 + b_2 i) = (a_1 a_2 - b_1 b_2) + (a_1 b_2 + b_1 a_2)i. \tag{2}$$

Решая соответствующие уравнения, находим также, что

$$-(a + bi) = (-a) + (-b)i, \tag{3}$$

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

Формулы (1)–(4) показывают, что K — подполе. Так как K, очевидно, содержит $\mathbb{R}$ и i, то $K = \mathbb{C}$.

Таким образом, каждый элемент поля $\mathbb{C}$ представляется в виде $a+bi$, где $a, b \in \mathbb{R}$. Докажем, что такое представление единственно. Пусть

$$a_1 + b_1 i = a_2 + b_2 i \quad (a_1, b_1, a_2, b_2 \in \mathbb{R}).$$

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

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

что и требовалось доказать.

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

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

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

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

$$(a_1, b_1) + (a_2, b_2) = (a_1 + a_2, b_1 + b_2),$$

$$(a_1, b_1)(a_2, b_2) = (a_1 a_2 - b_1 b_2, a_1 b_2 + b_1 a_2),$$

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. 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

$$(a, b)(a/(a^2 + b^2), -b/(a^2 + b^2)) = (1, 0).$$

Следовательно, $\mathbb{C}$ — поле.

Далее, имеем

$$(a_1, 0) + (a_2, 0) = (a_1 + a_2, 0),$$

$$(a_1, 0)(a_2, 0) = (a_1 a_2, 0),$$

т. е. операции над парами вида $(a, 0)$ сводятся к соответствующим операциям над их первыми компонентами. Условимся отождествлять пару $(a, 0)$ с действительным 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 $a,b \in \mathbb{R}$. Поэтому, если какое-либо подполе $K \subset \mathbb{C}$ содержит $\mathbb{R}$ и i, то $K = \mathbb{C}$. Следовательно, $\mathbb{C}$ — поле 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 *действительной частью* числа c и обозначается через $\operatorname{Re} c$, а число b называется *мнимой частью* числа c и обозначается через $\operatorname{Im} c$. Комплексные числа, не являющиеся действительными, называются *мнимыми*; числа вида bi, где $b \in \mathbb{R}$, называются *чисто мнимыми*.

Если в первой части доказательства теоремы в качестве $\mathbb{C}'$ взять то же поле $\mathbb{C}$, а в качестве i' — элемент $-i$, то мы получим, что отображение

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

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 $c \mapsto \overline{c}$ есть автоморфизм поля комплексных чисел. Очевидно, что $\overline{\overline{c}} = c$.

Действительные числа характеризуются тем, что они совпадают со своими сопряженными. Отсюда следует, что для любого $c \in \mathbb{C}$ числа $c + \overline{c}$ и $c\overline{c}$ действительны. В самом деле,

$$\overline{c + \overline{c}} = \overline{c} + c = c + \overline{c}, \quad \overline{c\overline{c}} = \overline{c}c = c\overline{c}.$$

Легко видеть, что если $c = a + bi$ $(a,b \in \mathbb{R})$, то

$$c + \overline{c} = 2a, \quad c\overline{c} = a^2 + b^2. \tag{5}$$

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 бывает удобнее представление комплексных чисел точками, иногда

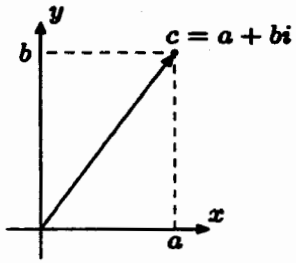

Рис. 1

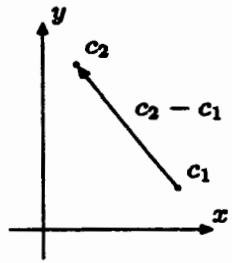

Рис. 2

— 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 изображается 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 $\arg c$.

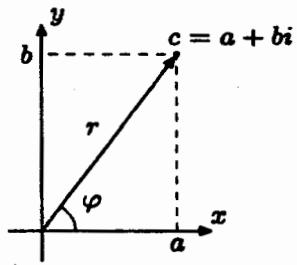

Рис. 3

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

Очевидно, что

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

откуда

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

Такое представление комплексного числа называется его *тригонометрической формой*. Так как тригонометрическая форма дан-

ного комплексного числа определена однозначно с точностью до прибавления к φ целого кратного 2π, то при $r_1, r_2 > 0$

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

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*:

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

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

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,

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

Окончательно получаем

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

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 $k = 0, 1, \ldots, n-1$. В геометрическом изображении эти числа располагаются в вершинах *правильного n-угольника* с центром в начале координат (см. рис. 4, где изображен случай $n = 6$).

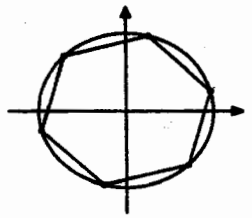

Рис. 4

§1.6. 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 определяются 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-нибудь 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.

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

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

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

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

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 выполняются следующие свойства:

1) aRa (*рефлексивность*);

2) $aRb \Longrightarrow bRa$ (*симметричность*);

3) aRb и $bRc \Longrightarrow aRc$ (*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 $a \underset{R}{\sim} b$ или просто $a \sim b$.

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

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

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

$$R(a) \cap R(b) \neq \varnothing \implies R(a) = R(b).$$

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

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

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

называется *отображением факторизации*.

Так, в третьем из приведенных выше примеров классы эквивалентности — это множества жильцов одного дома. Факторомножество отождествляется с множеством домов, а отображение факторизации — это отображение, сопоставляющее каждому человеку дом, в котором он живет. В пятом примере классы эквивалентности — это множества окружностей одного радиуса, факторомножество отождествляется с множеством положительных чисел, а отображение факторизации — это отображение, сопоставляющее 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

$$a \underset{R}{\sim} a', \; b \underset{R}{\sim} b' \implies a * b \underset{R}{\sim} a' * b'.$$

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

$$R(a) * R(b) = R(a * b). \tag{6}$$

В словесном выражении это определение выглядит так: чтобы произвести операцию над какими-либо двумя классами эквивалентности, надо выбрать из них произвольных представителей, произвести операцию над ними и взять тот класс, в котором будет лежать получившийся элемент. То, 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 $0, 1, \ldots, n-1$ 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.

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

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

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

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

$$a \equiv a' \pmod{n}, \qquad b \equiv b' \pmod{n}.$$

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 $\mathbb{Z}_{125}$:

$$[2]^7 = [128] = [3],$$
$$[2]^{35} = ([2]^7)^5 = [3]^5 = [243] = [-7],$$
$$[2]^{50} = [2]^{35}([2]^7)^2[2] = [-7][3]^2[2] = [-126] = [-1],$$
$$[2]^{100} = ([2]^{50})^2 = [1].$$

Полученный результат означает, что

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

Отсюда нетрудно вывести, что

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

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 не обратим.

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

$$[0]_n, [a]_n, [2a]_n, \ldots, [(n-1)a]_n. \tag{7}$$

Докажем, что все они различны. В самом деле, если $[ka]_n = [la]_n$ $(0 \leqslant k < l \leqslant n-1)$, 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 ряду элементов (7) встречаются все элементы кольца $\mathbb{Z}_n$, в том числе $[1]_n$, а это и означает, что элемент $[a]_n$ обратим.

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

$$\underbrace{1+1+\ldots+1}_{n} = 0. \tag{8}$$

(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 $\operatorname{char} K$.

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

$$\underbrace{a + a + \ldots + a}_{n} = \underbrace{(1 + 1 + \ldots + 1)}_{n} a = 0a = 0.$$

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

$$\underbrace{1 + 1 + \ldots + 1}_{n} = \underbrace{(1 + 1 + \ldots + 1)}_{k}\underbrace{(1 + 1 + \ldots + 1)}_{l} = 0$$

и, значит, либо $\underbrace{1 + 1 + \ldots + 1}_{k} = 0$, либо $\underbrace{1 + 1 + \ldots + 1}_{l} = 0$, 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 n справедливо тождество

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

В самом деле, по формуле бинома Ньютона

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

Однако при $0 < k < n$

$$C_n^k = \frac{n(n-1)\cdots(n-k+1)}{k!}$$

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

Задача. Вывести отсюда, что в поле $\mathbb{Z}_n$ (n простое) $a^n = a$ для любого a. (Другое доказательство последнего факта, называемого малой теоремой Ферма, будет дано в §4.4.)

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 $\underbrace{1 + 1 + \ldots + 1}_{k}$ данного 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,

$$x_1 = \frac{[-1] + [4]}{[2]} = \frac{[3]}{[2]} = \frac{[14]}{[2]} = [7],$$

$$x_2 = \frac{[-1] - [4]}{[2]} = \frac{[-5]}{[2]} = \frac{[6]}{[2]} = [3].$$

§1.7. 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) сопоставляет вектор. В общем определении векторного пространства дело обстоит так же, но действительные числа заменяются элементами произвольного (но фиксированного) поля.

Определение 1. *Векторным 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) $\lambda(a+b)=\lambda a+\lambda b \quad \forall \lambda \in K,\ a,b \in V$;
3) $(\lambda+\mu)a=\lambda a+\mu a \quad \forall \lambda,\mu \in K,\ a \in V$;
4) $(\lambda\mu)a=\lambda(\mu a) \quad \forall \lambda,\mu \in K,\ a \in V$;
5) $1a=a \quad \forall a \in V$.

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 (т. е. подполе поля комплексных чисел).

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 (соотв. трехмерного евклидова пространства) мы будем обозначать через E^2 (соотв. 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

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

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

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

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 следствием только аксиом абелевой группы. Все они доказываются аналогично похожим на них следствиям аксиом кольца (см. §1.3). 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) $\lambda 0=0 \quad \forall \lambda \in K$ (здесь 0 — нулевой вектор).
2) $\lambda(-a)=-\lambda a \quad \forall \lambda \in K,\ a \in V$.
3) $\lambda(a-b)=\lambda a-\lambda b \quad \forall \lambda \in K,\ a,b \in V$.
4) $0a=0 \quad \forall a \in V$ (здесь 0 слева — число, справа — вектор).
5) $(-1)a=-a \quad \forall a \in V$.
6) $(\lambda-\mu)a=\lambda a-\mu a \quad \forall \lambda,\mu \in K,\ a \in V$.

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

1) U является подгруппой аддитивной группы V;
2) $a \in U \implies \lambda a \in U \quad \forall \lambda \in K$.

Замечание. В определении подгруппы требуется, чтобы

$$a \in U \implies -a \in U.$$

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

отображение

$$f\colon V \to U,$$

что

1) $f(a+b) = f(a) + f(b) \quad \forall a, b \in V$;
2) $f(\lambda a) = \lambda f(a) \quad \forall \lambda \in K,\ a \in V$.

Само отображение f 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

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

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

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

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

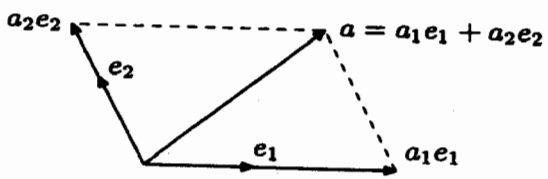

Рис. 5

Пример 7. *Единичные строки*

$$e_1 = (1, 0, \ldots, 0),$$
$$e_2 = (0, 1, \ldots, 0),$$
$$\ldots\ldots\ldots\ldots$$
$$e_n = (0, 0, \ldots, 1)$$

составляют базис пространства K^n. Координатами строки $a = (a_1, a_2, \ldots, a_n)$ в этом базисе служат числа $a_1, a_2, \ldots, a_n$. 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 действительная и мнимая части.

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

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

$$f\colon V \to K^n,$$

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

$$a = a_1e_1 + a_2e_2 + \ldots + a_ne_n,$$
$$b = b_1e_1 + b_2e_2 + \ldots + b_ne_n,$$

то

$$a + b = (a_1 + b_1)e_1 + (a_2 + b_2)e_2 + \ldots + (a_n + b_n)e_n,$$
$$\lambda a = (\lambda a)e_1 + (\lambda a_2)e_2 + \ldots + (\lambda a_n)e_n.$$

Отсюда следует, что f — изоморфизм.

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

§1.8. 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. *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) $(\lambda a)b = a(\lambda b) = \lambda(ab) \quad \forall \lambda \in K,\ a, b \in A$.

Замечание. 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 7.2) 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. Доказать, что кольцо 2^M из задачи 3.1 превращается в алгебру над полем $\mathbb{Z}_2$, если определить в нем умножение на элементы этого поля по правилам

$$0A = \varnothing, \quad 1A = A \quad \forall A \in 2^M.$$

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

$$a = a_1e_1 + a_2e_2 + \ldots + a_ne_n = \sum_{i=1}^{n} a_ie_i,$$

$$b = b_1e_1 + b_2e_2 + \ldots + b_ne_n = \sum_{i=1}^{n} b_ie_i$$

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

$$ab = \sum_{i=1}^{n} a_i(e_ib) = \sum_{i=1}^{n} a_i\left(\sum_{j=1}^{n} b_j(e_ie_j)\right) = \sum_{i,j=1}^{n} a_ib_j(e_ie_j).$$

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

Если умножение базисных векторов коммутативно, т. е.

$$e_ie_j = e_je_i \quad \forall i, j,$$

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

$$ab = \sum_{i,j} a_ib_j(e_ie_j) = \sum_{i,j} b_ja_i(e_je_i) = ba.$$

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

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

то и умножение в алгебре A в целом ассоциативно.

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

$$ab = \sum_{i,j} a_ib_je_{ij}$$

и тем самым превратить V в алгебру.

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

Проверка коммутативности и ассоциативности умножения в $\mathbb{C}$ сводится к тривиальной проверке коммутативности и ассоциативности умножения элементов 1 и i.

Пример 5. В ортонормированном (т. е. состоящее 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 антикоммутативно 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 в поле, всякий ненулевой элемент обратим. Таким образом, это «некоммутативное поле».

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

§1.9. 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

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

(Смысл этого определения выяснится в §2.3)

Подчеркнем, что произведение двух матриц определено только тогда, когда их размеры согласованы, а именно, когда число столбцов первой матрицы равно числу строк второй.

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.

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

Умножение матриц ассоциативно в том смысле, что

$$(AB)C = A(BC), \tag{9}$$

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:

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

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, находящееся вне (главной) диагонали, равны нулю. Умножение на диагональные матрицы выглядит особенно просто:

$$\begin{pmatrix}a_1 & & 0\\ & a_2 & \\ & & \ddots\\ 0 & & a_n\end{pmatrix}\begin{pmatrix}b_{11} & b_{12} & \dots & b_{1p}\\ b_{21} & b_{22} & \dots & b_{2p}\\ \dots & \dots & \dots & \dots\\ b_{n1} & b_{n2} & \dots & b_{np}\end{pmatrix} = \begin{pmatrix}a_1b_{11} & a_1b_{12} & \dots & a_1b_{1p}\\ a_2b_{21} & a_2b_{22} & \dots & a_2b_{2p}\\ \dots & \dots & \dots & \dots\\ a_nb_{n1} & a_nb_{n2} & \dots & a_nb_{np}\end{pmatrix}.$$

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

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

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

Диагональная матрица вида

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

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

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

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

Следующие очевидные свойства связывают операцию умножения матриц с другими операциями:

$$A(B+C) = AB + AC, \quad (A+B)C = AC + BC, \tag{11}$$
$$(\lambda A)B = A(\lambda B) = \lambda(AB) \quad \forall\lambda \in K. \tag{12}$$

(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 $L_n(K)$.

Отметим некоторые «отрицательные» свойства алгебры $L_n(K)$ при $n \geqslant 2$. (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 §1.3). 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} $(i,j = 1,\dots,n)$ 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

$$\begin{pmatrix}a&-b\\b&a\end{pmatrix},\quad a,b\in\mathbb{R},$$

образуют подалгебру, изоморфную алгебре комплексных чисел.

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

$$\begin{pmatrix}a&-\overline{b}\\b&\overline{a}\end{pmatrix},\quad a,b\in\mathbb{C},$$

образуют подалгебру, изоморфную алгебре кватернионов (см. пример 8.6).

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

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

Очевидно также, что

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

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. 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».

Глава 2. Начала линейной алгебры

§2.1. Системы линейных уравнений

Пусть K — произвольное (но фиксированное) поле. Допуская вольность речи, мы будем обычно называть его элементы числами. Если читателю трудно представить себе произвольное поле, он может считать, что $K = \mathbb{R}$, хотя объективно этот случай ничуть не проще общего.

Линейным уравнением с неизвестными $x_1, x_2, \ldots, x_n$ над полем K называется уравнение вида

$$a_1x_1 + a_2x_2 + \ldots + a_nx_n = b,$$

где *коэффициенты* $a_1, a_2, \ldots, a_n$ и *свободный член* b суть элементы поля K. Линейное уравнение называется *однородным*, если $b = 0$.

Система m линейных уравнений с n неизвестными в общем виде записывается следующим образом:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m. \end{cases} \tag{13}$$

Матрица

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

называется *матрицей коэффициентов*, а матрица

$$\widetilde{A} = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} & b_1 \\ a_{21} & a_{22} & \ldots & a_{2n} & b_2 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ a_{m1} & a_{m2} & \ldots & a_{mn} & b_m \end{pmatrix}$$

— *расширенной матрицей* системы (13).

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 1-го 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 (которое 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.

Определение 1′. *Элементарными преобразованиями строк* матрицы A называются преобразования следующих трех типов:

1) прибавление к одной строке другой, умноженной на число;

2) перестановка двух строк;

3) умножение одной строки на число, отличное от нуля.

Очевидно, что всякое элементарное преобразование системы линейных уравнений приводит к соответствующему элементарному преобразованию ее матрицы коэффициентов и расширенной матрицы.

Покажем теперь, что с помощью элементарных преобразований строк любую матрицу можно привести к достаточно простому виду.

Назовем первый отличный от нуля элемент ненулевой строки $(a_1, a_2, \ldots, a_n) \in K^n$ ее *ведущим элементом*.

Определение 2. Матрица называется *ступенчатой*, если

1) номера ведущих элементов ее ненулевых строк образуют строго возрастающую последовательность;

2) нулевые строки, если они есть, стоят в конце.

Таким образом, ступенчатая матрица — это матрица вида

$$\begin{pmatrix} & \boxed{a_{1j_1}} & & & \\ & & \boxed{a_{2j_2}} & & \\ & & & \ddots & \\ 0 & & & & \boxed{a_{rj_r}} \end{pmatrix}, \qquad (14)$$

в которой элементы $a_{1j_1}, a_{2j_2}, \ldots, a_{rj_r}$, находящиеся в углах ступенчатой линии, отличны от нуля, а все элементы, находящиеся слева и снизу от этой линии, равны нулю. При этом $j_1 < j_2 < \ldots < j_r$.

Теорема 1. *Всякую матрицу путем элементарных преобразований строк можно привести к ступенчатому виду.*

Доказательство. Если данная матрица нулевая, то она уже ступенчатая. Если она ненулевая, то пусть j_1 — номер ее первого ненулевого столбца. Переставив, если нужно, строки, добьемся того, чтобы $a_{1j_1} \neq 0$. После этого прибавим к каждой строке, начиная со второй, первую строку, умноженную на подходящее число, с таким расчетом, чтобы все элементы j_1-го столбца, кроме первого, стали равными нулю. Мы получим матрицу вида

$$\begin{pmatrix} \boxed{a_{1j_1}} & \\ 0 & \boxed{A_1} \end{pmatrix}.$$

Поступая таким же образом с матрицей A_1, мы в конце концов получим матрицу вида (14).

Замечание 1. В этом доказательстве мы обошлись без элементарных преобразований 3-го 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 $j_1, j_2, \ldots, j_r$ не были просто первыми r членами натурального ряда. Такая ситуация является в определенном смысле исключительной. Например, $j_1 \neq 1$ только при условии, что первый столбец исходной матрицы нулевой. Как правило,

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

В этом случае матрица (14) называется *трапецоидальной*.

Применим доказанную теорему к решению систем линейных уравнений.

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

Из теоремы 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 *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.

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

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

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 — x_{n-1} и т. д. Следовательно, система имеет единственное решение.

3-й случай: $\tilde{r} = r < n$. Пусть в этом случае $j_1, j_2, \ldots, j_r$ — номера ведущих коэффициентов ненулевых уравнений системы. Неизвестные $x_{j_1}, x_{j_2}, \ldots, x_{j_r}$ назовем *главными*, а остальные — *свободными*. После отбрасывания нулевых уравнений и перенесения членов со свободными неизвестными в правую часть получается строго треугольная система относительно главных неизвестных. Решая ее, как в предыдущем случае, находим выражения главных неизвестных через свободные. Эти выражения называют *общим решением* системы. Все решения системы получаются из общего решения подстановкой каких-то значений свободных неизвестных. Поскольку эти значения могут выбираться произвольно, система имеет, во всяком случае, более одного решения, а если поле K бесконечно — то бесконечно много решений.

Совместная система линейных уравнений называется *определенной*, если она имеет единственное решение, и *неопределенной*, если она имеет более одного решения. В последнем случае, как следует из проведенного выше анализа, она имеет бесконечно много решений, если только поле K бесконечно. Ее общее решение с точностью до перенумерации неизвестных имеет вид

$$\begin{cases} x_1 = c_{11}x_{r+1} + c_{12}x_{r+2} + \ldots + c_{1,n-r}x_n + d_1, \\ x_2 = c_{21}x_{r+1} + c_{22}x_{r+2} + \ldots + c_{2,n-r}x_n + d_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ x_r = c_{r1}x_{r+1} + c_{r2}x_{r+2} + \ldots + c_{r,n-r}x_n + d_r. \end{cases} \tag{15}$$

Пример 2. Решим систему уравнений

$$\begin{cases} x_1 + 2x_2 + x_3 = 2, \\ x_1 + 3x_2 + 2x_3 - x_4 = 4, \\ 2x_1 + x_2 - x_3 + 3x_4 = -2, \\ 2x_1 - 2x_3 + x_4 = 1. \end{cases}$$

расширенной матрицей которой служит матрица из примера 1. Вычисления, проведенные в примере 1, показывают, что данная система эквивалентна ступенчатой системе

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

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

Замечание 4. 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

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

Перенося 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 определенна, то она имеет только нулевое решение, если неопределенна — 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 $r \leqslant m$, 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.*

Неопределенные системы линейных уравнений могут иметь разную «степень неопределенности», каковой естественно считать число свободных неизвестных в общем решении системы. Так, прямая в пространстве задается системой линейных уравнений с одним свободным неизвестным, а плоскость — системой (из одного уравнения) с двумя свободными неизвестными. Ясно, что это принципиально разные случаи. Однако одна и та же система линейных уравнений может допускать различные общие решения, в которых разные неизвестные играют роль свободных, и закономерен вопрос, будет ли число свободных неизвестных всегда одним и тем же. Положительный ответ на этот вопрос дается с помощью понятия размерности векторного пространства, которое будет введено в следующем параграфе.

В оставшейся части этого параграфа мы интерпретируем метод Гаусса на языке умножения матриц.

Прежде всего, если обозначить через X столбец неизвестных, а через B — столбец свободных членов, то систему (13) можно переписать в следующей матричной форме:

$$AX = B. \tag{16}$$

Действительно, матрица AX согласно правилу умножения матриц есть столбец высоты m, i-й элемент которого равен

$$a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{in}x_n.$$

Приравнивая этот элемент i-му элементу столбца B, мы получаем как раз i-е уравнение системы (13).

Пусть U — какая-либо квадратная матрица порядка m. Умножая обе части уравнения (16) слева на U, мы получаем уравнение

$$UAX = UB. \tag{17}$$

Очевидно, что всякое решение уравнения (16) удовлетворяет и уравнению (17). Если же матрица U обратима, то умножение слева на U^{-1} осуществляет обратный переход от уравнения (17) к уравнению (16) и, следовательно, эти уравнения эквивалентны.

Уравнению (17) соответствует система линейных уравнений с матрицей коэффициентов UA и столбцом свободных членов UB. Легко видеть, что расширенная матрица этой системы равна $U\widetilde{A}$.

Далее, непосредственно проверяется, что элементарные преобразования строк какой-либо матрицы A равносильны ее умножению слева на так называемые ***элементарные матрицы*** следующих трех типов:

1) $\begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & \cdots & c \\ & & & \ddots & \vdots \\ & & & & 1 \end{pmatrix}$ (i-я строка, j-й столбец) $= E + cE_{ij} \quad (i \neq j)$;

2) $\begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & 0 & \cdots & 1 \\ & & \vdots & \ddots & \vdots \\ & & 1 & \cdots & 0 \\ & & & & & \ddots \\ & & & & & & 1 \end{pmatrix}$ (i-я, j-й строки; i-й, j-й столбцы) $= P_{ij} \quad (i \neq j)$;

3) $\begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & c & & \\ & & & \ddots & \\ & & & & 1 \end{pmatrix}$ (i-я строка, i-й столбец) $= Q_i(c) \quad (c \neq 0)$.

(Все элементы этих матриц, не выписанные явно, — такие же, как у единичной матрицы.)

Так, например, умножение матрицы A слева на $E + cE_{ij}$ $(i \neq j)$ приводит к тому, что к i-й строке прибавляется j-я строка, умноженная на c (а прочие строки не изменяются).

Все элементарные матрицы обратимы, причем обратные к ним матрицы суть элементарные матрицы, соответствующие обратным элементарным преобразованиям:

$$(E + cE_{ij})^{-1} = E - cE_{ij},$$
$$P_{ij}^{-1} = P_{ij}, \qquad Q_i(c)^{-1} = Q_i(c^{-1}).$$

Метод Гаусса в матричной интерпретации состоит в последовательном умножении уравнения (16) слева на элементарные матрицы, имеющем целью приведение матрицы A (а также расширенной матрицы $\widetilde{A}$) к ступенчатому виду.

Используя вместо элементарных матриц какие-либо другие матрицы, можно получить другие методы решения систем линейных уравнений, которые, быть может, не столь просты в теоретическом отношении, но, скажем, более надежны при приближенных вычислениях (в случае $K = \mathbb{R}$). Таков, например, метод вращений, при котором в качестве U берутся матрицы вида

$$\begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & \cos\alpha & \cdots & -\sin\alpha & & \\ & & \vdots & \ddots & \vdots & & \\ & & \sin\alpha & \cdots & \cos\alpha & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix} \begin{matrix} \\ \\ i \\ \\ j \\ \\ \\ \end{matrix}$$
$$\qquad\qquad i \qquad\qquad j$$

§2.2. Базис и размерность векторного пространства

Представление о размерности пространства есть одна из фундаментальных идей математики. В разных разделах математики оно (как и представление о самом пространстве) принимает разные формы. В этом параграфе мы дадим определение размерности векторного пространства и исследуем связанные с этим вопросы.

В §1.7 мы ввели понятие базиса векторного пространства и доказали, что векторное пространство над полем K, имеющее базис из n векторов, изоморфно пространству строк K^n. Размерность векторного пространства определяется как число векторов в его

базисе. Однако перед тем как дать такое определение, необходимо ответить 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 некоторые утверждения, которые важны и сами по себе.

Пусть V — векторное пространство над полем K.

Линейная комбинация

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

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

Определение 1. Векторы $a_1, a_2, \ldots, a_n$ 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 $\{1, 2, \ldots, n\}$ 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 $\lambda_1, \lambda_2, \ldots, \lambda_n$, 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 $a_1, a_2, \ldots, a_n$ означает, иными словами, что равенство

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

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

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

Доказательство. 1) Пусть, например,

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

тогда

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

что показывает линейную зависимость векторов $a_1, a_2, \ldots, a_n$.

2) Обратно, пусть

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

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

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

т. е. a_1 линейно выражается через $a_2, \ldots, a_n$.

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 $a_1, a_2, \ldots, a_n$ линейно независимы. Вектор b линейно выражается через $a_1, a_2, \ldots, a_n$ тогда и только тогда, когда векторы $a_1, a_2, \ldots, a_n, b$ линейно зависимы.*

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

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

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

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

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

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

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

Тогда

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

есть линейная зависимость между $a_1, a_2, \ldots, a_n$.

2) Обратно, пусть

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

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

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

то также

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

что дает другое выражение b через $a_1, a_2, \ldots, a_n$.

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.

Предложение 1 (Основная лемма о линейной зависимости). *Если векторное пространство V порождается n векторами, то всякие $m > n$ векторов пространства V линейно зависимы.*

Доказательство. Пусть $V = \langle a_1, a_2, \ldots, a_n \rangle$ и $b_1, b_2, \ldots, b_m$ $(m > n)$ — какие-то векторы пространства V. Выразим их через $a_1, a_2, \ldots, a_n$:

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

Для любых $\lambda_1, \lambda_2, \ldots, \lambda_m \in K$ получаем отсюда

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

Рассмотрим систему n однородных линейных уравнений с m неизвестными:

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

Если $(\lambda_1, \lambda_2, \ldots, \lambda_m)$ — любое решение этой системы, то

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

С другой стороны, по теореме 2.2 эта система имеет ненулевое решение. Следовательно, векторы $b_1, b_2, \ldots, b_m$ линейно зависимы.

Ввиду леммы 3 определение 1.7.4 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.

Теорема 2. *Все базисы конечномерного векторного пространства V содержат одно и тоже 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.

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 (соотв. E^3) имеет размерность 2 (соотв. 3).

Пример 5. Ввиду примера 1.7.7 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 1.8.6) — 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 1.7.2) имеет размерность n. В самом деле, рассмотрим так называемые *δ-функции* δ_a ($a \in X$) определяемые формулами

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

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, $\delta_{a_1}, \delta_{a_2}, \ldots, \delta_{a_n}$, где $a_1, a_2, \ldots, a_n \in X$ различны; 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 конечномерно, то действительное число определялось бы конечным набором рациональных чисел — своих координат в некотором базисе этого пространства. Но тогда множество всех действительных чисел было бы счетно, 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 в любом (конечном или бесконечном) множестве S векторов конечномерного векторного пространства V имеется максимальное линейно независимое подмножество, т. е. такое линейно независимое подмножество, которое становиться 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 $\{e_1, \ldots, e_k\}$ множества S является базисом линейной оболочки $\langle S \rangle$ этого множества.*

Доказательство. Нужно доказать, что каждый вектор из $\langle S \rangle$ линейно выражается через $e_1, \ldots, e_k$. 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 $e_1, \ldots, e_k$. Для $a \in \{e_1, \ldots, e_k\}$ это очевидно. Для $a \notin \{e_1, \ldots, e_k\}$ это следует из леммы 2.

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 $\dim U \leqslant \dim V$. Более того, если $U \neq V$, то $\dim U < \dim V$.*

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

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. Пусть $f\colon V \to U$ — изоморфизм векторных пространств и $\{e_1, e_2, \ldots, e_n\}$ — базис V, тогда $\{f(e_1), f(e_2), \ldots, f(e_n)\}$ — базис U, так что $\dim V = \dim U$. Обратно, согласно предложению 1.7.1 всякое n-мерное векторное пространство над полем K изоморфно K^n; следовательно, все такие пространства изоморфны между собой.

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 1.7.7). 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 1.7.1) 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.

Совокупность всех базисов n-мерного векторного пространства V может быть описана следующим образом. Пусть $\{e_1, \ldots, e_n\}$ — какой-либо фиксировнный базис. Любая система n векторов $\{e'_1, \ldots, e'_n\}$ может быть тогда задана квадратной матрицей $C = (c_{ij})$, определяемой равенствами

$$e'_j = \sum_i e_i c_{ij} \quad (j = 1, \ldots, n), \tag{18}$$

и называемой *матрицей перехода* от базиса $\{e_1, \ldots, e_n\}$ к системе $\{e'_1, \ldots, e'_n\}$. Согласно этому определению, j-й столбец матрицы C есть столбец координат вектора e'_j в базисе $\{e_1, \ldots, e_n\}$. Поэтому векторы $e'_1, \ldots, e'_n$ линейно независимы (и, значит, составляют базис) тогда и только тогда, когда столбцы матрицы C линейно независимы, т. е. когда матрица C невырожденна. Тем самым установлено взаимно однозначное соответствие между множеством всех базисов пространства V и множеством невырожденных матриц порядка 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 (18) могут быть переписаны в следующей матричной форме:

$$(e'_1, \ldots, e'_n) = (e_1, \ldots, e_n)C. \tag{19}$$

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

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

Положим

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

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 $\{e'_1, \dots, e'_n\}$:

$$X = CX', \tag{20}$$

или, более подробно,

$$x_i = \sum_j c_{ij} x'_j \quad (i = 1, \dots, n). \tag{21}$$

Понятия базиса и размерности могут быть распространены на бесконечномерные векторные пространства. Чтобы это сделать, надо определить, что такое линейная комбинация бесконечной системы векторов. В чисто алгебраической систуации 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 тем самым имеет смысл. На основе этого определения линейной комбинации точно так же, как в случае конечных систем векторов, определяются понятия линейной выражаемости, линейной зависимости и базиса.

Мощность базиса называется размерностью пространства. В частности, 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

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

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

Так же, как предложение 1.7.1, доказывается, 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.

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.

С понятием размерности тесно связаны понятия ранга системы векторов и ранга матрицы.

Определение 4. *Рангом системы векторов* называется размерность ее линейной оболочки. *Рангом матрицы* называется ранг системы ее строк.

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

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 3. *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 (14) с коэффициентами $\lambda_1, \lambda_2, \ldots, \lambda_r$ 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 $\lambda_1, \lambda_2, \ldots, \lambda_r$ равны нулю, что и требовалось доказать.

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.

Задача 10. Доказать *теорему Кронекера – Капелли*: система линейных уравнений совместна тогда и только тогда, когда ранг ее расширенной матрицы равен рангу матрицы коэффициентов.

§2.3. Линейные отображения

В любой алгебраической теории наряду с изоморфизмами рассматривают более общие отображения, называемые в общем случае гомоморфизмами, а в случае векторных пространств — линейными отображениями. В то время как изоморфизмы полностью сохраняют внутренние свойства алгебраических структур и их элементов, гомоморфизмы сохраняют их лишь частично.

Определение 1. Пусть V и U — векторные пространства над полем K. Отображение

$$f: V \to U$$

называется *линейным*, если

1) $f(a+b) = f(a) + f(b) \quad \forall a, b \in V$;
2) $f(\lambda a) = \lambda f(a) \quad \forall \lambda \in K, \ a \in V$.

Это определение отличается от определения изоморфизма векторных пространств тем, что в нем не требуется биективности.

Отметим, что при линейном отображении нулевой вектор переходит в нулевой, а противоположный — в противоположный. В самом деле,

$$f(0) = f(0 \cdot 0) = 0 f(0) = 0,$$
$$f(-a) = f((-1)a) = (-1) f(a) = -f(a).$$

Легко также доказать, что

$$f(a-b) = f(a) - f(b).$$

Пример 1. Поворот есть линейное отображение (и даже изоморфизм) пространства E^2 в себя (рис. 6).

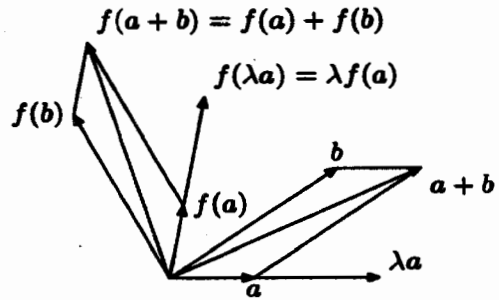

Рис. 6

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 этой прямой.

Пример 4. Отображение

$$\varphi \mapsto \int_a^b \varphi(x)\,dx$$

является линейным отображением пространства непрерывных функций на отрезке $[a,b]$ в поле $\mathbb{R}$, рассматриваемое как векторное пространство над самим собой.

Линейное отображение $f\colon V \to U$ 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

$$f(x) = \sum_i x_i f(e_i).$$

С другой стороны, если $u_i \in U$ $(i \in I)$ — произвольные векторы, то отображение $f\colon V \to U$, определяемое по формуле

$$f(x) = \sum_i x_i u_i,$$

как легко видеть, является линейным и $f(e_i) = u_i$.

Эти соображения позволяют получить аналитическое описание линейных отображений. Сделаем это для пространств строк. Пусть

$$f\colon K^n \to K^m$$

— линейное отображение. Применим его к единичным строкам $e_1, e_2, \ldots, e_n$ пространства K^n (см. пример 1.7.7). Мы получим какие-то строки

$$f(e_j) = (a_{1j}, a_{2j}, \ldots, a_{mj}) \in K^m \quad (j = 1, 2, \ldots, n).$$

Числа a_{ij} $(i = 1, 2, \ldots, m;\ j = 1, 2, \ldots, n)$ образуют матрицу A размера $m \times n$, которая называется *матрицей линейного отображения* f. (Обратите внимание, что координаты строки $f(e_j)$ записываются в j-м столбце матрицы A.)

Для любой строки

$$x = (x_1, x_2, \ldots, x_n) = \sum_j x_j e_j \in K^n$$

имеем:

$$f(x) = \sum_j x_j f(e_j) = \left(\sum_j a_{1j}x_j, \sum_j a_{2j}x_j, \ldots, \sum_j a_{mj}x_j\right).$$

Таким образом, если положить

$$f(x) = y = (y_1, y_2, \ldots, y_m),$$

то $y_1, y_2, \ldots, y_m$ выражается через $x_1, x_2, \ldots, x_n$ по формулам

$$y_i = \sum_{j=1}^{n} a_{ij}x_j \quad (i = 1, 2, \ldots, m). \tag{22}$$

Обратно, если $A = (a_{ij})$ — произвольная матрица размера $m \times n$, то отображение $f\colon K^n \to K^m$, определяемое формулой (22), линейно и его матрица есть A. Тем самым устанавливается взаимно однозначное соответствие между линейными отображениями K^n в K^m и матрицами размера $m \times n$.

Соответственно этому матрица линейного отображения $f\colon V \to U$ произвольных конечномерных векторных пространств определяется следующим образом: в ее j-м столбце стоят координаты образа j-го базисного вектора пространства V. Эта матрица, естественно, зависит от выбора базисов в пространствах V и U.

Пример 5. В пространстве E^2 выберем ортонормированный базис $\{e_1, e_2\}$. Пусть f — поворот на угол α. Тогда (рис. 7)

$$f(e_1) = e_1\cos\alpha + e_2\sin\alpha,$$
$$f(e_2) = -e_1\sin\alpha + e_2\cos\alpha.$$

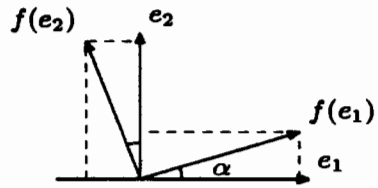

Рис. 7

Это означает, что матрица f есть

$$\begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}. \tag{23}$$

Заметим, что в данном случае $V = U$ и мы использовали один и тот же базис $\{e_1, e_2\}$ в двух качествах: как базис V и как базис U, хотя согласно определению не обязаны были этого делать.

Пример 6. 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

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

В отличие от изоморфизма линейное отображение не обязано быть ни сюръективным, ни инъективным. Нарушение этих свойств приводит к возможности связать с каждым линейным отображением два подпространства — его образ и ядро.

Определение 2. *Образом* линейного отображения $f: V \to U$ называется подмножество

$$\operatorname{Im} f = \{f(a) : a \in V\} \subset U,$$

а *ядром* — подмножество

$$\operatorname{Ker} f = \{\, a \in V : f(a) = 0 \,\} \subset V.$$

Легко видеть, что $\operatorname{Im} f$ — подпространство в U, а $\operatorname{Ker} f$ — подпространство в V. Докажем, например, второе. Если $a, b \in \operatorname{Ker} f$, т. е. $f(a) = f(b) = 0$, то

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

т. е. $a + b \in \operatorname{Ker} f$. Далее, если $a \in \operatorname{Ker} f$, т. е. $f(a) = 0$, то для любого $\lambda \in K$

$$f(\lambda a) = \lambda f(a) = \lambda 0 = 0,$$

т. е. $\lambda a \in \operatorname{Ker} f$. Наконец, $0 \in \operatorname{Ker} f$, так как по доказанному выше $f(0) = 0$.

Пример 7. Ядром отображения проектирования из примера 2 является совокупность векторов, ортогональных плоскости проектирования.

Пример 8. 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.

Теорема 1. *Линейное отображение $f: V \to U$ инъективно тогда и только тогда, когда $\operatorname{Ker} f = 0$. Более точно, для любого $b \in \operatorname{Im} f$ множество решений уравнения*

$$f(x) = b \tag{24}$$

имеет вид $a + \operatorname{Ker} f$, где a — какое-то одно решение этого уравнения.

(Здесь $a + \operatorname{Ker} f$ понимается как совокупность сумм вида $a + y$, где $y \in \operatorname{Ker} f$.)

Заметим, что $\operatorname{Ker} f$ согласно определению есть множество решений уравнения

$$f(x) = 0. \tag{25}$$

Доказательство. Инъективность отображения f означает, что для любого $b \in \operatorname{Im} f$ уравнение (24) имеет единственное решение. Поэтому нам достаточно доказать второе утверждение теоремы.

Пусть $f(a) = b$. Если $y \in \operatorname{Ker} f$, то

$$f(a+y) = f(a) + f(y) = b + 0 = b.$$

Обратно, если $f(x) = b$, то

$$f(x-a) = f(x) - f(a) = b - b = 0,$$

т. е. $y = x - a \in \operatorname{Ker} f$; следовательно,

$$x = a + y \in a + \operatorname{Ker} f.$$

Если $f: K^n \to K^m$ — линейное отображение с матрицей A и $b = (b_1, b_2, \ldots, b_m)$, то уравнение (24) в координатной форме — это не что иное, как система линейных уравнений (13), а уравнение (25) — это система однородных линейных уравнений с теми же коэффициентами при неизвестных:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = 0, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = 0, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = 0. \end{cases} \tag{26}$$

Таким образом, множество решений системы уравнений (26) есть подпространство пространства K^n, а множество решений системы (13), если оно не пусто, есть сумма какого-нибудь одного решения и этого подпространства.

Какова размерность пространства решений системы (26)? Ответ на этот вопрос дает

Теорема 2. *Пусть $f\colon K^n \to K^m$ — линейной отображение с матрицей A. Тогда*

$$\dim \operatorname{Ker} f = n - \operatorname{rk} A.$$

Доказательство. С помощью элементарных преобразований приведем систему (26) к ступенчатому виду. В силу предложения 2 число ненулевых уравнений в этом ступенчатом виде будет равно $r = \operatorname{rk} A$. Поэтому общее решение будет содержать r главных неизвестных и с точностью до перенумерации неизвестных будет иметь вид (ср. (15)):

$$\begin{cases} x_1 = c_{11}x_{r+1} + c_{12}x_{r+2} + \ldots + c_{1,n-r}x_n, \\ x_2 = c_{21}x_{r+1} + c_{22}x_{r+2} + \ldots + c_{2,n-r}x_n, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ x_r = c_{r1}x_{r+1} + c_{r2}x_{r+2} + \ldots + c_{r,n-r}x_n. \end{cases} \tag{27}$$

Придавая по очереди одному из свободных неизвестных $x_{r+1}, x_{r+2}, \ldots, x_n$ значение 1, а остальным — значения 0, мы получим следующие решения системы (26):

$$\begin{aligned} u_1 &= (c_{11}, c_{21}, \ldots, c_{r1}, 1, 0, \ldots, 0), \\ u_2 &= (c_{12}, c_{22}, \ldots, c_{r2}, 0, 1, \ldots, 0), \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ u_{n-r} &= (c_{1,n-r}, c_{2,n-r}, \ldots, c_{r,n-r}, 0, 0, \ldots, 1). \end{aligned}$$

Докажем, что они составляют базис пространства $\operatorname{Ker} f$, откуда и будет следовать утверждение теоремы.

Для любых $\lambda_1, \lambda_2, \ldots, \lambda_{n-r} \in K$ линейная комбинация

$$u = \lambda_1 u_1 + \lambda_2 u_2 + \ldots + \lambda_{n-r} u_{n-r}$$

является решением системы (26), в котором свободные неизвестные имеют значения $\lambda_1, \lambda_2, \ldots, \lambda_{n-r}$. Так как значения главных неизвестных однозначно определяются значениями свободных неизвестных (по формулам (27)), то любое решение системы (26) является линейной комбинацией $u_1, u_2, \ldots, u_{n-r}$. С другой стороны, если $u = 0$, то $\lambda_1 = \lambda_2 = \ldots = \lambda_{n-r} = 0$; следовательно, $u_1, u_2, \ldots, u_{n-r}$ линейно независимы.

Всякий базис пространства решений системы однородных линейных уравнений называется *фундаментальной системой решений*. Предыдущее доказательство дает практический способ построения такой системы решений.

Пусть $f\colon V \to U$ — линейное отображение конечномерных векторных пространств и $\{e_1, e_2, \ldots, e_n\}$ — базис пространства V. Для любого

$$a = a_1e_1 + a_2e_2 + \ldots + a_ne_n \in V$$

имеем

$$f(a) = a_1f(e_1) + a_2f(e_2) + \ldots + a_nf(e_n).$$

Следовательно,

$$\operatorname{Im} f = \langle f(e_1), f(e_2), \ldots, f(e_n)\rangle. \tag{28}$$

Теорема 3. $\dim \operatorname{Im} f + \dim \operatorname{Ker} f = \dim V$.

Доказательство. Выберем базис пространства V специальным образом: сначала выберем базис $\{e_1, \ldots, e_k\}$ подпространства $\operatorname{Ker} f$, а затем дополним его какими-то векторами $e_{k+1}, \ldots, e_n$ до базиса V. Так как по построению $f(e_1) = \ldots = f(e_k) = 0$, то из (28) следует, что

$$\operatorname{Im} f = \langle f(e_{k+1}), \ldots, f(e_n)\rangle.$$

Докажем, что векторы $f(e_{k+1}), \ldots, f(e_n)$ линейно независимы, откуда и будет следовать утверждение теоремы.

Пусть

$$\lambda_1 f(e_{k+1}) + \ldots + \lambda_{n-k} f(e_n) = 0.$$

Рассмотрим вектор

$$a = \lambda_1 e_{k+1} + \ldots + \lambda_{n-k} e_n.$$

Предыдущее равенство означает, что $f(a) = 0$, т. е.

$$a \in \operatorname{Ker} f = \langle e_1, \ldots, e_k\rangle.$$

Так как $e_1, \ldots, e_k, e_{k+1}, \ldots, e_n$ линейно независимы, то это возможно только при $\lambda_1 = \ldots = \lambda_{n-k} = 0$, что и требовалось доказать.

Следствие 1. *Если $f\colon K^n \to K^m$ — линейное отображение с матрицей A, то*

$$\dim \operatorname{Im} f = \operatorname{rk} A.$$

Доказательство получается сравнением теорем 2 и 3.

Следствие 2. *Ранг системы столбцов любой матрицы равен рангу системы ее строк.*

Доказательство. Пусть $f\colon K^n \to K^m$ — линейное отображение с матрицей A и $e_1, e_2, \ldots, e_n$ — единичные строки пространства K^n. Из (28) следует, что размерность пространства $\operatorname{Im} f$ равна рангу системы столбцов матрицы A. Сравнение этого с предыдущим следствием и дает желаемый результат.

Пример 9. Поле K можно рассматривать как (одномерное) векторное пространство над самим собой. Линейное отображение $f\colon V \to K$ называется *линейной функцией* на V. Если f — ненулевая линейная функция, то $\operatorname{Im} f = K$ и, если $\dim V = n$, теорема 3 дает, что

$$\dim \operatorname{Ker} f = n - 1.$$

Пример 10. Пусть X — множество ребер тетраэдра и Y — множество его граней. Каждой функции φ на X со значениями в поле K сопоставим функцию ψ на Y, определяемую следующим образом:

$$\psi(y) = \sum_{x \subset y} \varphi(x),$$

т. е. значение функции ψ на какой-либо грани равно сумме значений функции φ на сторонах этой грани. Этим определяется линейное отображение

$$f\colon F(X, K) \to F(Y, K)$$

(см. пример 1.7.2). Докажем, что если $\operatorname{char} K \neq 2$, то оно сюръективно. Для этого достаточно показать, что $\operatorname{Im} f$ содержит δ-функции всех граней. Функция φ, для которой $f(\varphi)$ есть δ-функция нижней грани, изображена на рис. 8а (значения φ на неотмеченных ребрах равны нулю). Так как

$$\dim F(X, K) = 6, \quad \dim F(Y, K) = 4,$$

то по теореме 3

$$\dim \operatorname{Ker} f = 6 - 4 = 2.$$

Функции, составляющие базис $\operatorname{Ker} f$, изображены на рис. 8б.

Задача 1. Для отображения f из предыдущего примера найти $\dim \operatorname{Ker} f$ в случае, когда $\operatorname{char} K = 2$.

Так как столбцы матрицы A — это строки транспонированной матрицы A^T (см. §1.9), то следствие 2 §2.3 означает, что

$$\operatorname{rk} A^T = \operatorname{rk} A.$$

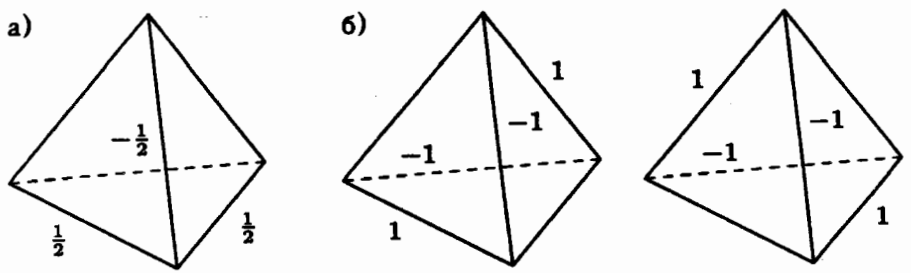

Рис. 8

Аналогично элементарным преобразованием строк матрицы определяются элементарные преобразования столбцов. Им соответствуют элементарные преобразования строк транспонированной матрицы. Поэтому ранг матрицы не изменяется не только при элементарных преобразованиях строк, но и при элементарных преобразованиях столбцов.

Замечание. Элементарные преобразования столбцов матрицы равносильны ее умножению на элементарные матрицы справа.

Обратимся теперь к операциям над линейными отображениями.

Линейные отображения $V \to U$ можно складывать и умножать на числа, как обычные функции:

$$(f + g)(a) = f(a) + g(a),$$
$$(\lambda f)(a) = \lambda f(a).$$

Относительно этих операций они образуют векторное пространство.

Далее, если

$$f\colon V \to U, \quad g\colon W \to V$$

— линейные отображения, то их произведение (композиция)

$$fg\colon W \to U$$

есть также линейное отображение. В самом деле,

$$(fg)(a + b) = f(g(a + b)) = f(g(a) + g(b)) =$$
$$= f(g(a)) + f(g(b)) = (fg)(a) + (fg)(b),$$
$$(fg)(\lambda a) = f(g(\lambda a)) = f(\lambda g(a)) = \lambda f(g(a)) = \lambda (fg)(a).$$

Умножение линейных отображений связано с линейными операциями свойствами

$$(g+h)=fg+fh, \quad (f+g)h=fh+gh,$$
$$(\lambda f)g=f(\lambda g)=\lambda(fg) \quad \forall \lambda \in K.$$

Докажем, например, первое из свойств дистрибутивности. Пусть

$$f\colon V \to U, \quad g\colon W \to V, \quad h\colon W \to V$$

— линейные отображения. Для любого $a \in W$ имеем:

$$(f(g+h))(a)=f((g+h)(a))=f(g(a)+h(a))=$$
$$=f(g(a))+f(h(a))=(fg)(a)+(fh)(a)=(fg+fh)(a).$$

Умножение линейных отображений ассоциативно, как и вообще умножение любых отображений. В самом деле, пусть M, N, P, Q — какие-то множества и

$$f\colon M \to N, \quad g\colon P \to N, \quad h\colon Q \to P$$

— какие-то отображения. Тогда для любого $a \in Q$ имеем:

$$((fg)h)(a)=(fg)(h(a))=f(g(h(a))),$$
$$(f(gh))(a)=f((gh)(a))=f(g(h(a))),$$

откуда

$$(fg)h=f(gh).$$

Операции над линейными отображениями пространств строк соответствуют таким же операциям над их матрицами. Для линейных операций это очевидно. Докажем это для умножения. Пусть

$$f\colon K^n \to K^m, \quad g\colon K^p \to K^n$$

— линейные отображения с матрицами $A=(a_{ij})$ и $B=(b_{jk})$ соответственно. Пусть $e_1, e_2, \ldots, e_p$ — единичные строки пространства K^p. Имеем тогда:

$$g(e_k)=(b_{1k}, b_{2k}, \ldots, b_{nk}),$$
$$(fg)(e_k)=f(g(e_k))=\left(\sum_j a_{1j}b_{jk}, \sum_j a_{2j}b_{jk}, \ldots, \sum_j a_{mj}b_{jk}\right).$$

Следовательно, матрица отображения fg есть $C=(c_{ik})$, где

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

Это означает, что $C=AB$, что и требовалось доказать.

Пример 11. Матричное равенство, доказанное в примере 1.9.2, на языке линейных отображений означает, что произведение поворотов плоскости на углы α и β есть поворот на угол $\alpha+\beta$ (см. пример 3). Поскольку последнее утверждение геометрически очевидно, это дает доказательство формул для косинуса и синуса суммы двух углов.

Свойства операций над матрицами, полученные нами в §1.9 прямыми вычислениями, могут быть теперь выведены из соответствующих свойств операций над линейными отображениями.

Очевидно, что тождественное отображение

$$\mathrm{id}\colon V \to V$$

линейно. Матрица тождественного отображения $\mathrm{id}\colon K^n \to K^n$ есть единичная матрица E порядка n. Поэтому свойства (10) единичной матрицы есть просто перевод на матричный язык очевидных равенств

$$f \cdot \mathrm{id}=f, \quad \mathrm{id} \cdot f=f,$$

где $f\colon K^n \to K^m$ — линейное отображение, задаваемое матрицей A, а id в первом случае обозначает тождественное отображение пространства K^n, а во втором — тождественное отображение пространства K^m.

Если $f\colon V \to U$ — биективное линейное отображение, то обратное отображение $f^{-1}\colon U \to V$ также линейно. В самом деле, для любых $a, b \in U$ пусть $c, d \in V$ — такие векторы, что $f(c)=a$, $f(d)=b$; тогда $f(c+d)=a+b$ и, следовательно,

$$f^{-1}(a+b)=c+d=f^{-1}(a)+f^{-1}(b).$$

Аналогично проверяется и второе свойство линейности.

Применим эти соображения к проблеме обратимости матриц.

Определение 3. Квадратная матрица A порядка n называется *невырожденной*, если $\mathrm{rk}\, A=n$.

Иными словами, матрица A невырожденна, если ее строки (или столбцы) линейно независимы.

Теорема 4. *Квадратная матрица обратима тогда и только тогда, когда она невырожденна.*

(Определение обратимого элемента кольца с единицей см. в §1.3.)

Доказательство. Пусть $f\colon K^n \to K^n$ — линейное отображение, задаваемое матрицей A. Согласно предыдущему матри-

ца A обратима тогда и только тогда, когда отображение f биективно. Последнее в силу теоремы 1 имеет место тогда и только тогда, когда

$$\operatorname{Im} f = K^n, \quad \operatorname{Ker} f = 0.$$

Ввиду теоремы 2 и следствия 1 теоремы 3 каждое из этих условий эквивалентно тому, что $\operatorname{rk} A = n$.

Нахождение матрицы, обратной к A, можно рассматривать как решение матричного уравнения

$$AX = E$$

(где X — неизвестная квадратная матрица). Такое уравнение можно решать, как и уравнение (16), с помощью умножения слева на элементарные матрицы, что равносильно элементарным преобразованиям строк «расширенной» матрицы $(A \mid E)$. Приведя левую половину этой матрицы к единичной матрице (что возможно в силу невырожденности матрицы A), в правой половине мы получим обратную матрицу.

Пример 12. Найдем матрицу, обратную к матрице

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

Для этого проделаем 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,

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

Задача 2. Используя линейные отображения, доказать, что ранг произведения двух матриц (не обязательно квадратных) не превосходит ранга каждой из них, а если одна из этих матриц невырожденна, то ранг произведения равен рангу другой матрицы.

§2.4. 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 невырожденна.

Поясним идею получения такого условия на примере геометрических векторов.

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 $\operatorname{area}(a_1, a_2)$ 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 $\operatorname{area}(a_1, a_2) = 0$. Величина $|\operatorname{area}(a_1, a_2)|$ может служить мерой линейной независимости векторов a_1 и a_2.

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

1) она линейна по a_1 и по a_2 (см. пример 2.3.9);

2) $\operatorname{area}(a_2, a_1) = -\operatorname{area}(a_1, a_2)$;

3) если $\{e_1, e_2\}$ — положительно ориентированный ортонормированный базис, то $\operatorname{area}(e_1, e_2) = 1$.

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

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

где $|a_1|$ — длина вектора a_1, а h_2 — проекция вектора a_2 на прямую, ортогональную a_1 (рис. 9). Так как проектирование есть линейное отображение, то отсюда следует линейность $\operatorname{area}(a_1, a_2)$ по a_2. Аналогично, взяв за основание a_2, можно доказать линейность по a_1.

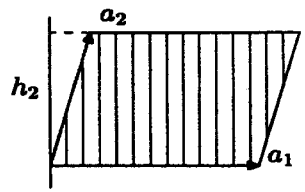

Рис. 9

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

$$a_1 = a_{11}e_1 + a_{12}e_2,$$
$$a_2 = a_{21}e_1 + a_{22}e_2.$$

Имеем тогда:

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

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 $\operatorname{vol}(a_1, a_2, a_3)$ 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 $\operatorname{vol}(e_1, e_2, e_3) = 1$.

(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 $\operatorname{vol}(a_1, a_2, a_3)$ через координаты векторов a_1, a_2, a_3 в положительно ориентированном ортонормированном базисе (проделайте это!):

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

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. 10 схематически изображено, какие из этих произведений берутся со знаком плюс и какие со знаком минус.

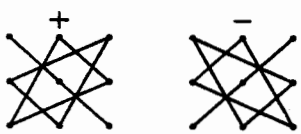

Рис. 10

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.

$$\begin{vmatrix}1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9\end{vmatrix} = 1\cdot5\cdot9 + 2\cdot6\cdot7 + 3\cdot4\cdot8 - 3\cdot5\cdot7 - 2\cdot4\cdot9 - 1\cdot6\cdot8 =$$

$$= 45 + 84 + 96 - 105 - 72 - 48 = 0.$$

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 полем K и $f(a_1, a_2, \ldots, a_m)$ — функция со значениями в K от m векторов пространства V.

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

Например, линейность по первому аргументу означает, что

$$f(a_1' + a_1'', a_2, \ldots, a_m) = f(a_1', a_2, \ldots, a_m) + f(a_1'', a_2, \ldots, a_m),$$
$$f(\lambda a_1, a_2, \ldots, a_m) = \lambda f(a_1, a_2, \ldots, a_m).$$

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

Важное свойство кососимметрической полилинейной функции состоит в том, что, если только $\operatorname{char} K \neq 2$, 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 $\operatorname{char} K = 2$, 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 счи-

тать 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

$$f(b,a) = -f(b,a).$$

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

Последовательность $(k_1, k_2, \ldots, k_n)$ чисел $1, 2, \ldots, n$, 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 $n-1$ значение, k_3 при заданных k_1 и k_2 может принимать $n-2$ значения и т. д., то имеется всего

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

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

Замечание 2. Слово «перестановка» в математической литературе (в частности, в этой книге) иногда употребляется в общечеловеческом 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* (соотв. *нечетной*), если число инверсий в ней четно (соотв. нечетно). Наряду с этим определяется *знак* перестановки, равный 1, если перестановка четна, и -1, если она нечетна. Знак перестановки $(k_1, k_2, \ldots, k_n)$ обозначается через $\operatorname{sgn}(k_1, k_2, \ldots, k_n)$.

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 — $(1,3,2)$ (одна инверсия), $(3,2,1)$ (три инверсии) и $(2,1,3)$ (одна инверсия).

Пример 4. Тривиальная перестановка не имеет инверсий и поэтому четна. Напротив, в перестановке $(n, n-1, \ldots, 2, 1)$ 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,

$$\operatorname{sgn}(n, n-1, \ldots, 2, 1) = (-1)^{n(n-1)/2} = (-1)^{[n/2]}.$$

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.

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.

Теперь мы в состоянии сформулировать и доказать основную теорему.

Теорема 1. *В пространстве K^n существует единственная кососимметрическая n-линейная функция* det, *удовлетворяющая условию*

$$\det(e_1, e_2, \ldots, e_n) = 1 \tag{29}$$

(*где $e_1, e_2, \ldots, e_n$ — единичные строки*). *Она имеет вид*

$$\det(a_1, a_2, \ldots, a_n) = \sum_{(k_1, k_2, \ldots, k_n)} \operatorname{sgn}(k_1, k_2, \ldots, k_n)\, a_{1k_1} a_{2k_2} \ldots a_{nk_n}, \tag{30}$$

где a_{ik} обозначает k-ю компоненту строки a_i, а суммирование происходит по всем перестановкам из n элементов.

Доказательство. 1) Предположим, что функция det удовлетворяет условиям теоремы. Тогда

$$\det(a_1, a_2, \ldots, a_n) = \det\Big(\sum_{k_1} a_{1k_1} e_{k_1}, \sum_{k_2} a_{2k_2} e_{k_2}, \ldots, \sum_{k_n} a_{nk_n} e_{k_n}\Big) = \\ = \sum_{k_1, k_2, \ldots, k_n} a_{1k_1} a_{2k_2} \ldots a_{nk_n} \det(e_{k_1}, e_{k_2}, \ldots, e_{k_n}).$$

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

$$\det(e_{k_1}, e_{k_2}, \ldots, e_{k_n}) = \operatorname{sgn}(k_1, k_2, \ldots, k_n).$$

В самом деле, если это равенство верно для какой-то перестановки $(j_1, j_2, \ldots, j_n)$, то оно верно и для любой перестановки, получаемой из нее транспозицией, так как при транспозиции обе части равенства умножаются на -1. По условию теоремы оно верно для тривиальной перестановки. Но очевидно, что любую перестановку можно получить из тривиальной последовательными транспозициями. Следовательно, доказываемое равенство верно для любой перестановки, и мы получаем для $\det(a_1, a_2, \ldots, a_n)$ выражение (30). Таким образом, если функция det, удовлетворяющая условиям теоремы, существует, то она имеет вид (30).

2) Докажем теперь, что функция det, определяемая равенством (30), удовлетворяет условиям теоремы. Линейность по каждому из аргументов очевидна, поскольку для любого i равенство (30) можно представить в виде

$$\det(a_1, a_2, \ldots, a_n) = \sum_j a_{ij} u_j,$$

где $u_1, \ldots, u_n$ не зависят от a_i. Условие (29) также выполнено, поскольку в выражении для $\det(e_1, e_2, \ldots, e_n)$ слагаемое, отвечающее тривиальной перестановке, равно 1, а все остальные слагаемые равны нулю. Остается проверить кососимметричность.

Посмотрим, что происходит при перестановке аргументов a_i и a_j. Мы можем разбить множество всех перестановок на пары перестановок, получаемых друг из друга транспозицией k_i и k_j. Согласно предложению 1 произведения $a_{1k_1} a_{2k_2} \ldots a_{nk_n}$, соответствующие перестановкам одной такой пары, входят в выражение

(30) с противоположными знаками. При перестановке a_i и a_j они меняются местами и, следовательно, все выражение умножается на -1.

Замечание 3. Если $\operatorname{char} K = 2$, то кососимметричность следует понимать в смысле замечания 1. Ее доказательство в этом случае состоит в том, что при $a_i = a_j$ члены выражения (30), соответствующие перестановкам каждой из описанных выше пар, взаимно уничтожаются.

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

$$\det A = \det(a_1, a_2, \ldots, a_n),$$

где $a_1, a_1, \ldots, a_n$ — строки матрицы A.

Таким образом,

$$\det A = \sum_{(k_1, k_2, \ldots, k_n)} \operatorname{sgn}(k_1, k_2, \ldots, k_n)\, a_{1k_1} a_{2k_2} \ldots a_{nk_n}. \qquad (31)$$

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

При $n \geqslant 4$ вычисление определителя непосредственно по формуле (31) 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 1-го типа.*

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

$$\det A' = \det(a_1 + ca_2, a_2, \ldots, a_n) = \\ = \det(a_1, a_2, \ldots, a_n) + c \det(a_2, a_2, \ldots, a_n) = \det A.$$

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 (31) 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

$$k_1 \geqslant 1,\ k_2 \geqslant 2,\ \dots,\ k_n \geqslant n;$$

но так как

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

то это возможно только при

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

Помимо того, что они дают практический способ вычисления определителей, предложения 1 и 2 позволяют нам ответить на вопрос, ради которого мы и ввели понятие определителя.

Теорема 2. *Квадратная матрица A невырожденна 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 2-го или 3-го типов, то определитель может измениться, но, во всяком случае, его равенство или неравенство нулю сохранится. Матрица A невырожденна тогда и только тогда, когда полученная ступенчатая матрица является строго треугольной; но это равносильно тому, что ее определитель отличен от нуля.

Продолжим изучение свойств определителей.

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.

Для того чтобы выяснить, с каким знаком входит в $\det A^T$ произведение $a_{1k_1}a_{2k_2}\dots a_{nk_n}$, нужно расположить его множители по порядку номеров столбцов. Этого можно достичь, последовательно меняя местами два множителя. 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 $a_{l_1 1}a_{l_2 2}\dots a_{l_n n}$, то

$$\operatorname{sgn}(k_1, k_2, \dots, k_n) = \operatorname{sgn}(l_1, l_2, \dots, l_n),$$

а это и означает, что рассматриваемое произведение входит в $\det A$ и $\det A^T$ с одним и тем же знаком.

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. Пусть матрица B имеет порядок m, матрица C — порядок n. Если $\det C = 0$, то элементарными преобразованиями строк матрицы C можно получить нулевую строку. Проделав такие же элементарные преобразования последних n строк матрицы A, мы получим нулевую строку во всей матрице A. Следовательно, в этом случае $\det A = 0$ и доказываемое равенство верно. Аналогичное рассуждение, но с элементарными преобразованиями первых m столбцов, показывает, что если $\det B = 0$, то $\det A = 0$ и доказываемое равенство верно.

Пусть теперь $\det B \neq 0$ и $\det C \neq 0$. Рассмотрим отношение

$$\frac{\det A}{\det B \cdot \det C}. \tag{32}$$

Нам нужно доказать, что оно равно 1. При любом элементарном преобразовании последних n строк матрицы A $\det A$ и $\det C$ умножаются на одно и то же число и, следовательно, отношение (32) не изменяется. Такими преобразованиями можно привести матрицу C к треугольному виду. Аналогично, отношение (32) не изменяется при любых элементарных преобразованиях первых m столбцов матрицы A. Такими преобразованиями можно привести матрицу B к треугольному виду (методом Гаусса, отраженным относительно побочной диагонали). Поэтому нам достаточно доказать, что отношение (32) равно 1 в случае, когда B и C — треугольные матрицы; но это очевидно.

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

Пример 5. Вычислим так называемый *определитель Вандермонда*

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

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

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

$$= (x_2 - x_1) \cdots (x_n - x_1)\, V(x_2, \ldots, x_n).$$

Продолжая так дальше, в конце концов получаем:

$$V(x_1, x_2, \ldots, x_n) = \prod_{i>j} (x_i - x_j). \qquad (33)$$

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

(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

$$\begin{vmatrix} a_{ij} & 0 & \ldots & 0 \\ a_{1j} & a_{11} & \ldots & a_{1n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{nj} & a_{n1} & \ldots & a_{nn} \end{vmatrix},$$

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.

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 (31) 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.

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, B*

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

Доказательство. Для любой матрицы U имеем

$$(UA)B = U(AB).$$

В частности, если U — элементарная матрица, это означает, что, когда мы делаем какое-нибудь элементарное преобразование строк матрицы A, в матрице AB происходит такое же преобразование.

Если $\det A = 0$, то с помощью элементарных преобразований строк мы можем получить в матрице A нулевую строку. Соответствующая строка произведения AB, как легко видеть, также станет нулевой. Следовательно, $\det AB = 0$, так что доказываемое равенство в этом случае верно.

Пусть теперь $\det A \neq 0$. Рассмотрим отношение

$$\frac{\det AB}{\det A}. \tag{34}$$

Нам нужно доказать, что оно равно $\det B$. Из предыдущего следует, что оно не изменяется при любых элементарных преобразованиях строк матрицы A. Так как с помощью таких преобразований матрицу A можно привести к единичной матрице E, то нам достаточно доказать, что отношение (34) равно $\det B$ в случае, когда $A = E$; но это очевидно.

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. 11)

$$\alpha_1 = \widehat{a_2 a_3},\ \alpha_2 = \widehat{a_3 a_1},\ \alpha_3 = \widehat{a_1 a_2}.$$

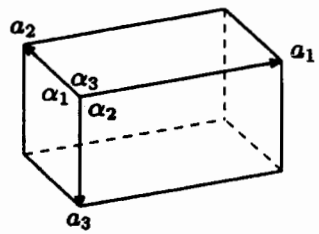

Рис. 11

Пусть $A = (a_{ij})$ — матрица, составленная из координат векторов a_i в ортонормированном базисе. Мы знаем (см. начало параграфа), что

$$V = \pm \det A.$$

Следовательно,

$$V^2 = (\det A)^2 = \det A \cdot \det A^T = \det AA^T.$$

Из правила умножения матриц следует, что (i,j)-й элемент матрицы AA^T есть скалярное произведение

$$(a_i, a_j) = |a_i||a_j| \cos \widehat{a_i a_j}.$$

Таким образом,

$$V^2 = \begin{vmatrix} |a_1|^2 & |a_1||a_2|\cos\alpha_3 & |a_1||a_3|\cos\alpha_2 \\ |a_2||a_1|\cos\alpha_3 & |a_2|^2 & |a_2||a_3|\cos\alpha_1 \\ |a_3||a_1|\cos\alpha_2 & |a_3||a_2|\cos\alpha_1 & |a_3|^2 \end{vmatrix} =$$

$$= |a_1|^2|a_2|^2|a_3|^2 \begin{vmatrix} 1 & \cos\alpha_3 & \cos\alpha_2 \\ \cos\alpha_3 & 1 & \cos\alpha_1 \\ \cos\alpha_2 & \cos\alpha_1 & 1 \end{vmatrix}$$

и, значит,

$$V = |a_1||a_2||a_3| \times$$
$$\times \sqrt{1 + 2\cos\alpha_1\cos\alpha_2\cos\alpha_3 - \cos^2\alpha_1 - \cos^2\alpha_2 - \cos^2\alpha_3}.$$

§2.5. Некоторые приложения определителей

Как мы видели в предыдущем параграфе (теорема 4.2), определители дают ответ на вопрос о невырожденности (и, тем самым, об обратимости) квадратной матрицы, который служил нам поводом для их введения. Вариации на эту тему приводят к многочисленным приложениям определителей в теории линейных уравнений и теории матриц. Первые из таких приложений будут рассмотрены в этом параграфе.

Рассмотрим квадратную систему линейных уравнений

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n = b_n. \end{cases} \tag{35}$$

Обозначим через A ее матрицу коэффициентов и через A_i $(i = 1,2,\ldots,n)$ матрицу, полученную из A заменой ее i-го столбца столбцом свободных членов.

Теорема 1. *Если* $\det A \neq 0$, *то система* (35) *имеет единственное решение, которое может быть найдено по формулам*

$$x_i = \frac{\det A_i}{\det A} \quad (i = 1,2,\ldots,n).$$

Эти формулы называются *формулами Крамера*.

Доказательство. При любом элементарном преобразовании системы (35) в матрицах A и A_i $(i = 1,2,\ldots,n)$ 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

$$\begin{cases} x_1 = b_1, \\ x_2 = b_2, \\ \ddots \\ x_n = b_n. \end{cases}$$

Она, очевидно, имеет единственное решение $x_i = b_i$ $(i = 1, 2,\ldots,n)$. С другой стороны,

$$\det A = \det E = 1,$$

$$\det A_i = \begin{vmatrix} 1 & 0 & \ldots & b_1 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & b_2 & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & b_i & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & b_{n-1} & \ldots & 1 & 0 \\ 0 & 0 & \ldots & b_n & \ldots & 0 & 1 \end{vmatrix} = b_i,$$

i

так что формулы Крамера в этом случае действительно верны.

Если $\det A = 0$, то ступенчатый вид матрицы A не будет строго треугольным и, следовательно, система (35) 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 (35) несовместна.

Задача 2. Показать, что если

$$\det A = \det A_1 = \ldots = \det A_n = 0,$$

то система (35) может быть как несовместна, так и неопределенна. (Привести примеры на оба случая.)

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. §2.4.)

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:

$$AX_j = E_j, \tag{36}$$

где E_j — j-й столбец матрицы E.

В координатной записи уравнение (36) 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

$$x_{ij} = \frac{1}{\det A}\begin{vmatrix} a_{11} & \dots & 0 & \dots & a_{1n} \\ \dots & \dots & \dots & \dots & \dots \\ a_{j1} & \dots & 1 & \dots & a_{jn} \\ \dots & \dots & \dots & \dots & \dots \\ a_{n1} & \dots & 0 & \dots & a_{nn} \end{vmatrix} = \frac{A_{ji}}{\det A},$$
$$\qquad\qquad i$$

что и требовалось доказать.

Пример. 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.

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 $\operatorname{rk} A = r$.

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.

Глава 3. Начала алгебры многочленов

§3.1. Построение и основные свойства алгебры многочленов

Функция действительной переменной x называется *многочленом*, если она может быть представлена в виде

$$f(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n,$$

где $a_0, a_1, a_2, \ldots, a_n$ — какие-то действительные числа (некоторые из которых 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

$$a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n = b_0 + b_1 x + b_2 x^2 + \ldots + b_n x^n \quad \forall x \in \mathbb{R},$$

то $a_k = b_k$ при $k = 0, 1, 2, \ldots, n$.

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 действительной переменной (см. пример 1.8.3). Эта подалгебра называется алгеброй многочленов над $\mathbb{R}$ и обозначается $\mathbb{R}[x]$.

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 2.2.6). Условимся нумеровать члены последовательностей, начиная с нуля, и пусть e_k ($k = 0, 1, 2, \ldots$) обозначает последовательность, k-й член которой равен 1, а все остальные члены равны 0. Последовательности $e_0, e_1, e_2, \ldots$ 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 ae_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

$$(a_0, a_1, a_2, \ldots, a_n, 0, \ldots) = a_0 e_0 + a_1 e_1 + a_2 e_2 + \ldots + a_n e_n = \\ = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n.$$

Числа $a_0, a_1, a_2, \ldots$ 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

$$\deg(f + g) \leqslant \max\{\deg f, \deg g\}, \tag{37}$$

$$\deg fg = \deg f + \deg g. \tag{38}$$

Докажем, например, последнее равенство. Пусть

$$f = a_0 + a_1 x + \ldots + a_n x^n \quad (a_n \neq 0),$$
$$g = b_0 + b_1 x + \ldots + b_m x^m \quad (b_m \neq 0).$$

Тогда при перемножении f и g получается только один член степени $n + m$, а именно, $a_n b_m x^{n+m}$, а членов большей степени не получается вообще. Так как в поле нет делителей нуля, то $a_n b_m \neq 0$ 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:

$$f = a_0x^n + a_1x^{n-1} + \ldots + a_{n-1}x + a_n.$$

Замечание 3. В качестве K можно взять любое коммутативное ассоциативное кольцо с единицей (ср. замечание 1.9.1). В этом случае все предыдущее остается без изменений, за исключением последней части, связанной с формулой (38), где нужно дополнительно потребовать, чтобы в кольце K не было делителей нуля.

Замечание 4. Произведение финитных последовательностей $(a_0, a_1, a_2, \ldots)$ и $(b_0, b_1, b_2, \ldots)$ в кольце $K[x]$ есть последовательность $(c_0, c_1, c_2, \ldots)$, 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

$$a_0 + a_1x + a_2x^2 + \ldots$$

Алгебра $K[[x]]$, как и $K[x]$, не имеет делителей нуля, но доказывается это по-другому (попробуйте это сделать!).

Каждый многочлен

$$f = a_0 + a_1x + a_2x^2 + \ldots + a_nx^n \tag{39}$$

определяет функцию на K со значениями в K, значение которой в точке $c \in K$ по определению равно

$$f(c) = a_0 + a_1c + a_2c^2 + \ldots + a_nc^n.$$

Так как сумма и произведение многочленов, а также произведение многочлена на число, приводятся к каноническому виду (39) 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

$$(f+g)(c) = f(c) + g(c),$$
$$(fg)(c) = f(c)g(c).$$
$$(\lambda f)(c) = \lambda f(c),$$

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

$$h = a_0 + a_1x + a_2x^2 + \ldots + a_{n-1}x^{n-1} \quad (a_{n-1} \neq 0).$$

Возьмем различные $x_1, x_2, \ldots, x_n \in K$ (здесь используется бесконечность поля K). Совокупность верных равенств

$$\begin{cases} a_0 + a_1x_1 + a_2x_1^2 + \ldots + a_{n-1}x_1^{n-1} = 0, \\ a_0 + a_1x_2 + a_2x_2^2 + \ldots + a_{n-1}x_2^{n-1} = 0, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_0 + a_1x_n + a_2x_n^2 + \ldots + a_{n-1}x_n^{n-1} = 0 \end{cases}$$

рассмотрим как (квадратную) систему однородных линейных уравнений относительно $a_0, a_1, a_2, \ldots, a_{n-1}$. Определитель матрицы коэффициентов этой системы есть определитель Вандермонда $V(x_1, x_2, \ldots, x_n)$ (см. пример 2.4.5) и потому отличен от нуля. Следовательно, система имеет только нулевое решение, что противоречит нашему предположению.

Замечание 5. Если поле K конечно, то множество всех многочленов над K тем не мение бесконечно (но счетно), в то время как множество всех функций на K со значениями в K конечно. Поэтому в этом случае обязательно должны существовать разные многочлены, определяющие одну и ту же функцию.

Задача. Так называемая *задача интерполяции* состоит в нахождении многочлена степени $< n$, принимающего в заданных (различных) точках $x_1, x_2, \ldots, x_n \in K$ заданные значения $y_1, y_2, \ldots, y_n \in K$. (В частности, при $n = 2$ это называется линейной интерполяцией.) Доказать, что задача интерполяции имеет единственное решение при любых $x_1, x_2, \ldots, x_n$ и $y_1, y_2 \ldots, y_n$.

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* 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 $\deg f \geqslant \deg g$, то q и r находятся обычной процедурой «деления уголком». А именно, пусть

$$f = a_0x^n + a_1x^{n-1} + \ldots + a_{n-1}x + a_n,$$
$$g = b_0x^m + b_1x^{m-1} + \ldots + b_{m-1}x + b_m,$$

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 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

$$q = c_0x^{n-m} + c_1x^{n-m-1} + \ldots + c_{n-m},$$

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

$$f = q_1g + r_1 = q_2g + r_2,$$

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

$$\deg(r_1 - r_2) = \deg(q_2 - q_1) + \deg g \geqslant \deg g,$$

что, очевидно, неверно. Следовательно, $q_1 = q_2$ и $r_1 = r_2$.

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

$$a_0x^n + a_1x^{n-1} + \ldots + a_{n-1}x + a_n =$$
$$= (x - c)(b_0x^{n-1} + b_1x^{n-2} + \ldots + b_{n-2}x + b_{n-1}) + r.$$

Приравнивая коэффициенты при соответствующих степенях x, по-

лучаем цепочку равенств

$$\begin{aligned} a_0 &= b_0, \\ a_1 &= b_1 - cb_0, \\ a_2 &= b_2 - cb_1, \\ &\dots\dots\dots\dots \\ a_{n-1} &= b_{n-1} - cb_{n-2}, \\ a_n &= r - cb_{n-1}, \end{aligned}$$

откуда находим следующие рекуррентные формулы для $b_0, b_1, \dots, b_{n-1}$ и r:

$$\begin{aligned} b_0 &= a_0, \\ b_1 &= a_1 + cb_0, \\ b_2 &= a_2 + cb_1, \\ &\dots\dots\dots\dots \\ b_{n-1} &= a_{n-1} + cb_{n-2}, \\ r &= a_n + cb_{n-1}. \end{aligned}$$

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. 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, $f(3) = 20$.

§3.2. 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

$$f = (x - c_1)(x - c_2)\dots(x - c_m)g, \tag{40}$$

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

$$m = \deg f - \deg g \leqslant \deg f.$$

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 — корень кратности k многочлена f тогда и только тогда, когда

$$f = (x-c)^k g, \tag{41}$$

где $g(c) \neq 0$.

Теперь мы докажем уточнение теоремы 2.

Теорема 3. *Число корней ненулевого многочлена с учетом их кратностей (т. е. считая каждый корень 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 (40), объединив одинаковые множители:

$$f = (x-c_1)^{k_1}(x-c_2)^{k_2}\dots(x-c_s)^{k_s} g \tag{42}$$

($c_1, c_2, \dots, c_s$ различны). Ясно, что $c_1, c_2, \dots, c_s$ — это все корни многочлена. Далее, выделяя в (42) 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.

Замечание 2. Условно считается, что многочлен нулевой степени разлагается в произведение пустого множества линейных множителей.

Если многочлен

$$f = a_0 x^n + a_1 x^{n-1} + \dots + a_{n-1}x + a_n$$

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*:

$$\begin{gathered}
c_1 + c_2 + \dots + c_n = -\frac{a_1}{a_0}, \\
c_1c_2 + c_1c_3 + \dots + c_{n-1}c_n = \frac{a_2}{a_0}, \\
\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\
\textstyle\sum_{i_1<i_2<\dots<i_k} c_{i_1}c_{i_2}\dots c_{i_k} = (-1)^k \dfrac{a_k}{a_0}, \\
\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\
c_1c_2\dots c_n = (-1)^n \frac{a_n}{a_0}.
\end{gathered}$$

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^k в произведении $(x-c_1)\times$ $\times(x-c_2)\dots(x-c_n)$.

Пример 1. Комплексные корни 5-й степени из 1

$$\epsilon_k = \cos\frac{2\pi k}{5} + i\sin\frac{2\pi k}{5} \quad (k = 0, 1, 2, 3, 4)$$

(рис. 12) суть корни многочлена $x^5 - 1$. По первой из формул Виета их сумма равна нулю. Приравнивая нулю сумму их действительных 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

$$\cos\frac{2\pi}{5} = \frac{\sqrt{5}-1}{4}, \quad \cos\frac{4\pi}{5} = -\frac{\sqrt{5}+1}{4}.$$

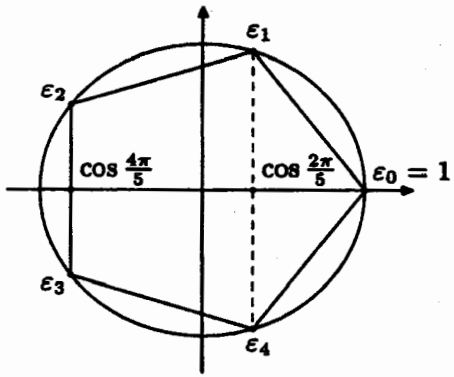

Рис. 12

Задача. Пусть n — простое число. Пользуясь задачей из §1.6 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 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

$$\begin{aligned} -a_1 &= 1+1+2+3 = 7,\\ a_2 &= 1\cdot1+1\cdot2+1\cdot3+1\cdot2+1\cdot3+2\cdot3 = 17,\\ -a_3 &= 1\cdot1\cdot2+1\cdot1\cdot3+1\cdot2\cdot3+1\cdot2\cdot3 = 17,\\ a_4 &= 1\cdot1\cdot1\cdot2\cdot3 = 6. \end{aligned}$$

Таким образом,

$$f = x^4 - 7x^3 + 17x^2 - 17x + 6.$$

Кратность корня многочлена может быть истолкована и другим способом, отличным от данного выше определения, по крайней мере в случае $\operatorname{char} K = 0$. Для этого нужно определить дифференцирование многочленов.

Из правил дифференцирования функций действительной переменной следует, что производная многочлена есть также многочлен. Обозначим через D отображение алгебры $\mathbb{R}[x]$ в себя, сопоставляющее 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. *Существует единственное отображение* $D\colon K[x] \to K[x]$, *обладающее свойствами* 1)–3).

Доказательство. Пусть D — такое отображение. Тогда

$$D1 = D(1\cdot1) = (D1)\cdot1 + 1\cdot(D1) = D1 + D1,$$

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

$$\begin{aligned} Dx^n = D(x^{n-1}x) = (Dx^{n-1})x + x^{n-1}(Dx) =\\ = (n-1)x^{n-2}\cdot x + x^{n-1} = nx^{n-1}. \end{aligned}$$

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 $D\colon K[x] \to K[x]$, задав его на базисных векторах по формулам

$$D1 = 0, \quad Dx^n = nx^{n-1} \quad (n = 1, 2, \dots),$$

и проверим, что оно обладает свойством 2). В силу линейности достаточно проверить это свойство для базисных векторов. Имеем:

$$D(x^mx^n) = Dx^{m+n} = (m+n)x^{m+n-1},$$
$$(Dx^m)x^n + x^m(Dx^n) = mx^{m-1}x^n + nx^mx^{n-1} = (m+n)x^{m+n-1}.$$

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 $x - c$:

$$f = b_0 + b_1(x-c) + b_2(x-c)^2 + \ldots + b_n(x-c)^n. \qquad (43)$$

Очевидно, что если c — корень многочлена f, то его кратность равна номеру первого отличного от нуля коэффициента этого разложения.

Предложение 2. *Если* $\operatorname{char} K = 0$, *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 (43) k раз и подставим $x = c$.

Таким образом,

$$f = f(c) + \frac{f'(c)}{1!}(x-c) + \frac{f''(c)}{2!}(x-c)^2 + \ldots + \frac{f^{(n)}(c)}{n!}(x-c)^n.$$

Эта формула называется *формулой Тейлора* для многочленов.

Из формулы Тейлора и сделанного выше замечания следует

Теорема 4. *При условии, что* $\operatorname{char} K = 0$, *кратность корня с 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 $\operatorname{char} K > 0$ кратность корня c может быть меньше указанного в теореме 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-го порядка. Кроме того, знак $f(x)$ при прохождении точки c при нечетном k меняется, а при четном k — не меняется (см. рис. 13).

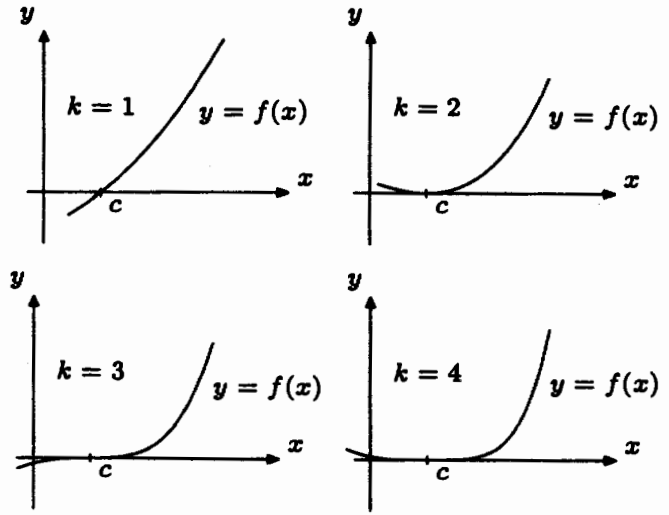

Рис. 13

Коэффициенты разложения (43), а, значит, и значения производных многочлена f в точке c (в случае $\operatorname{char} K = 0$), могут быть найдены последовательными делениями с остатком многочлена f на $x - c$. А именно, при первом делении получается остаток b_0 и неполное частное

$$f_1 = b_1 + b_2(x-c) + \ldots + b_n(x-c)^n;$$

при делении f_1 на $x - c$ получается остаток b_1 и т. д.

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,

$$f^{\mathrm{III}}(2) = 3! \cdot 7 = 42,$$
$$f^{\mathrm{IV}}(2) = 4! \cdot 5 = 120,$$
$$f^{\mathrm{V}}(2) = 5! \cdot 1 = 120.$$

§3.3. 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 $\mathbb{R}$ действительных 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.*

Существует несколько доказательств этой теоремы. Любое из них включает элементы анализа, так как оно должно как-то использовать определение поля действительных чисел, которое не является чисто алгебраическим. Доказательство, приводимое ниже, является почти полностью аналитическим.

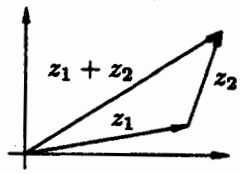

Рис. 14

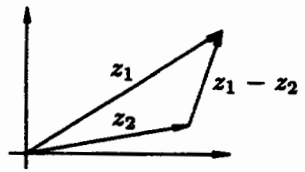

Рис. 15

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. 14 и 15), что

$$|z_1 + z_2| \leqslant |z_1| + |z_2|,$$
$$\big||z_1| - |z_2|\big| \leqslant |z_1 - z_2|.$$

(Равенства могут иметь место, когда соответствующий треугольник вырождается в отрезок.)

Определение. Последовательность комплексных чисел z_k ($k \in \mathbb{N}$) называется сходящейся к комплексному числу z (обозначение: $z_k \to z$), если $|z_k - z| \to 0$.

Лемма 1. *Пусть* $z_k = x_k + y_k i$, $z = x + yi$ $(x_k, y_k, x, y \in \mathbb{R})$. *Тогда*

$$z_k \to z \iff x_k \to x \text{ и } y_k \to y.$$

Доказательство. Имеем (см. рис. 16)

$$|z_k - z| = \sqrt{|x_k - x|^2 + |y_k - y|^2},$$

так что

$$x_k \to x \text{ и } y_k \to y \implies z_k \to z.$$

Обратная импликация вытекает из неравенств

$$|x_k - x| \leqslant |z_k - z|, \quad |y_k - y| \leqslant |z_k - z|.$$

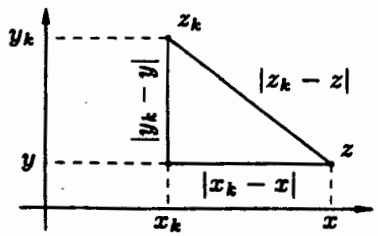

Рис. 16

Лемма 2. $z_k \to z$ *и* $w_k \to w \implies z_k + w_k \to z + w$ *и* $z_k w_k \to zw$.

Доказательство. Представим комплексные числа z_k, w_k, z, w в алгебраической форме:

$$z_k = x_k + y_k i, \quad w_k = u_k + v_k i, \quad z = x + yi, \quad w = u + vi.$$

Тогда

$$z_k w_k = (x_k u_k - y_k v_k) + (x_k v_k + y_k u_k)i,$$
$$zw = (xu - yv) + (xv + yu)i.$$

По лемме 1

$$x_k \to x, \quad y_k \to y, \quad u_k \to u, \quad v_k \to v$$

и нам нужно доказать, что

$$x_k u_k - y_k v_k \to xu - yv, \quad x_k v_k + y_k u_k \to xv + yu;$$

но это следует из свойств пределов последовательностей действительных чисел, известных из анализа. Аналогично доказывается, что $z_k + w_k \to z + w$.

Следствие. *Пусть* $z_k \to z$ *и* $f \in \mathbb{C}[z]$ *— любой многочлен. Тогда* $f(z_k) \to f(z)$.

(Здесь мы допускаем вольность в обозначениях, обычную в анализе, когда значение переменной обозначается так же, как сама переменная.)

Лемма 3. $z_k \to z \implies |z_k| \to |z|$.

Доказательство. следует из того, что

$$\big||z_k| - |z|\big| \leqslant |z_k - z|.$$

Лемма 4 (о возрастании модуля многочлена). *Если* $|z_k| \to \infty$ *и* $f \in \mathbb{C}[z]$ *— многочлен положительной степени, то* $|f(z_k)| \to \infty$.

Доказательство. Пусть

$$f = a_0 z^n + a_1 z^{n-1} + \ldots + a_{n-1} z + a_n \quad (a_0 \neq 0);$$

тогда

$$|f(z_k)| = |z_k|^n \left|a_0 + \frac{a_1}{z_k} + \ldots + \frac{a_{n-1}}{z_k^{n-1}} + \frac{a_n}{z_k^n}\right| \geqslant$$
$$\geqslant |z_k|^n \left(|a_0| - \frac{|a_1|}{|z_k|} - \ldots - \frac{|a_{n-1}|}{|z_k|^{n-1}} - \frac{|a_n|}{|z_k|^n}\right).$$

Выражение, стоящее в скобках, стремится к $|a_0|$. Следовательно, все произведение и, тем более, $|f(z_k)|$ стремится к бесконечности.

Следующая лемма является ключевой для доказательства основной теоремы.

Лемма 5 (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

$$\frac{f(z)}{f(z_0)} = 1 + c_p(z-z_0)^p + c_{p+1}(z-z_0)^{p+1} + \ldots + c_n(z-z_0)^n \quad (c_p \neq 0).$$

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. 17), где $t \in (0, 1)$, а z_1 — комплексное число, удовлетворяющее условию

$$c_p z_1^p = -1.$$

Имеем тогда:

$$\frac{f(z)}{f(z_0)} = 1 - t^p + t^{p+1}\varphi(t),$$

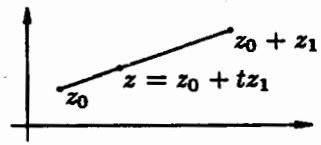

Рис. 17

где φ — некоторый многочлен степени $n-p-1$ (с комплексными коэффициентами). Если C — максимум модулей коэффициентов многочлена φ, то

$$|\varphi(t)| \leqslant A = (n-p)C$$

и, следовательно,

$$\left|\frac{f(z)}{f(z_0)}\right| \leqslant 1 - t^p + At^{p+1} = 1 - t^p(1-At) < 1$$

при $t < \frac{1}{A}$.

Доказательство теоремы. 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

$$|f(z_k)| \to M. \tag{44}$$

Если последовательность $|z_k|$ неограниченна, то из нее можно выбрать подпоследовательность, сходящуюся к бесконечности; но тогда в силу леммы 4 мы придем в противоречие с (44).

Таким образом, существует такое $C > 0$, что

$$|z_k| \leqslant C \quad \forall k.$$

Представим z_k в алгебраической форме:

$$z_k = x_k + y_k i.$$

Тогда

$$|x_k| \leqslant |z_k| \leqslant C, \quad |y_k| \leqslant |z_k| \leqslant C.$$

По теореме Больцано – Вейерштрасса из последовательности x_k можно выбрать сходящуюся подпоследовательность. Перейдя к этой подпоследовательности и изменив обозначения, можно считать, что

$$x_k \to x_0.$$

Аналогичным образом, перейдя еще раз к подпоследовательности, можно считать, что

$$y_k \to y_0.$$

Но тогда по лемме 1

$$z_k \to z_0 = x_0 + y_0 i$$

и, следовательно,

$$|f(z_k)| \to |f(z_0)| = M.$$

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 (40) 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).*

§3.4. Корни многочленов с действительными коэффициентами

Многочлен степени n с действительными коэффициентами может иметь $< n$ (в частности вообще не иметь) действительных корней, но, как и всякий многочлен с комплексными коэффициентами, он всегда имеет ровно n комплексных корней (с учетом кратностей). Мнимые корни многочлена с действительными коэффициентами обладают специальным свойством.

Теорема 1. *Если c — мнимый корень многочлена $f \in \mathbb{R}[x]$, то $\overline{c}$ также является корнем этого многочлена, причем той же кратности, что и c.*

Доказательство. Пусть

$$f = a_0x^n + a_1x^{n-1} + \ldots + a_{n-1}x + a_n \quad (a_0, a_1, \ldots, a_n \in \mathbb{R}).$$

Если $f(c) = 0$, то, поскольку комплексное сопряжение является автоморфизмом поля $\mathbb{C}$ (см. §1.5),

$$f(\overline{c}) = a_0\overline{c}^n + a_1\overline{c}^{n-1} + \ldots + a_{n-1}\overline{c} + a_n =$$
$$= \overline{a}_0\overline{c}^n + \overline{a}_1\overline{c}^{n-1} + \ldots + \overline{a}_{n-1}\overline{c} + \overline{a}_n = \overline{f(c)} = \overline{0} = 0,$$

т. е. $\overline{c}$ — также корень многочлена f. Аналогично доказывается, что

$$f^{(k)}(c) = 0 \iff f^{(k)}(\overline{c}) = 0.$$

Следовательно, кратности корней c и $\overline{c}$ одинаковы.

Следствие 1. *В алгебре $\mathbb{R}[x]$ всякий ненулевой многочлен разлагается на линейные множители и квадратные множители с отрицательным дискриминантом.*

Доказательство. Заметим, что если c — мнимое число, то в силу (5) квадратный трехчлен

$$(x-c)(x-\overline{c}) = x^2 - (c+\overline{c})x + c\overline{c}$$

имеет действительные коэффициенты; его дискриминант, очевидно, отрицателен.

Пусть теперь

$$c_1, \ldots, c_s, c_{s+1}, \ldots, c_{s+t}, \overline{c}_{s+1}, \ldots, \overline{c}_{s+t}$$

— это все (различные) комплексные корни многочлена $f \in \mathbb{R}[x]$, причем

$$c_1, \ldots, c_s \in \mathbb{R}, \quad c_{s+1}, \ldots, c_{s+t} \notin \mathbb{R}.$$

Если кратность корня c_i равна k_i, то

$$f = a_0(x-c_1)^{k_1} \ldots (x-c_s)^{k_s}[(x-c_{s+1})(x-\overline{c}_{s+1})]^{k_{s+1}} \ldots \\ \ldots [(x-c_{s+t})(x-\overline{c}_{s+t})]^{k_{s+t}},$$

(где a_0 — старший коэффициент многочлена f). Перемножая линейные множители в квадратных скобках, получаем искомое разложение.

Пример 1.

$$x^5 - 1 =$$
$$= (x-1)\left(x - \left(\cos\frac{2\pi}{5} + i\sin\frac{2\pi}{5}\right)\right)\left(x - \left(\cos\frac{2\pi}{5} - i\sin\frac{2\pi}{5}\right)\right) \times$$
$$\times \left(x - \left(\cos\frac{4\pi}{5} + i\sin\frac{4\pi}{5}\right)\right)\left(x - \left(\cos\frac{4\pi}{5} - i\sin\frac{4\pi}{5}\right)\right) =$$
$$= (x-1)\left(x^2 - 2x\cos\frac{2\pi}{5} + 1\right)\left(x^2 - 2x\cos\frac{4\pi}{5} + 1\right) =$$
$$= (x-1)\left(x^2 - \frac{\sqrt{5}-1}{2}x + 1\right)\left(x^2 + \frac{\sqrt{5}+1}{2}x + 1\right)$$

(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 действительный корень. Впрочем, это легко доказать и по-другому. А именно, если старший коэффициент многочлена f положителен, то

$$\lim_{x\to+\infty} f(x) = +\infty, \qquad \lim_{x\to-\infty} f(x) = -\infty$$

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 действительных 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 действительных 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 $x \leqslant 0$, а также при $x \geqslant 2$. Следовательно, все действительные корни многочлена f лежат в интервале (0,2). Однако, точное их число остается неясным, 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 действительных корней любого многочлена с действительными 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 точный ответ, но зато не требует никаких вычислений. Она говорит не просто о числе всех действительных 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 действительных чисел

$$a_0, a_1, a_2, \ldots, a_n.$$

Говорят, что на k-м месте в этой последовательности имеется *перемена знака*, если $a_k \neq 0$ и знак 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 (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 действительны, 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

$$f = a_0x^n + a_1x^{n-1} + \ldots + a_{n-1}x + a_n \quad (a_0 > 0,\ a_n \neq 0).$$

Тогда $f(0) = a_n$ и $f(x) > 0$ при достаточно больших x. Когда мы двигаемся вправо по числовой прямой, то при прохождении каждого корня $f(x)$ меняет свой 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 $L(f)$.

Лемма 2. $N(f) \leqslant N(f') + 1$.

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 $N(f') \geqslant N(f) - 1$.

Лемма 3. $L(f') \leqslant L(f)$.

Доказательство очевидно.

Число отрицательных корней многочлена f равно числу положительных корней многочлена

$$\overline{f}(x) = f(-x).$$

Лемма 4. $L(f) + L(\overline{f}) \leqslant n = \deg f$.

Доказательство. Коэффициенты многочлена $\overline{f}$ получаются из коэффициентов многочлена f попеременным умножением на ± 1. Предположим вначале, что все коэффициенты $a_0, a_1, \ldots, a_n$ многочлена f отличны от нуля. Если тогда на k-м месте в последовательности $a_0, a_1, \ldots, a_n$ имеется перемена знака, то на том же месте в последовательности коэффициентов многочлена $\overline{f}$ перемены знака нет, и наоборот. Поэтому в этом случае $L(f) + L(\overline{f}) = n$. В общем случае, когда среди коэффициентов $a_0, a_1, \ldots, a_n$ могут быть нули, при их замене произвольными числами, отличными от нуля, числа $L(f)$ и $L(\overline{f})$ могут только увеличиться. Так как после этого их сумма по доказанному станет равной n, то $L(f) + L(\overline{f}) \leqslant n$.

Доказательство теоремы. Докажем неравенство $N(f) \leqslant L(f)$ 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:

$$N(f) \leqslant N(f') + 1 \leqslant L(f') + 1 \leqslant L(f) + 1.$$

Однако равенство $N(f) = L(f) + 1$ невозможно ввиду леммы 1. Следовательно, $N(f) \leqslant L(f)$.

Пусть теперь известно, что все корни многочлена f действительны. Мы можем считать, что 0 не является корнем. Имеем тогда в силу уже доказанного неравенства и леммы 4:

$$n = N(f) + N(\overline{f}) \leqslant L(f) + L(\overline{f}) \leqslant n,$$

откуда

$$N(f) = L(f), \quad N(\overline{f}) = L(\overline{f}).$$

Пример 4. Для многочлена f из примера 3 имеем $L(f) = 2$, так что $N(f) \leqslant 2$. Но мы уже установили, что $N(f) \geqslant 2$. Следовательно, $N(f) = 2$.

Пример 5. Многочлен $f = x^2 - x + 1$ не имеет положительных (и вообще действительных) корней, но $L(f) = 2$, так что в этом случае $N(f) < L(f)$.

Применяя теорему Декарта к многочлену

$$g(x) = f(c + x) = f(c) + \frac{f'(c)}{1!}x + \frac{f'(c)}{2!}x^2 + \ldots + \frac{f^{(n)}(c)}{n!}x^n,$$

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 (тривиальной случай теоремы Декарта), а это означает, что все действительные корни многочлена f не превосходят c.

Пример 6. Найдем границы действительных корней многочлена

$$f = x^5 - 5x^3 - 10x^2 + 2.$$

Пользуясь схемой Горнера, вычислим $f(3)$:

	1	0	−5	−10	0	2
3	1	3	4	2	6	20

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 действительные корни меньше 3. Рассмотрим теперь многочлен

$$\overline{f}(x) = -f(-x) = x^5 - 5x^3 + 10x^2 - 2.$$

Вычислим значения многочлена $\overline{f}$ 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

$$\overline{f}(1) = 4 > 0, \quad \overline{f}'(1) = 10 > 0, \quad \overline{f}''(1) = 2 \cdot 5 > 0.$$

Значения следующих производных также положительны, поскольку последняя строка таблицы состоит только из положительных чисел. Следовательно, все действительные корни многочлена $\overline{f}$ меньше 1, а это означает, что все действительные корни многочлена f больше -1. Таким образом, все действительные корни многочлена f лежат в интервале $(-1, 3)$.

Задача. Исследовав производную многочлена $\overline{f}$, 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 алгебраическим 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. 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 си-

туациями, изображенными на рис. 18а, а в случае трехкратного — между ситуациями, изображенными на рис. 18б.

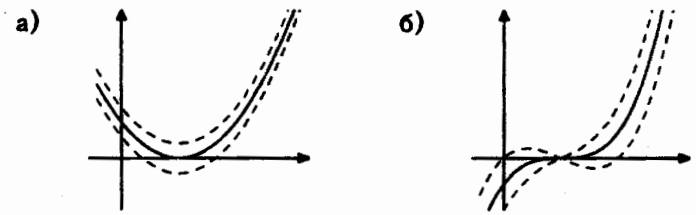

Рис. 18

§3.5. Теория делимости в евклидовых кольцах

Разложение многочленов над $\mathbb{C}$ на линейные множители и многочленов над $\mathbb{R}$ на линейные и квадратные 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).

Пусть A — целостное кольцо. Говорят, что элемент $a \in A$ *делится* на элемент $b \in A$ (обозначение: $a \vdots b$) или, иначе, что b *делит* a (обозначение: $b \mid a$), 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) $b \mid a$ и $a \mid b$;

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

$$N\colon A \setminus \{0\} \to \mathbb{Z}_+$$

(называемая *нормой*), удовлетворяющая следующим условиям:

1) $N(ab) \geqslant N(a)$, 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 1. Условие 2) означает возможность «деления с остатком». Его единственности (т. е. однозначной определенности пары (q, r)) не требуется.

Замечание 2. 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

$$N(a) \leqslant N(r) < N(ab).$$

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, во втором — степень многочлена.

Существуют и другие евклидовы кольца.

Пример. 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 $\left|\frac{c}{d} - q\right| \leqslant \frac{1}{\sqrt{2}}$ (см. рис. 19). Положим $r = c - qd$. Тогда $c = qd + r$ и

$$N(r) = |c - qd|^2 = \left|\frac{c}{d} - q\right|^2 |d|^2 \leqslant \frac{1}{2} N(d) < N(d).$$

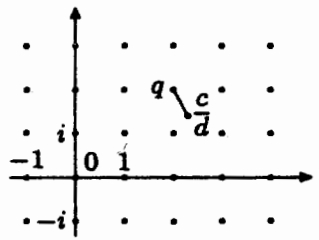

Рис. 19

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 НОД(a, b).

Наибольший общий делитель, если он существует, определен однозначно с точностью до ассоциированности. Однако он может не существовать. Например, элементы 2 и x в кольце $\mathbb{Z}[x]$ 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 с остатком a на b, затем b на полученный остаток, затем первый остаток на второй остаток и т. д. Так как нормы остатков убывают, то в конце концов деление произойдет без остатка. Получим

цепочку равенств:

$$\begin{aligned} a &= q_1 b + r_1, \\ b &= q_2 r_1 + r_2, \\ r_1 &= q_3 r_2 + r_3, \\ &\dots\dots\dots\dots \\ r_{n-2} &= q_n r_{n-1} + r_n, \\ r_{n-1} &= q_{n+1} r_n. \end{aligned}$$

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

$$r_n \mid r_{n-1}, \quad r_n \mid r_{n-2}, \quad \dots, \quad r_n \mid r_1, \quad r_n \mid b, \quad r_n \mid a.$$

Таким образом, r_n — общий делитель элементов a и b.

Двигаясь по той же цепочке равенств сверху вниз, получаем последовательно

$$\begin{aligned} r_1 &= au_1 + bv_1, \\ r_2 &= au_2 + bv_2, \\ r_3 &= au_3 + bv_3, \\ &\dots\dots\dots\dots \\ r_n &= au_n + bv_n, \end{aligned}$$

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 $r_n = au + bv$. Отсюда, в свою очередь, следует, что r_n делится на любой общий делитель элементов a и b.

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 следует, что неприводимые многочлены над $\mathbb{C}$ — это только многочлены первой степени, а из 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 $a_1a_2\dots a_n$, то p 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 $a_1a_2\dots a_n$ 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 $2, \dots, n$.

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 3. 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

$$a = p_1p_2\dots p_n = q_1q_2\dots q_m, \tag{45}$$

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 $p_1 \mid q_1q_2\dots q_m$ и по лемме 1 существует такой номер i, что $p_1 \mid q_i$. Тогда $p_1 \sim q_i$. Можно считать, что $i = 1$ и $p_1 = q_1$. Сокращая равенство (45) 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.

Следствие 1. *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*

$$d = cp_1^{l_1}\dots p_s^{l_s},$$

где $0 \leqslant l_i \leqslant k_i$ $(i = 1, \dots, s)$, 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.

Задача 1. Доказать, что в евклидовом кольце

а) $b \mid a$, $c \mid a$ и $(b, c) = 1 \implies bc \mid a$;

б) $c \mid ab$ и $(b, c) = 1 \implies c \mid a$.

Задача 2. *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 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 3. 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 $p_1, p_2, \ldots, p_n$ — это все простые числа. Тогда число $p_1, p_2, \ldots, p_n + 1$ 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 4. Перечислить неприводимые многочлены степеней $\leqslant 4$ над полем $\mathbb{Z}_2$ и доказать, что существует ровно 6 неприводимых многочленов степени 5.

§3.6. Многочлены с рациональными коэффициентами

Из однозначности разложения целого числа на простые множители вытекает

Теорема 1. *Если многочлен*

$$f = a_0x^n + a_1x^{n-1} + \ldots + a_{n-1}x + a_n \in \mathbb{Z}[x]$$

имеет рациональный корень $\frac{u}{v}$, *где* $u, v \in \mathbb{Z}$, $(u, v) = 1$, *то* $u \mid a_n$ *и* $v \mid a_0$.

Доказательство. Имеем:

$$0 = v^n f\left(\frac{u}{v}\right) = a_0u^n + a_1u^{n-1}v + \ldots + a_{n-1}uv^{n-1} + a_nv^n.$$

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 5.1б). Аналогично доказывается, что a_0 делится на v.

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

$$\pm\frac{1}{2},\ \pm1,\ \pm\frac{3}{2},\ \pm3.$$

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 λg, $\lambda^{-1}h \in \mathbb{Z}[x]$.

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 многочлен с целыми коэффициентами, а значит, и всякий 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

$$f = a_0x^n + a_1x^{n-1} + \ldots + a_{n-1}x + a_n \in \mathbb{Z}[x]$$

как многочлен

$$[f]_p = [a_0]_p x^n + [a_1]_p x^{n-1} + \ldots + [a_{n-1}]_p x + [a_n]_p \in \mathbb{Z}_p[x],$$

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. 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

$$vf = ug_1h_1$$

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 $\mathbb{Z}_p[x]$.

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»

$$f = x^{p-1} + x^{p-2} + \ldots + x + 1 = \frac{x^p - 1}{x - 1}.$$

(Комплексными корнями этого многочлена являются все нетривиальные корни p-й степени из 1, которые вместе с 1 делят окружность $|z| = 1$ на p равных частей.) Как следует из формулы бинома Ньютона (см. §1.6), 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 $m = \left[\frac{n}{2}\right]$.

Будем придавать x различные целые значения $x_0, x_1, \ldots, x_m$. Из равенств

$$f(x_i) = g(x_i)h(x_i)$$

следует, что $g(x_i) \mid f(x_i)$ при $i = 0, 1, \ldots, m$. Многочлен g однозначно определяется своими значениями в точках $x_0, x_1, \ldots, x_m$.

Выбирая всевозможные наборы делителей $d_0, d_1, \ldots, d_m$ целых чисел $f(x_0), f(x_1), \ldots, f(x_m)$ и находя для каждого из них интерполяционный многочлен степени $\leqslant m$, принимающий в точках $x_0, x_1, \ldots, x_m$ значения $d_0, d_1, \ldots, d_m$, можно найти всех кандидатов на роль g (их будет конечное число). Тех 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.

§3.7. Многочлены от нескольких переменных

Функция действительных переменных $x_1, x_2, \ldots, x_n$ называется *многочленом*, если она может быть представлена в виде

$$f(x_1, x_2, \ldots, x_n) = \sum_{k_1, k_2, \ldots, k_n} a_{k_1 k_2 \ldots k_n} x_1^{k_1} x_2^{k_2} \ldots x_n^{k_n}, \qquad (46)$$

где суммирование происходит по конечному числу наборов $(k_1, k_2, \ldots, k_n)$ неотрицательных целых чисел. (Формально можно считать, что суммирование происходит по всем таким наборам, но лишь конечное число коэффициентов $a_{k_1 k_2 \ldots k_n}$ отлично от нуля.) Многочлены образуют подалгебру в алгебре всех функций от $x_1, x_2, \ldots, x_n$. Она называется алгеброй многочленов от $x_1, x_2, \ldots, x_n$ над $\mathbb{R}$ и обозначается $\mathbb{R}[x_1, x_2, \ldots, x_n]$.

Можно показать (см. теорему 6.1), что представление многочлена над $\mathbb{R}$ в виде (46) 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

$$\{ e_{k_1 k_2 \ldots k_n} : k_1, k_2, \ldots, k_n \in \mathbb{Z}_+ \}$$

и таблицей умножения

$$e_{k_1 k_2 \ldots k_n} e_{l_1 l_2 \ldots l_n} = e_{k_1 + l_1, k_2 + l_2, \ldots, k_n + l_n}.$$

Очевидно, что эта алгебра коммутативна и ассоциативна и что элемент $e_{00\ldots0}$ является ее единицей. Она называется *алгеброй многочленов* над K и обозначается через $K[x_1, x_2, \ldots, x_n]$.

Условимся отождествлять элементы вида $ae_{00\ldots0}$ $(a \in K)$ с соответствующими элементами поля K и введем обозначения

$$\begin{aligned} e_{10\ldots0} &= x_1, \\ e_{01\ldots0} &= x_2, \\ &\ldots\ldots\ldots \\ e_{00\ldots1} &= x_n. \end{aligned}$$

Тогда

$$e_{k_1 k_2 \ldots k_n} = x_1^{k_1} x_2^{k_2} \ldots x_n^{k_n}$$

и любой элемент

$$\sum_{k_1, k_2, \ldots, k_n} a_{k_1 k_2 \ldots k_n} e_{k_1 k_2 \ldots k_n} \in K[x_1, x_2, \ldots, x_n]$$

записывается в обычном виде (46).

Многочлен (46) называется *однородным* степени d, если

$$a_{k_1 k_2 \ldots k_n} = 0 \text{ при } k_1 + k_2 + \ldots + k_n \neq d.$$

Однородные многочлены заданной степени d образуют конечномерное подпространство, так как имеется лишь конечное число наборов $(k_1, k_2, \ldots, k_n)$ целых неотрицательных чисел, удовлетворяющих условию

$$k_1 + k_2 + \ldots + k_n = d.$$

Задача. Доказать, что размерность пространства однородных многочленов степени d от n переменных равна

$$C_n^d = \frac{n(n+1)\ldots(n+d-1)}{d!}$$

(*число сочетаний с повторениями из n по d*).

Любой многочлен однозначно представляется в виде суммы однородных многочленов степеней $0, 1, 2, \ldots$, 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, максимальная из степеней его ненулевых однородных компонент. Степень многочлена f обозначается через $\deg f$. Справедливы следующие соотношения:

$$\deg(f+g) \leqslant \deg f + \deg g,$$
$$\deg(fg) = \deg f + \deg g.$$

Первое из них очевидно, второе мы докажем чуть позже.

С другой стороны, каждый многочлен $f \in K[x_1, x_2, \ldots, x_n]$ однозначно представляется в виде

$$f(x_1, x_2, \ldots, x_n) = \sum_{k=0}^{\infty} f_k(x_2, \ldots, x_n) x_1^k, \qquad (47)$$

где $f_0, f_1, f_2, \ldots$ — какие-то многочлены от $x_2, \ldots, x_n$, 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 (47), можно рассматривать кольцо $K[x_1, x_2, \ldots, x_n]$ как кольцо многочленов от x_1 с коэффициентами из $K[x_2, \ldots, x_n]$:

$$K[x_1, x_2, \ldots, x_n] = K[x_2, \ldots, x_n][x_1]. \qquad (48)$$

Замечание. Мы говорим о кольцах, а не об алгебрах, так как $K[x_1, x_2, \ldots, x_n]$ по определению есть алгебра над K, в то время как $K[x_2, \ldots, x_n][x_1]$ есть алгебра над $K[x_2, \ldots, x_n]$. Однако если рассматривать $K[x_2, \ldots, x_n][x_1]$ как алгебру над K (пользуясь тем, что $K[x_2, \ldots, x_n] \supset K$), то можно говорить о равенстве алгебр.

Предложение 1. *Алгебра $K[x_1, x_2, \ldots, x_n]$ не имеет делителей нуля.*

Доказательство. В §3.1 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 (48) позволяет доказать наше утверждение индуктивным путем, начиная с поля K.

Теперь мы в состоянии доказать соотношение (48). Разложим многочлены f и g на однородные компоненты:

$$f = f_0 + f_1 + \ldots + f_d \quad (\deg f_k = k,\ f_d \neq 0),$$
$$g = g_0 + g_1 + \ldots + g_e \quad (\deg g_k = k,\ g_e \neq 0).$$

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 2.1. Предположим теперь, что многочлен $f \in K[x_1, x_2, \ldots, x_n]$ ($n > 1$) определяет нулевую функцию. Представим его в виде (47) и придадим какие-то значения переменным $x_2, \ldots, x_n$. 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 $f_k \in K[x_2, \ldots, x_n]$ обращается в нуль при любых значениях $x_2, \ldots, x_n$, т. е. определяет нулевую функцию. По предположению индукции отсюда следует, что $f_k = 0$ при любом k; но тогда и $f = 0$.

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 упорядочение* (т. е. подобное упорядочению слов в словаре), 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 большим 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

$$k \geqslant l \geqslant m,$$

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:

$$u_1u_2 \succ v_1u_2 \succ v_1v_2.$$

Пример. Следующий многочлен расположен по лексикографическому убыванию членов:

$$x_1^2x_2 + x_1x_2^2x_3 + 2x_1x_3^2 + x_2x_3^3 - x_2x_3^2 + 3.$$

Обратите внимание на то, что член $x_1x_2^2x_3$ лексикографически младше $x_1^2x_2$, хотя его степень больше.

Среди ненулевых членов любого ненулевого многочлена $f \in K[x_1, x_2, \ldots, x_n]$ 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, 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

$$u_1u_2 \succ v_1v_2.$$

Следовательно, после приведения подобных членов в произведении f_1f_2 произведение u_1u_2 сохранится в качестве ненулевого члена, который старше всех остальных.

§3.8. Симметрические многочлены

Определение. Многочлен $f \in K[x_1, x_2, \ldots, x_n]$ называется *симметрическим*, если он не изменяется при любых 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*

$$s_k = x_1^k + x_2^k + \ldots + x_n^k \quad (k = 1, 2, \ldots),$$

очевидно, являются симметрическими многочленами.

Пример 2. Следующие симметрические многочлены называются *элементарными симметрическими многочленами*:

$$\begin{aligned}
&\sigma_1 = x_1 + x_2 + \ldots + x_n,\\
&\sigma_2 = x_1x_2 + x_1x_3 + \ldots + x_{n-1}x_n,\\
&\ldots\ldots\ldots\ldots\ldots\ldots\\
&\sigma_k = \sum_{i_1<i_2<\ldots<i_k} x_{i_1}x_{i_2}\ldots x_{i_k},\\
&\ldots\ldots\ldots\ldots\ldots\ldots\\
&\sigma_n = x_1x_2\ldots x_n.
\end{aligned}$$

Пример 3. Определитель Вандермонда

$$V(x_1, x_2, \ldots, x_n) = \prod_{i>j} (x_i - x_j)$$

(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 случаев равно числу инверсий в соответствующей перестановке. Следовательно,

$$V(x_{k_1}, x_{k_2}, \ldots, x_{k_n}) = \operatorname{sgn}(k_1, k_2, \ldots, k_n) V(x_1, x_2, \ldots, x_n).$$

Таким образом, сам определитель Вандермонда не является симме-

трическим многочленом, но таковым является его квадрат

$$V(x_1, x_2, \ldots, x_n)^2 = \prod_{i>j} (x_i - x_j)^2.$$

Пример 4. При любых перестановках переменных x_1, x_2, x_3, x_4 многочлены

$$h_1 = x_1x_2 + x_3x_4,\ h_2 = x_1x_3 + x_2x_4,\ h_3 = x_1x_4 + x_2x_3$$

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 к алгебраическим уравнениям с одним неизвестным благодаря формулам Виета (см. §3.2), 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 $s_2 = x_1^2 + x_2^2 + \ldots + x_n^2$ является симметрическим. Легко видеть, что

$$s_2 = \sigma_1^2 - 2\sigma_2. \tag{49}$$

Поэтому сумма квадратов корней алгебраического уравнения

$$x^n + a_1x^{n-1} + a_2x^{n-2} + \ldots + a_{n-1}x + a_n = 0$$

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 $F \in K[X_1, X_2, \ldots, X_m]$ — произвольный многочлен от m переменных и $f_1, f_2, \ldots, f_m \in K[x_1, x_2, \ldots, x_n]$ — какие-то симметрические многочлены, то $F(f_1, f_2, \ldots, f_m)$ — также симметрический многочлен от $x_1, x_2, \ldots, x_n$. Естественно поставить вопрос, нельзя ли найти такие симметрические многочлены $f_1, f_2, \ldots, f_m$, чтобы всякий симметрический многочлен можно было выразить через них указанным способом. Оказывается, что в качестве таких многочленов можно взять элементарные симметрические многочлены $\sigma_1, \sigma_2, \ldots, \sigma_n$.

Теорема 1. *Всякий симметрический многочлен единственным образом представляется в виде многочлена от элементарных симметрических многочленов.*

Доказательству теоремы предпошлем две леммы.

Лемма 1. *Пусть $u = ax_1^{k_1}x_2^{k_2}\ldots x_n^{k_n}$ — старший член симметрического многочлена f. Тогда*

$$k_1 \geqslant k_2 \geqslant \ldots \geqslant k_n. \tag{50}$$

Доказательство. Предположим, что $k_i < k_{i+1}$ для некоторого i. Наряду с членом u многочлен f должен содержать член

$$u' = ax_1^{k_1}\ldots x_i^{k_{i+1}}x_{i+1}^{k_i}\ldots x_n^{k_n},$$

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.

Лемма 2. *Для любого одночлена $u = x_1^{k_1}x_2^{k_2}\ldots x_n^{k_n}$, показатели которого удовлетворяют неравенствам* (50), *существуют такие неотрицательные целые числа $l_1, l_2, \ldots, l_n$, что старший член многочлена $\sigma_1^{l_1}\sigma_2^{l_2}\ldots\sigma_n^{l_n}$ совпадает с u. Числа $l_1, l_2, \ldots, l_n$ определены этим условием однозначно.*

Доказательство. Старший член многочлена σ_k равен $x_1x_2\ldots x_k$. В силу предложения 3 старший член многочлена $\sigma_1^{l_1}\sigma_2^{l_2}\ldots\sigma_n^{l_n}$ равен

$$x_1^{l_1}(x_1x_2)^{l_2}\ldots(x_1x_2\ldots x_n)^{l_n} = x_1^{l_1+l_2+\ldots+l_n}x_2^{l_2+\ldots+l_n}\ldots x_n^{l_n}.$$

Приравнивая его одночлену u, получаем систему линейных уравнений

$$\begin{cases} l_1 + l_2 + \ldots + l_n = k_1, \\ l_2 + \ldots + l_n = k_2, \\ \ldots\ldots\ldots \\ l_n = k_n, \end{cases}$$

которая, очевидно, имеет единственное решение

$$l_i = k_i - k_{i+1} \ (i = 1, 2, \ldots, n-1), \quad l_n = k_n. \tag{51}$$

Из условия леммы следует, что определенные таким образом числа $l_1, l_2, \ldots, l_n$ неотрицательны.

Замечание 1. Уравнение $l_1 + l_2 + \ldots + l_n = k_1$ показывает, что степень одночлена $X_1^{l_1} X_2^{l_2} \ldots X_n^{l_n}$ по совокупности переменных равна степени одночлена u по x_1.

Доказательство теоремы. Пусть $f \in K[x_1, x_2, \ldots, x_n]$ — симметрический многочлен. Нам нужно найти такой многочлен $F \in K[X_1, X_2, \ldots, X_n]$, что

$$F(\sigma_1, \sigma_2, \ldots, \sigma_n) = f.$$

Если $f = 0$, то можно взять $F = 0$. В противном случае пусть $u_1 = a x_1^{k_1} x_2^{k_2} \ldots x_n^{k_n}$ — старший член многочлена f. По лемме 1 выполняются неравенства (50). По лемме 2 существует такой одночлен $F_1 \in K[X_1, X_2, \ldots, X_n]$, что старший член многочлена $F_1(\sigma_1, \sigma_2, \ldots, \sigma_n)$ равен u_1. Рассмотрим симметрический многочлен

$$f_1 = f - F_1(\sigma_1, \sigma_2, \ldots, \sigma_n).$$

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 $F_2 \in K[X_1, X_2, \ldots, X_n]$, что старший член многочлена $F_2(\sigma_1, \sigma_2, \ldots, \sigma_n)$ равен u_2. Рассмотрим симметрический многочлен

$$f_2 = f_1 - F_2(\sigma_1, \sigma_2, \ldots, \sigma_n).$$

Если $f_2 = 0$, то можно взять $F = F_1 + F_2$. В противном случае, продолжая процесс, получаем последовательность симметрических многочленов $f, f_1, f_2, \ldots$, старшие члены которых удовлетворяют неравенствам

$$u_1 \succ u_2 \succ \ldots$$

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 $F_1 + F_2 + \ldots + F_M$.

Докажем теперь, что многочлен F определен однозначно. Предположим, что F и G — такие многочлены, что

$$F(\sigma_1, \sigma_2, \ldots, \sigma_n) = G(\sigma_1, \sigma_2, \ldots, \sigma_n).$$

Рассмотрим их разность $H = F - G$. Тогда

$$H(\sigma_1, \sigma_2, \ldots, \sigma_n) = 0.$$

Нам нужно доказать, что $H = 0$. Предположим, что это не так, и пусть $H_1, H_2, \ldots, H_s$ — все ненулевые члены многочлена H. Обозначим через w_i $(i = 1, 2, \ldots, s)$ старший член многочлена

$$H_i(\sigma_1, \sigma_2, \ldots, \sigma_n) \in K[x_1, x_2, \ldots, x_n].$$

В силу леммы 2 среди одночленов $w_1, w_2, \ldots, w_s$ нет пропорциональных. Выберем из них старший. Пусть это будет w_1. По построению одночлен w_1 старше всех остальных членов многочлена $H_1(\sigma_1, \sigma_2, \ldots, \sigma_n)$ и всех членов многочленов $H_i(\sigma_1, \sigma_2, \ldots, \sigma_n)$ $(i = 2, \ldots, s)$. Поэтому после приведения подобных членов в сумме

$$H_1(\sigma_1, \sigma_2, \ldots, \sigma_n) + H_2(\sigma_1, \sigma_2, \ldots, \sigma_n) + \ldots + H_s(\sigma_1, \sigma_2, \ldots, \sigma_n) = H(\sigma_1, \sigma_2, \ldots, \sigma_n)$$

член w_1 сохранится, так что эта сумма не будет равна нулю, что противоречит нашему предположению.

Замечание 2. Согласно замечанию 1 для любого m

$$\deg F_m = \deg_{x_1} u_m \leqslant \deg_{x_1} u_1 = \deg_{x_1} f (= k_1).$$

Следовательно,

$$\deg F = \deg_{x_1} f. \tag{52}$$

Следуя доказательству этой теоремы, можно в принципе найти выражение любого конкретного симметрического многочлена через $\sigma_1, \sigma_2, \ldots, \sigma_n$.

Пример 6. Выразим через $\sigma_1, \sigma_2, \ldots, \sigma_n$ многочлен

$$f = s_3 = x_1^3 + x_2^3 + \ldots + x_n^3.$$

Представим вычисления в виде таблицы.

m	u_m	$F_m(\sigma_1, \sigma_2, \ldots, \sigma_n)$	f_m
1	x_1^3	$\sigma_1^3 = \sum x_i^3 + 3\sum_{i\neq j} x_i^2 x_j + 6\sum_{i<j<k} x_i x_j x_k$	$-3\sum_{i\neq j} x_i^2 x_j - 6\sum_{i<j<k} x_i x_j x_k$
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,

$$s_3 = \sigma_1^3 - 3\sigma_1\sigma_2 + 3\sigma_3 \qquad (53)$$

На практике для однородных симметрических многочленов удобнее применять другой способ, который мы поясним на следующем ниже примере.

Пример 7. Выразим через $\sigma_1, \sigma_2, \sigma_3, \sigma_4$ многочлен

$$f = (x_1x_2 + x_3x_4)(x_1x_3 + x_2x_4)(x_1x_4 + x_2x_3)$$

из примера 4. В обозначениях доказательства теоремы, 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 $u_2, u_3, \ldots$ 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 u_1. Справа выпишем соответствующие произведения элементарных симметрических многочленов, найденные по

формулам (51).

$$\begin{array}{cccc|l} 3 & 1 & 1 & 1 & \sigma_1^2\sigma_4 \\ 2 & 2 & 2 & 0 & \sigma_3^2 \\ 2 & 2 & 1 & 1 & \sigma_2\sigma_4 \end{array}$$

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 $K = \mathbf{Z}$ 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 $f \in K[x_1, x_2, \ldots, x_n]$ — симметрический многочлен и $F \in K[X_1, X_2, \ldots, X_n]$ — такой многочлен, что

$$f = F(\sigma_1, \sigma_2, \ldots, \sigma_n).$$

Пусть далее, $c_1, c_2, \ldots, c_n$ — корни алгебраического уравнения

$$a_0x^n + a_1x^{n-1} + \ldots + a_{n-1}x + a_n = 0 \quad (a_0 \neq 0).$$

Тогда

$$f(c_1, c_2, \ldots, c_n) = F\left(-\frac{a_1}{a_0}, \frac{a_2}{a_0}, \ldots, (-1)^n \frac{a_n}{a_0}\right). \qquad (54)$$

Замечание 4. Пусть $\deg_{x_i} f = k$. Тогда $\deg F = k$ (см. замечание 2) и, домножив равенство (54) на a_0^k, мы получим в правой части однородный многочлен степени k от $a_0, a_1, a_2, \ldots, a_n$.

Пример 8. Пусть c_1, c_2, c_3, c_4 — корни уравнения

$$x^4 + px^2 + qx + r = 0. \qquad (55)$$

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

$$\begin{aligned} a_1 &= -(d_1 + d_2 + d_3), \\ a_2 &= d_1d_2 + d_1d_3 + d_2d_3, \\ a_3 &= -d_1d_2d_3. \end{aligned}$$

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:

$$h_1 + h_2 + h_3 = \sigma_2,$$

$$h_1h_2 + h_1h_3 + h_2h_3 = \sum_{\substack{i\neq j,k \\ j<k}} x_i^2x_jx_k = \sigma_1\sigma_3 - 4\sigma_4,$$

$$h_1h_2h_3 = \sigma_1^2\sigma_4 + \sigma_3^2 - 4\sigma_2\sigma_4.$$

(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,

$$a_1 = -p, \quad a_2 = -4r, \quad a_3 = q^2 - 4pr,$$

т. е. искомое уравнение имеет вид

$$y^3 - py^2 - 4ry + (4pr - q^2) = 0. \qquad (56)$$

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,

$$(c_1 + c_2 - c_3 - c_4)(c_1 - c_2 + c_3 - c_4)(c_1 - c_2 - c_3 + c_4) = -8q \qquad (57)$$

(см. задачу 1).

Пользуясь результатами этой задачи, можно свести решение уравнения (55) к решению уравнения (56) (при условии, что $\operatorname{char} K \neq 2$). 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 (57)).

Уравнение (56) называется *кубической резольвентой* уравнения (55).

§3.9. 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 действительных чисел) — о числе действительных 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

$$D(\varphi) = a_1^2 - 4a_0a_2 = a_0^2\left[\left(\frac{a_1}{a_0}\right)^2 - \frac{4a_2}{a_0}\right] = \\ = a_0^2[(c_1+c_2)^2 - 4c_1c_2] = a_0^2(c_1-c_2)^2.$$

В случае, когда $a_0, a_1, a_2 \in \mathbb{R}$, полученная формула хорошо объясняет ту связь между дискриминантом и свойствами корней, о которой 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 действительное число и $D(\varphi) > 0$;

2) $c_1 = c_2 \in \mathbb{R}$; тогда $c_1 - c_2 = 0$ и $D(\varphi) = 0$;

3) $c_1 = \overline{c}_2 \notin \mathbb{R}$; 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

$$\varphi = a_0x^n + a_1x^{n-1} + \ldots + a_{n-1}x + a_n \in K[x] \quad (a_0 \neq 0).$$

Предположим вначале, что многочлен φ имеет n корней $c_1, c_2, \ldots, c_n \in K$. Определим тогда его *дискриминант* $D(\varphi)$ по формуле

$$D(\varphi) = a_0^{2n-2}\prod_{i>j}(c_i - c_j)^2. \tag{58}$$

(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 φ. Описанная в §3.8 процедура позволяет выразить $D(\varphi)$ через коэффициенты 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 $a_0, a_1, \ldots, a_n$:

$$D(\varphi) = \Delta(a_0, a_1, \ldots, a_n). \tag{59}$$

Для нахождения многочлена Δ нет необходимости знать, что многочлен φ имеет n корней в K. Это позволяет определить дискриминант любого многочлена φ по формуле (59).

Замечание 1. Так как f имеет целые коэффициенты, то и Δ имеет целые коэффициенты (см. замечание 8.3).

Замечание 2. Как можно доказать, 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 (58), если в качестве $c_1, c_2, \ldots, c_n$ взять корни многочлена φ в поле L.

Из определения (58) 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 выбрать коэффициенты многочлена наудачу, то вероятность того, что он будет иметь кратные корни, равна нулю.

Пусть теперь φ — кубический многочлен с действительными 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 действительные числа; тогда $D(\varphi) > 0$;

2) $c_1, c_2, c_3 \in \mathbb{R}$, $c_2 = c_3$; тогда $D(\varphi) = 0$;

3) $c_1 \in \mathbb{R}$, $c_2 = \overline{c_3} \notin \mathbb{R}$; тогда

$$D(\varphi) = a_0^4[(c_1-c_2)(c_1-\overline{c_2})]^2(c_2-\overline{c_2})^2 = \\ = a_0^4|c_1-c_2|^4(c_2-\overline{c_2})^2 < 0.$$

Таким образом, мы приходим к тому же выводу, что и в случае квадратного трехчлена: все корни многочлена φ действительны тогда и только тогда, когда $D(\varphi) \geqslant 0$.

Задача. Доказать, что если φ — многочлен любой степени с действительными 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

$$\varphi = x^n + a_1x^{n-1} + a_2x^{n-2} + \ldots + a_{n-1}x + a_n$$

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

$$\psi = y^n + b_2y^{n-2} + \ldots + b_{n-1}y + b_n,$$

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

$$\varphi = x^3 + px + q. \tag{60}$$

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 $x_1^4x_2^2$. Выпишем наборы показателей старших членов симметрических многочленов, которые могут встретиться в процессе, описанном в доказательстве теоремы 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

$$f = \sigma_1^2\sigma_2^2 + a\sigma_1^3\sigma_3 + b\sigma_2^3 + c\sigma_1\sigma_2\sigma_3 + d\sigma_3^2. \tag{61}$$

Для вычисления $D(\varphi)$ мы должны будем сделать в выражении (61) 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 (60) 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

$$D(\varphi) = -4p^3 - 27q^2. \tag{62}$$

Пример 1. Найдем число действительных 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 находим:

$$D(\varphi) = D(\psi) = 4 \cdot 4{,}33^3 - 27 \cdot 3{,}468^2 = 0{,}013 > 0.$$

Следовательно, многочлен φ имеет 3 различных действительных 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, w. Тогда $1, w, w^{-1}$ — это все кубические корни из единицы, и по формуле Виета получаем

$$w + w^{-1} = -1. \tag{63}$$

Рассмотрим линейные многочлены

$$h_1 = x_1 + wx_2 + w^{-1}x_3, \quad h_2 = x_1 + w^{-1}x_2 + wx_3.$$

При перестановке x_2 и x_3 они меняются местами, а при перестановке x_1 и x_2 многочлен h_1 переходит в wh_2, а h_2 — в $w^{-1}h_1$. Отсюда следует, что многочлены

$$f = h_1^3 + h_2^3, \quad g = h_1h_2$$

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 (60). Положим

$$d_1 = c_1 + wc_2 + w^{-1}c_3, \quad d_2 = c_1 + w^{-1}c_2 + wc_3.$$

Из предыдущего следует, что

$$d_1^3 + d_2^3 = -27q, \quad d_1d_2 = -3p$$

и, значит,

$$d_1^3d_2^3 = -27p^3.$$

Таким образом, d_1^3 и d_2^3 — это корни квадратного уравнения

$$x^2 + 27qx - 27p^3 = 0.$$

Решая его, находим:

$$d_1^3 = 27\left(-\frac{q}{2} + \sqrt{\frac{p^3}{27} + \frac{q^2}{4}}\right), \tag{64}$$

$$d_2^3 = 27\left(-\frac{q}{2} - \sqrt{\frac{p^3}{27} + \frac{q^2}{4}}\right). \tag{65}$$

Заметим, что выражение, стоящее под знаком радикала, лишь множителем $-\frac{1}{108}$ отличается от дискриминанта многочлена (60).

Складывая равенства

$$c_1 + c_2 + c_3 = 0,$$
$$c_1 + wc_2 + w^{-1}c_3 = d_1,$$
$$c_1 + w^{-1}c_2 + wc_3 = d_2,$$

с учетом соотношения (63) 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 (64) и (65), связанные полученным выше соотношением

$$d_1d_2 = -3p. \tag{66}$$

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 действительными коэффициентами, то нам, вообще говоря, придется работать с комплексными числами, даже если нас интересуют только действительные корни. Именно так обстоит дело в случае положительного дискриминанта, когда все три корня действительны: 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

$$-\frac{q}{2}+\sqrt{\frac{p^3}{27}+\frac{q^2}{4}}\approx -1{,}734+0{,}00347i\approx$$
$$\approx 1{,}73400[\cos(\pi-0{,}00200)+i\sin(\pi-0{,}00200)].$$

Под знаком другого кубического радикала будет стоять комплексно сопряженное число. Условие (66) означает в данном случае, что при извлечении кубических корней следует комбинировать их комплексно сопряженные значения. При сложении комплексно сопряженных чисел получается их удвоенная действительная часть. Таким образом,

$$c_1\approx 2\sqrt[3]{1{,}73400}\cos\frac{\pi-0{,}00200}{3}\approx 1{,}20278,$$
$$c_2\approx 2\sqrt[3]{1{,}73400}\cos\frac{\pi+0{,}00200}{3}\approx 1{,}20001,$$
$$c_3\approx -2\sqrt[3]{1{,}73400}\cos\frac{0{,}00200}{3}\approx -2{,}40277.$$

§3.10. 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

$$(a_1,b_1)\sim(a_2,b_2)\iff a_1b_2=a_2b_1.$$

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

$$a_1b_2b_3=a_2b_1b_3=a_3b_1b_2,$$

откуда после сокращения на b_2 получаем:

$$a_1b_3=a_3b_1,$$

т. е. $(a_1,b_1)\sim(a_3,b_3)$.

Из данного определения следует, что

$$(a,b)\sim(ac,bc) \tag{67}$$

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 (67):

$$(a_1,b_1)\sim(a_1b_2,b_1b_2)=(a_2b_1,b_1b_2)\sim(a_2,b_2.)$$

(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:

$$(a_1,b_1)+(a_2,b_2)=(a_1b_2+a_2b_1,b_1b_2),$$
$$(a_1,b_1)(a_2,b_2)=(a_1a_2,b_1b_2).$$

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» $\frac{a}{b}$ (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:

$$\frac{a_1}{b_1}+\frac{a_2}{b_2}=\frac{a_1b_2+a_2b_1}{b_1b_2},\qquad \frac{a_1}{b_1}\frac{a_2}{b_2}=\frac{a_1a_2}{b_1b_2}.$$

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 умно-

жения дробей относительно сложения:

$$\left(\frac{a_1}{b}+\frac{a_2}{b}\right)\frac{a_3}{b_3}=\frac{(a_1+a_2)a_3}{bb_3}=\frac{a_1a_3+a_2a_3}{bb_3}=\frac{a_1}{b}\frac{a_3}{b_3}+\frac{a_2}{b}\frac{a_3}{b_3}.$$

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 $\operatorname{Quot} A$.

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 $\operatorname{Quot} A$. 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 $\operatorname{Quot} A$. Таким образом, обозначение $\frac{a}{b}$ можно теперь понимать содержательным образом.

В силу (67) 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.) Несократимый вид дроби 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 1. *Всякая рациональная дробь единственным образом представляется в виде суммы многочлена и правильной дроби.*

Доказательство. Пусть $f,g\in K[x]$, $g\neq 0$. Разделим f на g с остатком в кольце $K[x]$:

$$f=qg+r\quad(q,r\in K[x],\ \deg r<\deg g). \tag{68}$$

Тогда

$$\frac{f}{g}=q+\frac{r}{g}, \tag{69}$$

причем $\frac{r}{g}$ — правильная дробь. Обратно, из равенства (69) следует равенство (68), так что единственность искомого представления рациональной дроби вытекает из однозначности деления с остатком в кольце многочленов.

Изложим теперь теорию, используемую в математическом анализе при интегрировании рациональных функций.

Определение. 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 $\frac{f}{g}$ единственным образом 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. Возможность указанного разложения докажем индукцией по числу неприводимых множителей многочлена g. Если g — неприводимый многочлен, то сама дробь $\frac{f}{g}$ является простейшей. Если $g = p^k$, где p — неприводимый многочлен, то разделим f на p с остатком:

$$f = qp + r, \quad \deg r < \deg p.$$

Мы тогда получим:

$$\frac{f}{g} = \frac{q}{p^{k-1}} + \frac{r}{p^k}.$$

Второе из слагаемых является простейшей дробью, а первое может быть разложено в сумму простейших дробей со знаменателями $p, p^2, \dots, p^{k-1}$ по предположению индукции.

Пусть теперь $s \geqslant 2$. Так как многочлены $p_1^{k_1}$ и $h = p_2^{k_2} \dots p_s^{k_s}$ взаимно просты, то существуют такие многочлены u и v, что

$$f = up_1^{k_1} + vh. \tag{70}$$

Мы можем добиться того, чтобы

$$\deg u < \deg h. \tag{71}$$

В самом деле, если это не так, разделим u на h с остатком:

$$u = qh + r, \quad \deg r < \deg h.$$

Равенство (70) может быть тогда переписано в виде

$$f = rp_1^{k_1} + (v + qp_1^{k_1})h = u_1 p_1^{k_1} + v_1 h,$$

где $u_1 = r$, $v_1 = v + qp_1^{k_1}$ и $\deg u_1 < \deg h$.

В предположении, что выполнено (71), разделим равенство (70) на g. Мы тогда получим:

$$\frac{f}{g} = \frac{u}{h} + \frac{v}{p_1^{k_1}}.$$

В силу (71) первое из слагаемых является правильной дробью. По предположению индукции оно может быть разложено в сумму простейших дробей со знаменателями

$$p_2, p_2^2, \dots, p_2^{k_2}, \dots, p_s, p_s^2, \dots, p_s^{k_s}.$$

Второе слагаемое является правильной дробью как разность правильных дробей и по доказанному выше может быть разложено в сумму простейших дробей со знаменателями

$$p_1, p_1^2, \dots, p_1^{k_1}.$$

Это и дает нужное разложение дроби $\frac{f}{g}$.

Докажем теперь единственность такого разложения. Если имеются два таких разложения, то, вычитая одно из другого, получаем равенство вида

$$\frac{u_{11}}{p_1} + \frac{u_{12}}{p_1^2} + \dots + \frac{u_{1k_1}}{p_1^{k_1}} + \frac{u_{21}}{p_2} + \frac{u_{22}}{p_2^2} + \dots + \frac{u_{2k_2}}{p_2^{k_2}} + \\ + \frac{u_{s1}}{p_s} + \frac{u_{s2}}{p_s^2} + \dots + \frac{u_{sk_s}}{p_s^{k_s}} = 0,$$

где $\deg u_{ij} < \deg p_i$. Нам нужно доказать, что все многочлены u_{ij} равны нулю. Предположим, что это не так. Пусть, например, среди многочленов $u_{11}, u_{12}, \dots, u_{1k_1}$ имеются ненулевые, и пусть u_{1k} $(k \leqslant k_1)$ — последний из них. Умножив тогда предыдущее равенство на $p_1^k p_2^{k_2} \dots p_s^{k_s}$, мы получим в левой части сумму многочленов, все слагаемые которой, кроме $u_{1k} p_2^{k_2} \dots p_s^{k_s}$, делятся на p_1. Следовательно, и это слагаемое должно делиться на p_1, но это невозможно, так как $\deg u_{1k} < \deg p_1$.

Пример 1. Предположим, что

$$g = (x - c_1)(x - c_2) \dots (x - c_n),$$

где $c_1, c_2, \ldots, c_n$ различны. Тогда

$$\frac{f}{g} = \frac{a_1}{x - c_1} + \frac{a_2}{x - c_2} + \ldots + \frac{a_n}{x - c_n},$$

где $a_1, a_2, \ldots, a_n \in K$. Для нахождения a_i умножим обе части предыдущего равенства на g и положим $x = c_i$. Мы получим тогда:

$$f(c_i) = a_i(c_i - c_1)\ldots(c_i - c_{i-1})(c_i - c_{i+1})\ldots(c_i - c_n) = a_i g'(c_i),$$

откуда

$$a_i = \frac{f(c_i)}{g'(c_i)}.$$

Итак,

$$\frac{f}{g} = \sum_{i=1}^{n} \frac{f(c_i)}{g'(c_i)(x - c_i)} \tag{72}$$

(при условии, что $\deg f < \deg g$). Интересно отметить, что, умножив обе части этого равенства на g, мы получим *интерполяционную формулу Лагранжа*

$$f = \sum_{i=1}^{n} b_i \frac{(x - c_1)\ldots(x - c_{i-1})(x - c_{i+1})\ldots(x - c_n)}{(c_i - c_1)\ldots(c_i - c_{i-1})(c_i - c_{i+1})\ldots(c_1 - c_n)},$$

задающую многочлен f степени $< n$, принимающий в точках $c_1, c_2, \ldots, c_n$ значения $b_1, b_2, \ldots, b_n$.

Задача 2. Доказать равенство

$$\frac{1}{x^n - 1} = \frac{1}{n} \sum_{i=0}^{n-1} \frac{\varepsilon_i}{x - \varepsilon_i},$$

где $\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{n-1}$ — комплексные корни n-й степени из единицы.

Задача 3. Разложить в сумму простейших дробей над полем $\mathbb{Z}_p$ (p — простое) дробь $\dfrac{1}{x^p - x}$.

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:

$$\frac{x}{(x+1)(x^2+1)^2} = \frac{a}{x+1} + \frac{bx+c}{x^2+1} + \frac{dx+e}{(x^2+1)^2},$$

где a, b, c, d, e — какие-то действительные 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:

$$-1 = 4a, \quad i = (di+e)(i+1) = (e-d) + (d+e)i,$$

откуда

$$a = -\frac{1}{4}, \quad d = e = \frac{1}{2}.$$

Далее, сравнив свободные члены и коэффициенты при x^4, получим:

$$0 = a + c + e, \quad 0 = a + b,$$

откуда

$$b = \frac{1}{4}, \quad c = -\frac{1}{4}.$$

Таким образом,

$$\frac{x}{(x+1)(x^2+1)^2} = -\frac{1}{4(x+1)} + \frac{x-1}{4(x^2+1)} + \frac{x+1}{2(x^2+1)^2}.$$

Глава 4. Начала теории групп

§4.1. 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 самого 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) $\mathrm{id} \in G$.

(Здесь $\varphi\psi$ обозначает произведение (композицию) преобразований φ и ψ, 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 $X = \{1, 2, \ldots, n\}$; 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

$$\sigma = \begin{pmatrix} i_1 & i_2 & \ldots & i_n \\ j_1 & j_2 & \ldots & j_n \end{pmatrix},$$

в первой строке которой выписаны в каком-то порядке числа $1, 2, \ldots, n$, 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 в начале 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 (соотв. евклидова пространства E^3) образуют группу преобразований, обозначаемую через $\mathrm{Isom}\, E^2$ (соотв. $\mathrm{Isom}\, E^3$). Это свойство является аксиомой в той версии аксиоматики евклидовой геометрии, в которой понятие движения является одним из неопределяемых понятий. В другой версии, берущей за основу понятие расстояния между точками, движение определяется как преобразование, сохраняющее расстояния, а сформулированное выше свойство является очевидной теоремой.

Замечание 1. В предыдущих главах мы обозначали через E^2 (соотв. E^3) множество векторов евклидовой плоскости (соотв. пространства). Здесь же символ E^2 (соотв. E^3) использован для обозначения самой евклидовой плоскости (соотв. пространства). Впрочем, если в плоскости (соотв. в пространстве) фиксирована некоторая точка o (которую мы будем в дальнейшем называть *началом координат*), 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 движения, утверждение о том, что всякое биективное преобразование, сохраняющее расстояния, является движением, является (несложной) теоремой.

Пример 3. Ввиду свойств линейных отображений, доказанных в §2.3, биективные линейные преобразования векторного пространства V образуют группу преобразований. Она называется *полной линейной группой* пространства V и обозначается через $GL(V)$.

Пример 4. Назовем *параллельным переносом* векторного пространства V на вектор $a \in V$ преобразование

$$t_a \colon x \mapsto x + a.$$

Легко видеть, что

$$t_a t_b = t_{a+b}, \quad t_a^{-1} = t_{-a}, \quad \mathrm{id} = t_0. \tag{73}$$

Эти формулы показывают, что совокупность $\mathbf{Tran}(V)$ 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) $(ab)c = a(bc)$ $\forall a, b, c \in G$ (*ассоциативность*);

2) существует такой элемент $e \in G$ (*единица*), что $ae = ea = a$ $\forall a \in G$;

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 $a^{-1}b$, и 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 квадратными матрицами порядка n и линейными преобразованиями пространства K^n, причем невырожденным матрицам отвечают обратимые линейные преобразования, а умножению матриц соответствует умножение линейных преобразований, группа $GL_n(K)$ изоморфна группе $GL(K^n)$ (и, тем самым, группе $GL(V)$, где V — любое n-мерное векторное пространство над полем K).

Пример 6. Как показывают формулы (73), группа $\mathrm{Tran}(V)$ 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 $G = \{А, Б, В, Г, Д, Е\}$ с операцией, заданной таблицей

	$А$	$Б$	$В$	$Г$	$Д$	$Е$
$А$	$Е$	$Д$	$Г$	$В$	$Б$	$А$
$Б$	$В$	$Г$	$Д$	$Е$	$А$	$Б$
$В$	$Б$	$А$	$Е$	$Д$	$Г$	$В$
$Г$	$Д$	$Е$	$А$	$Б$	$В$	$Г$
$Д$	$Г$	$В$	$Б$	$А$	$Е$	$Д$
$Е$	$А$	$Б$	$В$	$Г$	$Д$	$Е$

является группой, изоморфной S_3.

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 3. Так как $aa^{-1} = e$, то 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

$$\operatorname{Sym} f = \{\, \sigma \in S_n : f(x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(n)}) = f(x_1, x_2, \ldots, x_n) \,\}$$

есть подгруппа группы S_n. В самом деле, пусть $\sigma, \tau \in \operatorname{Sym} f$. Положим $x_{\sigma(i)} = y_i$; тогда

$$f(x_{\sigma\tau(1)}, x_{\sigma\tau(2)}, \ldots, x_{\sigma\tau(n)}) = f(y_{\tau(1)}, y_{\tau(2)}, \ldots, y_{\tau(n)}) =$$
$$= f(y_1, y_2, \ldots, y_n) = f(x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(n)}) = f(x_1, x_2, \ldots, x_n).$$

Остальные две аксиомы подгруппы выполнены очевидным образом. Определение симметрического многочлена, данное в §3.8, означает, что $\operatorname{Sym} f = S_n$. 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 $\operatorname{Sym} f$ 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, линейные преобразования пространства K^n, сохраняющие какой-либо заданный многочлен от n переменных, образуют подгруппу группы $GL_n(K)$. Линейные преобразования пространства $\mathbb{R}^n$, сохраняющие многочлен $x_1^2 + x_2^2 + \ldots + x_n^2$, называются *ортогональными преобразованиями*; они образуют подгруппу группы $GL_n(\mathbb{R})$, которая называется *ортогональной группой* и обозначается через O_n. Так как в декартовых координатах пространства E^2 (соотв. E^3) многочлен $x^2 + y^2$ (соотв. $x^2 + y^2 + z^2$) выражает квадрат длины вектора, то ортогональные преобразования пространства E^2 (соотв. E^3) — это не что иное,

как линейные преобразования, сохраняющие длину вектора. Дадим геометрическое описание ортогональных преобразований пространства E^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.

$$a^2 + c^2 = b^2 + d^2 = 1, \quad ab + cd = 0.$$

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,

$$\varphi = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \quad \text{или} \quad \begin{pmatrix} \cos\alpha & \sin\alpha \\ \sin\alpha & -\cos\alpha \end{pmatrix}.$$

В первом случае, как мы уже знаем (см. пример 2.3.5) φ есть поворот на угол α. Во втором случае φ — зеркальное отражение относительно прямой l, образующей угол $\frac{\alpha}{2}$ с осью x (см. рис. 20). Эти два случая отличаются друг от друга тем, что в первом случае φ сохраняет ориентацию плоскости, а во втором — меняет. В курсе линейной алгебры доказывается, что всякое ортогональное преобразование пространства E^3, сохраняющее ориентацию, есть поворот вокруг некоторой прямой.

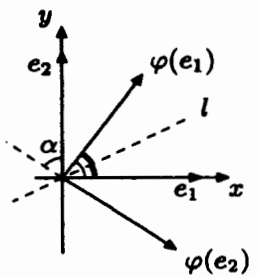

Рис. 20

Пример 10. Движения евклидовой плоскости, оставляющие на месте начало координат o, образуют подгруппу группы $\operatorname{Isom} E^2$. Обозначим ее через H. Так как сложение векторов и их умножение на числа определяются в инвариантных геометрических терминах, то всякое движение, оставляющее на месте точку o, является линейным преобразованием. Более того, так как оно сохраняет длины

векторов, то оно является ортогональным преобразованием. Обратно, поскольку расстояние между точками a и b есть длина вектора $a-b$, то всякое ортогональное преобразование сохраняет расстояние между точками и, значит, является движением. Таким образом, $H = O_2$. Аналогично, группа движений евклидова пространства, оставляющих на месте начало координат, совпадает с O_3.

Пример 11. Пусть F — какая-либо фигура на евклидовой плоскости. Тогда

$$\operatorname{Sym} F = \{\, \varphi \in \operatorname{Isom} E^2 : \varphi(F) = F \,\}$$

есть подгруппа группы $\operatorname{Isom} E^2$; она называется *группой симметрии* фигуры F. Так, группа симметрии окружности с центром в начале координат 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 $\frac{2\pi}{n}$, 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 $GL_n(K)$. Эта подгруппа называется *унимодулярной группой* и обозначается через $SL_n(K)$.

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 3. Доказать, что множество строго треугольных квадратных матриц порядка n является подгруппой группы $GL_n(K)$.

§4.2. Группы в геометрии и физике

Цель этого параграфа — дать общее представление о роли групп в геометрии и физике.

В XIX веке 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 изучения свойств фигур, инвариантных 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 $F_1 \underset{G}{\sim} F_2$, если существует такой преобразование $\varphi \in G$, что $F_2 = \varphi(F_1)$. Проверим, что это действительно отношение эквивалентности:

1) $F \underset{G}{\sim} F$, так как $F = \mathrm{id}(F)$ и $\mathrm{id} \in G$;

2) если $F_1 \underset{G}{\sim} F_2$, т. е. $F_2 = \varphi(F_1)$, где $\varphi \in G$, то $F_2 \underset{G}{\sim} F_1$, так как $F_1 = \varphi^{-1}(F_2)$ и $\varphi^{-1} \in G$;

3) если $F_1 \underset{G}{\sim} F_2$ и $F_2 \underset{G}{\sim} F_3$, т. е. $F_2 = \varphi(F_1)$ и $F_3 = \psi(F_2)$, где $\varphi, \psi \in G$, то $F_1 \underset{G}{\sim} F_3$, 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. 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. Группа $\mathrm{Tran}(V)$ параллельных переносов векторного пространства V (см. пример 1.4) транзитивна. В самом деле, для любых $x, y \in V$ имеем

$$y = t_{y-x}x.$$

Однако группа $\mathrm{Tran}(V)$ все еще слишком мала, чтобы определять интересную геометрию. В качестве примера интересной геометрии, отличной от евклидовой, приведем аффинную геометрию.

Пусть V — какое-либо векторное пространство, $\varphi \in GL(V)$ и $a \in V$. Тогда

$$\varphi t_a \varphi^{-1} = t_{\varphi(a)}. \tag{74}$$

В самом деле, для любого $x \in V$ имеем:

$$(\varphi t_a \varphi^{-1})(x) = \varphi(\varphi^{-1}(x) + a) = x + \varphi(a) = t_{\varphi(a)}x.$$

Предложение 1. *Для любой подгруппы $G \subset GL(V)$ множество*

$$\mathrm{Tran}(V) \cdot G = \{t_a\varphi : a \in V,\ \varphi \in G\}$$

является транзитивной группой преобразований пространства V.

Доказательство. При $a, b \in V$, $\varphi, \psi \in GL(V)$ имеем ввиду формул (73) и (74):

$$(t_a\varphi)(t_b\psi) = t_a(\varphi t_b \varphi^{-1})\varphi\psi = t_{a+\varphi(b)}\varphi\psi \in \mathrm{Tran}(V) \cdot G.$$

Отсюда следует, что

$$(t_a\varphi)^{-1} = t_{-\varphi^{-1}(a)}\varphi^{-1} \in \mathrm{Tran}(V) \cdot G.$$

Таким образом, $\mathrm{Tran}(V) \cdot G$ — группа преобразований. Она транзитивна, поскольку уже ее подгруппа $\mathrm{Tran}(V)$ транзитивна.

В частности, мы можем взять $G = GL(V)$. Полученная группа

$$GA(V) = \mathrm{Tran}(V) \cdot GL(V) \tag{75}$$

называется *полной аффинной группой* пространства V, а ее элементы — (биективными) *аффинными преобразованиями*. Связанная с ней геометрия называется *аффинной геометрией*.

В случае $V = E^2$ мы получаем аффинную геометрию евклидовой плоскости.

Предложение 2. *Группа движений евклидовой плоскости (соотв. пространства) есть подгруппа группы $GA(E^2)$, равная $\mathrm{Tran}(E^2) \cdot O_2$.*

Доказательство. Прежде всего, заметим, что все параллельные переносы и все ортогональные преобразования являются движениями. Пусть теперь α — какое-либо движение. Положим $a = \alpha(o)$. Тогда движение $\varphi = t_a^{-1}\alpha$ оставляет на месте точку o и, значит, принадлежит группе O_2 (см. пример 1.10). Таким образом,

$$\alpha = t_a\varphi \in \mathrm{Tran}(E^2) \cdot O_2$$

Следствие. *Если фигуры $F_1, F_2 \subset E^2$ равны в евклидовой геометрии, то они равны и в аффинной геометрии.*

Группа $GA(E^2)$ 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 $GA(E^2)$ 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.

Аналогичным образом определяется аффинная геометрия евклидова пространства.

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:

$$G = \mathrm{Tran}(\mathbb{R}^4) \cdot H \cdot O_3,$$

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*

$$G = \mathrm{Tran}(\mathbb{R}^4) \cdot O_{3,1},$$

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*

$$(x, y, z, t) \mapsto (x \operatorname{ch} a + t \operatorname{sh} a, y, z, x \operatorname{sh} a + t \operatorname{ch} a),$$

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.

§4.3. Циклические группы

Так же, как в группе $\mathbb{R}^*$, в любой группе G могут быть определены *степени* элемента $g \in G$ с целыми показателями:

$$g^k = \begin{cases} \underbrace{gg\ldots g}_{k}, & \text{если } k > 0, \\ e, & \text{если } k = 0, \\ \underbrace{g^{-1}g^{-1}\ldots g^{-1}}_{-k}, & \text{если } k < 0. \end{cases}$$

Имеет место свойство

$$g^k g^l = g^{k+l}. \tag{76}$$

Это очевидно, если $k, l > 0$. Рассмотрим случай, когда $k > 0$, $l < 0$, $k + l > 0$. Тогда

$$g^k g^l = \underbrace{gg\ldots g}_{k}\underbrace{g^{-1}g^{-1}\ldots g^{-1}}_{-l} = \underbrace{gg\ldots g}_{k+l} = g^{k+l}.$$

Аналогично рассматриваются остальные случаи.

Из (76) 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 $\operatorname{ord} g$.

Предложение 1. *Если* $\operatorname{ord} g = n$, *то*
1) $g^m = e \iff n \mid m$;
2) $g^k = g^l \iff k \equiv l \pmod n$.

Доказательство. 1) Разделим m на n с остатком:

$$m = qn + r, \quad 0 \leqslant r < n.$$

Имеем тогда в силу определения порядка:

$$g^m = (g^n)^q \cdot g^r = g^r = e \iff r = 0.$$

2) Имеем в силу предыдущего:

$$g^k = g^l \iff g^{k-l} = e \iff n \mid (k-l) \iff k \equiv l \pmod n.$$

Следствие. *Если* $\operatorname{ord} g = n$, *то подгруппа* $\langle g\rangle$ *содержит* n *элементов*.

Доказательство. Действительно,

$$\langle g\rangle = \{e, g, g^2, \ldots, g^{n-1}\}, \tag{77}$$

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 $\operatorname{ord} g = \infty$. Отметим, что $\operatorname{ord} e = 1$; 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. Соответственно этому порядок элемента g аддитивной группы G — это наименьшее из натуральных чисел m (если такие существуют), для которых

$$mg \doteqdot \underbrace{g + g + \ldots + g}_{m} = 0.$$

Пример 1. Характеристика поля (см. §1.6) 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 $(i_1 i_2 \ldots i_p)$, если она циклически переставляет $i_1, i_2, \ldots, i_p$, т. е.

$$\tau(i_1) = i_2,\ \tau(i_2) = i_3,\ \ldots,\ \tau(i_p) = i_1,$$

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,

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 5 & 6 & 7 & 4 & 8 & 3 & 2 & 1 \end{pmatrix} = (2637)(158),$$

что наглядно показано на рис. 21, где действие подстановки σ изображено стрелками. Если подстановка σ разлагается в произведение независимых циклов длин $p_1, p_2, \ldots, p_s$, то

$$\operatorname{ord}\sigma = \text{НОК}\{p_1, p_2, \ldots, p_s\}.$$

Например, для подстановки σ, изображенной на рис. 21, $\operatorname{ord}\sigma = 12$.

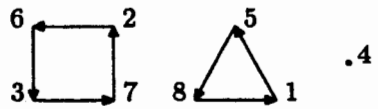

Рис. 21

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. $\dfrac{\arg c}{\pi} \in \mathbb{Q}$.

Задача 2. Доказать, что $\operatorname{arctg}\dfrac{3}{4}$ несоизмерим с π.

Пример 4. Найдем элементы конечного порядка в группе $\operatorname{Isom} E^2$ движений плоскости. Пусть $\varphi \in \operatorname{Isom} E^2$, $\varphi^n = \mathrm{id}$. Для

любой точки $p \in E^2$ точки

$$p, \varphi p, \varphi^2 p, \ldots, \varphi^{n-1} p$$

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 $GL_2(\mathbb{R})$. 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 $A \in GL_2(\mathbb{R})$ будет конечен, равна нулю.

Предложение 2. *Если* $\operatorname{ord} g = n$, *то*

$$\operatorname{ord} g^k = \frac{n}{(n,k)}. \tag{78}$$

Доказательство. Пусть

$$(n,k) = d, \quad n = n_1 d, \quad k = k_1 d,$$

так что $(n_1, k_1) = 1$. Имеем:

$$(g^k)^m = e \iff n \mid km \iff n_1 \mid k_1 m \iff n_1 \mid m.$$

Следовательно, $\operatorname{ord} g^k = n_1$.

Определение. 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

$$\varepsilon_k = \cos\frac{2\pi k}{n} + i \sin\frac{2\pi k}{n} \quad (k = 0, 1, \ldots, n-1).$$

Ясно, что $\varepsilon_k = \varepsilon_1^k$. Следовательно, группа C_n порождается элементом ε_1.

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 (76) отображение $f\colon \mathbb{Z} \to G$, $k \mapsto g^k$ есть изоморфизм.

Пусть $G = \langle g \rangle$ — конечная циклическая группа порядка n. Рассмотрим отображение

$$f\colon \mathbb{Z}_n \to G, \quad [k] \mapsto g^k \quad (k \in \mathbb{Z}).$$

Так как

$$[k] = [l] \iff k \equiv l \pmod{n} \iff g^k = g^l,$$

то отображение f корректно определено и биективно. Свойство

$$f(k+l) = f(k)f(l)$$

вытекает из той же формулы (76). Таким образом, f — изоморфизм.

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:

$$k = qm + r, \quad 0 \leqslant r < m.$$

Имеем:

$$g^r = g^k (g^m)^{-q} \in H,$$

откуда в силу определения 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 $n = qm$. Подгруппа H имеет вид

$$H = \{e, g^m, g^{2m}, \ldots, g^{(q-1)m}\} \tag{79}$$

и является единственной подгруппой порядка q в группе G. Обратно, если q — любой делитель числа n и $n = qm$, то подмножество H, определяемое равенством (79), является подгруппой порядка q.

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 $m \geqslant 0$.

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 $q \mid n$.

§4.4. Системы порождающих

Пусть S — какое-либо подмножество группы G. Обозначим через $\langle S \rangle$ совокупность всевозможных произведений вида

$$g_1^{\varepsilon_1} g_2^{\varepsilon_2} \ldots g_k^{\varepsilon_k} \quad (g_1, g_2, \ldots g_k \in S;\ \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_k = \pm 1). \tag{80}$$

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

$$(g_1^{\varepsilon_1} g_2^{\varepsilon_2} \ldots g_k^{\varepsilon_k})(g_{k+1}^{\varepsilon_{k+1}} g_{k+2}^{\varepsilon_{k+2}} \ldots g_{k+l}^{\varepsilon_{k+l}}) = g_1^{\varepsilon_1} g_2^{\varepsilon_2} \ldots g_{k+l}^{\varepsilon_{k+l}},$$

$$(g_1^{\varepsilon_1} g_2^{\varepsilon_2} \ldots g_k^{\varepsilon_k})^{-1} = g_k^{-\varepsilon_k} \ldots g_2^{-\varepsilon_2} g_1^{-\varepsilon_1}.$$

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. Удобно считать, что в число произведений (80) входит пустое произведение ($k = 0$), которое по определению равно e.

Определение. 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. 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

$$\sigma = \begin{pmatrix} 1 & 2 & \dots & n \\ k_1 & k_2 & \dots & k_n \end{pmatrix} \tag{81}$$

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

$$\tau_s \dots \tau_2 \tau_1 \sigma = \mathrm{id}$$

и, значит,

$$\sigma = \tau_1 \tau_2 \dots \tau_s.$$

Задача 1. Доказать, что группа S_n порождается *смежными транспозициями* $(12), (23), \dots, ((n-1)n)$, 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 (81).

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,

$$A = U_1^{-1} U_2^{-1} \dots U_s^{-1}.$$

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 произведением двух отражений.)

§4.5. 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

$$g_1^{-1} g_2 \in H, \tag{82}$$

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 (82) аналогичное ему условие

$$g_2 g_1^{-1} \in H. \tag{83}$$

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 действительной оси (рис. 22а).

Пример 2. Смежные классы мультипликативной группы $\mathbb{C}^*$ по подгруппе $\mathbb{R}_+^*$ положительных чисел — это лучи, исходящие из начала координат (рис. 22б).

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. 22в).

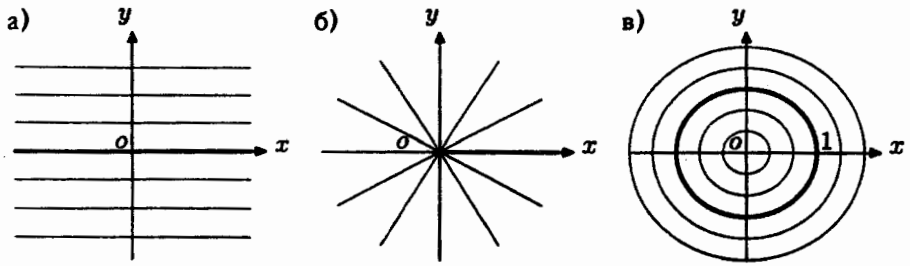

Рис. 22

Пример 4. В случае $G = GL_n(K)$, $H = SL_n(K)$ (см. пример 1.12) условие (82), равно как и (83), 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 $H \simeq S_{n-1}$, 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 $P_1, P_2, \ldots, P_n$, 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 $Q_1, Q_2, \ldots, Q_n$, где

$$Q_k = \{\sigma \in S_n : \sigma(k) = n\}.$$

Мы видим, что правые смежные классы отличны от левых (за исключением $Q_n = P_n = H$).

Множество левых смежных классов группы G по подгруппе H обозначается через G/H. Число смежных классов G по H (левых или правых, безразлично), если оно конечно, называется *индексом* подгруппы H и обозначается через $|G : H|$.

Теорема 1 (Лагранж). *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 $|H|$.

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.*

Действительно, порядок элемента равен порядку порождаемой им циклической подгруппы.

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.

Следствие 4. *Если* $|G| = n$, *то* $g^n = e$ *для любого* $g \in G$.

Доказательство. Пусть $\operatorname{ord} g = m$. В силу следствия 2 $m \mid n$. Следовательно, $g^n = e$.

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*.

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 $x \underset{G}{\sim} y$, 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 координат является ортогональная группа O_2 (см. пример 1.10).

Пример 8. Орбиты группы O_2 суть окружности с центром в начале координат o и сама точка o. Стабилизатор точки $p \neq o$ состоит из тождественного преобразования и отражения относительно прямой op, а стабилизатор точки o есть вся группа O_2.

Пример 9. Группа S_n транзитивна на множестве $\{1, 2, \ldots, n\}$. 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:

$$g_1 \equiv g_2 \pmod{G}_x \iff$$
$$\iff g_1^{-1} g_2 \in G_x \iff g_1^{-1} g_2 x = x \iff g_1 x = g_2 x.$$

Таким образом, элементы одного смежного класса 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.

Число элементов орбиты Gx, если оно конечно, называется ее *длиной* и обозначается через $|Gx|$.

Следствие. *Если* G — *конечная группа, то*

$$|G| = |Gx|\,|G_x|. \tag{84}$$

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

$$G_{gx} = gG_x g^{-1}.$$

Пример 10. Пусть $K \subset E^3$ — куб. Рассмотрим группу его симметрии

$$G = \operatorname{Sym} K = \{\varphi \in \operatorname{Isom} E^3 : \varphi(K) = K\}.$$

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,

$$|G| = 8|G_v|,$$

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. 23). Таким образом,

$$|\mathrm{Sym}\, K| = 8 \cdot 3 \cdot 2 = 48.$$

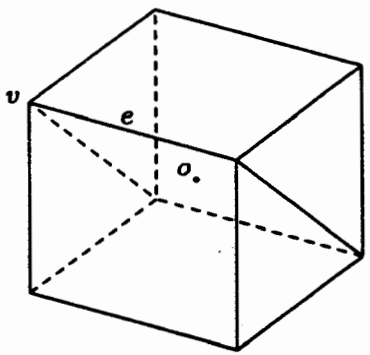

Рис. 23

Задача 2. Получить тот же результат еще двумя способами, рассмотрев группу $\mathrm{Sym}\, K$ как группу преобразований множества граней и множества ребер куба соответственно.

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 $f = x_1x_2 + x_3x_4$. Перестановками 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 (84) 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 что иное, как подгруппа, обозначенная в примере 1.8 через $\mathrm{Sym}\, f$.

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. Подгруппа H группы G называется *нормальной*, если

$$gH = Hg \quad \forall g \in G \tag{85}$$

или, что эквивалентно,

$$gHg^{-1} = H \quad \forall g \in G. \tag{86}$$

В этом случае пишут $H \lhd G$ (или $G \rhd H$).

Для того чтобы подгруппа H была нормальна, 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, что

$$g_1 \equiv g_1' \pmod H, \quad g_2 \equiv g_2' \pmod H \Longrightarrow$$
$$\Longrightarrow g_1g_2 \equiv g_1'g_2' \pmod H$$

или, что эквивалентно, для любых $g_1, g_2 \in G$ и $h_1, h_2 \in H$

$$(g_1h_1)(g_2h_2) \equiv g_1g_2 \pmod{H}.$$

Последнее условие согласно определению переписывается в виде

$$g_2^{-1}h_1g_2 \in H.$$

Так как g_2 может быть любым элементом группы G, а h_1 — любым элементом подгруппы H, то это равносильно условию нормальности (86).

Задача 3. Доказать, что всякое отношение эквивалентности в группе, согласованное с операцией, есть отношение сравнимости по модулю некоторой подгруппы.

Таким образом, если $H \lhd G$, то операция умножения в группе G определяет операцию умножения в множестве G/H по правилу

$$(g_1H)(g_2H) = g_1g_2H.$$

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 $\mathbb{C}$ 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

$$\mathbb{C}/\mathbb{R} \simeq \mathbb{R}.$$

Пример 14. Смежные классы $\mathbb{C}^*$ 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 $C_rC_s = C_{rs}$, так что

$$\mathbb{C}^*/\mathbb{T} \simeq \mathbb{R}_+^*.$$

Пример 15. Как мы видели выше (см. пример 4), левые смежные классы $GL_n(K)$ 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 $M_aM_b = M_{ab}$, так что

$$GL_n(K)/SL_n(K) \simeq K^*.$$

Пример 16. Подгруппа $H \simeq S_{n-1}$ группы S_n, рассмотренная в примере 5, не является нормальной при $n \geqslant 3$.

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 $n \geqslant 2$ и $|K| \geqslant 3$.

§4.6. Гомоморфизмы

Связи между различными алгебраическими структурами одного типа устанавливаются при помощи гомоморфизмов. Понятие гомоморфизма отличается от понятия изоморфизма тем, что оно не требует биективности. В одном случае мы уже встречались с этим понятием. А именно, гомоморфизмы векторных пространств — это не что иное, как их линейные отображения.

Дадим точное определение гомоморфизма групп.

Определение. *Гомоморфизмом* группы G в группу H называется отображение $f\colon G \to H$, 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) $\operatorname{Im} f = \{f(a) : a \in G\}$ есть подгруппа группы H (называемая *образом* гомоморфизма f). Это следует из определения гомоморфизма и предыдущих свойств.

4) $\operatorname{Ker} f = \{a \in G : f(a) = e\}$ есть нормальная подгруппа группы G (называемая *ядром* гомоморфизма f). Действительно,

$$a, b \in \operatorname{Ker} f \implies f(ab) = f(a)f(b) = e^2 = e \implies ab \in \operatorname{Ker} f,$$

$$a \in \operatorname{Ker} f \implies f(a^{-1}) = f(a)^{-1} = e^{-1} = e \implies a^{-1} \in \operatorname{Ker} f,$$

$$e \in \operatorname{Ker} f,$$

$$a \in \operatorname{Ker} f,\ g \in G \implies f(gag^{-1}) = f(g)f(a)f(g)^{-1} =$$

$$= f(g)ef(g)^{-1} = f(g)f(g)^{-1} = e \implies gag^{-1} \in \operatorname{Ker} f.$$

Гомоморфизм группы в саму себя называется ее *эндоморфизмом*. Изоморфизм группы на саму 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 (соотв. мультипликативная) абелева группа. Тогда для любого $n \in \mathbb{Z}$ отображение $x \mapsto nx$ (соотв. $x \mapsto x^n$) является эндоморфизмом группы G. (Для неабелевой группы это, вообще говоря, неверно.) В случае $G=\mathbb{C}^*$ ядром этого гомоморфизма является группа C_n корней n-ой степени из 1.

Пример 3. Согласно основному свойству экспоненты отображение $x \mapsto e^x$ является гомоморфизмом аддитивной группы $\mathbb{R}$ в мультипликативную группу $\mathbb{R}^*$. Его образ есть подгруппа $\mathbb{R}^*_+$ положительных чисел, а ядро тривиально. Это отображение может быть продолжено до гомоморфизма $\mathbb{C} \to \mathbb{C}^*$ по формуле

$$z=x+yi \mapsto e^z = e^x(\cos y + i \sin y) \quad (x,y \in \mathbb{R}).$$

В самом деле, пусть $z_1 = x_1 + y_1 i$, $z_2 = x_2 + y_2 i$; тогда

$$\begin{aligned} e^{z_1+z_2} &= e^{x_1+x_2}(\cos(y_1+y_2)+i\sin(y_1+y_2)) = \\ &= [e^{x_1}(\cos y_1 + i \sin y_1)][e^{x_2}(\cos y_2 + i \sin y_2)] = e^{z_1}e^{z_2}. \end{aligned}$$

Образ последнего гомоморфизма есть вся группа $\mathbb{C}^*$, а его ядро есть бесконечная циклическая подгруппа $2\pi i\mathbb{Z}$ группы $\mathbb{C}$.

Пример 4. Формула умножения определителей означает, что отображение

$$\det\colon GL_n(K) \to K^*, \quad A \mapsto \det A,$$

является гомоморфизмом. Его ядро — это унимодулярная группа $SL_n(K)$.

Пример 5. Назовем *знаком* подстановки $\sigma \in S_n$ и обозначим через $\operatorname{sgn}\sigma$ произведение знаков верхней и нижней перестановки в ее записи (см. пример 1.1):

$$\operatorname{sgn}\begin{pmatrix} i_1 & i_2 & \dots & i_n \\ j_1 & j_2 & \dots & j_n \end{pmatrix} = \operatorname{sgn}(i_1, i_2, \dots, i_n) \cdot \operatorname{sgn}(j_1, j_2, \dots, j_n).$$

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

$$\operatorname{sgn}\colon S_n \to C_2 = \{\pm 1\}, \quad \sigma \mapsto \operatorname{sgn}\sigma,$$

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

$$\begin{aligned} \operatorname{sgn}\sigma\tau &= \operatorname{sgn}(i_1, i_2, \dots, i_n) \cdot \operatorname{sgn}(k_1, k_2, \dots, k_n) = \\ &= [\operatorname{sgn}(i_1, i_2, \dots, i_n)\operatorname{sgn}(j_1, j_2, \dots, j_n)] \times \\ &\quad \times [\operatorname{sgn}(j_1, j_2, \dots, j_n)\operatorname{sgn}(k_1, k_2, \dots, k_n)] = \\ &= \operatorname{sgn}\tau \cdot \operatorname{sgn}\sigma = \operatorname{sgn}\sigma \cdot \operatorname{sgn}\tau. \end{aligned}$$

Ядро гомоморфизма sgn называется *знакопеременной группой* и обозначается через A_n. Употребляется также следующая терминология: подстановки σ, для которых $\operatorname{sgn}\sigma = 1$ (соотв. $\operatorname{sgn}\sigma = -1$), называются *четными* (соотв. *нечетными*). Таким образом, A_n — это подгруппа четных подстановок.

Задача 1. Вывести следующую формулу для знака циклической подстановки:

$$\operatorname{sgn}(i_1 i_2 \dots i_p) = (-1)^{p-1}.$$

Пользуясь этим, доказать, что знак любой подстановки равен $(-1)^{m-s}$, где m — число фактически переставляемых ею (т. е. не оставляемых на месте) элементов, а s — число независимых циклов, в произведение которых она разлагается.

Следующая теорема носит название *теоремы о гомоморфизме групп*.

Теорема 1. *Пусть $f\colon G \to H$ — гомоморфизм групп. Тогда*

$$\operatorname{Im} f \simeq G/\operatorname{Ker} f.$$

Более точно, имеется изоморфизм

$$\varphi\colon \operatorname{Im} f \to G/\operatorname{Ker} f,$$

сопоставляющий каждому элементу $h = f(g) \in \operatorname{Im} f$ смежный класс $g \operatorname{Ker} f$.

Доказательство. Доказательство этой теоремы аналогично доказательству теоремы 5.2. А именно,

$$g_1 \equiv g_2 \pmod{\operatorname{Ker} f} \iff g_1^{-1}g_2 \in \operatorname{Ker} f \iff$$
$$\iff f(g_1^{-1}g_2) = e \iff f(g_1) = f(g_2),$$

так что элементы одного смежного класса G по $\operatorname{Ker} f$ характеризуются тем, что их образом служит один и тот же элемент группы H (принадлежащий $\operatorname{Im} f$). Более точно, все элементы смежного класса $g \operatorname{Ker} f$, и только они, переходят при гомоморфизме f в элемент $h = f(g) \in \operatorname{Im} f$. 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 $f(g_1g_2) = h_1h_2$, и

$$\varphi(h_1h_2) = g_1g_2 \operatorname{Ker} f = (g_1 \operatorname{Ker} f)(g_2 \operatorname{Ker} f) = \varphi(h_1)\varphi(h_2),$$

что и требовалось доказать.

Следствие 1. *Гомоморфизм f является изоморфизмом тогда и только тогда, когда $\operatorname{Im} f = H$ и $\operatorname{Ker} f = \{e\}$.*

Следствие 2. *Если группа G конечна, то*

$$|G| = |\operatorname{Im} f|\,|\operatorname{Ker} f|.$$

(Интересно сравнить эту формулу с формулой (84))

Пример 6. Рассмотрим гомоморфизм

$$f\colon \mathbb{C} \to \mathbb{R}, \quad z \mapsto \operatorname{Im} z.$$

Имеем $\operatorname{Im} f = \mathbb{R}$, $\operatorname{Ker} f = \mathbb{R}$, так что

$$\mathbb{C}/\mathbb{R} \simeq \mathbb{R}$$

— результат, уже полученный нами в примере 5.13.

Пример 7. Рассмотрим гомоморфизм

$$f\colon \mathbb{C}^* \to \mathbb{R}_+^*, \quad z \mapsto |z|.$$

Имеем $\operatorname{Im} f = \mathbb{R}_+^*$, $\operatorname{Ker} f = \mathbb{T} = \{z \in \mathbb{C}^* : |z| = 1\}$, так что

$$\mathbb{C}^*/\mathbb{T} \simeq \mathbb{R}_+^*$$

— результат, уже полученный нами в примере 5.14.

Пример 8. Отображение

$$f\colon \mathbb{C}^* \to \mathbb{T}, \quad z \mapsto \frac{z}{|z|},$$

также является гомоморфизмом, причем $\operatorname{Im} f = \mathbb{T}$, $\operatorname{Ker} f = \mathbb{R}_+^*$. Следовательно,

$$\mathbb{C}^*/\mathbb{R}_+^* \simeq \mathbb{T}.$$

(Соответствующее разбиение на смежные классы было описана в примере 5.2.)

Пример 9. Аналогичным образом рассмотрение гомоморфизма det из примера 4 приводит к тому, что

$$GL_n(K)/SL_n(K) \simeq K^*$$

— результат, уже полученный в примере 5.15.

Пример 10. Рассмотрение гомоморфизма sgn из примера 5 приводит к тому, что

$$S_n/A_n \simeq C_2.$$

В частности, отсюда следует, что

$$|A_n| = \frac{1}{2}n!$$

Пример 11. Согласно определению (см. §2) всякое аффинное преобразование α есть произведение параллельного переноса и линейного преобразования φ. Последнее называется *линейной частью*, или *дифференциалом* преобразования α, и обозначается через $d\alpha$. Формула

$$(t_a\varphi)(t_b\psi) = t_{a+\varphi(b)}\,\varphi\psi,$$

полученная при доказательстве предложения 2.1, показывает, что отображение

$$d\colon GA(V) \to GL(V), \quad \alpha \mapsto d\alpha,$$

является гомоморфизмом. Очевидно, что

$$\operatorname{Im} d = GL(V), \quad \operatorname{Ker} d = \operatorname{Tran}(V),$$

так что

$$GA(V)/\operatorname{Tran}(V) \simeq GL(V).$$

Пример 12. Пусть $\Delta = A_1A_2A_3$ — правильный треугольник. Сопоставив каждому движению $\varphi \in \operatorname{Sym}\Delta$ подстановку $\sigma \in S_3$ по правилу

$$\varphi(A_i) = A_{\sigma(i)},$$

мы получим гомоморфизм

$$f\colon \operatorname{Sym}\Delta \to S_3.$$

Поскольку всякое движение плоскости, оставляющее на месте 3 точки, не лежащие на одной прямой, тождественно, $\operatorname{Ker} f = \{\mathrm{id}\}$. Докажем, что $\operatorname{Im} f = S_3$. Так как $\operatorname{Im} f$ — подгруппа группы S_3 и группа S_3 порождается транспозициями, то достаточно проверить, что любая транспозиция принадлежит $\operatorname{Im} f$, т. е. может быть осуществлена некоторым движением $\varphi \in \operatorname{Sym}\Delta$. Но это действительно так: например, транспозиция (12) осуществляется отражением относительно прямой l, показанной на рис. 24. Таким образом,

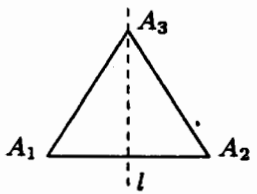

Рис. 24

$$\operatorname{Sym}\Delta \simeq S_3.$$

Аналогично доказывается, что группа симметрии правильного тетраэдра изоморфна S_4 (проделайте это!).

Пример 13. При перестановках переменных x_1, x_2, x_3, x_4 многочлены

$$x_1x_2 + x_3x_4, \ x_1x_3 + x_2x_4, \ x_1x_4 + x_2x_3 \tag{87}$$

переставляются между собой. Занумеровав их каким-либо образом, мы получим гомоморфизм

$$f\colon S_4 \to S_3.$$

Докажем, что $\operatorname{Im} f = S_3$. Для этого достаточно проверить, что любая транспозиция многочленов (87) может быть осуществлена некоторой перестановкой x_1, x_2, x_3, x_4. Но это действительно так:

например, транспозиция первых двух многочленов (87) может быть осуществлена транспозицией x_2 и x_3.

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 $GL_2(\mathbb{Z}_p)$ и $SL_2(\mathbb{Z}_p)$.

Очевидно, что композиция гомоморфизмов $F \to G$ и $G \to H$ есть гомоморфизм $F \to H$.

Пример 14. Рассмотрим композицию гомоморфизмов

$$\det\colon GL_n(\mathbb{R}) \to \mathbb{R}^* \quad \text{и} \quad \operatorname{sgn}\colon \mathbb{R}^* \to C_2 = \{\pm 1\},$$

где sgn обозначает знак действительного числа. Мы получим таким образом гомоморфизм

$$\varepsilon\colon GL_n(\mathbb{R}) \to C_2.$$

При $n = 2$ он имеет следующий геометрический смысл: если $\varepsilon(A) = 1$ (соотв. $\varepsilon(A) = -1$), то линейное преобразование пространства E^2, определяемое матрицей A, сохраняет (соотв. меняет) ориентацию в том смысле, что любой положительно ориентированный базис оно переводит в положительно (соотв. отрицательно) ориентированный базис. Аналогичная интерпретация возможна и при $n = 3$.

Пример 15. Композиция гомоморфизмов

$$d\colon GA(\mathbb{R}^n) \to GL(\mathbb{R}^n) = GL_n(\mathbb{R}) \quad \text{и} \quad \varepsilon\colon GL_n(\mathbb{R}) \to C_2$$

есть гомоморфизм

$$GA(\mathbb{R}^n) \to C_2. \tag{88}$$

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 (соотв. меняет) ориентацию, если его дифференциал сохраняет (соотв. меняет) ориентацию. В частности, можно говорить о движениях, сохраняющих или меняющих ориентацию (что мы уже делали раньше, не давая точного определения).

Пример 16. Пусть $G \subset \operatorname{Isom} E^n$ ($n = 2$ или 3) — какая-либо подгруппа, содержащая движения, меняющие ориентацию. Рассматривая ограничение на G гомоморфизма (88), мы получаем, что

подмножество движений из G, сохраняющих ориентацию, есть подгруппа индекса 2. Мы будем обозначать эту подгруппу через G_+.

Пример 17. В частности, подгруппу $\mathrm{Sym}_+ K \subset \mathrm{Sym}\, K$ будем называть *группой вращений куба* K. Так как $|\mathrm{Sym}\, K| = 48$ (см. пример 5.10), а $\mathrm{Sym}_+ K$ есть подгруппа индекса 2, то

$$|\mathrm{Sym}_+ K| = 24.$$

Докажем, что

$$\mathrm{Sym}_+ K \simeq S_4.$$

Для этого занумеруем каким-либо образом 4 диагонали куба K и сопоставим каждому движению $\varphi \in \mathrm{Sym}_+ K$ подстановку, осуществляемую им на множестве диагоналей. Мы получим гомоморфизм

$$f\colon \mathrm{Sym}_+ K \to S_4.$$

Докажем, что $\mathrm{Im}\, f = S_4$, откуда уже будет следовать, что f — изоморфизм, поскольку $|\mathrm{Sym}_+ K| = |S_4|$. Для этого достаточно проверить, что любая транспозиция принадлежит $\mathrm{Im}\, f$. Но это действительно так: например, транспозиция (12) осуществляется поворотом на π вокруг прямой l, изображенной на рис. 25.

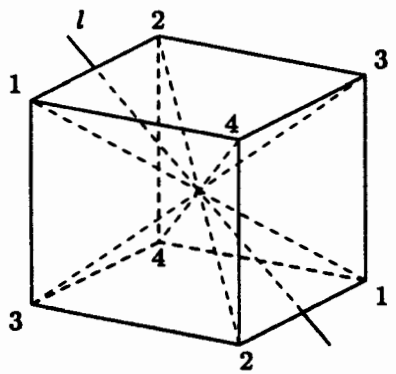

Рис. 25

Задача 4. Доказать, что группа D_4 (группа симметрии квадрата) изоморфна группе $\mathrm{Sym}(x_1x_2 + x_3x_4)$ (см. примеры 1.8 и 5.11).

Задача 5. Доказать, что $SL_2(\mathbf{Z}_2) \simeq S_3$.

Словарь сокращений английских слов, употребляемых в обозначениях

сокращение	от слова	перевод
area	area	площадь
arg	argument	аргумент
char	characteristic	характеристика
deg	degree	степень
det	determinant	определитель
dim	dimension	размерность
id	identity	тождество
Im	image	образ
Im	imaginary	мнимый
inf	infimum(лат.)	нижняя грань
Isom	isometry	изометрия
Ker	kernel	ядро
lim	limit	предел
mod	modulo(лат.)	по модулю
ord	order	порядок
Quot	quotient	частное
Re	real	действительный
rk	rank	ранг
sgn	signum(лат.)	знак
Sym	symmetry	симметрия
Tran	translation	перенос
vol	volume	объем