

А. И. КОСТРИКИН

ОСНОВЫ АЛГЕБРЫ



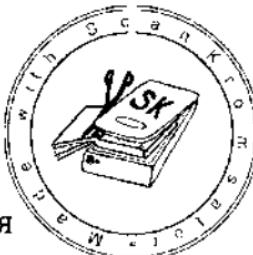
1
ОСНОВЫ АЛГЕБРЫ

А.И. Кострикин

ВВЕДЕНИЕ В АЛГЕБРУ

Часть I ОСНОВЫ АЛГЕБРЫ

*Рекомендовано Министерством общего и специального образования
Российской Федерации в качестве учебника
для студентов университетов, обучающихся по специальностям
“Математика” и “Прикладная математика”*



**МОСКВА
ФИЗИКО-МАТЕМАТИЧЕСКАЯ
ЛИТЕРАТУРА
2000**

УДК 512 (075.8)
ББК 22.143
К71

Кострикин А.И. Введение в алгебру. Часть I. Основы алгебры: Учебник для вузов. — М.: Физико-математическая литература, 2000. — 272 с. — ISBN 5-9221-0017-3.

Рассмотрены системы линейных уравнений, элементарная теория матриц, теория определителей, простейшие свойства групп, колец и полей, комплексные числа и корни многочленов. Помещено большое число упражнений различной степени трудности. Специальный раздел посвящён обсуждению некоторых нерешённых задач о многочленах.

Для студентов младших курсов университетов и вузов с повышенными требованиями по математике.

Ил. 28.

ТП-2000-I-73
ISBN 5-9221-0017-3 (Т. I)
ISBN 5-9221-0016-5



© А.И. Кострикин, 2000

ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ	7
СОВЕТЫ ЧИТАТЕЛЮ	10

ГЛАВА 1 ИСТОКИ АЛГЕБРЫ

§ 1. Алгебра вкратце	12
§ 2. Некоторые модельные задачи	15
1. Задача о разрешимости уравнений в радикалах (15). 2. Задача о состояниях многоатомной молекулы (17). 3. Задача о кодировании сообщения (18). 4. Задача о нагретой пластинке (18).	
§ 3. Системы линейных уравнений. Первые шаги	19
1. Терминология (20). 2. Эквивалентность линейных систем (21). 3. Приведение к ступенчатому виду (23). 4. Исследование системы линейных уравнений (24). 5. Отдельные замечания и примеры (26).	
§ 4. Определители небольших порядков	29
Упражнения (33).	
§ 5. Множества и отображения	33
1. Множества (33). 2. Отображения (35). Упражнения (40).	
§ 6. Отношения эквивалентности. Факторизация отображений . .	41
1. Бинарные отношения (41). 2. Отношение эквивалентности (41). 3. Факторизация отображений (42). 4. Упорядоченные множества (44). Упражнения (45).	
§ 7. Принцип математической индукции	46
Упражнения (50).	
§ 8. Перестановки	50
1. Стандартная запись перестановки (50). 2. Цикловая структура перестановки (52). 3. Знак перестановки (56). 4. Действие S_n на функциях (58). Упражнения (60).	
§ 9. Арифметика целых чисел	61
1. Основная теорема арифметики (61). 2. НОД и НОК в \mathbb{Z} (63). 3. Алгоритм деления в \mathbb{Z} (63). Упражнения (64).	

ГЛАВА 2 МАТРИЦЫ

§ 1. Векторные пространства строк и столбцов	66
1. Мотивировка (65). 2. Основные определения (66). 3. Линейные комбинации. Линейная оболочка (67). 4. Линейная зависимость (68). 5. Базис. Размерность (69). Упражнения (72).	
§ 2. Ранг матрицы	72
1. Возвращение к уравнениям (72). 2. Ранг матрицы (74). 3. Критерий совместности (76). Упражнения (77).	
§ 3. Линейные отображения. Действия с матрицами	78
1. Матрицы и отображения (78). 2. Произведение матриц (81). 3. Транспонирование матриц (83). 4. Ранг произведения матриц (84). 5. Квадратные матрицы (86). 6. Классы эквивалентных матриц (91). 7. Вычисление обратной матрицы (93). 8. Пространство решений (96). Упражнения (98).	

ГЛАВА 3 ОПРЕДЕЛИТЕЛИ

§ 1. Определители: построение и основные свойства	102
1. Геометрическая мотивировка (102). 2. Комбинаторно-аналитический подход (104). 3. Основные свойства определителей (105). Упражнения (112).	
§ 2. Дальнейшие свойства определителей	113
1. Разложение определителя по элементам столбца или строки (113). 2. Определители специальных матриц (116). Упражнения (119).	
§ 3. Применения определителей	121
1. Критерий невырожденности матрицы (121). 2. Формулы Крамера (123). 3. Метод окаймляющих миноров (125). Упражнения (128).	
§ 4. К построению теории определителей	130
1. Первое аксиоматическое построение (130). 2. Второе аксиоматическое построение (131). 3. Построение методом полной индукции (131). 4. Характеризация мультипликативными свойствами (131). Упражнения (133).	

ГЛАВА 4 ГРУППЫ. КОЛЬЦА. ПОЛЯ

§ 1. Множества с алгебраическими операциями	134
1. Бинарные операции (134). 2. Полугруппы и моноиды (135). 3. Обобщённая ассоциативность; степени (136). 4. Обратимые элементы (138). Упражнения (139).	

§ 2. Группы	139
1. Определение и примеры (139). 2. Циклические группы (142).	
3. Изоморфизмы (143). 4. Гомоморфизмы (147). 5. Словарик.	
Примеры (148). Упражнения (149).	
§ 3. Кольца и поля	151
1. Определение и общие свойства колец (151). 2. Сравнения.	
Кольцо классов вычетов (155). 3. Гомоморфизмы колец (156).	
4. Типы колец. Поля (157). 5. Характеристика поля (161). 6. За-	
мечание о линейных системах (163). Упражнения (165).	

ГЛАВА 5 КОМПЛЕКСНЫЕ ЧИСЛА И МНОГОЧЛЕНЫ

§ 1. Поле комплексных чисел	167
1. Вспомогательная конструкция (167). 2. Плоскость комплекс-	
ных чисел (168). 3. Геометрическое истолкование действий с	
комплексными числами (169). 4. Возведение в степень и извле-	
чение корня (173). 5. Теорема единственности (175). 6. Элемен-	
тарная геометрия комплексных чисел (176). Упражнения (179).	
§ 2. Кольцо многочленов	180
1. Многочлены от одной переменной (181). 2. Многочлены	
от многих переменных (185). 3. Алгоритм деления с остат-	
ком (187). Упражнения (188).	
§ 3. Разложение в кольце многочленов	190
1. Элементарные свойства делимости (190). 2. НОД и НОК	
в кольцах (192). 3. Факториальность евклидовых колец (194).	
4. Неприводимые многочлены (197). Упражнения (200).	
§ 4. Поле отношений	201
1. Построение поля отношений целостного кольца (201). 2. По-	
ле рациональных дробей (203). 3. Простейшие дроби (204).	
Упражнения (207).	

ГЛАВА 6 КОРНИ МНОГОЧЛЕНОВ

§ 1. Общие свойства корней	208
1. Корни и линейные множители (208). 2. Полиномиаль-	
ные функции (210). 3. Дифференцирования кольца многочле-	
нов (212). 4. Кратные множители (214). 5. Формулы Вие-	
та (216). Упражнения (218).	
§ 2. Симметрические многочлены	220
1. Кольцо симметрических многочленов (220). 2. Основная тео-	
рема о симметрических многочленах (221). 3. Метод неопре-	
делённых коэффициентов (224). 4. Дискриминант многочле-	
на (226). 5. Результант (228). Упражнения (231).	

§ 3. Алгебраическая замкнутость поля \mathbb{C}	232
1. Формулировка основной теоремы (232). 2. Доказательство основной теоремы (234). 3. Ещё одно доказательство основной теоремы (237).	
§ 4. Многочлены с вещественными коэффициентами	241
1. Разложение на неприводимые множители в $\mathbb{R}[X]$ (241). 2. Простейшие дроби над \mathbb{C} и \mathbb{R} (242). 3. Проблема локализации корней многочлена (244). 4. Вещественные многочлены с вещественными корнями (249). 5. Устойчивые многочлены (251). 6. Зависимость корней многочлена от коэффициентов (252). 7. Вычисление корней многочлена (254). 8. Рациональные корни целочисленных многочленов (255). Упражнения (257).	

**ПРИЛОЖЕНИЕ
НЕРЕШЁННЫЕ ЗАДАЧИ О МНОГОЧЛЕНАХ**

1. Проблема якобиана	259
2. Задача о дискриминанте	261
3. Задача о двух порождающих кольца многочленов	261
4. Задачи о критических точках и критических значениях	262
5. Задача о глобальной сходимости метода Ньютона	263
ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ	266

Алгебра щедра — зачастую
она даёт больше, чем у неё
спрашивают.

Даламбер

ПРЕДИСЛОВИЕ

Необходимость в едином изложении курсов алгебры, линейной алгебры и геометрии ощущалась давно. Во всяком случае, учебник "Введение в алгебру" (М., Наука, 1977) 22-летней давности с самого начала рассматривался лишь как первый шаг к интегрированному подходу. Алгебра — живая ветвь математики, обладающая значительной притягательной силой и основывающаяся на небольшом числе ясных, интуитивных начал. Смысл алгебраического понятия может иметь теоретико-числовую или геометрическую природу, а зачастую его корни лежат в вычислительных аспектах математики и в решении уравнений. Возникающие из такого исторического понимания принципы и требования, предъявляемые к современному университетскому учебнику по алгебре, стали общепринятыми. Вся трудность падает на реализацию более или менее известных идей. Естественная эволюция стандартных программ — то в сторону объединения курсов линейной алгебры и многомерной аналитической геометрии, то в сторону их разделения и вкрапления элементов теории чисел в курс алгебры — нашла отражение на страницах предлагаемого "Введения в алгебру", написанного на базе упомянутого одноимённого учебника, но сильно расширенного и разбитого для удобства читателя на три части. Само собой разумеется, что объединение этих частей заведомо содержит устойчивое ядро указанных курсов — тот минимум, которому должен удовлетворять всякий учебник. С другой стороны, распределение материала по частям соответствует реально сложившемуся за последние десятилетия порядку чтения курсов студентам механико-математического факультета МГУ: первый семестр — "Основы алгебры"; второй семестр — "Линейная алгебра и геометрия"; третий семестр — "Основные структуры алгебры" (алгебра на уровне элементарных, но довольно содержательных сведений об алгебраических системах, ставших принадлежностью каждого математика наших дней). В дальнейшем для удобства ссылок на эти книги будут использоваться соответственно сокращения [ВА I], [ВА II], [ВА III]. На этот порядок, равно как и на принцип подачи материала, наложили свой отпечаток не только здравый смысл, но и мудрый совет Горация: "Надо сегодня сказать лишь то, что уместно сегодня. Прочее всё отложить и сказать в подходящее время". Другими словами, мы придерживаемся концентрического стиля изложения, не боясь возвращаться к одной

и той же теме, к одному и тому же примеру много раз. Так, понятия группы, кольца, поля, изоморфизма возникают в [ВА I] и обсуждаются на уровне примеров, накапливаемых затем в [ВА II]; более основательное изучение этих понятий проводится лишь в [ВА III]. Абстрактные векторные пространства и линейные операторы на них исследуются в [ВА II], хотя их конкретные аналоги, сопровождающие теорию систем линейных уравнений, появляются на первых страницах настоящей книги. Разумеется, только читатель вправе судить, приближает ли такой подход то понимание предмета, о котором писал великий математик А. Пуанкаре в своем замечательном сочинении "Наука и метод" (гл. 2. Математические определения и преподавание). Реально читаемым курсом (три часа лекций в неделю в первом семестре, четыре — во втором и два — в третьем), по опыту самого автора, заведомо невозможно охватить весь материал учебника, да к этому и не следует стремиться. По своему замыслу он рассчитан на свободное творчество лектора (разумеется, в известных рамках). Хотелось бы рассматривать его также как своего рода справочник и как источник для дополнительного чтения студентами. Многообразие современной алгебры невозможно уложить в прокрустово ложе какого-либо "Введения в алгебру", однако импульсом к творческой работе мысли учебник послужить должен. Этому способствуют многочисленные упражнения, рассчитанные в какой-то мере на развитие основной темы. Кроме того, в каждой части имеется раздел, где перечислены, с необходимыми пояснениями, некоторые нерешённые или трудно решаемые задачи, непосредственно примыкающие (во всяком случае, по своей постановке) к программному материалу и лежащие почти что на поверхности. Вряд ли эти задачи станут предметом повального увлечения, но будет прекрасно, если в ком-то они зажгут огонёк поиска математической истины.

Несколько слов о [ВА I]. Эту книгу можно считать алгеброй в миниатюре. Фундаментальные понятия группы, кольца, поля, новые для большинства студентов, вводятся по возможности неформально и в минимальных дозах, хотя общее количество производных понятий получается довольно большим. Их не нужно запоминать: они станут привычными после самостоятельной работы над задачами и упражнениями. Для удобства выделяется несколько наиболее употребительных алгебраических систем таких, как группы $(\mathbb{Z}, +)$, S_n , A_n , GL_n , SL_n , кольцо многочленов, поля \mathbb{Q} , \mathbb{R} , \mathbb{C} и \mathbb{Z}_p , на фоне которых демонстрируется язык алгебры. По традиции и по соображениям преемственности между школой и вузом вначале излагается техника матриц и определителей, используемая для отыскания и исследования решений систем линейных уравнений. На этом пути естественным образом возникают и основные алгебраические структуры. Их

более обстоятельному изучению посвящена книга [ВА III], а пока в нашу задачу входит лишь накопление "живых" примеров.

Следует обратить особое внимание на книгу И.Р. Шафаревича [4] из дополнительного списка литературы, в которой развивается свежий и в высшей мере нетрадиционный взгляд на алгебру, а также на математику в целом.

Я благодарен всем читателям старого учебника "Введение в алгебру", его переводчикам на английский, болгарский, испанский, польский, французский, китайский языки и рецензентам, сообщившим свои замечания, а также членам кафедры высшей алгебры МГУ, где учебник продолжает подвергаться ежегодному испытанию.

Я рад выразить глубокую благодарность А.Я. Кострикиной, а также Н.К. Ильиной и В.В. Острику за неоценимую помощь при оформлении рукописи.

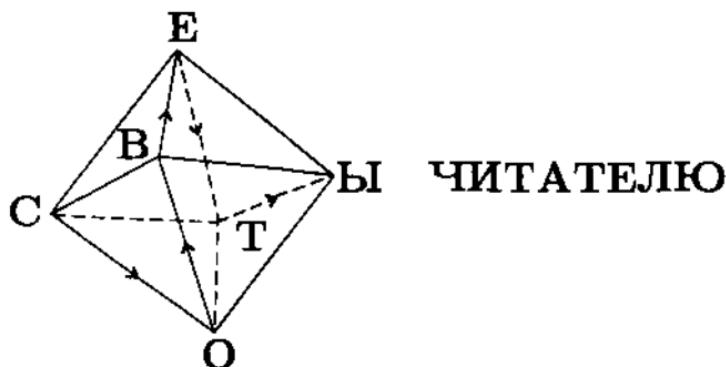
А.И. Кострикин

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

1. Сборник задач по алгебре / Под редакцией А.И. Кострикина. — М.: Факториал, 1995.
2. Курош А.Г. Курс высшей алгебры. — 10-е изд. — М.: Наука, 1971.
3. Фаддеев Д.К. Лекции по алгебре. — М.: Наука, 1984.
4. Шафаревич И.Р. Основные понятия алгебры. — М.: ВИНИТИ, 1986.

ГРЕЧЕСКИЙ АЛФАВИТ

α альфа	β бета	γ гамма	δ дельта	ϵ эпсилон	ζ дзета
η эта	θ тэта	ι йота	κ каппа	λ ламбда	μ мю
ν ню	ξ кси	\omicron омикрон	π пи	ρ ро	σ сигма
τ тау	υ ипсилон	ϕ фи	χ хи	ψ пси	ω омега



Согласно общему плану, изложенному в предисловии, схема зависимости глав в книге линейна. Фактически студенту-первокурснику полезно читать всё подряд, обращая особое внимание на многочисленные примеры и на упражнения, значительная часть которых обычно предлагается во время экзамена.

Искушённому читателю (скажем, преподавателю или студенту второго курса) будет нетрудно начинать чтение практически с любого места, естественно — при наличии готовности обращаться время от времени к определениям в предыдущих параграфах и главах. Не все новые понятия вводятся в абзацах, начинающихся словом “определение”. Подробное оглавление и предметный указатель помогут найти нужное место в книге.

Каждая глава разбита на несколько параграфов, а каждый параграф — на несколько пунктов с собственными названиями. Внутри параграфа теоремы, предложения, леммы, следствия имеют свою собственную нумерацию: теорема 1, теорема 2, ...; лемма 1, лемма 2, ... С этой примитивной, но весьма наглядной и экономной нумерацией при ссылках на утверждения из другого параграфа приходится писать “теорема i § j ” или даже “теорема i § j гл. k ”, однако это не вызывает неудобств.

Конец доказательства отмечается знаком \square .

Для сокращения используются простейшие логические символы. Знак импликации \Rightarrow в записи $A \Rightarrow B$ имеет простую смысловую нагрузку, что “ A влечёт B ” или “из A следует B ”, в то время как $A \Leftrightarrow B$ означает эквивалентность высказываний A и B , т.е. (... тогда и только тогда, когда ...). Квантор всеобщности \forall служит заменой выражения “для любого”. Остальные обозначения понятны из контекста.

Выше приведен целиком греческий алфавит с указанием произношения букв. Наблюдаемая здесь путаница досадна, поскольку буквы греческого алфавита весьма употребительны в математике.

Глава 1

ИСТОКИ АЛГЕБРЫ

С чего начинается алгебра? С некоторым приближением можно сказать, что истоки алгебры кроются в искусстве складывать, умножать и возводить в степень целые числа. Формальная, но далеко не очевидная и не однозначная замена чисел буквами позволяет действовать по аналогичным правилам в пределах гораздо более общих алгебраических систем. Стало быть, попытка ответить исчерпывающим образом на поставленный вопрос увела бы нас не только в глубь веков, в тайны зарождения математической мысли. Более трудная часть ответа была бы связана с описанием основных структур алгебры наших дней: групп, колец, полей, модулей и т. п. Но этому как раз и посвящена значительная часть книги, так что цель главы 1 кажется пока недостижимой.

К счастью, под абстрактной оболочкой большинства аксиоматических теорий алгебры скрываются вполне конкретные задачи теоретического или практического характера, решение которых служило в свое время счастливым, а иногда и неизбежным поводом к далеко идущим обобщениям. В свою очередь развитая теория давала импульс и средства к решению новых задач. Сложное взаимодействие теоретических и прикладных аспектов теории, присущее всей математике, в алгебре проступает весьма отчетливо и делает в какой-то мере оправданным принятый нами концентрический стиль изложения.

После кратких общих замечаний, связанных с историей предмета, мы сформулируем несколько задач, предваряющих содержание последующих глав. Одна из этих задач послужит отправной точкой для изучения систем линейных уравнений, теории матриц и теории определителей. Мы изложим метод Гаусса и получим первые сведения о решениях линейных систем.

Уже на этом этапе полезно ввести стандартные обозначения и терминологию, для чего мы дадим сжатый обзор теории множеств и отображений.

Будут введены важные понятия отношения эквивалентности и факторизации отображений. Далее в связи с разъяснением принципа математической индукции устанавливаются элементарные комбинаторные соотношения. Особое место отводится перестановкам, на которых базируется теория определителей.

Наконец, приводимые в последнем параграфе простейшие арифметические свойства системы целых чисел не только используются в дальнейшем, но и являются прототипом для построения аналогичной арифметики в более сложных алгебраических системах.

Материал этой главы не выходит далеко за пределы школьной программы. От читателя требуется лишь готовность встать на несколько более общую точку зрения. Чтение можно начинать с § 3.

§ 1. Алгебра вкратце

В наши дни не без основания говорят об “алгебраизации” математики, т.е. о проникновении идей и методов алгебры как в теоретические, так и в прикладные разделы математики. Такое положение вещей, ставшее совершенно отчётым к середине XX столетия, наблюдалось отнюдь не всегда. Как всякая область человеческой деятельности математика подвержена влиянию моды. Мода на алгебраические методы вызвана существом дела, хотя увлечение ею иногда переходит разумные границы. А так как алгебраическая оболочка, затмевающая содержание, не меньшая беда, чем элементарное забвение алгебры, то не случайно достоинством той или иной книги уже считается (вполне резонно) умение её автора избежать перегруженности алгебраическим формализмом.

Если отвлечься от крайностей, то алгебра издревле составляла существенную часть математики. То же самое следовало бы сказать и о геометрии, но мы скроемся за крылатой фразой Софи Жермен (XIX век): “Алгебра — не что иное, как записанная в символах геометрия, а геометрия — это просто алгебра, воплощённая в фигурах”. С тех пор положение изменилось, но, кажется, “признано, что “природа” математических объектов есть, в сущности, дело второстепенное и что довольно неважно, например, представили ли мы результат в виде теоремы “чистой” геометрии или при помощи аналитической геометрии в виде алгебраической теоремы” (Н. Бурбаки).

В соответствии с принципом “важны не математические объекты, а отношения между ними” алгебра определяется (несколько тавтологически и совершенно непонятно для непосвящённого) как наука об алгебраических операциях, выполняемых над элементами различных множеств. Самы алгебраические операции выросли из элементарной арифметики. В свою очередь на основе алгебраических соображений получаются наиболее естественные доказательства многих фактов из “высшей арифметики” — теории чисел.

Но значение алгебраических структур, т.е. множеств с алгебраическими операциями, далеко выходит за рамки теоретико-числовых применений. Многие математические объекты (топологические пространства, функции нескольких комплексных переменных и др.) изучаются путём построения надлежащих алгебраических структур, если и не адекватных изучаемым объектам, то во всяком случае отражающих их существенные стороны. Нечто подобное относится и к объектам реального мира.

Определённое мнение на этот счет было высказано более 45 лет назад одним из творцов квантовой механики П. Дираком: "Современная физика требует всё более абстрактной математики и развития ее основ. Так, неевклидова геометрия и некоммутативная алгебра, считавшиеся одно время просто плодом воображения или увлечения логическими рассуждениями, теперь признаны весьма необходимыми для описания общей картины физического мира".

Алгебраические средства весьма полезны при исследовании элементарных частиц в квантовой механике, свойств твёрдого тела и кристаллов (в этой связи особенно важна теория представлений групп), при анализе модельных задач экономики, при конструировании современных ЭВМ и т.д. и т.п.

В свою очередь алгебра питается живительными соками других дисциплин, в том числе математических. Так, гомологические методы алгебры выросли из недр топологии и алгебраической теории чисел. Не удивительно поэтому, что облик алгебры и точка зрения на алгебру менялись в разные эпохи. Мы не имеем возможности проследить подробно за этими изменениями не только из-за недостатка места, но главным образом потому, что описание истории предмета должно быть конкретным, — требование, которому можно удовлетворить лишь при основательном знакомстве с самим предметом. Ограничимся схематическим перечислением имён и периодов.

Древние цивилизации Вавилона и Египта. Греческая цивилизация "Арифметика" Диофанта (III в. н. э.).	Арифметические действия на множествах целых и рациональных положительных чисел. Алгебраические формулы в геометрических и астрономических расчётах. Формулировка задач на построение (об удвоении куба и трисекции угла), занимавших алгебраические умы в гораздо более позднее время.
Восточная цивилизация средних веков. Сочинение уроженца Хивы Мухаммеда ибн Муса ал-хорезми (ок. 825г.) "Хисаб ал-джабр ва-л-мукабала".	Алгебраические уравнения первой и второй степени. Возникновение самого термина "алгебра".
Эпоха Возрождения. С. Ферро (1465–1526) Н. Тарталья (1500–1557) И. Кардано (1501–1576) Л. Феррари (1522–1565) Ф. Виет (1540–1603) Р. Бомбелли (1530–1572)	Решение общих алгебраических уравнений третьей и четвертой степени.
XVII–XVIII вв. Р. Декарт (1596–1650) П. Ферма (1601–1665) И. Ньютона (1643–1727)	Создание современной алгебраической символики. Возникновение аналитической геометрии — прочного мостика между геометрией и алгеброй. Оживление деятельности в теории чисел.

Г. Лейбниц (1646–1716) Л. Эйлер (1707–1783) Ж. Даламбер (1717–1783) Ж.-Л. Лагранж (1736–1813) Г. Крамер (1704–1752) П. Лаплас (1749–1827) Вандермонд (1735–1796)	Развитие алгебры многочленов. Интенсивные поиски общих формул для решений алгебраических уравнений. Первые подходы к доказательству существования корня уравнения с числовыми коэффициентами. Начала теории определителей.
XIX в. – начало XX в. К. Ф. Гаусс (1777–1855) П. Дирихле (1805–1859) Э. Куммер (1810–1893) Л. Кронекер (1823–1891) Р. Дедекинд (1831–1916) Е. И. Золотарёв (1847–1878) Г. Ф. Вороной (1868–1908) А. А. Марков (1856–1922)	Доказательство основной теоремы о существовании корней уравнений с числовыми коэффициентами. Интенсивное развитие теории алгебраических чисел.
П. Л. Чебышев (1821–1894) Ш. Эрмит (1822–1901) Н. И. Лобачевский (1792–1856) А. Гурвиц (1859–1919)	Поиски методов приближенного решения алгебраических уравнений. Условия на коэффициенты, обеспечивающие заданное расположение корней.
А. Руффини (1765–1822) Н. Х. Абель (1802–1829) К. Якоби (1804–1851) Э. Галуа (1811–1832) Б. Риман (1826–1866) О. Коши (1789–1857) К. Жордан (1838–1922) Л. Сильвестр (1832–1918)	Решение проблемы о неразрешимости общих уравнений степени $n \geq 5$ в радикалах. Развитие теории алгебраических функций. Создание теории Галуа. Начала теории конечных групп, преимущественно на базе групп перестановок.
Г. Гассман (1809–1877) Д. Сильвестр (1814–1897) А. Кэли (1821–1895) У. Гамильтон (1805–1865) Дж. Буль (1815–1864) С. Ли (1842–1899) Г. Фробениус (1849–1918) Ж. Серре (1819–1885) М. Нёттер (1844–1922) Д. А. Граве (1863–1939) А. Пуанкаре (1854–1912) Ф. Клейн (1849–1925) У. Бернсайд (1852–1927) Ф. Э. Молин (1861–1941) И. Шур (1875–1941) Г. Вейль (1885–1955) Ф. Энрикаес (1871–1946)	Интенсивное развитие методов линейной алгебры. Возникновение, после открытия кватернионов, теории гиперкомплексных систем (такие системы теперь называются алгебрами). В частности, в связи с развитием теории непрерывных групп (групп Ли) были заложены основы теории алгебр Ли. Важными главами математики стали алгебраическая геометрия и теория инвариантов. В XIX в. математика ещё не достигла тонкой дифференциации, и многие крупные учёные творчески работали в различных её областях.
Дж. фон Нейман (1903–1957) Д. Гильберт (1862–1943) Э. Картан (1869–1951) К. Гензель (1861–1941) Э. Штейниц (1871–1928) Э. Нёттер (1882–1935)	Первая половина XX в. была отмечена коренной перестройкой всего здания математики. Алгебра, отказавшаяся от привилегии быть наукой об алгебраических уравнениях, решительно встала на аксиоматический и гораздо более

Э. Артин (1898–1962) Н. Бурбаки “Элементы математики”.	абстрактный путь развития.
Всё входит в язык теории колец, модулей, категорий, гомологий. Многие разрозненные теории оказались уложены в общую схему универсальной алгебры. На стыке алгебры и математической логики родилась теория моделей. Старые теории обновились, расширив область своих применений. Примером здесь могут служить современная алгебраическая геометрия, алгебраическая топология, алгебраическая K -теория, теория алгебраических групп. Несколько ярких взлётов испытала теория конечных групп.	

Вся алгебра находится сейчас в состоянии динамического развития. Крупные заслуги в этом принадлежат математикам России. Высокий уровень алгебраических исследований в нашей стране многим обязан таким учёным, как Н.Г. Чеботарёв (1894–1947), О.Ю. Шмидт (1891–1956), А.И. Мальцев (1909–1967), А.Г. Курош (1908–1971), П.С. Новиков (1901–1975), Д.К. Фаддеев (1907–1989).

§ 2. Некоторые модельные задачи

Формулируемые ниже четыре задачи стоят на разных уровнях. Первые три, сами по себе тоже неравнозначные, предназначены исключительно для мотивировки исследования полей разных типов, линейных пространств, групп и их представлений, т. е. тех алгебраических теорий, о которых речь будет ниже. “Решениям” этих задач посвящено много специальных монографий. Четвёртую задачу, предваряющую изучение линейных систем, полезно попробовать тут же решить, не заглядывая в следующий параграф, где приводится нужное рассуждение.

1. Задача о разрешимости уравнений в радикалах. Из элементарной алгебры известна формула

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \quad (1)$$

для решений x_1, x_2 квадратного уравнения $ax^2 + bx + c = 0$.

Уравнение третьей степени

$$x^3 + ax^2 + bx + c = 0$$

подстановкой $x \mapsto x - a/3$ приводится к виду $x^3 + px + q = 0$. Корни x_1, x_2, x_3 этого уравнения следующим образом выражаются через его коэффициенты. Если положить

$$D = -4p^3 - 27q^2, \quad \varepsilon = \frac{-1 + \sqrt{-3}}{2}, \quad (2)$$

$$u = \sqrt[3]{-\frac{27}{2}q + \frac{3}{2}\sqrt{-3D}}, \quad v = \sqrt[3]{-\frac{27}{2}q - \frac{3}{2}\sqrt{-3D}}$$

(кубические корни выбираются так, что $uv = -3p$), то можно показать, что

$$x_1 = \frac{1}{3}(u + v), \quad x_2 = \frac{1}{3}(\varepsilon^2 u + \varepsilon v), \quad x_3 = \frac{1}{3}(\varepsilon u + \varepsilon^2 v). \quad (3)$$

Формулы (2) и (3), называемые *формулами Кардано* (1545 г.) и ассоциирующиеся также с именами других итальянских математиков эпохи Возрождения (С. Ферро, Н. Тарталья), равно как и формула (1), справедливы при любых буквенных коэффициентах a, b, c, p, q , которым можно придать, например, произвольные рациональные значения. Аналогичные формулы были найдены для корней уравнения четвёртой степени, и на протяжении почти трёхсот лет предпринимались безуспешные попытки “решить в радикалах” общее уравнение пятой степени. Лишь в 1813 г. А. Руффини (в первом приближении) и в 1827 г. Н. Абель (независимо и совершенно строго) доказали теорему о том, что общее уравнение

$$x^n + a_1 x^{n-1} + \dots + a_n = 0$$

при $n > 4$ не разрешимо в радикалах.

Фундаментальное открытие в этой области было сделано двадцатилетним Эваристом Галуа в 1831 г. (оно стало известным лишь в 1846 г.), когда он дал универсальный критерий для разрешимости в радикалах любого (например, с рациональными коэффициентами), а не только общего уравнения степени n . Каждому многочлену (уравнению) степени n он сопоставил поле разложения и конечное семейство (мощности не более $n!$) автоморфизмов этого поля, называемое теперь *группой Галуа* поля (или исходного многочлена).

Хотя мы и лишены возможности останавливаться более подробно на теории Галуа, в [ВА III] будет выделен чисто внутренними свойствами специальный класс так называемых разрешимых групп. Оказывается, уравнение степени n с рациональными коэффициентами разрешимо в радикалах в точности тогда, когда разрешима соответствующая ему группа Галуа. Пусть, например, дано уравнение пятой степени

$$x^5 - ax - 1 = 0,$$

где a — некоторое целое число. Ему отвечает группа Галуа G_a , зависящая каким-то сложным образом от a ; G_0 — циклическая группа порядка 4 (а все циклические группы разрешимы по определению) и уравнение

$$x^5 - 1 = 0$$

разрешимо в радикалах. Напротив, G_1 имеет то же строение, что и симметрическая группа S_5 порядка 120, а последняя, как показано в [ВА III], неразрешима. Следовательно, неразрешимо в радикалах и уравнение

$$x^5 - x - 1 = 0.$$

Отметим в заключение, что для практических нужд возможность выразить корень алгебраического уравнения в явном виде через радикалы существенного значения не имеет; более актуальны разные приближённые методы вычисления корней. Но это обстоятельство не умаляет красоты достижения Галуа, оказавшего сильнейшее идеиное воздействие на последующее развитие математики. Начать с того, что именно Галуа заложил основы теории групп. Установленное Э. Галуа взаимно однозначное соответствие между подполями поля разложения и подгруппами его группы Галуа в XX веке обогатилось новыми абстрактными конструкциями и стало незаменимым средством исследования математических объектов.

2. Задача о состояниях многоатомной молекулы. Каждую молекулу можно рассматривать как систему частиц — атомных ядер (окружённых электронами). Если в начальный момент времени конфигурация системы близка к равновесной, то при определённых условиях частицы, входящие в систему, всегда будут оставаться вблизи положений равновесия и не будут приобретать больших скоростей. Движения такого типа называются *колебаниями относительно равновесной конфигурации*, а система — *устойчивой*.

Известно, что любое малое колебание молекулы вблизи положения устойчивого равновесия является суперпозицией так называемых нормальных колебаний. Во многих случаях удается определить потенциальную энергию молекулы и её нормальные частоты, принимая во внимание внутреннюю симметрию молекулы. Симметрия молекулярной структуры описывается точечной группой молекулы. Различные реализации этой конечной группы (её неприводимые представления) и связанные с этими реализациями функции на группе (характеры представлений) определяют параметры колебаний молекулы.

Например, молекуле воды H_2O (рис. 1) отвечает четверная группа Клейна (прямое произведение двух циклических групп второго порядка), а молекуле фосфора P_4 (рис. 2), имеющей вид правильного

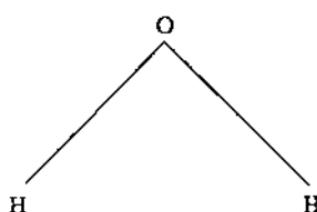


Рис. 1

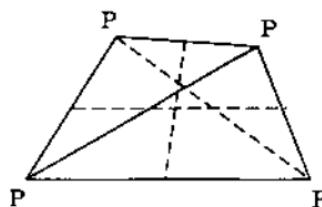


Рис. 2

тетраэдра, в вершинах которого расположены атомы фосфора, — симметрическая группа S_4 порядка 24. Неприводимые представления этих групп будут изучены в [ВА III].

В настоящее время развитие структурной теории молекул трудно себе представить без помощи теории групп. Гораздо более ранние применения теории групп относятся к кристаллографии. Ещё в 1891 г. великий русский кристаллограф Е. С. Фёдоров, а затем немецкий ученый А. Шёнфлис нашли 230 пространственных кристаллографических групп, описывающих все имеющиеся в природе симметрии кристаллов. С тех пор теория групп постоянно используется для исследования влияния симметрии на физические свойства кристалла.

3. Задача о кодировании сообщения. В конструировании автоматических систем связи, наземных или космических, обычно в качестве элементарного сообщения берётся упорядоченная последовательность — строка (или слово)

$$a = (a_1, a_2, \dots, a_n)$$

длины n , где $a_i = 0$ или $a_i = 1$. Так как обычные операции сложения и умножения по модулю 2 хорошо приспособлены для выполнения на электронной машине, а сами символы 0, 1 удобны для передачи в виде электрических сигналов (1 и 0 отличаются фазой разделённых по времени сигналов или их наличием и отсутствием), то неудивительно, что поле $GF(2)$ (см. § 3 гл. 4) — необходимый атрибут специалиста по переработке информации. Иногда удобно использовать в качестве a_i элементы других конечных полей.

С целью исключения влияния помех (атмосферных разрядов, космических шумов и т. д.), способных превратить 0 в 1 и обратно, приходится брать a достаточно длинным и использовать специальную систему кодирования — выбор такого подмножества (кода) S_0 передаваемых строк (кодовых слов) из всего их множества S , чтобы было возможно восстановить a по полученному искажённому слову a' при условии, что произошло не слишком много ошибок. Так возникают *коды, исправляющие ошибки*.

Алгебраическая теория кодирования, сильно развившаяся за последние годы и предложившая много остроумных методов кодирования, имеет дело в основном со специальными линейными кодами, когда выбор S_0 связан с построением специальных прямоугольных матриц и решением систем линейных уравнений, коэффициенты которых принадлежат заданному конечному полю. Простой пример такого кода будет приведён в гл. 4.

4. Задача о нагретой пластинке. Плоская прямоугольная пластинка с тремя отверстиями (рис. 3) используется в качестве клапана одного фантастического устройства для получения низких температур.

На клапан нанесена квадратная сетка (решётка). Её вершины, лежащие на четырёх контурах, называются *граничными*, а все остальные вершины — *внутренними*. Непосредственное измерение

показывает, что при любом нагревании или охлаждении температура в каждой внутренней вершине является средней арифметической величиной от температур ближайших четырёх вершин — неважно, граничных или внутренних. Ожидается, что детали устройства,

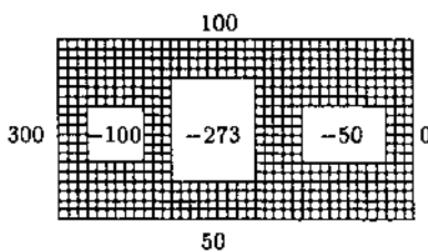


Рис. 3

соприкасаясь с различными участками контуров, сообщают соответствующим граничным точкам указанную на рис. 3 температуру. Возможно ли это, а если возможно, то однозначно ли при этом распределение температуры во внутренних точках?

§ 3. Системы линейных уравнений. Первые шаги

Линейные уравнения $ax = b$ и системы вида

$$\begin{aligned} ax + by &= e, \\ cx + dy &= f \end{aligned} \tag{1}$$

с вещественными (действительными) коэффициентами a, b, c, d, e, f “решаются” в средней школе. Наша цель — научиться оперировать с *системой линейных алгебраических уравнений* (или, коротко, с *линейной системой*) самого общего вида

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

Здесь m и n — произвольные целые положительные числа.

Будучи, казалось бы, чисто количественным, усложнение, получающееся при переходе от (1) к (2), имеет на самом деле принципиальное значение. Системы вида (2) встречаются буквально во всех разделах математики, и так называемые линейные методы, конечным продуктом которых часто являются решения линейных систем, составляют её наиболее развитую часть. Достаточно упомянуть, что теория систем вида (2) послужила в конце XIX века прототипом для создания теории интегральных уравнений, играющей исключительно

важную роль в механике и физике. Решение большого числа практических задач на ЭВМ также сводится к системам (2).

1. Терминология. Следует обратить внимание на весьма экономное и удобное обозначение коэффициентов системы (2): *коэффициент* a_{ij} (читается а-и-жи; например, a_{12} есть а-один-два, но никак не а-двенадцать) стоит в i -м уравнении при j -й неизвестной x_j . Число b_i называется *свободным членом* i -го уравнения. Система (2) называется *однородной*, если $b_i = 0$ для $i = 1, 2, \dots, m$. При любых b_i линейную систему

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= 0, \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &= 0, \\ \dots &\dots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n &= 0 \end{aligned} \tag{2^0}$$

называют *однородной системой, ассоциированной* с системой (2) или *ещё — приведённой системой* для системы (2). Коэффициенты при неизвестных составляют прямоугольную таблицу

$$\left| \begin{array}{cccc} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{array} \right|, \tag{3}$$

называемую *матрицей размера $m \times n$* ($m \times n$ -матрицей или *квадратной матрицей порядка n* при $m = n$) и сокращённо обозначаемую символом (a_{ij}) или просто буквой A . Естественно говорить об i -й строке $(a_{i1}, a_{i2}, \dots, a_{in})$ матрицы (3) и о j -м столбце

$$\left| \begin{array}{c} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{array} \right|,$$

который в дальнейшем, ради экономии места, будет изображаться строкой, заключённой в квадратные скобки: $[a_{1j}, a_{2j}, \dots, a_{mj}]$. В случае квадратной матрицы говорят *ещё о главной диагонали*, состоящей из элементов $a_{11}, a_{22}, \dots, a_{nn}$. Матрица (a_{ij}) , у которой все элементы вне главной диагонали равны нулю, обозначается иногда

$$\text{diag}(a_{11}, a_{22}, \dots, a_{nn})$$

и называется *диагональной* матрицей, а при $a_{11} = a_{22} = \dots = a_{nn} = a$ обозначается $\text{diag}_n(a)$ (*скалярная* матрица). Для матрицы $\text{diag}_n(1)$, называемой *единичной* матрицей, обычно используется обозначение E_n или E , когда размер матрицы фиксирован.

Наряду с матрицей (3) рассматривают и *расширенную* матрицу $(a_{ij} \mid b_i)$ системы (2), получаемую из (3) добавлением столбца $[b_1, b_2, \dots, b_m]$ свободных членов; для ясности он отделён от остальных столбцов вертикальной чертой.

Если каждое из уравнений системы (2) обращается в тождество после замены неизвестных x_i числами x_i^0 , то упорядоченный набор из n чисел $x_1^0, x_2^0, \dots, x_n^0$ называется *решением* системы (2), а x_i^0 — его i -й компонентой. Говорят также, что набор $x_1^0, x_2^0, \dots, x_n^0$ удовлетворяет всем уравнениям системы (2). Система, не имеющая ни одного решения, называется *несовместной*. Если же у системы есть решения, то она называется *совместной* и притом *определенной*, коль скоро решение единственно. Решений может быть и более одного, тогда система называется *неопределенной*. Совместна ли данная система линейных уравнений, а если совместна, то каковы все её решения — вот ближайшие вопросы, на которые нужно получить ответ.

Посмотрим ещё раз на задачу п. 4 § 2. Пронумеруем все внутренние точки пластиинки произвольным образом от 1 до 416 (именно столько их на рис. 3), добавим к ним 204 номера граничных точек и в соответствии с заданным правилом вычисления температуры t_i во внутренней точке с номером i составим 416 соотношений типа

$$t_e = \frac{t_a + t_b + t_c + t_d}{4}.$$

f	b	g
α		c
	e	
k	d	h

Рис. 4

Пусть, скажем, $a, b, c \leq 416$, $d > 416$. Тогда это соотношение можно переписать в виде линейного уравнения

$$-t_a - t_b - t_c + 4t_e = t_d$$

с правой частью $t_d = -273, -100, -50, 0, 50, 100, 300$ (возможны и другие варианты). Взятые вместе эти уравнения составят квадратную линейную систему вида (2) с $n = m = 416$. Коэффициенты при неизвестных t_i равны 0 (их большинство), -1 или 4 . Является ли эта система совместной и определённой?

Мы получили иную, математически точную формулировку задачи качественного характера. Вопрос о существовании и единственности весьма типичен для многих разделов математики, связанных с изучением физических явлений.

2. Эквивалентность линейных систем. Пусть нам дана ещё одна линейная система того же размера

$$\begin{aligned} a'_{11}x_1 + a'_{12}x_2 + \dots + a'_{1n}x_n &= b'_1, \\ a'_{21}x_1 + a'_{22}x_2 + \dots + a'_{2n}x_n &= b'_2, \\ \dots &\dots \\ a'_{m1}x_1 + a'_{m2}x_2 + \dots + a'_{mn}x_n &= b'_m. \end{aligned} \tag{2'}$$

Будем говорить, что система (2') получена из (2) при помощи *элементарного преобразования типа (I)*, если в системе (2) все уравнения, кроме i -го и k -го, остались прежними, а i -е и k -е уравнения поменялись местами. Если же в (2') все уравнения, кроме i -го, те же,

что и в (2), а i -е уравнение имеет вид

$$(a_{i1} + ca_{k1})x_1 + \dots + (a_{in} + ca_{kn})x_n = b_i + cb_k, \quad (*)$$

где c — какое-то число (т.е. $a'_{ij} = a_{ij} + ca^{kj}$, $b'_i = b_i + cb_k$), то полагаем, что к системе (2) применено элементарное преобразование типа (II).

Линейные системы (2) и (2') называются эквивалентными, если обе они либо несовместны, либо совместны и обладают одними и теми же решениями. Условившись обозначать эквивалентность систем (a) и (b) так: $(a) \sim (b)$, мы замечаем, что $(a) \sim (a)$, из $(a) \sim (b)$ следует $(b) \sim (a)$, а из $(a) \sim (b)$ и $(b) \sim (c)$ следует $(a) \sim (c)$.

Достаточный признак эквивалентности систем содержится в следующем утверждении.

Теорема 1. Две линейные системы эквивалентны, если одна получается из другой путём применения конечной последовательности элементарных преобразований.

Доказательство. Достаточно установить эквивалентность системы (2) и системы (2'), полученной из (2) путём применения одного элементарного преобразования.

Заметим, что система (2) получается из (2') также в результате применения одного элементарного преобразования, поскольку эти преобразования обратимы. Другими словами, в случае (I), переставив ещё раз местами уравнения с номерами i и k , мы вернёмся к первоначальной системе; аналогично в случае типа (II), прибавив к i -му уравнению в (2') k -е, умноженное на $(-c)$, мы получим i -е уравнение системы (2).

Докажем теперь, что любое решение (x_1^0, \dots, x_n^0) системы (2) является также решением системы (2'). Если было произведено элементарное преобразование типа (I), то сами уравнения вообще не изменились (изменился только порядок их записи). Поэтому числа $x_1^0, x_2^0, \dots, x_n^0$, удовлетворявшие им ранее, будут удовлетворять им и после преобразования. В случае элементарного преобразования типа (II) уравнения, кроме i -го, не изменились, и поэтому решение $(x_1^0, x_2^0, \dots, x_n^0)$ им по-прежнему удовлетворяет. Что касается i -го уравнения, то оно приобрело вид (*). Так как наше решение удовлетворяет i -му и k -му уравнениям системы (2), то

$$a_{i1}x_1^0 + \dots + a_{in}x_n^0 = b_i,$$

$$a_{k1}x_1^0 + \dots + a_{kn}x_n^0 = b_k.$$

Умножив обе части последнего тождества на c и прибавив его к первому, мы получим, группируя члены, тождество вида (*) с $x_i = x_i^0$.

В силу отмеченной выше обратимости элементарных преобразований проведённое рассуждение показывает также, что, обратно, любое решение системы (2') будет решением системы (2).

Осталось заметить, что несовместность одной системы влечёт несовместность другой (рассуждение от противного). \square

3. Приведение к ступенчатому виду. Путём последовательного применения элементарных преобразований можно перейти от заданной системы уравнений к системе более простого вида.

Во-первых, заметим, что среди коэффициентов a_{11} имеется хотя бы один, отличный от нуля. В противном случае не имело бы смысла упоминать о неизвестной x_1 . Если $a_{11} = 0$, то поменяем местами (преобразование типа (I)) первое уравнение с таким j -м, что $a_{j1} \neq 0$. Теперь коэффициент в первом уравнении при первой неизвестной отличен от нуля. Обозначим его через a'_{11} . Вычтем из i -го уравнения ($i = 2, 3, \dots, m$) новой системы первое уравнение, обе части которого умножены на такой коэффициент c_i , чтобы после вычитания коэффициент при x_1 обратился в 0 ($m - 1$ элементарных преобразований типа (II)). Очевидно, что для этого нужно положить $c_i = a_{i1}/a'_{11}$. В результате мы получим систему, в которой x_1 входит только в первое уравнение. При этом может оказаться, что вторая неизвестная также не входит во все уравнения с номером $i > 1$. Пусть x_k — неизвестная с наименьшим номером, которая входит в какое-нибудь уравнение, не считая первого. Мы получим систему

$$\begin{aligned} a'_{11}x_1 + \dots + a'_{1n}x_n &= b'_1, \\ a'_{2k}x_k + \dots + a'_{2n}x_n &= b'_2, \\ \dots & \\ a'_{mk}x_k + \dots + a'_{mn}x_n &= b'_m, \\ k > 1, \quad a'_{11} &\neq 0. \end{aligned}$$

Не обращая теперь внимания на первое уравнение, применим ко всем оставшимся те же рассуждения, что и ранее. После ряда элементарных преобразований исходная система примет вид

$$\begin{aligned} a''_{11}x_1 + \dots + a''_{1n}x_n &= b''_1, \\ a''_{2k}x_k + \dots + a''_{2n}x_n &= b''_2, \\ a''_{3l}x_l + \dots + a''_{3n}x_n &= b''_3, \\ \dots & \\ a''_{m1}x_1 + \dots + a''_{mn}x_n &= b''_m, \\ l > k > 1, \quad a''_{11} &\neq 0, \quad a''_{2k} \neq 0. \end{aligned}$$

Разумеется, здесь $a''_{1j} = a'_{1j}$, $b''_1 = b'_1$, ибо первое уравнение не затрагивалось. Будем применять этот процесс до тех пор, пока возможно. Ясно, что мы будем вынуждены остановиться, когда станут равными нулю не только коэффициенты при очередной неизвестной (скажем,

s -й), но и коэффициенты при всех следующих неизвестных вплоть до n -й. При этом система (2) примет вид

Здесь

$$\bar{a}_{11}\bar{a}_{2k}\bar{a}_{3l}\dots\bar{a}_{rs} \neq 0, \quad \quad 1 < k < l < \dots < s.$$

Может оказаться, что $\tau = m$, и поэтому уравнений вида $0 = \bar{b}_t$ в системе (4) не будет. Про систему уравнений вида (4) говорят, что она имеет *ступенчатый* вид.

Этот термин не является общепринятым; здесь можно было бы говорить о *трапециевидном* или о *квазитреугольном* виде и т.п., что не так уж существенно.

Теорема 2. Всякая система линейных уравнений эквивалентна системе, имеющей ступенчатый вид.

Доказательство непосредственно вытекает из предшествующих рассуждений.

Элементарные преобразования иногда удобно производить не над системой, а над её расширенной матрицей $(a_{ij} | b_i)$. Точно так же, как и теорема 2, доказывается

Теорема 2'. Всякую матрицу можно при помощи элементарных преобразований привести к ступенчатому виду.

4. Исследование системы линейных уравнений. Ввиду теорем 1 и 2 вопросы совместности и определённости достаточно исследовать для систем ступенчатого вида (4).

Начнём с вопроса о совместности. Очевидно, что если система (4) содержит уравнение вида $0 = \bar{b}_t$ с $\bar{b}_t \neq 0$, то эта система несовместна, так как равенство $0 = \bar{b}_t$ нельзя удовлетворить никакими значениями для неизвестных. Докажем, что если таких уравнений в системе (4) нет, то эта система совместна.

Итак, пусть $b_t = 0$ при $t > r$. Назовём неизвестные $x_1, x_k, x_l, \dots, x_s$, с которых начинаются первое, второе, \dots , r -е уравнения, главными, а остальные неизвестные, если таковые имеются, — свободными. Главных неизвестных по определению всего r .

Придадим свободным неизвестным произвольные значения и подставим их в уравнения системы (4). Тогда для x_1 получится одно (r -е)

уравнение вида $ax_s = b$ с $a = \bar{a}_{rs} \neq 0$, которое имеет единственное решение. Подставляя найденное значение $x_s = x_s^0$ в первые $r - 1$ уравнений и поднимаясь так снизу вверх по системе (4), мы убедимся в том, что значения для главных неизвестных определяются однозначно при любых заданных значениях для свободных неизвестных.

Нами доказана

Теорема 3. Для совместности системы линейных уравнений необходимо и достаточно, чтобы после приведения к ступенчатому виду в ней не оказалось уравнений вида $0 = \bar{b}_t$ с $\bar{b}_t \neq 0$. Если это условие выполнено, то свободным неизвестным можно придать произвольные значения; главные неизвестные (при заданных значениях для свободных) однозначно определяются из системы.

Выясним теперь, когда система будет определённой, в предположении, что введённое нами условие совместности выполнено. Если в системе (4) имеются свободные неизвестные, то система заведомо неопределёна: мы можем придать свободным неизвестным любые значения, выражая через них (по теореме 3) главные неизвестные. Если же свободных неизвестных нет, и все неизвестные, стало быть, главные, то по теореме 3 они определяются из системы однозначно, так что система является определённой.

Остается заметить, что отсутствие свободных неизвестных равносильно условию $r = n$.

Мы доказали следующее утверждение.

Теорема 4. Совместная линейная система (2) является определённой тогда и только тогда, когда в полученной из неё ступенчатой системе (4) выполняется равенство $r = n$.

При $m = n$ линейную систему, приведённую к ступенчатому виду, можно записать ещё так (треугольный вид):

$$\begin{aligned} \bar{a}_{11}x_1 + \bar{a}_{12}x_2 + \dots + \bar{a}_{1n}x_n &= \bar{b}_1, \\ \bar{a}_{22}x_2 + \dots + \bar{a}_{2n}x_n &= \bar{b}_2, \\ \dots & \\ \bar{a}_{nn}x_n &= \bar{b}_n, \end{aligned} \tag{5}$$

если не заботиться о том, чтобы выполнялось условие $\bar{a}_{ii} \neq 0$ для всех i . Действительно, запись (5) означает, что в системе k -е уравнение не содержит неизвестных x_i с $i < k$, а это условие заведомо выполнено для систем ступенчатого вида.

Заметим на будущее, что матрица (a_{ij}) с элементами $a_{ij} = 0$ при $i > j$ называется *верхней треугольной*. Аналогично определяется *нижняя треугольная* матрица.

Из теорем 3 и 4 вытекает

Следствие 1. Линейная система (2) в случае $m = n$ является совместной и определённой тогда и только тогда, когда после

приведения к ступенчатому виду получится система (5) с $\bar{a}_{11}\bar{a}_{22} \dots \bar{a}_{nn} \neq 0$.

Обратим внимание на тот факт, что это условие не зависит от правых частей системы. Поэтому при $m = n$ система (2) тогда и только тогда совместна и определённа, когда это верно для ассоциированной с ней однородной системы (2⁰). Но однородная система всегда совместна: она имеет, например, нулевое решение

$$x_1^0 = 0, \dots, x_n^0 = 0.$$

Условие $\bar{a}_{11}\bar{a}_{22} \dots \bar{a}_{nn} \neq 0$ означает, что однородная система обладает только нулевым решением. Мы приходим к иной форме следствия 1, не связанной с её ступенчатым видом.

Следствие 1'. *Линейная система (2) в случае $m = n$ является совместной и определённой тогда и только тогда, когда ассоциированная с ней однородная система (2⁰) имеет только нулевое решение.*

Специального внимания заслуживает случай $n > m$.

Следствие 2. *Совместная система (2) при $n > m$ является неопределенной. В частности, однородная система при $n > m$ всегда имеет ненулевое решение.*

Доказательство. Действительно, в любом случае $r \leq m$, поскольку в системе (4) не больше уравнений, чем в системе (2) (уравнения с тождественно равными нулю левыми и правыми частями отброшены). Поэтому неравенство $n > m$ влечёт $n > r$, что по теореме 4 означает неопределенность системы (2). Остаётся заметить, что неопределенность однородной системы равносильна существованию у неё ненулевого решения. \square

Часть полученных нами результатов отражена в следующей таблице.

Тип линейной системы				
общая	однородная	$n > m$	$n > m$	
Число решений	0, 1, ∞	1, ∞	0, ∞	∞

5. Отдельные замечания и примеры. Изложенный нами метод решения систем линейных уравнений называется *методом Гаусса* или *методом последовательного исключения неизвестных*. Весьма удобный при небольших n , он годится и для осуществления на ЭВМ, хотя по разным причинам более практическими зачастую оказываются другие способы решения, например итерационные. Это относится в особенности к тому случаю, когда коэффициенты даны, а решения ищутся с определённой степенью точности. В теоретических исследованиях, однако, первостепенное значение приобретают формулировка условий совместности или определённости линейной

системы, а также нахождение общих формул для решений в терминах коэффициентов и свободных членов без приведения системы к ступенчатому виду. В какой-то мере одному из этих требований отвечает следствие 1'.

Пример 1. Вновь обратимся к задаче о нагретой пластинке из § 2. Как мы видели в п. 1, интересующий нас вопрос выражается в свойствах вполне конкретной линейной системы (для определённости назовём ее НП) с довольно большим числом неизвестных t_i . Следуя критерию, сформулированному в следствии 1', рассмотрим однородную линейную систему (ОНП), ассоциированную с НП. Другими словами, температура всех граничных точек пластиинки принимается теперь равной нулю. Пусть e — номер внутренней точки с максимальным значением $|t_e|$. Тогда из условия

$$t_e = \frac{t_a + t_b + t_c + t_d}{4}$$

вытекает $|t_e| = |t_a| = |t_b| = |t_c| = |t_d|$. Сдвигаясь на один шаг решётки в любом из четырёх направлений, мы будем проходить через точки с тем же значением $|t_i| = |t_e|$, пока не достигнем граничной точки с нулевой температурой. Значит, $|t_e| = 0$, а поэтому и $t_i = 0$ для всех i . Итак, система ОНП имеет лишь нулевое решение, и, стало быть, НП — совместная и определённая линейная система. Задача о нагретой пластиинке в первоначальной её постановке тем самым решена.

Пример 2. Даны линейная система

$$\begin{aligned} x_1 &= 1, \\ x_2 &= 1, \\ -x_1 - x_2 + x_3 &= 0, \\ \dots & \\ -x_{n-2} - x_{n-1} + x_n &= 0. \end{aligned}$$

Очевидно, что это — совместная определённая система, уже приведённая к ступенчатому (треугольному) виду. Только, решая ее, нужно двигаться не снизу вверх, а сверху вниз. Решением является по определению последовательность первых n чисел Фибоначчи f_1, f_2, \dots, f_n . Эти числа связаны с одним ботаническим явлением, так называемым *филотаксисом* (расположением листьев на растениях). Однако при $n = 1000$ или даже при произвольном n хотелось бы указать общее выражение (аналитическую формулу) для n -го числа Фибоначчи. Вы можете возразить, сказав, что у вас хватит терпения указать и f_{1000} , следуя индуктивному определению этих чисел. Но это не будет математическим решением вопроса. В гл. 2 и гл. 3 мы укажем два выражения для f_n , хотя, конечно, эту конкретную задачу можно решить и более прямыми средствами.

Замечание 1. Иногда бывает удобнее находить решения линейной системы, не приводя её к ступенчатому виду. Это особенно относится к тому случаю, когда матрица системы содержит много нулей. Небольшая практика здесь предпочтительнее длинных объяснений.

Замечание 2. Какое количество Γ_n арифметических операций необходимо выполнить для решения системы n линейных уравнений с n неизвестными методом Гаусса? Это не праздный вопрос, поскольку ставшему обыденным в наши дни использованию ЭВМ при больших n должны предшествовать априорные оценки машинного времени, требуемого для решения задачи.

Так как умножение двух чисел более трудоёмко, чем сложение, то рекомендуется подсчитывать только количество умножений и, разумеется, делений, называемых далее просто операциями. Без ограничения общности можно предполагать, что решение линейной системы единственno, т.е. все неизвестные — главные. Правые части уравнений пока игнорируем. Тогда для исключения неизвестной x_1 из уравнения с номером $i > 1$ нужно заготовить число $l_i = a_{1i}/a_{11}$ (одно деление) и вычислить ещё $n - 1$ произведений $l_i a_{ij}$, $j = 2, 3, \dots, n$, т.е. всего требуется n операций. Процедурой вычитания из i -го уравнения первого, умноженного на l_i , мы условились пренебречь. Так как $i = 2, 3, \dots, n$, то для исключения x_1 понадобилось $n(n - 1)$ операций. На втором шаге, когда мы имеем дело с системой порядка $n - 1$, понадобится $(n - 1)(n - 2)$ операций, на третьем — соответственно $(n - 2)(n - 3)$ и т.д. Общее число операций для приведения левых частей системы к треугольному виду (5) равно сумме

$$\Gamma(n) = n(n - 1) + (n - 1)(n - 2) + \dots + 1(1 - 1).$$

Нетрудно убедиться (докажите сами или загляните в § 7), что

$$\Gamma(n) = \frac{n^3 - n}{3}.$$

Процесс нахождения компонент $x_n^0, x_{n-1}^0, \dots, x_1^0$ решения (движение снизу вверх по системе (5)) требует всего

$$1 + 2 + 3 + \dots + n = \frac{n(n + 1)}{2}$$

операций. При больших n это не внесёт существенного вклада в общую сумму операций. Итак, вполне удовлетворительной оценкой числа операций является (гауссова) величина $\Gamma_n = n^3/3$.

В 1969 г. Штрассеном разработан метод (подробности см. в [ВА II]), требующий только

$$\mathbb{W}_n = C \cdot n^{\log_2 7} \approx C \cdot n^{2.81}$$

операций, — значительный выигрыш при очень больших n , полученный, правда, за счёт увеличения числа операций сложения. Но константа C в \mathbb{W}_n чрезвычайно велика, а программа реализации логически сложна, поэтому речь идёт скорее о выигрыше в теоретическом плане.

Оба упомянутых нами метода являются типичными математическими алгоритмами, приспособленными для решения массовых задач. Позднее мы встретимся с другими примерами алгоритмов. Их роль в наш век сплошной компьютеризации весьма велика. При этом важны не только сами алгоритмы, но и оценки их сложности.

§ 4. Определители небольших порядков

Излагая метод Гаусса, мы не слишком заботились о значениях коэффициентов при главных неизвестных. Важно было лишь то, что эти коэффициенты отличны от нуля. Проведём теперь более аккуратно процесс исключения неизвестных хотя бы в случае квадратных линейных систем небольших размеров. Это даст нам пищу для размышлений и исходный материал для построения общей теории определителей в гл. 3.

Как и в § 3, рассмотрим систему двух уравнений с двумя неизвестными

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 &= b_1, \\ a_{21}x_1 + a_{22}x_2 &= b_2 \end{aligned} \quad (1)$$

и постараемся найти общие формулы для компонент x_1^0, x_2^0 её решения.

Назовём *определителем* матрицы $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$ выражение $a_{11}a_{22} - a_{21}a_{12}$ и обозначим его

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}.$$

Тем самым квадратной матрице сопоставляется число

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{21}a_{12}. \quad (2)$$

Если мы попытаемся исключить x_2 из системы (1), умножив первое уравнение на a_{22} и прибавив к нему второе, умноженное на $-a_{12}$, то получим

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} x_1 = b_1 a_{22} - b_2 a_{12}.$$

Правую часть также можно рассматривать как определитель матрицы $\begin{vmatrix} b_1 & a_{12} \\ b_2 & a_{22} \end{vmatrix}$. Предположим, что $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \neq 0$. Тогда мы имеем

$$x_1 = \frac{\begin{vmatrix} b_1 & a_{12} \\ b_2 & a_{22} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}}, \quad x_2 = \frac{\begin{vmatrix} a_{11} & b_1 \\ a_{21} & b_2 \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}}. \quad (3)$$

Имея формулы для решения системы двух линейных уравнений с двумя неизвестными, мы можем решать и некоторые другие системы (решать системы — значит находить их решения). Рассмотрим,

например, систему двух однородных уравнений с тремя неизвестными

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 &= 0, \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 &= 0. \end{aligned} \quad (4)$$

Нас интересует ненулевое решение этой системы, так что хотя бы одно из x_i не равно нулю. Пусть, например, $x_3 \neq 0$. Разделив обе части на $-x_3$ и положив $y_1 = -x_1/x_3$, $y_2 = -x_2/x_3$, запишем систему (4) в том же виде

$$a_{11}y_1 + a_{12}y_2 = a_{13},$$

$$a_{21}y_1 + a_{22}y_2 = a_{23},$$

что и (1). При предположении $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \neq 0$ формулы (3) дают

$$y_1 = -\frac{x_1}{x_3} = \frac{\begin{vmatrix} a_{13} & a_{12} \\ a_{23} & a_{22} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}}, \quad y_2 = -\frac{x_2}{x_3} = \frac{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{23} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}}.$$

Неудивительно, что мы нашли из системы (4) не сами x_1, x_2, x_3 , а только их отношения: из однородности системы легко следует, что если (x_1^0, x_2^0, x_3^0) — решение и c — любое число, то (cx_1^0, cx_2^0, cx_3^0) тоже будет решением. Поэтому мы можем положить

$$x_1 = -\begin{vmatrix} a_{13} & a_{12} \\ a_{23} & a_{22} \end{vmatrix}, \quad x_2 = -\begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix}, \quad x_3 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \quad (5)$$

и сказать, что любое решение получается из указанного умножением всех x_i на некоторое число c . Чтобы придать ответу несколько более симметричный вид, заметим, что всегда

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = -\begin{vmatrix} b & a \\ d & c \end{vmatrix},$$

как это непосредственно видно из формулы (2). Поэтому (5) можно записать в виде

$$x_1 = \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix}, \quad x_2 = -\begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix}, \quad x_3 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}. \quad (6)$$

Эти формулы выведены в предположении, что $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \neq 0$. Нетрудно проверить, что доказанное утверждение верно, если хоть один из входящих в выражения (6) определителей отличен от нуля. Если же все три определителя равны нулю, то, конечно, формулы (6) дают решение (а именно нулевое), но мы не можем утверждать, что

все решения получаются из него умножением на число (рассмотрите систему, состоящую из двух совпадающих уравнений $x_1 + x_2 + x_3 = 0$).

Перейдём теперь к случаю системы трёх уравнений с тремя неизвестными

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = 0,$$

$$a_{21}x_1 + a_{22}x_2 + a_{23}x_3 = 0,$$

$$a_{31}x_1 + a_{32}x_2 + a_{33}x_3 = 0.$$

Мы хотим исключить из этой системы x_2 и x_3 , чтобы получить значение для x_1 . С этой целью умножим первое уравнение на c_1 , второе на c_2 , третье на c_3 и сложим их. Подберём c_1, c_2, c_3 так, чтобы в получившемся уравнении члены с x_2 и x_3 обратились в нуль. Приравнивая нулю соответствующие коэффициенты, мы получим для c_1, c_2 и c_3 систему уравнений

$$a_{12}c_1 + a_{22}c_2 + a_{32}c_3 = 0,$$

$$a_{13}c_1 + a_{23}c_2 + a_{33}c_3 = 0,$$

относящуюся к тому же типу, что и (4). Поэтому можно взять

$$c_1 = \begin{vmatrix} a_{22} & a_{32} \\ a_{23} & a_{33} \end{vmatrix}, \quad c_2 = -\begin{vmatrix} a_{12} & a_{32} \\ a_{13} & a_{33} \end{vmatrix}, \quad c_3 = \begin{vmatrix} a_{12} & a_{22} \\ a_{13} & a_{23} \end{vmatrix}.$$

После очевидных изменений мы получаем для x_1 выражение

$$\left(\begin{array}{c|cc} a_{11} & a_{22} & a_{23} \\ & a_{32} & a_{33} \end{array} \right) - a_{21} \left(\begin{array}{c|cc} a_{12} & a_{13} \\ a_{32} & a_{33} \end{array} \right) + a_{31} \left(\begin{array}{c|cc} a_{12} & a_{13} \\ a_{22} & a_{23} \end{array} \right) x_1 = \\ = b_1 \left(\begin{array}{c|cc} a_{22} & a_{23} \\ a_{32} & a_{33} \end{array} \right) - b_2 \left(\begin{array}{c|cc} a_{12} & a_{13} \\ a_{32} & a_{33} \end{array} \right) + b_3 \left(\begin{array}{c|cc} a_{12} & a_{13} \\ a_{22} & a_{23} \end{array} \right). \quad (7)$$

Коэффициент при x_1 называется *определителем* матрицы

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

и обозначается

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}.$$

Таким образом, за определитель третьего порядка мы берём выра-

жение

$$\begin{aligned}
 & \left| \begin{array}{ccc} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{array} \right| = \\
 & = a_{11} \left| \begin{array}{cc} a_{22} & a_{23} \\ a_{32} & a_{33} \end{array} \right| - a_{21} \left| \begin{array}{cc} a_{12} & a_{13} \\ a_{32} & a_{33} \end{array} \right| + a_{31} \left| \begin{array}{cc} a_{12} & a_{13} \\ a_{22} & a_{23} \end{array} \right| = \\
 & = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}, \tag{8}
 \end{aligned}$$

задаваемое при помощи определителей второго порядка. Легко заметить, что правая часть в равенстве (7) получается из коэффициента при x_1 заменой a_{11} на b_1 , a_{21} на b_2 и a_{31} на b_3 . Поэтому равенство (7) можно записать в виде

$$\left| \begin{array}{ccc} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{array} \right| x_1 = \left| \begin{array}{ccc} b_1 & a_{12} & a_{13} \\ b_2 & a_{22} & a_{23} \\ b_3 & a_{32} & a_{33} \end{array} \right|.$$

Предположим, что коэффициент при x_1 отличен от нуля. Тогда, проведя аналогичные вычисления для x_2 и x_3 , мы выразим соответственно x_1, x_2, x_3 в виде

$$\left| \begin{array}{ccc} b_1 & a_{12} & a_{13} \\ b_2 & a_{22} & a_{23} \\ b_3 & a_{32} & a_{33} \end{array} \right|, \quad \left| \begin{array}{ccc} a_{11} & b_1 & a_{13} \\ a_{21} & b_2 & a_{23} \\ a_{31} & b_3 & a_{33} \end{array} \right|, \quad \left| \begin{array}{ccc} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \\ a_{31} & a_{32} & b_3 \end{array} \right|, \quad \left| \begin{array}{ccc} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{array} \right|. \tag{9}$$

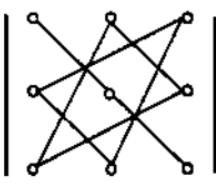
Очевидно, что те же самые рассуждения применимы к системе из четырёх, пяти и т.д. уравнений с тем же числом неизвестных. Для этого нам надо сначала вывести формулы, аналогичные (6), для решений однородной системы трёх уравнений с четырьмя неизвестными; потом в системе четырёх уравнений с четырьмя неизвестными исключить x_2, x_3, x_4 , умножая уравнения на c_1, c_2, c_3, c_4 и складывая их. Мы найдём значения c_i ($i = 1, 2, 3, 4$) из системы трёх однородных уравнений.

Коэффициент, получающийся при x_1 и строящийся из определителей третьего порядка по образцу (8), мы назовём *определителем четвёртого порядка*.

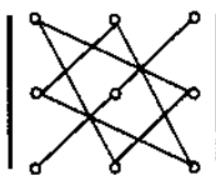
Проводя те же рассуждения с x_2, x_3, x_4 , мы найдём для x_i формулы, аналогичные (9). Так можно продолжать неограниченно. Уверенность в том, что мы когда-нибудь достигнем цели, нам даёт общий принцип, широко используемый в математике, а именно принцип математической индукции (см. § 7).

УПРАЖНЕНИЯ

1. Формулу (8) легче запомнить, если воспользоваться следующим наглядным правилом знаков для выписывания произведений, входящих в разложение определителя третьего порядка:



+



-

Найти аналогичное правило для определителя четвёртого порядка.

2. Показать, что все шесть членов в разложении определителя третьего порядка не могут быть одновременно положительными.

3. Проверить, что

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{vmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{vmatrix}, \quad \begin{vmatrix} 0 & a & b \\ -a & 0 & c \\ -b & -c & 0 \end{vmatrix} = 0.$$

§ 5. Множества и отображения

В предыдущих двух параграфах мы встретились с множествами элементов разной природы, равно как и с отображениями множеств. Множество решений данной системы линейных уравнений или правило, ставящее в соответствие каждой матрице второго порядка её определитель, — это лишь частные проявления того круга формальных понятий, знакомство с которым (хотя бы на интуитивном уровне) полезно для дальнейшего.

1. Множества. Под **множеством**, понимают любую совокупность объектов, называемых **элементами** множества.

Множества с конечным числом различных элементов могут быть описаны путём явного перечисления всех их элементов; обычно эти элементы заключаются в фигурные скобки. Например, $\{1, 2, 4, 8\}$ — множество степеней двойки, заключённых между 1 и 10. Как правило, множество обозначается прописной буквой какого-либо алфавита, а его элементы — строчными буквами того же или другого алфавита.

Для некоторых особенно важных множеств приняты стандартные обозначения, которых стоит придерживаться. Так, буквами N , Z , Q , R обозначают соответственно множество положительных целых чисел (натуральные числа), множество всех целых чисел, множество рациональных чисел и множество вещественных чисел.

При заданном множестве S включение $a \in S$ указывает на то, что a — элемент множества S ; в противном случае пишут $a \notin S$.

Говорят, что S — подмножество множества T и записывают это $S \subset T$ (S содержится в T), когда имеет место импликация

$$\forall x \quad x \in S \Rightarrow x \in T.$$

(По поводу обозначений см. раздел “Советы читателю”.)

Два множества S и T совпадают (или равны), если у них одни и те же элементы. Символически это записывается так:

$$S = T \Leftrightarrow S \subset T, T \subset S$$

(\Leftrightarrow — “тогда и только тогда, когда” или “влечёт в обе стороны”).

Пустое множество \emptyset , совсем не содержащее элементов, по определению входит в число подмножеств любого множества. Если $S \subset T$, но $S \neq \emptyset$ и $S \neq T$, то S — собственное подмножество в T . Для выделения подмножества $S \subset T$ часто используют какое-либо свойство, присущее только элементам из S . Например,

$$\{n \in \mathbb{Z} \mid n = 2m \text{ для некоторого } m \in \mathbb{Z}\}$$

— множество всех чётных целых чисел, а

$$\mathbb{N} = \{n \in \mathbb{Z} \mid n > 0\}$$

— множество натуральных чисел.

Под *пересечением* двух множеств S и T понимают множество

$$S \cap T = \{x \mid x \in S, x \in T\},$$

а под их *объединением* — множество

$$S \cup T = \{x \mid x \in S \text{ или } x \in T\}.$$

Пересечение $S \cap T$ может быть пустым множеством. Тогда говорят, что S и T — непересекающиеся множества. Операции пересечения и объединения удовлетворяют тождествам типа

$$R \cap (S \cup T) = (R \cap S) \cup (R \cap T),$$

$$R \cup (S \cap T) = (R \cup S) \cap (R \cup T),$$

проверку которых мы оставляем читателю в качестве упражнения. Рис. 5 поможет провести несложное рассуждение.



Рис. 5

Разностью $S \setminus T$ множеств S и T называется совокупность тех элементов из S , которые не содержатся в T . При этом, вообще говоря, не предполагается, что $T \subset S$. Вместо $S \setminus T$ пишут также $S - T$.

Если T — подмножество в S , то запись $S \setminus T$ обозначает ещё *дополнение к T в S* . Положив $R = S \setminus T$, будем иметь $R \cap T = \emptyset$, $R \cup T = S$. Обратим внимание на соответствие между операциями пересечения, объединения, дополнения и логическими связками “и”, “или”, “нет”.

Пусть далее X и Y — произвольные множества. Пару (x, y) элементов $x \in X$, $y \in Y$, взятых в данном порядке, будем называть *упорядоченной парой*, считая при этом, что $(x_1, y_1) = (x_2, y_2)$ тогда и только тогда, когда $x_1 = x_2$, $y_1 = y_2$.

Декартовым произведением двух множеств X и Y называется множество всех упорядоченных пар (x, y) :

$$X \times Y = \{(x, y) \mid x \in X, y \in Y\}.$$

Пусть, например, \mathbb{R} — множество всех вещественных чисел. Тогда *декартов квадрат* $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ есть просто множество всех декартовых координат точек на плоскости относительно заданных координатных осей. Аналогичным образом можно было бы ввести декартово произведение $X_1 \times X_2 \times X_3$ трёх множеств ($= (X_1 \times X_2) \times X_3 = X_1 \times (X_2 \times X_3)$), четырёх и т.д.

При $X_1 = X_2 = \dots = X_k$ пишут сокращённо

$$X^k = X \times X \times \dots \times X$$

и говорят о *k -й декартовой степени* множества X . Элементами X^k являются последовательности (или строки) (x_1, x_2, \dots, x_k) длины k .

Чтобы почувствовать различие между множествами $X \times Y$ и $X \cup Y$, возьмём за X и Y множества конечной мощности (*cardinality* (англ.)):

$$|X| = \text{Card } X = n, \quad |Y| = \text{Card } Y = m.$$

Тогда

$$|X \times Y| = nm, \quad |X \cup Y| = n + m - |X \cap Y|.$$

Если это не ясно, то нужно перечитать заново все определения.

2. Отображения. Понятие *отображения* или *функции* играет центральную роль в математике. При заданных множествах X и Y отображение f с *областью определения* X и *областью значений* Y сопоставляет каждому элементу $x \in X$ элемент $f(x) \in Y$, обозначаемый также f_x или f_x . В случае $Y = X$ говорят ещё о *преобразовании* f множества X в себя. Символически отображение записывается в виде $f : X \rightarrow Y$ или $X \xrightarrow{f} Y$.

Образом при отображении f называется множество всех элементов вида $f(x)$:

$$\text{Im } f = \{f(x) \mid x \in X\} = f(X) \subset Y$$

(Im — от *image* (англ.)).

Множество

$$f^{-1}(y) = \{x \in X \mid f(x) = y\}$$

называется *прообразом* элемента $y \in Y$. Более общо: для $Y_0 \subset Y$ положим

$$f^{-1}(Y_0) = \{x \in X \mid f(x) \in Y_0\} = \bigcup_{y \in Y_0} f^{-1}(y).$$

Если $y \in Y \setminus \text{Im } f$, то, очевидно, $f^{-1}(y) = \emptyset$.

Отображение $f: X \rightarrow Y$ называется *сюръективным* (*surjective* (фр.)) или *отображением на*, когда $\text{Im } f = Y$; оно называется *инъективным* (*injective* (фр.)), когда из $x \neq x'$ следует $f(x) \neq f(x')$. Наконец, $f: X \rightarrow Y$ — *биективное* (*bijective* (фр.)) или *взаимно однозначное* отображение, когда оно одновременно сюръективно и инъективно.

Равенство $f = g$ двух отображений означает по определению, что их соответствующие области совпадают:

$$X \xrightarrow{f} Y, \quad X \xrightarrow{g} Y,$$

причём $\forall x \in X \quad f(x) = g(x)$. Сопоставление “аргументу” x , т.е. элементу $x \in X$, значения $f(x) \in Y$ принято обозначать при помощи *ограниченной стрелки*: $x \mapsto f(x)$.

Пусть, например, f_n — число Фибоначчи (см. § 3) с номером n . Соответствие $n \mapsto f_n$ определяет отображение $\mathbb{N} \rightarrow \mathbb{N}$, не являющееся ни сюръективным, что очевидно, ни инъективным, поскольку $f_1 = f_2 = 1$. Если \mathbb{R}_+ — множество положительных вещественных чисел, то отображения

$$f: \mathbb{R} \rightarrow \mathbb{R}, \quad g: \mathbb{R} \rightarrow \mathbb{R}_+ \cup \{0\}, \quad h: \mathbb{R}_+ \rightarrow \mathbb{R}_+,$$

определенные одним и тем же правилом $x \rightarrow x^2$, все различны: f ни сюръективно, ни инъективно, g сюръективно, но не инъективно, а отображение h биективно. Таким образом, задание области определения и области значения — существенная часть определения отображения (функции).

Единичным (или *тождественным*) отображением $e_X: X \rightarrow X$ называется отображение, переводящее каждый элемент $x \in X$ в себя. Если X — подмножество в Y ; $X \subset Y$, то иногда бывает полезным специальное отображение — *вложение* $I: X \rightarrow Y$, которое каждому элементу $x \in X$ сопоставляет тот же самый элемент, но уже во множестве Y . Отображение $f: X \rightarrow Y$ называется *сужением* (или *ограничением*) отображения $g: X' \rightarrow Y'$, когда $X \subset X'$, $Y \subset Y'$ и $\forall x \in X \quad f(x) = g(x)$. В свою очередь g называется *продолжением* отображения f . Например, вложение $I: X \rightarrow Y$ есть ограничение единичного отображения $e_Y: Y \rightarrow Y$.

Нам представится также случай говорить о функциях многих переменных. Полезно уяснить себе, что введённое выше понятие декартовой степени X^n множества X даёт возможность говорить о

функции $f(x_1, \dots, x_n)$ многих переменных $x_i \in X, i = 1, \dots, n$, как об обычном отображении $f: X^n \rightarrow Y$.

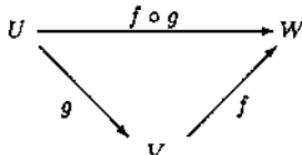
Произведением (суперпозицией или композицией) двух отображений $g: U \rightarrow V$ и $f: V \rightarrow W$ называется отображение

$$f \circ g: U \rightarrow W,$$

определенное условием

$$(f \circ g)(u) = f(g(u)) \quad \forall u \in U.$$

То же самое наглядно изображается *треугольной диаграммой*



Про эту диаграмму говорят, что она *коммутирует* (или *коммутативна*), т.е. результат перехода от U к W не зависит от того, сделаем ли мы это прямо при помощи $f \circ g$ или воспользуемся промежуточным этапом V . Заметим, что композиция определена не для любых отображений f и g . Надо, чтобы в предшествующих обозначениях у них было общим множество V . Но композиция двух преобразований множества X в себя всегда имеет смысл.

В дальнейшем вместо $f \circ g$ мы будем писать просто fg . Ясно, что

$$fex = f, \quad e_y f = f$$

для любого отображения $f: X \rightarrow Y$. Проверка этого свойства очевидна. Важное свойство композиции (произведения) отображений выражает следующая

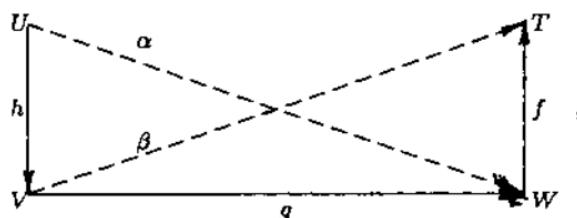
Теорема 1. *Композиция отображений подчиняется закону ассоциативности. Это значит, что если*

$$h: U \rightarrow V, \quad g: V \rightarrow W, \quad f: W \rightarrow T$$

— *три отображения, то*

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

Доказательство. Наглядно все необходимые рассуждения содержатся в следующей диаграмме:



где $\alpha = gh$, $\beta = fg$. В соответствии с формальным определением равенства отображений нужно просто сравнить значения отображений $f(gh) : U \rightarrow T$ и $(fg)h : U \rightarrow T$ в произвольной "точке" $u \in U$. Но согласно определению композиции отображений имеем

$$(f(gh))u = f((gh)u) = f(g(hu)) = (fg)(hu) = ((fg)h)u. \quad \square$$

Композиция отображений $X \rightarrow X$, вообще говоря, некоммутативна, т.е. $fg \neq gf$. В этом легко убедиться на примере, когда $X = \{a, b\}$ — множество из двух элементов, $f(a) = b$, $f(b) = a$, $g(a) = a$, $g(b) = a$. Другой пример: f и g — постоянные отображения из X в X , т.е. значения $f(x)$ и $g(x)$ не зависят от x . Тогда $f \neq g \Rightarrow fg \neq gf$.

Некоторые функции имеют *обратные*. Предположим, что $f : X \rightarrow Y$ и $g : Y \rightarrow X$ — какие-то отображения, так что композиции fg и gf определены. Если $fg = e_Y$, то f называется *левым обратным* к g , а g — *правым обратным* к f . Когда произведения в любом порядке являются единичными отображениями:

$$fg = e_Y, \quad gf = e_X, \quad (1)$$

g называется *двусторонним обратным* (или просто *обратным*) отображением для f или к f (а f — обратным отображением к g) и обозначается f^{-1} . Итак, $f(u) = v \Leftrightarrow f^{-1}(v) = u$.

Предположив существование ещё одного отображения $g' : Y \rightarrow X$, для которого

$$fg' = e_Y, \quad g'f = e_X, \quad (1')$$

мы, опираясь на равенства (1), (1') и на теорему 1, получим

$$g' = e_X g' = (gf)g' = g(fg') = g e_Y = g.$$

Таким образом, двустороннее обратное отображение к f , коль скоро оно существует, определено однозначно. Это и служит оправданием для обозначения f^{-1} .

Теорема 2. *Отображение $f : X \rightarrow Y$ тогда и только тогда имеет обратное, когда оно взаимно однозначно (биективно).*

Доказательство теоремы опирается на следующую лемму, представляющую самостоятельный интерес.

Лемма. *Если*

$$f : X \rightarrow Y, \quad g : Y \rightarrow X$$

— любые отображения, для которых $gf = e_X$, то f инъективно, а g сюръективно.

Доказательство. В самом деле, пусть $x, x' \in X$ и $f(x) = f(x')$. Тогда

$$x = e_X(x) = (gf)x = g(fx) = (gf)x' = e_X(x') = x'.$$

Стало быть, f инъективно. Если, далее, x — любой элемент из X , то

$$x = e_X(x) = (gf)x = g(fx),$$

а это доказывает сюръективность g . \square

Возвращаясь к теореме 2, предположим вначале, что f обладает обратным $g = f^{-1}$. Тогда из равенств (1) и из леммы вытекает как сюръективность, так и инъективность f . Другими словами, f биективно. Обратно: предположив f биективным, мы для любого $y \in Y$ найдём единственный элемент $x \in X$, для которого $f(x) = y$. Положив $g(y) = x$, мы определим отображение $g: Y \rightarrow X$, обладающее свойствами (1). Значит, $f^{-1} = g$. \square

Следствие. Из биективности отображения $f: X \rightarrow Y$ вытекает биективность f^{-1} , причём

$$(f^{-1})^{-1} = f. \quad (2)$$

Пусть, далее, $f: X \rightarrow Y$, $h: Y \rightarrow Z$ — биективные отображения.

Тогда биективна и их композиция hf , причём

$$(hf)^{-1} = f^{-1}h^{-1}. \quad (3)$$

Доказательство. По теореме 2 биективность f влечёт существование f^{-1} , что в силу той же теоремы эквивалентно биективности f^{-1} . Симметричность условий (1), переписанных в виде $ff^{-1} = e_Y$, $f^{-1}f = e_X$, даёт равенство (2). Далее, по условию и по теореме 2 существуют отображения

$$f^{-1}: Y \rightarrow X, \quad h^{-1}: Z \rightarrow Y$$

и их композиция

$$f^{-1}h^{-1}: Z \rightarrow X.$$

Из равенств

$$(hf)(f^{-1}h^{-1}) = ((hf)f^{-1})h^{-1} = (h(f f^{-1}))h^{-1} = hh^{-1} = e_Z,$$

$$(f^{-1}h^{-1})(hf)f^{-1}(h^{-1}(hf)) = f^{-1}((h^{-1}hh)f) = f^{-1}f = e_X$$

вытекает, что $f^{-1}h^{-1}$ — обратное отображение к f . \square

Отображение “следования” $\sigma: \mathbb{N} \rightarrow \mathbb{N}$, определённое правилом $\sigma(n) = n + 1$, инъективно, но не сюръективно, поскольку первый элемент (единица) не принадлежит $\text{Im } \sigma$. Интересно, что для конечных множеств подобная ситуация невозможна.

Теорема 3. Если X — конечное множество и преобразование $f: X \rightarrow X$ инъективно, то оно биективно.

Доказательство. Нужно лишь показать, что f сюръективно, т.е. для любого элемента $x \in X$ найдётся x' с $f(x') = x$. Положим

$$f^k(x) = f(f \dots (fx) \dots) = f(f^{k-1}x), \quad k = 0, 1, 2, \dots$$

В силу конечности X в этой последовательности элементов должны быть повторения. Пусть, скажем, $f^m(x) = f^n(x)$, $m > n$. Если $n > 0$,

то из $f(f^{m-1}x) = f(f^{n-1}x)$ и из инъективности f следует равенство $f^{m-1}(x) = f^{n-1}(x)$. Повторив достаточное число раз сокращение f , мы придём к равенству

$$f^{m-n}(x) = f^0(x) = e(x) = x.$$

А в таком случае $f(x') = x$, где $x' = f^{m-n-1}(x)$. \square

Как легко понять, сюръективное преобразование конечного множества в себя также биективно.

Несколько слов о мощности. Считается, что два множества X и Y имеют одинаковую **мощность** тогда и только тогда, когда существует биективное отображение $f : X \rightarrow Y$. Множества той же мощности, что и \mathbb{N} (или \mathbb{Z}), называются **счётными**.

УПРАЖНЕНИЯ

1. Пусть $\Omega = \{+, -, ++, +-, -+, --, +++, \dots\}$ — множество всех конечных последовательностей плюсов и минусов, а $f : \Omega \rightarrow \Omega$ — преобразование, переводящее элемент $\omega = \omega_1\omega_2\dots\omega_n \in \Omega$ в $\omega' = \omega_1\dot{\omega}_1\omega_2\dot{\omega}_2\dots\omega_n\dot{\omega}_n$, где $\dot{\omega}_k = -$, если $\omega_k = +$, и $\dot{\omega}_k = +$, если $\omega_k = -$. Показать, что в $f(f\omega)$ любой отрезок длины > 4 содержит $++$ или $--$.

2. Имеет ли отображение $f : \mathbb{N} \rightarrow \mathbb{N}$, заданное правилом $n \mapsto n^2$, правое обратное? Указать для f два левых обратных отображения.

3. Пусть $f : X \rightarrow Y$ — отображение и S, T — подмножества в X . Показать, что

$$f(S \cup T) = f(S) \cup f(T), \quad f(S \cap T) \subset f(S) \cap f(T).$$

Привести пример, показывающий, что последнее включение нельзя, вообще говоря, заменить равенством.

4. Множество всех подмножеств множества S обозначается

$$\mathcal{P}(S) = \{T \mid T \subset S\}.$$

Если, например, $S = \{s_1, s_2, \dots, s_n\}$ — конечное множество из n элементов, то $\mathcal{P}(S)$ состоит из пустого множества \emptyset , n однозлементных множеств $\{s_1\}, \{s_2\}, \dots, \{s_n\}$, $n(n-1)/2$ двухэлементных множеств $\{s_i, s_j\}$, $1 \leq i < j \leq n$, и т.д. вплоть до $T = S$. Какова мощность множества $\mathcal{P}(S)$?

5. Пусть $f : X \rightarrow Y$ — отображение и $b = f(a)$ для некоторого $a \in X$. Прообраз

$$f^{-1}(b) = f^{-1}(f(a)) = \{x \mid f(x) = f(a)\}$$

иногда называют *ещё слоем* над элементом $b \in \text{Im } f$. Показать, что всё множество X является объединением непересекающихся слоёв (т.е. разбиением множества X).

Предупреждение. Обозначение $f^{-1}(b)$ не следует ассоциировать с обратным отображением, которого может и не быть.

6. Показать, что конечная декартова степень счётного множества является счётным множеством.

7. Симметрическая разность двух множеств S и T обозначается $S \Delta T$: $S \Delta T = (S \setminus T) \cup (T \setminus S)$ (рис. 6). Показать, что $S \setminus T = (S \cup T) \setminus (S \cap T)$.

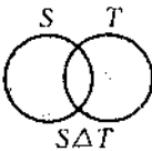


Рис. 6

§ 6. Отношения эквивалентности. Факторизация отображений

Эквивалентность систем линейных уравнений, введённая нами в § 3, наводит на мысль посмотреть на это понятие в общем плане, тем более что эквивалентностями разных типов мы пользуемся неосознанно как в логических рассуждениях, так и в обыденной жизни.

1. Бинарные отношения. Для любых двух множеств X и Y всякое подмножество $\omega \subset X \times Y$ называется **бинарным отношением** между X и Y (или просто на X , если $Y = X$). Для упорядоченной пары $(x, y) \in \omega$ используют обозначение $x\omega y$ и говорят, что x находится в отношении ω к y . Это удобно, поскольку, например, упорядочение $<$ на множестве вещественных чисел \mathbb{R} является бинарным отношением на \mathbb{R} , состоящим из всех точек плоскости \mathbb{R}^2 , которые лежат выше прямой $x - y = 0$ (рис. 7); громоздкое включение $(x, y) \in \omega(\omega <)$ заменяется обычным неравенством $x < y$.

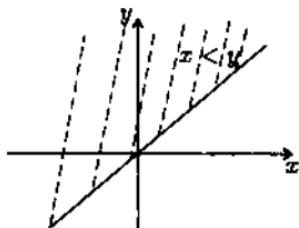


Рис. 7

Каждой функции $f : X \rightarrow Y$ сопоставляется её **график** — подмножество

$$\Gamma(f) = \{(x, y) \mid x \in X, y = f(x)\} \subset X \times Y,$$

являющееся отношением между X и Y . Изучение на \mathbb{R}^2 графиков функций $\mathbb{R} \rightarrow \mathbb{R}$ входит в курс математического анализа. Понятно, что не каждое отношение ω может служить графиком какого-либо отображения $X \rightarrow Y$. Необходимое и достаточное условие заключается в том, чтобы каждому $x \in X$ отвечал ровно один элемент y с $x\omega y$. Фактически задание X, Y и графика $\Gamma(f)$ восстанавливает f .

2. Отношение эквивалентности. Бинарное отношение \sim на X называется **отношением эквивалентности**, если для всех $x, x', x'' \in X$ выполнены условия:

- i) $x \sim x$ (рефлексивность);
- ii) $x \sim x' \implies x' \sim x$ (симметричность);
- iii) $x \sim x', x' \sim x'' \implies x \sim x''$ (транзитивность).

Запись $a \not\sim b$ выражает отрицание эквивалентности элементов $a, b \in X$.

Подмножество

$$\bar{x} = \{x' \in X \mid x' \sim x\} \subset X$$

всех элементов, эквивалентных данному x , называется **классом эквивалентности**, содержащим x . Так как $x \sim x$ (см. i)), то действительно $x \in \bar{x}$. Любой элемент $x' \in \bar{x}$ называется **представителем класса** \bar{x} .

Справедливо следующее утверждение.

Множество классов эквивалентности по отношению \sim является разбиением множества X в том смысле, что X является объединением непересекающихся подмножеств (это разбиение можно обозначить $\pi_{\sim}(X)$).

В самом деле, так как $x \in \bar{x}$, то $X = \bigcup_{x \in X} \bar{x}$. Если теперь $\bar{x}' \cap \bar{x}'' \neq \emptyset$ и $x \in \bar{x}' \cap \bar{x}''$, то $x \sim x'$ и $x \sim x''$, откуда в силу транзитивности iii) имеем $x' \sim x''$ и $\bar{x}' = \bar{x}''$. Значит, различные классы не пересекаются. \square

Пусть $\Pi = \mathbb{R}^2$ — вещественная плоскость с прямоугольной системой координат.

Взяв за свойство \sim принадлежность точек $P, P' \in \Pi$ одной горизонтальной прямой, мы получим, очевидно, отношение эквивалентности с классами — горизонтальными прямыми (рис. 8).

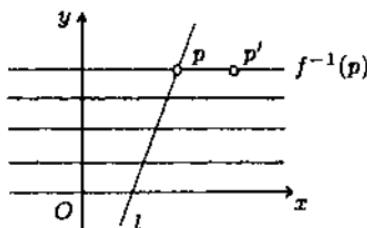


Рис. 8

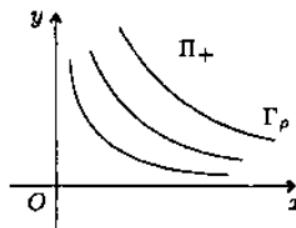


Рис. 9

Гиперболы Γ_ρ (рис. 9) вида $xy = \rho > 0$ определяют отношение эквивалентности в области $\Pi_+ \in \Pi$ точек $P(x, y)$ с координатами $x > 0, y > 0$. Эти геометрические примеры наглядно показывают, что верно следующее обратное утверждение.

Если имеется какое-то разбиение $\pi(X)$ множества X на непересекающиеся подмножества C_x , то C_x будут классами эквивалентности по некоторому отношению эквивалентности \sim .

В самом деле, по условию каждый элемент $x \in X$ содержится точно в одном подмножестве C_x . Достаточно считать $x \sim x'$ в том и только том случае, когда x и x' лежат в одном и том же подмножестве C_a . Очевидно, это отношение \sim рефлексивно, симметрично и транзитивно, т.е. является отношением эквивалентности. Далее, $x \in C_a \implies \bar{x} = C_a$ по определению \sim . Стало быть, $\pi(X) = \pi_{\sim}(X)$. \square

3. Факторизация отображений. Ввиду установленного выше взаимно однозначного соответствия между отношениями эквивалентности и разбиениями множества X принято разбиение, отвечающее отношению эквивалентности \sim , обозначать X/\sim и называть **фактормножеством** X относительно \sim (или по отношению \sim). Сюръективное отображение

$$p: x \mapsto p(x) = \bar{x} \quad (1)$$

называется **естественным отображением** (или **канонической проек-**

цией) X на фактормножество X/\sim .

Пусть X, Y — два множества и $f: X \rightarrow Y$ — отображение. Бинарное отношение ω_f :

$$\forall x, x' \in X \quad x\omega_fx' \Leftrightarrow f(x) = f(x'),$$

очевидно, рефлексивно ($f(x) = f(x)$), симметрично ($f(x') = f(x) \Rightarrow f(x) = f(x')$) и транзитивно ($f(x) = f(x') \& f(x') = f(x'') \Rightarrow f(x) = f(x'')$). Таким образом, ω_f — отношение эквивалентности на X . Соответствующие классы эквивалентности \bar{x} являются слоями (прообразами) в смысле упр. 5 § 5. Другими словами,

$$x = \{x' \mid f(x') = f(x)\}.$$

Отображение $f: X \rightarrow Y$ индуцирует отображение $\bar{f}: X/\omega_f \rightarrow Y$, определённое правилом

$$\bar{f}(\bar{x}) = f(x),$$

или, что то же самое,

$$\bar{f}p(x) = f(x), \tag{2}$$

где p — естественное отображение (1). Так как

$$\bar{x} = \bar{x}' \Leftrightarrow f(x) = f(x'),$$

то соотношение (2), задающее \bar{f} , не зависит от выбора представителя x класса \bar{x} . В таких случаях говорят, что определение \bar{f} является *правильным* или *корректным*. Коммутативная диаграмма

$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ & \searrow p & \swarrow \bar{f} \\ & X/\omega_f & \end{array}$$

наглядно описывает *факторизацию (разложение)*

$$f = \bar{f} \cdot p \tag{3}$$

отображения f в произведение сюръективного отображения p и инъективного отображения \bar{f} . Инъективность \bar{f} вытекает из того, что

$$\bar{f}(\bar{x}_1) = \bar{f}(\bar{x}_2) \Leftrightarrow f(x_1) = f(x_2) \Leftrightarrow x_1 = x_2.$$

Очевидно, сюръективность \bar{f} равносильна сюръективности f . Заметим, что если $f': X/\omega_f \rightarrow Y$ — ещё одно отображение, для которого выполнено соотношение (3): $f' \cdot p = f$, то из

$$f'(\bar{x}) = f'(px) = (f'p)x = f(x) = \bar{f}(\bar{x})$$

(см. (2)) следует на самом деле равенство $f' = f$. Стало быть, отображение \bar{f} , делающее указанную выше треугольную диаграмму коммутативной, единственno.

4. Упорядоченные множества. Упорядочением множества X (или порядком на X) называется бинарное отношение \leqslant на X , обладающее свойствами рефлексивности ($x \leqslant x$), антисимметричности (если $x \leqslant y$ и $y \leqslant x$, то $x = y$) и транзитивности (если $x \leqslant y$ и $y \leqslant z$, то $x \leqslant z$). При $x \leqslant y$ и $x \neq y$ пишут $x < y$. Вместо $x \leqslant y$ используется также запись $y \geqslant x$. Пара элементов $x, x' \in X$ может и не находиться в отношении \leqslant . Если, однако, $x \leqslant x'$ или $x' \leqslant x$ для каждой пары элементов из X , то X называется линейно упорядоченным множеством или цепью. В общем же случае говорят о частично упорядоченном множестве на X .

Множество $X = \mathcal{P}(S)$ подмножеств множества S (см. упр. 4 § 5) с обычным отношением включения $R \subset T$ между подмножествами, а также множество N натуральных чисел с отношением $d | n$ (n делится на d) являются примерами частично упорядоченных множеств.

Пусть X — произвольное частично упорядоченное множество, x и y — его элементы. Говорят, что y накрывает x , если $x < y$ и не существует z с условием $x < z < y$. В случае $\text{Card } X < \infty$ $x < y$ (т.е. x и y сравнимы) тогда и только тогда, когда найдётся цепочка элементов $x = x_1, x_2, \dots, x_{n-1}, x_n = y$, в которой x_{i+1} накрывает x_i . Понятие накрытия удобно при изображении конечного частично упорядоченного множества X плоской диаграммой. Элементы множества X изображаются точками. Если y накрывает x , то y помещается выше x и x соединяется с y прямолинейным отрезком. Сравнимость y и x изображается пониждающейся ломаной, соединяющей y с x , причём таких ломаных может быть несколько. Несравнимые x и y не соединяются. На двух из приводимых диаграмм (рис. 10) изображены "отрезок" натурального ряда чисел и множество $\mathcal{P}(\{a, b, c\})$ (N — естественное линейно упорядоченное множество, а упорядочение на $\mathcal{P}(S)$ было введено выше).

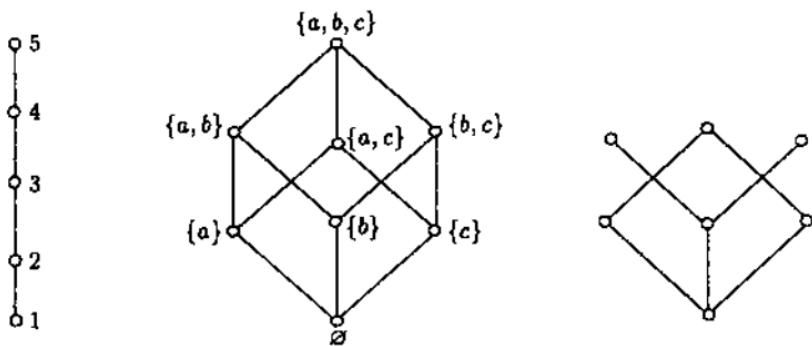


Рис. 10

Наибольшим элементом частично упорядоченного множества X называется элемент $n \in X$ такой, что $x \leqslant n$ для всех $x \in X$, а максимальным — элемент $m \in X$, для которого из $m \leqslant x \in X$ следует $x = m$. Наибольший элемент всегда максимальен, но не обратно.

Максимальных элементов может быть много, но наибольший элемент, если он существует, определён однозначно. Те же замечания относятся к наименьшему и минимальному элементам. На рис. 10 две диаграммы слева имеют наибольшие и наименьшие элементы, диаграмма справа — три максимальных элемента (один наименьший) но нет наибольшего элемента.

Теория частично упорядоченных алгебраических систем (булевы алгебры, решётки) насыщена содержательными результатами и занимает важное место в алгебре, но мы не имеем возможности её касаться. Этот параграф преследует скромную цель — познакомить читателя с ещё одним естественным бинарным отношением и дать представление о диаграммах, которые помогут в будущем понять взаимное расположение подгрупп в группах или, скажем, расположение подполей в полях.

УПРАЖНЕНИЯ

- Показать, что фактормножество \mathbb{R}^2 / \sim , получающееся из рис. 8, и любая прямая l , пересекающая ось Ox , находятся в биективном соответствии.
- Положить $P(x, y) \sim P(x', y')$ для точек вещественной координатной плоскости \mathbb{R}^2 в тоиности тогда, когда $x' - x \in \mathbb{Z}$ и $y' - y \in \mathbb{Z}$. Доказать, что \sim является отношением эквивалентности и что фактормножество \mathbb{R}^2 / \sim геометрически изображается точками на торе (поверхности бублика; рис. 11).

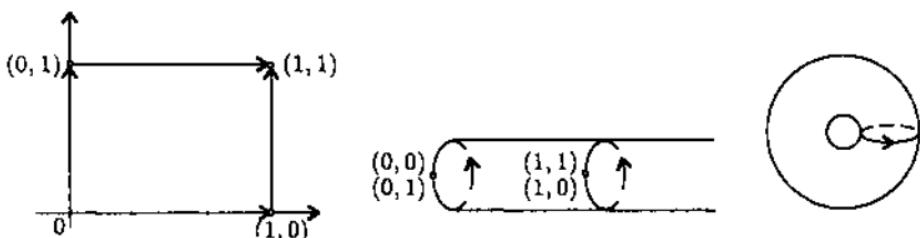


Рис. 11

- Показать, что множества из двух, трех и четырёх элементов имеют соответственно 2, 5 и 15 различных фактормножеств.

- Пусть \sim — отношение эквивалентности на множестве X и $f: X \rightarrow Y$ — отображение, для которого

$$x \sim x' \implies f(x) = f(x').$$

Показать, что это условие *совместимости* f с \sim (более слабое, чем рассмотренное в п. 2) позволяет правильно определить индуцированное отображение $\bar{f}: X/\sim \rightarrow Y$, приводящее к факторизации $f = \bar{f} \cdot p$, но \bar{f} уже не обязательно должно быть инъективным. В чём заключается условие инъективности \bar{f} ?

- Изобразить диаграммами частично упорядоченные множества:

- $\mathcal{P}(\{a, b, c, d\})$;
- множество всех делителей целого числа 24 (отношения порядка указаны в тексте).

§ 7. Принцип математической индукции

Считается, что нам известно множество $\mathbb{N} = \{1, 2, 3, \dots\}$ всех *натуральных* (или *целых положительных*) чисел. На самом деле отправной точкой для изучения \mathbb{N} служит аксиоматика Пеано (Дж. Пеано, 1858–1932). Из аксиом Пеано (мы их не приводим) вытекают свойства сложения, умножения и линейного упорядочения (см. п. 4 § 6) натуральных чисел, точнее, системы $\mathbb{N} \cup \{0\}$. В частности, доказывается интуитивно ясное утверждение: в *каждом непустом множестве* $S \subset \mathbb{N}$ *имеется наименьший элемент*, т.е. натуральное число $s \in S$, меньшее всех остальных чисел в S . С учётом этого утверждения из аксиом Пеано извлекается следующий

Принцип индукции. Предположим, что для каждого $n \in \mathbb{N}$ мы имеем некоторое утверждение $M(n)$. Предположим также, что мы располагаем правилом, позволяющим установить истинность $M(l)$ для данного l при условии, что $M(k)$ верно для всех $k < l$ (в частности, подразумевается, что мы можем проверить истинность $M(l)$).

Тогда $M(n)$ верно для всех $n \in \mathbb{N}$.

В самом деле, допустим, что подмножество

$$S = \{s \mid s \in \mathbb{N}, M(s) \text{ неверно}\} \subset \mathbb{N}$$

непусто. Согласно сказанному выше S содержит наименьший элемент s_0 . Тогда утверждение $M(s_0)$ ложно, а $M(s)$ истинно для каждого $s < s_0$. Это, однако, противоречит нашему предполагаемому умению доказывать истинность $M(s_0)$.

Здесь не место для всестороннего обсуждения принципа математической индукции. Мы ограничимся замечанием, что он отражает, так сказать, суть натурального ряда, а познание последнего не сводится к чему-либо существенно более простому. Стоит ещё обратить внимание на одно обстоятельство. Именно, непременным моментом “доказательства методом полной индукции” является установление базиса индукции, т.е. проверка того, что свойство или утверждение выполнено для небольших n . Без такой проверки можно приходить к произвольным умозаключениям типа “все студенты одинакового роста”. Вот и рассуждение. Пустое множество студентов и множество из одного студента обладают этим свойством. Делаем предположение индукции, что им обладает любое множество из $\leq n$ студентов. Во множестве из $n+1$ студентов первые n и последние n студентов одинакового роста по предположению индукции. Эти множества пересекаются по подмножеству из $n-1$ студентов тоже одинакового роста. Значит, все $n+1$ студентов одинакового роста. На самом деле первое содержательное утверждение относилось бы ко множеству из любых двух студентов, а здесь-то оно как раз и неверно. Насколько же длинным должно быть основание индукции? Обычно это ясно

из доказательства. В нашем элементарном примере важным является условие непустоты пересечения двух множеств, т.е. выполнение неравенства $n - 1 \geq 1$, откуда $n \geq 2$.

В более сложных ситуациях, в особенности когда приходится определять или строить объект по индукции при помощи рекуррентных соотношений, необходимо проявлять особую заботу о базисе индукции. Например, делимость на 5 числа Фибоначчи f_{5m} (см. пример 2 § 3) при любом целом $m \geq 1$ вытекает из равенства $f_5 = 5$ и из соотношения $f_{5(m+1)} = 5f_{5m+1} + 3f_{5m}$, которое ещё нужно получить. С другой стороны, нельзя впадать в иную крайность: убедившись в истинности $M(k)$ для всех k из достаточно длинного отрезка $1 \leq k \leq l$ натурального ряда, делать необоснованный вывод (это будет так называемая *неполная индукция*) об истинности $M(n)$ для всех $n \in \mathbb{N}$.

Вот — два обескураживающих примера.

Пример 1. П. Ферма полагал, что все числа вида $F_n = 2^n + 1$, $n = 0, 1, \dots$ (числа Ферма), простые. Первые пять чисел Ферма простые, но для F_5 Эйлер нашёл разложение $F_5 = 4294967297 = 641 \cdot 6700417$. Настойчивые усилия получить при помощи новейших ЭВМ хотя бы одно новое простое число Ферма пока не увенчались успехом. Одним из последних "достижений" в этом направлении является проверка того, что F_{1945} делится на $5 \cdot 2^{1947} + 1$.

Пример 2. Исследование при $n = 1, 2, \dots, 40$ чисел вида $n^2 - n + 41$ (многочлен, предложенный Эйлером) способно склонить к мысли о простоте этих чисел при любом n (о простых числах см. § 9). Однако $41^2 - 41 + 41 = 41^2$.

Примеров такого рода можно приводить сколь угодно много.

В рассуждениях по индукции иногда самое важное — придать надлежащую форму доказываемому утверждению. Предположим, что нужно найти сумму

$$p_k(n) = 1^k + 2^k + 3^k + \dots + (n-1)^k + n^k, \quad k = 1, 2, 3.$$

Задача значительно облегчается, когда вам скажут, что предполагаемый ответ содержится в выражениях

$$p_1(n) = \frac{n(n+1)}{2}, \quad p_2(n) = \frac{n(n+1)(2n+1)}{2}, \quad p_3(n) = \left[\frac{n(n+1)}{2} \right]^2.$$

Степенные суммы $p_k(n)$ самого общего вида будут ещё обсуждаться в связи с корнями многочленов (см. гл. 6), а сейчас заметим, что встретившаяся нам в п. 5 § 3 сумма $\Gamma(n)$ имеет вид

$$\Gamma(n) = n(n-1) + \dots + k(k-1) + \dots + 1 \cdot (1-1) =$$

$$= \sum_{k=1}^n k^2 - \sum_{k=1}^n k = p_2(n) - p_1(n)$$

(в дальнейшем знак суммирования \sum будет систематически использоваться). Опираясь на приведённые выше выражения для $p_2(n)$ и $p_1(n)$, получаем, что $\Gamma(n) = (n^3 - n)/3$. Разумеется, к тому же ре-

зультату нетрудно прийти, рассуждая по индукции непосредственно в $\Gamma(n)$.

Если до вида $p_1(n)$ додуматься нетрудно, то вид $p_2(n)$ и $p_3(n)$ уже не так тривиален, а соотношение

$$p_5 + p_7 = 2 \left[\frac{n(n+1)}{2} \right]^4$$

вообще нужно было бы искать по какому-то определённому плану. В данном случае такой план указать можно, но не в этом дело. Для обоснования всех указанных выше соотношений нужно провести прямыми вычислениями шаг индукции от n к $n+1$. Оставим это читателю в качестве полезного упражнения.

Кстати, в этом упражнении пригодится так называемая *биномиальная формула*

$$(a+b)^n = a^n + \binom{n}{1} a^{n-1} b + \dots + \binom{n}{k} a^{n-k} b^k + \dots + b^n. \quad (1)$$

Здесь под a и b подразумеваются произвольные числа, а *биномиальный коэффициент* $\binom{n}{k}$ при одночлене $a^{n-k} b^k$ имеет вид

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)\dots(n-k+1)}{k(k-1)\dots2\cdot1}, \quad (2)$$

где $n! = n(n-1)\dots2\cdot1$ (эн-факториал). Это — быстро растущая величина, например, $6! = 720$, $10! = 3628800$, а $100! > 10^{150}$. Полезно дополнить выражение (2) соглашениями $0! = 1$ и $\binom{n}{k} = 0$ при $k < 0$. Отметим ещё, что

$$\binom{n}{n-k} = \binom{n}{k}$$

(свойство симметричности биномиальных коэффициентов).

Формулу (1), очевидно, верную при $n = 1, 2$, мы докажем индукцией по n . Считая её справедливой для всех показателей $\leq n$, умножим обе части соотношения (1) на $a+b$. Получим

$$(a+b)^{n+1} = (a+b)^n(a+b) =$$

$$= a^n(a+b) + \dots + \binom{n}{k} a^{n-k} b^k (a+b) + \dots + b^n(a+b) =$$

$$= a^{n+1} + a^n b + \dots + \binom{n}{k-1} a^{n+2-k} b^{k-1} + \binom{n}{k-1} a^{n+1-k} b^k + \\ + \binom{n}{k} a^{n+1-k} b^k + \binom{n}{k} a^{n-k} b^{k+1} + \dots + a b^n + b^{n+1}.$$

Приведение подобных членов показывает, что коэффициентом при

одночлене $a^{n+1-k}b^k$ будет

$$\begin{aligned}\binom{n}{k-1} + \binom{n}{k} &= \frac{n!}{(k-1)!(n-k+1)!} + \frac{n!}{k!(n-k)!} = \\ &= \frac{n!}{(k-1)!(n-k)!} \left[\frac{1}{n-k+1} + \frac{1}{k} \right] = \\ &= \frac{n!}{(k-1)!(n-k)!} \cdot \frac{n+1}{k(n-k+1)} = \frac{(n+1)!}{k!(n+1-k)!} = \binom{n+1}{k},\end{aligned}$$

т.е. как раз биномиальный коэффициент вида (2) с верхним индексом, увеличенным на единицу. Тем самым справедливость формулы (1) доказана для всех $n \in \mathbb{N}$.

Если записать

$$(a+b)^n = (a+b)(a+b)\dots(a+b),$$

присвоив каждому множителю справа номер от 1 до n , и посмотреть на те подмножества номеров

$$1 \leq i_1 < i_2 < \dots < i_k \leq n,$$

которые отвечают при умножении одночлену $a^{n-k}b^k$, то мы придём к выводу, что $\binom{n}{k}$ есть не что иное, как число всех подмножеств мощности k множества из n элементов. Несколько "старомодный" термин — число

$$C_n^k = \binom{n}{k}$$

сочетаний из n по k — выражает по существу то же самое.

В частности, мощность множества $\mathcal{P}(\{s_1, \dots, s_n\})$ (см. упр. 4 § 5) равна

$$\binom{n}{0} + \binom{n}{1} + \dots + \binom{n}{n-1} + \binom{n}{n}.$$

Но, полагая $a = b = 1$ в формуле (1), получим

$$2^n = \binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \dots + \binom{n}{n-1} + \binom{n}{n}.$$

Таким образом,

$$\text{Card } \mathcal{P}(\{s_1, s_2, \dots, s_n\}) = 2^n.$$

Биномиальные коэффициенты — почти непременный атрибут элементарных комбинаторных рассуждений. Вот — наглядный геометрический пример.

Пример (Amer. Math. Monthly. — 1977. — V. 84, № 6). Известна задача об определении числа R_n областей, образуемых в круге $\binom{n}{2}$ хордами, которые соединяют n фиксированных точек на окружности, при предположении, что никакие три хорды не пересекаются внутри круга (рис. 12). Результат при $n = 1, 2, 3, 4, 5$ наводит на мысль, что $R_n = 2^{n-1}$; но на самом деле правильной будет формула $R_n = 1 + \binom{n}{2} + \binom{n}{4}$. Попробуйте это доказать.

Доказательство теоремы или построение объекта иногда удобно проводить, опираясь на более сложные формы индукции. Например, принцип *двойной индукции* заключается в следующем. Пусть любым

натуральным числам m и n отвечает некоторое утверждение $Y(m, n)$, причём:

- $Y(m, 1)$ и $Y(1, n)$ истинны для всех m и n ;
- если $Y(k-1, l)$ и $Y(k, l-1)$ истинны, то $Y(k, l)$ также истинно.

Это эквивалентно:

- если $Y(k', l')$ истинно при всех $k' \leq k$, $l' \leq l$, $k' + l' < k + l$, то $Y(k, l)$ также истинно.

Тогда утверждение $Y(m, n)$ истинно для всех натуральных m и n .

УПРАЖНЕНИЯ

1. Положим

$$\begin{aligned}s(n) &= \sin \varphi + \sin 2\varphi + \dots + \sin n\varphi, \\ c(n) &= \cos \varphi + \cos 2\varphi + \dots + \cos n\varphi.\end{aligned}$$

Индукцией по n доказать формулы

$$s(n) = \frac{\sin(n\varphi/2) \sin((n+1)\varphi/2)}{\sin(\varphi/2)}, \quad c(n) = \frac{\sin(n\varphi/2) \cos((n+1)\varphi/2)}{\sin(\varphi/2)}.$$

2. Имеют место формулы:

$$a) \sum_{k=1}^n \operatorname{ctg}^2 \left(\frac{k\pi}{2n+1} \right) = \frac{n(2n-1)}{3};$$

$$b) \sum_{k=0}^n \binom{2k}{k} \binom{2n-2k}{n-k} = 4^n.$$

Убедиться в их справедливости хотя бы при $n \leq 5$.

§ 8. Перестановки

1. **Стандартная запись перестановки.** Разберём немного тему, начатую в § 5, применительно к биективным преобразованиям конечных множеств. На этой базе естественным образом возникают важные алгебраические понятия.

Пусть Ω — конечное множество из n элементов. Поскольку природа его элементов для нас несущественна, удобно считать, что $\Omega = \{1, 2, \dots, n\}$. Элементы множества $S_n = S(\Omega)$ всех взаимно однозначных преобразований $\Omega \rightarrow \Omega$, обычно обозначаемые строчными

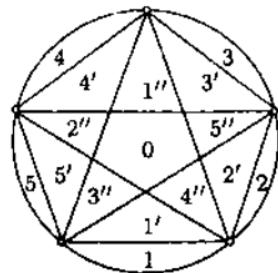


Рис. 12

буквами греческого алфавита, называются *перестановками*. Лишь за единичным преобразованием $e = e_\Omega$ сохранилась буква латинского алфавита.

В развернутой и наглядной форме произвольную перестановку $\pi: i \mapsto \pi(i)$, $i = 1, 2, \dots, n$, изображают в виде

$$\pi = \begin{pmatrix} 1 & 2 & \dots & n \\ i_1 & i_2 & \dots & i_n \end{pmatrix},$$

полностью указывая все образы:

$$\pi: \begin{matrix} 1 & 2 & \dots & n \\ \downarrow & \downarrow & & \downarrow \\ i_1 & i_2 & \dots & i_n \end{matrix},$$

где $i_k = \pi(k)$, $k = 1, \dots, n$ — переставленные символы $1, 2, \dots, n$.

Перестановки $\sigma, \tau \in S_n$ перемножаются в соответствии с общим правилом композиции отображений: $(\sigma\tau)(i) = \sigma(\tau(i))$. Например, для перестановок

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \end{pmatrix}, \quad \tau = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{pmatrix}$$

имеем

$$\sigma\tau = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{pmatrix} = \begin{matrix} 1 & 2 & 3 & 4 \\ \downarrow & \downarrow & \downarrow & \downarrow \\ 4 & 3 & 2 & 1 \\ \downarrow & \downarrow & \downarrow & \downarrow \\ 1 & 4 & 3 & 2 \end{matrix} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 4 & 3 & 2 \end{pmatrix}.$$

В то же время

$$\tau\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 1 & 4 \end{pmatrix},$$

так что $\sigma\tau \neq \tau\sigma$.

Согласно результатам § 5 умножение перестановок подчиняется следующим правилам.

i) Умножение ассоциативно, т.е. $(\alpha\beta)\gamma = \alpha(\beta\gamma)$ для всех $\alpha, \beta, \gamma \in S_n$.

ii) S_n обладает единичным элементом e : $\pi e = \pi = e\pi$ для всех $\pi \in S_n$.

iii) Для каждой перестановки $\pi \in S_n$ существует обратная перестановка π^{-1} : $\pi\pi^{-1} = \pi^{-1}\pi = e$.

Эти три свойства, дополненные общими принципами, на которых мы не хотим сейчас останавливаться (см. гл. 4), дают основание говорить о группе S_n . Точнее, множество S_n , рассматриваемое вместе с естественной операцией умножения его элементов (композицией перестановок), называется *симметрической группой степени n* .

(иначе, симметрической группой на n символах или на n точках). Для нас пока это всего лишь удобное терминологическое соглашение, смещающее акценты с множества S_n как такового на мультилинейные свойства перестановок, т.е. на то, что может быть выявлено при композиции элементов из S_n . Симметрическая группа S_n лежала у истоков общей теории групп и теории Галуа более 170 лет тому назад, и можно только поражаться связанным с ней обилию математических идей.

Замечание. Иногда элементы группы S_n называют подстановками, используя термин перестановка в качестве синонима расположения чисел $1, 2, \dots, n$ в каком-то фиксированном порядке. Так как между такими упорядочениями чисел и элементами группы S_n имеется взаимно однозначное соответствие, а слово “перестановка” ассоциируется в сознании скорее с действием, чем с застывшим упорядочением, то подстановки у нас из употребления исключены. Впрочем, ниже мы будем говорить, например, о подстановке числа в многочлен, но это служит лишь дополнительным аргументом в пользу указанного терминологического соглашения.

Если нужны ещё какие-то доводы, то их можно найти по меньшей мере: а) в научной литературе; б) в учебнике П.С. Александрова “Лекции по аналитической геометрии” (Наука, 1968, с. 767).

Найдём порядок $|S_n|$ группы S_n . Символ 1 можно подходящей перестановкой σ перевести в любой другой символ $\sigma(1)$, для чего существует в точности n различных возможностей. Но, зафиксировав $\sigma(1)$, мы имеем право брать в качестве $\sigma(2)$ лишь один из оставшихся $n - 1$ символов (всего различных пар $\sigma(1), \sigma(2)$ имеется $(n - 1) + (n - 1) + \dots + (n - 1) = n(n - 1)$), в качестве $\sigma(3)$ — соответственно $n - 2$ символов и т.д. Всего возможностей выбора $\sigma(1), \sigma(2), \dots, \sigma(n)$, а стало быть, и всех различных перестановок будет $n(n - 1) \dots 2 \cdot 1 = n!$. Таким образом,

$$\text{Card } S_n = |S_n| = n!.$$

2. Цикловая структура перестановки. Разложим теперь перестановки из S_n в произведения более простых перестановок. Идею разложения поясним схематически (рис. 13) на примере указанных выше перестановок $\sigma, \tau \in S_4$.

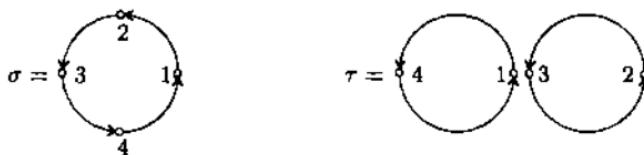


Рис. 13

Перестановка σ , кратко записываемая в виде $\sigma = (1\ 2\ 3\ 4)$, или,

что то же самое, в виде

$$\sigma = (2 \ 3 \ 4 \ 1) = (3 \ 4 \ 1 \ 2) = (4 \ 1 \ 2 \ 3),$$

называется циклом длины 4, а перестановка

$$\tau = (1 \ 4)(2 \ 3)$$

— произведением двух независимых (непересекающихся) циклов $(1 \ 4)$ и $(2 \ 3)$ длины 2. Заметим, что

$$\sigma^2 = (1 \ 3)(2 \ 4), \quad \sigma^4 = (\sigma^2)^2 = e, \quad \tau^2 = e.$$

Пусть теперь π — произвольная перестановка из S_n . Её степень π^s определяется по индукции (см. доказательство теоремы 3 § 5):

$$\pi^s = \begin{cases} \pi(\pi^{s-1}), & \text{если } s > 0, \\ e, & \text{если } s = 0, \\ \pi^{-1}((\pi^{-1})^{(-s-1)}), & \text{если } s < 0. \end{cases}$$

При таком определении, очевидно,

$$\pi^s \pi^t = \pi^{s+t} = \pi^t \pi^s, \quad s, t \in \mathbb{Z}$$

(последовательное приписывание π или π^{-1} при s и t одинакового знака и замена $\pi\pi^{-1}$, $\pi^{-1}\pi$ на e при s и t разных знаков). Так как $|\Omega| < \infty$, то на самом деле для каждой перестановки $\pi \in S_n$ найдётся однозначно определённое натуральное число $q = q(\pi)$ такое, что все различные степени содержатся во множестве $\langle \pi \rangle = \{e, \pi, \dots, \pi^{q-1}\}$ и $\pi^q = e$. Это число q называется ещё порядком перестановки π . Так, рассмотренные выше перестановки σ и τ имеют соответственно порядки 4 и 2.

Две точки $i, j \in \Omega$ назовём π -эквивалентными, если $j = \pi^s(i)$ для некоторого $s \in \mathbb{Z}$. Так как

$$\begin{aligned} i = \pi^0(i), \quad j = \pi^s(i) \implies i = \pi^{-s}(j), \quad j = \pi^s(i), \quad k = \pi^t(j) \implies \\ \implies k = \pi^{s+t}(i), \end{aligned}$$

то, очевидно, мы имеем дело с рефлексивным, симметричным и транзитивным отношением на Ω (см. п. 2 § 6). В соответствии с общим свойством отношений эквивалентности получаем разбиение

$$\Omega = \Omega_1 \cup \dots \cup \Omega_p \tag{1}$$

множества Ω на попарно непересекающиеся классы $\Omega_1, \dots, \Omega_p$, которые принято называть ещё π -орбитами. Название это вполне обосновано. Каждая точка $i \in \Omega$ принадлежит в точности одной орбите, и если $i \in \Omega_k$, то Ω_k состоит из образов точки i при действии степеней элемента π : $i, \pi(i), \pi^2(i), \dots, \pi^{l_k-1}(i)$. Здесь $l_k = |\Omega_k|$ — длина π -орбиты Ω_k . Очевидно, что

$$i_k \leq q = \text{Card}(\langle \pi \rangle), \quad \pi^{i_k}(i) = i,$$

причём i_k — наименьшее число, обладающее этим свойством. Положив

$$\pi_k = (i\pi(i) \dots \pi^{l_k-1}(i)) = \begin{pmatrix} i & \pi(i) & \dots & \pi^{l_k-2}(i) \\ \pi(i) & \pi^2(i) & \dots & \pi^{l_k-1}(i) \end{pmatrix},$$

мы придём как раз к перестановке, называемой циклом длины l_k .

Вопрос вкуса и удобства — писать $(1\ 2\ 3\ \dots\ l)$ или $(1, 2, 3, \dots, l)$.

Цикл π_k оставляет на месте все точки из множества $\Omega \setminus \Omega_k$, а $\pi(j) = \pi_k(j)$ для любой точки $j \in \Omega_k$. Это свойство даёт нам основание называть π_s , π_t , $s \neq t$, независимыми или непересекающимися циклами. Так как $\pi^{l_k}(i) = i$ для $i \in \Omega_k$, то $\pi_k^{l_k} = e$.

Итак, с разбиением (1) ассоциируется разложение перестановки π в произведение

$$\pi = \pi_1 \pi_2 \dots \pi_p, \quad (2)$$

где все циклы перестановочны: $\pi = \pi_1 \pi_2 \dots \pi_p = \pi_{l_1} \pi_{l_2} \dots \pi_{l_p}$. Можно считать, например, что $l_1 \geq l_2 \geq \dots \geq l_m > l_{m+1} = \dots = l_p = 1$.

Если цикл $\pi_k = (i)$ имеет длину 1, то он действует как единичная перестановка. Естественно такие циклы в произведении (2) опускать:

$$\pi = \pi_1 \pi_2 \dots \pi_m, \quad l_k > 1, \quad 1 \leq k \leq m. \quad (3)$$

Например, перестановку

$$\pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 2 & 3 & 4 & 5 & 1 & 7 & 6 & 8 \end{pmatrix} \in S_8$$

мы запишем в виде

$$\pi = (1\ 2\ 3\ 4\ 5)(6\ 7)(8) = (1\ 2\ 3\ 4\ 5)(6\ 7). \quad (4)$$

Некоторую неловкость вызывает то обстоятельство, что $(1\ 2\ 3\ 4\ 5)(6\ 7)$ можно интерпретировать как перестановку из S_n при любом $n \geq 7$, однако при фиксированном n никакой неоднозначности нет.

Более точно, пусть наряду с разложением (3) мы имеем ещё одно разложение $\pi = \alpha_1 \alpha_2 \dots \alpha_r$ в произведение независимых циклов, и пусть i — символ, не остающийся на месте при действии π . Тогда $\pi_s(i) \neq i$, $\alpha_t(i) \neq i$ для одного (и только одного) из циклов π_1, \dots, π_m и одного из $\alpha_1, \dots, \alpha_r$. Имеем $\pi_s(i) = \pi(i) = \alpha_t(i)$. Если мы уже знаем, что

$$\pi_s^k(i) = \pi^k(i) = \alpha_t^k(i), \quad (5)$$

то, применяя к этим равенствам перестановку π и используя перестановочность π с π_s^k и с α_t^k , получаем

$$\pi \pi_s^k = \pi^{k+1}(i) = \pi \alpha_t^k(i),$$

откуда $\pi_s^k \pi(i) = \pi^{k+1}(i) = \alpha_t^k \pi(i)$ и, наконец,

$$\pi_s^{k+1}(i) = \pi^{k+1}(i) = \alpha_t^{k+1}(i).$$

Значит, равенства (5) справедливы при любом $k = 0, 1, 2, \dots$. Но цикл однозначно определяется действием его степеней на любой символ, который не остаётся на месте. Следовательно, $\pi_s = \alpha_t$. Далее применяется индукция по t или r .

Итак, нами доказана

Теорема 1. Каждая перестановка $\pi \neq e$ в S_n является произведением независимых циклов длины ≥ 2 . Это разложение в произведение определено однозначно с точностью до порядка следования циклов.

Обратим внимание на циклы длины 2.

Определение. Цикл длины 2 называется *транспозицией*.

Любая транспозиция имеет вид $\tau = (ij)$ и оставляет на месте все символы, отличные от i, j . Из теоремы 1 вытекает

Следствие. Каждая перестановка $\pi \in S_n$ является произведением транспозиций.

Доказательство. В самом деле, в силу теоремы 1 достаточно записать в виде произведения транспозиций каждый из циклов. Но это можно сделать, например, так:

$$(1 \ 2 \ \dots \ l-1 \ l) = (1 \ l)(1 \ l-1) \dots (1 \ 3)(1 \ 2). \square$$

Формулировки теоремы 1 и её следствия нуждаются в пояснении. Как следует из определения цикла $\sigma = (i_1 \ i_2 \ i_3 \ \dots \ i_{l-1} \ i_l)$,

$$i_1 \mapsto i_2, \quad i_2 \mapsto i_3, \quad \dots, \quad i_{l-1} \mapsto i_l, \quad i_l \mapsto i_1$$

и

$$j \mapsto j, \quad j \in \Omega \setminus \{i_1, i_2, \dots, i_{l-1}, i_l\},$$

и потому ничто не изменится, если мы запишем $\sigma = (i_2 \ i_3 \ \dots \ i_l \ i_1)$, т.е. произведём циклический сдвиг номеров, входящих в σ . Таким образом, утверждение единственности в теореме 1 носит по существу абсолютный характер. С другой стороны, в следствии ни о какой единственности записи перестановки через транспозиции не может быть и речи. Скажем,

$$\sigma = (i_1 \ i_2 \ i_3 \ \dots \ i_{l-1} \ i_l) = (i_1 \ i_l)(i_1 \ i_{l-1}) \dots (i_1 \ i_3)(i_1 \ i_2),$$

$$\sigma = (i_2 \ i_3 \ \dots \ i_{l-1} \ i_l \ i_1) = (i_2 \ i_1)(i_2 \ i_l)(i_2 \ i_{l-1}) \dots (i_2 \ i_3).$$

Эти две записи одной и той же перестановки σ содержат по однаковому числу $l - 1$ совершенно разных транспозиций (лишь $(i_2 \ i_1) = (i_1 \ i_2)$). Более того, транспозиции, вообще говоря, не перестановочные, а их число не является инвариантом перестановки. Например, в S_4 имеем

$$(1 \ 2 \ 3) = (1 \ 3)(1 \ 2) = (2 \ 3)(1 \ 3) = (1 \ 3)(2 \ 4)(1 \ 2)(1 \ 4).$$

3. Знак перестановки. Справедлива следующая важная
Теорема 2. Пусть π — перестановка из S_n ,

$$\pi = \tau_1 \tau_2 \dots \tau_k \quad (6)$$

— произвольное разложение π в произведение транспозиций.
Тогда число

$$\varepsilon_\pi = (-1)^k, \quad (7)$$

называемое знаком π (иначе: сигнатурой или чётностью), полностью определяется перестановкой π и не зависит от способа разложения (6), т.е. чётность целого числа k для данной перестановки π всегда одна и та же. Кроме того,

$$\varepsilon_{\alpha\beta} = \varepsilon_\alpha \varepsilon_\beta \quad (8)$$

для всех $\alpha, \beta \in S_n$.

Доказательство. 1) Предположим, что наряду с (6) мы имеем также разложение

$$\pi = \tau'_1 \tau'_2 \dots \tau'_{k'}, \quad (6')$$

причём четности k и k' различны. Это значит, что целое число $k + k'$ нечётно. Так как $(\tau'_s)^2 = e$, то, последовательно умножая справа обе части равенства $\tau_1 \tau_2 \dots \tau_k = \tau'_1 \tau'_2 \dots \tau'_{k'}$, вытекающего из (6) и (6'), на $\tau_{k'}, \dots, \tau'_2, \tau'_1$, получим $\tau_1 \tau_2 \dots \tau_k \tau_{k'} \dots \tau'_2 \tau'_1 = e$. Мы свели нашу задачу к следующей. Пусть

$$e = \sigma_1 \sigma_2 \dots \sigma_{m-1} \sigma_m, \quad m > 0, \quad (9)$$

— запись единичной перестановки в виде произведения $m > 0$ транспозиций. Нужно показать, что обязательно m — чётное число.

С этой целью будет установлено, что от записи (9) мы можем перейти к записи e в виде произведения $m-2$ транспозиций. Продолжив этот спуск, мы пришли бы при нечётном m к одной транспозиции τ . Но, очевидно, $e \neq \tau$. Итак, нам нужно обосновать спуск в (9) от m к $m-2$ множителям.

2) Пусть s , $1 \leq s \leq n$, — любое фиксированное натуральное число, входящее в одну из транспозиций $\sigma_2, \dots, \sigma_m$. Для определённости считаем, что

$$e = \sigma_1 \dots \sigma_{p-1} \sigma_p \sigma_{p+1} \dots \sigma_m,$$

где $\sigma_p = (st)$, а $\sigma_{p+1}, \dots, \sigma_m$ не содержат s . Для σ_{p-1} имеются четыре возможности:

a) $\sigma_{p-1} = (st)$; тогда отрезок $\sigma_{p-1} \sigma_p = (st)(st)$ из записи e удаляется, и мы приходим к $m-2$ транспозициям;

b) $\sigma_{p-1} = (sr)$, $r \neq s, t$; здесь

$$\sigma_{p-1} \sigma_p = (sr)(st) = (st)(rt),$$

и мы сдвинули вхождение s на одну позицию влево, не изменив m ;

в) $\sigma_{p-1} = (tr)$, $r \neq s, t$; здесь

$$\sigma_{p-1}\sigma_p = (tr)(st) = (sr)(tr),$$

и снова, как в случае б), произошёл сдвиг s влево без изменения t ;
г) $\sigma_{p-1} = (qr)$, $\{q, r\} \cap \{s, t\} = \emptyset$; здесь

$$\sigma_{p-1}\sigma_p = (qr)(st) = (st)(qr).$$

В случае а) наша цель достигнута. В случаях б)–г) повторяем процесс, сдвигая вхождение s на одну позицию влево. В конечном счёте мы придем либо к случаю а), либо к экстремальному случаю, когда $e = \sigma'_1\sigma'_2 \dots \sigma'_m$, причём $\sigma'_1 = (st')$ и s не имеет вхождений в $\sigma'_2, \dots, \sigma'_m$. Значит, $\sigma'_k(s) = s$ при $k > 1$ и $s = e(s) = \sigma'_1(s) = t' \neq s$. Полученное противоречие доказывает утверждение об инвариантности ε_{pi} .

3) Если $\alpha = \tau_1 \dots \tau_k$, $\beta = \tau_{k+1} \dots \tau_{k+l}$, то $\alpha\beta = \tau_1 \dots \tau_k \tau_{k+1} \dots \tau_{k+l}$ и $\varepsilon_\alpha = (-1)^k$, $\varepsilon_\beta = (-1)^l$, $\varepsilon_{\alpha\beta} = (-1)^{k+l} = (-1)^k(-1)^l = \varepsilon_\alpha \varepsilon_\beta$. \square

Определение Перестановка $\pi \in S_n$ называется чётной, если $\varepsilon_\pi = 1$, и нечётной, если $\varepsilon_\pi = -1$.

Из определения вытекает, что все транспозиции — нечётные перестановки, а $\varepsilon_e = 1$.

Следствие. Пусть перестановка $\pi \in S_n$ разложена в произведение независимых циклов длины l_1, l_2, \dots, l_m . Тогда

$$\varepsilon_\pi = (-1)^{\sum_{k=1}^m (l_k - 1)}.$$

Действительно, по теореме 2 имеем

$$\varepsilon_\pi = \varepsilon_{\pi_1 \dots \pi_m} = \varepsilon_{\pi_1} \dots \varepsilon_{\pi_m}.$$

Кроме того, $\varepsilon_{\pi_k} = (-1)^{l_k - 1}$, поскольку π_k записывается в виде произведения $l_k - 1$ транспозиций (см. доказательство следствия теоремы 1). Окончательно

$$\varepsilon_\pi = (-1)^{l_1 - 1} \dots (-1)^{l_m - 1} = (-1)^{\sum_{k=1}^m (l_k - 1)}. \quad \square$$

Пример. $\pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 \\ 5 & 4 & 6 & 7 & 2 & 9 & 1 & 3 & 8 & 11 & 10 \end{pmatrix}$. Имеем $\pi = (1\ 5\ 2\ 4\ 7)(3\ 6\ 9\ 8)(10\ 11)$, откуда $l_1 = 5$, $l_2 = 4$, $l_3 = 2$ и $\varepsilon_\pi = (-1)^{4+3+1} = 1$.

Запишем S_n в виде объединения $S_n = A_n \cup \bar{A}_n$, где

$$A_n = \{\pi \in S_n \mid \varepsilon_\pi = 1\}$$

— множество всех чётных перестановок, $\bar{A}_n = S_n \setminus A_n$ — множество нечётных перестановок. Пусть $\tau = (ij)$ — любая транспозиция. Отображение S_n в себя, определённое правилом $L_\tau : \pi \mapsto \tau\pi$, биективно. (Оно инъективно: $\tau\alpha = \tau\beta \Rightarrow \alpha = \beta$; далее применить теорему 3 из § 5. Можно просто заметить, что L_τ^2 — единичное отображение и $L_\tau^{-1} = L_\tau$.) Для наглядности изобразим L_τ в виде перестановки

степени $N = n!$ на множестве $S_n = \{\pi_1 = e, \pi_2, \pi_3, \dots, \pi_N\}$:

$$L_\tau = \begin{pmatrix} \pi_1 & \pi_2 & \pi_3 & \dots & \pi_N \\ \tau\pi_1 & \tau\pi_2 & \tau\pi_3 & \dots & \tau\pi_N \end{pmatrix}. \quad (10)$$

Аналогично,

$$R_\tau = \begin{pmatrix} \pi_1 & \pi_2 & \pi_3 & \dots & \pi_N \\ \pi_1\tau & \pi_2\tau & \pi_3\tau & \dots & \pi_N\tau \end{pmatrix}. \quad (10')$$

— перестановка на S_n . Отображения (10) и (10') будут использоваться нами впоследствии, причём даже в более общем контексте. А сейчас заметим, что $\varepsilon_{\tau\pi} = \varepsilon_\tau \varepsilon_\pi = -\varepsilon_\pi$, поэтому

$$L_\tau(A_n) = \bar{A}_n, \quad L_\tau(\bar{A}_n) = A_n.$$

Значит, число чётных перестановок в S_n совпадает с числом нечётных перестановок, откуда

$$|A_n| = \frac{1}{2}|S_n| = \frac{n!}{2}. \quad (11)$$

4. Действие S_n на функциях. К важному понятию знака перестановки $\sigma \in S_n$ можно подойти несколько иначе, подсчитывая число так называемых σ -инверсий (см. упр. 5 в конце параграфа). Но вместо этого мы дадим сейчас альтернативное доказательство теоремы 2, которое опирается на понятие кососимметрической функции, важное само по себе и полезное для дальнейшего.

Определение. Пусть $\pi \in S_n$ и f — функция от любых n аргументов. Полагаем

$$(\pi \circ f)(x_1, \dots, x_n) = f(x_{\pi 1}, \dots, x_{\pi n}). \quad (12)$$

Говорят, что функция $g = \pi \circ f$ получается действием π на f .

Лемма 1. Пусть α, β — любые перестановки из S_n . Тогда

$$(\alpha\beta) \circ f = \alpha \circ (\beta \circ f).$$

Доказательство. В соответствии с определяющим соотношением (12) имеем

$$(\alpha \circ (\beta \circ f))(x_1, \dots, x_n) = (\beta \circ f)(x_{\alpha 1}, \dots, x_{\alpha n}),$$

или, полагая $y_k = x_{\alpha k}$ и замечая, что $y_{\beta i} = x_{\alpha(\beta i)}$,

$$(\alpha \circ (\beta \circ f))(x_1, \dots, x_n) = (\beta \circ f)(y_1, \dots, y_n) =$$

$$= f(y_{\beta 1}, \dots, y_{\beta n}) = f(x_{\alpha(\beta 1)}, \dots, x_{\alpha(\beta n)}) =$$

$$= f(x_{(\alpha\beta)1}, \dots, x_{(\alpha\beta)n}) = (\alpha\beta) \circ f)(x_1, \dots, x_n). \quad \square$$

Определение. Функция f от n аргументов называется *кососимметрической*, если

$$f(\dots, x_k, x_{k+1}, \dots) = -f(\dots, x_{k+1}, x_k, \dots),$$

т.е. при перестановке местами любых двух соседних аргументов значение f меняет знак на противоположный.

Лемма 2. *При перестановке местами любых двух аргументов кососимметрическая функция меняет знак на противоположный.*

Доказательство. Пусть переставлены i -й и j -й аргументы, причём $i < j$. Проводим индукцию по числу $l = j - i - 1$ аргументов между переставляемой парой. При $l = 0$ утверждение леммы совпадает с определением кососимметрической функции. Пусть лемма верна при всех $j - i - 1 < l$. Тогда

$$\begin{aligned} f(\dots, x_i, x_{i+1}, \dots, x_{j-1}, x_j, \dots) &= \\ &= -f(\dots, x_{i+1}, x_i, \dots, x_{j-1}, x_j, \dots) = \\ &= f(\dots, x_{i+1}, x_j, \dots, x_{j-1}, x_i, \dots) = \\ &= -f(\dots, x_j, x_{i+1}, \dots, x_{j-1}, x_i, \dots). \quad \square \end{aligned}$$

Надо ещё быть уверенным в том, что не все кососимметрические функции тождественно равны нулю. Простейшим является следующий пример.

Пример. Пусть

$$\Delta_n = \Delta_n(x_1, x_2, \dots, x_n) = \prod_{1 \leq j < i \leq n} (x_i - x_j).$$

Символ \prod при записи произведения играет ту же роль, что и \sum при записи суммы. Выделив любые два рядом стоящих аргумента x_k, x_{k+1} , будем иметь

$$\Delta_n = (x_{k+1} - x_k)[(x_{k+1} - x_{k-1}) \dots (x_{k+1} - x_1)(x_k - x_{k-1}) \dots (x_k - x_1)] \cdot A \cdot B,$$

где

$$A = \prod_{1 \leq j < i < k} (x_i - x_j),$$

$$B = \prod_{s=k+2}^n [(x_s - x_{s-1}) \dots (x_s - x_{k+1})(x_s - x_k) \dots (x_s - x_1)].$$

При перестановке местами x_k и x_{k+1} множители

$$[(x_{k+1} - x_{k-1}) \dots (x_{k+1} - x_1) \cdot (x_k - x_{k-1}) \dots (x_k - x_1)],$$

A и B , очевидно, не меняют своих значений, в то время как

$$(x_k - x_{k+1}) = -(x_{k+1} - x_k).$$

Это и значит, что

$$\Delta_n(\dots, x_k, x_{k+1}, \dots) = -\Delta_n(\dots, x_{k+1}, x_k, \dots), \quad 1 \leq k \leq n-1.$$

По лемме 2 имеем также

$$\Delta_n(\dots, x_i, \dots, x_j, \dots) = -\Delta_n(\dots, x_j, \dots, x_i, \dots).$$

Кроме того, $\Delta_n(x_1, \dots, x_n) \neq 0$ при попарно различных x_1, \dots, x_n .

Второе доказательство теоремы 2. Возьмём произвольную кососимметрическую функцию f от n аргументов x_1, \dots, x_n . По лемме 1 действие $\pi = \tau_1 \tau_2 \dots \tau_k$ на f сводится к последовательному применению транспозиций $\tau_k, \tau_{k-1}, \dots, \tau_1$, т.е. к k -кратному умножению f на -1 :

$$\pi \circ f = (\tau_1 \dots \tau_{k-1}) \circ (\tau_k \circ f) = -(\tau_1 \dots \tau_{k-1}) \circ f = \dots = (-1)^k f = \epsilon_\pi f.$$

Так как левая часть этого соотношения зависит от π , но не от какого-либо его разложения, то и отображение $\epsilon : \pi \mapsto \epsilon_\pi$, заданное равенством (7), должно полностью определяться перестановкой π при условии, конечно, что f не тождественно равна нулю функция. В качестве такой функции можно взять, например, только что рассмотренную функцию $f = \Delta_n$.

Применение к такой функции f перестановки $\alpha\beta$ по правилу, изложенному в лемме 1, дает

$$\begin{aligned} \epsilon_{\alpha\beta} f &= (\alpha\beta) \circ f = \alpha \circ (\beta \circ f) = \alpha \circ (\epsilon_\beta f) = \epsilon_\beta (\alpha \circ f) = \\ &= \epsilon_\beta (\epsilon_\alpha f) = (\epsilon_\beta \epsilon_\alpha) f, \end{aligned}$$

откуда получается соотношение (8). \square

Замечание. К действию S_n на функциях мы будем обращаться неоднократно, а в [ВА III] увидим, что это лишь частное проявление гораздо более общей закономерности. Пока наше маленькое достижение заключается в том, что словесное выражение “поменяем в $f(x_1, \dots, x_n)$ местами x_i и x_j ” сведено к символьной записи $\tau \circ f$ с транспозицией $\tau = (ij)$.

УПРАЖНЕНИЯ

1. В курсе математического анализа доказывается формула Стирлинга

$$n! \sim \sqrt{2\pi n} n^n e^{-n},$$

где $e = 2,718281\dots$ — основание натурального логарифма, $\pi = 3,141592\dots$; символ \sim здесь означает, что отношение $\sqrt{2\pi n} n^n e^{-n} / n!$ стремится к 1 при $n \rightarrow \infty$.

При помощи формулы Стирлинга, дающей приближение с недостатком, проверить, что $100! > (9,33\dots)10^{157}$. Сколько в S_{100} циклов длины 100?

2. Найти порядок перестановки (4) и перестановки

$$\pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 3 & 6 & 8 & 2 & 1 & 4 & 5 & 7 \end{pmatrix}.$$

3. Перестановка π вида (3) с m независимыми циклами оставляет

$$m' = n - \sum_{k=1}^m l_k$$

символов (или точек) на месте. Число $d(\pi) = n - (m + m')$ называется *декрементом* перестановки π . Проверить, что $\epsilon_\pi = (-1)^{d(\pi)}$.

4. Найти знак перестановки

$$\pi = \begin{pmatrix} 1 & 2 & 3 & \dots & n-1 & n \\ n & n-1 & n-2 & \dots & 2 & 1 \end{pmatrix}.$$

5. Пусть $\Omega = \{1, 2, \dots, n\}$, $\Omega \times \Omega$ — декартов квадрат. Будем называть пару $(i, j) \in \Omega \times \Omega$ инверсией относительно перестановки $\sigma \in S_n$ (или, короче: σ -инверсией), если $i < j$, но $\sigma(i) > \sigma(j)$. Положим

$$\operatorname{sgn} \sigma = \prod_{1 \leq i < j \leq n} \frac{\sigma(j) - \sigma(i)}{j - i}.$$

Так как $(\sigma(j) - \sigma(i))/(j - i)$ — отличное от нуля рациональное число, являющееся отрицательным в точности тогда, когда (i, j) будет σ -инверсией, и так как $\sigma : \Omega \rightarrow \Omega$ — биективное отображение, то $\operatorname{sgn} \sigma = (-1)^k$, где k — общее число σ -инверсий. Если $\tau = (ij)$ — транспозиция, то $\operatorname{sgn} \tau = -1$. Как легко видеть,

$$(\sigma(j) \sigma(i))\sigma =$$

$$= \begin{pmatrix} \dots & \sigma(j) & \dots & \sigma(i) & \dots \\ \dots & \sigma(i) & \dots & \sigma(j) & \dots \end{pmatrix} \begin{pmatrix} \dots & i & \dots & j & \dots \\ \dots & \sigma(i) & \dots & \sigma(j) & \dots \end{pmatrix} = \\ = \begin{pmatrix} \dots & i & \dots & j & \dots \\ \dots & \sigma(j) & \dots & \sigma(i) & \dots \end{pmatrix}.$$

так что σ -инверсия (i, j) перестает быть инверсией относительно перестановки $\tau\sigma$, где $\tau = (\sigma(j) \sigma(i))$ — транспозиция.

Показать, что найдутся k транспозиций τ_1, \dots, τ_k , для которых

$$\tau_k \tau_{k-1} \dots \tau_1 \sigma = e$$

— единичная перестановка. Стало быть, $\sigma = \tau_1 \dots \tau_{k-1} \tau_k$ и $\operatorname{sgn} \sigma = (-1)^k = \varepsilon_\sigma$ — два равноправных обозначения одного и того же инварианта перестановки; sgn (от *signum* (лат.)) — знак. Мы получили еще один удобный способ определения знака перестановки. Скажем, относительно перестановки (4) множество инверсий состоит из пяти пар $(1, 5), (2, 5), (3, 5), (4, 5), (6, 7)$, так что $\operatorname{sgn} \pi = -1$. Практически дело сводится к подсчету в нижней строке перестановки π количества чисел j , больших i , но стоящих перед i , для $i = 1, 2, \dots, n-1$.

§ 9. Арифметика целых чисел

Задачей этого параграфа является краткое описание тех простейших свойств делимости целых чисел, на которые удобно по разным поводам ссылаться в дальнейшем. Дополнительные факты будут приведены в гл. 5, где теория делимости переносится на более общие алгебраические системы.

1. Основная теорема арифметики. Целое число s называется делителем (или множителем) целого числа n , если $n = st$ для некоторого $t \in \mathbb{Z}$. В свою очередь n называется кратным s . Делимость n на s обозначается символом $s|n$, а отрицание делимости — символом $s \not| n$. Делимость — транзитивное отношение на \mathbb{Z} . Если, далее,

$t|n$ и $n|m$, то $n = \pm t$, и целые числа n, m называются *ассоциированными*. Целое число p , делители которого исчерпываются числами $\pm p, \pm 1$ (*не собственные делители*), называется *простым*. Обычно в качестве простых берутся положительные простые числа > 1 .

Фундаментальную роль простых чисел вскрывает

Основная теорема арифметики. Каждое положительное целое число $n \neq 1$ может быть записано в виде произведения простых чисел: $n = p_1 p_2 \dots p_s$. Эта запись единственна с точностью до порядка множителей.

Собрав вместе одинаковые простые множители и изменив обозначения, получим запись

$$n = p_1^{\varepsilon_1} p_2^{\varepsilon_2} \dots p_k^{\varepsilon_k}, \quad \varepsilon_i > 0, \quad 1 \leq i \leq k.$$

Для любого рационального числа $a = n/m \in \mathbb{Q}$ имеет место аналогичное разложение, но с показателями ε_i , как положительными, так и отрицательными.

Заметим, что множество

$$P = \{2, 3, 5, 7, 11, 13, \dots\}$$

всех простых чисел бесконечно (теорема Евклида). Действительно, если бы существовало лишь конечное число простых чисел, скажем, p_1, p_2, \dots, p_t , то по основной теореме число $c = p_1 p_2 \dots p_t + 1$ делилось бы по крайней мере на одно из p_i . Без ограничения общности считаем $c = p_1 c'$. Тогда $p_1(c' - p_2 \dots p_t) = 1$, а это невозможно, поскольку делителями единицы в \mathbb{Z} являются лишь ± 1 . \square

Доказательство основной теоремы откладывается до гл. 5. На первый взгляд, её вообще не надо доказывать, настолько она кажется очевидной. Между тем, хотя речь идёт о мультипликативных свойствах (свойствах делимости) целых чисел, основную теорему невозможно доказать, не используя одновременно операций умножения и сложения в \mathbb{Z} .

В качестве иллюстрации этого утверждения рассмотрим в \mathbb{N} подмножество

$$S = \{4k + 1 \mid k = 0, 1, 2, \dots\}.$$

Оно замкнуто относительно умножения:

$$(4k_1 + 1)(4k_2 + 1) = 4k_3 + 1.$$

Индукцией по $n \in S$ нетрудно установить существование разложения (первая часть основной теоремы) $n = q_1 \dots q_s$, где q_i — далее неразложимые элементы из S . Мы назовём их *квазипростыми числами*. Выпишем несколько таких чисел: 5, 9, 13, 17, 21, 49.

Вторая часть основной теоремы для системы S неверна, поскольку, например, число $441 \in S$ имеет два существенно разных разложения в произведение квазипростых чисел: $441 = 9 \cdot 49 = 21^2$.

2. НОД и НОК в \mathbb{Z} . Любые два целых числа n и m можно записать в виде произведения степеней одних и тех же простых чисел

$$n = \pm p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_k^{\alpha_k}, \quad m = \pm p_1^{\beta_1} p_2^{\beta_2} \cdots p_k^{\beta_k},$$

если условиться допускать нулевые показатели (как всегда, считая $p_i^0 = 1$). Введём в рассмотрение два целых числа

$$\begin{aligned} \text{НОД}(n, m) &= p_1^{\gamma_1} p_2^{\gamma_2} \cdots p_k^{\gamma_k}, \\ \text{НОК}(n, m) &= p_1^{\delta_1} p_2^{\delta_2} \cdots p_k^{\delta_k}, \end{aligned} \tag{1}$$

где $\gamma_i = \min(\alpha_i, \beta_i)$, $\delta_i = \max(\alpha_i, \beta_i)$, $i = 1, 2, \dots, k$.

Так как $d|n \implies d = \pm p_1^{\alpha'_1} \cdots p_k^{\alpha'_k}$, $0 \leq \alpha'_i \leq \alpha_i$, то из (1) вытекают следующие утверждения.

- i) $\text{НОД}(n, m)|n$, $\text{НОД}(n, m)|m$, и если $d|n$, $d|m$, то $d|\text{НОД}(n, m)$.
- ii) $n|\text{НОК}(n, m)$, $m|\text{НОК}(n, m)$, и если $n|u$, $m|u$, то $\text{НОК}(n, m)|u$.

Свойства i) и ii) оправдывают сокращённые обозначения НОД и НОК наибольшего общего делителя и наименьшего общего кратного целых чисел n , m . При $n > 0$, $m > 0$ выполнено соотношение

$$\text{НОД}(n, m) \cdot \text{НОК}(n, m) = nm. \tag{2}$$

Целые числа n , m называются *взаимно простыми*, если $\text{НОД}(n, m) = 1$. В этом случае соотношение (2) принимает вид $\text{НОК}(n, m) = nm$.

3. Алгоритм деления в \mathbb{Z} . При заданных $a, b \in \mathbb{Z}$, $b > 0$, всегда найдутся $q, r \in \mathbb{Z}$ такие, что

$$a = bq + r, \quad 0 \leq r < b$$

(если считать лишь $b \neq 0$, то будет выполнено неравенство $0 \leq r < |b|$).

В самом деле, множество

$$S = \{a - bs \mid s \in \mathbb{Z}, a - bs \geq 0\},$$

очевидно, непусто (например, $a - b(-a^2) > 0$). Стало быть, S содержит наименьший элемент; обозначим его $r = a - bq$. По условию $r \geq 0$. Предположив $r \geq b$, мы получили бы элемент $r - b = a - b(q+1) \in S$, меньший, чем r . Это противоречие устраивается лишь при $r < b$. \square

Проведённое несложное рассуждение даёт также предписание, т.е. *алгоритм* (или *алгорифм*), для нахождения *частного* b и *остатка* r в конечное число шагов. Алгоритм деления в \mathbb{Z} используется для иного определения НОД, а следовательно, и НОК, если принять во внимание соотношение (2).

Именно, при заданных целых числах n , m , одновременно не равных нулю, положим

$$J = \{nu + mv \mid u, v \in \mathbb{Z}\}. \tag{3}$$

Выберем в J наименьший положительный элемент $d = nu_0 + mv_0$. Используя алгоритм деления, запишем $n = dq + r$, $0 \leq r < d$. Ввиду выбора d включение

$$r = n - dq = n - (nu_0 + mv_0)q = n(1 - u_0q) + m(-v_0q) \in J$$

влечёт равенство $r = 0$. Стало быть, $d|n$. Аналогично доказывается, что $d|m$. Пусть теперь d' — любой делитель чисел n и m . Тогда

$$d'|n, d'|m \Rightarrow d'|nu_0, d'|mv_0 \Rightarrow d'|(nu_0 + mv_0) \Rightarrow d'|d.$$

Итак, d обладает всеми свойствами наибольшего общего делителя, и поэтому $d = \text{НОД}(n, m)$. Мы приходим к следующему утверждению.

Наибольший общий делитель двух целых чисел n, m , не равных одновременно нулю, всегда записывается в виде

$$\text{НОД}(n, m) = nu + mv, \quad u, v \in \mathbb{Z}. \quad (4)$$

В частности, целые числа n, m взаимно просты тогда и только тогда, когда

$$nu + mv = 1 \quad (4')$$

при некоторых $u, v \in \mathbb{Z}$.

Было проверено, что взаимная простота n, m влечёт соотношение (4'). Обратно: если n, m таковы, что имеет место (4'), то

$$d|n, d|m \Rightarrow d|nu, d|mv \Rightarrow d|(nu + mv) \Rightarrow d|1 \Rightarrow d = \pm 1. \quad \square$$

Доказательство соотношений (4) и (4') довольно эффективно. Нужно взять любой положительный элемент из множества J (см. (3)), а затем уменьшать его при помощи алгоритма деления до тех пор, пока не получится наименьший элемент, который и будет наибольшим общим делителем.

УПРАЖНЕНИЯ

1. Каждое нечётное простое число имеет вид $4k + 1$ или $4k - 1$. Используя мультипликативность множества S из п. 1, доказать бесконечность множества простых чисел вида $4k - 1$.

2. Доказать, что существует бесконечно много простых чисел вида $4k + 1$, опираясь на следующее нетривиальное утверждение.

Если $n, m \in \mathbb{Z}$, $\text{НОД}(n, m) = 1$, и если p — простое число, делящее $n^2 + m^2$, то $p = 4k + 1$.

3. Если натуральное число n делится в точности на r различных простых чисел p_1, \dots, p_r , то количество чисел, меньших n и взаимно простых с n , равно

$$\varphi(n) = n \left(1 - \frac{1}{p_1}\right) \dots \left(1 - \frac{1}{p_r}\right).$$

Функция $\varphi: \mathbb{N} \rightarrow \mathbb{N}$ называется функцией Эйлера.

Проверить справедливость формулы для значений $\varphi(n)$ при $n \leq 25$ и при $n = p^m$.

4. Используя биномиальную формулу, индукцией по n доказать, что если p — простое число, то $n^p - n$ делится на p при любом $n \in \mathbb{Z}$.

Глава 2

МАТРИЦЫ

Прямоугольные матрицы, введённые в § 3 гл. 1, встречаются настолько часто, что с течением времени возник самостоятельный раздел математики — *теория матриц*. Её становление относят к середине прошлого века, но полноту и изящество она приобрела позднее, вместе с развитием линейной алгебры. До сих пор теория матриц остаётся важным инструментом исследования, хорошо приспособленным и к запросам практики, и к абстрактным конструкциям современной математики. Здесь будут изложены простейшие результаты теории матриц.

Матрицы являются естественными спутниками линейных отображений векторных пространств. В курсе линейной алгебры и геометрии [ВА II] этому утверждению будет придан точный смысл. В настоящей главе понятия пространства, вектора, линейной зависимости, ранга системы и т.п. развиваются в чисто алгебраическом аспекте и ровно настолько, насколько они необходимы для наших непосредственных целей.

§ 1. Векторные пространства строк и столбцов

1. **Мотивировка.** В связи с системами линейных уравнений нам приходилось рассматривать строки длины n , в которые вкладывался разный смысл. Это были строки $(a_{i1}, a_{i2}, \dots, a_{in})$, $1 \leq i \leq m$, матрицы $A = (a_{ij})$ размера $m \times n$ и решения $(x_1^0, x_2^0, \dots, x_n^0)$ линейной системы с матрицей A . Приведение в § 3 гл. 1 системы или матрицы к ступенчатому виду включало, помимо элементарного преобразования типа (I), два важных акта: умножение строки на число и сложение двух строк. Те же действия можно производить и с решениями однородной линейной системы. Действительно, если $(x'_1, x'_2, \dots, x'_n)$ и $(x''_1, x''_2, \dots, x''_n)$ — два решения системы

$$a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n = 0, \quad i = 1, 2, \dots, m,$$

а α, β — два любых вещественных числа, то строка

$$(\alpha x'_1 + \beta x''_1, \alpha x'_2 + \beta x''_2, \dots, \alpha x'_n + \beta x''_n)$$

тоже будет решением нашей системы:

$$\begin{aligned} a_{i1}(\alpha x'_1 + \beta x''_1) + a_{i2}(\alpha x'_2 + \beta x''_2) + \dots + a_{in}(\alpha x'_n + \beta x''_n) &= \\ = \alpha(a_{i1}x'_1 + a_{i2}x'_2 + \dots + a_{in}x'_n) + \beta(a_{i1}x''_1 + a_{i2}x''_2 + \dots + a_{in}x''_n) &= 0. \end{aligned}$$

С другой стороны, любая строка, что бы она ни выражала, является элементом "универсального" множества \mathbb{R}^n , т.е. n -й декартовой степени множества \mathbb{R} вещественных чисел. Поэтому желательно изучить общий объект, свойства которого автоматически переносились бы на матрицы и на решения однородных систем.

2. Основные определения. Пусть n — какое-то фиксированное натуральное число. Векторным пространством строк длины n над \mathbb{R} называется множество \mathbb{R}^n (его элементами являются векторы-строки или просто векторы), рассматриваемое вместе с операциями сложения векторов и умножения их на скаляры — вещественные числа. Скаляры обозначаются строчными буквами латинского или греческого алфавита, а векторы — заглавными латинскими буквами, как матрицы. По существу на вектор $X = (x_1, x_2, \dots, x_n)$ можно смотреть как на $1 \times n$ -матрицу. Пусть $Y = (y_1, y_2, \dots, y_n)$ — еще один вектор, λ — скаляр. По определению

$$X + Y = (x_1 + y_1, x_2 + y_2, \dots, x_n + y_n), \\ \lambda X = (\lambda x_1, \lambda x_2, \dots, \lambda x_n).$$

Нулевой вектор $(0, 0, \dots, 0)$ обозначается в дальнейшем обычным символом нуля 0. Далее, \mathbb{R}^1 принято отождествлять с \mathbb{R} .

Формальные правила действий с вещественными числами, безусловно, известные читателю, переносятся на \mathbb{R}^n . Их перечисление, хотя и скучное, даёт точное представление о том, что следует понимать под *абстрактным векторным пространством*, которое изучается в более позднем курсе линейной алгебры и геометрии:

ВП₁: $X + Y = Y + X$ для любых векторов $X, Y \in \mathbb{R}^n$ (*закон коммутативности*);

ВП₂: $(X + Y) + Z = X + (Y + Z)$ для любых трех векторов $X, Y, Z \in \mathbb{R}^n$ (*закон ассоциативности*);

ВП₃: существует специальный (нулевой) вектор 0 такой, что $X + 0 = X$ для всех $X \in \mathbb{R}^n$;

ВП₄: каждому $X \in \mathbb{R}^n$ отвечает противоположный (или обратный) вектор $-X$ такой, что $X + (-X) = 0$;

ВП₅: $1X = X$ для всех $X \in \mathbb{R}^n$;

ВП₆: $(\alpha\beta)X = \alpha(\beta X)$ для всех $\alpha, \beta \in \mathbb{R}, X \in \mathbb{R}^n$;

ВП₇: $(\alpha + \beta)X = \alpha X + \beta X$;

ВП₈: $\alpha(X + Y) = \alpha X + \alpha Y$.

Единственность векторов 0 и $-X$, о которых говорится в ВП₃ и ВП₄, равно как и другие простые следствия из указанных правил (или аксиом, если имеется в виду абстрактное векторное пространство), мы не будем выводить, считая их достаточно прозрачными.

Происхождение термина "векторное (или еще линейное) пространство" разъясняется в курсе аналитической геометрии (читаемом также в первом семестре), где устанавливается взаимно однозначное

соответствие между точками (векторами) пространства — декартовой плоскости — и их координатами (x, y) . Сложению векторов по правилу параллелограмма и умножению их на число соответствуют как раз действия с векторами-строками в \mathbb{R}^2 .

Наряду с векторным пространством строк длины n рассматривается также векторное пространство столбцов высоты n

$$\begin{vmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{vmatrix} = [x_1, x_2, \dots, x_n],$$

как мы их условились обозначать в § 3 гл. 1. Понятно, что различие между пространствами строк и столбцов чисто условное, но мы вскоре убедимся, что полезно иметь оба варианта пространства. Из контекста обычно ясно, о каких векторах, столбцах или строках идёт речь, поэтому никаких специальных обозначений не вводится.

3. Линейные комбинации. Линейная оболочка. Пусть X_1, X_2, \dots, X_k — векторы пространства \mathbb{R}^n и $\alpha_1, \alpha_2, \dots, \alpha_k$ — скаляры. Вектор $X = \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_k X_k$ называется *линейной комбинацией* векторов X_i с коэффициентами α_i . Например,

$$(2, 3, 5, 5) - 3(1, 1, 1, 1) + 2(1, 0, -1, -1) = (1, 0, 0, 0).$$

Пусть, далее, $Y = \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$ — линейная комбинация тех же векторов X_i с коэффициентами β_i , а $\alpha, \beta \in \mathbb{R}$. Тогда

$$\alpha X + \beta Y =$$

$$\begin{aligned} &= \alpha(\alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_k X_k) + \beta(\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k) = \\ &= (\alpha\alpha_1 + \beta\beta_1)X_1 + (\alpha\alpha_2 + \beta\beta_2)X_2 + \dots + (\alpha\alpha_k + \beta\beta_k)X_k \end{aligned}$$

— снова линейная комбинация векторов X_i с коэффициентами $\alpha\alpha_i + \beta\beta_i$. Мы видим, что множество V всех линейных комбинаций данной системы векторов X_1, X_2, \dots, X_k обладает свойством

$$X, Y \in V \implies \alpha X + \beta Y \in V \tag{1}$$

для всех $\alpha, \beta \in \mathbb{R}$. В частности, нулевой вектор всегда содержиться в V .

Обычно V обозначают символом $\langle X_1, X_2, \dots, X_k \rangle$ и называют *линейной оболочкой* (или просто *оболочкой*) системы векторов X_1, X_2, \dots, X_k . Говорят ещё, что оболочка $\langle X_1, X_2, \dots, X_k \rangle$ *натянута на* X_1, X_2, \dots, X_k или *порождена* векторами X_1, X_2, \dots, X_k .

Можно определить линейную оболочку любого подмножества $S \subset \mathbb{R}^n$, понимая под $\langle S \rangle$ совокупность всех линейных комбинаций конечных систем векторов из S . Ясно, что если V — линейная оболочка в \mathbb{R}^n , то $\langle V \rangle = V$: любая линейная комбинация векторов

из V принадлежит V . В частности, $S \subset V \Rightarrow \langle S \rangle \subset V$, т.е. линейную оболочку $\langle S \rangle$ можно определить как пересечение всех оболочек, содержащих данное множество S векторов из \mathbb{R}^n :

$$\langle S \rangle = \bigcap_{S \subset V} V. \quad (2)$$

На первый взгляд не очевидно, что стоящее в правой части (2) пересечение $\cap V$ какого-то множества оболочек будет линейной оболочкой. Но если $X, Y \in \cap V$, то $X, Y \in V$ для каждой оболочки V , входящей в множество. Значит, $\alpha X + \beta Y \in V$ для всех $\alpha, \beta \in \mathbb{R}$, а это и даёт нужное включение $\alpha X + \beta Y \in \cap V$. Напротив, объединение $U \cup V$ оболочек U и V , вообще говоря, не является оболочкой, как показывает хотя бы пример $U = \{(\lambda, 0) \mid \lambda \in \mathbb{R}\}$, $V = \{(0, \lambda) \mid \lambda \in \mathbb{R}\}$ в \mathbb{R}^2 .

Рассмотрим два общих примера.

Пример 1. Пусть

$$U_m = \{(\lambda_1, \dots, \lambda_m, 0, \dots, 0) \mid \lambda_i \in \mathbb{R}\} \subset \mathbb{R}^n,$$

$$V_m = \{(0, \dots, 0, \lambda_{m+1}, \dots, \lambda_n) \mid \lambda_i \in \mathbb{R}\} \subset \mathbb{R}^n,$$

$0 < m < n$. Непосредственно проверяется, что U_m, V_m — линейные оболочки, причём $\langle U_m, V_m \rangle = \mathbb{R}^n$ и $U_m \cap V_m = \{0\}$.

Пример 2. В пространстве \mathbb{R}^n рассмотрим так называемые единичные векторы-строки

$$E_{(1)} = (1, 0, \dots, 0), \quad E_{(2)} = (0, 1, \dots, 0), \quad \dots, \quad E_{(n)} = (0, 0, \dots, 1). \quad (3)$$

Каждый вектор $X = (x_1, x_2, \dots, x_n)$ однозначно записывается в виде $X = x_1 E_{(1)} + x_2 E_{(2)} + \dots + x_n E_{(n)}$. Поэтому

$$\mathbb{R}^n = \langle E_{(1)}, E_{(2)}, \dots, E_{(n)} \rangle.$$

Единичные векторы-столбцы будем обозначать символами

$$E^{(1)} = [1, 0, \dots, 0], \quad E^{(2)} = [0, 1, \dots, 0], \quad \dots, \quad E^{(n)} = [0, 0, \dots, 1]. \quad (3')$$

4. Линейная зависимость. Система векторов X_1, \dots, X_k пространства \mathbb{R}^n называется линейно зависимой, если найдутся k чисел $\alpha_1, \dots, \alpha_k$, одновременно не равных нулю и таких, что

$$\alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_k X_k = 0 \quad (4)$$

(справа стоит нулевой вектор). Будем говорить также, что линейная зависимость (4) нетривиальна. Если же $\alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_k X_k = 0 \Rightarrow \alpha_1 = \alpha_2 = \dots = \alpha_k = 0$, то векторы X_1, X_2, \dots, X_k называются линейно независимыми.

Пример 2 в п. 3 показывает, что единичные векторы $E_{(1)}, E_{(2)}, \dots, E_{(n)}$ линейно независимы. Один вектор $X \neq 0$, очевидно, всегда линейно независим, поскольку $(\lambda X = 0, X \neq 0) \Rightarrow \lambda = 0$. Далее, свойство системы X_1, \dots, X_k быть линейно независимой никак не

связано с порядком векторов, так как слагаемые $\alpha_i X_i$ в равенстве (4) могут быть переставлены произвольным образом.

Теорема 1. Имеют место следующие утверждения:

i) система векторов $\{X_1, \dots, X_k\}$ с линейно зависимой подсистемой сама линейно зависима;

ii) любая часть линейно независимой системы векторов $\{X_1, \dots, X_k\}$ линейно независима;

iii) среди линейно зависимых векторов X_1, \dots, X_k хотя бы один является линейной комбинацией остальных;

iv) если один из векторов X_1, \dots, X_k выражается через остальные, то векторы X_1, \dots, X_k линейно зависимы;

v) если векторы X_1, \dots, X_k линейно независимы, а X_1, \dots, X_k, X линейно зависимы, то X — линейная комбинация векторов X_1, \dots, X_k ;

vi) если векторы X_1, \dots, X_k линейно независимы и вектор X_{k+1} нельзя через них выразить, то система X_1, \dots, X_k, X_{k+1} линейно независима.

Доказательство. i) Пусть, например, первые s векторов $X_1, \dots, X_s, s < k$, линейно зависимы, т.е.

$$\alpha_1 X_1 + \dots + \alpha_s X_s = 0,$$

где не все α_i равны нулю. Положив тогда $\alpha_{s+1} = \dots = \alpha_k = 0$, получим нетривиальную линейную зависимость

$$\alpha_1 X_1 + \dots + \alpha_s X_s + \alpha_{s+1} X_{s+1} + \dots + \alpha_k X_k = 0.$$

Утверждение ii) непосредственно следует из i) (рассуждение от противного).

iii) Пусть, например, $\alpha_k \neq 0$ в соотношении (4). Тогда

$$X_k = -\frac{\alpha_1}{\alpha_k} X_1 - \dots - \frac{\alpha_{k-1}}{\alpha_k} X_{k-1}.$$

iv) Пусть, например, $X_k = \beta_1 X_1 + \dots + \beta_{k-1} X_{k-1}$. Положив $\alpha_1 = \beta_1, \dots, \alpha_{k-1} = \beta_{k-1}, \alpha_k = -1$, придём к соотношению (4) с коэффициентом $\alpha_k \neq 0$.

v) Нетривиальное соотношение

$$\beta_1 X_1 + \dots + \beta_k X_k + \beta X = 0$$

с $\beta \neq 0$ даёт в силу iii) то, что нужно. Если, однако, $\beta = 0$, то $\beta_1 = \dots = \beta_k = 0$, поскольку X_1, \dots, X_k по условию линейно независимы.

Утверждение vi) непосредственно следует из v). \square

5. Базис. Размерность. Дадим теперь важное

Определение. Пусть V — ненулевая линейная оболочка в \mathbb{R}^n . Система векторов $X_1, \dots, X_r \in V$ называется базисом для V (или

в V), если она линейно независима и её линейная оболочка совпадает с V :

$$\langle X_1, \dots, X_r \rangle = V.$$

Из определений базиса и линейной оболочки системы векторов следует, что каждый вектор $X \in V$ записывается единственным образом в виде $X = \alpha_1 X_1 + \dots + \alpha_r X_r$. Коэффициенты $\alpha_1, \dots, \alpha_r \in \mathbb{R}$ называются координатами вектора X относительно базиса X_1, \dots, X_r .

Как мы уже видели, линейно независимые единичные векторы (3) порождают \mathbb{R}^n . Стало быть, $\{E_{(1)}, E_{(2)}, \dots, E_{(n)}\}$ — базис пространства \mathbb{R}^n . Но этот так называемый *стандартный* базис — далеко не единственный базис в \mathbb{R}^n . Например, векторы

$$E'_{(1)} = E_{(1)}, \quad E'_{(2)} = E_{(1)} + E_{(2)},$$

$$E'_{(3)} = E_{(1)} + E_{(2)} + E_{(3)}, \quad \dots, \quad E'_{(n)} = E_{(1)} + E_{(2)} + \dots + E_{(n)}$$

тоже составляют базис пространства \mathbb{R}^n (проверьте это аккуратно). С другой стороны, пока не ясно, каждая ли линейная оболочка в \mathbb{R}^n обладает базисом, а если да, то будет ли количество базисных векторов постоянным. Ответы на оба вопроса оказываются положительными. Наши рассуждения будут основаны на следующей лемме.

Лемма. Пусть V — линейная оболочка в \mathbb{R}^n с базисом X_1, \dots, X_r , и Y_1, Y_2, \dots, Y_s — линейно независимая система векторов из V . Тогда $s \leq r$.

Доказательство. Как и все векторы из V , Y_1, \dots, Y_s являются линейными комбинациями базисных векторов. Пусть

$$Y_1 = a_{11} X_1 + a_{12} X_2 + \dots + a_{1r} X_r,$$

$$Y_2 = a_{21} X_1 + a_{22} X_2 + \dots + a_{2r} X_r,$$

$$\dots$$

$$Y_s = a_{s1} X_1 + a_{s2} X_2 + \dots + a_{sr} X_r,$$

где a_{ij} — какие-то скаляры (являясь координатами векторов Y_j , они однозначно определены, но это пока несущественно для нас).

Рассуждаем от противного. Предположим, что $s > r$. Составим линейную комбинацию векторов Y_j , с коэффициентами x_j :

$$x_1 Y_1 + \dots + x_s Y_s =$$

$$= (a_{11} x_1 + a_{12} x_2 + \dots + a_{1s} x_s) X_1 + \dots + (a_{r1} x_1 + a_{r2} x_2 + \dots + a_{rs} x_s) X_r.$$

и рассмотрим систему из r линейных уравнений с s неизвестными

$$a_{11} x_1 + a_{12} x_2 + \dots + a_{1s} x_s = 0,$$

$$\dots$$

$$a_{r1} x_1 + a_{r2} x_2 + \dots + a_{rs} x_s = 0.$$

Так как по предположению $s > r$, то применимо следствие 2 § 3 гл. 1, согласно которому наша система обладает ненулевым решением

(x_1^0, \dots, x_s^0) . Мы приходим к нетривиальной линейной зависимости

$$x_1^0 Y_1 + x_2^0 Y_2 + \dots + x_s^0 Y_s = 0,$$

наличие которой, однако, противоречит условию леммы. Значит, $s \leq r$. \square

Теорема 2. Каждая ненулевая линейная оболочка $V \subset \mathbb{R}^n$ обладает конечным базисом. Все базисы оболочки V состоят из одинакового числа $r \leq n$ векторов (это число называется размерностью оболочки V и обозначается $\dim_{\mathbb{R}} V$ или просто $\dim V$).

Доказательство. В соответствии с условием V содержит хотя бы один ненулевой вектор X_1 (строку или столбец). Пусть мы нашли в V линейно независимую систему векторов X_1, \dots, X_k . Если линейная оболочка $\langle X_1, \dots, X_k \rangle$ не совпадает с V , то выберем в V вектор $X_{k+1} \notin \langle X_1, \dots, X_k \rangle$. Другими словами, X_{k+1} не является линейной комбинацией векторов X_1, \dots, X_k . По теореме 1, vi) система X_1, \dots, X_k, X_{k+1} линейно независима. Мы могли бы продолжать неограниченно процесс расширения линейно независимой системы, но все её векторы X_i лежат в $\mathbb{R}^n = \langle E_{(1)}, E_{(2)}, \dots, E_{(n)} \rangle$, а по только что доказанной лемме всякая линейно независимая система в \mathbb{R}^n содержит не более n векторов. Стало быть, при некотором натуральном $r \leq n$ линейно независимая система $X_1, \dots, X_r \in V$ станет максимальной, т.е. мы получим линейно зависимую систему X_1, \dots, X_r, X , каков бы ни был вектор $X \neq 0$ из V . По теореме 1, v) будем иметь включение $X \in \langle X_1, \dots, X_r \rangle$. Значит, $V = \langle X_1, \dots, X_r \rangle$, и векторы X_1, \dots, X_r составляют базис для V .

Предположим теперь, что Y_1, \dots, Y_s — ещё один базис для V . По лемме мы имеем неравенство $s \leq r$. Поменяв местами системы X_1, \dots, X_r и Y_1, \dots, Y_s , мы получим по той же лемме неравенство $r \leq s$. Стало быть, $s = r$, и теорема доказана. \square

Заметим, хотя в этом и нет большой необходимости, что все наши рассуждения в равной мере относились как к пространству строк, так и к пространству столбцов.

Итак, с каждой линейной оболочкой V в \mathbb{R}^n ассоциируется целое положительное число $r \leq n$, которое мы назвали её размерностью: $r = \dim V$. В частности, $\dim \mathbb{R}^n = n$. Этот важный числовой параметр пространства можно характеризовать разными другими способами. Один из вариантов определения размерности основан на понятии ранга системы векторов. Именно, если $\{X_1, X_2, \dots\}$ — какая-то, возможно, бесконечная, система векторов в пространстве \mathbb{R}^n , то, как мы знаем, размерность линейной оболочки $\langle X_1, \dots \rangle$ не превосходит n . Её называют *рангом* системы $\{X_1, X_2, \dots\}$:

$$\text{rank } \{X_1, X_2, \dots\} = \dim \langle X_1, X_2, \dots \rangle.$$

В случае $V = \{0\}$ принято считать $\dim V = 0$.

УПРАЖНЕНИЯ

1. Линейная оболочка $(U \cup V)$ называется *суммой* подпространств U и V :

$$U + V = \{u + v \mid u \in U, v \in V\}.$$

Если $U \cap V = 0$, то говорят, что сумма $U + V$ *прямая*, и пишут $U \oplus V$.

Пусть $V = V_1 \oplus V_2$ и $X = X_1 + X_2 = X'_1 + X'_2$ — два выражения вектора $X \in V$ в виде линейной комбинации векторов $X_1, X'_1 \in V_1$ и $X_2, X'_2 \in V_2$. Тогда имеем $X_1 - X'_1 = X'_2 - X_2 \in V_1 \cap V_2$, а так как $V_1 \cap V_2 = 0$, то $X_1 = X'_1$, $X_2 = X'_2$.

Доказать обратное: если запись $X = X_1 + X_2$, $X_i \in V_i$, $i = 1, 2$, единственна для каждого вектора $X \in V$, то сумма $V = V_1 + V_2$ прямая. Более общо: сумма V подпространств $V_1, \dots, V_k \subset \mathbb{R}^n$ называется *прямой суммой* $V = V_1 \oplus \dots \oplus V_k$, если каждый вектор $X \in V$ имеет однозначное выражение вида $X = X_1 + \dots + X_k$ с $X_i \in V_i$.

2. Пусть V, V_1 и V_2 — линейные оболочки в \mathbb{R}^n , причём $V \subset V_1 + V_2$. Всегда ли верно, что $V = V \cap V_1 + V \cap V_2$? Что можно сказать про это соотношение в частном случае $V_1 \subset V$?

3. Пусть V — линейная оболочка в \mathbb{R}^n . Если $V = U \oplus W$ — разложение в прямую сумму, то оболочка W называется *дополнением* к U , а U — дополнением к W в V . Однозначно ли определено дополнение к U в V ? Сравнить W с теоретико-множественным понятием дополнения V/U (см. § 5 гл. 1).

4. Показать, что векторы $X_1 = (1, 2, 3)$, $X_2 = (3, 2, 1)$ линейно независимы; рассмотреть линейную оболочку $V(X_1, X_2)$; показать, что вектор $X = (-5, 2, 9)$ содержится в V , и найти его координаты в базисе X_1, X_2 ; найти в \mathbb{R}^3 хотя бы одно дополнение к V .

5. Показать, что система векторов X_1, \dots, X_n из \mathbb{R}^n тогда и только тогда порождает \mathbb{R}^n , когда она линейно независима.

6. Показать, что всякую линейно независимую систему векторов X_1, \dots, X_k из линейной оболочки $V \subset \mathbb{R}^n$ можно вложить в некоторую базисную систему для V .

7. Пусть U и V — линейные оболочки в \mathbb{R}^n . Доказать, что если $U \cap V = 0$, то $\dim(U + V) = \dim U + \dim V$.

8. Найти ранг системы векторов $(0, 1, 1), (1, 0, 1), (1, 1, 0)$.

§ 2. Ранг матрицы

1. **Возвращение к уравнениям.** В векторном пространстве \mathbb{R}^m столбцов высоты m рассмотрим n векторов

$$A^{(j)} = [a_{1j}, a_{2j}, \dots, a_{mj}], \quad j = 1, 2, \dots, n,$$

и их линейную оболочку $V = \langle A^{(1)}, A^{(2)}, \dots, A^{(n)} \rangle$. Пусть дан ещё один вектор $B = [b_1, b_2, \dots, b_m]$. Спрашивается, принадлежит ли B линейной оболочке $V \subset \mathbb{R}^m$, а если принадлежит, то каким образом его координаты b_1, \dots, b_m (относительно стандартного базиса (3') из § 1) выражаются через координаты векторов $A^{(j)}$? В случае $\dim V = n$ вторая часть вопроса относится к значениям координат вектора B в базисе $A^{(1)}, \dots, A^{(n)}$. Мы берём линейную комбинацию векторов

$A^{(j)}$ с произвольными коэффициентами x_j и составляем уравнение $x_1 A^{(1)} + \dots + x_n A^{(n)} = B$. Наглядный вид этого уравнения

$$\begin{array}{c|c||c|c||c|c||c|c} & a_{11} & & a_{12} & & a_{1n} & & b_1 \\ x_1 & a_{21} & + x_2 & a_{22} & + \dots + x_n & a_{2n} & \vdots & b_2 \\ & \vdots & & \vdots & & \vdots & & \vdots \\ & a_{m1} & & a_{m2} & & a_{mn} & & b_m \end{array} \quad (1)$$

есть лишь иная запись системы из m линейных уравнений с n неизвестными

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

Именно такую систему мы и встретили впервые в § 3 гл. 1. Там же были введены простая и расширенная матрицы

$$\begin{aligned} A &= \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{vmatrix}, \\ (A|B) &= \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{vmatrix} \end{aligned} \quad (3)$$

линейной системы (2). Первое впечатление таково, что мы вернулись к исходным позициям, потеряв время и ничего не выиграв. На самом же деле мы располагаем теперь рядом важных понятий. Осталось приобрести навыки в обращении с ними.

В этом месте удобно ещё раз остановиться на обозначениях. Для сокращения записи мы часто будем сумму $s_1 + s_2 + \dots + s_n$ обозначать $\sum_{i=1}^n s_i$. При этом s_1, \dots, s_n — величины произвольной природы (числа, векторы-строки и т.д.), для которых выполнены все законы сложения чисел или векторов. Правила

$$\sum_{i=1}^n t s_i = t \sum_{i=1}^n s_i, \quad \sum_{i=1}^n (s_i + t_i) = \sum_{i=1}^n s_i + \sum_{i=1}^n t_i$$

достаточно понятны, чтобы их нужно было разъяснить.

Будут рассматриваться также *двойные суммы*

$$\sum_{j=1}^n \sum_{i=1}^m a_{ij} = \sum_{j=1}^n \left(\sum_{i=1}^m a_{ij} \right) = \sum_{i=1}^m \left(\sum_{j=1}^n a_{ij} \right) = \sum_{i,j} a_{ij},$$

в которых порядок суммирования (по первому и второму индексу) можно выбирать по своему желанию. Это легко понять, если расположить величины a_{ij} в прямоугольную матрицу размера $m \times n$: в нашей воле начинать суммирование элементов матрицы по строкам или по столбцам.

Другие возможные типы суммирования будут разъяснены в нужном месте.

2. Ранг матрицы. Назовём пространством столбцов прямоугольной матрицы A размера $m \times n$ (см. (3)) введённую выше линейную оболочку $V = \langle A^{(1)}, A^{(2)}, \dots, A^{(n)} \rangle$. Будем теперь V обозначать $V_B(A)$ или просто V_B (B — вертикальный). Размерность $r_B(A) = \dim V_B$ назовём *rangом по столбцам* матрицы A . Аналогично вводится *rang по строкам* матрицы A : $r_F(A) = \dim V_F$, где $V_F = \langle A_{(1)}, A_{(2)}, \dots, A_{(m)} \rangle$ — пространство строк матрицы A , т.е. линейная оболочка в \mathbb{R}^n , натянутая на векторы-строки $A_{(i)} = (a_{i1}, a_{i2}, \dots, a_{in})$, $i = 1, 2, \dots, m$ (F — горизонтальный). Другими словами,

$$r_B(A) = \text{rank} \{A^{(1)}, A^{(2)}, \dots, A^{(n)}\},$$

$$r_F(A) = \text{rank} \{A_{(1)}, A_{(2)}, \dots, A_{(m)}\}$$

— ранги систем векторов-столбцов и векторов-строк соответственно. По теореме 2 § 1 величины $r_B(A)$ и $r_F(A)$ определены правильно.

Следуя терминологии, введённой в § 3 гл. 1, будем говорить, что матрица A' получена из A *элементарным преобразованием типа (I)*, если $A'_{(s)} = A_{(t)}$, $A'_{(t)} = A_{(s)}$ для какой-то пары индексов $s \neq t$ и $A'_{(i)} = A_{(i)}$ для $i \neq s, t$. Если же $A'_{(i)} = A_{(i)}$ для всех $i \neq s$ и $A'_{(s)} = A_{(s)} + \lambda A_{(t)}$, $s \neq t$, $\lambda \in \mathbb{R}$, то говорим, что к A применено *элементарное преобразование типа (II)*. Здесь имеются в виду элементарные преобразования над строками матрицы A .

Заметим, что элементарные преобразования обоих типов обратимы, т.е. матрица A' , получающаяся из A при помощи одного элементарного преобразования, переходит снова в A путём применения одного элементарного преобразования, причём того же типа.

Лемма. *Если матрица A' получена из прямоугольной матрицы A путём применения конечной последовательности элементарных преобразований над строками, то имеют место равенства:*

- i) $r_F(A') = r_F(A)$;
- ii) $r_B(A') = r_B(A)$.

Доказательство. Достаточно рассмотреть тот случай, когда A' получена из A путём применения одного элементарного преобразования (э.п.).

- i) Так как

$$\langle A_{(1)}, \dots, A_{(s)}, \dots, A_{(t)}, \dots, A_{(m)} \rangle = \langle A_{(1)}, \dots, A_{(t)}, \dots, A_{(s)}, \dots, A_{(m)} \rangle,$$

то э.п. типа (I) не меняет $r_{\Gamma}(A)$. Далее,

$$A'_{(s)} = A_{(s)} + \lambda A_{(t)} \implies A_{(s)} = A'_{(s)} - \lambda A_{(t)},$$

и, следовательно,

$$\begin{aligned} \langle A_{(1)}, \dots, A_{(s)} + \lambda A_{(t)}, \dots, A_{(t)}, \dots, A_{(m)} \rangle = \\ = \langle A_{(1)}, \dots, A_{(s)}, \dots, A_{(t)}, \dots, A_{(m)} \rangle, \end{aligned}$$

так что $r_{\Gamma}(A)$ не меняется и при э.п. типа (II).

ii) Пусть $A'^{(j)}, 1 \leq j \leq n$, — столбцы матрицы A' . Докажем, что

$$\sum_{j=1}^n \lambda_j A^{(j)} = 0 \iff \sum_{j=1}^n \lambda_j A'(j) = 0. \quad (4)$$

С этой целью рассмотрим две линейные однородные системы ЛОС и ЛОС' с матрицами A и A' соответственно, записанные в виде (1) (столбцы свободных членов нулевые):

$$\text{ЛОС : } \sum_{j=1}^n x_j A^{(j)} = 0, \quad \text{ЛОС' : } \sum_{j=1}^n x_j A'(j) = 0.$$

Матрицы A и A' у нас таковы, что ЛОС' получается из ЛОС при помощи э.п. типа (I) или (II). По теореме 1 § 3 гл. 1 системы ЛОС и ЛОС' эквивалентны, т.е. всякое решение $(\lambda_1, \lambda_2, \dots, \lambda_n)$ одной системы будет решением другой, а это и есть импликация (4).

Таким образом, всякой, в том числе и максимальной, независимой системе столбцов одной матрицы будет отвечать независимая система столбцов с теми же номерами другой матрицы, чем и устанавливается равенство $r_{\text{B}}(A') = r_{\text{B}}(A)$. \square

Основным результатом этого параграфа является следующее утверждение.

Теорема 1. Для любой прямоугольной $m \times n$ -матрицы A справедливо равенство $r_{\text{B}}(A) = r_{\Gamma}(A)$ (это число называется рангом матрицы A и обозначается $\text{rank } A$).

Доказательство. По теореме 2 § 3 гл. 1 конечным числом элементарных преобразований, совершаемых над строками A_i , матрицу A можно привести к ступенчатому виду

$$\bar{A} = \left| \begin{array}{ccccccc} \bar{a}_{11} & \dots & \bar{a}_{1k} & \dots & \bar{a}_{1l} & \dots & \bar{a}_{1s} & \dots & \bar{a}_{1n} \\ 0 & \dots & \bar{a}_{2k} & \dots & \bar{a}_{2l} & \dots & \bar{a}_{2s} & \dots & \bar{a}_{2n} \\ 0 & \dots & 0 & \dots & \bar{a}_{3l} & \dots & \bar{a}_{3s} & \dots & \bar{a}_{3n} \\ 0 & \dots & 0 & \dots & 0 & \dots & \bar{a}_{rs} & \dots & \bar{a}_{rn} \\ 0 & \dots & 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ 0 & \dots & 0 & \dots & 0 & \dots & 0 & \dots & 0 \end{array} \right| \quad (5)$$

с $\bar{a}_{11}\bar{a}_{2k}\bar{a}_{3l}\dots\bar{a}_{rs} \neq 0$. Согласно лемме

$$r_B(A) = R_B(\bar{A}), \quad r_G(A) = r_G(\bar{A}),$$

так что нам достаточно доказать равенство $r_B(\bar{A}) = r_G(\bar{A})$.

Столбцы матриц A и \bar{A} с номерами $1, k, l, \dots, s$, отвечающими главным неизвестным $x_1, x_k, x_l, \dots, x_s$ линейной системы (2), будем называть *базисными столбцами*. Эта терминология вполне оправдана. Предположив наличие соотношения

$$\lambda_1\bar{A}^{(1)} + \lambda_k\bar{A}^{(k)} + \lambda_l\bar{A}^{(l)} + \dots + \lambda_s\bar{A}^{(s)} = 0,$$

связывающего векторы-столбцы

$$\begin{aligned} \bar{A}^{(1)} &= [\bar{a}_{11}, 0, \dots, 0], \quad \bar{A}^{(k)} = [\bar{a}_{1k}, \bar{a}_{2k}, 0, \dots, 0], \quad \dots \\ &\dots, \quad \bar{A}^{(s)} = [\bar{a}_{1s}, \bar{a}_{2s}, \dots, \bar{a}_{rs}, 0, \dots, 0] \end{aligned}$$

матрицы (5), получим

$$\lambda_s\bar{a}_{rs} = 0, \quad \dots, \quad \lambda_1\bar{a}_{3l} = 0, \quad \lambda_k\bar{a}_{2k} = 0, \quad \lambda_1\bar{a}_{11} = 0,$$

а так как $\bar{a}_{11}\bar{a}_{2k}\bar{a}_{3l}\dots\bar{a}_{rs} \neq 0$, то $\lambda_1 = \lambda_k = \lambda_l = \dots = \lambda_s = 0$. Значит, $\text{rank}\{\bar{A}^{(1)}, \bar{A}^{(k)}, \bar{A}^{(l)}, \dots, \bar{A}^{(s)}\} = r$ и $r_B(\bar{A}) \geq r$. Но пространство \bar{V}_B , порождённое столбцами матрицы \bar{A} , отождествляется с пространством столбцов матрицы, которая получается из \bar{A} удалением последних $m - r$ нулевых строк. Поэтому $r_B(\bar{A}) = \dim \bar{V}_B \leq \dim \mathbb{R}^r = r$. Сопоставление двух неравенств показывает, что $r_B(\bar{A}) = r$ (неравенство $r_B(\bar{A}) \leq r$ вытекает также из того очевидного соображения, что все столбцы матрицы \bar{A} являются линейными комбинациями базисных; проделайте это самостоятельно в качестве упражнения).

С другой стороны, все ненулевые строки матрицы \bar{A} линейно независимы: любое гипотетическое соотношение

$$\lambda_1\bar{A}_{(1)} + \lambda_2\bar{A}_{(2)} + \dots + \lambda_r\bar{A}_{(r)} = 0, \quad \lambda_i \in \mathbb{R},$$

как и в случае со столбцами, даёт последовательно

$$\lambda_1\bar{a}_{11} = 0, \quad \lambda_2\bar{a}_{2k} = 0, \quad \dots, \quad \lambda_r\bar{a}_{rs} = 0,$$

откуда $\lambda_1 = \lambda_2 = \dots = \lambda_r = 0$. Стало быть, $r_G(\bar{A}) = r = r_B(\bar{A})$. \square

3. Критерий совместности. Ступенчатый вид матрицы A , дающий ответ на ряд вопросов относительно линейных систем (см. § 3 гл. 1), содержит элементы произвола, связанные, например, с выбором базисных столбцов, или, что эквивалентно, с выбором главных неизвестных системы (2). В то же время из теоремы 1 и из её доказательства извлекается

Следствие. Число главных неизвестных линейной системы (2) не зависит от способа приведения её к ступенчатому виду и равно $\text{rank } A$, где A — матрица системы.

Действительно, мы видели, что число главных неизвестных равно числу ненулевых строк матрицы \tilde{A} (см. (5)), совпадающему с рангом матрицы A . Ранг определялся нами совершенно инвариантным образом. (Этими словами выражается тот факт, что ранг матрицы служит её внутренней характеристикой, не зависящей от каких-либо приводящих обстоятельств.) \square

В следующей главе мы получим эффективное средство для вычисления ранга матрицы A , устраниющее необходимость приведения A к ступенчатому виду. Это, несомненно, повысит ценность утверждений, основанных на понятии ранга. В качестве простого, но полезного примера сформулируем критерий разрешимости линейной системы, речь о котором шла ещё в гл. 1.

Теорема 2 (Кронекер—Капелли). *Система линейных уравнений (2) совместна тогда и только тогда, когда ранг её матрицы, совпадает с рангом расширенной матрицы (см. (3)).*

Доказательство. Совместность линейной системы (2), записанной в виде (1), можно трактовать (с этого начинался настоящий параграф) как вопрос о представлении вектора-столбца B свободных членов в виде линейной комбинации векторов-столбцов $A^{(j)}$ матрицы A . Если такое представление возможно (т.е. система (2) совместна), то $B \in \langle A^{(1)}, \dots, A^{(n)} \rangle$ и $\text{rank}\{A^{(1)}, \dots, A^{(n)}\} = \text{rank}\{A^{(1)}, \dots, A^{(n)}, B\}$, откуда $\text{rank}A = r_B(A) = r_B((A|B)) = \text{rank}(A|B)$ (см. формулировку теоремы 1).

Обратно: если ранги матриц A и $(A|B)$ совпадают и $\{A^{(j_1)}, \dots, A^{(j_r)}\}$ — какая-то максимальная линейно независимая система столбцов матрицы A , то расширенная система $\{A^{(j_1)}, \dots, A^{(j_r)}, B\}$ будет линейно зависимой, а это по теореме 1, в) § 1 означает, что B — линейная комбинация базисных (и тем более всех) столбцов $A^{(j)}$. Стало быть, система (2) совместна. \square

УПРАЖНЕНИЯ

1. Доказать теорему 1, не приводя $m \times n$ -матрицу $A = (a_{ij})$ к ступенчатому виду.

Указание. Пусть $\dim V_T(A) = r$, $\dim V_B(A) = s$. Выбрать r базисных строк; без ограничения общности можно считать, что ими являются первые r строк $A_{(1)}, A_{(2)}, \dots, A_{(r)}$. Рассмотреть укороченную $r \times n$ -матрицу $\tilde{A} = [A_{(1)}, A_{(2)}, \dots, A_{(r)}]$, составленную из первых r строк матрицы A . Выбрать в \tilde{A} t базисных столбцов, $t = \dim V_B(\tilde{A})$. Пусть ими будут $\tilde{A}^{(1)}, \dots, \tilde{A}^{(t)}$. Так как $V_B(\tilde{A}) \subset \mathbb{R}^r$, то $t \leq r$. Для каждого столбца $A^{(k)}$, $k > t$, нужно найти скаляры $\lambda_1, \dots, \lambda_t \in \mathbb{R}$ такие, что $A^{(k)} = \lambda_1 A^{(1)} + \dots + \lambda_t A^{(t)}$, т.е. $a_{ik} = \sum_{p=1}^t \lambda_p a_{ip}$, $1 \leq i \leq m$. При $i \leq r$ это, наверное, так, ибо имеется соотношение $\tilde{A}^{(k)} = \lambda_1 \tilde{A}^{(1)} + \dots + \lambda_t \tilde{A}^{(t)}$ для укороченных столбцов. При $i > r$ использовать выражение $A_{(i)} = \mu_1 A_{(1)} + \dots + \mu_r A_{(r)}$ для i -й строки через первые r строк. Из него следует, что $a_{ik} = \sum_{l=1}^r \mu_l a_{il} = \sum_{l=1}^r \mu_l \sum_{p=1}^t \lambda_p a_{ip} = \sum_{p=1}^t \lambda_p \sum_{l=1}^r \mu_l a_{il} = \sum_{p=1}^t \lambda_p a_{ip}$. Установленная ли-

нейная зависимость столбцов показывает, что $s \leq t$, а так как $t \leq r$, то $s \leq r$. Рассмотреть, далее, так называемую *транспонированную* матрицу

$${}^t A = \begin{vmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{vmatrix}$$

размера $n \times m$. Имеют место равенства $r_{{}^t A} = r_A$, $r_{{}^t A} = r_A$, поэтому по доказанному $r \leq s$. Стало быть, $r = s$.

2. Как и в случае строк, перестановку столбцов с номерами s и t матрицы A называют элементарным преобразованием (э.п.) типа (I), а прибавление к s -му столбцу t -го столбца, умноженного на скаляр λ , — э.п. типа (II).

Указать ступенчатый вид матрицы A по столбцам. Элементарными преобразованиями столбцов привести матрицу \tilde{A} (см. (5)) к виду

$$\tilde{A} = \text{diag}(\tilde{a}_{11}, \tilde{a}_{22}, \dots, \tilde{a}_{rr}, 0, \dots, 0),$$

где $\tilde{a}_{11} = a_{11}$, $\tilde{a}_{22} = a_{22}$, $\tilde{a}_{33} = a_{3r}$, \dots , $\tilde{a}_{rr} = a_{rs}$; $\prod_{i=1}^r \tilde{a}_{ii} \neq 0$.

3. Показать, что при $a_0 \neq 0$ квадратная матрица

$$A = \begin{vmatrix} 0 & 0 & \dots & 0 & 0 & a_0 \\ 1 & 0 & \dots & 0 & 0 & a_1 \\ 0 & 1 & \dots & 0 & 0 & a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 & a_{n-2} \\ 0 & 0 & \dots & 0 & 1 & a_{n-1} \end{vmatrix}$$

имеет ранг n .

4. Условие равенства рангов двух матриц

$$A = \begin{vmatrix} \alpha_1 & \alpha_2 & \dots & \alpha_n \\ \beta_1 & \beta_2 & \dots & \beta_n \end{vmatrix}, \quad B = \begin{vmatrix} \alpha_1 & \alpha_2 & \dots & \alpha_n \\ \beta_1 & \beta_2 & \dots & \beta_n \\ \gamma_1 & \gamma_2 & \dots & \gamma_n \end{vmatrix}$$

выразить геометрическим свойством множества n прямых на плоскости.

§ 3. Линейные отображения. Действия с матрицами

1. **Матрицы и отображения.** Пусть \mathbb{R}^n и \mathbb{R}^m — векторные пространства столбцов высоты n и m соответственно. Пусть, далее, $A = (a_{ij})$ — матрица размера $m \times n$. Определим отображение $\varphi_A : \mathbb{R}^n \rightarrow \mathbb{R}^m$, полагая для любого $X = [x_1, x_2, \dots, x_n] \in \mathbb{R}^n$

$$\varphi_A(X) = x_1 A^{(1)} + x_2 A^{(2)} + \dots + x_n A^{(n)}, \quad (1)$$

где $A^{(1)}, \dots, A^{(n)}$ — столбцы матрицы A (сравнить с (1) § 2). Так как они имеют высоту m , то в правой части (1) стоит вектор-столбец $Y = [y_1, y_2, \dots, y_m] \in \mathbb{R}^m$. Более подробно (1) переписывается в виде

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

Если $X = X' + X'' = [x'_1 + x''_1, x'_2 + x''_2, \dots, x'_n + x''_n]$, то

$$\varphi_A(X' + X'') = \sum_{i=1}^n (x'_i + x''_i) A^{(i)} = \sum_{i=1}^n x'_i A^{(i)} + \sum_{i=1}^n x''_i A^{(i)} = \\ = \varphi_A(X') + \varphi_A(X'').$$

Аналогично,

$$\varphi_A(\lambda X) = \sum_{i=1}^n \lambda x_i A^{(i)} = \lambda \sum_{i=1}^n x_i A^{(i)} = \lambda \varphi_A(X), \quad \lambda \in \mathbb{R}.$$

Обратно, предположим, что $\varphi: \mathbb{R}^n \rightarrow \mathbb{R}^m$ — отображение множества в смысле § 5 гл. 1, обладающее следующими двумя свойствами:

- i) $\varphi(X' + X'') = \varphi(X') + \varphi(X'')$ для всех $X', X'' \in \mathbb{R}^n$;
- ii) $\varphi(\lambda X) = \lambda \varphi(X)$ для всех $X \in \mathbb{R}^n, \lambda \in \mathbb{R}$.

Как мы знаем (см. п. 3 § 1), $\mathbb{R}^n = \langle E^{(1)}, \dots, E^{(n)} \rangle$ — линейная оболочка стандартных базисных столбцов, так что

$$X = [x_1, x_2, \dots, x_n] = \sum_{j=1}^n x_j E^{(j)}.$$

Согласно свойствам i), ii) имеем

$$\varphi(X) = \varphi\left(\sum_{j=1}^n x_j E^{(j)}\right) = \sum_{j=1}^n x_j \varphi(E^{(j)}). \quad (2)$$

Соотношение (2) показывает, что отображение φ полностью определяется своими значениями на базисных векторах-столбцах. Положив

$$\varphi(E^{(j)}) = [a_{1j}, a_{2j}, \dots, a_{mj}] = A^{(j)} \in \mathbb{R}^m, \quad (3)$$

мы обнаруживаем, что задание φ равносильно заданию прямоугольной матрицы $A = (a_{ij})$ размера $m \times n$ со столбцами $A^{(1)}, \dots, A^{(n)}$, а соотношения (1) и (2) фактически совпадают. Стало быть, можно положить $\varphi = \varphi_A$.

Определение. Отображение $\varphi = \varphi_A: \mathbb{R}^n \rightarrow \mathbb{R}^m$, обладающее свойствами i), ii), называется *линейным отображением из \mathbb{R}^n в \mathbb{R}^m* . Часто, в особенности при $n = m$, говорят о *линейном преобразовании*. Матрица A называется *матрицей линейного отображения* φ_A .

Пусть $\varphi_A, \varphi_{A'}$ — два линейных отображения $\mathbb{R}^n \rightarrow \mathbb{R}^m$ с матрицами $A = (a_{ij})$ и $A' = (a'_{ij})$. Тогда равенство $\varphi_A = \varphi_{A'}$ равносильно совпадению значений $\varphi_A(X) = \varphi_{A'}(X)$ для всех $X \in \mathbb{R}^n$. В частности, $A'(j) = \varphi_{A'}(E^{(j)}) = \varphi_A(E^{(j)}) = A^{(j)}, 1 \leq j \leq n$, откуда $a'_{ij} = a_{ij}$ и $A' = A$.

Резюмируем наши результаты.

Теорема 1. Между линейными отображениями \mathbb{R}^n в \mathbb{R}^m и матрицами размера $m \times n$ существует взаимно однозначное соответствие.

Следует подчеркнуть, что бессмысленно говорить о линейных отображениях $S \rightarrow T$ произвольных множеств S и T . Условия i), ii) предполагают, что S и T — линейные оболочки в \mathbb{R}^n и \mathbb{R}^m соответственно.

Обратим внимание на специальный случай $m = 1$, когда линейное отображение $\varphi: \mathbb{R}^n \rightarrow \mathbb{R}$, обычно называемое линейной функцией от n переменных, задается n скалярами a_1, a_2, \dots, a_n :

$$\varphi(X) = \varphi(x_1, x_2, \dots, x_n) = a_1x_1 + a_2x_2 + \dots + a_nx_n. \quad (4)$$

Замечание. Наша терминология отличается от той, которая принята в средней школе, где (в случае одной переменной x) линейной называют функцию $x \mapsto ax + b$.

Линейные функции (4), равно как и произвольные линейные отображения $\mathbb{R}^n \rightarrow \mathbb{R}^m$ при фиксированных n и m можно складывать и умножать на скаляры. В самом деле, пусть $\varphi_A, \varphi_B: \mathbb{R}^n \rightarrow \mathbb{R}^m$ — два линейных отображения. Отображение

$$\varphi = \alpha\varphi_A + \beta\varphi_B: \mathbb{R}^n \rightarrow \mathbb{R}^m, \quad \alpha, \beta \in \mathbb{R},$$

определяется своими значениями:

$$\varphi(X) = \alpha\varphi_A(X) + \beta\varphi_B(X).$$

В правой части стоит обычная линейная комбинация векторов-столбцов.

Так как

$$\begin{aligned} \varphi(X' + X'') &= \alpha\varphi_A(X' + X'') + \beta\varphi_B(X' + X'') = \\ &= \alpha\{\varphi_A(X') + \varphi_A(X'')\} + \beta\{\varphi_B(X') + \varphi_B(X'')\} = \\ &= \{\alpha\varphi_A(X') + \beta\varphi_B(X')\} + \{\alpha\varphi_A(X'') + \beta\varphi_B(X'')\} = \varphi(X') + \varphi(X''), \\ \varphi(\lambda X) &= \alpha\varphi_A(\lambda X) + \beta\varphi_B(\lambda X) = \alpha\lambda\varphi_A(X) + \beta\lambda\varphi_B(X) = \\ &= \lambda\{\alpha\varphi_A(X) + \beta\varphi_B(X)\} = \lambda\varphi(X) \end{aligned}$$

(здесь мы неявно пользовались правилами ВП₁–ВП₈ из § 1), то φ — линейное отображение. По теореме 1 можно говорить о его матрице C : $\varphi = \varphi_C$. Чтобы найти C , выпишем, следуя (3), столбец с номером j :

$$\begin{aligned} [c_{1j}, c_{2j}, \dots, c_{mj}] &= C^{(j)} = \varphi_C(E^{(j)}) = \\ &= \alpha\varphi_A(E^{(j)}) + \beta\varphi_B(E^{(j)}) = \alpha A^{(j)} + \beta B^{(j)} = \\ &= [\alpha a_{1j} + \beta b_{1j}, \alpha a_{2j} + \beta b_{2j}, \dots, \alpha a_{mj} + \beta b_{mj}]. \end{aligned}$$

Матрицу $C = (c_{ij})$ с элементами $c_{ij} = \alpha a_{ij} + \beta b_{ij}$ естественно назвать линейной комбинацией матриц A и B с коэффициентами α и β :

$$\begin{aligned} \alpha \left| \begin{array}{ccc} a_{11} & \dots & a_{1n} \\ a_{m1} & \dots & a_{mn} \end{array} \right| + \beta \left| \begin{array}{ccc} b_{11} & \dots & b_{1n} \\ b_{m1} & \dots & b_{mn} \end{array} \right| &= \\ &= \left| \begin{array}{ccc} \alpha a_{11} + \beta b_{11} & \dots & \alpha a_{1n} + \beta b_{1n} \\ \alpha a_{m1} + \beta b_{m1} & \dots & \alpha a_{mn} + \beta b_{mn} \end{array} \right|. \quad (5) \end{aligned}$$

Итак,

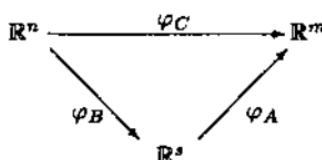
$$\alpha \varphi_A + \beta \varphi_B = \varphi_{\alpha A + \beta B}. \quad (6)$$

Особенно часто нами будет использоваться тот факт, что линейные комбинации линейных функций снова являются линейными функциями.

В заключение этого пункта отметим, что если правила ВП₁–ВП₈ из § 1 для векторных пространств переписать, заменив всюду векторы-строки X, Y, Z на матрицы размера $m \times n$, то в соответствии с определяющим соотношением (5) получатся правила ВМ₁–ВМ₈, которые дают основание говорить о векторном пространстве матриц размера $m \times n$. Если угодно, его можно считать компактной записью векторного пространства $\mathbb{R}^{m \times n}$ строк длины mn (строки разбиты на отрезки длины n , расположенные друг под другом).

2. Произведение матриц. Соотношения (5) и (6) выражают согласованность действий сложения и умножения на скаляры в множествах матриц размера $m \times n$ и отображений $\mathbb{R}^n \rightarrow \mathbb{R}^m$. В случае произвольных множеств имеется ещё важное понятие произведения (композиции) отображений (см. п. 2 § 5 гл. 1). Разумно ожидать, что композиция двух линейных отображений должна выражаться неким согласованным образом в терминах матриц. Посмотрим, как это делается.

Пусть $\varphi_B : \mathbb{R}^n \rightarrow \mathbb{R}^s$, $\varphi_A : \mathbb{R}^s \rightarrow \mathbb{R}^m$ — линейные отображения, $\varphi_C = \varphi_A \circ \varphi_B$ — их композиция:



Вообще говоря, нам следовало бы предварительно проверить, что $\varphi = \varphi_A \circ \varphi_B$ — линейное отображение, но это довольно ясно:

$$\begin{aligned} i) \varphi(X' + X'') &= \varphi_A(\varphi_B(X' + X'')) = \varphi_A(\varphi_B(X') + \varphi_B(X'')) = \\ &= \varphi_A(\varphi_B(X')) + \varphi_A(\varphi_B(X'')) = \varphi(X') + \varphi(X''); \end{aligned}$$

ii) $\varphi(\lambda X) = \varphi_A(\varphi_B(\lambda X)) = \varphi_A(\lambda\varphi_B(X)) = \lambda\varphi_A(\varphi_B(X)) = \lambda\varphi(X)$;
поэтому по теореме 1 с φ ассоциируется вполне определённая матрица C .

Действие отображений на столбцы в цепочке

$$[x_1, \dots, x_n] \xrightarrow{\varphi_B} [y_1, \dots, y_s] \xrightarrow{\varphi_A} [z_1, \dots, z_m]$$

запишем в явном виде по формуле (1'):

$$z_i = \sum_{k=1}^s a_{ik} y_k = \sum_{k=1}^s a_{ik} \sum_{j=1}^n b_{kj} x_j = \sum_{j=1}^n \left(\sum_{k=1}^s a_{ik} b_{kj} \right) x_j.$$

С другой стороны,

$$z_i = \sum_{j=1}^n c_{ij} x_j, \quad i = 1, 2, \dots, m.$$

Сравнивая полученные выражения и памятую о том, что x_j , ($j = 1, 2, \dots, n$) — произвольные вещественные числа, мы приходим к соотношению

$$c_{ij} = \sum_{k=1}^s a_{ik} b_{kj}, \quad 1 \leq i \leq m, 1 \leq j \leq n. \quad (7)$$

Будем говорить, что матрица $C = (c_{ij})$ получается в результате умножения матрицы A на матрицу B . Принято писать

$$C = AB.$$

Таким образом, произведением прямоугольной матрицы (a_{ik}) размера $m \times s$ и прямоугольной матрицы (b_{kj}) размера $s \times n$ называется прямоугольная матрица (c_{ij}) размера $m \times n$ с элементами c_{ij} , заданными соотношением (7).

Нами доказана

Теорема 2. Произведение $\varphi_A \varphi_B$ двух линейных отображений с матрицами A и B является линейным отображением с матрицей $C = AB$. Другими словами,

$$\varphi_A \varphi_B = \varphi_{AB}. \quad (8)$$

Соотношение (8) — естественное дополнение к соотношению (6).

Мы можем забыть о линейных отображениях и находить произведение AB двух произвольных матриц A, B , имея в виду, однако, что символ AB имеет смысл только в том случае, когда число столбцов в матрице A совпадает с числом строк в матрице B . Именно при этом условии работает правило (7) умножения i -й строки $A_{(i)}$ на j -й столбец $B^{(j)}$, согласно которому

$$c_{ij} = (a_{i1}, \dots, a_{is})[b_{1j}, \dots, b_{sj}] = A_{(i)} B^{(j)}. \quad (9)$$

Число строк матрицы AB равно числу строк матрицы A , а число столбцов — числу столбцов матрицы B . В частности, произведение квадратных матриц одинаковых порядков всегда определено, но даже в этом случае, вообще говоря, $AB \neq BA$, как показывает хотя бы следующий пример:

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

Умножение матриц, конечно, можно было бы вводить многими другими способами (умножать, например, строки на строки), но ни один из этих способов не сравним по важности с рассмотренным выше. Это и понятно, поскольку мы пришли к нему при изучении естественной композиции (суперпозиции) отображений, а само понятие отображения относится к числу наиболее фундаментальных в математике.

Следствие. Умножение матриц ассоциативно:

$$A(BC) = (AB)C.$$

Действительно, произведение матриц соответствует произведению линейных отображений (теорема 2 и соотношение (8)), а по теореме 1 § 5 гл. 1 произведение любых отображений ассоциативно. К тому же результату можно прийти вычислительным путём, используя непосредственно соотношение (7). \square

Обратим ещё внимание на так называемые законы дистрибутивности:

$$(A + B)C = AC + BC, \quad D(A + B) = DA + DB, \quad (10)$$

где A, B, C, D — произвольные матрицы размеров соответственно $m \times s, m \times s, s \times n, n \times m$.

Действительно, полагая $A = (a_{ij})$, $B = (b_{ij})$, $C = (c_{ij})$, мы получим для любых i, j равенство (используя дистрибутивность в \mathbb{R})

$$\sum_{k=1}^n (a_{ik} + b_{ik}) c_{kj} = \sum_{k=1}^n a_{ik} c_{kj} + \sum_{k=1}^n b_{ik} c_{kj},$$

левая часть которого даёт элемент g_{ij} матрицы $(A + B)C$, а правая — элементы h_{ij} и h'_{ij} матриц AC и соответственно BC . Второй закон дистрибутивности (10) проверяется совершенно аналогично.

3. Транспонирование матриц. Будем говорить, что матрицы

$$A = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{vmatrix}, \quad {}^t A = \begin{vmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{vmatrix}$$

размеров $m \times n$ и $n \times m$ соответственно получаются друг из друга *транспонированием* — заменой строк на столбцы, а столбцов на

строки (внимательный читатель заметит, что понятие транспонирования уже встречалось в упр. 1 § 2). Непосредственно видно, что

$${}^t({}^t A) = A, \quad {}^t(A + B) = {}^t A + {}^t B, \quad {}^t(\lambda A) = \lambda {}^t A.$$

Транспонирование произведения матриц подчиняется более интересной закономерности. Если

$$A = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1s} \\ a_{21} & a_{22} & \dots & a_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{ms} \end{vmatrix}, \quad B = \begin{vmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{s1} & b_{s2} & \dots & b_{sn} \end{vmatrix}$$

и

$${}^t A = (a'_{ki}), \quad {}^t B = (b'_{jk}),$$

то

$$a'_{ki} = a_{ik}, \quad b'_{jk} = b_{kj}.$$

Вычисление коэффициентов матриц

$$C = AB = \begin{vmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \dots & c_{mn} \end{vmatrix},$$

$$D = {}^t B \cdot {}^t A = \begin{vmatrix} d_{11} & d_{12} & \dots & d_{1m} \\ d_{21} & d_{22} & \dots & d_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n1} & d_{n2} & \dots & d_{nm} \end{vmatrix}$$

по формуле (7):

$$c_{ij} = \sum_{k=1}^n a_{ik} b_{kj}, \quad d_{ji} = \sum_{k=1}^n b'_{jk} a'_{ki} = \sum_{k=1}^n a_{ik} b_{kj},$$

показывает, что $d_{ji} = c_{ij}$ при всех $1 \leq i \leq m, 1 \leq j \leq n$. Значит, ${}^t C = D$, или, в исходных обозначениях,

$${}^t(AB) = {}^t B \cdot {}^t A.$$

Более общо: если определено произведение матриц A_1, A_2, \dots, A_r , то

$${}^t(A_1 A_2 \dots A_r) = {}^t A_r \dots {}^t A_2 {}^t A_1.$$

В силу теоремы 1 § 2 выполнено также свойство $\text{rank } {}^t A = \text{rank } A$.

4. Ранг произведения матриц. Пусть A и B — произвольные матрицы размеров $m \times s$ и $s \times n$. Что можно сказать о величине $\text{rank } AB$?

Теорема 3. Справедливо неравенство

$$\text{rank } AB \leq \min \{\text{rank } A, \text{rank } B\}.$$

Доказательство. Для строк $C_{(i)}$ и столбцов $C^{(j)}$ матрицы $C = AB$ мы в соответствии с (7) имеем выражения

$$C_{(i)} = A_{(i)}B, \quad C^{(j)} = AB^{(j)}. \quad (11)$$

Интерпретируя теперь ранг матрицы A как

$$r_1 = \text{rank } A = \dim\langle A_{(1)}, A_{(2)}, \dots, A_{(m)} \rangle,$$

считаем без ограничения общности базисными строки $A_{(1)}, \dots, A_{(r_1)}$, поскольку необходимая перестановка строк в A будет сопровождаться точно такой же перестановкой строк матрицы C , а это преобразование (э.п. типа (I)) не меняет ни $\text{rank } A$, ни $\text{rank } C$. Итак,

$$A_{(k)} = \sum_{i=1}^{r_1} \lambda_{ki} A_{(i)}, \quad r_1 < k \leq m,$$

откуда (используя дистрибутивность (10)) получаем

$$C_{(k)} = A_{(k)}B = \left(\sum_{i=1}^{r_1} \lambda_{ki} A_{(i)} \right) B = \sum_{i=1}^{r_1} \lambda_{ki} (A_{(i)}B) = \sum_{i=1}^{r_1} \lambda_{ki} C_{(i)},$$

и, стало быть,

$$\langle C_{(1)}, \dots, C_{(m)} \rangle = \langle C_{(1)}, \dots, C_{(r_1)} \rangle.$$

Таким образом,

$$\text{rank } C = \dim\langle C_{(1)}, \dots, C_{(m)} \rangle \leq r_1 = \text{rank } A.$$

Аналогично, интерпретируя ранг матрицы B как

$$r_2 = \text{rank } B = \dim\langle B^{(1)}, B^{(2)}, \dots, B^{(n)} \rangle$$

и считая без ограничения общности базисными столбцы $B^{(1)}, \dots, B^{(r_2)}$, будем иметь

$$B^{(k)} = \sum_{j=1}^{r_2} \mu_{kj} B^{(j)},$$

$$C^{(k)} = AB^{(k)} = A \left(\sum_{j=1}^{r_2} \mu_{kj} B^{(j)} \right) = \sum_{j=1}^{r_2} \mu_{kj} AB^{(j)} = \sum_{j=1}^{r_2} \mu_{kj} C^{(j)},$$

$$r_2 < k \leq n,$$

откуда

$$\text{rank } C = \dim\langle C^{(1)}, \dots, C^{(n)} \rangle \leq r_2 = \text{rank } B. \quad \square$$

Заметим, что в случае каких-то специальных матриц A, B доказанное неравенство может становиться строгим. Так будет, скажем, при $A \neq 0, B \neq 0, AB = 0$ (см. пример в п. 2). В общем случае теорема 3 просто утверждает, что при умножении матриц ранг не может увеличиться.

5. Квадратные матрицы. Множество всех квадратных матриц (a_{ij}) порядка n с вещественными коэффициентами a_{ij} , обычно обозначается $M_n(\mathbb{R})$ (или M_n). Как уже отмечалось в конце п. 1, можно говорить о векторном пространстве $M_n(\mathbb{R})$. Согласно п. 2 произведение любых двух матриц из $M_n(\mathbb{R})$ снова принадлежит $M_n(\mathbb{R})$. При этом выполнены свойства ассоциативности и дистрибутивности.

Определение. Говорят, что квадратные матрицы фиксированного порядка n образуют **матричное (ассоциативное) кольцо**; а с учётом легко проверяемых правил $\lambda AB = (\lambda A)B = A(\lambda B)$ умножения на скаляры $\lambda \in \mathbb{R}$ множество $M_n(\mathbb{R})$ называют также **алгеброй матриц над \mathbb{R}** .

К этим наименованиям предстоит ещё привыкнуть (см. гл. 4 по поводу систематизации терминологических новшеств), а сейчас мы обратим внимание на единичную матрицу $E = (\delta_{kj})$, где

$$\delta_{kj} = \begin{cases} 1, & \text{если } k = j, \\ 0, & \text{если } k \neq j, \end{cases}$$

— **символ Кронекера**. Очевидно, что $\text{rank } E = n$. Правило умножения матриц (7), в котором следует заменить b_{kj} на δ_{kj} , показывает, что справедливы соотношения

$$EA = A = AE, \quad A \in M_n(\mathbb{R}).$$

Более общо:

$$\text{diag}_n(\lambda)A = \lambda A = A \text{diag}_n(\lambda), \quad (12)$$

где

$$\text{diag}_n(\lambda) = \lambda E = \begin{vmatrix} \lambda & 0 & \dots & 0 \\ 0 & \lambda & \dots & 0 \\ 0 & 0 & \dots & \lambda \end{vmatrix}$$

— известная нам скалярная матрица (см. § 3 гл. 1). Таким образом, умножение матрицы A на скаляр λ равносильно умножению A на скалярную матрицу.

В равенстве (12) отражён легко проверяемый факт перестановочности $\text{diag}_n(\lambda)$ с любой матрицей A . Весьма важным для приложений является следующее его обращение.

Теорема 4. *Матрица из M_n , перестановочная со всеми матрицами в M_n , должна быть скалярной.*

Доказательство. Введём матрицу E_{ij} , в которой на пересечении i -й строки и j -го столбца стоит 1, а все остальные элементы нулевые. Если $Z = (z_{ij})$ — матрица, о которой идёт речь в теореме, то она перестановочна, в частности, со всеми E_{ij} :

$$ZE_{ij} = E_{ij}Z, \quad i, j = 1, 2, \dots, n.$$

Перемножая матрицы в левой и правой частях этого равенства, мы получим матрицы

$$\left| \begin{array}{cccc} 0 & \dots & z_{1i} & \dots & 0 \\ 0 & \dots & z_{2i} & \dots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \dots & z_{ni} & \dots & 0 \end{array} \right|_j, \quad \left| \begin{array}{cccc} 0 & 0 & \dots & 0 \\ z_{j1} & z_{j2} & \dots & z_{jn} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 0 \end{array} \right|_i$$

с единственным ненулевым j -м столбцом и соответственно с единственной ненулевой i -й строкой. Их сравнение немедленно приводит к соотношениям $z_{ki} = 0$ при $k \neq i$ и $z_{ii} = z_{jj}$. Меняя i и j , получаем требуемое. \square

Для данной матрицы $A \in M_n(\mathbb{R})$ можно попробовать найти такую матрицу $A' \in M_n(\mathbb{R})$, чтобы выполнялись соотношения $AA' = E = A'A$. Сразу же заметим, что

$$AA' = E = A''A \implies A'' = A'. \quad (13)$$

Действительно, $A'' = A''E = A''(AA') = (A''A)A' = EA' = A'$. Таким образом, матрица A' , коль скоро она существует, единственна. Её называют матрицей, обратной к A , и обозначают A^{-1} :

$$AA^{-1} = E = A^{-1}A. \quad (14)$$

При выполнении (14) говорят ещё, что матрица A *обратима*.

Определение. Матрица $A \in M_n(\mathbb{R})$ называется *невырожденной*, если система её строк (а тем самым и столбцов) линейно независима, т.е. $\text{rank } A = n$. Если $\text{rank } A < n$, то A называется *вырожденной*.

Теорема 5. *Матрица $A \in M_n(\mathbb{R})$ обратима тогда и только тогда, когда она невырождена.*

Доказательство. 1) Если $AB = E$ (или $BA = E$), то по теореме 3 имеем

$$n = \text{rank } E = \text{rank } AB \leq \min\{\text{rank } A, \text{rank } B\} \leq n,$$

откуда $\text{rank } A = n$.

2) Если $\text{rank } A = n$, то

$$\langle E^{(1)}, \dots, E^{(n)} \rangle = \mathbb{R}^n = \langle A^{(1)}, \dots, A^{(n)} \rangle,$$

и, стало быть,

$$E^{(j)} = \sum_{i=1}^n a'_{ij} A^{(i)}, \quad 1 \leq j \leq n, \quad (15)$$

причём коэффициенты a'_{ij} , составляющие матрицу $A' = (a'_{ij}) \in M_n(\mathbb{R})$, определены однозначно. Согласно п. 1 § 2 (см. там уравнения (1) и (2)) соотношения (15) переписываются в виде

$$E^{(j)} = AA'^{(j)}, \quad 1 \leq j \leq n,$$

откуда

$$E = (E^{(1)}, \dots, E^{(n)}) = (AA'^{(1)}, \dots, AA'(n)) = AA'.$$

Здесь мы интерпретировали матрицы E и AA' как объединения отвечающих им столбцов.

Заметим теперь (см. п. 3), что вместе с A невырожденной является и транспонированная матрица ${}^t A$. Поэтому в силу доказанного найдётся матрица B такая, что ${}^t A \cdot B = E$. Снова обращаясь к п. 3 и полагая $A'' = {}^t B$, находим

$$E = {}^t E = {}^t ({}^t AB) = {}^t B {}^t ({}^t A) = A'' A.$$

Итак,

$$AA' = E = A'' A.$$

Остается заметить (см. (13)), что $A'' = A'$, а поэтому в соответствии с (14) $A' = A^{-1}$, т.е. матрица A обратима. \square

Следствие 1. *Если B и C — невырожденные квадратные матрицы порядков m и n соответственно, а A — произвольная $m \times n$ -матрица, то*

$$\operatorname{rank} BAC = \operatorname{rank} A.$$

Доказательство. В силу теорем 3 и 5 имеем

$$\begin{aligned} \operatorname{rank} BAC &\leq \operatorname{rank} BA = \operatorname{rank} BA(CC^{-1}) = \\ &= \operatorname{rank}(BAC)C^{-1} \leq \operatorname{rank} BAC, \end{aligned}$$

откуда $\operatorname{rank} BAC = \operatorname{rank} BA$. Аналогично устанавливается равенство

$$\operatorname{rank} BA = \operatorname{rank} A. \quad \square$$

Следствие 2. *Если $A, B \in M_n(\mathbb{R})$ и $AB = E$ или $BA = E$, то $B = A^{-1}$.*

Доказательство. Как показано в части 1) доказательства теоремы 5, $AB = E \implies \operatorname{rank} A = n$, т.е. A невырождена и, следовательно, обратима. \square

Следствие 3. *Если A, B, \dots, C, D — невырожденные $n \times n$ -матрицы, то произведение $AB \dots CD$ также невырождено и*

$$(AB \dots CD)^{-1} = D^{-1}C^{-1} \dots B^{-1}A^{-1}.$$

Доказательство. Невырожденность матрицы $G = AB \dots CD$ видна из следствия 1, а равенство $G^{-1} = D^{-1}C^{-1} \dots B^{-1}A^{-1}$ проверяется непосредственно:

$$\begin{aligned} G(D^{-1}C^{-1} \dots B^{-1}A^{-1}) &= AB \dots C(DD^{-1})C^{-1} \dots B^{-1}A^{-1} = \\ &= AB \dots (CC^{-1}) \dots B^{-1}A^{-1} = \dots = E. \quad \square \end{aligned}$$

Удобный способ вычисления обратной матрицы, обычно используемый на практике, будет приведён в п. 7. Одновременно получится ещё одно доказательство теоремы 5.

Явную формулу для A^{-1} мы укажем в гл. 3. Сейчас лишь заметим, что фактическое вычисление A^{-1} для матрицы A с числовыми коэффициентами или вычисление произведения двух матриц обычно требует выполнения большого числа операций. На практике встречаются матрицы порядка $n = 100$ и более. Если A и B — две такие матрицы, то для вычисления $C = AB$ нужно найти n^2 элементов c_{ij} по формуле (7) (или (9)), что в каждом случае требует $2n - 1$ умножений и сложений чисел. Всего нужно произвести $(2n - 1)n^2$ операций, т.е. около двух миллионов операций при $n = 100$. Для современных ЭВМ это — сравнительно лёгкая задача, но реальные трудности возникнут, если потребуется найти степень A^m матрицы A с показателем $m \geq 1000$. Здесь по определению $A^m = AA^{m-1}$; фактически $A^m = A^k A^{m-k}$, $0 \leq k \leq m$, — лёгкое следствие ассоциативности (см. следствие теоремы 2), как это будет показано в гл. 4 в более общем контексте. Для вычисления A^m используют разные дополнительные приёмы, либо основанные на специфике матрицы A , либо заимствованные из курса линейной алгебры. В качестве иллюстрации рассмотрим три примера.

Пример 1. Если

$$A = \text{diag}(\alpha_1, \dots, \alpha_n) = \begin{vmatrix} \alpha_1 & \dots & 0 \\ 0 & \dots & \alpha_n \end{vmatrix},$$

то, очевидно,

$$A^m = \text{diag}(\alpha_1^m, \dots, \alpha_n^m) = \begin{vmatrix} \alpha_1^m & \dots & 0 \\ 0 & \dots & \alpha_n^m \end{vmatrix}.$$

Пример 2. Пусть

$$A = \begin{vmatrix} a & c \\ 0 & b \end{vmatrix}.$$

Тогда индукция по m показывает, что

$$A^m = \begin{vmatrix} a^m & c \frac{a^m - b^m}{a - b} \\ 0 & b^m \end{vmatrix},$$

где

$$\frac{a^m - b^m}{a - b} = a^{m-1} + a^{m-2}b + \dots + ab^{m-2} + b^{m-1}.$$

В частности, при $a = b$ имеем

$$\begin{vmatrix} a & c \\ 0 & b \end{vmatrix}^m = \begin{vmatrix} a^m & ma^{m-1}c \\ 0 & b^m \end{vmatrix}.$$

Пример 3. Индукцией по m нетрудно убедиться в том, что m -я степень матрицы

$$A = \begin{vmatrix} 0 & 1 \\ 1 & 1 \end{vmatrix}$$

имеет вид

$$A^m = \begin{vmatrix} f_{m+1} & f_m \\ f_m & f_{m+1} \end{vmatrix}, \quad (16)$$

где целые числа $f_0 = 0, f_1 = 1, f_2 = 1, f_3 = 2, \dots$ определяются рекуррентным соотношением

$$f_{m+1} = f_m + f_{m-1}.$$

Это не что иное, как числа Фибоначчи (см. пример 2 в конце § 3 гл. 1).

Введём матрицу

$$B = \begin{vmatrix} -\frac{\lambda_2}{5} & \frac{1}{5} \\ -\sqrt{5}\lambda_1 & \sqrt{5} \end{vmatrix}$$

с определителем 1 (см. § 4 гл. 1), где

$$\lambda_1 = \frac{1 + \sqrt{5}}{2}, \quad \lambda_2 = \frac{1 - \sqrt{5}}{2}.$$

Небольшое вычисление показывает, что

$$B^{-1} = \begin{vmatrix} \sqrt{5} & -\frac{1}{5} \\ \sqrt{5}\lambda_1 & -\frac{\lambda_2}{5} \end{vmatrix}, \quad A = B^{-1} \cdot \begin{vmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{vmatrix} \cdot B.$$

Но если три произвольные $n \times n$ -матрицы A, B, C , из коих B невырожденная, связаны соотношением $A = B^{-1}CB$, то

$$A^m = B^{-1}CB \cdot B^{-1}CB \cdot B^{-1}CB \cdots B^{-1}CB = B^{-1}C^mB$$

(внутренние множители BB^{-1} , заменённые на E , сократились). В нашем случае с учётом примера 1 и соотношения (16) имеем

$$\begin{aligned} \begin{vmatrix} f_{m-1} & f_m \\ f_m & f_{m+1} \end{vmatrix} &= A^m = B^{-1} \begin{vmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{vmatrix}^m B = \\ &= B^{-1} \begin{vmatrix} \lambda_1^m & 0 \\ 0 & \lambda_2^m \end{vmatrix} B = \begin{vmatrix} \sqrt{5} & -\frac{1}{5} \\ \sqrt{5}\lambda_1 & -\frac{\lambda_2}{5} \end{vmatrix} \begin{vmatrix} \lambda_1^m & 0 \\ 0 & \lambda_2^m \end{vmatrix} B = \\ &= \begin{vmatrix} \sqrt{5}\lambda_1^m & -\frac{\lambda_2^m}{5} \\ \sqrt{5}\lambda_1^{m+1} & -\frac{\lambda_2^{m+1}}{5} \end{vmatrix} \begin{vmatrix} -\frac{\lambda_2}{5} & \frac{1}{5} \\ -\sqrt{5}\lambda_1 & \sqrt{5} \end{vmatrix} = \begin{vmatrix} * & \frac{1}{\sqrt{5}}(\lambda_1^m - \lambda_2^m) \\ ** & * \end{vmatrix} \end{aligned}$$

(звёздочками отмечены не интересующие нас члены).

Сравнивая коэффициенты матриц в левой и правой частях этого равенства, получаем для числа Фибоначчи с номером m значение

$$f_m = \frac{\lambda_1^m - \lambda_2^m}{\sqrt{5}} = \frac{1}{\sqrt{5}} \left\{ \left(\frac{1 + \sqrt{5}}{2} \right)^m - \left(\frac{1 - \sqrt{5}}{2} \right)^m \right\}.$$

Мы видим, что $f_m \sim \frac{1}{\sqrt{5}} \lambda_1^m$ при больших m (геометрическая прогрессия), поскольку $\lim_{m \rightarrow \infty} \left(\frac{1 - \sqrt{5}}{2} \right)^m = 0$.

6. Классы эквивалентных матриц. Как и при доказательстве теоремы 4, обозначим через E_{st} матрицу размера $m \times m$, в которой на пересечении s -й строки и t -го столбца стоит 1, а все остальные элементы нулевые (такие матрицы называются иногда *матричными единицами*). Рассмотрим в $M_m(\mathbb{R})$ так называемые *элементарные матрицы* следующих типов:

$$F_{s,t} = E - E_{ss} - E_{tt} + E_{st} + E_{ts} =$$

$$= \begin{vmatrix} 1 & & & \\ & \ddots & & \\ & & 0 & 1 \\ & & & \ddots & 1 \\ & & & & 1 \\ & & & & & \ddots & 0 \\ & & & & & & \ddots & 1 \end{vmatrix}, \quad s \neq t; \quad (\text{I})$$

$$F_{s,t} = E + \lambda E_{st} = \begin{vmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \dots & \lambda & \dots \\ & & & \ddots & & \\ & & & & & 1 \end{vmatrix}, \quad s \neq t; \quad (\text{II})$$

$$F_s(\lambda) = E + (\lambda - 1)E_{ss} = \text{diag}\{1, \dots, 1, \lambda, 1, \dots, 1\}, \quad \lambda \neq 0. \quad (\text{III})$$

Пусть A — произвольная $m \times n$ -матрица. Тогда непосредственно проверяется, что матрица $A' = FA$ получается из A посредством элементарного преобразования (э.п.) над строками типа (I) или (II) в зависимости от того, будет $F = F_{st}$ или $F = F_{st}(\lambda)$.

В случае $F = F_s(\lambda)$ будем говорить об э.п. типа (III) (умножение s -й строки $A_{(s)}$ на λ). Аналогично, матрица $A'' = AF$ получается из A посредством э.п. столбцов. Мы уже знаем из п. 2 § 2 и из упр. 2 из § 2, что э.п. типов (I) и (II), совершаемыми над строками и столбцами, A приводится к матрице с диагональной невырожденной подматрицей

в левом верхнем углу размера $r \times r$, где $r = \text{rank } A$ (при $r = 0$ матрица A нулевая). Так как

$$\begin{array}{c} \left\| \begin{array}{ccccc} a_1 & a_2 & & 0 & \\ & & \ddots & & \\ & & a_r & 0 & \\ 0 & & & & 0 \end{array} \right\| = \\ = F_1(a_1)F_2(a_2)\dots F_r(a_r) \left\| \begin{array}{ccccc} 1 & 1 & & 0 & \\ & & \ddots & & \\ & & 1 & 0 & \\ 0 & & & & 0 \end{array} \right\|, \end{array}$$

то привлечение э.п. типа (III) даёт возможность получить из A матрицу вида

$$\left\| \begin{array}{cc} E_r & 0 \\ 0 & 0 \end{array} \right\| \quad (17)$$

(здесь E_r — единичная матрица в $M_r(\mathbb{R})$; нули обозначают матрицы размеров $r \times (n-r)$, $(m-r) \times r$ и $(m-r) \times (n-r)$). Таким образом,

$$P_k P_{k-1} \dots P_1 A Q_1 Q_2 \dots Q_l = \left\| \begin{array}{cc} E_r & 0 \\ 0 & 0 \end{array} \right\|, \quad (18)$$

где P_i , (соответственно Q_j) — элементарные матрицы порядка m (соответственно n).

Не раз отмечалось, что элементарные операции обратимы. Это согласуется с существованием обратных матриц

$$(F_{s,t})^{-1} = F_{s,t}, \quad F_{s,t}(\lambda)^{-1} = F_{s,t}(-\lambda), \\ F_s(\lambda)^{-1} = F_s(\lambda^{-1}).$$

В соответствии со следствием 3 теоремы 5 матрицы $P = P_k P_{k-1} \dots P_1$ и $Q = Q_1 Q_2 \dots Q_l$ тоже обратимы:

$$P^{-1} = P_1^{-1} \dots P_{k-1}^{-1} P_k^{-1}, \quad Q^{-1} = Q_l^{-1} \dots Q_2^{-1} Q_1^{-1}.$$

Заметим, что P_i^{-1} , Q_j^{-1} — элементарные матрицы.

Две матрицы A, B размера $m \times n$ назовём эквивалентными и запишем $A \sim B$, если найдутся невырожденные матрицы порядков m и n соответственно такие, что $B = PAQ$.

Как легко понять, \sim является отношением эквивалентности:

- $A \sim A$ ($P = E_m, Q = E_n$);

- ii) $A \sim B \Rightarrow B \sim A$, поскольку $B = PAQ \Rightarrow A = P^{-1}BQ^{-1}$;
 iii) $B = P'AQ'$, $C = P''BQ'' \Rightarrow C = PAQ$, где $P = P''P'$, $Q = Q'Q''$.

Согласно общим принципам (см. § 6 гл. 1) множество всех $m \times n$ -матриц разбивается по отношению \sim на непересекающиеся классы эквивалентных матриц. Так как ранги эквивалентных матриц равны (см. следствие 1 теоремы 5), то рассуждение, приведшее нас к равенству (18), показывает, что в качестве представителей классов можно брать матрицы (17).

Мы получаем следующее утверждение.

Теорема 6. *Множество матриц размера $m \times n$ разбивается на $r = \min(m, n) + 1$ классов эквивалентности. Все матрицы ранга r попадают в один класс с представителем (17).*

Следствие. *Всякая невырожденная $n \times n$ -матрица записывается в виде произведения элементарных матриц.*

Действительно, все невырожденные матрицы порядка n попадают в один класс с представителем — единичной матрицей, поскольку их ранги равны n . Соотношение (18)

$$P_k P_{k-1} \dots P_1 A Q_1 Q_2 \dots Q_l = E,$$

переписанное в виде

$$A = P_1^{-1} \dots P_{k-1}^{-1} P_k^{-1} Q_l^{-1} \dots Q_2^{-1} Q_1^{-1}, \quad (19)$$

даёт нужное утверждение. \square

Не утверждается, что запись A в виде произведения элементарных матриц единственна, но сам факт существования такой записи весьма полезен. В частности, его можно использовать для отыскания обратной матрицы. В самом деле, из (19) мы находим

$$A^{-1} = Q_1 Q_2 \dots Q_l P_k P_{k-1} \dots P_1 = QP.$$

7. Вычисление обратной матрицы. Если в рассуждениях предыдущего пункта ограничиться преобразованиями над строками и рассмотреть с самого начала расширенную матрицу $(A|E)$ размера $n \times 2n$, то в случае невырожденной матрицы $A \in M_n(\mathbb{R})$ возникнет цепочка

$$(A|E) \xrightarrow{P_1} (P_1 A|P_1 E) \xrightarrow{P_2} \dots \xrightarrow{P_k} (P_k \dots P_2 P_1 A|P_k \dots P_2 P_1 E) = (E|A').$$

Она оборвётся на k -м шаге, когда в левой половине расширенной матрицы место A заполнит единичная матрица E . В правой половине при этом получится однозначный ответ: $A' = A^{-1}$. В случае вырожденной матрицы A процесс оборвётся, возможно, раньше — приведением A к ступенчатому виду и вычислением ранга $r = \text{rank } A$.

В матричной реализации, с которой начинался п. 6, при $n = 3$ имеем

$$F_{1,2}(-3) = \begin{vmatrix} 1 & -3 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix}, \quad F_{3,2}(4) = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 4 & 1 \end{vmatrix},$$

$$F_{1,3} = \begin{vmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{vmatrix}.$$

Реально элементарные $n \times n$ -матрицы P_i слева не приписываются. На них следует смотреть как на предписания и выполнять соответствующие им э.п. над строками.

Напомним ещё раз о значении символов:

$P_i = F_{s,t}$ — переставить местами строки с номерами s и t ;

$P_i = F_{s,t}(\lambda)$ — прибавить к s -й строке t -ю строку, умноженную на λ ;

$P_i = F_s(\lambda)$ — умножить s -ю строку на λ .

Пример 4. Пусть

$$A = \begin{vmatrix} 0 & 2 & 0 \\ 1 & 1 & -1 \\ 2 & 1 & -1 \end{vmatrix}.$$

Имеем

$$(A|E) = \left\| \begin{array}{ccc|ccc} 0 & 2 & 0 & 1 & 0 & 0 \\ 1 & 1 & -1 & 0 & 1 & 0 \\ 2 & 1 & -1 & 0 & 0 & 1 \end{array} \right\| \xrightarrow{F_{1,2}} \left\| \begin{array}{ccc|ccc} 1 & 1 & -1 & 0 & 1 & 0 \\ 0 & 2 & 0 & 1 & 0 & 0 \\ 2 & 1 & -1 & 0 & 0 & 1 \end{array} \right\| \xrightarrow{F_2(1/2)} \left\| \begin{array}{ccc|ccc} 1 & 1 & -1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1/2 & 0 & 0 \\ 2 & 1 & -1 & 0 & -2 & 1 \end{array} \right\| \xrightarrow{F_{1,2}(-1)} \left\| \begin{array}{ccc|ccc} 1 & 0 & -1 & -1/2 & 1 & 0 \\ 0 & 1 & 0 & 1/2 & 0 & 0 \\ 2 & 1 & -1 & 0 & -2 & 1 \end{array} \right\| \xrightarrow{F_{3,2}(1)} \left\| \begin{array}{ccc|ccc} 1 & 0 & -1 & -1/2 & 1 & 0 \\ 0 & 1 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 1 & 1/2 & -2 & 1 \end{array} \right\| \xrightarrow{F_{1,3}(1)} \left\| \begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & -1 & 1 \\ 0 & 1 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 1 & 1/2 & -2 & 1 \end{array} \right\|.$$

Таким образом,

$$A^{-1} = \begin{vmatrix} 0 & -1 & 1 \\ 1/2 & 0 & 0 \\ 1/2 & -2 & 1 \end{vmatrix}.$$

Для экономии места целесообразно выполнять сразу серию однотипных преобразований.

Пример 5. Пусть

$$A = \begin{vmatrix} -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \end{vmatrix}.$$

Имеем

$$(A|E) = \left(\begin{array}{cccc|cccc} -1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & -1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & -1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & -1 & 0 & 0 & 0 & 1 \end{array} \right) \xrightarrow{\substack{F_{1,4}(1) \\ F_{1,3}(1)}} \left(\begin{array}{cccc|cccc} 2 & 2 & 2 & 2 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & -1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & -1 & 0 & 0 & 0 & 1 \end{array} \right) \xrightarrow{F_1(1/2)} \left(\begin{array}{cccc|cccc} 1 & 1 & 1 & 1 & 1/2 & 1/2 & 1/2 & 1/2 \\ 1 & -1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & -1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & -1 & 0 & 0 & 0 & 1 \end{array} \right) \xrightarrow{\substack{F_{4,1}(-1) \\ F_{3,1}(-1)}} \left(\begin{array}{cccc|cccc} 1 & 1 & 1 & 1 & 1/2 & 1/2 & 1/2 & 1/2 \\ 0 & -2 & 0 & 0 & -1/2 & 1/2 & -1/2 & -1/2 \\ 0 & 0 & -2 & 0 & -1/2 & -1/2 & 1/2 & -1/2 \\ 0 & 0 & 0 & -2 & -1/2 & -1/2 & -1/2 & 1/2 \end{array} \right) \xrightarrow{\substack{F_4(-1/2) \\ F_3(-1/2)}} \left(\begin{array}{cccc|cccc} 1 & 1 & 1 & 1 & 1/2 & 1/2 & 1/2 & 1/2 \\ 0 & 1 & 0 & 0 & 1/4 & -1/4 & 1/4 & 1/4 \\ 0 & 0 & 1 & 0 & 1/4 & 1/4 & -1/4 & 1/4 \\ 0 & 0 & 0 & 1 & 1/4 & 1/4 & 1/4 & -1/4 \end{array} \right) \xrightarrow{\substack{F_{1,4}(-1) \\ F_{1,3}(-1)}} \left(\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & -1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 1 & 0 & 0 & 1/4 & -1/4 & 1/4 & 1/4 \\ 0 & 0 & 1 & 0 & 1/4 & 1/4 & -1/4 & 1/4 \\ 0 & 0 & 0 & 1 & 1/4 & 1/4 & 1/4 & -1/4 \end{array} \right).$$

Таким образом, $A^{-1} = (1/4)A$.

Впрочем, в данном случае вычислений можно было бы избежать. Замечая, что произведение вырожденной матрицы и произвольной матрицы всегда вырожденно (теорема 3), в то время как

$$A^2 = \begin{vmatrix} 4 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 4 \end{vmatrix} = 4E,$$

мы делаем заключение о невырожденности A и, следовательно, о существовании A^{-1} . Но коль скоро это так, то

$$A = A^2 A^{-1} = 4E \cdot A^{-1} = 4A^{-1} \Rightarrow A^{-1} = \frac{1}{4}A.$$

Замечание. При выполнении серии преобразований над строками следует избегать типичной ошибки — прибавления в неизменном виде строки, изменившейся в ходе предыдущих преобразований. Например, предписание

$$\begin{array}{ccc} F_{2,1}(1) \\ A & \longrightarrow & A' \\ F_{1,2}(1) \end{array}$$

двусмысленно: не ясно, в каком порядке действовать — сначала $F_{1,2}(1)$, потом $F_{2,1}(1)$; сначала $F_{2,1}(1)$, потом $F_{1,2}(1)$ или одновременно? Каждый раз будут получаться различные выражения для строк $A'_{(1)}, A'_{(2)}$. В примере 5 мы объединяли лишь однотипные преобразования, а если ставить своей целью вычисление на ЭВМ по указанному методу, то естественно всю последовательность элементарных преобразований линейно упорядочить.

Рассмотренный нами метод вычисления ранга, а также обратной матрицы называется *P-приведением* или, более общо, *(P, Q)-приведением матриц к нормальному виду* (17).

8. Пространство решений. Из вводных замечаний в начале § 2 и § 3 следует, что система линейных уравнений с матрицей A размера $m \times n$ и столбцом свободных членов $B \in \mathbb{R}^m$ может быть записана коротко в виде

$$AX = B \quad (20)$$

($X = [x_1, \dots, x_n]$ — столбец высоты n). Представив, что $m = n$ и квадратная матрица A невырождена (см. п. 5), мы получим, и при этом единственное, решение системы (20), умножая обе части матричного соотношения слева на A^{-1} : $X = EX = (A^{-1}A)X = A^{-1}(AX) = = A^{-1}B$. Эта удобная символическая запись решений определённой системы не избавляет нас от вычислений, поскольку матрица A^{-1} заранее не дана. Но не откажем себе в удовольствии заметить, что матричный аппарат доставляет по меньшей мере эстетическое наслаждение. Воспользуемся им теперь для обозрения всех решений линейной однородной системы (ЛОС):

$$AX = 0. \quad (21)$$

По существу мы уже знаем, что если $X^{(1)}, X^{(2)}$ — решения нашей ЛОС, то и любая их линейная комбинация тоже будет решением:

$$A(\alpha_1 X^{(1)} + \alpha_2 X^{(2)}) = \alpha_1 AX^{(1)} + \alpha_2 AX^{(2)} = 0.$$

Поэтому можно говорить о *пространстве решений* ЛОС — линейной оболочке

$$V_A = \langle X \in \mathbb{R}^n \mid AX = 0 \rangle \subset \mathbb{R}^n.$$

Пусть $s = \dim V_A$, $r = \operatorname{rank} A$. По определению $s \leq n$, $r \leq \min(m, n)$. Какая связь существует между s и r ?

Теорема 7. Имеет место равенство $r + s = n$.

Доказательство. Выберем базис $X^{(1)}, \dots, X^{(s)}$ линейной оболочки V_A и дополним его до базиса $X^{(1)}, \dots, X^{(s)}, X^{(s+1)}, \dots, X^{(n)}$ всего пространства \mathbb{R}^n . Это всегда можно сделать, как показывает доказательство теоремы 2 § 1 (и упр. 6 из § 1). Для любого вектора $X = \sum_{i=1}^n \alpha_i X^{(i)} \in \mathbb{R}^n$ имеем

$$AX = \sum_{i=1}^n \alpha_i AX^{(i)} = \alpha_{s+1} AX^{(s+1)} + \dots + \alpha_n AX^{(n)},$$

так что в соответствии с § 2 линейная оболочка

$$\begin{aligned} V_B(A) &= \langle A^{(1)}, \dots, A^{(n)} \rangle = \langle x_1 A^{(1)} + \dots + x_n A^{(n)} \mid x_i \in \mathbb{R}^n \rangle = \\ &= \langle AX \mid X \in \mathbb{R}^n \rangle \subset \mathbb{R}^m, \end{aligned}$$

называемая *пространством столбцов* матрицы A , совпадает с линейной оболочкой $\langle AX^{(s+1)}, \dots, AX^{(n)} \rangle$.

В частности, $r = \dim V_B(A) \leq n - s$. Но векторы $AX^{(s+1)}, \dots, AX^{(n)}$ линейно независимы, поскольку из

$$0 = \sum_{k \geq s+1} \beta_k AX^{(k)} = A \left(\sum_{k \geq s+1} \beta_k X^{(k)} \right)$$

следует $\sum_{k \geq s+1} \beta_k X^{(k)} \in V_A$, а это в силу выбора $X^{(s+1)}, \dots, X^{(n)}$ возможно только при $\beta_{s+1} = \dots = \beta_n = 0$. Значит, $r = n - s$. \square

Замечание. Если использовать язык линейных отображений (см. п. 1 § 3), то, очевидно,

$$V_A = \operatorname{Ker} \varphi_A, \quad V_B(A) = \operatorname{Im} \varphi_A$$

— ядро и образ отображения $\varphi_A : \mathbb{R}^n \rightarrow \mathbb{R}^m$, отвечающего A . Для нас, однако, этот подход служит лишь мотивировкой для введения матричных понятий.

Чтобы найти базис пространства V_A , выберем в A r базисных столбцов одним из способов — приведением A к ступенчатому виду или так, как это указано в гл. 3. Перестановкой столбцов или, что равносильно, перенумерацией неизвестных можно добиться, чтобы базисными были r первых столбцов $A^{(1)}, \dots, A^{(r)}$. При этом в новой системе неизвестных x'_1, x'_2, \dots, x'_n главными неизвестными станут x'_1, \dots, x'_{r+k} . Любая система из $r+1$ столбцов $A^{(1)}, \dots, A^{(r)}, A^{(r+k)}$, $k > 0$, будет линейно зависимой, и на основании теоремы 1, в) из § 1 можно выписать систему соотношений

$$x_1^{(k)} A^{(1)} + x_2^{(k)} A^{(2)} + \dots + x_r^{(k)} A^{(r)} + A^{(r+k)} = 0, \quad k = 1, 2, \dots, n - r.$$

Векторы-столбцы

$$\begin{aligned} X^{(1)} &= [x_1^{(1)}, x_2^{(1)}, \dots, x_r^{(1)}, 1, 0, \dots, 0], \\ X^{(2)} &= [x_1^{(2)}, x_2^{(2)}, \dots, x_r^{(2)}, 0, 1, \dots, 0], \\ &\dots \\ X^{(n-r)} &= [x_1^{(n-r)}, x_2^{(n-r)}, \dots, x_r^{(n-r)}, 0, 1, \dots, 0] \end{aligned} \quad (22)$$

в количестве $n - r$ штук, очевидно, линейно независимы (из-за специального вида своих последних $n - r$ компонент) и, будучи решениями ЛОС (21), составляют по теореме 7 базис пространства V_A всех её решений. Понятно, что решение $X^{(k)}$ получается, если новым (штрихованным) свободным неизвестным придать значения

$$x'_{r+1} = 0, \dots, x'_{r+k} = 1, \dots, x'_n = 0.$$

Любой базис пространства решений однородной системы $AX = 0$ ранга r называется *фундаментальной системой решений*. Систему (22) называют ещё *нормальной фундаментальной системой*. Согласно следствию теоремы 1 § 2 её ранг $s = \dim V_A = n - r$ равен числу свободных неизвестных линейной системы.

УПРАЖНЕНИЯ

1. Даны отображения:

- a) $[x_1, x_2, \dots, x_n] \mapsto [x_n, \dots, x_2, x_1]$;
 б) $[x_1, x_2, \dots, x_n] \mapsto [x_1, x_2^2, \dots, x_n^n]$;
 в) $[x_1, x_2, \dots, x_n] \mapsto [x_1, x_1 + x_2, \dots, x_1 + x_2 + \dots + x_n]$.

Какие из них являются линейными?

2. Доказать, что

$$\left| \begin{array}{ccc} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{array} \right|^m = \left| \begin{array}{ccc} 1 & ma & \frac{m(m-1)}{2}ab + mc \\ 0 & 1 & mb \\ 0 & 0 & 1 \end{array} \right|.$$

Найти для $\left| \begin{array}{ccc} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{array} \right|$ обратную матрицу.

3. Проверить, что $\left| \begin{array}{cc} 0 & -1 \\ 1 & -1 \end{array} \right|^3 = E$.

4. В приложениях большую роль играют *марковские* (или *стохастические*) матрицы

$$P = (P_{ij}), \quad p_{ij} \geq 0, \quad \sum_{j=1}^n p_{ij} = 1, \quad i = 1, 2, \dots, n.$$

Линейные отображения φ_P , ассоциированные с марковскими матрицами, обычно применяют к специальным так называемым вероятностным векторам-столбцам

$$X = [x_1, \dots, x_n], \quad x_i \geq 0, \quad \sum_{i=1}^n x_i = 1.$$

Согласованность этих определений, диктуемых естественнонаучными задачами, видна из следующих утверждений, которые нужно доказать хотя бы при $n = 2$.

а) Матрица $P \in M_n(\mathbb{R})$ является марковской в точности тогда, когда вместе с любым вероятностным вектором X вектор PX также является вероятностным (здесь $PX = \varphi_P(X)$).

б) Если P — положительная марковская матрица ($\forall i, j \ p_{ij} > 0$), то любому вероятностному вектору X отвечает положительный вероятностный вектор PX (все компоненты строго больше нуля).

в) Если P и Q — марковские матрицы, то марковской будет также матрица PQ . Это означает, в частности, что любая степень P^k марковской матрицы является марковской.

5. Найти ${}^t H \cdot H$, если

$$H = \begin{vmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{vmatrix}.$$

6. Ассоциировав с циклом длины n в S_n (см. § 8 гл. 1) матрицу перестановки (строк единичной матрицы E_n)

$$P = \begin{vmatrix} 0 & 0 & \dots & 0 & 1 \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 1 & 0 \end{vmatrix},$$

проверить, что $P^n = E$.

7. Показать, что

$$\operatorname{rank}(A + B) \leq \operatorname{rank} A + \operatorname{rank} B$$

для любых двух $m \times n$ -матриц A и B .

8. Показать, что для любой $m \times s$ -матрицы A и любой $s \times n$ -матрицы B имеет место неравенство

$$\operatorname{rank} A + \operatorname{rank} B - s \leq \operatorname{rank} AB.$$

9. Показать, что если $ABC = 0$ для квадратных матриц A, B, C порядка n , то

$$\operatorname{rank} A + \operatorname{rank} B + \operatorname{rank} C \leq 2n.$$

10. Найти ранг матрицы

$$A = \begin{vmatrix} x_1 y_1 & x_1 y_2 & \dots & x_1 y_n \\ x_2 y_1 & x_2 y_2 & \dots & x_2 y_n \\ \vdots & \vdots & \ddots & \vdots \\ x_n y_1 & x_n y_2 & \dots & x_n y_n \end{vmatrix}.$$

Указание. Показать, что $A = [x_1, \dots, x_n][y_1, \dots, y_n]$.

11. Показать, что если $A = (a_{ij})$ — невырожденная симметрическая матрица ($a_{ij} = a_{ji}$), то и A^{-1} — симметрическая матрица.

12. Найти A^{-1} и F^{-1} , если

$$A = \begin{vmatrix} 5 & 4 & 3 & 2 & 1 \\ 4 & 8 & 6 & 4 & 2 \\ 3 & 6 & 9 & 6 & 3 \\ 2 & 4 & 6 & 8 & 4 \\ 1 & 2 & 3 & 4 & 5 \end{vmatrix}, \quad F = \begin{vmatrix} 2 & 3 & 2 & 1 \\ 3 & 6 & 4 & 2 \\ 4 & 8 & 6 & 3 \\ 2 & 4 & 3 & 2 \end{vmatrix}.$$

13. Проверить, что

$$A = \begin{vmatrix} a & b \\ c & d \end{vmatrix}, ad - bc \neq 0 \implies A^{-1} = \frac{1}{ad - bc} \begin{vmatrix} d & -b \\ -c & a \end{vmatrix}.$$

В частности,

$$ad - bc = 1 \implies A^{-1} = \begin{vmatrix} d & -b \\ -c & a \end{vmatrix}.$$

Существует ли A^{-1} при $ad - bc = 0$?

14. Доказать, что для любой матрицы

$$A = \begin{vmatrix} a & b \\ c & d \end{vmatrix}$$

выполнено соотношение

$$A^2 = (a + d)A - (ad - bc)E \quad (23)$$

(другими словами, A является "корнем" квадратного уравнения $x^2 - (a + d)x + (ad - bc) = 0$).

15. При $ad - bc \neq 0$ использовать соотношение (23) для нахождения обратной матрицы A^{-1} .

16. Доказать, что если $\begin{vmatrix} a & b \\ c & d \end{vmatrix}^m = 0$, то $\begin{vmatrix} a & b \\ c & d \end{vmatrix}^2 = 0$.

17. Обосновать следующее рассуждение. Пусть $m \times s$ -матрица X разбита горизонтальными и вертикальными прямыми на блоки (или клетки), так что

$$X = \begin{vmatrix} X_{11} & X_{12} & \dots & X_{1k} \\ X_{21} & X_{22} & \dots & X_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ X_{l1} & X_{l2} & \dots & X_{lk} \end{vmatrix}$$

где X_{i1}, \dots, X_{ik} — матрицы с одинаковым числом m_i строк ($m_1 + \dots + m_k = m$), а X_{1j}, \dots, X_{lj} — матрицы с одинаковым числом s_j столбцов ($s_1 + \dots + s_k = s$). Если теперь

$$Y = \begin{vmatrix} Y_{11} & Y_{12} & \dots & Y_{1r} \\ Y_{21} & Y_{22} & \dots & Y_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{k1} & Y_{k2} & \dots & Y_{kr} \end{vmatrix}$$

— $s \times n$ -матрица с блоками Y_{ij} размеров $s_i \times n_j$ ($n_1 + \dots + n_r = n$), то имеет смысл говорить о произведении $Z = XY$, причём матрицу $Z = (z_{ij})$ тоже можно считать блочной с блоками Z_{ij} , вычисляемыми формально по формуле (7):

$$Z_{ij} = X_{i1}Y_{1j} + X_{i2}Y_{2j} + \dots + X_{ir}Y_{rj}.$$

По условию размеры матриц X_{iv}, Y_{vj} таковы, что произведение $X_{iv}Y_{vj}$ имеет смысл. Приём разбиения матриц на блоки удобен даже в таком простейшем случае, как

$$\begin{vmatrix} E & A \\ 0 & E \end{vmatrix} \begin{vmatrix} A & 0 \\ -E & B \end{vmatrix} = \begin{vmatrix} 0 & AB \\ -E & B \end{vmatrix},$$

где $A, B, E, 0 \in M_n(\mathbb{R})$ (E — единичная, а 0 — нулевая матрица).

18. Показать, что умножение матрицы

$$X = (x_{ij}) \in M_n(\mathbb{R})$$

на $T = (t_{ij}) \in M_n(\mathbb{R})$ слева равносильно линейному комбинированию строк $X_{(1)}, \dots, X_{(n)}$, а справа — линейному комбинированию столбцов $X^{(1)}, \dots, X^{(n)}$. В частности, обратить внимание на то, что если

$$T = \begin{vmatrix} 1 & t_{12} & t_{13} & \dots & t_{1n} \\ 0 & 1 & t_{23} & \dots & t_{2n} \\ 0 & 0 & 0 & \dots & 1 \end{vmatrix}$$

— верхняя унитрэугольная матрица, то

$$TX = \begin{vmatrix} X_{(1)} + t_{12}X_{(2)} + \dots + t_{1n}X_{(n)} \\ X_{(2)} + \dots + t_{2n}X_{(n)} \\ \dots \\ X_{(n)} \end{vmatrix}$$

— матрица, полученная из X посредством цепочки элементарных преобразований типа (II) над строками.

Глава 3

ОПРЕДЕЛИТЕЛИ

Формулы (3) и (9) из § 4 гл. 1 для решений квадратных линейных систем порядков $n = 2, 3$ наводят на мысль о существовании подобных формул при любом n .

В конечном счёте речь идёт о правильной интерпретации в каждой из упомянутых формул числителя и знаменателя. Мы будем смотреть на них как на значения некоторой “универсальной” функции $\det : M_n(\mathbb{R}) \rightarrow \mathbb{R}$ из множества квадратных матриц порядка n в \mathbb{R} . Эффективное построение функции \det (определителя) даст ответ также на многие другие вопросы о матрицах, поднятые в гл. 2. На самом деле роль теории определителей в математике гораздо шире затронутой нами темы, и каждое из применений этой теории подсказывает собственный путь её построения. Один из наиболее естественных подходов — геометрический, основанный на аналогии “определители матриц — объёмы многомерных фигур” и на внешних n -формах. Так как для этого нужно чуточку больше техники, то мы остановимся на аналитическом пути, апеллируя к геометрической интуиции лишь в самом начале.

§ 1. Определители: построение и основные свойства

1. Геометрическая мотивировка. Ничто сейчас не мешает ввести общее понятие определителя, но попытаемся на время забыть о нашей задаче, обратившись к вычислению объёмов простейших геометрических фигур — параллелепипедов. Квадратной матрице $A = (a_{ij})$ порядка n поставим в соответствие параллелепипед

$$\Pi(A) = \Pi(A^{(1)}, A^{(2)}, \dots, A^{(n)}),$$

ребра которого задаются столбцами матрицы $A^{(1)}, A^{(2)}, \dots, A^{(n)}$, т.е. векторами (или точками) $A^{(j)} = [a_{1j}, a_{2j}, \dots, a_{nj}] \in \mathbb{R}^n$. Под $\Pi(A)$ нужно понимать подмножество в \mathbb{R}^n , состоящее из всех точек вида

$$x_1 A^{(1)} + \dots + x_n A^{(n)}, \quad 0 \leq x_i \leq 1$$

(мы незаметно перешли к отождествлению векторов-столбцов с их концевыми точками в пространстве с прямоугольной системой координат). При $n = 1$ параллелепипед называется *отрезком*, а при $n = 2$ — *параллелограммом*.

Объём $v(\Pi(A))$ n -мерного параллелепипеда определяется по индукции как произведение объёма $v(\Pi(A^{(1)}, \dots, A^{(n-1)}))$ $(n-1)$ -мерного основания в \mathbb{R}^{n-1} и длины h перпендикуляра $A^{(n)}P$, опущенного на гиперплоскость этого основания из точки $A^{(n)}$. Под объёмом

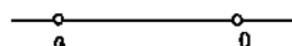
отрезка ($n = 1$) понимается, конечно, его *длина*, а под объёмом параллелограмма ($n = 2$) — его *площадь*. В общую теорию измерений объёмов мы сейчас не входим.

Прямые вычисления показывают, что с точностью до знака

$$\begin{aligned} n = 2 : \quad v(\Pi(A^{(1)}, A^{(2)})) &= \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}; \\ n = 3 : \quad v(\Pi(A^{(1)}, A^{(2)}, A^{(3)})) &= \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \end{aligned} \quad (1)$$

(определители матриц порядков 2 и 3 вводятся соответственно формулами (2) и (8) из § 4 гл. 1).

Соблазнительно было бы сохранить формулы типа (1) без оговорок, т.е. при любом расположении точек $A^{(1)}, A^{(2)}, \dots$, но это возможно только в том случае, если пользоваться понятием *ориентированного объёма* параллелепипеда с допустимыми отрицательными значениями. В частности, для отрезка



ориентированной длиной будет $a < 0$. Для параллелограмма $\Pi(A^{(1)}, A^{(2)})$ площадь берётся со знаком плюс, если упорядоченная пара векторов $(A^{(1)}, A^{(2)})$ задает ту же ориентацию плоскости \mathbb{R}^2 , что и базисная пара векторов (e_1, e_2) ; в противном случае — со знаком минус. При таком понимании естественно обратить формулу (1) и считать при любом n определителем $\det A$ матрицы A ориентированный объём параллелепипеда, обозначаемый тем же символом:

$$\det A = v(\Pi(A)).$$

Базисный вектор e_j , отвечает стандартному столбцу $E^{(j)} = [0, \dots, 1, \dots, 0]$, так что

$$A^{(j)} = \varphi_A(E^{(j)})$$

— образ единичного вектора при линейном отображении $\varphi_A : X \mapsto \mapsto AX$ (см. § 3 гл. 2). Образом единичного куба $\Pi(E)$ при отображении φ_A будет как раз параллелепипед $\Pi(A)$, а поскольку $v(\Pi(E)) = 1$, определитель $\det \varphi_A = \det A$ равен коэффициенту изменения ориентированного объёма. На самом деле при применении φ_A ориентированный объём любой фигуры, а не только единичного куба, меняется в $\det A$ раз (см. [ВА II]).

Обратим внимание на легко проверяемые свойства ориентированной площади параллелограмма:

$$1) \quad v(\Pi(A^{(1)}, A^{(2)})) = -v(\Pi(A^{(2)}, A^{(1)}));$$

$$2) \quad v(\Pi(A^{(1)} + \lambda A^{(2)}, A^{(2)})) = v(\Pi(A^{(1)}, A^{(2)}));$$

3) $v(\Pi(E)) = 1$.

О свойствах 1) и 3) говорилось выше, а свойство 2) проиллюстрировано (при $n = 2$) на рис. 14 и основано на идее равносоставленности. При $n > 3$ свойства 1)-3) объёмов параллелепипедов уже менее наглядны, но совершенно очевидно, что при любом подходе к теории определителей отмеченные три свойства должны выполняться.

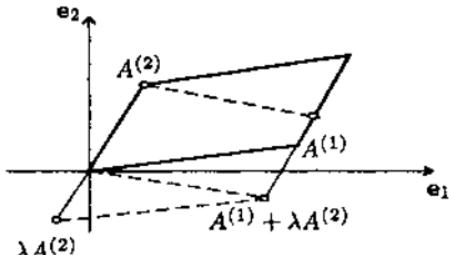


Рис. 14

Кроме того, должны быть получены и другие свойства определителей так, чтобы вычисление $\det A$ для любой фиксированной квадратной матрицы A , а следовательно, и вычисление объёма $v(\Pi(A))$, было алгоритмически реализуемым и легко осуществимым актом.

2. Комбинаторно-аналитический подход. Близкие обозначения

$$A = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix}, \quad \det A = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix}, \quad (2)$$

которые для нас не новы и которыми мы будем постоянно пользоваться в дальнейшем, существенно различны. Если A — квадратная таблица, заполненная своими коэффициентами (обычно числами), то определитель порядка n как та же таблица, но ограниченная вертикальными чёрточками, — это число (или выражение), приписываемое матрице A и определённое *формулой полного развертывания*

$$\det A = \sum_{\sigma \in S_n} \varepsilon_\sigma a_{1,\sigma 1} a_{2,\sigma 2} \dots a_{n,\sigma n}. \quad (3)$$

Другими словами, *определителем* $\det A$ матрицы $A = (a_{ij})$ называется алгебраическая сумма всевозможных произведений коэффициентов a_{ij} , взятых по одному из каждой строки и из каждого столбца. В каждом произведении сомножители записываются в порядке следования строк, а номера столбцов определяются образами $\sigma 1, \sigma 2, \dots, \sigma n$ номеров строк при перестановке $\sigma \in S_n$. Всего под знаком суммы в (3) стоит $n!$ слагаемых; слагаемые, отвечающие чётным перестановкам, входят со знаком плюс, а отвечающие нечётным перестановкам, — со знаком минус. Тех и других, согласно соотношению (11) из § 8 гл. 1, — одинаковое число $n!/2$.

Как показывает несложная проверка, формула (3) при $n = 2$ и $n = 3$ приводит к известным нам выражениям. Пусть $n = 4$, и пусть $\sigma = (1\ 2)(3\ 4)$. Тогда $\varepsilon_\sigma = 1$, а $a_{1,\sigma 1} a_{2,\sigma 2} a_{3,\sigma 3} a_{4,\sigma 4} = a_{12} a_{21} a_{34} a_{43}$.

Это значит, что в определитель четвёртого порядка слагаемое $a_{12}a_{21}a_{34}a_{43}$ входит со знаком плюс. В качестве полезного упражнения, рассчитанного на прочное владение материалом § 8 гл. 1, стоит выписать все 24 члена этого определителя и внимательно проследить за расстановкой знаков. Кстати, уже при $n = 5$ подобное задание с выписыванием 120 членов выглядело бы бессмысленным. Между тем, следуя наводящим соображениям из п. 1, мы хотели бы извлечь из исходной формулы (3) все нужные нам свойства определителей любого порядка.

3. Основные свойства определителей. Этих свойств немногого, но для формулировки и, главное, для их понимания нужно уговориться о терминологии и обозначениях.

В дальнейшем, как и в гл. 2, символами

$$A_{(i)} = (a_{i1}, a_{i2}, \dots, a_{in}), \quad i = 1, 2, \dots, n,$$

$$A^{(j)} = [a_{1j}, a_{2j}, \dots, a_{nj}], \quad j = 1, 2, \dots, n,$$

будут обозначаться соответственно i -я строка и j -й столбец матрицы $A = (a_{ij})$. Сама матрица A представляется либо как объединение своих строк:

$$A = [A_{(1)}, A_{(2)}, \dots, A_{(n)}]$$

(столбец строк), либо как объединение своих столбцов:

$$A = (A^{(1)}, A^{(2)}, \dots, A^{(n)})$$

(строка столбцов). Условимся впредь строки и столбцы $n \times n$ -матрицы A называть также *строками* и *столбцами определителя* $|a_{ij}|$ порядка n .

Согласно определению $| | = \det$ (от англ. *determinant*) — функция, сопоставляющая квадратной матрице A некоторое число $|A| = \det A$. Наша задача — изучить поведение этой функции при изменении строк или столбцов матрицы A , рассматриваемых как элементы (векторы) линейного пространства \mathbb{R}^n . Если угодно, для нас $\det A$ — сокращённое обозначение (в духе п. 2 § 5 гл. 1) функции $\det[A_{(1)}, \dots, A_{(n)}]$ или $\det(A^{(1)}, \dots, A^{(n)})$ n переменных, коими являются векторы из \mathbb{R}^n .

Произвольную функцию $D : [A_{(1)}, \dots, A_{(n)}] \mapsto D(A_{(1)}, \dots, A_{(n)})$ мы будем называть *полилинейной*, если она линейна по каждому аргументу $A_{(i)}$, т.е.

$$D(A_{(1)}, \dots, \alpha A'_{(i)} + \beta A''_{(i)}, \dots, A_{(n)}) =$$

$$= \alpha D(A_{(1)}, \dots, A'_{(i)}, \dots, A_{(n)}) + \beta D(A_{(1)}, \dots, A''_{(i)}, \dots, A_{(n)})$$

(ср. с п. 1 § 3 гл. 2). Та же функция D называется *кососимметри-*

ческой (см. п. 4 § 8 гл. 1), если

$$\begin{aligned} \mathcal{D}(A_{(1)}, \dots, A_{(i)}, A_{(i+1)}, \dots, A_{(n)}) = \\ = -\mathcal{D}(A_{(1)}, \dots, A_{(i+1)}, A_{(i)}, \dots, A_{(n)}), \quad 1 \leq i \leq n-1. \end{aligned} \quad (4)$$

Замечание 1. Из определения линейных функций (см. (4) § 3 гл. 2) можно заключить, что функция \mathcal{D} полилинейна ровно тогда, когда при фиксированных $A_{(1)}, \dots, A_{(i-1)}, A_{(i+1)}, \dots, A_{(n)}$ и при $A_{(i)} = X = (x_1, \dots, x_n)$ мы имеем

$$\mathcal{D}(A_{(1)}, \dots, A_{(n)}) = \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_n x_n,$$

где $\alpha_1, \dots, \alpha_n$ — скаляры, не зависящие от x_1, \dots, x_n .

Замечание 2. Кососимметричность полилинейной функции \mathcal{D} эквивалентна выполнению соотношений

$$\mathcal{D}(A_{(1)}, \dots, A_{(i-1)}, X, X, A_{(i+1)}, \dots, A_{(n)}) = 0, \quad 1 \leq i \leq n-1. \quad (4')$$

В самом деле, положив $A_{(i)} = A_{(i+1)} = X$ в (4), мы придём к (4'). Обратно, при $X = A_{(i)} + A_{(i+1)}$ из (4') вытекает в силу полилинейности \mathcal{D} соотношение

$$\begin{aligned} \mathcal{D}(\dots, A_{(i)}, A_{(i)}, \dots) + \mathcal{D}(\dots, A_{(i+1)}, A_{(i+1)}, \dots) + \\ + \mathcal{D}(\dots, A_{(i)}, A_{(i+1)}, \dots) + \mathcal{D}(\dots, A_{(i+1)}, A_{(i)}, \dots) = \\ = \mathcal{D}(\dots, A_{(i)} + A_{(i+1)}, A_{(i)} + A_{(i+1)}, \dots) = 0. \end{aligned}$$

Первые два члена равны нулю (положить в (4') соответственно $X = A_{(i)}$ и $X = A_{(i+1)}$), поэтому равна нулю сумма двух последних членов, что является лишь иной записью соотношения (4).

Те же определения и замечания относятся к функции $\mathcal{D}(A^{(1)}, \dots, A^{(n)})$ векторов-столбцов. Более того, условие (2) кососимметричности применимо к любой функции $\mathcal{D}: M^n \rightarrow \mathbb{R}$, где M^n — декартова степень некоторого множества M . Напомним ещё, что согласно лемме 2 гл. 1 при перестановке местами любых двух аргументов **кососимметрическая функция меняет знак на противоположный**.

Обратим внимание на то обстоятельство, что в формулу (3) строки и столбцы матрицы A входят, на первый взгляд, “неравноправным” образом. Но если в A поменять местами строки и столбцы, то получится транспонированная матрица ${}^t A$ (см. п. 3 § 3 гл. 2). Стало быть, речь идёт о сравнении двух величин: $\det A$ и $\det {}^t A$. Ответ даёт

Теорема 1. *Определители любой квадратной матрицы A и транспонированной с ней матрицы ${}^t A$ совпадают:*

$$\det {}^t A = \det A.$$

Доказательство. Положив $A = (a_{ij})$, ${}^t A = (a'_{ij})$, где $a'_{ij} = a_{ji}$, и заметив, что $k = \pi(\pi^{-1}k)$ для любой перестановки $\pi \in S_n$ и для

любого номера $k \in \{1, 2, \dots, n\}$, мы видим, что упорядочение множителей произведения $a'_{1,\pi 1} \dots a'_{n,\pi n}$ в соответствии с перестановкой π^{-1} даёт

$$\begin{aligned} a'_{1,\pi 1} \dots a'_{n,\pi n} &= a'_{\pi^{-1}1,\pi(\pi^{-1}1)} \dots a'_{\pi^{-1}n,\pi(\pi^{-1}n)} = \\ &= a'_{\pi^{-1}1,1} \dots a'_{\pi^{-1}n,n} = a_{1,\pi^{-1}1} \dots a_{n,\pi^{-1}n}. \end{aligned}$$

Если учесть ещё, что $\varepsilon_\pi = \varepsilon_{\pi^{-1}}$ ($\varepsilon_\pi \varepsilon_{\pi^{-1}} = \varepsilon_{\pi\pi^{-1}} = \varepsilon_e = 1$), а $\{\pi^{-1} \mid \pi \in S_n\} = \{\pi \mid \pi \in S_n\} = S_n$ (поскольку $\pi \mapsto \pi^{-1}$ — биективное отображение из S_n в S_n), то по формуле (3) имеем

$$\begin{aligned} \det {}^t A &= \sum_{\pi \in S_n} \varepsilon_\pi a'_{1,\pi 1} \dots a'_{n,\pi n} = \sum_{\pi \in S_n} \varepsilon_{\pi^{-1}} a'_{1,\pi^{-1}1} \dots a'_{n,\pi^{-1}n} = \\ &= \sum_{\sigma \in S_n} \varepsilon_\sigma a_{1,\sigma 1} \dots a_{n,\sigma n} = \det A. \quad \square \end{aligned}$$

Замечание 3. Утверждение теоремы 1 интерпретируется так: если для определителей выполнено какое-то свойство относительно строк (столбцов), то оно имеет место и относительно столбцов (строк).

Теорема 2. Функция $\det: A \mapsto \det A$ на множестве $M_n(\mathbb{R})$ обладает следующими свойствами.

D1. $\det A$ — кососимметрическая функция строк матрицы A (т.е. при перестановке местами любых двух строк определитель меняет знак на противоположный).

D2. $\det A$ — полилинейная функция строк матрицы A (т.е. определитель матрицы A является линейной функцией элементов любой её строки $A_{(k)}$).

D3. $\det E = 1$.

Доказательство. D1. Пусть A' — матрица, получающаяся из A перестановкой строк $A_{(s)}, A_{(t)}$, т.е. $A'_{(s)} = A_{(t)}, A'_{(t)} = A_{(s)}, A'_{(i)} = A_{(i)}$ при $i \neq s, t$. Тогда, записав любую перестановку $\pi \in S_n$ в виде $\pi = \sigma\tau$ с транспозицией $\tau = (s, t)$ (см. в п. 3 § 8 гл. 1 выражение (10'), определяющее перестановку R_τ), будем иметь

$$\begin{aligned} \det A' &= \sum_{\pi \in S_n} \varepsilon_\pi a'_{1,\pi 1} \dots a'_{n,\pi n} = \\ &= \sum_{\sigma \in S_n} \varepsilon_{\sigma\tau} a'_{1,\sigma 1} \dots a'_{s,\sigma ts} \dots a'_{t,\sigma rt} \dots a'_{n,\sigma tn} = \\ &= \sum_{\sigma \in S_n} \varepsilon_{\sigma\tau} a'_{1,\sigma 1} \dots a'_{s,\sigma t} \dots a'_{t,\sigma s} \dots a'_{n,\sigma tn} = \end{aligned}$$

$$= \sum_{\sigma \in S_n} \varepsilon_{\sigma\tau} a_{1,\sigma 1} \dots a_{t,\sigma t} \dots a_{s,\sigma s} \dots a'_{n,\sigma n} = \\ = - \sum_{\sigma \in S_n} \varepsilon_{\sigma} a_{1,\sigma 1} \dots a_{n,\sigma n} = - \det A.$$

D2. Пусть $A = (a_{ij})$, и пусть $A_{(k)} = \lambda' A'_{(k)} + \lambda'' A''_{(k)}$, где штрихи указывают на вспомогательные матрицы

$$A' = [A_{(1)}, \dots, A_{(k-1)}, A'_{(k)}, A_{(k+1)}, \dots, A_{(n)}], \\ A'' = [A_{(1)}, \dots, A''_{(k)}, \dots, A_{(n)}].$$

По условию

$$a_{kj} = \lambda' a'_{kj} + \lambda'' a''_{kj}, \quad j = 1, 2, \dots, n.$$

Основываясь на замечании 1, свойство линейности $\det A$ относительно элементов k -й строки $A_{(k)}$ можно установить следующим образом. По определению

$$\det [A_{(1)}, \dots, A_{(k)}, \dots, A_{(n)}] = \det A =$$

$$= \sum_{\sigma \in S_n} \varepsilon_{\sigma} a_{1,\sigma 1} \dots a_{k,\sigma k} \dots a_{n,\sigma n} = \sum_{\sigma \in S_n} p_{\sigma} a_{k,\sigma k},$$

где p_{σ} , $\sigma \in S_n$, — коэффициенты, не зависящие от элементов строки $A_{(k)}$. Собирая подобные члены, отвечающие тем $\sigma \in S_n$, для которых $\sigma k = j$, и полагая $\alpha_j = \sum_{\sigma k=j} p_{\sigma}$, получим нужное свойство линейности

$$\det [\dots, A_{(k)}, \dots] = \sum_{j=1}^n \alpha_j a_{kj},$$

$$\det [\dots, \lambda' A'_{(k)} + \lambda'' A''_{(k)}, \dots] = \\ = \sum_{j=1}^n \alpha_j (\lambda' a_{kj} + \lambda'' a''_{kj}) = \lambda' \sum_{j=1}^n \alpha_j a'_{kj} + \lambda'' \sum_{j=1}^n \alpha_j a''_{kj} = \\ = \lambda' \det [\dots, A'_{(k)}, \dots] + \lambda'' \det [\dots, A''_{(k)}, \dots].$$

Короче:

$$\det A = \lambda' \det A' + \lambda'' \det A''.$$

D3. Очевидно, $\det E = \sum_{\sigma \in S_n} \varepsilon_{\sigma} \delta_{1,\sigma 1} \dots \delta_{n,\sigma n} = \varepsilon_{\sigma} \delta_{1,1} \dots \delta_{n,n} = 1$. \square

Из теоремы 2 вытекает несколько простых утверждений, которые мы сформулируем в виде свойств определителей, но доказывать их будем в более общей ситуации — для любой функции $\mathcal{D}: M_n(\mathbb{R}) \rightarrow \mathbb{R}$, обладающей свойствами D1–D2.

D4. Пусть $A \in M_n(\mathbb{R})$, $\lambda \in \mathbb{R}$. Тогда

$$\det \lambda A = \lambda^n \det A.$$

Действительно, в силу свойства D2, применённого последовательно к строкам с номерами 1, 2, ..., имеем

$$\begin{aligned} \mathcal{D}(\lambda A) &= \mathcal{D}[\lambda A_{(1)}, \lambda A_{(2)}, \dots, \lambda A_{(n)}] = \\ &= \lambda \mathcal{D}[A_{(1)}, \lambda A_{(2)}, \dots, \lambda A_{(n)}] = \lambda^2 \mathcal{D}[A_{(1)}, A_{(2)}, \dots, \lambda A_{(n)}] = \dots \\ &\dots = \lambda^n \mathcal{D}[A_{(1)}, A_{(2)}, \dots, A_{(n)}] = \lambda^n \mathcal{D}(A). \quad \square \end{aligned}$$

D5. Определитель с нулевой строкой равен нулю.

Пусть, например, $A_{(k)} = (0, 0, \dots, 0)$. Тогда и $2A_{(k)} = (0, 0, \dots, 0)$. Следовательно, по D2

$$\begin{aligned} \mathcal{D}(A) &= \mathcal{D}[A_{(1)}, \dots, A_{(k)}, \dots, A_{(n)}] = \mathcal{D}[A_{(1)}, \dots, 2A_{(k)}, \dots, A_{(n)}] = \\ &= 2\mathcal{D}[A_{(1)}, \dots, A_{(k)}, \dots, A_{(n)}] = 2\mathcal{D}(A), \end{aligned}$$

откуда $\mathcal{D}(A) = 0$. \square

D6. Если в квадратной матрице A две строки совпадают, то её определитель равен нулю.

Берём опять произвольную функцию \mathcal{D} со свойствами D1–D2. Поменяв местами две совпадающие строки $A_{(s)}, A_{(t)}$ в A , мы получим ту же матрицу A . С другой стороны, согласно свойству D1 для \mathcal{D} значение $\mathcal{D}(A)$ примет противоположный знак. Таким образом, $\mathcal{D}(A) = -\mathcal{D}(A)$, откуда $2\mathcal{D}(A) = 0$ и $\mathcal{D}(A) = 0$. \square

D7. Определитель не меняется, если над его строками совершаются элементарные преобразования типа (II).

Достаточно рассмотреть случай применения одного элементарного преобразования. Пусть после прибавления к s -й строке матрицы A её t -й строки, умноженной на λ , получилась матрица A' . Тогда в соответствии со свойствами D1 и D6 для \mathcal{D} имеем

$$\begin{aligned} \mathcal{D}(A') &= \mathcal{D}[A_{(1)}, \dots, A_{(s)} + \lambda A_{(t)}, \dots, A_{(t)}, \dots] = \\ &= \mathcal{D}[A_{(1)}, \dots, A_{(s)}, \dots, A_{(t)}, \dots] + \lambda \mathcal{D}[A_{(1)}, \dots, A_{(t)}, \dots, A_{(t)}, \dots] = \\ &= \mathcal{D}[A_{(1)}, \dots, A_{(s)}, \dots, A_{(t)}, \dots] = \mathcal{D}(A). \quad \square \end{aligned}$$

Замечание 4. Проведённые доказательства показывают, что любая функция $\mathcal{D}: M_n(\mathbb{R}) \rightarrow \mathbb{R}$ со свойствами D1–D2 обладает также свойствами D4–D7 (заменить символ \det на \mathcal{D}).

Предложение 1. Пусть

$$\bar{A} = \left| \begin{array}{cccc} \bar{a}_{11} & \bar{a}_{12} & \dots & \bar{a}_{1n} \\ 0 & \bar{a}_{22} & \dots & \bar{a}_{2n} \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \bar{a}_{nn} \end{array} \right| \quad (5)$$

— верхняя треугольная матрица порядка n , E — единичная матрица и $D: M_n(\mathbb{R}) \rightarrow \mathbb{R}$ — любая функция, обладающая свойствами D1–D2. Тогда

$$D(\bar{A}) = D(E) \bar{a}_{11} \bar{a}_{22} \dots \bar{a}_{nn}.$$

Доказательство. Согласно замечанию 4 мы можем опираться на свойства D2, D7. На основании D2 вынесем \bar{a}_{nn} за знак D :

$$D(\bar{a}) = \bar{a}_{nn} D\left(\begin{array}{cccc} \bar{a}_{11} & \dots & \bar{a}_{1,n-1} & \bar{a}_{1n} \\ \dots & \dots & \dots & \dots \\ 0 & \dots & \bar{a}_{n-1,n-1} & \bar{a}_{n-1,n} \\ 0 & \dots & 0 & 1 \end{array}\right).$$

Применим теперь к \bar{A} элементарное преобразование типа (II): вычтем из i -й строки стоящей под знаком D матрицы последнюю строку, умноженную предварительно на \bar{a}_{in} . При этом элементы последнего столбца обратятся в нуль (кроме $\bar{a}_{nn} = 1$), а все другие элементы матрицы останутся без изменения. Применим то же самое рассуждение к предпоследней строке вновь полученной матрицы и т.д. Каждый раз очередной элемент \bar{a}_{ii} выносится за знак D и рассуждение возобновляется. Проделав его n раз, мы убеждаемся в том, что

$$D(\bar{A}) = \bar{a}_{nn} \dots \bar{a}_{11} \cdot D\left(\begin{array}{ccc} 1 & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & 1 \end{array}\right),$$

а это и есть искомая формула. \square

Следствие. Если \bar{A} — матрица вида (5), то

$$\det \bar{A} = \bar{a}_{11} \bar{a}_{22} \dots \bar{a}_{nn}. \quad (6)$$

Доказательство непосредственно вытекает из предложения 1, если заметить, что $\det E = 1$ (свойство D3). \square

Полезно привести ещё один вывод формулы (6), опирающийся на более общее утверждение, которым мы воспользуемся позднее. Предварительно дадим следующее

Определение. Определитель матрицы, получающейся из $A = (a_{ij})$ вычёркиванием i -й строки и j -го столбца, обозначается M_{ij} и называется **минором** матрицы A , соответствующим элементу a_{ij} . Величина $A_{ij} = (-1)^{i+j} M_{ij}$ называется **алгебраическим дополнением** элемента a_{ij} .

Предложение 2. Если

$$A = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ 0 & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ 0 & a_{n2} & \dots & a_{nn} \end{vmatrix},$$

то

$$\det A = a_{11} M_{11} = a_{11} A_{11}.$$

Доказательство. Так как $\det A = \det {}^t A$ (теорема 1) и так как a_{11} — единственный отличный от нуля элемент первого столбца $A^{(1)}$, то $a_{\pi 1,1} = 0$ при $\pi 1 \neq 1$ и

$$\det A = \sum_{\pi \in S_n} \varepsilon_\pi a_{\pi 1,1} a_{\pi 2,2} \dots a_{\pi n,n} = \sum_{\pi \in S_n, \pi 1=1} \varepsilon_\pi a_{1,1} a_{\pi 2,2} \dots a_{\pi n,n}.$$

Совокупность всех перестановок $\pi \in S_n$, оставляющих на месте символ 1, отождествляется с множеством S_{n-1} перестановок, действующих на множестве $\{2, 3, \dots, n\}$. Таким образом,

$$\begin{aligned} \det A &= a_{11} \sum_{\sigma \in S_{n-1}} \varepsilon_\sigma a_{\sigma 2,2} \dots a_{\sigma n,n} = \\ &= a_{11} \begin{vmatrix} a_{22} & \dots & a_{2n} \\ \dots & \ddots & \dots \\ a_{n2} & \dots & a_{nn} \end{vmatrix} = a_{11} M_{11}. \quad \square \end{aligned}$$

В применении к верхней треугольной матрице \bar{A} предложение 2 даёт равенство $\det \bar{A} = \bar{a}_{11} \bar{M}_{11}$, где

$$\bar{M}_{11} = \begin{vmatrix} \bar{a}_{22} & & * \\ & \ddots & \\ 0 & & \bar{a}_{nn} \end{vmatrix}$$

— определитель того же вида, но на единицу меньшего порядка. Очевидное рассуждение по индукции приводит к формуле (6).

Доказанные свойства дают возможность сравнительно просто вычислить определитель порядка n . Один из методов заключается в следующем. Матрицу $A = (a_{ij})$ следует привести элементарными преобразованиями к треугольному виду (см. § 3 гл. 1). Пусть мы получим матрицу \bar{A} вида (5). Предположим, что в процессе приведения было совершено q элементарных преобразований типа (I) и какое-то количество преобразований типа (II). Так как последние не меняют определителя (свойство D7), а каждое преобразование типа (I) умножает его на -1 , то $\det \bar{A} = (-1)^q \det A$. По формуле (6) мы имеем

$$\det \bar{A} = \bar{a}_{11} \bar{a}_{22} \dots \bar{a}_{nn}.$$

В таком случае

$$\det A = (-1)^q \bar{a}_{11} \bar{a}_{22} \dots \bar{a}_{nn}. \quad (7)$$

Это и есть одна из формул для вычисления $\det A$.

Теперь, опираясь на формулу (7), мы установим важный факт, касающийся роли свойств D1–D3 определителя. Именно, имеет место

Теорема 3. Пусть $\mathcal{D}: M_n(\mathbb{R}) \rightarrow \mathbb{R}$ — функция, обладающая следующими свойствами:

i) при перестановке местами любых двух соседних строк матрицы $A \in M_n(\mathbb{R})$ значение $\mathcal{D}(A)$ меняет знак на противоположный;

ii) $\mathcal{D}(A)$ является линейной функцией элементов каждой строки матрицы A (другими словами, $\mathcal{D}(A)$ — кососимметрическая полилинейная функция строк матрицы).

Тогда

$$\mathcal{D}(A) = \mathcal{D}(E) \cdot \det A.$$

Доказательство. Как мы знаем, свойство i) эквивалентно тому, что $\mathcal{D}(A)$ меняет знак на противоположный при перестановке любых двух строк, т.е. при любом элементарном преобразовании типа (I). Далее, согласно замечанию 4 $\mathcal{D}(A)$ обладает также свойствами D4–D7. В частности, $\mathcal{D}(A)$ не меняется, если строки матрицы A подвергнуть элементарному преобразованию типа (II).

Приведём матрицу A при помощи элементарных преобразований к треугольному виду (5), где, конечно, некоторые из \bar{A}_{ii} могут равняться нулю. С учётом вышесказанного мы имеем формулы (см. (7))

$$\det A = (-1)^q \det \bar{A} = (-1)^q \bar{a}_{11} \bar{a}_{22} \dots \bar{a}_{nn},$$

$$\mathcal{D}(A) = (-1)^q \mathcal{D}(\bar{A}),$$

где q — число элементарных преобразований типа (I), совершенных при переходе от A к \bar{A} . Нужное нам соотношение $\mathcal{D}(A) = \mathcal{D}(E) \det A$ вытекает теперь непосредственно из предложения 1. \square

Итак, свойствами D1–D3 функция \det характеризуется однозначно. По этой причине мы отнесли их к основным свойствам определителей. Можно было с самого начала назвать определителем функцию \mathcal{D} , обладающую свойствами D1–D3, но в таком случае нужно установить её существование. У нас существование обеспечивается самой конструкцией функции \det — формулой (3).

Имея в виду дальнейшие применения теоремы 3, мы не включили в её формулировку нормировочное условие $\mathcal{D}(E) = 1$.

УПРАЖНЕНИЯ

1. Кососимметрическую функцию $\Delta: \mathbb{R}^3 \rightarrow \mathbb{R}$ трёх переменных

$$\Delta(x, y, z) = (y - x)(z - x)(z - y)$$

записать в виде определителя третьего порядка.

2. Пусть $A = (a_{ij})$, $A' = (a'_{ij})$ — две $n \times n$ -матрицы, Δ , Δ' — их определители. Сравнить Δ и Δ' в случаях:

a) $a'_{ij} = 2^{i-j} a_{ij};$

b) $a'_{ij} = a_{n+1-i, j};$

c) $a'_{ij} = a_{n+1-i, n+1-j}.$

3. Показать, что

$$\begin{vmatrix} 1 & 1 & 1 & \dots & 1 & 1 \\ 1 & 2 & 1 & \dots & 1 & 1 \\ 1 & 1 & 3 & \dots & 1 & 1 \\ 1 & 1 & 1 & \dots & n & 1 \\ 1 & 1 & 1 & \dots & 1 & n+1 \end{vmatrix} = n!.$$

§ 2. Дальнейшие свойства определителей

1. Разложение определителя по элементам столбца или строки. Существует регулярный способ вычисления определителей, основанный на редукции к определителям меньшего порядка. При этом используются понятия минора M_{ij} и алгебраического дополнения A_{ij} (см. определение в § 1).

Теорема 1. Пусть $A = (a_{ij}) \in M_n(\mathbb{R})$. Справедливы следующие формулы:

$$\det A = \sum_{i=1}^n (-1)^{i+j} a_{ij} M_{ij} = \sum_{i=1}^n a_{ij} A_{ij} \quad (1)$$

(разложение определителя по элементам j -го столбца);

$$\det A = \sum_{j=1}^n (-1)^{i+j} a_{ij} M_{ij} = \sum_{j=1}^n a_{ij} A_{ij} \quad (2)$$

(разложение определителя по элементам i -й строки).

Иначе говоря, определитель матрицы A равен сумме произведений всех элементов некоторого столбца (некоторой строки) на их алгебраические дополнения.

Доказательство. 1) Опираясь на основные свойства D1 и D2 определителей (сначала относительно столбцов, а затем относительно строк), выпишем цепочку равенств:

$$\begin{aligned} \det A &= \begin{vmatrix} a_{11} & \dots & a_{1j} & \dots & a_{1n} \\ a_{21} & \dots & a_{2j} & \dots & a_{2n} \\ \dots & & \dots & & \dots \\ a_{n1} & \dots & a_{nj} & \dots & a_{nn} \end{vmatrix} = \\ &= \begin{vmatrix} a_{11} & \dots & a_{1j} & \dots & a_{1n} \\ a_{21} & \dots & 0 & \dots & a_{2n} \\ \dots & & \dots & & \dots \\ a_{n1} & \dots & 0 & \dots & a_{nn} \end{vmatrix} + \begin{vmatrix} a_{11} & \dots & 0 & \dots & a_{1n} \\ a_{21} & \dots & a_{2j} & \dots & a_{2n} \\ \dots & & \dots & & \dots \\ a_{n1} & \dots & 0 & \dots & a_{nn} \end{vmatrix} + \dots \end{aligned}$$

$$\begin{aligned}
 & \dots + \left| \begin{array}{cccccc} a_{11} & \dots & 0 & \dots & a_{1n} \\ a_{21} & \dots & 0 & \dots & a_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ a_{n1} & \dots & a_{nj} & \dots & a_{nn} \end{array} \right| = \\
 & = \sum_{i=1}^n \left| \begin{array}{ccccc} a_{11} & \dots & a_{1,j-1} & 0 & a_{1,j+1} \dots a_{1n} \\ \dots & \dots & \dots & \dots & \dots \\ a_{i1} & \dots & a_{i,j-1} & a_{ij} & a_{i,j+1} \dots a_{in} \\ \dots & \dots & \dots & \dots & \dots \\ a_{n1} & \dots & a_{n,j-1} & 0 & a_{n,j+1} \dots a_{nn} \end{array} \right| = \\
 & = \sum_{i=1}^n (-1)^{j-1} \left| \begin{array}{ccccc} 0 & a_{11} & \dots & a_{1,j-1} & a_{1,j+1} \dots a_{1n} \\ a_{ij} & a_{i1} & \dots & a_{i,j-1} & a_{i,j+1} \dots a_{in} \\ \dots & \dots & \dots & \dots & \dots \\ 0 & a_{n1} & \dots & a_{n,j-1} & a_{n,j+1} \dots a_{nn} \end{array} \right| = \\
 & = \sum_{i=1}^n (-1)^{(j-1)+(i-1)} \times \\
 & \quad \left| \begin{array}{cccccc} a_{ij} & a_{i1} & \dots & a_{i,j-1} & a_{i,j+1} & \dots & a_{in} \\ 0 & a_{i1} & \dots & a_{1,j-1} & a_{1,j+1} & \dots & a_{1n} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & a_{i-1,1} & \dots & a_{i-1,j-1} & a_{i-1,j+1} & \dots & a_{i-1,n} \\ 0 & a_{i+1,1} & \dots & a_{i+1,j-1} & a_{i+1,j+1} & \dots & a_{i+1,n} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & a_{n1} & \dots & a_{n,j-1} & a_{n,j+1} & \dots & a_{nn} \end{array} \right| = \\
 & \quad = \sum_{i=1}^n (-1)^{i+j} a_{ij} M_{ij}.
 \end{aligned}$$

Последнее равенство основано на предложении 2 § 1, применённом к матрице

$$A' = \left\| \begin{array}{cccc} a'_{11} & a'_{12} & \dots & a'_{1n} \\ 0 & a'_{22} & \dots & a'_{2n} \\ \dots & \dots & \dots & \dots \\ 0 & a'_{n2} & \dots & a'_{nn} \end{array} \right\|$$

с $a'_{11} = a_{ij}$, $a'_{12} = a_{i1}$, ..., $a'_{1n} = a_{in}$, $M'_{11} = M_{ij}$. Остаётся вспомнить, что по определению $A_{ij} = (-1)^{i+j} M_{ij}$. Формула (1) доказана.

2) Положим ${}^t A = (a'_{ji})$, $a'_{ji} = a_{ij}$. Заметим ещё, что минором, соответствующим элементу a'_{ji} в $\det {}^t A$, будет $M'_{ji} = M_{ij}$. Как было показано в 1), $\det A = \det {}^t A = \sum_{j=1}^n (-1)^{j+i} a'_{ji} M'_{ji} = \sum_{j=1}^n (-1)^{i+j} a_{ij} M_{ij}$,

т.е. мы пришли к формуле (2). Можно было рассуждать проще, сказавшись сразу на замечание 3 из § 1. \square

Следующие два примера служат иллюстрацией полученных нами свойств определителей.

Пример 1. Определитель

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

связываемый с именем Вандермонда, вычисляется по формуле

$$\Delta_n = \prod_{1 \leq i < j \leq n} (x_j - x_i),$$

или, в более подробной записи,

$$\Delta_n = (x_2 - x_1)(x_3 - x_1) \dots (x_n - x_1)(x_3 - x_2) \dots (x_n - x_2) \dots (x_n - x_{n-1})$$

(в связи с этой формулой полезно вернуться к упр. 1 из § 1). В частности, при попарно различных элементах x_1, \dots, x_n определитель Вандермонда отличен от нуля. Этим его свойством часто пользуются. По теореме 1 § 1 имеем также

$$\Delta_n = \begin{vmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \dots & x_2^{n-1} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_n & x_n^2 & \dots & x_n^{n-1} \end{vmatrix}.$$

Для доказательства формулы (3) применим индукцию по n . Считая, что $\Delta_m, m < n$, вычисляется по формуле (3) и опиралась на свойство D7, вычтем из каждой i -й строки определителя Δ_n ($i-1$)-ю строку, умноженную на x_1 :

$$\Delta_n = \begin{vmatrix} 1 & 1 & \dots & 1 \\ 0 & x_2 - x_1 & \dots & x_n - x_1 \\ 0 & x_2^2 - x_2 x_1 & \dots & x_n^2 - x_n x_1 \\ \dots & \dots & \dots & \dots \\ 0 & x_2^{n-1} - x_2^{n-2} x_1 & \dots & x_n^{n-1} - x_n^{n-2} x_1 \end{vmatrix}.$$

Напрашивается мысль разложить теперь Δ_n по первому столбцу, а в получившемся определителе порядка $n-1$ вынести из j -го столбца ($j = 1, 2, \dots, n-1$) за знак определителя общий множитель $x_{j+1} - x_1$ (свойство D1 для столбцов). Мы придём к выражению

$$\begin{aligned} \Delta_n &= (x_n - x_1)(x_{n-1} - x_1) \dots (x_2 - x_1) \begin{vmatrix} 1 & 1 & \dots & 1 \\ x_2 & x_3 & \dots & x_n \\ \dots & \dots & \dots & \dots \\ x_2^{n-2} & x_3^{n-2} & \dots & x_n^{n-2} \end{vmatrix} = \\ &= (x_n - x_1)(x_{n-1} - x_1) \dots (x_2 - x_1) \cdot \Delta(x_2, x_3, \dots, x_n), \end{aligned}$$

совпадающему с (3), поскольку по предположению индукции

$$\Delta(x_2, \dots, x_n) = \prod_{2 \leq i < j \leq n} (x_j - x_i).$$

Пример 2. Матрица $A = (a_{ij})$ вида

$$A = \begin{vmatrix} 0 & a_{12} & a_{13} & \dots & a_{1n} \\ -a_{12} & 0 & a_{23} & \dots & a_{2n} \\ -a_{13} & -a_{23} & 0 & \dots & a_{3n} \\ \dots & \dots & \dots & \dots & \dots \\ -a_{1n} & -a_{2n} & -a_{3n} & \dots & 0 \end{vmatrix}$$

называется кососимметрической (о её определителе тоже говорят, что он кососимметрический). Другими словами, ${}^t A = -A$. С учётом теоремы 1 из § 1 имеем

$$\det A = \det {}^t A = \det(-A) = (-1)^n \det A,$$

откуда $[1 + (-1)^{n-1}] \det A = 0$. При нечётном n получаем $\det A = 0$, т.е. определитель любой кососимметрической матрицы нечётного порядка равен нулю.

2. Определители специальных матриц. Чем больше нулей среди элементов матрицы A и "чем лучше" они расположены, тем легче вычислять определитель $\det A$. Это интуитивное представление находит в некоторых случаях точное количественное выражение. Например, мы знаем (см. (6) из § 1), что определитель треугольной матрицы (верхней или нижней) равен произведению элементов, стоящих на главной диагонали. Другой важный частный случай содержит

Теорема 2. Для определителя D порядка $n+m$, у которого на пересечении первых n столбцов и последних m строк, стоят нули, имеет место формула

$$\begin{vmatrix} a_{11} & \dots & a_{1n} & a_{1,n+1} & \dots & a_{1,n+m} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ a_{n1} & \dots & a_{nn} & a_{n,n+1} & \dots & a_{n,n+m} \\ 0 & \dots & 0 & b_{11} & \dots & b_{1m} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & b_{m1} & \dots & b_{mm} \end{vmatrix} = \\ = \begin{vmatrix} a_{11} & \dots & a_{1n} \\ \dots & \dots & \dots \\ a_{n1} & \dots & a_{nn} \end{vmatrix} \cdot \begin{vmatrix} b_{11} & \dots & b_{1m} \\ \dots & \dots & \dots \\ b_{m1} & \dots & b_{mm} \end{vmatrix}$$

(определитель в левой части этого равенства называется квазитреугольным или определителем с углом нулей).

Доказательство. Зафиксируем сначала $n(n+m)$ элементов a_{ij} и рассмотрим определитель D как функцию элементов b_{kl} , которые образуют квадратную матрицу B порядка m . На полученную функцию можно смотреть как на функцию матрицы B : $D = D(B)$.

Ясно, что полилинейность и кососимметричность определителя D относительно последних m строк эквивалентна тем же свойствам $D(B)$ относительно строк матрицы B . Значит, правомерно применить к $D(B)$ теорему 3 § 1, согласно которой $D(B) = D(E) \det B$. По

определению функции \mathcal{D} имеем

$$\mathcal{D}(E) = \begin{vmatrix} a_{11} & \dots & a_{1n} & a_{1,n+1} & \dots & a_{1,n+m} \\ \dots & & \dots & \dots & & \dots \\ a_{n1} & \dots & a_{nn} & a_{n,n+1} & \dots & a_{n,n+m} \\ 0 & \dots & 0 & 1 & \dots & 0 \\ \dots & & \dots & \dots & & \dots \\ 0 & \dots & 0 & 0 & \dots & 1 \end{vmatrix}.$$

Разложим $\mathcal{D}(E)$ по последней строке (см. формулу (2)), затем по предпоследней и т.д. Повторив эту операцию m раз, мы убедимся в том, что $\mathcal{D}(E) = \det A$, где

$$A = \begin{vmatrix} a_{11} & \dots & a_{1n} \\ \dots & & \dots \\ a_{n1} & \dots & a_{nn} \end{vmatrix}.$$

Окончательно получаем $D = \mathcal{D}(B) = \det A \cdot \det B$. \square

В новых обозначениях формула из теоремы 2 принимает более компактный вид

$$\det \begin{vmatrix} A & C \\ 0 & B \end{vmatrix} = \det A \cdot \det B. \quad (4)$$

Здесь A и B — квадратные матрицы, а нулевая матрица 0 и матрица C прямоугольные. Опираясь на теорему 1 из § 1 и теорему 2 или на рассуждения, использованные в ходе доказательства теоремы 2, мы без труда устанавливаем, что

$$\det \begin{vmatrix} A & 0 \\ C & B \end{vmatrix} = \det A \cdot \det B. \quad \square$$

Иногда пытаются написать в точности такое же выражение для определителя $\det \begin{vmatrix} C & A \\ B & 0 \end{vmatrix}$, хотя сразу же напрашивается простейший контрпример $\begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} = -1$. Всё дело в знаке. Правильный результат получается путём перестановки строк или столбцов, приводящей матрицу $\begin{vmatrix} C & A \\ B & 0 \end{vmatrix}$ к виду $\begin{vmatrix} B & 0 \\ C & A \end{vmatrix}$ или $\begin{vmatrix} A & C \\ 0 & B \end{vmatrix}$.

Более простые рассуждения основаны на той же теореме 3 из § 1, которую мы неоднократно использовали. Действительно,

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

Далее по формуле (2), применённой m раз, находим

$$\det \begin{vmatrix} C & A \\ E_m & 0 \end{vmatrix} = \begin{vmatrix} * & a_{11} & \dots & a_{1n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & \dots & a_{nn} \\ 1 & \dots & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 1 & 0 & \dots & 0 \end{vmatrix} = (-1)^{(n+2)+(n+4)+\dots+(n+2m)} \det A = (-1)^{nm} \det A.$$

Окончательно приходим к выводу, что если A, B — квадратные матрицы порядков n и m соответственно, то

$$\det \begin{vmatrix} C & A \\ B & 0 \end{vmatrix} = (-1)^{nm} \det A \cdot \det B. \quad (5)$$

Формулы (4) и (5) охватываются общей теоремой Лапласа о разложении определителей. Эта теорема, однако, употребляется сравнительно редко, и мы на ней не останавливаемся, отсылая любознательного читателя к упражнениям в конце следующего параграфа.

Исключительно важное в теоретическом плане утверждение об определителях матриц содержит

Теорема 3. Пусть A и B — квадратные матрицы порядка n . Тогда

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

Доказательство. Согласно формулам (7) и (9) § 3 гл. 2, выражающим коэффициенты c_{ij} матрицы $(c_{ij}) = AB = (a_{ij})(b_{ij})$ через коэффициенты матриц A и B , i -я строка $(AB)_{(i)}$ записывается в виде

$$(AB)_{(i)} = (A_{(i)}B^{(1)}, A_{(i)}B^{(2)}, \dots, A_{(i)}B^{(n)}),$$

$$A_{(i)}B^{(j)} = \sum_{k=1}^n a_{ik} b_{kj}.$$

Фиксируем матрицу B и для любой матрицы A положим

$$\mathcal{D}_B(A) = \det AB.$$

Докажем, что функция $\mathcal{D} = \mathcal{D}_B$ удовлетворяет условиям i), ii) теоремы 3 из § 1. В самом деле, поменяем $A_{(s)}$ и $A_{(t)}$ местами. Так как s -я и t -я строки матрицы AB имеют вид

$$(A_{(s)}B^{(1)}, \dots, A_{(s)}B^{(n)}),$$

$$(A_{(t)}B^{(1)}, \dots, A_{(t)}B^{(n)}),$$

то при этом они тоже поменяются местами и, значит, то теореме 1

$$\begin{aligned} \mathcal{D}(\dots, A_{(s)}, \dots, A_{(t)}, \dots) &= \mathcal{D}(A) = \\ &= \det AB = \det[\dots, (AB)_{(s)}, \dots, (AB)_{(t)}, \dots] = \\ &= -\det[\dots, (AB)_{(t)}, \dots, (AB)_{(s)}, \dots] = -\mathcal{D}(\dots, A_{(t)}, \dots, A_{(s)}, \dots). \end{aligned}$$

Далее, как известно, $\det AB$ — линейная функция элементов i -й строки $(AB)_{(i)}$:

$$\det AB = \lambda_1 A_{(i)} B^{(1)} + \lambda_2 A_{(i)} B^{(2)} + \dots + \lambda_n A_{(i)} B^{(n)}.$$

Поэтому

$$\mathcal{D}(A) = \sum_{j=1}^n \lambda_j \sum_{k=1}^n a_{ik} b_{kj} = \sum_{k=1}^n a_{ik} \sum_{j=1}^n \lambda_j b_{kj} = \sum_{k=1}^n \mu_k a_{ik},$$

где $\mu_k = \sum_{j=1}^n \lambda_j b_{kj}$ — скаляр, не зависящий от элементов i -й строки $A_{(i)}$ матрицы A .

Мы видим, что \mathcal{D} линейно зависит от элементов i -й строки матрицы A .

Таким образом, выполнены оба условия теоремы 3 § 1, согласно которой $\mathcal{D}(A) = \mathcal{D}(E) \cdot \det A$. Но по определению $\mathcal{D}(E) = \det EB = \det B$. Отсюда вытекает искомая формула. \square

Непосредственная проверка теоремы 3, сравнительно легко выполняемая при $n = 2$, уже при $n = 3$ сопряжена со значительными трудностями. Однако и в общем случае можно указать обходной маневр, основанный непосредственно на свойствах D1-D2, а также на привлечении теорем 1 и 2 (см. упр. 3).

УПРАЖНЕНИЯ

1. Целые числа 1798, 2139, 3255, 4867 делятся на 31. Без всяких вычислений показать, что определитель четвёртого порядка

$$\left| \begin{array}{rrrr} 1 & 7 & 9 & 8 \\ 2 & 1 & 3 & 9 \\ 3 & 2 & 5 & 5 \\ 4 & 8 & 6 & 7 \end{array} \right|$$

также делится на 31.

2. Показать, что любой кососимметрический определитель $|a_{ij}|$ четвертого порядка с $a_{ij} \in \mathbb{Z}$ является квадратом целого числа.

Замечание. Это верно для кососимметрического определителя произвольного порядка.

3. Доказать соотношение $\det AB = \det A \cdot \det B$ (теорема 3) путём приведения элементарными преобразованиями типа (II) над строками вспомогательной матрицы $C = \begin{vmatrix} E & B \\ -A & 0 \end{vmatrix}$ размера $2n \times 2n$ к виду $C' = \begin{vmatrix} E & B \\ 0 & AB \end{vmatrix}$.

Указание. Воспользоваться равенством $\det C = \det C'$ и соотношениями (4), (5).

То же доказательство провести, основываясь на упр. 17, 18 из § 3 гл. 2 и на замечании, что $\begin{vmatrix} E & A \\ 0 & E \end{vmatrix}$ — верхняя унитреугольная матрица.

4 (Захаров В.И. — Тула, 1984). В задачах по моделированию случайных стационарных процессов возникают определители вида

$$\Delta_n(k_1, x_1; \dots; k_m, x_m) = \begin{vmatrix} M_{k_1}^n(x_1) \\ M_{k_2}^n(x_2) \\ \dots \\ M_{k_m}^n(x_m) \end{vmatrix},$$

где x_1, \dots, x_m — любые переменные; k_1, \dots, k_m — натуральные числа, $k_1 + k_2 + \dots + k_m = n$; $M_k^n(x)$ — $k \times n$ -матрица вида

$$M_k^n(x) = \begin{vmatrix} 1 & x & x^2 & \dots & x^{n-1} \\ 0 & 1 & \binom{x}{2} & \dots & \binom{n-1}{1} x^{n-2} \\ 0 & 0 & 1 & \dots & \binom{n-1}{2} x^{n-3} \\ \dots & & & & \\ 0 & 0 & 0 & \dots & \binom{n-1}{k-1} x^{n-k} \end{vmatrix}.$$

Доказать, что

$$\Delta_n(k_1, x_1; \dots; k_m, x_m) = \prod_{1 \leq j < i \leq m} (x_i - x_j)^{k_i k_j}.$$

Указание. При $k_1 = \dots = k_m = 1$, т.е. при $m = n$, получается определитель Вандермонда.

5. Показать, что

$$B_n(s, t) = \begin{vmatrix} \binom{s}{t} & \binom{s}{t+1} & \dots & \binom{s}{t+n-1} \\ \binom{s+1}{t} & \binom{s+1}{t+1} & \dots & \binom{s+1}{t+n-1} \\ \dots & & & \\ \binom{s+n-1}{t} & \binom{s+n-1}{t+1} & \dots & \binom{s+n-1}{t+n-1} \end{vmatrix} =$$

$$= \frac{\binom{n+s-1}{n} \binom{n+s-2}{n} \dots \binom{n+s-t}{n}}{\binom{n+t-1}{n} \binom{n+t-2}{n} \dots \binom{n}{n}}.$$

Указание. Вынести последовательно $s+k-1$ из k -й строки при $k = 1, 2, \dots, n$, а затем $1/(t+l-1)$ из l -го столбца при $l = 1, 2, \dots, n$. Действовать так до тех пор, пока в первом столбце не будут стоять только 1.

6. Пусть

$$C_n(\lambda_1, \dots, \lambda_n) = \begin{vmatrix} \lambda_1 & 1 & 0 & \dots & 0 & 0 & 0 \\ -1 & \lambda_2 & 1 & \dots & 0 & 0 & 0 \\ \dots & & & & & & \\ 0 & 0 & 0 & \dots & \lambda_{n-2} & 1 & 0 \\ 0 & 0 & 0 & \dots & -1 & \lambda_{n-1} & 1 \\ 0 & 0 & 0 & \dots & 0 & -1 & \lambda_n \end{vmatrix}.$$

Показать, что $\det C_n = \lambda_n \det C_{n-1} + \det C_{n-2}$. При $\lambda_1 = \lambda_2 = \dots = \lambda_n = 1$ найти численное значение $\det C_n$.

Указание. Вспомнить пример 3 из п. 3 § 3 гл. 2 и обратить внимание на тот факт, что $\det C_n(1, \dots, 1) = (-1)^n \det C_n(-1, \dots, -1)$.

7. Показать, что определитель $n \times n$ -матрицы

$$A_n = \begin{vmatrix} 2 & -1 & 0 & 0 & \dots & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & \dots & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & \dots & 0 & 0 & 0 \\ \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & \dots & 0 & -1 & 2 \end{vmatrix}$$

равен $n + 1$.

8. Пусть A, B — любые квадратные матрицы порядка n . Показать, что

$$\det \begin{vmatrix} A & B \\ B & A \end{vmatrix} = \det(A + B) \cdot \det(A - B).$$

9. Пусть X — матрица размера $n \times k$, а Y — размера $k \times n$. Доказать, что

$$\det(E_n + XY) = \det(E_k + YX).$$

Указание. Использовать соотношение

$$\begin{vmatrix} E_k + YX & 0 \\ X & E_n \end{vmatrix} \begin{vmatrix} E_k & Y \\ 0 & E_n \end{vmatrix} = \begin{vmatrix} E_k & Y \\ 0 & E_n \end{vmatrix} \begin{vmatrix} E_k & 0 \\ X & E_n + XY \end{vmatrix}.$$

§ 3. Применения определителей

1. Критерий невырожденности матрицы. По теореме 5 из § 3 гл. 2 условие невырожденности матрицы $A \in M_n(\mathbb{R})$ (т.е. равенство $\text{rank } A = n$) эквивалентно её обратимости. Применяя теорему 3 из § 2 к соотношению $AA^{-1} = A^{-1}A = E$, мы получаем, что

$$\det A \cdot \det(A^{-1}) = 1.$$

Стало быть, определитель невырожденной матрицы отличен от нуля и

$$\det(A^{-1}) = (\det A)^{-1}.$$

Наряду с матрицей A рассмотрим её присоединённую (или взаимную) матрицу

$$A^\vee = \begin{vmatrix} A_{11} & \dots & A_{n1} \\ \dots & \dots & \dots \\ A_{1n} & \dots & A_{nn} \end{vmatrix}.$$

Чтобы получить A^\vee , надо поставить на место каждого элемента a_{ij} матрицы A его алгебраическое дополнение A_{ij} ($i, j = 1, \dots, n$), а затем перейти к транспонированной матрице.

Теорема 1. Матрица $A \in M_n(\mathbb{R})$ невырождена (обратима) тогда и только тогда, когда $\det A \neq 0$. Если $\det A \neq 0$, то

$$A^{-1} = (\det A)^{-1} A^{\vee},$$

или, в более подробной записи,

$$\begin{vmatrix} a_{11} & \dots & a_{n1} \\ \dots & \dots & \dots \\ a_{1n} & \dots & a_{nn} \end{vmatrix}^{-1} = \begin{vmatrix} \frac{A_{11}}{\det A} & \dots & \frac{A_{n1}}{\det A} \\ \dots & \dots & \dots \\ \frac{A_{1n}}{\det A} & \dots & \frac{A_{nn}}{\det A} \end{vmatrix}.$$

Доказательству теоремы предпошлём лемму.

Лемма. Пусть $A \in M_n(\mathbb{R})$. Имеют место соотношения

$$a_{i1}A_{j1} + a_{i2}A_{j2} + \dots + a_{in}A_{jn} = \delta_{ij} \det A, \quad (1)$$

$$a_{1i}A_{1j} + a_{2i}A_{2j} + \dots + a_{ni}A_{nj} = \delta_{ij} \det A, \quad (2)$$

где δ_{ij} — символ Кронекера (при $i \neq j$ говорят о разложении определителя $\det A$ по чужой строке или соответственно по чужому столбцу).

Доказательство. При $i = j$ утверждение леммы совпадает с теоремой 1 из § 2. Поэтому остаётся рассмотреть случай $i \neq j$, когда $\delta_{ij} = 0$. С этой целью введём матрицу

$$A' = [A_{(1)}, \dots, A_{(i)}, \dots, A_{(i)}, \dots, A_{(n)}] = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \dots & \dots & \dots & \dots \\ a_{i1} & a_{i2} & \dots & a_{in} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix},$$

получающуюся из $A = [\dots, A_{(i)}, \dots, A_{(j)}, \dots]$ заменой j -й строки на i -ю (i -я строка остается на месте). Как и у всякой другой квадратной матрицы с двумя одинаковыми строками, $\det A' = 0$. С другой стороны, алгебраическое дополнение A'_{jk} ($k = 1, \dots, n$) образуется путём зачёркивания j -й строки $A'_{(j)} = A_{(i)}$ и k -го столбца определителя, так что $A'_{jk} = A_{jk}$. Формальное разложение определителя матрицы $A' = (a'_{st})$ по j -й строке даст нам соотношение

$$0 = \det A' = \sum_{k=1}^n a'_{jk} A'_{jk} = \sum_{k=1}^n a_{ik} A_{jk},$$

совпадающее с соотношением (1) в формулировке леммы. Второе соотношение получается из аналогичных соображений, относящихся к столбцам. \square

Обращаясь к доказательству теоремы, мы просто замечаем, что левая часть соотношения (1) есть не что иное, как элемент c_{ij} матрицы $C = AA^\vee$:

$$\left\| \begin{array}{cccc} c_{11} & \dots & c_{n1} \\ \dots & \dots & \dots \\ c_{1n} & \dots & c_{nn} \end{array} \right\| = \left\| \begin{array}{cccc} a_{11} & \dots & a_{n1} \\ \dots & \dots & \dots \\ a_{1n} & \dots & a_{nn} \end{array} \right\| = \left\| \begin{array}{cccc} A_{11} & \dots & A_{n1} \\ \dots & \dots & \dots \\ A_{1n} & \dots & A_{nn} \end{array} \right\|.$$

Согласно соотношению (1) $(c_{ij}) = (\delta_{ij} \det A) = (\det A)E$. Таким образом,

$$AA^\vee = (\det A)E,$$

откуда при $\det A \neq 0$ получаем

$$(\det A)^{-1}(AA^\vee) = A(\det A)^{-1}A^\vee = E.$$

Левая часть соотношения (2) является выражением элемента c'_{ji} матрицы $C' = A^\vee A$. Так как правые части в (1) и (2) совпадают, то в случае $\det A \neq 0$ мы приходим к соотношениям

$$A(\det A)^{-1}A^\vee = (\det A)^{-1}A^\vee A = E,$$

означающим, что $A^{-1} = (\det A)^{-1}A^\vee$. \square

Следствие. Определитель равен нулю тогда и только тогда, когда его строки (и столбцы) линейно зависимы.

Доказательство. Линейная зависимость строк (или столбцов) матрицы $A \in M_n(\mathbb{R})$ эквивалентна неравенству $\text{rank } A < n$, т.е. вырожденности матрицы A , что по теореме 1 равносильно условию $\det A = 0$. \square

Замечание. Импликация $\text{rank } A < n \implies \det A = 0$ является, конечно, непосредственным следствием основных свойств определителей (см. D2, D6 в § 2).

Теорема 1 имеет скорее теоретическое значение. С вычислительной точки зрения, в особенности при больших размерах матриц, для отыскания матрицы A^{-1} удобнее пользоваться методом (P, Q) -приведения, описанным в п. 7 гл. 2.

2. Формулы Крамера. Выведем теперь формулы для решения системы из n линейных уравнений с n неизвестными, ради которых, в частности, и была первоначально развита теория определителей.

Теорема 2 (Крамер). Если линейная система

$$\begin{aligned} a_{11}x_1 + \dots + a_{1n}x_n &= b_1, \\ \dots &\dots \\ a_{n1}x_1 + \dots + a_{nn}x_n &= b_n \end{aligned}$$

имеет отличный от нуля определитель (т.е. $\det(a_{ij}) \neq 0$), то её

единственное решение задаётся формулами

$$x_k^0 = \frac{\begin{vmatrix} a_{11} & \dots & b_1 & \dots & a_{1n} \\ \dots & \dots & \dots & \dots & \dots \\ a_{n1} & \dots & b_n & \dots & a_{nn} \\ \hline a_{11} & \dots & a_{1k} & \dots & a_{1n} \\ \dots & \dots & \dots & \dots & \dots \\ a_{n1} & \dots & a_{nk} & \dots & a_{nn} \end{vmatrix}}{\begin{vmatrix} a_{11} & \dots & a_{1k} & \dots & a_{1n} \\ \dots & \dots & \dots & \dots & \dots \\ a_{n1} & \dots & a_{nk} & \dots & a_{nn} \end{vmatrix}}, \quad k = 1, 2, \dots, n$$

(числитель D_k получается заменой k -го столбца в $D = \det(a_{ij})$ столбцом свободных членов).

Доказательство. По теореме 1 матрица $A = (a_{ij})$ обратима. Поэтому, записав нашу систему в виде $AX = B$, мы, как и в п. 8 § 3 гл. 2, будем иметь

$$\left\| \begin{array}{c} x_1^0 \\ \vdots \\ x_k^0 \\ \vdots \\ x_n^0 \end{array} \right\| = A^{-1}B = \frac{1}{\det A} \left\| \begin{array}{cccc} A_{11} & A_{21} & \dots & A_{n1} \\ A_{12} & A_{22} & \dots & A_{n2} \\ \dots & \dots & \dots & \dots \\ A_{1n} & A_{2n} & \dots & A_{nn} \end{array} \right\| \left\| \begin{array}{c} b_1 \\ b_2 \\ \vdots \\ b_n \end{array} \right\|,$$

откуда

$$\begin{aligned} x_k^0 &= \frac{1}{\det A} \sum_{i=1}^n A_{ik} b_i = \\ &= \frac{1}{\det A} (b_1 A_{1k} + b_2 A_{2k} + \dots + b_n A_{nk}), \quad k = 1, 2, \dots, n. \end{aligned}$$

Именно такое выражение в числителе мы получим, разложив определитель D_k по k -му столбцу (см. (2)).

Выполнение всех преобразований в обратном порядке показывает, что набор $(D_1 / \det A, \dots, D_n / \det A)$ действительно является решением нашей системы. \square

Заметим, что формулы (3), (9) из § 4 гл. 1 совпадают как раз с формулами Крамера при $n = 2$ и $n = 3$ соответственно. Удобные при небольших n формулы Крамера несут в общем чисто теоретическую функцию. Например, их применение к линейной системе из примера 2 в п. 5 § 3 гл. 1 даёт (с учётом равенства $\det A = 1$) для чисел Фибоначчи выражение

$$f_n = \frac{\begin{vmatrix} 1 & 0 & 0 & \dots & 0 & 0 & 1 \\ 0 & 1 & 0 & \dots & 0 & 0 & 1 \\ -1 & -1 & 1 & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & -1 & 1 & 0 \\ 0 & 0 & 0 & \dots & -1 & -1 & 0 \end{vmatrix}}{\begin{vmatrix} 1 & 0 & 0 & \dots & 0 & 0 & 1 \\ 0 & 1 & 0 & \dots & 0 & 0 & 1 \\ -1 & -1 & 1 & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & -1 & 1 & 0 \\ 0 & 0 & 0 & \dots & -1 & -1 & 0 \end{vmatrix}}.$$

Понятно, что оно весьма далеко от того явного выражения для f_n , которое мы нашли в п. 5 § 3 гл. 2.

Надо сказать ещё, что необходимое для применения формул Крамера условие $\det A \neq 0$ неустойчиво в следующем смысле. Для реальных квадратных линейных систем с приближённо вычисленными коэффициентами увеличение точности вычислений может радикально изменить картину. Если, например,

$$A_\epsilon = \begin{vmatrix} -1 & 10 & 0 & \dots & 0 & 0 \\ 0 & -1 & 10 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & -1 & 10 \\ \epsilon & 0 & 0 & \dots & 0 & -1 \end{vmatrix} \in M_{10}(\mathbb{R}),$$

то $\det A_\epsilon = 1 - \epsilon \cdot 10^9$ (разложить определитель по элементам первого столбца). При $\epsilon = 10^{-9}$ имеем $\det A_\epsilon = 0$. В то же время, вычисляя коэффициенты матрицы всего лишь с точностью до одной миллионной, мы могли “не заметить ϵ ” (т.е. посчитать $\epsilon = 0$, а $\det A_0 = 1$). Таким образом, условия применимости формул Крамера чувствительны к малому “шевелению” коэффициентов системы.

3. Метод окаймляющих миноров. В § 3 гл. 2 содержится всё необходимое для описания совокупности решений прямоугольной системы линейных уравнений. Важнейшая роль в этом описании принадлежит понятию ранга матрицы. Нам осталось лишь перевести его на язык теории определителей, чтобы получить в своё распоряжение ещё один метод вычисления ранга и удобное средство для выражения факта линейной независимости системы векторов линейного пространства \mathbb{R}^m .

Итак, пусть

$$A = \begin{vmatrix} a_{11} & \dots & a_{1r} & \dots & a_{1n} \\ \dots & \dots & \dots & \dots & \dots \\ a_{r1} & \dots & a_{rr} & \dots & a_{rn} \\ \dots & \dots & \dots & \dots & \dots \\ a_{m1} & \dots & a_{mr} & \dots & a_{mn} \end{vmatrix}$$

— произвольная прямоугольная матрица размера $m \times n$ с коэффициентами $a_{ij} \in \mathbb{R}$.

Определение. Элементы, стоящие на пересечении каких-то выделенных k строк и k столбцов $m \times n$ -матрицы A ($k \leq \min(m, n)$), составляют квадратную матрицу, определитель которой называется **минором k -го порядка** для A . Иногда говорят о миноре

$$M \begin{pmatrix} i_1 & \dots & i_k \\ j_1 & \dots & j_k \end{pmatrix},$$

если i_1, \dots, i_k и j_1, \dots, j_k — номера выделенных строк и столбцов.

При $k = n - 1$ мы приходим к ранее введённому понятию минора M_{ij} для $n \times n$ -матрицы A .

Минор \tilde{M} называется окаймляющим для M , если M получается из \tilde{M} вычёркиванием одной крайней строки (первой или последней) и одного крайнего столбца.

Теорема (метод окаймляющих миноров). При вычислении ранга матрицы A следует переходить от миноров меньших порядков к минорам больших порядков. Если для A уже найден минор $M \neq 0$ порядка r , то требуют вычисления лишь миноры порядка $r+1$, окаймляющие минор M . Если все они равны нулю, то $\text{rank } A = r$.

Доказательство. Рассуждение основано на простом замечании, что если все миноры k -го порядка матрицы A равны нулю, то равны нулю и все миноры более высоких порядков. Для этого согласно теореме 1 § 2 достаточно рассмотреть разложение любого минора порядка $k+1$ по элементам какого-нибудь столбца (например, первого или последнего, если ограничиться рассмотрением только миноров, полученных посредством окаймления), затем перейти к минорам порядка $k+2$ и т.д.

Действуя теперь по схеме, указанной в формулировке теоремы, мы дойдём до какого-то минора $M \neq 0$ порядка r . Без ограничения общности считаем, что M отвечает матрице, стоящей в левом верхнем углу нашей матрицы:

$$A = \left| \begin{array}{cccc|cccc} a_{11} & \dots & a_{1r} & & \dots & a_{1j} & \dots & a_{1n} \\ M & & & & \dots & & & \\ \hline a_{r1} & \dots & a_{rr} & & \dots & a_{rj} & \dots & a_{rn} \\ \dots & & \dots & & \dots & \dots & & \dots \\ a_{i1} & \dots & a_{ir} & & a_{ii} & \dots & a_{in} \\ \dots & & \dots & & \dots & & \dots \\ a_{m1} & \dots & a_{mr} & & a_{mj} & \dots & a_{mn} \end{array} \right|.$$

Этого всегда можно достичь перестановкой строк и столбцов, не меняющей, как нам известно, ранга матрицы A .

Выделим теперь в A строку $A_{(i)}$ и столбец $A^{(j)}$ с совершенно произвольными номерами i, j (возможно, $i \leq r$ или $j \leq r$). Составим при помощи элементов из $A_{(i)}$ и $A^{(j)}$ минор \tilde{M} порядка $r+1$, окаймляющий M :

$$\tilde{M} = \left| \begin{array}{cccc} a_{11} & \dots & a_{1r} & a_{1j} \\ \dots & & \dots & \dots \\ a_{r1} & \dots & a_{rr} & a_{rj} \\ a_{i1} & \dots & a_{ir} & a_{ij} \end{array} \right|.$$

Если $\tilde{M} \neq 0$, переходим к минорам, окаймляющим \tilde{M} . Критический момент наступит, когда все окаймляющие M миноры будут равны нулю.

Итак, пусть $\widetilde{M} = 0$ при любом выборе i, j . Разлагая M по элементам последней строки, придём к соотношению

$$a_{11}M_1 + a_{12}M_2 + \dots + a_{ir}M_r + a_{ij}M = 0$$

с коэффициентами

$$M_s = (-1)^{r+s+1} \begin{vmatrix} a_{11} & \dots & a_{1,s-1} & a_{1,s+1} & \dots & a_{1r} & a_{1j} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ a_{r1} & \dots & a_{r,s-1} & a_{r,s+1} & \dots & a_{rr} & a_{rj} \end{vmatrix},$$

не зависящими от i . Так как $M \neq 0$, то

$$a_{ij} = \lambda_1 a_{11} + \lambda_2 a_{12} + \dots + \lambda_r a_{ir}$$

для $i = 1, 2, \dots, r$ с одними и теми же коэффициентами $\lambda_s = -M_s/M$, $1 \leq s \leq r$. Стало быть,

$$A^{(j)} = \lambda_1 A^{(1)} + \lambda_2 A^{(2)} + \dots + \lambda_r A^{(r)},$$

т.е. любой столбец $A^{(j)}$ является линейной комбинацией первых r столбцов. Это значит, что $\text{rank } A \leq r$. Но из $M \neq 0$ вытекает линейная независимость столбцов в M и тем более — соответствующих более длинных столбцов в A . Мы приходим к заключению, что $\text{rank } A = r$. \square

Следствие. Ранг всякой матрицы совпадает с наивысшим порядком её отличных от нуля миноров.

Для следствия можно указать короткое независимое доказательство. Именно, пусть ранг матрицы A равен r . Согласно теореме 1 § 2 гл. 2 это значит, что r — максимальное число линейно независимых строк и максимальное число линейно независимых столбцов матрицы A . Обращаясь к теореме 6 § 3 гл. 2, мы замечаем, что

$$A = B \left\| \begin{array}{cc} E_r & 0 \\ 0 & 0 \end{array} \right\| C,$$

где B и C — невырожденные матрицы порядков r и n соответственно, записываемые в виде произведения элементарных матриц.

Так как у матрицы $\left\| \begin{array}{cc} E_r & 0 \\ 0 & 0 \end{array} \right\|$ имеется отличный от нуля минор $M = |E_r| = 1$ порядка r , но нет ненулевых миноров порядка $> r$, и так как это свойство сохраняется при элементарных преобразованиях строк и столбцов, то мы приходим к нужному утверждению. \square

Метод окаймляющих миноров достаточно практичен, особенно тогда, когда мы хотим знать не только ранг, но и те столбцы или строки матрицы A , которые составляют максимальную линейно независимую систему. При элементарных преобразованиях эта информация, конечно, утрачивается.

УПРАЖНЕНИЯ

1. Показать, что выполнены следующие соотношения:

$$(AB)^V = B^V A^V; \quad (^t A)^V = ^t (A^V); \quad (\lambda A)^V = \lambda^{n-1} A^V;$$

$$(A^V)^V = (\det A)^{n-2} A.$$

2. Выразить $\operatorname{rank} A^V$ через $\operatorname{rank} A$.

3. Доказать, что квадратная система линейных однородных уравнений тогда и только тогда обладает нетривиальными решениями, когда определитель системы равен нулю.

4. Опираясь на результаты п. 8 § 3 гл. 2 и на теорему 2, показать, что фундаментальная система решений однородной системы

$$a_{11}x_1 + \dots + a_{1n}x_n = 0,$$

· · · · ·

$$a_{n-1,1}x_1 + \dots + a_{n-1,n}x_n = 0$$

ранга $r = n - 1$ будет состоять из одного вектора-столбца

$$X^0 = [D_1, -D_2, D_3, \dots, (-1)^{n-1} D_n],$$

где D_i — определитель матрицы, получающейся из $A = (a_{ij})$ вычёркиванием её i -го столбца. Любое решение системы имеет вид $X = \lambda X^0$.

5. Пусть $A = (a_{ij}) \in M_n(\mathbb{R})$ и $(n-1)|a_{ij}| < |a_{ii}|$ для всех $i \neq j$. Доказать, что $\det A \neq 0$.

Указание. Предположив противное, воспользоваться критерием, сформулированным в упр. 3. Именно, если $[x_1^0, \dots, x_n^0]$ — нетривиальное решение линейной системы $AX = 0$ и x_k^0 — его компонента, имеющая максимальный модуль, то из k -го уравнения

$$a_{kk}x_k^0 + \sum_{j \neq k} a_{kj}x_j^0 = 0$$

следует оценка

$$(n-1)|a_{kk}| |x_k^0| = (n-1) \left| \sum_{j \neq k} a_{kj}x_j^0 \right| < (n-1)|a_{kk}| |x_k^0|,$$

дающая нужное противоречие.

6. Доказать следующее утверждение (теорема Бине—Коши). Пусть $A = (a_{ij})$, $B = (b_{kl})$ — матрицы размеров $n \times m$ и $m \times n$ соответственно, и пусть $C = AB$. Тогда

$$\det C = \sum_{1 \leq j_1 < \dots < j_n \leq m} \begin{vmatrix} a_{1j_1} & a_{2j_1} & \dots & a_{nj_1} \\ a_{1j_2} & a_{2j_2} & \dots & a_{nj_2} \\ \dots & \dots & \dots & \dots \\ a_{1j_n} & a_{2j_n} & \dots & a_{nj_n} \end{vmatrix} \times \begin{vmatrix} b_{j_1 1} & b_{j_1 2} & \dots & a_{j_1 n} \\ b_{j_2 1} & b_{j_2 2} & \dots & a_{j_2 n} \\ \dots & \dots & \dots & \dots \\ b_{j_n 1} & b_{j_n 2} & \dots & a_{j_n n} \end{vmatrix}.$$

Суммирование в правой части проходит по всем $\binom{m}{n}$ возможным комбинациям по n элементов $\{j_1, j_2, \dots, j_n\}$ из $1, 2, \dots, m$. В частности, $\det C = \det A \cdot \det B$ при $m = n$ и $\det C = 0$ при $n > m$.

Указание. Так как

$$C = (c_{ij}), \quad c_{ij} = \sum_{k=1}^m a_{ik}b_{kj},$$

то многократное применение свойства D2 определителей (теорема 2 § 1) даёт

$$\det C = \sum_{k_1, \dots, k_n=1}^m \begin{vmatrix} a_{1k_1} & a_{1k_2} & \dots & a_{1k_n} \\ a_{2k_1} & a_{2k_2} & \dots & a_{2k_n} \\ \dots & \dots & \dots & \dots \\ a_{nk_1} & a_{nk_2} & \dots & a_{nk_n} \end{vmatrix} b_{k_1 1} b_{k_2 2} \dots b_{k_n n},$$

где суммирование проводится по всем попарно различным k_1, \dots, k_n . При $m < n$ таких индексов нет и, следовательно, $\det C = 0$. Если же $m \geq n$, то k_1, \dots, k_n — выборка элементов $\{j_1, \dots, j_n\}$, взятых в каком-то порядке из $1, 2, \dots, m$. Следует собрать все члены, соответствующие фиксированной комбинации $\{j_1, \dots, j_n\}$, и при помощи формулы (3) § 1 получить нужное выражение:

$$\sum \begin{vmatrix} a_{1k_1} & \dots & a_{nk_1} \\ \dots & \dots & \dots \\ a_{1k_n} & \dots & a_{nk_n} \end{vmatrix} b_{k_1 1} \dots b_{k_n n} = \begin{vmatrix} a_{1j_1} & \dots & a_{nj_1} \\ \dots & \dots & \dots \\ a_{1j_n} & \dots & a_{nj_n} \end{vmatrix} \cdot \sum \varepsilon_\pi b_{k_1 1} \dots b_{k_n n} = \\ = \begin{vmatrix} a_{1j_1} & \dots & a_{nj_1} \\ \dots & \dots & \dots \\ a_{1j_n} & \dots & a_{nj_n} \end{vmatrix} \begin{vmatrix} b_{j_1 1} & \dots & b_{j_1 n} \\ \dots & \dots & \dots \\ b_{j_n 1} & \dots & b_{j_n n} \end{vmatrix},$$

где $\pi = \begin{pmatrix} j_1 & \dots & j_n \\ k_1 & \dots & k_n \end{pmatrix}$.

7. Используя предыдущее упражнение, показать, что если A — $m \times n$ -матрица над \mathbb{R} , $m \geq n$, то

$$\det {}^t A A = \sum_M M^2,$$

где M пробегает по всем $\binom{m}{n}$ минорам порядка n матрицы A .

8. Данному минору

$$M \begin{pmatrix} i_1 & \dots & i_k \\ j_1 & \dots & j_k \end{pmatrix}$$

порядка k для $n \times n$ -матрицы $A = (a_{ij})$ (см. определение в п. 3) отвечает дополнительный минор $\overline{M} \begin{pmatrix} i_1 & \dots & i_k \\ j_1 & \dots & j_k \end{pmatrix}$ порядка $n - k$, матрица которого получается из A вычёркиванием строк с номерами i_1, \dots, i_k и столбцов с номерами j_1, \dots, j_k . Выражение

$$A \begin{pmatrix} i_1 & \dots & i_k \\ j_1 & \dots & j_k \end{pmatrix} = (-1)^s(M) \overline{M} \begin{pmatrix} i_1 & \dots & i_k \\ j_1 & \dots & j_k \end{pmatrix},$$

$$s(M) = (i_1 + \dots + i_k) + (j_1 + \dots + j_k),$$

называется алгебраическим дополнением к $M \begin{pmatrix} i_1 & \dots & i_k \\ j_1 & \dots & j_k \end{pmatrix}$. При $k = n - 1$ мы возвращаемся к обычному определению. Используя последовательно разложение определителя по элементам строк с номерами i_1, \dots, i_k , показать, что справедлива следующая

Теорема (Лаплас). Пусть в матрице $A = (a_{ij}) \in M_n(\mathbb{R})$ выбраны k строк с номерами i_1, \dots, i_k . Тогда

$$\det A = \sum_{1 \leq j_1 < \dots < j_k \leq n} M \begin{pmatrix} i_1 & \dots & i_k \\ j_1 & \dots & j_k \end{pmatrix} \cdot A \begin{pmatrix} i_1 & \dots & i_k \\ j_1 & \dots & j_k \end{pmatrix}.$$

При произвольном n теорема Лапласа известна нам в двух частных случаях:
 1) $k = 1$; 2) A — матрица с углом нулей размера $(n - k) \times k$. В случае неудачи полезно убедиться в правильности теоремы Лапласа хотя бы при $n = 4$, $i_1 = 1$, $i_2 = 2$:

$$\det A = \left| \begin{array}{cc} a_{11} & a_{12} \\ a_{21} & a_{22} \end{array} \right| \cdot \left| \begin{array}{cc} a_{33} & a_{34} \\ a_{43} & a_{44} \end{array} \right| - \left| \begin{array}{cc} a_{11} & a_{13} \\ a_{21} & a_{23} \end{array} \right| \cdot \left| \begin{array}{cc} a_{32} & a_{34} \\ a_{42} & a_{44} \end{array} \right| + \\ + \left| \begin{array}{cc} a_{11} & a_{14} \\ a_{21} & a_{24} \end{array} \right| \cdot \left| \begin{array}{cc} a_{32} & a_{33} \\ a_{42} & a_{43} \end{array} \right| + \left| \begin{array}{cc} a_{12} & a_{13} \\ a_{22} & a_{23} \end{array} \right| \cdot \left| \begin{array}{cc} a_{31} & a_{34} \\ a_{41} & a_{44} \end{array} \right| - \\ - \left| \begin{array}{cc} a_{12} & a_{14} \\ a_{22} & a_{24} \end{array} \right| \cdot \left| \begin{array}{cc} a_{31} & a_{33} \\ a_{41} & a_{43} \end{array} \right| + \left| \begin{array}{cc} a_{13} & a_{14} \\ a_{23} & a_{24} \end{array} \right| \cdot \left| \begin{array}{cc} a_{31} & a_{32} \\ a_{41} & a_{42} \end{array} \right|.$$

9. Пусть $A \in M_n(\mathbb{R})$, $B \in M_m(\mathbb{R})$ — невырожденные матрицы, C — произвольная $n \times m$ -матрица. Используя приём блочного умножения матриц (см. упр. 17 из § 3 гл. 2), показать, что

$$\left\| \begin{array}{cc} A & C \\ 0 & B \end{array} \right\|^{-1} = \left\| \begin{array}{cc} A^{-1} & -A^{-1}CB^{-1} \\ 0 & B^{-1} \end{array} \right\|.$$

10. Показать, что если $A, B, C, D \in M_n(\mathbb{R})$, $\det A \neq 0$, то

$$\det \left\| \begin{array}{cc} A & B \\ C & D \end{array} \right\| = \det(AD - ACA^{-1}B) = (\det A) \cdot \det(D - CA^{-1}B).$$

Кроме того, проверить, что

$$\det \left\| \begin{array}{cc} A & B \\ C & D \end{array} \right\| = \begin{cases} \det(AD - CB), & \text{если } AC = CA, \\ \det(DA - CB), & \text{если } AB = BA. \end{cases}$$

§ 4. К построению теории определителей

Теоремы 2 и 3 из § 1 дают по существу аксиоматическое описание функции \det , хотя начинали мы с чисто конструктивного её задания. Укажем ещё несколько подходов к теории определителей, каждый раз ограничиваясь наброском канвы рассуждений. (Полное их проведение является хорошим упражнением.)

1. Первое аксиоматическое построение. Будем считать определителем любую функцию $D: M_n(\mathbb{R}) \rightarrow \mathbb{R}$, обладающую следующими тремя свойствами:

- 1.1) $D(A)$ — кососимметрическая функция строк матрицы A ;
- 1.2) $D(A)$ — полилинейная функция строк матрицы A ;
- 1.3) $D(E) = 1$.

Мы видели, что свойствами 1.1)-1.3) функция D однозначно характеризуется и совпадает с функцией \det , определённой формулой полного развертывания (3) § 1. Единственное, о чём нужно позаботиться, это дать независимое доказательство факту $D(^t A) = D(A)$. Сама формула (3) § 1, если угодно, также нуждается в выводе.

2. Второе аксиоматическое построение. Определителем считаем любую функцию $\mathcal{D}: M_n(\mathbb{R}) \rightarrow \mathbb{R}$, обладающую тремя свойствами:

2.1) $\mathcal{D}(\dots, \lambda A_{(i)}, \dots) = \lambda \mathcal{D}(\dots, A_{(i)}, \dots)$, т.е. если одну из строк $A_{(i)}$ матрицы A умножить на λ , то значение $\mathcal{D}(A)$ также умножается на λ ;

2.2) $\mathcal{D}(\dots, A_{(i)}, \dots, A_{(j)}, \dots) = \mathcal{D}(\dots, A_{(i)} + A_{(j)}, \dots, A_{(j)}, \dots)$;

2.3) $\mathcal{D}(E) = 1$.

Последовательно проверяется, что:

а) значение $\mathcal{D}(A)$ не меняется при элементарных преобразованиях типа (II) над строками матрицы A ;

б) $\mathcal{D}(A)$ — полилинейная функция строк матрицы A ;

в) $\mathcal{D}(A) = 0$ при равенстве двух строк матрицы A и, следовательно, $\mathcal{D}(A)$ — кососимметрическая функция строк.

Мы вернулись, очевидно, к первому аксиоматическому построению. Нормировочное свойство $\mathcal{D}(E) = 1$ в обоих случаях необходимо.

3. Построение методом полной индукции. Возьмём в качестве определителя матрицы (a_{11}) порядка 1 число a_{11} . Определители матриц порядков 2 и 3 вводятся соответственно формулами (2) и (8) из § 4 гл. 1. Пусть определители матриц порядков $1, 2, \dots, n-1$ уже введены. Назовём определителем матрицы $A = (a_{ij})$ порядка n величину

$$\mathcal{D}(A) = a_{11}M_{11} - a_{21}M_{21} + \dots + (-1)^{n-1}a_{n1}M_{n1},$$

где M_{ij} — “минор” матрицы A , соответствующий элементу a_{ij} и являющийся определителем $\mathcal{D}(\bar{A})$ матрицы \bar{A} порядка $n-1$, которая получается из A вычёркиванием строки с номером i и столбца с номером j . Таким образом, в качестве исходного свойства берётся разложение определителя по элементам первого столбца (частный случай теоремы 1 § 2).

Используя индукцию по n , нужно установить свойства 1.1)-1.3) функции \mathcal{D} применительно к матрицам порядка n , памятая, что для M_{ij} эти свойства выполнены. Реализация этой программы, закрепляющей навыки в грамотном применении метода индукции, не очень сложна. С деталями можно познакомиться по учебнику “Введение в алгебру” (1977 г.).

4. Характеризация мультипликативными свойствами. Пусть мы имеем функцию $\mathcal{D}: M_n(\mathbb{R}) \rightarrow \mathbb{R}$, обладающую следующими свойствами:

i) $\mathcal{D}(AB) = \mathcal{D}(A)\mathcal{D}(B)$ для любых матриц $A, B \in M_n(\mathbb{R})$;

ii) $\mathcal{D}(F_{s,t}) = -1$ для каждой элементарной матрицы $F_{s,t}$ (см. п. 6 § 3 гл. 2);

iii) $\mathcal{D}(A) = \lambda$ для верхних треугольных матриц вида

$$A = \begin{vmatrix} \lambda & & * \\ & 1 & \\ 0 & \ddots & 1 \end{vmatrix}, \quad \lambda \in \mathbb{R}.$$

В частности, $\mathcal{D}(F_1(\lambda)) = \lambda$.

Утверждается, что $\mathcal{D} = \det$. В самом деле, воспользовавшись свойствами i) и ii) в применении к матрице

$$F_s(\lambda) = F_{1,s} \cdot F_1(\lambda) \cdot F_{1,s},$$

мы получим

$$\mathcal{D}(F_s(\lambda)) = (-1) \cdot \lambda \cdot (-1) = \lambda,$$

причём это верно при любом $\lambda \in \mathbb{R}$, а не только при $\lambda \neq 0$, когда по определению матрица $F_s(\lambda)$ элементарна. Отсюда для

$$\begin{vmatrix} E_r & 0 \\ 0 & 0 \end{vmatrix} = F_{r+1}(0) \dots F_n(0)$$

имеем

$$\mathcal{D}\left(\begin{vmatrix} E_r & 0 \\ 0 & 0 \end{vmatrix}\right) = \begin{cases} 0, & \text{если } r < n, \\ 1, & \text{если } r = n. \end{cases}$$

Согласно iii) $\mathcal{D}(F_{s,t}(\lambda)) = 1$ для элементарной матрицы $F_{s,t}(\lambda)$ с $s < t$. Так как

$$F_{s,t} F_{s,t}(\lambda) F_{s,t} = F_{t,s}(\lambda),$$

то и $\mathcal{D}(F_{t,s}(\lambda)) = 1$, а поэтому

$$\mathcal{D}(F_{s,t}(\lambda)) = 1$$

при любых индексах $s \neq t$.

Итак,

$$\begin{aligned} \mathcal{D}(F_{s,t}) &= -1 = \det F_{s,t}, & \mathcal{D}(F_{s,t}(\lambda)) &= 1 = \det F_{s,t}(\lambda), \\ \mathcal{D}(F_s(\lambda)) &= \lambda = \det F_s(\lambda). \end{aligned}$$

Поскольку любая матрица $A \in M_n(\mathbb{R})$ записывается в виде

$$A = P \begin{vmatrix} E_r & 0 \\ 0 & 0 \end{vmatrix} Q, \quad r \leq n,$$

где P и Q — произведения элементарных матриц (см. рассуждения перед теоремой 6 § 3 гл. 2), свойство i) позволяет заключить, что $\mathcal{D}(A) = \det A$.

УПРАЖНЕНИЯ

1 (J. Browkin, Poland). Пусть $f: \mathbb{R} \rightarrow \mathbb{R}$ — произвольная функция с условием $f(0) = 0$.

Доказать, что существует, и притом только одна, функция $\mathcal{D}: M_n(\mathbb{R}) \rightarrow \mathbb{R}$, обладающая следующими свойствами:

- i) если A содержит столбец нулей, то $\mathcal{D}(A) = 0$;
- ii) если A' получается из A элементарным преобразованием типа (II) над столбцами, то $\mathcal{D}(A') = \mathcal{D}(A)$;
- iii) если $A = \text{diag}(\lambda, 1, 1, \dots, 1)$ — диагональная матрица, то $\mathcal{D}(A) = f(\lambda)$.

При $f(\lambda) = \lambda$ получаем $\mathcal{D} = \det$, но произвол в выборе f полезен в других приложениях.

2. Читателю предлагается выдвинуть и обосновать собственные варианты аксиоматического описания функции \det .

ГРУППЫ. КОЛЬЦА. ПОЛЯ

В предыдущих главах накопилось довольно много конкретного материала, который необходимо осмыслить с более общих позиций. С этой целью мы введём и изучим (пока на элементарном уровне), фундаментальные для всей алгебры понятия группы, кольца и поля.

§ 1. Множества с алгебраическими операциями

1. Бинарные операции. Пусть X — произвольное множество. *Бинарной алгебраической операцией* (или *законом композиции*) на X называется произвольное (но фиксированное) отображение $\tau: X \times X \rightarrow X$ декартова квадрата $X^2 = X \times X$ в X . Таким образом, любой упорядоченной паре (a, b) элементов $a, b \in X$ ставится в соответствие однозначно определённый элемент $\tau(a, b)$ того же множества X . Иногда вместо $\tau(a, b)$ пишут ab , а ещё чаще бинарную операцию на X обозначают каким-нибудь специальным символом: $*$, \circ , \cdot или $+$. Последуем и мы по тому же пути, называя $a \cdot b$ (или просто ab , без всякого значка между a и b) *произведением*, а $a + b$ — *суммой элементов* $a, b \in X$. Понятно, что эти названия в большинстве случаев условны.

На X может быть задано, вообще говоря, много разных операций. Желая выделить одну из них, используют скобки: $(X, *)$, и говорят, что операция $*$ определяет на X *алгебраическую структуру* или что $(X, *)$ — *алгебраическая структура* (*алгебраическая система*). Так, например, на множестве \mathbb{Z} целых чисел, помимо естественных операций $+, \cdot$ (сложения и умножения), легко указать получающиеся при помощи $+$ (или $-$) и \cdot “производные” операции: $n \circ m = n + m - nm$, $n * m = -n - m$ и т.д. Мы приходим к различным алгебраическим структурам $(\mathbb{Z}, +)$, (\mathbb{Z}, \cdot) , (\mathbb{Z}, \circ) , $(\mathbb{Z}, *)$.

Наряду с бинарными алгебраическими операциями не лишены интереса гораздо более общие n -арные операции (унарные при $n = 1$, тернарные при $n = 3$ и т.д.), равно как и их комбинации. Связанные с ними алгебраические структуры составляют специальную теорию универсальных алгебр. Впрочем, мы упоминаем об этом только для того, чтобы лишний раз подчеркнуть принципиальную важность для математики, казалось бы, частных разделов теории универсальных алгебр — алгебраических структур с бинарными операциями.

В направлении конструирования разных бинарных операций на множестве X также, очевидно, открывается неограниченный простор для фантазии. Но задача изучения произвольных алгебраических структур слишком обща, чтобы представлять реальную ценность. По

этой причине её рассматривают при различных естественных ограничениях.

2. Полугруппы и моноиды. Бинарная операция $*$ на множестве X называется *ассоциативной*, если

$$(a * b) * c = a * (b * c)$$

для всех $a, b, c \in X$; она называется *коммутативной*, если

$$a * b = b * a.$$

Те же названия присваиваются и соответствующей алгебраической структуре $(X, *)$.

Требования ассоциативности и коммутативности независимы. В самом деле, операция $*$ на \mathbb{Z} , заданная правилом

$$n * m = -n - m,$$

очевидно, коммутативна, но

$$(1 * 2) * 3 = (-1 - 2) * 3 = -(-1 - 2) - 3 = 0 \neq 4 = 1 * (2 * 3),$$

так что условие ассоциативности не выполняется. Далее, на множестве $M_n(\mathbb{R})$ всех квадратных матриц порядка $n > 1$ определена операция умножения — ассоциативная, но некоммутативная (см. п. 2 § 3 гл. 2).

Элемент $e \in X$ называется *единичным* (или *нейтральным*) относительно рассматриваемой бинарной операции $*$, если $e * x = x * e = x$ для всех $x \in X$. Если e' — ещё один единичный элемент, то, как следует из определения, $e' = e' * e = e$. Стало быть, в алгебраической структуре $(X, *)$ может существовать не более одного единичного элемента.

Множество X с заданной на нём бинарной ассоциативной операцией называется *полугруппой*. Полугруппу с единичным (нейтральным) элементом принято называть *ещё моноидом* (или *полугруппой с единицей*).

Как и для всякого множества, мощность моноида $M = (M, *)$ обозначается символом $\text{Card } M$ или $|M|$. В случае конечности числа содержащихся в нём элементов говорят о конечном моноиде M порядка $|M|$. Приведём несколько примеров полугрупп и моноидов.

Пример 1. Пусть Ω — произвольное множество и $M(\Omega)$ — множество всех его преобразований (отображений Ω в себя). Из свойств множеств и отображений, отмеченных в § 5 гл. 1, следует, что $M(\Omega)$ — моноид. Имеется в виду, конечно, тройка $(M(\Omega), \circ, e_\Omega)$ где \circ — естественная композиция отображений, а e_Ω — тождественное отображение.

Выделим тот частный случай, когда Ω — конечное множество из $|\Omega| = n$ элементов, обозначаемых просто натуральными числами $1, 2, \dots, n$. Каждое преобразование $f : \Omega \rightarrow \Omega$ определяется указанием упорядоченной последовательности $f(1), f(2), \dots, f(n)$, где в качестве образа $f(i)$ может стоять любой элемент из Ω . Не исключаются совпадения $f(i) = f(j)$ при $i \neq j$. Выбирая все возможные последовательности, мы получим ровно n^n преобразований. Значит,

$|M(\Omega)| = \text{Card } M(\Omega) = n^n$. Пусть, скажем, $n = 2$. Элементы e, f, g, h монида $M(\{1, 2\})$ и их попарные произведения полностью задаются следующими двумя таблицами:

	1	2		·	e	f	g	h
e	1	2	e	e	f	g	h	h
f	2	1	f	f	e	h	g	g
g	1	1	g	g	g	g	g	g
h	2	2	h	h	h	h	h	h

Непосредственно видно, что $M(\{1, 2\})$ — некоммутативный монида.

Пример 2. Пусть снова Ω — произвольное множество и $\mathcal{P}(\Omega)$ — множество всех его подмножеств (см. упр. 4 из § 5 гл. 1). Так как $(A \cap B) \cup C = A \cap (B \cup C)$ и $(A \cup B) \cap C = A \cup (B \cap C)$, то на $\mathcal{P}(\Omega)$ определены две естественные ассоциативные бинарные операции. Очевидно, что $\emptyset \cup A = A$ и $A \cap \Omega = A$. Мы имеем два коммутативных монида: $(\mathcal{P}(\Omega), \cup, \emptyset)$ и $(\mathcal{P}(\Omega), \cap, \Omega)$. Как известно, $|\mathcal{P}(\Omega)| = 2^n$, если $|\Omega| = n$.

Пример 3. $(M_n(\mathbb{R}), +, 0)$ — коммутативный монида с нейтральным элементом — нулевой матрицей, а $(M_n(\mathbb{R}), \cdot, E)$ — некоммутативный монида с нейтральным элементом — единичной матрицей E . Это непосредственно вытекает из свойства сложения и умножения матриц, с которыми мы познакомились в гл. 2.

Пример 4. Пусть $n\mathbb{Z} = \{nm \mid m \in \mathbb{Z}\}$ — множество целых чисел, делящихся на n . Ясно, что $(n\mathbb{Z}, +, 0)$ — коммутативный монида, а $(n\mathbb{Z}, \cdot)$ — коммутативная полугруппа без единицы ($n > 1$).

Пример 5. Множество $P_n(\mathbb{R})$ стохастических матриц порядка n (см. упр. 4 из § 3 гл. 2) является мониодом с обычной операцией умножения матриц.

Подмножество S' полугруппы S с операцией $*$ называется подполугруппой, если $x * y \in S'$ для всех $x, y \in S'$. В этом случае говорят ещё, что подмножество $S' \subset S$ замкнуто относительно операции $*$. Если $(M, *)$ — мониод, а подмножество $M' \subset M$ не только замкнуто относительно операции $*$, но и содержит единичный элемент, то M' называется подмониодом в M . Например, $(n\mathbb{Z}, \cdot)$ — подполугруппа в (\mathbb{Z}, \cdot) , а $(n\mathbb{Z}, +, 0)$ — подмониод в $(\mathbb{Z}, +, 0)$. Всякий подмониод мониода $M(\Omega)$ называется мониодом преобразований (множества Ω).

3. Обобщённая ассоциативность; степени. Пусть (X, \cdot) — произвольная алгебраическая структура с бинарной операцией \cdot , которую мы ради простоты будем опускать, записывая xy вместо $x \cdot y$. Пусть, далее, x_1, \dots, x_n — упорядоченная последовательность элементов из X . Не меняя порядка, мы можем многими разными способами составлять произведения длины n . Пусть l_n — число таких способов:

$$l_2 = 1: x_1 x_2;$$

$$l_3 = 2: (x_1 x_2) x_3, x_1 (x_2 x_3);$$

$$l_4 = 5: ((x_1 x_2) x_3) x_4, (x_1 (x_2 x_3)) x_4, x_1 ((x_2 x_3) x_4), x_1 (x_2 (x_3 x_4)), (x_1 x_2) (x_3 x_4);$$

Очевидно, что, перебирая всевозможные произведения $x_1 \dots x_k, x_{k+1} \dots x_n$ длины k и $n - k$, $1 \leq k \leq n - 1$, а затем соединяя их нашей бинарной операцией в данном порядке, мы исчерпаем все l_n возмож-

ностей. Замечательно, что в моноидах (и полугруппах) расстановка скобок оказывается излишней.

Теорема 1. Если бинарная операция на X ассоциативна, то результат её последовательного применения к n элементам множества X не зависит от расстановки скобок.

Доказательство. При $n = 1, 2$ доказывать нечего. При $n = 3$ утверждение теоремы совпадает с законом ассоциативности. Далее рассуждаем индукцией по n . Предположим, что $n > 3$ и что для числа элементов $< n$ справедливость утверждения установлена. Нам нужно лишь показать, что

$$(x_1 \dots x_k)(x_{k+1} \dots x_n) = (x_1 \dots x_l)(x_{l+1} \dots x_n) \quad (1)$$

при любых k, l , $1 \leq k, l \leq n - 1$. Мы выписали только внешние пары скобок, поскольку по предположению индукции расстановка внутренних скобок несущественна. В частности, $x_1 x_2 \dots x_k = (\dots ((x_1 x_2) x_3) \dots x_{k-1}) x_k$ — произведение, называемое левонормированным. Рассмотрим два случая:

а) $k = n - 1$; тогда $(x_1 \dots x_{n-1}) x_n = (\dots (x_1 x_2) \dots x_{n-1}) x_n$ — левонормированное произведение;

б) $k < n - 1$; ввиду ассоциативности имеем

$$\begin{aligned} (x_1 \dots x_k)(x_{k+1} \dots x_n) &= (x_1 \dots x_k)((x_{k+1} \dots x_{n-1}) x_n) = \\ &= ((x_1 \dots x_k)(x_{k+1} \dots x_{n-1})) x_n = \\ &= (\dots ((\dots (x_1 x_2) \dots x_k) x_{k+1}) \dots x_{n-1}) x_n, \end{aligned}$$

т.е. снова левонормированное произведение. К тому же виду приводится и правая часть доказываемого равенства (1). \square

Ранее был введён знак суммирования $\sum x_i$. Очевидно, его можно использовать и в любом аддитивном коммутативном моноиде. В мультипликативном моноиде аналогом служит знак кратного произведения:

$$\prod_{i=1}^2 x_i = x_1 x_2, \quad \prod_{i=1}^3 x_i = (x_1 x_2) x_3, \quad \prod_{i=1}^n = \left(\prod_{i=1}^{n-1} x_i \right) x_n.$$

В силу теоремы 1 при записи (или при вычислении) произведения элементов $x_1 x_2 \dots x_n$ монида скобки излишни. Единственная забота должна проявляться о порядке множителей, да и то лишь в случае, когда они не все перестановочны между собой. В частности, при $x_1 = x_2 = \dots = x_n = x$ произведение $xx \dots x$ обозначают, как и при действиях с числами, символом x^n , называя его n -й степенью элемента x . Следствием теоремы 1 являются соотношения

$$x^m x^n = x^{m+n}, \quad (x^m)^n = x^{mn}, \quad m, n \in \mathbb{N}. \quad (2)$$

В монидах (M, \cdot, e) для любого $x \in M$ полагают ещё $x^0 = e$.

Степеням $x^n \in (M, \cdot, e)$ в моноиде $(M, +, 0)$ соответствуют *кратные* $nx = x+x+\dots+x$ элемента x . Правила (2) становятся правилами для кратных:

$$mx + nx = (m+n)x, \quad n(mx) = (nm)x. \quad (2')$$

Отметим ещё один полезный факт. Если $xy = yx$ в моноиде M , то

$$(xy)^n = x^n y^n, \quad n = 0, 1, 2, \dots \quad (3)$$

В частности, это всегда так в коммутативном моноиде. Соотношение (3) доказывается индукцией по n :

$$\begin{aligned} (xy)^n &= (xy)^{n-1}(xy) = (x^{n-1}y^{n-1})(xy) = (x^{n-1}y^{n-1}x)y = \\ &= (x^{n-1}xy^{n-1})y = (x^{n-1}x)(y^{n-1}y) = x^n y^n. \end{aligned}$$

Более общо: при $x_i x_j = x_j x_i$, $i, j = 1, \dots, m$, опираясь на соотношение (3) и используя индукцию по m , получаем

$$(x_1 \dots x_m)^n = x_1^n \dots x_m^n. \quad (4)$$

Аналогично, если $x + y = y + x$ и $x_i + x_j = x_j + x_i$ при $i, j = 1, \dots, m$, то

$$n(x + y) = nx + ny, \quad n = 0, 1, 2, \dots, \quad (3')$$

$$n(x_1 + \dots + x_m) = nx_1 + \dots + nx_m, \quad n = 0, 1, 2, \dots \quad (4')$$

Обычно моноид (M, \cdot, e) называют *мультипликативным*, а $(M, +, 0)$ — *аддитивным*. Аддитивная запись используется преимущественно в коммутативных моноидах.

4. Обратимые элементы. Элемент a моноида (M, \cdot, e) называется *обратимым*, если найдётся элемент $b \in M$, для которого $ab = e = ba$ (понятно, что элемент b тоже будет обратимым). Если ещё и $ab' = e = b'a$, то $b' = eb' = (ba)b' = b(ab') = be = b$. Это даёт нам основание говорить просто об *обратном элементе* a^{-1} к (обратимому) элементу $a \in M$: $a^{-1}a = e = aa^{-1}$.

Разумеется, $(a^{-1})^{-1} = a$. Понятие обратимого элемента моноида служит, очевидно, естественным обобщением понятия обратимой матрицы в мультипликативном моноиде $(M_n(\mathbb{R}), \cdot, E)$.

Так как $(xy)(y^{-1}x^{-1}) = x(yy^{-1})x^{-1} = xx^{-1} = e$ и, аналогично, $(y^{-1}x^{-1})(xy) = e$, то $(xy)^{-1} = y^{-1}x^{-1}$. Стало быть, множество всех обратимых элементов моноида (M, \cdot, e) замкнуто относительно операции и составляет подмоноид в M .

УПРАЖНЕНИЯ

1. В п. 2 в качестве примера на \mathbb{Z} вводилась операция $*: n * m = -n - m$, коммутативная, но неассоциативная. В алгебраической структуре $(\mathbb{Z}, *)$ выполняются соотношения $(n * m) * m = n$, $m * (m * n) = n$. Пусть теперь нам дана произвольная алгебраическая структура $(X, *)$, в которой $(x * y) * y = x$, $y * (y * x) = x$ для любых $x, y \in X$. Доказать, что $x * y = y * x$, т.е. операция $*$ коммутативна. Никаких указаний к решению не даётся, поскольку это одно из самых бесполезных упражнений в книге. Но все-таки!

2. Показать, что множество

$$M_n^0(\mathbb{R}) = \left\{ A = (a_{ij}) \in M_n(\mathbb{R}) \mid \sum_{j=1}^n a_{ij} = 0, \quad i = 1, 2, \dots, n \right\}$$

с обычной операцией умножения матриц является полугруппой. Является ли $(M_n^0(\mathbb{R}), \cdot)$ моноидом?

3. В мультиликативном моноиде M выбирается произвольный элемент t и вводится новая операция $*: x * y = xt y$. Показать, что $(M, *)$ — полугруппа и что обратимость элемента t в M — необходимое и достаточное условие, при выполнении которого $(M, *)$ — моноид с нейтральным (единичным) элементом t^{-1} .

4. Показать, что множество \mathbb{Z} с операцией $\circ: n \circ m = n + m + nm = (1+n) \times (1+m) - 1$, является коммутативным моноидом. Что служит в (\mathbb{Z}, \circ) нейтральным элементом? Найти в (\mathbb{Z}, \circ) все обратимые элементы.

§ 2. Группы

1. Определение и примеры. Рассмотрим множество $GL_n(\mathbb{R})$ всех $n \times n$ -матриц с вещественными коэффициентами и с отличным от нуля определителем. Согласно теореме 3 из § 2 гл. 3 $\det A \neq 0$, $\det B \neq 0 \Rightarrow \det AB \neq 0$. Мы видим, что $A, B \in GL_n(\mathbb{R}) \Rightarrow AB \in GL_n(\mathbb{R})$. Далее, $(AB)C = A(BC)$ и существует выделенная матрица E такая, что $AE = EA = A$ для всех $A \in GL_n(\mathbb{R})$. Кроме того, у каждой матрицы $A \in GL_n(\mathbb{R})$ имеется “антисимметрическая” — обратная матрица A^{-1} , для которой $AA^{-1} = A^{-1}A = E$.

Множество $GL_n(\mathbb{R})$, рассматриваемое вместе с законом композиции (бинарной операцией) $(A, B) \mapsto AB$ и называемое полной линейной группой степени n над \mathbb{R} , можно было бы коротко определить, следя терминологии § 1, как подмоноид всех обратимых элементов моноида $(M_n(\mathbb{R}), \cdot, E)$. Но этот подмоноид настолько важен, что он заслуживает специального названия и даёт веский повод ввести общее

Определение. Моноид G , все элементы которого обратимы, называется группой. Другими словами, предполагаются выполненные следующие аксиомы.

G0) На множестве G определена бинарная операция $(x, y) \mapsto xy$.

G1) Операция ассоциативна: $(xy)z = x(yz)$ для всех $x, y, z \in G$.

G2) G обладает нейтральным (единичным) элементом e : $xe = ex = x$ для всех $x \in G$.

G3) Для каждого элемента $x \in G$ существует обратный x^{-1} :
 $xx^{-1} = x^{-1}x = e$.

Мы видели в § 8 гл. 1, что указанным аксиомам удовлетворяет алгебраическая система S_n , названная нами симметрической группой перестановок степени n . Фактически этим важнейшим примером мы предварили общее определение группы.

Удивительно, что одна из старейших и богатейших по результатам область алгебры, играющая фундаментальную роль в геометрии и в приложениях математики к вопросам естествознания, основывается на столь простых аксиомах. Небольшой анализ показывает, что их можно ещё упростить, но эта задача для нас не принципиальна.

Группа с коммутативной операцией называется, естественно, коммутативной, а ещё чаще — абелевой (в честь норвежского математика Абеля). Сам термин “группа” принадлежит французскому математику Галуа — подлинному создателю теории групп. Идеи теории групп “носились в воздухе” (как это часто бывает с основополагающими математическими идеями) задолго до Галуа, и некоторые из её теорем в наивной форме были доказаны еще Лагранжем. Гениальные работы Галуа оказались непонятными, и возрождение интереса к ним началось только после книги К. Жордана “Курс теории перестановок и алгебраических уравнений” (1870 г.). Лишь к концу XIX века в теории групп “совершенно отказываются от фантазии. Взамен этого тщательно препарируется логический скелет” (Ф. Клейн, “Лекции о развитии математики в XIX столетии”).

Для обозначения числа элементов в группе G (точнее, мощности группы) используются равноправные символы $\text{Card } G$, $|G|$ и $(G : e)$. Почти всё сказанное в § 1 о моноидах переносится на группы. Следует лишь производить надлежащую замену слов. В частности, подмножество $H \subset G$ называется подгруппой в G , если $e \in H$; $h_1, h_2 \in H \Rightarrow h_1h_2 \in H$ и $h \in H \Rightarrow h^{-1} \in H$. Подгруппа $H \subset G$ собственная, если $H \neq \{e\}$ и $H \neq G$.

Приведём ещё несколько примеров групп.

Пример 1. В уже известной нам полной линейной группе $\text{GL}_n(\mathbb{R})$ рассмотрим подмножество $\text{SL}_n(\mathbb{R})$ матриц с определителем 1:

$$\text{SL}_n(\mathbb{R}) = \{A \in \text{GL}_n(\mathbb{R}) \mid \det A = 1\}.$$

Очевидно, что $E \in \text{SL}_n(\mathbb{R})$. Согласно общим результатам гл. 3 об определителях

$$\det A = 1, \det B = 1 \Rightarrow \det AB = 1$$

$$\det A^{-1} = (\det A)^{-1} = 1.$$

Поэтому $\text{SL}_n(\mathbb{R})$ — подгруппа в $\text{GL}_n(\mathbb{R})$; она носит название специальной линейной группы степени n над \mathbb{R} . Её называют ещё и унимодуллярной группой, хотя к последней часто причисляют матрицы с определителем ± 1 .

Надо сказать, что группа $GL_n(\mathbb{R})$, будучи вместилищем многих интересных групп, является для математиков разных поколений как бы нескончаемым источником новых идей и нерешённых задач.

Пример 2. Используя рациональные числа вместо вещественных, мы придём к полной линейной группе $GL_n(\mathbb{Q})$ степени n над \mathbb{Q} и к её подгруппе $SL_n(\mathbb{Q})$. В свою очередь $SL_n(\mathbb{Q})$ содержит интересную подгруппу $SL_n(\mathbb{Z})$ целочисленных матриц с определителем 1. Теорема 1 § 3 гл. 3, предлагающая явную формулу для коэффициентов обратной матрицы, показывает, что $SL_n(\mathbb{Z})$ действительно является группой. Группы $SL_n(\mathbb{Q})$ и $SL_n(\mathbb{Z})$ занимают почетное место в теории чисел. Частично упорядоченное множество (см. п. 4 § 6 гл. 1) рассмотренных подгрупп группы $GL_n(\mathbb{R})$ изображается помещённой здесь диаграммой (рис. 15).

Пример 3. Положив в примерах 1 и 2 $n = 1$, мы придём, во-первых, к мультипликативным группам

$$\mathbb{R}^* = \mathbb{R} \setminus \{0\} = GL_1(\mathbb{R}), \quad \mathbb{Q}^* = \mathbb{Q} \setminus \{0\} = GL_1(\mathbb{Q})$$

вещественных и рациональных чисел. Эти группы, очевидно, бесконечны. Так как в $(\mathbb{Z}, \cdot, 1)$ обратимыми элементами являются только 1 и -1 , то $GL_1(\mathbb{Z}) = \{\pm 1\}$. Далее, $SL_1(\mathbb{R}) = SL_1(\mathbb{Q}) = SL_1(\mathbb{Z}) = 1$. Но уже при $n = 2$ группа $SL_2(\mathbb{Z})$ бесконечна: ей принадлежат, например, все матрицы

$$\begin{vmatrix} 1 & m \\ 0 & 1 \end{vmatrix}, \quad \begin{vmatrix} 1 & 0 \\ m & 1 \end{vmatrix}, \quad \begin{vmatrix} m & m-1 \\ 1 & 1 \end{vmatrix}, \quad m \in \mathbb{Z}.$$

Отметим ещё бесконечные аддитивные группы:

$$(\mathbb{R}, +, 0), \quad (\mathbb{Q}, +, 0), \quad (\mathbb{Z}, +, 0).$$

Пример 4. Пусть Ω — произвольное множество, а $S(\Omega)$ — множество всех биективных (взаимно однозначных) преобразований $f : \Omega \rightarrow \Omega$. Обратившись к результатам § 5 гл. 1 об отображениях множеств (теоремы 1, 2 и следствие теоремы 2), мы немедленно делаем заключение, что $S(\Omega)$ — группа с естественной бинарной операцией, являющейся композицией преобразований. Разумеется, $S(\Omega)$ — подмоноид всех обратимых элементов моноида $M(\Omega)$ из примера 1 § 1, но это обстоятельство мы не склонны подчёркивать. Сама по себе группа $S(\Omega)$ и в особенности различные её подгруппы, называемые группами преобразований, — стартовая площадка, с которой начинаются всевозможные применения теории групп. Достаточно упомянуть о знаменитой «Эрлангенской программе» Ф. Клейна (1872 г.), положившей понятие группы преобразований в основу классификации различных типов геометрий (более подробно см. по этому поводу [ВА II]).

Взяв за Ω линейное пространство \mathbb{R}^n , мы придём к «большой» и малоизученной группе $S(\mathbb{R}^n)$. Но в $S(\mathbb{R}^n)$ содержится подгруппа обратимых (биективных) линейных преобразований $\varphi_A : \mathbb{R}^n \rightarrow \mathbb{R}^n$, находящихся во взаимно однозначном соответствии с невырожденными матрицами A порядка n (см. § 3 гл. 2).

Таким образом, получается вложение $GL_n(\mathbb{R})$ в $S(\mathbb{R}^n)$.

Смысль этого вложения станет яснее, когда будет введено важное понятие изоморфизма групп.

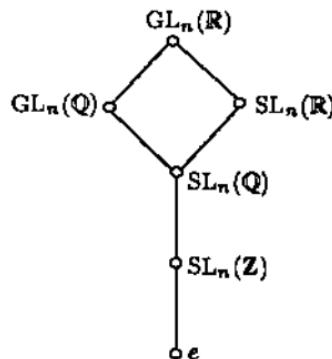


Рис. 15

2. Циклические группы. Пусть G — мультиликативная группа (т.е. с операцией умножения), a — её фиксированный элемент. Если любой элемент $g \in G$ записывается в виде $g = a^n$ для некоторого $n \in \mathbb{Z}$, то говорят, что $G = \langle a \rangle$ — циклическая группа с образующим a (или циклическая группа, порождённая элементом a). Аналогично циклическая группа определяется в аддитивном случае: $\langle a \rangle = \{na \mid n \in \mathbb{Z}\}$. Это, конечно, не означает, что все элементы a^n или na попарно различны. Условимся обозначать $(a^{-1})^k = a^{-k}$ и убедимся в справедливости следующего утверждения.

Теорема 1. *Каковы бы ни были $m, n \in \mathbb{Z}$,*

$$a^m a^n = a^{m+n}, \quad (a^m)^n = a^{mn}$$

(соответственно $ma + na = (m+n)a$, $n(ma) = (nm)a$).

Доказательство. При неотрицательных m, n см. соотношения (2), (2') из п. 3 § 1. Если $m < 0$, $n < 0$, то $m' = -m > 0$, $n' = -n > 0$ и

$$a^m a^n = (a^{-1})^{m'} (a^{-1})^{n'} = (a^{-1})^{m'+n'} = a^{-(m'+n')} = a^{m+n}.$$

При $m' = -m > 0$, $n > 0$ имеем

$$a^m a^n = (a^{-1})^{m'} a^n = (\underbrace{a^{-1} \dots a^{-1}}_{m'})(\underbrace{a \dots a}_n) = a^{n-m'} = a^{m+n}.$$

Если $m' \geq n$, то $a^{n-m'} = (a^{-1})^{m'-n} = a^{m+n}$.

Аналогично рассматривается случай $m > 0$, $n < 0$. Равенство $(a^m)^n = a^{mn}$ вытекает из предыдущего и достаточно очевидно из определения степеней. \square

Простейшим примером циклической группы служит аддитивная группа целых чисел $(\mathbb{Z}, +, 0)$, порождённая обычной единицей 1 или -1 .

Легко проверить, далее, что матрица $\begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix}$ порождает в $\text{SL}_2(\mathbb{Z})$ бесконечную циклическую подгруппу. Множество $\{1, -1\}$ является по умножению циклической группой порядка 2.

Пример циклической группы порядка n получается, если рассмотреть все вращения на плоскости вокруг некоторой точки O , совмещающие с собой правильный n -угольник P_n с центром в точке O . Очевидно, что эти вращения образуют группу: под их произведением следует понимать последовательное выполнение преобразований. Наша группа C_n содержит вращения $\varphi_0, \varphi_1, \dots, \varphi_{n-1}$ против часовой стрелки на углы $0, 2\pi/n, \dots, (n-1)2\pi/n$. При этом $\varphi_a = \varphi_b^t$, а из геометрических соображений ясно, что $\varphi_a^{-1} = \varphi_{n-a}^t$ и $\varphi_a^t = \varphi_0$ (единичное преобразование). Итак, $|C_n| = n$ и $C_n = \langle \varphi_1 \rangle$. Заметим, что циклическая группа C_n является собственной подгруппой группы D_n всех преобразований симметрии n -угольника P_n (т.е. совмещений P_n с собой).

Пусть снова G — произвольная группа, a — некоторый её элемент. Имеются две возможности.

1) Все степени элемента a различны, т.е. $m \neq n \Rightarrow a^m \neq a^n$. В этом случае говорят, что элемент $a \in G$ имеет бесконечный порядок.

2) Имеются совпадения $a^m = a^n$ при $m \neq n$. Если, например, $m > n$, то $a^{m-n} = e$, т.е. существуют положительные степени элемента $a \in G$, равные единичному элементу. Пусть q — наименьший положительный показатель, для которого $a^q = e$. Тогда говорят, что a — элемент *конечного порядка* q .

В конечной группе G ($\text{Card } G < \infty$) все элементы, разумеется, будут конечного порядка.

Предостережение. Слово “порядок” в математике многозначно. Мы говорили раньше о квадратных матрицах порядка n (матрица размера $n \times n$), но невырожденная матрица A , рассматриваемая как элемент группы $\text{GL}_n(\mathbb{R})$, имеет также порядок (возможно, бесконечный) в только что указанном смысле. Каждый раз будет ясно из контекста, о чём идёт речь.

На фоне приведённого выше примера циклической группы порядка n следующее утверждение почти очевидно.

Теорема 2. *Порядок любого элемента $a \in G$ (G — абстрактная группа) равен $\text{Card } \langle a \rangle$.*

Если a — элемент конечного порядка q , то $\langle a \rangle = \{e, a, \dots, a^{q-1}\}$,

$$a^k = e \iff k = lq, \quad l \in \mathbb{Z}.$$

Доказательство. В случае элемента бесконечного порядка доказывать нечего. Если a — элемент порядка q , то по определению все элементы $e, a, a^2, \dots, a^{q-1}$ различны. Любая другая степень a^k совпадает с одним из этих элементов, т.е. $\langle a \rangle = \{e, a, \dots, a^{q-1}\}$. В самом деле, воспользовавшись алгоритмом деления в \mathbb{Z} (п. 3 § 9 гл. 1), запишем показатель k в виде

$$k = lq + r, \quad 0 \leq r \leq q - 1,$$

после чего, оперируя со степенями по правилам, изложенными в теореме 1, получим

$$a^k = (a^q)^l a^r = e a^r = a^r.$$

В частности, $a^k = e \implies r = 0 \implies k = lq$. \square

3. Изоморфизмы. Как уже отмечалось ранее, три вращения $\varphi_0, \varphi_1, \varphi_2$ против часовой стрелки на углы $0^\circ, 120^\circ, 240^\circ$ соответственно переводят правильный треугольник P_3 в себя. Но имеются ещё три осевые преобразования симметрии (отражения) ψ_1, ψ_2, ψ_3 с указанными на рис. 16 осями симметрии $1-1'$, $2-2'$, $3-3'$. Всем шести преобразованиям симметрии соответствуют перестановки на множество вершин треугольника. Мы

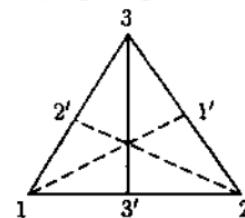


Рис. 16

получаем

$$\begin{array}{lll} \varphi_0 \sim e, & \varphi_1 \sim (1\ 2\ 3), & \varphi_2 \sim (1\ 3\ 2), \\ \psi_1 \sim (2\ 3), & \psi_2 \sim (1\ 3), & \psi_3 \sim (1\ 2). \end{array}$$

Так как других перестановок степени 3 нет, то можно утверждать, что группа D_3 всех преобразований симметрии правильного треугольника обнаруживает большое сходство с симметрической группой S_3 .

В том же смысле близки друг к другу циклические группы C_n (см. пример в п. 2) и $\{(1\ 2\ \dots\ n)\} \subset S_n$. Эти факты, а также общие размышления о группах не могут не приводить к весьма естественному вопросу о наиболее существенных свойствах групп. На первый взгляд, полная информация содержится в таблице умножения группы G , называемой *таблицей Кэли*:

	g_1	g_2	\dots	g_n	\dots
g_1	g_1g_1	g_1g_2	\dots	g_1g_n	\dots
g_2	g_2g_1	g_2g_2	\dots	g_2g_n	\dots
\dots	\dots	\dots	\dots	\dots	\dots
g_n	g_ng_1	g_ng_2	\dots	g_ng_n	\dots
\dots	\dots	\dots	\dots	\dots	\dots

Действительно, многие закономерности группы можно уловить из рассмотрения её таблицы Кэли или, что то же самое, матрицы $M = (m_{ij})$ (размера $n \times n$, если $n = (G : e)$) с элементами $m_{ij} = g_i g_j \in G$. Мы замечаем, например, что среди элементов каждой строки и каждого столбца матрицы M любой элемент группы G встречается ровно один раз (см. ниже доказательство теоремы 4). Группа G абелева тогда и только тогда, когда матрица M симметрическая, т.е. $m_{ij} = m_{ji}$. Этот список свойств можно было бы продолжить, но всё-таки сравнивать две таблицы для групп G, G' одинакового порядка довольно затруднительно, потому что вид матрицы M зависит от нумерации (расположения) элементов группы, а уж в случае бесконечных групп ситуация ещё более усложняется.

Самый правильный и самый радикальный подход к различению (или, напротив, к отождествлению) групп G и G' предлагает понятие изоморфизма.

Определение. Две группы G и G' с операциями $*$ и \circ называются *изоморфными*, если существует отображение $f: G \rightarrow G'$ такое, что:

- i) $f(a * b) = f(a) \circ f(b)$ для всех $a, b \in G$;
- ii) f биективно.

Факт изоморфизма групп часто обозначается символически $G \cong G'$.

Отметим простейшие свойства изоморфизма.

1) *Единица переходит в единицу.* Действительно, если e — единица группы G , то $e * a = a * e = a$, и, значит, $f(e) \circ f(a) = f(a) \circ f(e) = f(a)$, откуда следует, что $f(e) = e'$ — единица группы G' . В этом рассуждении использованы, хотя и частично, оба свойства f . Для i) это очевидно, а свойство ii) обеспечивает сюръективность f , так что элементами $f(g)$ исчерпываются вся группа G' .

2) $f(a^{-1}) = f(a)^{-1}$. В самом деле, согласно 1) $f(a) \circ f(a^{-1}) = f(a * a^{-1}) = f(e) = e'$ — единица в G' , откуда

$$\begin{aligned} f(a)^{-1} &= f(a)^{-1} \circ e' = f(a)^{-1} \circ (f(a) \circ f(a^{-1})) = \\ &= (f(a)^{-1} \circ f(a)) \circ f(a^{-1}) = e' \circ f(a^{-1}) = f(a^{-1}). \end{aligned}$$

3) *Обратное отображение* $f^{-1}: G' \rightarrow G$ (существующее в силу свойства ii)) тоже является изоморфизмом. В силу следствия теоремы 2 § 5 гл. 1 надо убедиться лишь в справедливости свойства i) для f^{-1} . Пусть $a', b' \in G'$. Тогда ввиду биективности f имеем $a' = f(a), b' = f(b)$ для каких-то $a, b \in G$. Поскольку f — изоморфизм, $a' \circ b' = f(a) \circ f(b) = f(a * b)$. Отсюда имеем $a * b = f^{-1}(a' \circ b')$, а так как в свою очередь $a = f^{-1}(a'), b = f^{-1}(b')$, то $f^{-1}(a' \circ b') = f^{-1}(a') * f^{-1}(b')$.

Замечание. Несложная проверка показывает, что установленное нами соответствие \sim между группами D_3 и S_3 является на самом деле изоморфизмом.

В качестве изоморфного отображения f мультиликативной группы (\mathbb{R}_+, \cdot) положительных вещественных чисел на аддитивную группу $(\mathbb{R}, +)$ всех вещественных чисел может служить $f := \ln$. Известное свойство логарифма $\ln ab = \ln a + \ln b$ как раз моделирует свойство i) в определении изоморфизма. Обратным к f служит отображение $x \mapsto e^x$.

Докажем теперь две общие теоремы, иллюстрирующие роль изоморфизма в теории групп.

Теорема 3. Все циклические группы одного и того же порядка (в том числе и бесконечного) изоморфны.

Доказательство. В самом деле, если $\langle g \rangle$ — бесконечная циклическая группа, то все степени g^n образующего g различны, и мы получим изоморфизм $f: \langle g \rangle \rightarrow (\mathbb{Z}, +)$, полагая $g^n \mapsto f(g^n) = n$. Биективность f очевидна, а свойство $f(g^m g^n) = f(g^n) + f(g^m)$ вытекает из теоремы 1.

Пусть теперь $G = \{e, g, \dots, g^{q-1}\}$ и $G' = \{e', g', \dots, (g')^{q-1}\}$ — две циклические группы порядка q (операции в G и G' не различаем). Определим биективное отображение

$$f: g^k \mapsto (g')^k, \quad k = 0, 1, \dots, q-1.$$

Полагая $n+m = lq+r$, $0 \leq r \leq q-1$, для любых $n, m = 0, 1, \dots, q-1$

и рассуждая как при доказательстве теоремы 2, будем иметь

$$f(g^{n+m}) = f(g^r) = (g')^r = (g')^{n+m} = (g')^n(g')^m = f(g^n)f(g^m). \quad \square$$

Теорема 4 (Кэли). Любая конечная группа порядка n изоморфна некоторой подгруппе симметрической группы S_n .

Доказательство. Пусть G — наша группа, $n = |G|$. Можно считать, что S_n — группа всех биективных отображений множества G на себя, так как природа элементов, переставляемых элементами из S_n , несущественна.

Для любого элемента $a \in G$ рассмотрим отображение $L_a : G \rightarrow G$, определённое формулой

$$L_a(g) = ag$$

(очевидно, мы повторяем определение из п. 3 § 8 гл. 1). Если $e = g_1, g_2, \dots, g_n$ — все элементы группы G , то a, ag_2, \dots, ag_n будут теми же элементами, но расположеными в каком-то другом порядке (вспомним таблицу Кэли). Это и понятно, поскольку

$$ag_i = ag_j \implies a^{-1}(ag_i) = a^{-1}(ag_j) \implies (a^{-1}a)g_i = (a^{-1}a)g_j \implies g_i = g_j.$$

Значит, L_a — биективное отображение (перестановка), обратным к которому будет $L_a^{-1} = L_{a^{-1}}$. Единичным отображением является, естественно, L_e .

Используя вновь ассоциативность умножения в G , получаем $L_{ab}(g) = (ab)g = a(bg) = L_a(L_bg)$, т.е. $L_{ab} = L_aL_b$.

Итак, множество $L_e, L_{g_2}, \dots, L_{g_n}$ образует подгруппу, скажем, H , в группе $S(G)$ всех биективных отображений множества G на себя, т.е. в S_n . Мы имеем включение $H \subset S_n$ и имеем соответствие $L : a \mapsto L_a \in H$, обладающее по сказанному выше всеми свойствами изоморфизма. \square

Теорема Кэли, несмотря на свою простоту, имеет важное значение в теории групп. Она выделяет некий универсальный объект (семейство $\{S_n \mid n = 1, 2, \dots\}$ симметрических групп) — вместилище всех вообще конечных групп, рассматриваемых с точностью до изоморфизма. Фраза “с точностью до изоморфизма” отражает сущность не только теории групп, стремящейся объединить в один класс все изоморфные группы, но и математики в целом, которая без таких обобщений была бы лишена смысла.

Положив $G' = G$ в определении изоморфизма, мы получим изоморфное отображение $\varphi : G \rightarrow G$ группы G на себя. Оно называется *автоморфизмом* группы G . Например, единичное отображение $e_G : g \mapsto g$ (далее обозначаемое просто через 1) — автоморфизм, но, как правило, G обладает и нетривиальными автоморфизмами. Свойство 3) изоморфных отображений показывает, что отображение, обратное к автоморфизму, тоже будет автоморфизмом. Если, далее, φ, ψ — автоморфизмы группы G , то $(\varphi \circ \psi)(ab) = \varphi(\psi(ab)) =$

$= \varphi(\psi(a)\psi(b)) = (\varphi \circ \psi)(a) \cdot (\varphi \circ \psi)(b)$ для любых $a, b \in G$. Стало быть, множество $\text{Aut}(G)$ всех автоморфизмов группы G образует группу — подгруппу группы $S(G)$ всех биективных отображений $G \rightarrow G$.

4. Гомоморфизмы. В группе автоморфизмов $\text{Aut}(G)$ группы G содержится одна особая подгруппа. Она обозначается символом $\text{Inn}(G)$ и называется группой внутренних автоморфизмов. Её элементами являются отображения

$$I_a : g \mapsto aga^{-1}.$$

Небольшое упражнение показывает, что I_a действительно удовлетворяет всем свойствам, требуемым от автоморфизмов, причём $I_a^{-1} = I_{a^{-1}}$, $I_e = 1$ — единичный автоморфизм, $I_a \circ I_b = I_{ab}$ (так как $(I_a \circ I_b)(g) = I_a(I_b(g)) = I_a(bgb^{-1}) = abgb^{-1}a^{-1} = abg(ab)^{-1} = I_{ab}(g)$).

Последнее соотношение показывает, что отображение

$$f : G \rightarrow \text{Inn}(G)$$

группы G на группу $\text{Inn}(G)$ ее внутренних автоморфизмов, определённое формулой $f(a) = I_a$, $a \in G$, обладает свойством i) изоморфного отображения: $f(a) \circ f(b) = f(ab)$. Однако свойство ii) при этом не обязательно выполняется. Если, например, G — абелева группа, то $aga^{-1} = g$ для всех $a, g \in G$, так что $I_a = I_e$, и вся группа $\text{Inn}(G)$ состоит из одного единичного элемента I_e . Это обстоятельство делает естественным следующее общее

Определение. Отображение $f : G \rightarrow G'$ группы $(G, *)$ в (G', \circ) называется гомоморфизмом, если

$$\forall a, b \in G \quad f(a * b) = f(a) \circ f(b)$$

(другими словами, выполняется только свойство i) из определения изоморфизма).

Ядром гомоморфизма f называется множество

$$\text{Ker } f = \{g \in G \mid f(g) = e' \text{ — единица группы } G'\}.$$

Гомоморфное отображение группы в себя называется ещё её эндоморфизмом.

В этом определении от f не требуется не только биективности, но и сюръективности (т.е. быть отображением “на”), что, впрочем, не очень существенно, поскольку всегда можно ограничиться рассмотрением образа $\text{Im } f \subset G'$, являющегося, очевидно, подгруппой в G' . Главное отличие гомоморфизма f от изоморфизма заключается в наличии нетривиального ядра $\text{Ker } f$, являющегося, так сказать, мерой неинъективности f . Если же $\text{Ker } f = \{e\}$, то $f : G \rightarrow \text{Im } f$ — изоморфизм.

Заметим, что

$$\begin{aligned} f(a) = e', f(b) = e' \implies f(a * b) &= f(a) \circ f(b) = e' \circ e' = e', \\ f(a^{-1}) &= f(a)^{-1} = (e')^{-1} = e'. \end{aligned}$$

Поэтому ядро $\text{Ker } f$ — подгруппа в G .

5. Словарик. Примеры. Стоит отметить, что термины *сюръективное отображение* (отображение “на”), *инъективное* (отображение вложение), *биективное* (взаимно однозначное отображение), применимые к отображениям любых множеств (без операций), в случае групп (и в случае других алгебраических структур) заменяются соответственно терминами *эпиморфизм* (гомоморфизм “на”), *мономорфизм* (гомоморфизм с единичным ядром), *изоморфизм* (взаимно однозначный гомоморфизм — эпиморфизм и мономорфизм одновременно). Имеется тенденция к замене гомоморфизма термином *морфизм*. Этот словарик полезно иметь в виду при чтении математической литературы, но на первых порах желающие могут обойтись двумя терминами: изоморфизм и гомоморфизм с добавлениями “в” и “на”.

В дополнение к рассмотренным выше приведём ещё несколько примеров морфизмов групп.

Пример 5. Аддитивная группа целых чисел \mathbb{Z} гомоморфно отображается на конечную циклическую группу $\langle g \rangle$ порядка q , если положить $f: n \mapsto g^n$ (см. теорему 2 § 2). В этом случае, очевидно, $\text{Ker } f = \{lq \mid l \in \mathbb{Z}\}$. В самом деле, ясно, что $\{lq\} \subset \text{Ker } f$. Обратное включение следует из теоремы 1.

Пример 6. Отображение $f: \mathbb{R} \rightarrow T = SO(2)$ аддитивной группы вещественных чисел на группу T вращений плоскости с неподвижной точкой 0, задаваемое формулой $f(\lambda) = \Phi_\lambda$ (Φ_λ — вращение против часовой стрелки на угол $2\pi\lambda$), гомоморфно, так как $\Phi_\lambda \circ \Phi_\mu = \Phi_{\lambda+\mu}$. Вращение на угол, целочисленно кратный 2π , совпадает с единичным вращением (на нулевой угол), поэтому $\text{Ker } f = \mathbb{Z}$. Говорят также, что f — гомоморфизм \mathbb{R} на окружность S^1 единичного радиуса, поскольку имеется взаимно однозначное соответствие между Φ_λ и точкой на S^1 с полярными координатами $(1, 2\pi\lambda)$, $0 \leq \lambda < 1$.

Пример 7. Полная линейная группа $GL_m(\mathbb{R})$ вещественных матриц A (т.е. матриц с коэффициентами в \mathbb{R} с не равным нулю определителем $\det A$) гомоморфно отображается на мультиликативную группу \mathbb{R}^* отличных от нуля вещественных чисел, если положить $f := \det$. Условие гомоморфизма $f(AB) = f(A)f(B)$ — лишь иная формулировка теоремы 3 § 2 гл. 3. По определению $SL_m(\mathbb{R}) = \text{Ker } f$.

Пример 8. Рассмотрим циклическую группу $C_2 = \langle -1 \rangle = \{1, -1\}$ порядка 2. Если угодно, её можно задать абстрактно таблицей Кэли:

	·	1	-1
1	1	1	-1
-1	-1	-1	1

Отображение $S_n \rightarrow C_2$ при помощи известной нам функции $\epsilon = \text{sgn}: \pi \mapsto \epsilon_\pi$ (знак перестановки π) является гомоморфизмом симметрической группы S_n на C_2 . Ядро $\text{Ker } \epsilon = A_n$ порядка $n!/2$ (см. п. 3 § 8 гл. 1) называется *знакопеременной группой*.

Пример 9. Бесконечная группа может быть изоморфна своей истинной (собственной) подгруппе. В самом деле, аддитивная группа $(\mathbb{Z}, +)$ содержит собственную подгруппу $n\mathbb{Z} = \{nk \mid k \in \mathbb{Z}\}$, где $n > 1$ — фиксированное натуральное число. Легко проверяется, что отображение $g_n : \mathbb{Z} \rightarrow n\mathbb{Z}$, определённое соотношением $g_n(k) = nk$, является изоморфизмом. Попутно заметим, что \mathbb{Z} и $n\mathbb{Z}$ — бесконечные циклические группы, в которых образующими служит соответственно 1 или -1 и n или $-n$; поэтому g_n и отображение $k \mapsto -nk$ исчерпывают все изоморфизмы $\mathbb{Z} \rightarrow n\mathbb{Z}$.

Пример 10. Группа $\text{Aut}(G)$ и даже отдельный неединичный элемент $\varphi \in \text{Aut}(G)$ могут служить источником важных сведений о группе G . Вот яркий пример такого рода. Пусть G — конечная группа, на которой действует автоморфизм φ порядка 2 ($\varphi^2 = 1$) без неподвижных точек:

$$a \neq e \implies \varphi(a) \neq a.$$

Предположим, что $\varphi(a)a^{-1} = \varphi(b)b^{-1}$ для каких-то $a, b \in G$. Тогда после умножения этого равенства слева на $\varphi(b)^{-1}$ и справа на a получим $\varphi(b)^{-1}\varphi(a) = b^{-1}a$, т.е. $\varphi(b^{-1}a) = b^{-1}a$, откуда $b^{-1}a = e$ и $b = a$. Итак, $\varphi(a)a^{-1}$ пробегает вместе с a все элементы группы G , или, что равносильно, любой элемент $g \in G$ записывается в виде $g = \varphi(a)a^{-1}$. Но в таком случае $\varphi(g) = \varphi(\varphi(a))\varphi(a^{-1}) = \varphi^2(a)\varphi(a^{-1}) = a\varphi(a)^{-1} = (\varphi(a)a^{-1})^{-1} = g^{-1}$. Итак, φ совпадает с отображением $g \mapsto g^{-1}$. Зная это, получаем $ab = \varphi(a^{-1})\varphi(b^{-1}) = \varphi(a^{-1}b^{-1}) = (a^{-1}b^{-1})^{-1} = ba$, т.е. группа G оказывается абелевой. Кроме того, $(G : e)$ — нечётное число, ибо G состоит из e и непересекающихся пар элементов $g_i, g_i^{-1} = \varphi(g_i)$.

Пример 11. Насколько можно изменить операцию на группе, не меняя в смысле изоморфизма самой группы, показывает следующий пример (см. также упр. 3 из § 1). Пусть G — произвольная группа, t — её какой-то фиксированный элемент. Введём на множестве G новую операцию

$$(g, h) \mapsto g * h = gth.$$

Непосредственно проверяется, что $(g_1 * g_2) * g_3 = g_1 * (g_2 * g_3)$, т.е. операция $*$ ассоциативна. Кроме того, $g * t^{-1} = t^{-1} * g = g$ и $g * (t^{-1}g^{-1}t^{-1}) = (t^{-1}g^{-1}t^{-1}) * g = t^{-1}$, а это значит, что $(G, *)$ — группа с единичным элементом $e_* = t^{-1}$. Элементом, обратным к g в $(G, *)$, служит $g_*^{-1} = t^{-1}g^{-1}t^{-1}$. Отображение $f : g \mapsto gt^{-1}$ устанавливает изоморфизм групп (G, \cdot) и $(G, *)$, т.е. $f(gh) = f(g)*f(h)$.

Все указанные примеры служат, между прочим, иллюстрацией к одному общему правилу: изучение морфизмов группы G даёт значительную информацию о самой группе G .

УПРАЖНЕНИЯ

1. Доказать, что пересечение $\bigcap_{i \in I} H_i$ любого семейства $\{H_i \mid i \in I\}$ подгруппы группы G является подгруппой.

2. Говорят, что группа G порождается подмножеством S своих элементов, и пишут $G = \langle S \rangle$, если пересечение всех подгрупп H , содержащих S , совпадает с G (другими словами, в G нет хотя бы одной собственной подгруппы, содержащей S). Показать, что в случае $G = \langle S \rangle$ каждый элемент $g \in G$ имеет вид $g = t_1 t_2 \dots t_n$, $n = 1, 2, \dots$, где либо $t_i \in S$, либо $t_i^{-1} \in S$, $1 \leq i \leq n$.

3. Показать, что перестановочные элементы a, b произвольной группы G , имеющие взаимно простые порядки s, t , порождают в G циклическую подгруппу порядка st : $\langle a, b \rangle = \langle ab \rangle$.

Указание. Включение $\langle ab \rangle \subset \langle a, b \rangle = \{a^i b^j \mid 0 \leq i \leq s-1, 0 \leq j \leq t-1\}$ очевидно. Вместе с тем, согласно п. 3 § 9 гл. 1, из $\text{НОД}(s, t) = 1$ следует, что $tk + sl = 1$ для некоторых $k, l \in \mathbb{Z}$. Поэтому с учётом теоремы 1 $a = a^{1-sl} = a^{tk} = a^{tk}b^{tk} = (ab)^{tk} \in \langle ab \rangle$. Аналогично, $b \in \langle ab \rangle$, и, стало быть, $\langle a, b \rangle \in \langle ab \rangle$.

4. Показать, что если $M = \langle S \rangle$ — моноид, порождённый множеством S , и каждый элемент $s \in S$ обратим в M , то M — группа.

5. Группа — это моноид G , в котором уравнения вида $ax = b$, $ya = b$ однозначно разрешимы при любых $a, b \in G$. Доказать это утверждение.

6. Показать, что множество $A_1(\mathbb{R})$ так называемых *аффинных преобразований* $\varphi_{a,b}: x \mapsto ax + b$ ($a, b \in \mathbb{R}; a \neq 0$) вещественной прямой \mathbb{R} образует группу с законом умножения $\varphi_{a,b} \circ \varphi_{c,d} = \varphi_{ac,ad+b}$. В группе $A_1(\mathbb{R})$ содержится подгруппа $GL_1(\mathbb{R})$, оставляющая точку $x = 0$ на месте, и подгруппа "чистых сдвигов" $x \mapsto x + b$.

7. Группа $SL_2(\mathbb{Z})$ содержит элементы $A = \begin{vmatrix} 0 & 1 \\ -1 & 0 \end{vmatrix}$, $B = \begin{vmatrix} 0 & 1 \\ -1 & -1 \end{vmatrix}$ порядков 4 и 3 соответственно. Показать, что $\langle AB \rangle$ — бесконечная циклическая подгруппа в $SL_2(\mathbb{Z})$. Таким образом, произведение двух элементов конечного порядка в группе G не обязано быть элементом конечного порядка. А как обстоит дело в абелевой группе?

8. Доказать, что группа G чётного порядка $|G| = 2n$ обязательно содержит элемент $g \neq e$ порядка 2.

Указание. Рассмотреть разбиение G на пары g, g^{-1} .

9. Доказать, что $S_n = \langle (1\ 2), (1\ 3), \dots, (1\ n) \rangle$.

10. Доказать, что $S_n = \langle (1\ 2), (1\ 2\ 3 \dots n) \rangle$.

11. Доказать, что знакопеременная группа A_n , $n \geq 3$, порождается циклами длины 3, причём на самом деле

$$A_n = \langle (1\ 2\ 3), (1\ 2\ 4), \dots, (1\ 2n) \rangle.$$

12. Доказать, что k -я степень π^k цикла $\pi = (1\ 2\ \dots\ n) \in S_n$ является произведением $d = \text{НОД}(n, k)$ независимых циклов, каждый из которых имеет длину $q = n/d$.

13. Показать, что порядок перестановки $\pi \in S_n$ (порядок циклической подгруппы $\langle \pi \rangle$) равен наименьшему общему кратному длии независимых циклов, входящих в разложение π .

14. Пусть $A, B \in M_n(\mathbb{R})$ и $(AB)^m = E$ для некоторого целого числа m . Верно ли, что $(BA)^m = E$?

15. Доказать, что непустое подмножество H конечной (мультиплекативной) группы G является подгруппой, если H замкнуто относительно умножения. Значит, в данном случае требования существования в H единичного элемента e и обратного h^{-1} для каждого $h \in H$ излишни.

16. Какую систему образующих можно предложить для мультиплекативной группы (\mathbb{Q}_+, \cdot) положительных рациональных чисел?

Указание. Использовать основную теорему арифметики из § 9 гл. 1.

Существует ли в (\mathbb{Q}_+, \cdot) конечная система образующих?

17. Доказать, что с точностью до изоморфизма существует лишь конечное число $\rho(n)$ групп данного порядка n .

Указание. Оценить сверху число различных таблиц Кэли порядка n . Формальные рассуждения с использованием теоремы 4 ограничивают $\rho(n)$ числом $\binom{n!}{n}$ различных подмножеств в S_n из n элементов. На самом деле $\rho(n)$ значительно меньше, но хорошей оценки, приближающейся к точной, пока не найдено.

18. Используя упр. 10, показать, что каждая конечная группа может быть вложена (т.е. для неё существует мономорфизм) в конечную группу с двумя образующими.

19. Попробуйте убедиться, что на диаграмме (рис. 17) изображены все подгруппы знакопеременной группы A_4 . Символом V_4 обозначена так называемая

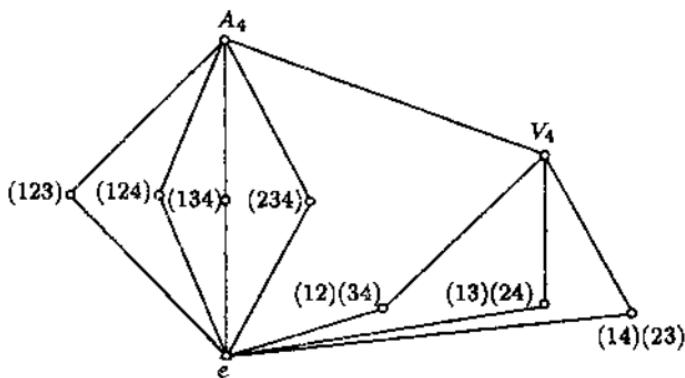


Рис. 17

четвертая группа (или группа Клейна) $V_4 = \{e, (12)(34), (13)(24), (14)(23)\}$, а возле других вершин диаграммы поставлены образующие циклических подгрупп.

20. Показать, что все группы порядка 4 абелевы и с точностью до изоморфизма исчерпываются группами перестановок $U = \langle(1234)\rangle, V_4$, или же группами матриц:

$$L_1 = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \right\} \subset GL_2(\mathbb{R}),$$

$$L_2 = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \right\} \subset GL_2(\mathbb{R}).$$

Выписать в явном виде изоморфизмы $U \rightarrow L_1, V_4 \rightarrow L_2$.

Указание. Если $x^2 = e$ для любого элемента $x \in G$, то $abab = e \implies ab = b^{-1}a^{-1} = b(b^{-1})^2(a^{-1})^2a = beeaa = ba$.

§ 3. Кольца и поля

1. Определение и общие свойства колец. Алгебраические структуры $(\mathbb{Z}, +)$, (\mathbb{Z}, \cdot) выступали у нас в качестве самых первых примеров моноидов, причём на $(\mathbb{Z}, +)$ мы смотрели позднее как на аддитивную абелеву (фактически циклическую) группу. В повседневной жизни, однако, эти структуры чаще всего объединяются и получается то, что в математике называется кольцом. Важная компонента элементарной арифметики заключена в дистрибутивном (или распределительном) законе $(a + b)c = ac + bc$, кажущемся тривиальным только в силу приобретённой привычки. Попытавшись, например, объединить алгебраические структуры $(\mathbb{Z}, +)$, (\mathbb{Z}, \circ) , где $n \circ m = n + m + nm$, мы уже не заметим столь хорошей согласованности между двумя бинарными операциями. Прежде чем переходить к дальнейшим примерам, дадим точное определение кольца.

Определение. Пусть K — непустое множество, на котором заданы две (бинарные алгебраические) операции $+$ (сложение) и \cdot (умножение), удовлетворяющие следующим условиям:

K1) $(K, +)$ — абелева группа;

K2) (K, \cdot) — полугруппа;

K3) операции сложения и умножения связаны дистрибутивными законами (другими словами, умножение дистрибутивно по сложению)

$$(a+b)c = ac + bc, \quad c(a+b) = ca + cb$$

для всех $a, b, c \in K$.

Тогда $(K, +, \cdot)$ называется *кольцом*.

Структура $(K, +)$ называется *аддитивной группой кольца*, а (K, \cdot) — его *мультипликативной полугруппой*. Если (K, \cdot) — моноид, то говорят, что $(K, +, \cdot)$ — *кольцо с единицей*.

Единичный элемент кольца принято обозначать обычной единицей 1. Существование 1 часто вносится в определение кольца, но мы этого делать не будем.

В приложениях и в общей теории колец (а такая теория, и притом чрезвычайно развитая, существует) рассматриваются алгебраические структуры, в которых аксиома K2) либо совсем устраивается, либо заменяется другой — в зависимости от конкретной задачи. В таких случаях говорят о *неассоциативных кольцах*. Пока у нас будут только обычные (ассоциативные) кольца. Это значит, что мы можем опираться на теорему 1 из § 1 и не заботиться о расстановке скобок в произведении $a_1 a_2 \dots a_k$ любого числа k элементов кольца.

Подмножество L кольца K называется *подкольцом*, если

$$x, y \in L \implies x - y \in L, \quad xy \in L,$$

т.е. если L — подгруппа аддитивной группы и подполугруппа мультипликативной полугруппы кольца.

Ясно, что пересечение любого семейства подколец в K является подкольцом (рассуждения те же, что и в упр. 1 из § 2) и, стало быть, имеет смысл говорить о подкольце $\langle T \rangle \subset K$, порождённом подмножеством $T \subset K$. По определению $\langle T \rangle$ — пересечение всех тех подколец в K , которые содержат T . Если с самого начала T было подкольцом, то $\langle T \rangle = T$.

Кольцо называется *коммутативным*, если $xy = yx$ для всех $x, y \in K$ (в отличие от групп, коммутативное кольцо не принято называть абелевым).

Понятие кольца в том виде, как оно введено нами, является весьма широким. Более того, класс коммутативных колец, кажущийся на первый взгляд довольно специальным, был предметом усиленного изучения в течение многих десятилетий, и в настоящее время теория коммутативных колец переплетается с алгебраической геомет-

рией — красивой математической дисциплиной, пограничной между алгеброй, геометрией и топологией.

Пример 1. $(\mathbb{Z}, +, \cdot)$ — кольцо целых чисел с обычными операциями сложения и умножения. Множество $m\mathbb{Z}$ целых чисел, делящихся на m , будет в \mathbb{Z} подкольцом (без единицы при $m > 1$). Аналогично, кольцами с единицей являются \mathbb{Q} и \mathbb{R} , причём естественные включения $\mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}$ определяют цепочки подколец кольца \mathbb{R} .

Пример 2. Свойства операций сложения и умножения в $M_n(\mathbb{R})$, введённые и подробно изученные нами в гл. 2, позволяют утверждать, что $M_n(\mathbb{R})$ — кольцо с единицей $1 = E$. Оно называется **полным матричным кольцом над \mathbb{R}** , а также **кольцом квадратных матриц порядка n** (или размера $n \times n$) над \mathbb{R} . Это один из самых важных примеров колец. Так как при $n > 1$ матрицы, как правило, неперестановочны, то $M_n(\mathbb{R})$ — некоммутативное кольцо. Оно содержит в качестве подколец кольца $M_n(\mathbb{Q})$ и $M_n(\mathbb{Z})$ квадратных матриц того же порядка над \mathbb{Q} и над \mathbb{Z} соответственно. Вообще, $M_n(\mathbb{R})$ насыщено всевозможными подкольцами. Время от времени некоторые из них будут возникать у нас естественным образом. Заметим ещё, что можно рассматривать кольцо квадратных матриц $M_n(K)$ над произвольным коммутативным кольцом K , поскольку при сложении и умножении двух матриц $A, B \in M_n(K)$ будет снова получаться матрица с коэффициентами из K , а законы дистрибутивности в $M_n(K)$ являются следствиями аналогичных законов в K . Всё это прямо вытекает из формальных правил действий с матрицами, подытоженных в пп. 2 и 5 из § 3 гл. 2.

Пример 3. Наряду с кольцом матриц в различных разделах математики широко используется также **кольцо функций**. Именно, пусть X — произвольное множество, K — произвольное кольцо. Пусть, далее, $K^X = \{X \rightarrow K\}$ — множество всех функций (или, что то же самое, отображений) $f : X \rightarrow K$, рассматриваемое вместе с двумя бинарными операциями — **поточечной суммой** $f + g$ и **поточечным произведением** fg , определёнными следующим образом:

$$(f + g)(x) = f(x) \oplus g(x), \\ (fg)(x) = f(x) \odot g(x)$$

(\oplus и \odot — операции сложения и умножения в K). Это, очевидно, не та композиция (суперпозиция) функций, которая привела нас в случае линейных отображений к кольцу M_n . Скорее мы становимся здесь на точку зрения, принятую в математическом анализе, когда, например, при $X = \mathbb{R}$, $K = \mathbb{R}$ произведением функций tg и \sin будет $\operatorname{tg} \cdot \sin : x \mapsto \operatorname{tg} x \cdot \sin x$, а не $\operatorname{tg} \circ \sin : x \mapsto \operatorname{tg}(\sin x)$.

Легко проверяется, что K^X удовлетворяет всем аксиомам кольца. Так, ввиду дистрибутивности операций в K имеем

$$[f(x) \oplus g(x)] \odot h(x) = f(x) \odot h(x) \oplus g(x) \odot h(x)$$

для любых трёх функций $f, g, h \in K^X$ и любого $x \in X$, а это по определению поточечных операций даёт $(f + g)h = fh + gh$. Справедливость второго дистрибутивного закона устанавливается аналогично. Если $0, 1$ — нулевой и единичный элементы в K , то

$$0_X : x \mapsto 0, \quad 1_X : x \mapsto 1$$

— постоянные функции, играющие роль нуля и единицы в K^X . В случае коммутативности K кольцо функций K^X также коммутативно.

Кольцо K^X содержит разнообразные подкольца, определяемые специальными свойствами функций. Пусть, например, $X = [0, 1]$ — замкнутый интервал в \mathbb{R} и $K = \mathbb{R}$. Тогда кольцо $\mathbb{R}^{[0,1]}$ всех вещественных функций, определённых на $[0, 1]$, содержит в качестве подколец кольцо $\mathbb{R}^{[0,1]_{\text{огр}}}$ всех ограниченных функций,

кольцо $\mathbb{R}_{\text{непр}}^{[0,1]}$ всех непрерывных функций, кольцо $\mathbb{R}_{\text{диф}}^{[0,1]}$ всех непрерывно дифференцируемых функций и т.д., поскольку все отмеченные свойства сохраняются при сложении (вычитании) и умножении функций.

Каждому числу $a \in K$ отвечает постоянная функция $a_X : x \mapsto a$, и отображение вложения $a \mapsto a_X$ позволяет рассматривать K как подкольцо в \mathbb{R}^X . Словом, почти каждому естественному классу функций соответствует свое подкольцо в \mathbb{R}^X .

Пример 4. На любой аддитивной абелевой группе $(A, +)$ соотношением $xy = 0$ для всех $x, y \in A$ устанавливается структура кольца с нулевым умножением.

Многие свойства колец являются переформулировками соответствующих свойств групп и вообще множеств с одной ассоциативной операцией. Например, $a^m a^n = a^{m+n}$, $(a^m)^n = a^{mn}$ для всех неотрицательных целых m, n и всех $a \in K$ (ср. с соотношением (2) § 1). Другие свойства, более специфические для колец и вытекающие прямо из аксиом кольца, моделируют по существу свойства \mathbb{Z} . Отметим некоторые из них. Во-первых, для всех $a \in K$

$$a \cdot 0 = 0 \cdot a = 0. \quad (1)$$

Действительно, $a + 0 = a \Rightarrow a(a + 0) = aa \Rightarrow a^2 + a \cdot 0 = a^2 \Rightarrow a^2 + a \cdot 0 = a^2 + 0 \Rightarrow a \cdot 0 = 0$ (аналогично, $0 \cdot a = 0$).

Теперь, предположив на момент, что $0 = 1$, мы получим $a = a \cdot 1 = a \cdot 0 = 0$ для всех $a \in K$, т.е. K состоит только из нуля. Стало быть, в нетривиальном кольце K всегда $0 \neq 1$. Далее,

$$(-a) \cdot b = a(-b) = -(ab), \quad (2)$$

поскольку, например, из (1) и аксиомы дистрибутивности следует

$$0 = a \cdot 0 = a(b - b) = ab + a(-b) \Rightarrow a(-b) = -(ab). \quad (3)$$

Так как $-(-a) = a$, то из (2) получаем равенства $(-a)(-b) = ab$ (например, $(-1)(-1) = 1$), $-a = (-1) \cdot a$.

Аксиома дистрибутивности имеет своим следствием общий закон дистрибутивности

$$(a_1 + \dots + a_n)(b_1 + \dots + b_m) = \sum_{i=1}^n \sum_{j=1}^m a_i b_j, \quad (4)$$

в чём нетрудно убедиться рассуждением по индукции сначала (при $m = 1$) по n , а затем по m . Используя теперь (1), (2) и (3), получим

$$n(ab) = (na)b = a(nb)$$

для всех $n \in \mathbb{Z}$ и $a, b \in K$.

Наконец, отметим биномиальную формулу (бином Ньютона)

$$(a + b)^n = \sum_{i=0}^n \binom{n}{i} a^i b^{n-i}, \quad (5)$$

справедливую для всех $a, b \in K$, но только в коммутативном кольце K . При доказательстве (5) нужно, опираясь на (4), действовать так же, как и в § 7 гл. I, где рассмотрен частный случай $K = \mathbb{Z}$.

2. Сравнения. Кольцо классов вычетов. Пусть m — фиксированное натуральное число, $m > 1$. Множество $m\mathbb{Z}$, очевидно, замкнуто не только относительно операции сложения, но и относительно операции умножения, и удовлетворяет всем трем аксиомам кольца.

Теперь, используя подкольцо $m\mathbb{Z} \subset \mathbb{Z}$, построим ненулевое кольцо, состоящее из конечного числа элементов. С этой целью введём

Определение. Два целых числа n, n' называются *сравнимыми по модулю m* , если при делении на m они дают одинаковые остатки. При этом пишут $n \equiv n' \pmod{m}$ или $n \equiv n' \pmod{m}$, а число m называют *модулем сравнения*.

Получается разбиение \mathbb{Z} на классы чисел, сравнимых между собой по модулю m и называемых *классами вычетов* по модулю m . Каждый класс вычетов имеет вид

$$\{r\}_m = r + m\mathbb{Z} = \{r + mk \mid k \in \mathbb{Z}\},$$

так что

$$\mathbb{Z} = \{0\}_m \cup \{1\}_m \cup \dots \cup \{m-1\}_m. \quad (6)$$

По определению $n \equiv n' \pmod{m} \iff n - n'$ делится на m . Удобство записи $n \equiv n' \pmod{m}$ для отношения делимости $m|(n - n')$ состоит в том, что с такими сравнениями можно оперировать совершенно так же, как с обычными равенствами. А именно, если $k \equiv k' \pmod{m}$ и $l \equiv l' \pmod{m}$, то $k \pm l \equiv k' \pm l' \pmod{m}$ и $kl \equiv k'l' \pmod{m}$. В частности, $k \equiv k' \pmod{m} \implies ks \equiv k's \pmod{m}$ для любого $s \in \mathbb{Z}$.

Таким образом, каждым двум классам $\{k\}_m$ и $\{l\}_m$ независимо от выбора в них представителей k, l можно сопоставить классы, являющиеся их суммой или произведением, т.е. на множестве $\mathbb{Z}_m = \mathbb{Z}/m\mathbb{Z}$ классов вычетов по модулю m однозначным образом индуцируются операции \oplus и \odot :

$$\begin{aligned} \{k\}_m \oplus \{l\}_m &= \{k + l\}_m, \\ \{k\}_m \odot \{l\}_m &= \{kl\}_m. \end{aligned} \quad (7)$$

Так как определения этих операций сводятся к соответствующим операциям над числами из классов вычетов, т.е. над элементами из \mathbb{Z} , то $\{\mathbb{Z}_m, \oplus, \odot\}$ будет также коммутативным кольцом с единицей $\{1\}_m = 1 + m\mathbb{Z}$. Оно называется *кольцом классов вычетов по модулю m* . При небольшом навыке (и фиксированном модуле) индекс m опускают и пишут \bar{k} вместо $\{k\}_m$, так что

$$\bar{k} \oplus \bar{l} = \overline{k + l},$$

$$\bar{k} \odot \bar{l} = \overline{kl}.$$

Высший этап освоения с \mathbb{Z}_m , кажущийся на первый взгляд кощунственным, но представляющий явные технические преимущества, заключается в том, что отказываются от чёрточек и кружочков и опирают с каким-нибудь фиксированным множеством представителей по модулю m , чаще всего — с множеством $\{0, 1, 2, \dots, m-1\}$ (оно называется *приведённой системой вычетов* по модулю m). Скажем, в соответствии с этим соглашением $-k = m - k$, $2(m-1) = -2 = m-2$.

Итак, конечные кольца существуют. Приведём три простейших примера, указывая отдельно таблицы сложения и умножения:

$\mathbb{Z}_2 :$	$+ \begin{array}{ c c c } \hline & 0 & 1 \\ \hline 0 & 0 & 1 \\ 1 & 1 & 0 \\ \hline \end{array}$	$\cdot \begin{array}{ c c c } \hline & 0 & 1 \\ \hline 0 & 0 & 0 \\ 1 & 0 & 1 \\ \hline \end{array}$
$\mathbb{Z}_3 :$	$+ \begin{array}{ c c c } \hline & 0 & 1 & 2 \\ \hline 0 & 0 & 1 & 2 \\ 1 & 1 & 2 & 0 \\ 2 & 2 & 0 & 1 \\ \hline \end{array}$	$\cdot \begin{array}{ c c c } \hline & 0 & 1 & 2 \\ \hline 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 2 \\ 2 & 0 & 2 & 1 \\ \hline \end{array}$
$\mathbb{Z}_4 :$	$+ \begin{array}{ c c c c } \hline & 0 & 1 & 2 & 3 \\ \hline 0 & 0 & 1 & 2 & 3 \\ 1 & 1 & 2 & 3 & 0 \\ 2 & 2 & 3 & 0 & 1 \\ 3 & 3 & 0 & 1 & 2 \\ \hline \end{array}$	$\cdot \begin{array}{ c c c c } \hline & 0 & 1 & 2 & 3 \\ \hline 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 2 & 3 \\ 2 & 0 & 2 & 0 & 2 \\ 3 & 0 & 3 & 2 & 1 \\ \hline \end{array}$

Кольцо вычетов \mathbb{Z}_m издавна привлекало внимание теоретико-числовиков, а в алгебре служило отправным пунктом для разного рода обобщений.

3. Гомоморфизмы колец. Отображение $f : n \mapsto \{n\}_m$ обладает в силу (7) следующими свойствами:

$$f(k+l) = f(k) \oplus f(l), \quad f(kl) = f(k) \odot f(l).$$

Это даёт нам основание говорить о гомоморфизме колец \mathbb{Z} и \mathbb{Z}_m в соответствии с общим определением.

Определение. Пусть $(K, +, \cdot)$ и (K', \oplus, \odot) — кольца. Отображение $f : K \rightarrow K'$ называется *гомоморфизмом*, если оно сохраняет все операции, т.е. если

$$f(a+b) = f(a) \oplus f(b),$$

$$f(ab) = f(a) \odot f(b).$$

При этом, конечно, $f(0) = 0'$ и $f(na) = nf(a)$, $n \in \mathbb{Z}$.

Ядром гомоморфизма f называется множество

$$\text{Ker } f = \{a \in K \mid f(a) = 0'\}.$$

Ясно, что $\text{Ker } f$ — подкольцо в K .

Как и в случае групп (см. словарик в п. 5 § 2), гомоморфизм

$$f: K \rightarrow K'$$

называется:

мономорфизмом, если $\text{Ker } f = 0$;

эпиморфизмом, если образ совпадает с K' , т.е.

$$\text{Im } f = f(K) = \{a' \in K' \mid a' = f(a)\} = K';$$

изоморфизмом, если отображение f мономорфно и эпиморфно.

Факт изоморфизма колец кратко записывают в виде $K \cong K'$.

Рассмотренное выше отображение $f: n \mapsto \{n\}_m$ является, очевидно, эпиморфизмом $\mathbb{Z} \rightarrow \mathbb{Z}_m$ с ядром $\text{Ker } f = m\mathbb{Z}$.

Если рассматривать только кольца с единицей, то в определение гомоморфизма $f: K \rightarrow K'$ целесообразно внести условие

$$f(1) = 1'.$$

При эпиморфизме это условие, конечно, автоматически выполняется.

Изоморфные кольца тождественны по своим алгебраическим свойствам, и подлинно математический интерес представляют только те свойства колец, которые сохраняются при изоморфных отображениях. Именно это обстоятельство имелось в виду, когда кольцо \mathbb{Z}_m мыслилось то как множество классов вычетов по модулю m , то как множество произвольным образом выбранных представителей этих классов.

4. Типы колец. Поле. В хорошо известных нам числовых кольцах \mathbb{Z} , \mathbb{Q} и \mathbb{R} из $ab = 0$ следует, что либо $a = 0$, либо $b = 0$. Но кольцо квадратных матриц M_n над любым из указанных колец этим свойством уже не обладает. Используя матрицы E_{ij} (см. доказательство теоремы 4 из § 3 гл. 2), мы приходим к равенствам $E_{ij}E_{ki} = 0$ при $j \neq k$, хотя, конечно, $E_{ij} \neq 0$ и $E_{ki} \neq 0$. Заметим, что $E_{ik}E_{kj} = E_{ij} \neq 0$. Можно было бы присвоить столь необычный для элементарной арифметики феномен некоммутативности кольца M_n , но это не так. Как мы видели в п. 2, в коммутативном кольце \mathbb{Z}_4 выполнено равенство $2 \odot 2 = 0$, вопреки общезвестной истине “дважды два — четыре”. Вот — ещё два примера.

Пример 5. Числовые пары (a, b) (где $a, b \in \mathbb{Z}, \mathbb{Q}, \mathbb{R}$) со сложением и умножением, определёнными формулами

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

$$(a_1, b_1) \cdot (a_2, b_2) = (a_1 a_2, b_1 b_2),$$

образуют, очевидно, коммутативное кольцо с единицей $(1, 1)$, в котором мы снова встречаемся с тем же явлением: $(1, 0) \cdot (0, 1) = (0, 0) = 0$.

Пример 6. В кольце $\mathbb{R}^{\mathbb{R}}$ вещественных функций (см. пример 3 в п. 1) функции $f: x \mapsto |x| + x$ и $g: x \mapsto |x| - x$ таковы, что $f(x) = 0$ для $x \leq 0$ и $g(x) = 0$ для $x \geq 0$, а поэтому их поточечным произведением fg будет нулевая функция, хотя $f \neq 0$ и $g \neq 0$.

Определение. Если $ab = 0$ при $a \neq 0$ и $b \neq 0$ в кольце K , то a называется *левым*, а b — *правым делителем нуля* (в коммутативном кольце K говорят просто о делителях нуля). Сам нуль в кольце $K \neq 0$ — тривиальный делитель нуля. Если других делителей нуля нет (кроме 0), то K называется *кольцом без делителей нуля*. Коммутативное кольцо с единицей $1 \neq 0$ и без делителей нуля называют *целостным кольцом* (*кольцом целостности* или *областью целостности*).

Теорема 1. *Нетривиальное коммутативное кольцо K с единицей является целостным тогда и только тогда, когда в нём выполнен закон сокращения*

$$ab = ac, a \neq 0 \implies b = c$$

для всех $a, b, c \in K$.

В самом деле, если в K имеет место закон сокращения, то из $ab = 0 = a \cdot 0$ следует, что либо $a = 0$, либо $a \neq 0$, но $b = 0$. Обратно: если K — область целостности, то

$$ab = ac, a \neq 0 \implies a(b - c) = 0 \implies b - c = 0 \implies b = c. \square$$

В кольце K с единицей естественно рассматривать множество обратимых элементов. Элемент a называется *обратимым* (или *делителем единицы*), если существует элемент a^{-1} , для которого $aa^{-1} = 1 = a^{-1}a$. Точнее, следовало бы говорить об элементах, *обратимых справа или слева* ($ab = 1$ или $ba = 1$), но в коммутативных кольцах, а также в кольцах без делителей нуля эти понятия совпадают. Действительно, из $ab = 1$ следует $aba = a$, откуда $a(ba - 1) = 0$. Так как $a \neq 0$, то $ba - 1 = 0$, т.е. $ba = 1$.

Нам известно, например, что в кольце M_n обратимые элементы — это в точности матрицы с отличным от нуля определителем. Обратимый элемент a не может быть делителем нуля:

$$ab = 0 \implies a^{-1}(ab) = 0 \implies (a^{-1}a)b = 0 \implies 1 \cdot b = 0 \implies b = 0$$

(аналогично, $ba = 0 \implies b = 0$). Неудивительно поэтому, что имеет место

Теорема 2. *Все обратимые элементы кольца K с единицей составляют группу $U(K)$ по умножению.*

В самом деле, так как множество $U(K)$ содержит единицу, $a \in U(K) \implies a^{-1} \in U(K)$ и ассоциативность по умножению в K выполнена, то нам нужно только убедиться в замкнутости множества $U(K)$, т.е. проверить, что произведение ab любых двух элементов a и b из $U(K)$ будет снова принадлежать $U(K)$. Но это очевидно, поскольку

$$(ab)^{-1} = b^{-1}a^{-1} \quad (ab \cdot b^{-1}a^{-1} = a(bb^{-1})a^{-1} = a \cdot 1 \cdot a^{-1} = aa^{-1} = 1),$$

и, значит, ab обратим. \square

Нетрудно видеть, что $U(\mathbb{Z}) = \{\pm 1\}$ — циклическая группа порядка 2.

Мы получим весьма интересный класс колец — так называемые *кольца с делением*, или *тела*, заменив в определении кольца аксиому K2) на существенно более сильное условие

K2') относительно операции умножения множество $K = K \setminus \{0\}$ является группой.

Кольцо с делением, стало быть, всегда без делителей нуля, и каждый ненулевой элемент в нём обратим. Операции сложения и умножения становятся почти полностью симметричными в коммутативном кольце с делением, которое называется *полям*.

Итак, дадим ещё раз

Определение. Поле P — это коммутативное кольцо с единицей $1 \neq 0$, в котором каждый элемент $a \neq 0$ обратим. Группа $P^* = U(P)$ называется *мультипликативной группой поля*.

Поле представляет собой гибрид двух абелевых групп — аддитивной и мультипликативной, связанных законом дистрибутивности (теперь уже одним виду коммутативности).

Произведение ab^{-1} записывается обычно в виде дроби (или отношения, частного) $\frac{a}{b}$, которую для экономии места на бумаге записывают ещё с помощью косой черты: a/b . Следовательно, дробь a/b , имеющая смысл только при $b \neq 0$, является единственным решением уравнения $bx = a$.

Действия с дробями подчиняются нескольким правилам:

$$\begin{aligned} \frac{a}{b} = \frac{c}{d} &\iff ad = bc, \quad b, d \neq 0, \\ \frac{a}{b} + \frac{c}{d} &= \frac{ad + bc}{bd}, \quad b, d \neq 0, \\ -\frac{a}{b} &= \frac{-a}{b} = \frac{a}{-b}, \quad b \neq 0, \\ \frac{a}{b} \cdot \frac{c}{d} &= \frac{ac}{bd}, \quad b, d \neq 0, \\ \left(\frac{a}{b}\right)^{-1} &= \frac{b}{a}, \quad a, b \neq 0. \end{aligned} \tag{8}$$

Это — обычные, “школьные” правила, но их надо не запоминать, а выводить из аксиом поля, что, впрочем, не представляет никаких трудностей. Вот рассуждения, достаточные для получения второго из правил (8). Пусть $x = a/b$ и $y = c/d$ — решения уравнений $bx = a$ и $dy = c$. Из этих уравнений следует

$$dbx = da, \quad bdy = bc \implies bd(x + y) = da + bc \implies t = x + y = \frac{da + bc}{bd}$$

— единственное решение уравнения $bdt = da + bc$.

Подполе F поля P называется подкольцо в P , само являющееся полем. Например, поле рациональных чисел \mathbb{Q} — подполе поля вещественных чисел \mathbb{R} .

В случае $F \subset P$ говорят также, что поле P является *расширением* своего под поля F . Из определения под поля следует, что нуль и единица поля P будут содержаться также в F и служить для F нулём и единицей. Если взять в P пересечение F_1 всех под полей, содержащих F и некоторый элемент $a \in P$, не принадлежащий F , то F_1 будет минимальным полем, содержащим множество $\{F, a\}$ (рассуждение такое же, как для групп в упр. 1 из § 2).

Говорят, что расширение F_1 поля F получено *присоединением* к F элемента a , и отражают этот факт записью $F_1 = F(a)$. Аналогично можно говорить о подполе $F_1 = F(a_1, \dots, a_n)$ поля P , полученном присоединением к F n элементов a_1, \dots, a_n поля P .

Небольшая проверка показывает, что $\mathbb{Q}(\sqrt{2})$ совпадает с множеством чисел $a + b\sqrt{2}$, где $a, b \in \mathbb{Q}$, поскольку $(\sqrt{2})^2 = 2$ и

$$\frac{1}{a+b\sqrt{2}} = \frac{a}{a^2 - 2b^2} - \frac{b}{a^2 - 2b^2}\sqrt{2}$$

при $a + b\sqrt{2} \neq 0$. То же самое относится к $\mathbb{Q}(\sqrt{3})$, $\mathbb{Q}(\sqrt{5})$ и т.д.

Поля P и P' называются *изоморфными*, если они изоморфны как кольца. По определению $f(0) = 0'$ и $f(1) = 1'$ для любого изоморфного отображения f . Не имеет смысла говорить о гомоморфизмах полей, так как

$\text{Ker } f \neq 0 \implies f(a) = 0, a \neq 0 \implies$

$\implies f(1) = f(aa^{-1}) = f(a)f(a^{-1}) = 0 \cdot f(a^{-1}) = 0 \implies$

$\implies \forall b \quad f(b) = f(1 \cdot b) = f(1)f(b) = 0 \cdot f(b) = 0 \implies \text{Ker } f = P$.

Напротив, *автоморфизмы*, т.е. изоморфные отображения поля P на себя, связаны с самыми глубокими свойствами полей и являются мощным инструментом для изучения этих свойств в рамках так называемой теории Галуа.

Понятие расширения полей вполне созвучно известному стремлению человечества увеличивать запас используемых чисел. Довольно медленный процесс, который условно изображается диаграммой

{один} \rightsquigarrow {один да один есть два} $\rightsquigarrow \mathbb{N} \rightsquigarrow \{\mathbb{N}, 0\} \rightsquigarrow$

$\rightsquigarrow \mathbb{Z} \rightsquigarrow \mathbb{Q} \rightsquigarrow \mathbb{Q}(\sqrt{2}) \rightsquigarrow \mathbb{R}$

и который продолжался вплоть до наших дней, привёл к чрезвычайно разветвлённой сети полей, весьма далеких от привычных числовых. Не все этапы этого процесса были чисто алгебраическими. Скажем, переход от рациональных чисел к вещественным (или действительным), основывающийся на понятии непрерывности и полноты (су-

ществование пределов у последовательностей Коши), и поныне разбирается в курсах математического анализа. В то же время совершенно аналогичная конструкция полей p -адических чисел, которой мы здесь не касаемся, и выросший на её основе современный p -адический анализ — достойные детища трёх областей — теории чисел, алгебры и анализа.

5. Характеристика поля. В п. 2 было построено конечное кольцо классов вычетов \mathbb{Z}_m с элементами

$$\bar{0}, \bar{1}, \bar{2}, \dots, \overline{m-1}$$

и операциями $\bar{k} + \bar{l} = \overline{k+l}$, $\bar{k} \cdot \bar{l} = \overline{kl}$ сложения и умножения (мы отказываемся от значков \oplus и \odot). Если $m = st$, $s > 1$, $t > 1$, то $\bar{s} \cdot \bar{t} = \bar{m} = \bar{0}$, т.е. \bar{s} и \bar{t} — делители нуля в \mathbb{Z}_m .

Пусть теперь $m = p$ — простое число. Утверждается, что \mathbb{Z}_p — поле (из p элементов). Для $p = 2, 3$ это непосредственно видно из таблиц умножения, выписанных в п. 2. В общем случае достаточно установить существование для каждого $\bar{s} \in \mathbb{Z}_p^*$ обратного элемента \bar{s}' (целые числа s и s' не должны, очевидно, делиться на p).

Рассмотрим элементы

$$\bar{s}, \overline{2s}, \dots, \overline{(p-1)s}. \quad (9)$$

Они все отличны от нуля, так как

$$s \not\equiv 0 \pmod{p} \implies ks \not\equiv 0 \pmod{p}$$

при $k = 1, 2, \dots, p-1$. (Здесь используется простота p .) По той же причине элементы (9) все различны: из $\overline{ks} = \overline{ls}$, $k < l$, следовало бы $\overline{(k-l)s} = \bar{0}$, что неверно. Итак, последовательность элементов (9) совпадает с последовательностью переставленных каким-то образом элементов

$$\bar{1}, \bar{2}, \dots, \overline{p-1}.$$

В частности, найдется $s', 1 \leq s' \leq p-1$, для которого $\overline{s's} = \bar{1}$. Но это значит, что $\bar{s}'\bar{s} = \bar{1}$, т.е. \bar{s}' — обратный к \bar{s} элемент. Нами доказана

Теорема 3. Кольцо классов вычетов \mathbb{Z}_m является полем тогда и только тогда, когда $m = p$ — простое число.

Следствие (малая теорема Ферма). Для любого целого числа m , не делящегося на простое число p , имеет место сравнение

$$m^{p-1} \equiv 1 \pmod{p}.$$

Доказательство. Как мы видели,

$$\{\bar{m}, \bar{2m}, \dots, \overline{(p-1)m}\} = \{\bar{1}, \bar{2}, \dots, \overline{p-1}\}$$

(заменить в (9) s на m и принять во внимание равенства $\overline{km} = \overline{k}\overline{m}$, $k = 1, \dots, p-1$). Поэтому, перемножая по отдельности все

элементы в левой и правой части, получим

$$\left(\prod_{k=1}^{p-1} \bar{k} \right) \bar{m}^{p-1} = \prod_{k=1}^{p-1} \bar{k}.$$

Поскольку \mathbb{Z}_p — кольцо без делителей нуля, по теореме 1 множитель $\prod_{k=1}^{p-1} \bar{k} \neq \bar{0}$ можно сократить: $\bar{m}^{p-1} = \bar{1}$. На языке сравнений имеем то, что нужно. \square

Справедлива более общая теорема Эйлера, но необходимость в ней возникнет лишь в [ВА III].

Поля $\mathbb{Z}_2, \mathbb{Z}_3, \mathbb{Z}_5, \dots$, столь не похожие на известные нам поля $\mathbb{Q}, \mathbb{Q}(\sqrt{2}), \mathbb{R}$, заняли в алгебраической иерархии полей место, вполне сопоставимое по своему значению с местом, давно отведённым для \mathbb{Q} . Дело здесь вот в чём. Пусть P — поле. Как мы уже отмечали, пересечение $\bigcap_i P_i$ любого семейства подполей $\{P_i | i \in I\}$ будет подполем в P .

Определение. Поле, не обладающее никаким собственным подполем, называется *простым*.

Теорема 4. В каждом поле P содержится одно и только одно простое поле P_0 . Это простое поле изоморфно либо \mathbb{Q} , либо \mathbb{Z}_p для некоторого простого p .

Доказательство. Допустив существование двух различных простых подполей $P', P'' \subset P$, мы неизбежно придём к выводу, что их пересечение $P' \cap P''$ (очевидно, непустое, поскольку 0 и 1 содержатся как в P' , так и в P'') будет полем, отличным от P' и P'' . Это, однако, невозможно ввиду их простоты. Стало быть, простое подполе $P_0 \subset P$ единственно.

В P_0 наряду с единичным элементом 1 содержатся все его кратные $n \cdot 1 = 1 + \dots + 1$. Из общих свойств операций сложения и умножения элементов в кольцах (см. конец п. 1) следует, что

$$s \cdot 1 + t \cdot 1 = (s+t) \cdot 1, \quad (s \cdot 1)(t \cdot 1) = (st) \cdot 1; \quad s, t \in \mathbb{Z}.$$

Поэтому отображение f кольца \mathbb{Z} в P , определённое правилом $f(n) = n \cdot 1$, является гомоморфизмом, ядро которого имеет вид $\text{Ker } f = m\mathbb{Z}$. Если $m = 0$, то f — мономорфизм, и дроби $(s \cdot 1)/(t \cdot 1)$, имеющие смысл в P (поскольку P — поле), образуют поле P_0 , изоморфное \mathbb{Q} . Оно и будет простым подполем в P .

Если же $m > 0$, то, очевидно, отображение f^* , определённое по правилу

$$f^*: \bar{k} = \{k\}_m \mapsto f(k),$$

будет изоморфным вложением $\mathbb{Z}_m \rightarrow P$. По теореме 3 это возможно только тогда, когда $m = p$ — простое число. Стало быть, $f^*(\mathbb{Z}_p)$ — простое подполе в P . \square

Определение. Говорят, что поле P имеет *характеристику нуль*, если его простое подполе P_0 изоморфно \mathbb{Q} ; P — поле *простой*

(или *конечной*) характеристики p , если $P_0 \cong \mathbb{Z}_p$. Соответственно пишут $\text{char}P = 0$ или $\text{char}P = p > 0$.

Вместо \mathbb{Z}_p обозначением "абстрактного" поля из p элементов служит обычно \mathbb{F}_p или $\text{GF}(p)$ (Galois Field — *поле Галуа*). Следует иметь в виду, что существует конечное поле $\text{GF}(q)$ с $q = p^n$ элементами, где p — простое, а n — любое целое положительное число. К этому интересному вопросу мы вернёмся в [ВА III], а сейчас ограничимся лишь примером поля из четырёх элементов $\{0, 1, \alpha, \beta\}$:

	+	0	1	α	β		*	0	1	α	β
$\text{GF}(4) :$	0	0	1	α	β		0	0	0	0	0
	1	1	0	β	α		1	0	1	α	β
	α	α	β	0	1		α	0	α	β	1
	β	β	α	1	0		β	0	β	1	α

Чем являются α и β , нас пока не интересует. Рекомендуется проверить выполнение закона дистрибутивности.

Иногда нулевую характеристику называют *бесконечной* в соответствии с её интерпретацией как порядка элемента 1 в аддитивной группе поля P . Аналогично, конечная характеристика p — общий порядок любого ненулевого элемента в аддитивной группе:

$$px = x + \dots + x = 1 \cdot x + \dots + 1 \cdot x = (1 + \dots + 1)x = (p \cdot 1)x = 0.$$

6. Замечание о линейных системах. Настала пора окинуть мысленным взором изложенную в предыдущих главах теорию систем линейных уравнений и выросшую из неё теорию определителей. Коэффициентами в линейных уравнениях и элементами матриц у нас были числа (рациональные или вещественные), но специфика этих чисел никак не использовалась. Нет никаких препятствий к тому, чтобы взять теперь вместо чисел элементы фиксированного поля P . При этом и результаты должны формулироваться в терминах поля P : компоненты решения линейной системы и значения функции \det будут лежать в P . Метод Гаусса решений систем линейных уравнений, теория определителей, правило Крамера остаются справедливыми (без существенных изменений) для произвольного поля P .

Пример 7. Пусть нам дана однородная система линейных уравнений $AX = 0$ с квадратной матрицей

$$A = (a_{ij}) = \begin{vmatrix} 1 & 2 & 3 & 4 \\ -10 & 13 & 14 & 15 \\ 12 & -9 & 14 & 15 \\ 12 & 13 & -8 & 15 \end{vmatrix}$$

и столбцом неизвестных $X = [x_1, x_2, x_3, x_4]$. Прямые вычисления показывают, что $\det A = 2^3 \cdot 11^3$. Следовательно, при $a_{ij}, x_k \in P$, где P — любое поле характеристики нуль или характеристики $p \neq 2, 11$ (в этом случае целые числа 1, 2, 3, 4, -10, ..., 15 заменяются на соответствующие классы вычетов), система является определённой и имеет только тривиальное решение $X = 0$.

Если $\text{char } P = 2$ (скажем, $P = \mathbb{Z}_2$), то из сравнения

$$\left| \begin{array}{cccc} 1 & 2 & 3 & 4 \\ -10 & 13 & 14 & 15 \\ 12 & -9 & 14 & 15 \\ 12 & 13 & -8 & 15 \end{array} \right| \equiv \left| \begin{array}{cccc} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{array} \right| \pmod{2}$$

мы заключаем, что ранг системы равен 2 и система допускает два независимых решения $X_1 = [1, 0, 1, 0]$, $X_2 = [0, 1, 0, 1]$. Во избежание недоразумений следовало бы писать $X_1 = [\bar{1}, \bar{0}, \bar{1}, \bar{0}]$, $X_2 = [\bar{0}, \bar{1}, \bar{0}, \bar{1}]$, но мы считаем себя достаточно подготовленными к восприятию упрощённой записи.

Если $\text{char } P = 11$, то из сравнения

$$\left| \begin{array}{cccc} 1 & 2 & 3 & 4 \\ -10 & 13 & 14 & 15 \\ 12 & -9 & 14 & 15 \\ 12 & 13 & -8 & 15 \end{array} \right| \equiv \left| \begin{array}{cccc} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{array} \right| \pmod{11}$$

вытекает, что система имеет три независимых решения

$$X_1 = [9, 1, 0, 0], \quad X_2 = [8, 0, 1, 0], \quad X_3 = [7, 0, 0, 1].$$

Как мы видим, ответ о числе решений существенно зависит от рассматриваемого поля P , но анализ системы ничем не отличается от обычного. Стало быть, одно из преимуществ перехода от \mathbb{R} и \mathbb{Q} к произвольному полю заключается в устраниении дублирования сходных рассуждений. Но имеются к тому и более веские причины.

Говоря о полной линейной группе, мы до сих пор считали её группой всех невырожденных матриц с коэффициентами из \mathbb{Q} или \mathbb{R} . Совокупность квадратных матриц порядка n с коэффициентами в произвольном поле P составляет кольцо матриц $M_n(P)$, а подмножество всех невырожденных матриц $A \in M_n(P)$ (матриц с $\det A \neq 0$) приводит к понятию полной линейной группы $GL_n(P)$ над полем P . Варьируя поле P , например, полагая $P = \mathbb{F}_p$, можно естественным путем получить ряд важных групп (см. [ВА III]).

Поля типа $\mathbb{R}, \mathbb{Q}, \mathbb{Q}(\sqrt{2})$ и прочие называются обычно **числовыми полями**. Поле \mathbb{F}_p — пример нечислового поля: было бы неправильным называть его элементы числами лишь на том основании, что они часто отождествляются с элементами множества $\{0, 1, \dots, p-1\}$.

В § 2 гл. 1 ставилась задача (под номером 3) по использованию конечных полей в теории кодирования. Мы приведём сейчас маленький пример на эту тему.

Пример 8. Для передачи лозунга МИРУ МИР в принципе достаточно повторения четырёх элементарных сообщений

$$M = (0, 0), \quad I = (1, 0), \quad R = (0, 1), \quad Y = (1, 1),$$

интерпретируемых как векторы-строки двумерного линейного пространства \mathbb{F}_2^2 над полем $\mathbb{F}_2 \cong \mathbb{Z}_2 = \{0, 1\}$ из двух элементов. Но во время передачи в канале связи возникают помехи (замены символа 0 на 1 или 1 на 0), в результате которых на приёмный конец канала может прийти, например, сообщение РИМУ РИМ. Согласно фундаментальной теореме Шеннона за счёт увеличения длины

элементарных сообщений (т.е. за счёт скорости передачи) влияние помех устранимо. Пусть, скажем, из условий передачи известно, что в каждом элементарном сообщении длины 5 происходит не более одного искажения. Возьмём тогда в линейном пространстве $S = \mathbb{F}_2^5$ подмножество

$$S_0 = \{M = (0, 0, 1, 1, 0), \quad I = (1, 0, 0, 1, 1), \quad R = (0, 1, 1, 0, 1), \quad Y = (1, 1, 0, 0, 0)\}$$

так называемых **кодовых векторов**. Из таблицы

Кодовые векторы	00110	10011	01101	11000
Векторы, получаемые из кодовых векторов в результате искажения	00010 00100 00111 01110 10110	00011 10001 10010 10111 11011	00101 01001 01100 01111 11101	01000 10000 11100 11001 11010

видно, что множества искажённых векторов из разных столбцов не пересекаются, и, стало быть, возможно правильное декодирование, т.е. восстановление истинного сообщения.

Мы получили код S_0 , исправляющий одну ошибку. Переходя к пространствам \mathbb{F}_2^n достаточно большой размерности n , можно сконструировать аналогичный код, способный безошибочно передать весь русский алфавит, т.е. любой текст. Чтобы декодирование не свелось к длительному и очень медленному перебору, S_0 приходится выбирать специальным образом. Для этого существует множество приёмов, в том числе и чисто алгебраических, основанных на использовании конечных полей \mathbb{F}_q .

УПРАЖНЕНИЯ

1. Развивая идею примера 2 из § 1, показать, что множество $\mathcal{P}(\Omega)$ с операциями

$$A + B = (A \cup B) \setminus (A \cap B), \quad AB = A \cap B, \quad A, B \in \Omega,$$

является кольцом с единицей, все элементы аддитивной группы которого имеют порядок 2.

2. Установить коммутативность произвольного кольца, в котором каждый элемент x удовлетворяет уравнению $x^2 = x$. Верно ли это при условии $x^3 = x$?

3. Изоморфны ли поля $\mathbb{Q}(\sqrt{2})$, $\mathbb{Q}(\sqrt{5})$?

4. Показать, что эпиморфный образ коммутативного кольца является коммутативным кольцом.

5. Показать, что любое конечное целостное кольцо K является полем.

6. Пусть p — простое число и K — коммутативное кольцо с единицей такое, что $px = 0$ для всех $x \in K$. Показать, что тогда

$$(x + y)^{p^m} = x^{p^m} + y^{p^m}, \quad m = 1, 2, \dots$$

Указание. Использовать индукцию по m и то обстоятельство, что биномиальный коэффициент $\binom{p}{k}$, $0 < k < p$, делится на p .

7. Доказать, что кольцо K , состоящее из пяти элементов, либо изоморфно \mathbb{Z}_5 , либо является кольцом с нулевым умножением.

8. Элемент $x \neq 0$ кольца K называется **нильпотентным**, если $x^n = 0$ для некоторого $n \in \mathbb{N}$. Показать, что:

1) nilpotентность элемента x влечёт обратимость элемента $1 - x$ в любом кольце с единицей;

2) кольцо $\mathbb{Z}_m = \mathbb{Z}/m\mathbb{Z}$ содержит нильпотентные элементы в точности тогда, когда m делится на квадрат натурального числа > 1 .

9. Доказать, что в кольце K с единицей и бесконечной мощности $|K|$ не может быть конечного числа $n \leq 1$ необратимых элементов $\neq 0$.

Указание. Использовать рассуждение от противного. Пусть $N = \{a_1, \dots, a_n\}$ — множество всех $\neq 0$ необратимых элементов кольца K . Отображение $\rho_x: a_i \mapsto xa_i$ является биекцией $N \rightarrow N$ для любого $x \in K \setminus (N \cup \{0\})$. Ядро $\text{Ker } \rho$ отображения $\rho: x \mapsto \rho_x$ бесконечно.

10. Пусть K — произвольное ассоциативное кольцо с единицей 1 и a, b — его элементы. Показать, что

$$(1 - ab)c = 1 = c(1 - ab) \implies (1 - ba)d = 1 = d(1 - ba),$$

где $d = 1 + bca$, т.е. обратимость $1 - ab$ в K влечёт обратимость $1 - ba$. Чему равен элемент $1 + adb$?

11. Показать, что матрицы $\begin{vmatrix} a & b \\ -b & a \end{vmatrix}$, где $b \in \mathbb{Z}_3$, образуют поле из девяти элементов и что мультипликативная группа этого поля циклическая порядка 8.

12. Способен ли код S_0 (из примера 2 в конце параграфа) исправить две ошибки?

КОМПЛЕКСНЫЕ ЧИСЛА И МНОГОЧЛЕНЫ

В этой главе будут рассмотрены вполне конкретные алгебраические системы, частично известные из школьной математики, но заслуживающие того, чтобы остановиться на них несколько подробнее. Точка зрения, выработанная в предыдущей главе, позволит нам бросить свежий взгляд на традиционное “поле деятельности” алгебры прошлых веков. В то же время на примере многочленов станут более понятными и осязаемыми такие проблемы, как расширение колец и однозначность разложения на простые множители в целостных кольцах (областях целостности).

§ 1. Поле комплексных чисел

История математики отмечена длительной борьбой сторонников и противников “мнимых” чисел, источником которых служит алгебраическое уравнение

$$x^2 + 1 = 0. \quad (1)$$

Можно занять упрощенную позицию и ограничиться формальной записью решений уравнения (1) в виде $\pm\sqrt{-1}$. Но такое немудрено было сделать и в более далёкие времена; оставалось лишь придать смысл указанной записи. Мы будем решать эту задачу на разных уровнях. Вначале приведём некоторые эвристические соображения.

1. Вспомогательная конструкция. Нам хочется расширить поле вещественных чисел \mathbb{R} так, чтобы в новом поле уравнение (1) обладало решением. Моделью такого расширения может служить множество P всех квадратных матриц

$$\begin{vmatrix} a & b \\ -b & a \end{vmatrix} \in M_2(\mathbb{R}). \quad (2)$$

Утверждается, что P — поле (ср. с упр. 11 из § 3 гл. 4).

В самом деле, в P содержатся нуль 0 и единица E кольца $M_2(\mathbb{R})$. Далее, из соотношений

$$\begin{aligned} \begin{vmatrix} a & b \\ -b & a \end{vmatrix} + \begin{vmatrix} c & d \\ -d & c \end{vmatrix} &= \begin{vmatrix} a+c & b+d \\ -(b+d) & a+c \end{vmatrix}, \\ - \begin{vmatrix} a & b \\ -b & a \end{vmatrix} &= \begin{vmatrix} -a & -b \\ -(-b) & -a \end{vmatrix}, \\ \begin{vmatrix} a & b \\ -b & a \end{vmatrix} \begin{vmatrix} c & d \\ -d & c \end{vmatrix} &= \begin{vmatrix} ac - bd & ad + bc \\ -(ad + bc) & ac - bd \end{vmatrix} \end{aligned} \quad (3)$$

вытекает замкнутость P относительно операций сложения и умножения. Ассоциативность этих операций является следствием их ассоциативности в $M_2(\mathbb{R})$. То же самое относится к законам дистрибутивности и коммутативности сложения. Таким образом, P — подкольцо в $M_2(\mathbb{R})$. Коммутативность умножения в P вытекает из третьей формулы (3), и остаётся доказать лишь существование в P матрицы, обратной к любой матрице (2) с определителем $a^2 + b^2 \neq 0$. Прямо по формуле для коэффициентов обратной матрицы (см. теорему 1 из § 3 гл. 3) или путём решения линейной системы

$$ax - by = 1,$$

$$bx + ay = 0,$$

возникающей из условия

$$\begin{vmatrix} a & b \\ -b & a \end{vmatrix} \begin{vmatrix} x & y \\ -y & x \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix},$$

находим, что

$$\begin{vmatrix} a & b \\ -b & a \end{vmatrix}^{-1} = \begin{vmatrix} c & d \\ -d & c \end{vmatrix}, \quad (4)$$

где

$$c = \frac{a}{a^2 + b^2}, \quad d = \frac{-b}{a^2 + b^2}.$$

Используя правило (5) из § 3 гл. 2 умножения матриц на числа, мы любой элемент поля P запишем в виде

$$\begin{vmatrix} a & b \\ -b & a \end{vmatrix} = aE + bJ, \quad a, b \in \mathbb{R}, J = \begin{vmatrix} 0 & 1 \\ -1 & 0 \end{vmatrix}. \quad (5)$$

Поле P содержит подполе $\{aE \mid a \in \mathbb{R}\} \cong \mathbb{R}$, а соотношение

$$J^2 + E = 0$$

показывает, что элемент $J \in P$ “с точностью до изоморфизма” является решением уравнения (1). Ни о какой мистике вокруг “мнимой величины J ” здесь не может быть и речи.

Не поле P , однако, называется полем комплексных чисел, а некий изоморфный ему объект, элементы которого изображаются точками плоскости. Желание иметь геометрическую реализацию поля P не случайно, если вспомнить, что и поле \mathbb{R} для нас не отделимо от “вещественной прямой” с фиксированной на ней точкой, изображающей 0, и фиксированным масштабом, определяемым положением числа 1.

2. Плоскость комплексных чисел. Итак, мы хотим построить поле \mathbb{C} , элементы которого были бы точками плоскости \mathbb{R}^2 , а сложение и умножение точек, подчиняясь всем правилам операций в поле, решали бы нашу задачу. Выберем на декартовой плоскости

прямоугольную систему координат с осью абсцисс x и осью ординат y . Будем писать (a, b) для точки с абсциссой a и ординатой b . Для точек (a, b) и (c, d) определим сумму и произведение по правилам

$$\begin{aligned}(a, b) + (c, d) &= (a + c, b + d), \\ (a, b)(c, d) &= (ac - bd, ad + bc)\end{aligned}\tag{6}$$

(использование тех же знаков $+$, \cdot , что и в поле \mathbb{R} , не должно приводить к путанице). Прямая, но довольно утомительная проверка убедила бы нас в том, что так определённые операции наделяют множество пар (точек плоскости) строением поля с нужными свойствами. В этой проверке, к счастью, нет необходимости. Сопоставление

$$(a, b) \mapsto \begin{vmatrix} a & b \\ -b & a \end{vmatrix}$$

точкам плоскости \mathbb{C} элементов построенного ранее поля P и беглый взгляд на формулы (3) и (6) убеждают нас в том, что мы имеем дело с изоморфизмом и что, следовательно, множество \mathbb{C} является полем. Оно и называется обычно *полем комплексных чисел*. Имея в виду геометрическую реализацию этого поля, \mathbb{C} называют ещё *плоскостью комплексных чисел* (а чаще, хотя и несколько двусмысленно, — *комплексной плоскостью*).

Выбранная нами ось абсцисс, т.е. множество точек $(a, 0)$, ничем не отличается по своим свойствам от вещественной прямой, и мы полагаем $(a, 0) = a$. Нуль $(0, 0)$ и единица $(1, 0)$ поля становятся при этом обычными вещественными числами. Для точки $(0, 1)$ на оси ординат вводится, со времён Эйлера и Гаусса, обозначение i “мнимой единицы”, являющейся корнем уравнения (1): $i^2 = (0, 1)(0, 1) = (-1, 0) = -1$. Произвольное комплексное число $z = (x, y) = (x, 0) + (0, 1)(y, 0)$ записывается теперь в традиционном виде

$$z = x + iy, \quad x, y \in \mathbb{R}, \tag{7}$$

весьма близком к виду (5) элементов поля P . Заметим, что $\mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$. Поэтому \mathbb{C} — поле нулевой характеристики (см. п. 5 § 3 гл. 4).

3. Геометрическое истолкование действий с комплексными числами. Ось абсцисс плоскости \mathbb{C} обычно называется *вещественной* (или *действительной*) осью, ось ординат — *мнимой* осью, а числа iy , лежащие на ней, — *чисто мнимыми* числами, хотя слово “мнимое” и утратило свой первоначальный смысл. Соответственно в записи (7) $x = \operatorname{Re} z$ — *вещественная часть*, а $y = \operatorname{Im} z$ — *мнимая часть* комплексного числа z . Рассмотрим отображение, которое сопоставляет каждому комплексному числу $z = x + iy$ комплексно

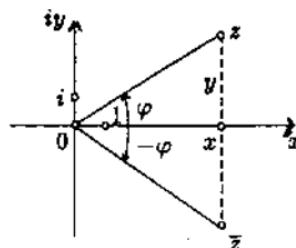


Рис. 18

сопряжённое с ним число $\bar{z} = x - iy$ (*операция комплексного сопряжения*). Геометрически оно сводится к отражению плоскости \mathbb{C} относительно вещественной оси (рис. 18). Весьма примечательно, что справедлива

Теорема 1. *Отображение $z \mapsto \bar{z}$ является автоморфизмом порядка 2 поля \mathbb{C} , оставляющим на месте все вещественные числа. Сумма и произведение комплексно сопряжённых чисел являются вещественными числами.*

Доказательство. Утверждение $\bar{x} = x$, $x \in \mathbb{R}$, очевидно из определения комплексно сопряжённого числа. В частности, $\bar{0} = 0$ и $\bar{1} = 1$. Столь же очевидно утверждение о порядке: $\overline{(\bar{z})} = z$. Нам остаётся проверить соотношения

$$\overline{z_1 + z_2} = \bar{z}_1 + \bar{z}_2, \quad \overline{z_1 z_2} = \bar{z}_1 \bar{z}_2, \quad (8)$$

но они прямо следуют из формул (6), которые нужно только переписать в виде

$$(x_1 + iy_1) + (x_2 + iy_2) = (x_1 + x_2) + i(y_1 + y_2), \\ (x_1 + iy_1) \cdot (x_2 + iy_2) = (x_1 x_2 - y_1 y_2) + i(x_1 y_2 + x_2 y_1). \quad (9)$$

Частным случаем формул (9) является утверждение о сумме и произведении числа $z = x + iy$ и комплексно сопряжённого с ним числа \bar{z} : $z + \bar{z} = 2x$, $z\bar{z} = x^2 + y^2$. \square

Замечание. Автоморфизм $z \mapsto \bar{z}$ выделяется среди многих других автоморфизмов поля \mathbb{C} тем, что он — единственный непрерывный автоморфизм (переводящий близкие точки плоскости \mathbb{C} в близкие). Мы не уточняем и не доказываем это утверждение.

Модулем комплексного числа $z = x + iy$ называется неотрицательное вещественное число $|z| = \sqrt{z\bar{z}} = \sqrt{x^2 + y^2}$. Положение точки z на плоскости, как известно, вполне определяется заданием её полярных координат: расстояния $r = |z|$ от начала координат до z и угла φ между положительным направлением оси абсцисс и направлением из начала координат на z (см. рис. 18). Угол φ называется *аргументом* числа z и обозначается $\arg z = \varphi$. По определению $\arg z$ может принимать любые положительные и отрицательные значения, но при заданном r углы, отличающиеся на целое кратное 2π , соответствуют одному и тому же числу. Аргумент не определён для числа 0 с модулем $|0| = 0$.

Отношения “больше” или “меньше” бессмысленны в применении к комплексным числам, т.е. их нельзя соединять знаком неравенства: в отличие от вещественных чисел, аргумент которых принимает лишь два главных значения — 0 (положительные числа) и π (отрицательные числа), — комплексные числа не упорядочены.

Более точно, на \mathbb{C} не существует отношения $>$ со свойствами:

- если $z \in \mathbb{C}$, то $z > 0$, $z = 0$ или $-z > 0$;
- из $u > 0$, $v > 0$ следует, что $u + v > 0$ и $uv > 0$.

Действительно, в противном случае из $z \neq 0$ следовало бы (как и в \mathbb{R}) $z^2 > 0$. В частности, $1^2 > 0$, $i^2 > 0$ и согласно ii) $0 = i^2 + 1 > 0$ — противоречие.

Полярные координаты r и φ определяют x и y по известным формулам

$$x = r \cos \varphi, \quad y = r \sin \varphi, \quad z = r(\cos \varphi + i \sin \varphi). \quad (10)$$

Это — так называемая *тригонометрическая форма* числа z .

Операция сложения комплексных чисел z, z' просто выражается в декартовых координатах, а именно по правилу параллелограмма, или, что равносильно, по правилу сложения направленных отрезков (векторов), выходящих из начала координат и соответствующих числом z, z' (рис. 19). Из рис. 19, сравнивая стороны треугольника с вершинами в точках $0, z$ и $z + z'$ (и отождествляя модули комплексных чисел с соответствующими геометрическими длинами), получаем важное неравенство

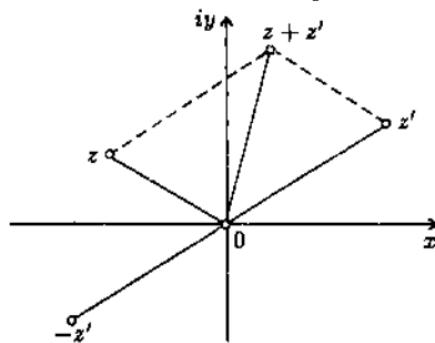


Рис. 19

$$|z + z'| \leq |z| + |z'|. \quad (11)$$

Заметим, что неравенство (11), которое можно было бы записать в более общей форме

$$|z| - |z'| \leq |z \pm z'| \leq |z| + |z'|,$$

совершенно аналогично соответствующему неравенству для вещественных чисел.

Операция умножения комплексных чисел удобно выражается в полярных координатах.

Теорема 2. Модуль произведения комплексных чисел z, z' равен произведению модулей, а аргумент — сумме аргументов множителей:

$$|zz'| = |z| \cdot |z'|, \quad \arg zz' = \arg z + \arg z'. \quad (12)$$

Аналогично,

$$\left| \frac{z}{z'} \right| = \frac{|z|}{|z'|}, \quad \arg \frac{z}{z'} = \arg z - \arg z'.$$

Доказательство. Действительно, пусть тригонометрической формой (10) для z и z' будет

$$z = r(\cos \varphi + i \sin \varphi), \quad z' = r'(\cos \varphi' + i \sin \varphi').$$

Непосредственным умножением или же по формуле (9) получаем

$$zz' = rr'[(\cos \varphi \cos \varphi' - \sin \varphi \sin \varphi') + i(\cos \varphi \sin \varphi' + \sin \varphi \cos \varphi')],$$

а это соотношение при помощи известных формул приводит к тригонометрической форме числа zz' :

$$zz' = |z| \cdot |z'| \cdot [\cos(\varphi + \varphi') + i \sin(\varphi + \varphi')].$$

Если, далее, $z'' = z/z'$, то $z = z'z''$. Поэтому, используя доказанные

формулы (12) для произведения $z'z''$, мы получим из них формулы для дроби z/z' . \square

В частности,

$$z^{-1} = |z|^{-1}[\cos(-\varphi) + i \sin(-\varphi)].$$

Чтобы получить z^{-1} на комплексной плоскости (рис. 20), надо, следовательно, применить к z инверсию относительно окружности единичного радиуса с центром в 0 (это даст точку z'), а затем — отражение относительно вещественной оси (или автоморфизм $z' \mapsto \bar{z}'$).

Фактически утверждения о модуле произведения и модуле суммы легко вытекают без обращения к геометрической интуиции из теоремы 1. В самом деле, во-первых,

$$|zz'|^2 = zz' \overline{zz'} = zz' \overline{zz'} = z \bar{z} z' \bar{z}' = |z|^2 |z'|^2,$$

откуда $|zz'| = |z| \cdot |z'|$. Далее, заметив, что $|z| = \sqrt{x^2 + y^2} \geq \sqrt{x^2} = |x|$, мы получаем

$$\begin{aligned} |1+z|^2 &= (1+z)(1+\bar{z}) = 1 + (z+\bar{z}) + z\bar{z} = \\ &= 1 + 2x + |z|^2 \leq 1 + 2|z| + |z|^2 = (1+|z|)^2, \end{aligned}$$

откуда $|1+z| \leq 1+|z|$. Если теперь $z \neq 0$ и $z' \neq 0$, то

$$\begin{aligned} |z+z'| &= |z(1+z^{-1}z')| = |z| \cdot |1+z^{-1}z'| \leq \\ &\leq |z| \cdot (1+|z^{-1}z'|) = |z|(1+|z|^{-1}|z'|) = |z| + |z'|. \end{aligned}$$

Из полученных результатов мы можем извлечь некий общий принцип: обычная форма (7) комплексных чисел приспособлена к выражению их аддитивных свойств, а тригонометрическая форма (10) — к

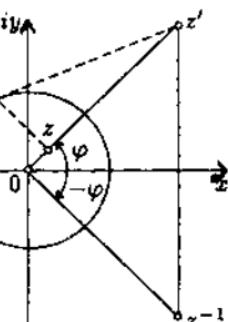


Рис. 20

выражению мультипликативных свойств. Нарушение этого принципа приводит к чрезвычайно сложным формулам, затуманивающим суть дела.

4. Возвведение в степень и извлечение корня. Из формулы (12) для умножения комплексных чисел, заданных в тригонометрической форме, вытекает *формула Муавра*

$$[r(\cos \varphi + i \sin \varphi)]^n = r^n (\cos n\varphi + i \sin n\varphi), \quad (13)$$

справедливая для всех $n \in \mathbf{Z}$ (в иной записи $|z^n| = |z|^n$, $\arg z^n = n \cdot \arg z$). Частный случай формулы (13) при $r = 1$, биномиальная формула (1) из § 7 гл. I и соотношения

$$i^2 = -1, \quad i^3 = -i, \quad i^4 = 1, \quad i^{4k+l} = i^l$$

дают возможность получить выражения для синусов и косинусов кратного угла:

$$\begin{aligned} \cos n\varphi &= \sum_{k \geq 0} (-1)^k \binom{n}{2k} \cos^{n-2k} \varphi \cdot \sin^{2k} \varphi, \\ \sin n\varphi &= \sum_{k \geq 0} (-1)^k \binom{n}{2k+1} \cos^{n-1-2k} \varphi \sin^{2k+1} \varphi. \end{aligned} \quad (14)$$

Справедливи ради стоит заметить, что частным случаем формул (14) при $n = 2$ мы воспользовались ранее — в ходе доказательства теоремы 2.

Замечание. Пусть $e^\alpha = \lim_{n \rightarrow \infty} (1 + \alpha/n)^n$. В анализе, путём разложения функции комплексной переменной в степенные ряды, доказывается *формула Эйлера*

$$e^{i\varphi} = \cos \varphi + i \sin \varphi, \quad (15)$$

из которой вытекают все полученные нами результаты. Стоит только заметить, что

$$e^{i\varphi} e^{i\varphi'} = e^{i(\varphi+\varphi')}, \quad (e^\varphi)^n = e^{in\varphi}.$$

Тригонометрическая форма комплексного числа z сводится к записи

$$z = |z| \cdot e^{i\varphi}.$$

Далее, мы хотели бы научиться извлекать корни произвольной степени из комплексных чисел, и основной вопрос, который здесь возникает: всегда ли это можно делать? Оказывается, что всегда, и формула Муавра даёт по существу полное решение этого вопроса. Пусть нам дано комплексное число $z = r(\cos \varphi + i \sin \varphi)$, а мы хотим найти число $z' = r'(\cos \varphi' + i \sin \varphi')$ такое, что $(z')^n = z$. Выражая $(z')^n$ по формуле Муавра, а затем сравнивая в обеих частях равенства $(z')^n = z$ модули и аргументы, мы находим $(r')^n = r$ и $n\varphi' = \varphi + 2\pi k$

(слагаемое $2\pi k$ — плата за неполную определённость аргумента). Итак,

$$r' = \sqrt[n]{r}, \quad \varphi' = \frac{\varphi + 2\pi k}{n}$$

(под $\sqrt[n]{r}$ подразумевается арифметическое значение корня n -й степени из положительного вещественного числа). Корень $\sqrt[n]{z}$, стало быть, существует, но определён неоднозначно. При $k = 0, 1, \dots, n-1$ для z' будет получено n различных значений, причём ими исчерпываются все корни, поскольку из $k = nq + r$, $0 \leq r \leq n-1$, следует

$$\varphi' = \frac{\varphi + 2\pi r}{n} + 2\pi q.$$

Нами доказана

Теорема 3. Извлечение корня n -й степени из комплексного числа $z = |z|(\cos \varphi + i \sin \varphi)$ всегда возможно. Все n значений корня n -й степени из z расположены в вершинах правильного n -угольника, вписанного в окружность с центром в нуле и радиусом $\sqrt[n]{|z|}$:

$$\sqrt[n]{z} = \sqrt[n]{|z|} \left(\cos \frac{\varphi + 2\pi k}{n} + i \sin \frac{\varphi + 2\pi k}{n} \right), \quad (16)$$

$k = 0, 1, \dots, n-1$.

Следствие. Корни n -й степени из 1 выражаются формулой

$$\sqrt[n]{1} = \varepsilon_k = \cos \frac{2\pi k}{n} + i \sin \frac{2\pi k}{n}, \quad (17)$$

$k = 0, 1, \dots, n-1$. Они расположены в вершинах правильного n -угольника, вписанного в окружность с центром в нуле и радиусом 1.

Из (16) и (17) непосредственно видно, что вещественных корней

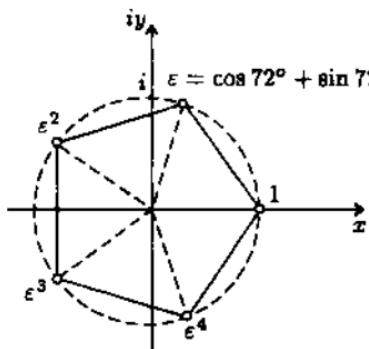


Рис. 21

$\sqrt[n]{z}$ будет нуль, один или два, а корней $\sqrt[n]{1}$ — один или два (на рис. 21 показаны корни из 1 степени 5).

Корень n -й степени из 1 называется **примитивным** (или **первообразным**), если он не является корнем из 1 никакой меньшей степени. Таковыми будут, например,

$$\varepsilon = \varepsilon_1 = \cos \frac{2\pi}{n} + i \sin \frac{2\pi}{n}, \quad \varepsilon_{n-1}.$$

Любой другой корень ε_k является степенью примитивного

$$\varepsilon_k = \varepsilon_1^k,$$

что опять-таки видно из формулы Муавра. Более того, $\varepsilon_k \varepsilon_l = \varepsilon_{k+l}$, если $k + l$ брать по модулю n . В частности, $\varepsilon_{n-k}^{-1} = \varepsilon_{n-k}$, $\varepsilon_0 = 1$. Уже

искушённые в теории групп, мы замечаем, таким образом, что корни n -й степени из 1 составляют циклическую группу $\langle \epsilon \rangle$ порядка n .

Тем самым получена ещё одна реализация циклической группы порядка n . Для каждого $d \mid n$ в $\langle \epsilon \rangle$ имеется ровно одна подгруппа $\langle \epsilon^{n/d} \rangle$ порядка d . Корень ϵ_m будет примитивным тогда и только тогда, когда $\langle \epsilon_m \rangle = \langle \epsilon \rangle$, т.е. $\text{Card}(\langle \epsilon^m \rangle) = n$, а это возможно только при m , взаимно простым с n . Например, при $n = 12$ примитивными корнями будут $\epsilon, \epsilon^5, \epsilon^7, \epsilon^{11}$. В случае простого $n = p$ все корни из единицы, отличные от 1, примитивные. С алгебраической точки зрения, без учёта геометрического изображения, все примитивные корни данной степени n равноправны.

Возвращаясь к вопросу об извлечении корня степени n из произвольного комплексного числа $z \neq 0$, заметим, что если z' — какой-нибудь фиксированный корень (скажем, $z' = \sqrt[n]{|z|}(\cos \frac{\varphi}{n} + i \sin \frac{\varphi}{n})$), то все другие корни имеют вид $z' \epsilon_k$, $k = 0, 1, \dots, n - 1$. Это утверждение находится в соответствии с формулой (16).

5. Теорема единственности. Преимущество поля \mathbb{C} перед \mathbb{R} мы сможем оценить полностью лишь впоследствии, но уже один тот факт, что \mathbb{C} содержит все корни из 1, оправдывает повышенный интерес к комплексным числам. Заметим, что по построению \mathbb{C} — двумерное векторное пространство над \mathbb{R} (в смысле определения из п. 2 § 1 гл. 2) с базисными элементами $1, i$: $\mathbb{C} = \langle 1, i \rangle_{\mathbb{R}}$.

Возникает естественный вопрос, насколько широко семейство полей, обладающих аналогичными свойствами. Оказывается, справедлива следующая теорема единственности поля комплексных чисел.

Теорема 4. Каждое ассоциативное коммутативное кольцо K с единицей 1 без делителей нуля, являющееся двумерным векторным пространством над \mathbb{R} , изоморфно полю \mathbb{C} .

Доказательство. Без ограничения общности отождествим $1 \cdot \mathbb{R}$ с \mathbb{R} и считаем \mathbb{R} вложенным в K . Так как $\dim_{\mathbb{R}} K = 2$, то существует $e \in K \setminus \mathbb{R}$ такой, что 1 и e составляют базис пространства K над \mathbb{R} . Очевидно, $e^2 = \alpha \cdot 1 + 2\beta \cdot e$ с $\alpha, \beta \in \mathbb{R}$. Для элемента $f = e - \beta \notin \mathbb{R}$ имеем $f^2 = \gamma$, где $\gamma = \alpha + \beta^2 \in \mathbb{R}$. Очевидно, $\gamma < 0$, поскольку иначе $\sqrt{\gamma} \in \mathbb{R}$, и мы имели бы $f = \pm \sqrt{\gamma}$. Таким образом, существует $\delta \in \mathbb{R}$, для которого $\delta^2 = -\gamma^{-1}$. Теперь $j^2 = -1$ для $j = \delta f$, и легко проверяется (как при построении \mathbb{C}), что каждый ненулевой элемент из K обратим, т.е. K — поле. Отображение $\varphi: \mathbb{C} \rightarrow K$, определённое соотношением $x + iy \mapsto x + jy$, является искомым изоморфизмом полей. \square

Где в этом доказательстве мы использовали условие, что K — кольцо без делителей нуля? Во-первых, могло бы случиться так, что $e^2 = 0$, и тогда $\alpha = \beta = 0 \Rightarrow \gamma = 0$. Далее, фактически утверждается, что $\gamma \geq 0 \Rightarrow f = \pm \sqrt{\gamma}$. Это действительно так, поскольку

$$0 = f^2 - \gamma = (f - \sqrt{\gamma})(f + \sqrt{\gamma}) \Rightarrow f - \sqrt{\gamma} = 0 \text{ или } f + \sqrt{\gamma} = 0.$$

В поле \mathbb{C} , кроме \mathbb{Q} и \mathbb{R} , содержится много других подполей. Особенно интересны расширения поля \mathbb{Q} , получающиеся присоединением какого-либо элемента из \mathbb{C} , не содержащегося в \mathbb{Q} .

Пример 1 (квадратичное поле). Пусть d — отличное от нуля целое число, возможно, отрицательное, такое, что $\sqrt{d} \notin \mathbb{Q}$. Поле $\mathbb{Q}(\sqrt{d}) \subset \mathbb{C}$ называется *вещественным квадратичным* при $d > 0$ и *мнимым квадратичным* при $d < 0$. О поле $\mathbb{Q}(\sqrt{2})$ упоминалось в § 3 гл. 4. Рассуждение, дословно повторяющее ход доказательства теоремы 4, если заменить там j на \sqrt{d} , а соотношение $j^2 = -1$ на $(\sqrt{d})^2 = d$, показывает, что

$$\mathbb{Q}(\sqrt{d}) = \{a + b\sqrt{d} \mid a, b \in \mathbb{Q}\}.$$

В частности,

$$\begin{aligned} (a_1 + b_1\sqrt{d}) + (a_2 + b_2\sqrt{d}) &= (a_1 + a_2) + (b_1 + b_2)\sqrt{d}, \\ (a_1 + b_1\sqrt{d})(a_2 + b_2\sqrt{d}) &= (a_1a_2 + b_1b_2d) + (a_1b_2 + a_2b_1)\sqrt{d}. \end{aligned} \tag{18}$$

Далее,

$$(a + b\sqrt{d})^{-1} = \frac{a}{a^2 - db^2} + \frac{-b}{a^2 - db^2}\sqrt{d}$$

при $a + b\sqrt{d} \neq 0$ (т.е. при a и b , одновременно не равных нулю).

Пользуясь (18), легко проверить, что отображение

$$f : a + b\sqrt{d} \mapsto a - b\sqrt{d}$$

является автоморфизмом поля $\mathbb{Q}(\sqrt{d})$ (аналог комплексного сопряжения).

Нормой числа $\alpha = a + b\sqrt{d}$ называется число

$$N(\alpha) = a^2 - db^2 = \alpha f(\alpha).$$

Очевидно, что $N(\alpha) = 0 \iff \alpha = 0$. Далее, так как f — автоморфизм, то

$$N(\alpha\beta) = \alpha\beta f(\alpha\beta) = \alpha\beta f(\alpha)f(\beta) = \alpha f(\alpha) \cdot \beta f(\beta) = N(\alpha) \cdot N(\beta).$$

В частности, $N(\alpha) \cdot N(\alpha^{-1}) = N(\alpha\alpha^{-1}) = N(1) = 1$. Поэтому норма обладает существенными свойствами (квадрата) модуля в поле \mathbb{C} .

6. Элементарная геометрия комплексных чисел. Вещественное векторное пространство $\mathbb{C} = \langle 1, i \rangle_{\mathbb{R}}$ является евклидовым: оно снабжено положительно определенным скалярным произведением

$$(z_1|z_2) = \operatorname{Re} z_1 \bar{z}_2 = x_1 x_2 + y_1 y_2,$$

где $z_k = x_k + iy_k$, $k = 1, 2$.

Справедливо неравенство Коши—Буняковского—Шварца

$$|(z_1|z_2)| \leq |z_1| \cdot |z_2|,$$

поскольку $|(z_1|z_2)| = |\operatorname{Re} z_1 \bar{z}_2| \leq |z_1 \bar{z}_2| = |z_1||\bar{z}_2| = |z_1||z_2|$.

Два вектора (комплексных числа) z_1, z_2 называются *ортогональными* или *перпендикулярными* друг другу, если $(z_1|z_2) = 0$.

Из соотношения (12) непосредственно вытекает, что *два вектора* $z, cz \in \mathbb{C}^*$ *ортогональны в точности тогда, когда* c — *чисто мнимое число*.

Прямая, проходящая через точки $u, v \in \mathbb{C}$, задаётся параметрически

$$w = u + (v - u)t, \quad t \in \mathbb{R}.$$

Поэтому ортогональность двух прямых $w = u + (v - u)t$, $w' = u' + (v' - u')t$ выражается соотношением $(v - u)(v' - u') = 0$. Ясно также, что три точки $z_1, z_2, z_3 \in \mathbb{C}$, $z_1 \neq z_2$, лежат на одной прямой в точности тогда, когда

$$\frac{z_3 - z_1}{z_2 - z_1} \in \mathbb{R}, \quad (19)$$

т.е. $z_3\bar{z}_2 - z_3\bar{z}_1 - z_1\bar{z}_2 \in \mathbb{R}$.

Вот маленькое рассуждение на тему ортогональности. Если расположить произвольный треугольник так, чтобы его две вершины α, β оказались на вещественной оси, а третья вершина $i\gamma$ — на мнимой, то легко проверяется, что *три высоты треугольника пересекаются в общей точке $i\delta$* , где $\delta = -\alpha\beta/\gamma$. Например, $(-\alpha + i\delta) - \beta + i\gamma = 0$ (рис. 22).

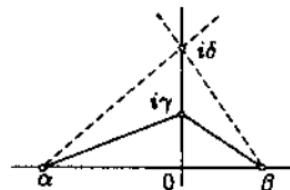


Рис. 22

Важную роль во многих геометрических вопросах играет понятие *двойного отношения* $[z_1, z_2, z_3, z_4]$ четырёх точек $z_1, z_2, z_3, z_4 \in \mathbb{C}$ с $z_1 \neq z_4, z_2 \neq z_3$ (детали см. в [ВА II]). По определению

$$\begin{aligned} [z_1, z_2, z_3, z_4] &= \frac{z_1 - z_2}{z_1 - z_4} : \frac{z_3 - z_2}{z_3 - z_4} = \\ &= \frac{(z_1 - z_2)(z_3 - z_4)}{(z_1 - z_4)(z_3 - z_4)} = \frac{(z_1 - z_2)(z_3 - z_4)(\bar{z}_1 - \bar{z}_4)(\bar{z}_3 - \bar{z}_2)}{|z_1 - z_4|^2 \cdot |z_3 - z_4|^2} \end{aligned} \quad (20)$$

— комплексное число, зависящее от порядка в последовательности z_1, z_2, z_3, z_4 . При циклической перестановке имеем

$$[z_2, z_3, z_4, z_1] = [z_1, z_2, z_3, z_4]^{-1}.$$

Заметим, что в соответствии с (20) двойное отношение не меняется при сдвигах $T_a : z \mapsto z + a$. Представим себе, что три точки z_1, z_2, z_3 не лежат на одной прямой. Это свойство тоже инвариантно относительно сдвигов. Поэтому центр окружности, в которую вписан треугольник с заданными вершинами z_1, z_2, z_3 , можно считать (при вычислении $[z_1, z_2, z_3, z_4]$) расположенным в начале координат. Но тогда $|z_1| = |z_2| = |z_3|$, и легко убедиться в том, что

$$\begin{aligned} (z_1 - z_2)(z_3 - z_4)(\bar{z}_1 - \bar{z}_4)(\bar{z}_3 - \bar{z}_2) - \\ - i(|z_3|^2 - |z_4|^2) \cdot \operatorname{Im}(z_3\bar{z}_2 - z_3\bar{z}_1 - z_1\bar{z}_2) \in \mathbb{R} \end{aligned}$$

(рекомендуем читателю проделать это в качестве упражнения). Согласно (19) должно выполняться условие $\operatorname{Im}(z_3\bar{z}_2 - z_3\bar{z}_1 - z_1\bar{z}_2) \neq 0$,

а в таком случае произведение $(z_1 - z_2)(z_3 - z_4)(\bar{z}_1 - \bar{z}_4)(\bar{z}_3 - \bar{z}_2)$ будет вещественно, или, что эквивалентно (см. (20)), $[z_1, z_2, z_3, z_4] \in \mathbb{R}$ тогда и только тогда, когда $|z_3|^2 - |z_4|^2 = 0$, т.е. $|z_3| = |z_4|$. Значит, z_k , $1 \leq k \leq 4$, — числа, равные по модулю и, стало быть, лежащие на одной окружности.

То же рассуждение действует и тогда, когда на одной прямой не лежат какие-то другие три точки из четырёх. Достаточно заметить, что вещественность $[z_1, z_2, z_3, z_4]$ сохраняется при циклической перестановке. Мы доказали следующее утверждение.

Теорема 5. Четыре точки $z_1, z_2, z_3, z_4 \in \mathbb{C}$ с $z_1 \neq z_4$, $z_2 \neq z_3$, не лежащие на одной прямой, лежат на одной окружности в точности тогда, когда их двойное отношение вещественно.

Это лишь одна из многих конфигураций, свойства которых выражаются на языке двойных отношений.

В заключение мы построим геометрическими средствами новые числовые поля, занимающие видное место в истории математики.

Пример (конструктивные числовые поля). На декартовой плоскости \mathbb{R}^2 считаем заданными точки $(0, 0)$ и $(1, 0)$. Все последующие конструкции осуществляются только при помощи циркуля и линейки. Построив точки P и Q , мы, естественно, можем считать построенным и соединяющий их отрезок PQ . Если построены точка P и отрезок r , то строится также окружность радиуса r с центром в точке P . Попарные пересечения уже построенных прямых (отрезков) и окружностей конструктивны в том же смысле.

Комплексное число $a + ib$ называется **конструктивным**, если при помощи конечной последовательности указанных выше (допустимых) конструкций мы можем построить, отправляясь от $(0, 0)$ и $(1, 0)$, точку $P = (a, b)$. Нетрудно видеть, что конструктивность $a + ib$ эквивалентна конструктивности $|a|$ и $|b|$. Множество точек плоскости, которые строятся при помощи циркуля и линейки, а следовательно, и множество всех конструктивных комплексных чисел обозначим CS .

Теорема 6. Множество CS является подполем поля \mathbb{C} .

Доказательство. Непосредственно из определения конструктивности чисел следует замкнутость CS относительно операции сложения (точка $z + z'$ строится как пересечение двух окружностей (радиуса $|z|$ с центром в z' и радиуса $|z'|$ с центром в z) и перехода от $z = x + iy \in CS$ к $-z = -x - iy$.

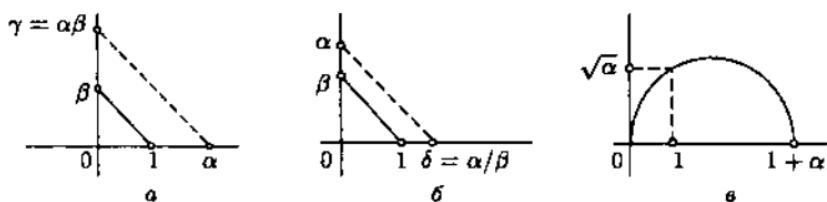


Рис. 23

Откладывая на осях координат отрезки конструктивных длин

$1, \alpha, \beta$ и рассматривая подобные треугольники, изображённые на рис. 23, а, б (штрихами показаны новые конструктивные отрезки), мы убеждаемся в конструктивности произведения $\gamma = \alpha\beta$ и частного $\delta = \alpha/\beta$. Так как построение

$$zz' = (x + iy)(x' + iy') = (xx' - yy') + i(xy' + x'y),$$

$$\frac{1}{z} = \frac{x}{x^2 + y^2} - i \frac{y}{x^2 + y^2}$$

сводится в конечном счёте к построению величин типа γ и δ , то конструктивность произведения zz' и частного $1/z$ также установлена. Вместе с тем доказана замкнутость множества CS относительно всех операций в поле C . \square

Замечание. 1) CS инвариантно относительно автоморфизма сопряжения $z \mapsto \bar{z}$.

2) На рис. 23, в показано, что извлечение квадратного корня $\sqrt{\alpha}$ из конструктивного вещественного числа $\alpha > 0$ конструктивно. На самом деле это высказывание относится к любому конструктивному числу z .

Всякое подполе $F \subset CS$ принято называть *конструктивным числовым полем*. Понятно, что $Q \subset CS$ и что любое конструктивное поле является полем нулевой характеристики. Согласно замечанию 2) всякое квадратичное поле (см. пример в п. 5) конструктивно.

УПРАЖНЕНИЯ

1. Найти все комплексные числа z , по модулю равные 1, при которых $z^2 + (1+i)z$ принимает чисто мнимые значения. Изобразить соответствующее геометрическое место точек на плоскости C .

2. Что можно сказать о поле $R(\delta)$, которое получено из R присоединением комплексного числа δ , удовлетворяющего равенству $\delta^4 = -1$?

3. Пусть $A, B \in M_n(\mathbb{R})$. Опираясь на теорему 1, доказать, что $\det(A + iB) = \det(A - iB)$ (черта означает сопряжение).

4. Пусть $A, B \in M_n(\mathbb{R})$,

$$C = \begin{vmatrix} A & B \\ -B & A \end{vmatrix} \in M_{2n}(\mathbb{R}).$$

Применяя к вещественной матрице C элементарные преобразования первого и второго типа над полем комплексных чисел C , показать, что

$$\det C = |\det(A + iB)|^2.$$

5 (Г. Поля и Г. Сегё). Используя упр. 3 и 4, дать объяснение следующему "странныому" факту. Однородная квадратная линейная система

$$\begin{aligned} d_{11}z_1 + \dots + d_{1n}z_n &= 0, \\ \dots &\dots \\ d_{n1}z_1 + \dots + d_{nn}z_n &= 0 \end{aligned} \tag{*}$$

с комплексными коэффициентами $d_{kl} = a_{kl} + ib_{kl}$ и неизвестными $z_i = x_i + iy_i$ имеет нетривиальное решение (z_1, \dots, z_n) в точности тогда, когда $\det(d_{kl}) =$

$= a + ib = 0$ (см. общие замечания по этому поводу в п. 6 § 3 гл. 4). Это условие приводит к двум уравнениям $a = 0$, $b = 0$, связывающим $2n^2$ вещественных величин a_{kl} , b_{kl} . С другой стороны, систему (*) можно представить в виде системы $2n$ линейных однородных уравнений с $2n$ вещественными неизвестными x_i , y_i . Теперь условие нетривиальности решения залишается в виде равенства нулю одного вещественного определителя размера $2n \times 2n$, что даст лишь одно уравнение между a_{kl} , b_{kl} . Как согласовать между собой эти два результата?

6. Имея в виду, что автоморфизмы квадратичного поля $\mathbb{Q}(\sqrt{d})$ должны оставлять на месте рациональные числа, найти автоморфизмы этого поля.

Ответ. Единичное отображение и $a + b\sqrt{d} \mapsto a - b\sqrt{d}$.

7. Чему равна сумма всех корней степени $n > 1$ из 1? Что можно сказать о сумме примитивных корней степени 12 и степени 15 из 1?

8. Показать, что $\zeta = (2+i)/(2-i)$ не является корнем из 1, хотя $|\zeta| = 1$.

Указание. $\zeta^n = 1 \implies (2-i)^n = (2+i)^n = (2-i+2i)^n = (2-i)^n + \dots + (2i)^n \implies (2-i)(a+bi) = (2i)^n \implies 5(a^2+b^2) = 2^{2n} \implies 5|2^{2n}$ — противоречие.

9. Множество $S^1 = \{e^{i\varphi} | \varphi \in \mathbb{R}\}$ (окружность единичного радиуса) образует относительно умножения в \mathbb{C} подгруппу группы (\mathbb{C}^*, \cdot) . Всякое \mathbb{R} -линейное отображение $f : \mathbb{C} \rightarrow \mathbb{C}$ называется *ортогональным*, если $(f(z)|f(z')) = (z|z')$, т.е. если оно сохраняет длины векторов (расстояния между точками). Доказать, что отображение $f : \mathbb{C} \rightarrow \mathbb{C}$ в точности тогда ортогонально, когда $f(z) = cz$ или $f(z) = c\bar{z}$, где $c \in S^1$.

10. Показать, что

$$\begin{vmatrix} x_0 & x_1 & x_2 & \dots & x_{n-1} \\ x_{n-1} & x_0 & x_1 & \dots & x_{n-2} \\ x_{n-2} & x_{n-1} & x_0 & \dots & x_{n-3} \\ \dots & \dots & \dots & \dots & \dots \\ x_1 & x_2 & x_3 & \dots & x_0 \end{vmatrix} = \prod_{k=0}^{n-1} (x_0 + \zeta^k x_1 + \zeta^{2k} x_2 + \dots + \zeta^{(n-1)k} x_{n-1}),$$

где ζ — примитивный корень степени n из 1.

§ 2. Кольцо многочленов

Наряду с линейными системами, рассмотренными нами в гл. 2 и гл. 3, многочлены составляют старый и хорошо изученный раздел традиционной алгебры. На языке многочленов формулируются или решаются самые различные задачи математики. Тому есть множество причин, и одна из них заключается в свойстве универсальности кольца многочленов, на чём мы коротко остановимся в п. 1.

Пусть K — коммутативное (и, как обычно, ассоциативное) кольцо с единицей 1, A — некоторое его подкольцо, содержащее 1. Если $t \in K$, то наименьшее подкольцо в K , содержащее A и t , будет, очевидно, состоять из элементов вида

$$a(t) = a_0 + a_1 t + a_2 t^2 + \dots + a_n t^n, \quad (*)$$

где $a_s \in A$, $n \in \mathbb{Z}$, $n \geq 0$. Мы обозначим его $A[t]$ и назовем кольцом, полученным из A присоединением элемента t , а выражение (*) — многочленом от t с коэффициентами в A . Что понимать под суммой и произведением многочленов, видно из простейших примеров

(скажем, при $n = 2$):

$$\begin{aligned} a(t) + b(t) &= (a_0 + a_1 t + a_2 t^2) + (b_0 + b_1 t + b_2 t^2) = \\ &= (a_0 + b_0) + (a_1 + b_1)t + (a_2 + b_2)t^2, \end{aligned}$$

$$\begin{aligned} a(t) \cdot b(t) &= a_0 b_0 + (a_0 b_1 + a_1 b_0)t + \\ &\quad + (a_0 b_2 + a_1 b_1 + a_2 b_0)t^2 + (a_1 b_2 + a_2 b_1)t^3 + a_2 b_2 t^4. \end{aligned}$$

Очевидно, что приведение подобных членов основано на попарной перестановочности всех элементов a_i, b_j, t^k .

Теперь настало время вспомнить, что t — наугад взятый элемент кольца K , и поэтому внешние различные выражения (*) могут на самом деле совпадать. Если, скажем, $A = \mathbb{Q}$, $t = \sqrt{2}$, то $t^2 = 2$ и $t^3 = 2t$ — соотношения, которые никоим образом не вытекают из формальных правил. Чтобы прийти к привычному понятию многочлена, необходимо освободиться от всех таких побочных соотношений, для чего под t следует понимать произвольный символ, не обязательно содержащийся в K . Он призван играть чисто вспомогательную роль. Гораздо большее значение имеют правила, по которым составляются коэффициенты выражений $a(t) + b(t)$, $a(t)b(t)$. Имея в виду эти предварительные замечания, перейдём к точному определению алгебраического объекта, называемого многочленом, и собрания таких объектов — кольца многочленов.

1. Многочлены от одной переменной. Пусть A — произвольное коммутативное кольцо с единицей. Построим новое кольцо B , элементами которого являются бесконечные упорядоченные последовательности

$$f = (f_0, f_1, f_2, \dots), \quad f_i \in A, \tag{1}$$

такие, что все f_i , кроме конечного их числа, равны нулю. Определим на множестве B операции сложения и умножения, полагая

$$f + g = (f_0, f_1, f_2, \dots) + (g_0, g_1, g_2, \dots) = (f_0 + g_0, f_1 + g_1, f_2 + g_2, \dots),$$

$$f \cdot g = h = (h_1, h_2, h_3, \dots),$$

где

$$h_k = \sum_{i+j=k} f_i g_j, \quad k = 0, 1, 2, \dots$$

Ясно, что в результате сложения и умножения получаются снова последовательности вида (1) с конечным числом отличных от нуля членов, т.е. элементы из B . Проверка всех аксиом кольца (см. § 3 гл. 4), кроме, разве, аксиомы ассоциативности, очевидна. В самом деле, поскольку сложение двух элементов из B сводится к сложению конечного числа элементов из кольца A , $(B, +)$ является коммутативной группой с нулевым элементом $(0, 0, 0, \dots)$ и элементом $-f =$

$= (-f_0, -f_1, -f_2, \dots)$, обратным к произвольному $f = (f_0, f_1, f_2, \dots)$. Далее, коммутативность умножения следует непосредственно из симметричности выражения элементов h_k через f_i и g_j . Это же выражение показывает, что в B выполнен закон дистрибутивности $(f+g)h = fh + gh$. Что касается ассоциативности операции умножения, то пусть

$$f = (f_0, f_1, f_2, \dots), \quad g = (g_0, g_1, g_2, \dots), \quad h = (h_0, h_1, h_2, \dots)$$

— три произвольных элемента множества B . Тогда $fg = d = (d_0, d_1, d_2, \dots)$, где $d_l = \sum_{i+j=l} f_i g_j$, $l = 0, 1, 2, \dots$, а $(fg)h = dh = e = (e_0, e_1, e_2, \dots)$, где $e_s = \sum_{l+k=s} d_l h_k = \sum_{l+k=s} \left(\sum_{i+j=l} f_i g_j \right) h_k = \sum_{i+j+k=s} f_i g_j h_k$. Вычисление $f(gh)$ даёт тот же результат. Итак, B — коммутативное ассоциативное кольцо с единицей $(1, 0, 0, \dots)$.

Последовательности $(a, 0, 0, \dots)$ складываются и умножаются так же, как элементы кольца A . Это позволяет отождествить такие последовательности с соответствующими элементами из A , т.е. положить $a = (a, 0, 0, \dots)$ для всех $a \in A$. Тем самым A становится подкольцом кольца B .

Обозначим, далее, $(0, 1, 0, 0, \dots)$ через X и назовем X *переменной* (или *неизвестной*) над A . Используя введённую на B операцию умножения, находим, что

$$\begin{aligned} X &= (0, 1, 0, 0, \dots), \\ X^2 &= (0, 0, 1, 0, \dots), \\ &\dots \dots \dots \\ X^n &= (0, 0, \dots, 0, 1, 0, \dots) \end{aligned} \tag{2}$$

Кроме того, ввиду (2) и ввиду включения $A \subset B$ имеем

$$(0, 0, \dots, 0, a, 0, \dots) = aX^n = X^n a.$$

Итак, если f_n — последний отличный от нуля член последовательности $f = (f_0, f_1, \dots, f_n, 0, 0, \dots)$, то в новых обозначениях

$$\begin{aligned} f &= (f_0, \dots, f_{n-1}, 0, 0, \dots) + f_n X^n = \\ &= (f_0, \dots, f_{n-2}, 0, 0, \dots) + f_{n-1} X^{n-1} + f_n X^n = \\ &= f_0 + f_1 X + f_2 X^2 + \dots + f_n X^n. \end{aligned}$$

Такое представление элемента f однозначно, поскольку f_0, \dots, f_n в правой части (3) — это члены последовательности $(f_0, \dots, f_n, 0, \dots)$, которая равна нулю тогда и только тогда, когда $f_0 = \dots = f_n = 0$.

Определение. Введённое выше кольцо B обозначается через $A[X]$ и называется *кольцом многочленов над A* от одной переменной X , а его элементы — *многочленами* (или *полиномами*).

Конечно, присвоение фиксированной букве X названия переменной или неизвестной не очень удачное терминологическое изобретение, но оно привилось, поскольку не приводит к недоразумениям.

Мы намеренно ввели заглавную букву X , чтобы отличить наш специально выделенный многочлен $f = X$ от теоретико-функциональной переменной x , пробегающей какое-то множество значений (чисто временное соглашение, придерживаться которого в будущем не обязательно). Более привычной является запись многочлена f в виде

$$f(X) = a_0 X^n + a_1 X^{n-1} + \dots + a_n,$$

т.е. по убывающим степеням X . В дальнейшем мы будем писать так, как это представляется удобным.

Элементы f_i (и a_i) называются *коэффициентами* многочлена f . Многочлен f *нулевой*, когда все его коэффициенты равны нулю. Коэффициент f_0 при X в нулевой степени называется *постоянным членом*. Если $f_n \neq 0$, то f_n называют *старшим коэффициентом*, а n — *степенью* многочлена и пишут $n = \deg f$. Нулевому многочлену приписывается степень $-\infty$ ($-\infty + (-\infty) = -\infty$, $-\infty + n = -\infty$, $-\infty < n$ для каждого $n \in \mathbb{N}$). Многочлены степени 1, 2, 3, … называются соответственно *линейными*, *квадратичными* (или *квадратными*), *кубичными* и т.д.

Роль единицы кольца $A[X]$ играет единичный элемент 1 кольца A , рассматриваемый как многочлен нулевой степени. Непосредственно из определения операций сложения и умножения в $A[X]$ следует, что для любых двух многочленов

$$f = f_0 + f_1 X + \dots + f_n X^n, \quad g = g_0 + g_1 X + \dots + g_m X^m \quad (4)$$

степеней n и m соответственно имеют место неравенства

$$\deg(f+g) \leq \max(\deg f, \deg g), \quad \deg(fg) \leq \deg f + \deg g. \quad (5)$$

Второе из неравенств (5) на самом деле заменяется равенством

$$\deg(fg) = \deg f + \deg g$$

всякий раз, когда произведение $f_n g_m$ старших коэффициентов многочленов (4) отлично от нуля, поскольку

$$fg = f_0 g_0 + (f_0 g_1 + f_1 g_0)X + \dots + (f_n g_m)X^{n+m}. \quad (6)$$

Но это значит, что верна

Теорема 1. *Если A — целостное кольцо, то и кольцо $A[X]$ является целостным.*

Место кольца многочленов среди коммутативных колец отчасти поясняет следующая

Теорема 2. *Пусть коммутативное кольцо K содержит A в качестве подкольца. Для каждого элемента $t \in K$ существует*

единственный гомоморфизм колец $\Pi_t : A[X] \rightarrow K$ такой, что

$$\forall a \in A \quad \Pi_t(a) = a, \quad \Pi_t(X) = t. \quad (7)$$

Доказательство. Предположим сначала, что такой гомоморфизм Π_t существует. Так как $\Pi_t(f_i) = f_i$ для каждого коэффициента многочлена f , записанного в стандартном виде (3), и $\Pi_t(X^k) = (\Pi_t(X))^k = t^k$ (свойство гомоморфизма и условие (7)), то

$$\Pi_t(f) = \Pi_t(f_0 + f_1 X + \dots + f_n X^n) = f_0 + f_1 X + \dots + f_n t^n, \quad (8)$$

т.е. $\Pi_t(f)$ определён однозначно и выражается формулой (8). Обратно: задав отображение Π_t формулой (8), мы, очевидно, удовлетворим условию (7) и получим гомоморфизм колец. Это ясно для отображения аддитивных групп колец, а что касается умножения, то применение Π_t к произведению (6), а затем использование (общего) закона дистрибутивности даёт

$$\Pi_t(fg) = f_0 g_0 + (f_0 g_1 + f_1 g_0)X + \dots + (f_n g_m)t^{n+m} =$$

$$= \left(\sum_{i=0}^n f_i t^i \right) \left(\sum_{j=0}^m g_j t^j \right) = \Pi_t(f) \cdot \Pi_t(g). \quad \square$$

Результат применения отображения Π_t , определённого формулой (8), к многочлену $f = f(X)$ называется *подстановкой* t в f вместо X или (с некоторой натяжкой) просто значением f при $X = t$, так что $\Pi_t(f) = f(t)$. Знать $\Pi_t(f)$ — значит уметь вычислить значение f при $X = t$. Гомоморфизмы Π_x , $x \in A$, служат связующим звеном между функциональной и алгебраической точками зрения на многочлен. По определению линейный многочлен $X - c = (-c, 1, 0, \dots)$ никогда не равен нулю, но ассоциированная с ним функция $x \mapsto x - c$ принимает нулевое значение при $x = c$. Другой пример: отличный от нуля многочлен $X^2 + X$ с коэффициентами из поля F_2 (где $1 + 1 = 0$) представляет нулевую функцию $\tilde{f} : F_2 \rightarrow F_2$, поскольку $0^2 + 0 = 0$ и $1^2 + 1 = 0$.

Элемент $t \in K$ называется *алгебраическим над* A , если $\Pi_t(f) = 0$ для некоторого $f \in A[X]$. Если же $\Pi_t : A[X] \rightarrow K$ — изоморфное вложение (мономорфизм), то t — *трансцендентный над* A элемент. В случае $A = \mathbb{Q}$ и $K = \mathbb{C}$ говорят просто об *алгебраических и трансцендентных* числах. Например, числа e и π , определяемые в анализе, являются трансцендентными, а числа $\sqrt{2}, \sqrt{3}, \sqrt{2} + \sqrt{3}$ — алгебраическими.

Гомоморфизм Π_t , собственно говоря, служит выражением *универсального свойства* кольца многочленов $A[X]$. Более полным образом универсальность кольца многочленов видна из следующего утверждения, обобщающего теорему 2.

Теорема 3. Пусть A и K — произвольные коммутативные кольца, t — элемент из K и $\varphi: A \rightarrow K$ — гомоморфизм.

Тогда существует, и притом единственное, продолжение φ до гомоморфизма $\varphi_t: A[X] \rightarrow K$ кольца многочленов $A[X]$ в K , переводящего переменную X в t .

Доказательство является незначительным видоизменением доказательства теоремы 2 и оставляется читателю в качестве упражнения. \square

2. Многочлены от многих переменных. Если в ситуации $A \subset K$, рассматривавшейся в начале параграфа, взять произвольные n элементов $t_1, \dots, t_n \in K$ и рассмотреть в K пересечение всех подкольца, содержащих A, t_1, \dots, t_n , то мы получим кольцо $A[t_1, \dots, t_n]$. Формальная запись его элементов подсказывает, как и в случае $n = 1$, необходимость введения в обиход кольца многочленов от n переменных. Делается это очень просто. Вспомним, что конструкция кольца $B = A[X]$ включала произвольное коммутативное кольцо A с единицей. Мы можем теперь заменить в нашей конструкции кольцо A на B и построить кольцо $C = B[Y]$, где Y — новая независимая переменная, играющая по отношению к B ту же роль, что и X по отношению к A . Элементы из C однозначно записываются в виде $\sum b_j Y^j$, $b_j \in B$, причём B отождествляется с подкольцом в C , а именно с множеством элементов $bY^0 = b \cdot 1$. Так как в свою очередь $b_j = \sum a_{ij} X^i$ — однозначная запись элементов $b_j \in B$, то любой элемент из C имеет вид

$$\sum_{i=0}^k \sum_{j=0}^l a_{ij} X^i Y^j, \quad a_{ij} \in A,$$

причём подразумевается (по смыслу конструкции), что a_{ij} перестановочны с X и Y , а переменная X перестановочна с Y . Кольцо C называется *кольцом многочленов над A от двух независимых переменных* (от двух неизвестных) X и Y .

Повторив достаточное число раз эту конструкцию, мы получим кольцо $A[X_1, \dots, X_n]$ многочленов (полиномов) над A от n независимых переменных (или неизвестных) X_1, \dots, X_n .

Набор $(i_1, \dots, i_n) \in \overline{\mathbb{N}}^n$ из n целых неотрицательных чисел i_1, \dots, i_n ($\overline{\mathbb{N}} = \mathbb{N} \cup \{0\}$) условимся сокращённо обозначать символом (i) . Тогда любой элемент $f \in A[X_1, \dots, X_n]$ запишется в виде

$$f = \sum_{(i)} a_{(i)} X^{(i)}, \quad a_{(i)} \in A, \tag{9}$$

где $X^{(i)} = X_1^{i_1} \dots X_n^{i_n}$ — одночлен (или моном), так что f — линейная комбинация одночленов с коэффициентами из A . В соответствии с определением многочленов все коэффициенты $a_{(i)}$ в (9), за исключе-

нием конечного их числа, равны нулю. Единственность записи (9) непосредственно вытекает из следующего утверждения.

Многочлен f равен нулю тогда и только тогда, когда равны нулю все его коэффициенты $a_{i_1 \dots i_n}$. При $n = 1$ это уже отмечалось в ходе построения кольца $A[X]$, а при $n > 1$ проще всего использовать индукцию по n . Именно, мы можем записать

$$f = \sum a_{i_1 \dots i_n} X_1^{i_1} \dots X_n^{i_n} = \sum b_{i_n} X_n^{i_n},$$

где

$$b_{i_n} = \sum_{i_1, \dots, i_{n-1}} a_{i_1 \dots i_{n-1} i_n} X_1^{i_1} \dots X_{n-1}^{i_{n-1}}$$

— многочлены от меньшего числа переменных. Утверждение для $n = 1$ и предположение индукции показывают, что

$$f = 0 \iff \forall i_n \quad b_{i_n} = 0 \iff \forall (i_1, \dots, i_n) \quad a_{i_1 \dots i_{n-1} i_n} = 0.$$

Теперь естественно считать два многочлена $f, g \in A[X_1, \dots, X_n]$ *равными*, если совпадают их коэффициенты при одинаковых одночленах (согласно высказанныму $(i_1, \dots, i_n) \neq (j_1, \dots, j_n) \implies \implies X_1^{i_1} \dots X_n^{i_n} \neq X_1^{j_1} \dots X_n^{j_n}$).

Под *степенью многочлена f относительно X_k* понимается наибольшее целое число, обозначаемое $\deg_k f$, которое встречается в качестве показателя при X_k в $a_{(i)} X^{(i)}$ с $a_{(i)} \neq 0$. Например, многочлен $1 + X + XY^3 + X^2Y^2$ имеет степень 2 относительно X и степень 3 относительно Y .

Целое число $i_1 + \dots + i_n$ называется (*полной*) *степенью одночлена* $X_1^{i_1} \dots X_n^{i_n}$.

Степенью $\deg f$ (или полной степенью) многочлена f будет максимальная из полных степеней его одночленов. Полагаем $\deg 0 = -\infty$. О старшем по степени члене многочлена f не имеет смысла говорить, поскольку таких членов (одночленов) может быть несколько.

На кольцо $A[X_1, \dots, X_n]$ переносятся многие результаты, полученные нами в п. 1 для $A[X]$. Например, опираясь на теорему 1 и используя индукцию по n , мы сразу же убеждаемся в том, что справедлива

Теорема 1'. Если A — целостное кольцо, то и кольцо $A[X_1, \dots, X_n]$ является целостным. В частности, кольцо многочленов от n переменных над любым полем P целостно.

Полезным уточнением теоремы 1' служит

Теорема 4. Пусть f и g — произвольные многочлены от n переменных над целостным кольцом A . Тогда

$$\deg(fg) = \deg f + \deg g.$$

Доказательство. Назовём однородным многочленом или *формой* степени m многочлен $h(X_1, \dots, X_n)$, все члены которого имеют одну и ту же полную степень m . Формы степеней 1, 2, 3 называются соответственно линейной, квадратичной и кубичной формами. Объединяя вместе все входящие в f (или, как ещё говорят, встречающиеся, имеющие ненулевые коэффициенты) одночлены одной и той же степени, мы однозначно представим многочлен $f = \sum a_{(i)} X^{(i)}$ в виде суммы нескольких форм f_m различных степеней

$$f = f_0 + f_1 + \dots + f_k, \quad k = \deg f.$$

Если теперь

$$g = g_0 + g_1 + \dots + g_l, \quad l = \deg g,$$

то, очевидно,

$$fg = f_0g_0 + (f_0g_1 + f_1g_0) + \dots + f_kg_l$$

(это похоже на соотношение (6), но f_i, g_j имеют там другой смысл), откуда $\deg fg \leq k+l$. По теореме 1' из $f_k \neq 0, g_l \neq 0$ следует $f_k g_l \neq 0$, т.е. $\deg(fg) = \deg(f_k g_l) = k+l = \deg f + \deg g$. \square

3. Алгоритм деления с остатком. В п. 3 § 9 гл. 1 для целых чисел был введён алгоритм деления с остатком. Оказывается, что совершенно аналогичный алгоритм имеет место и в кольце $A[X]$ над целостным кольцом A (для $A = \mathbb{R}$ это известно фактически из курса элементарной алгебры: вспомните деление уголком).

Теорема 5. Пусть A — целостное кольцо и g — многочлен в $A[X]$ со старшим коэффициентом, обратимым в A . Тогда каждому многочлену $f \in A[X]$ сопоставляется одна и только одна пара многочленов $q, r \in A[X]$, для которых

$$f = qg + r, \quad \deg r < \deg g. \quad (10)$$

Доказательство. Пусть

$$f = a_0X^n + a_1X^{n-1} + \dots + a_n,$$

$$g = b_0X^m + b_1X^{m-1} + \dots + b_m,$$

где $a_0b_0 \neq 0$ и $b_0|1$. Применим индукцию по n . Если $n = 0$ и $m = \deg g > \deg f = 0$, то положим $q = 0, r = f$, а если $n = m = 0$, то $r = 0$ и $q = a_0b_0^{-1}$. Допустим, что теорема доказана для всех многочленов степени $< n$ ($n > 0$). Без ограничения общности считаем $m \leq n$, поскольку в противном случае возьмём $q = 0$ и $r = f$. Раз это так, то

$$f = a_0b_0^{-1}X^{n-m} \cdot g + \bar{f},$$

где $\deg \bar{f} < n$. По индукции мы можем найти \bar{q} и r , для которых $\bar{f} = \bar{q}g + r$, причём $\deg r < m$. Положив

$$q = a_0b_0^{-1}X^{n-m} + \bar{q},$$

мы придём к паре многочленов с нужными свойствами.

Обращаясь к свойству единственности частного q и остатка r , предположим, что

$$qg + r = f = q'g + r'.$$

Тогда $(q' - q)g = r - r'$. По теореме 1 имеем $\deg(r - r') = \deg(q' - q) + \deg g$, что в наших условиях возможно только при $r' = r$ и $q' = q$ (напомним, что $\deg 0 = -\infty$ и что $-\infty + m = -\infty$).

Наконец, приведённые рассуждения показывают, что коэффициенты частного q и остатка r принадлежат тому же целостному кольцу A , т.е. $f, g \in A[X] \Rightarrow q, r \in A[X]$. \square

Замечание. Многочлены со старшим коэффициентом 1 часто называют нормализованными (ещё нормированными, унитарными). Указанный выше процесс деления многочлена f на g , называемый евклидовым, несколько упрощается, если g — нормализованный многочлен. Говорят, что f делится на g , если остаток r равен нулю (см. (10)): $f = qg$.

УПРАЖНЕНИЯ

1. Многочлены $f(X) = X^6 + 3X^4 + X^3 + 4X^2 - 3X - 1$, $g(X) = X^2 + X + 1$ можно считать принадлежащими кольцу $\mathbb{Z}[X]$ или, скажем, кольцу $\mathbb{Z}_5[X]$ в зависимости от того, как интерпретировать их коэффициенты. Применив алгоритм деления с остатком, показать, что в первом случае $f(X)$ не делится на $g(X)$, а во втором — делится. Возможна ли реализация противоположного варианта?

2. Доказать при помощи теоремы 3, что если F — поле, то группа всех автоморфизмов кольца $F[X]$, тождественных на F , изоморфна группе преобразований $X \mapsto aX + b$, где $a, b \in F$ и $a \neq 0$.

3. Показать, что многочлен $f \in F[X_1, \dots, X_n]$ является формой степени m (см. доказательство теоремы 4) тогда и только тогда, когда $f(tX_1, \dots, tX_n) = t^m f(X_1, \dots, X_n)$, где t — новая переменная.

4. Показать, что число различных одночленов от n независимых переменных полной степени m равно $\binom{m+n-1}{m}$.

Указание. Использовать принцип двойной индукции по n и m , опираясь на соотношение

$$\binom{m+(n-1)-1}{m} + \binom{(m-1)+n-1}{m-1} = \binom{m+n-1}{m}.$$

5. Возвращаясь к определениям п. 1, рассмотрим совокупность $A[[X]]$ так называемых *формальных степенных рядов* $f(X) = \sum_{i \geq 0} a_i X^i$ от переменной (неизвестной) X или, если угодно, последовательностей (a_0, a_1, a_2, \dots) с любым, возможно, бесконечным, числом коэффициентов $a_i \neq 0$, принадлежащих коммутативному кольцу A . Действия с формальными степенными рядами из $A[[X]]$ проводятся по тем же правилам, что и действия с многочленами:

$$\left(\sum a_i X^i \right) + \left(\sum b_i X^i \right) = \sum (a_i + b_i) X^i,$$

$$\left(\sum a_i X^i \right) \cdot \left(\sum b_i X^i \right) = \sum c_k X^k, \quad c_k = \sum_{i+j=k} a_i b_j.$$

Показать, что множество $A[[X]]$, рассматриваемое вместе с этими операциями, является ассоциативным и коммутативным кольцом с единицей $1 = (1, 0, \dots)$.

Так как в степенной ряд $f = \sum a_i X^i$ входят сколь угодно высокие степени X^i переменной X , то вместо степени $\deg f$, не имеющей теперь смысла, естественно рассматривать *порядок* $\omega(f)$ — целое число, равное наименьшему индексу n , для которого $a_n \neq 0$ (полагают ещё $\omega(0) = +\infty$).

Показать, что:

- i) $\omega(f - g) \geq \min\{\omega(f), \omega(g)\}$;
- ii) $\omega(fg) \geq \omega(f) + \omega(g)$.

Если A — целостное кольцо, то $\omega(fg) = \omega(f) + \omega(g)$. В частности, вместе с A целостным является и кольцо $A[[X]]$.

Показать также, что $A[X]$ — подкольцо в $A[[X]]$.

6. Многочлены и степенные ряды часто используются в качестве производящих функций различных числовых величин. Смысл оперирования с ними поясним на двух простых примерах.

а) Установить соотношение

$$\sum_{i=0}^k \binom{m}{i} \binom{n}{k-i} = \binom{m+n}{k},$$

исходя из биномиальной формулы $\sum \binom{n}{i} X^i = (1+X)^n$ в $\mathbb{Z}[X]$ и очевидного разложения $(1+X)^m(1+X)^n = (1+X)^{m+n}$.

б) Найти число l_n всевозможных расстановок скобок в произведении длины n элементов множества с одной бинарной операцией. С этой целью удобно ввести производящую функцию — формальный степенной ряд

$$l(X) = \sum_{n \geq 1} l_n X^n = X + X^2 + 2X^3 + \dots,$$

начальные коэффициенты которого были вычислены ещё в п. 3 § 1 гл. 4. Из очевидного рекуррентного соотношения

$$l_n = \sum_{k=1}^{n-1} l_k l_{n-k}$$

вытекает, что $l(X)^2 = l(X) - X$. Решая это квадратное уравнение, находим

$$l(X) = \frac{1 - \sqrt{1 - 4X}}{2}$$

(знак перед радикалом определяется условием $l_n > 0$). Но если степенной ряд $f(X)$ таков, что $f^r = 1 + \lambda X$, $r \in \mathbb{N}$, то

$$f(X) = 1 + \sum_{k=1}^{\infty} \left[\prod_{i=0}^{k-1} \left(\frac{1}{r} - i \right) \right] \frac{(\lambda X)^k}{k!}$$

(разложение в ряд Тейлора, которое можно принять пока на веру). В нашем случае $r = 2$, $\lambda = -4$, и простая подстановка даёт окончательное выражение

$$l_n = \frac{1}{n} \binom{2n-2}{n-1}$$

(заметим, что $l_n = C_{n-1}$ — классическое число Каталана).

Предлагается провести все промежуточные выкладки.

§ 3. Разложение в кольце многочленов

1. Элементарные свойства делимости. В разных местах, начиная с гл. 1, мы затрагивали вопросы делимости в кольце \mathbb{Z} целых чисел, но так называемая основная теорема арифметики у нас оставалась пока недоказанной. Теперь настала пора не только заполнить этот пробел, но и распространить соответствующие утверждения на более широкий класс колец. В первую очередь нас интересует кольцо многочленов $P[X]$ над полем P .

Начиная с произвольного целостного кольца K . Обратимые элементы в K были названы нами делителями единицы. Часто их называют ещё *регулярными* элементами. Совершенно очевидно, что многочлен $f \in A[X]$ обратим (регулярен) в точности тогда, когда $\deg f = 0$ и $f = f_0$ — обратимый элемент кольца A , поскольку $fg = 1 \Rightarrow \deg f + \deg g = \deg 1 = 0$.

Говорят, что элемент $b \in K$ делится на $a \in K$ (или b кратен a), если существует такой элемент $c \in K$, что $b = ac$ (это обозначается $a|b$). Если $a|b$ и $b|a$, то a и b называются *ассоциированными* элементами. Тогда $b = ua$, где $u|1$. В силу сделанного выше замечания ассоциированность многочленов $f, g \in A[X]$ означает, что они отличаются лишь обратимым множителем из A .

Элемент $p \in K$ называется *простым* (или *неразложимым*), если p необратим и его нельзя представить в виде $p = ab$, где a, b — не обратимые элементы. В поле P каждый ненулевой элемент обратим и в P нет простых элементов. Простой элемент кольца $A[X]$ называется чаще *неприводимым многочленом*.

Отметим следующие основные свойства отношения делимости в целостном кольце K .

1) *Если $a|b$, $b|c$, то $a|c$.* Действительно, мы имеем $b = ab'$, $c = bc'$, где $b', c' \in K$. Поэтому $c = (ab')c' = a(b'c')$.

2) *Если $c|a$ и $c|b$, то $c|(a \pm b)$.* В самом деле, по условию $a = ca'$, $b = cb'$ для некоторых $a', b' \in K$, и ввиду дистрибутивности $a \pm b = c(a' \pm b')$.

3) *Если $a|b$, то $a|bc$.* Ясно, что $b = ab' \Rightarrow bc = (ab')c = a(b'c)$.

Комбинируя 2) и 3), получаем

4) *Если каждый из элементов $b_1, b_2, \dots, b_m \in K$ делится на $a \in K$, то на a будет делиться также элемент $b_1c_1 + b_2c_2 + \dots + b_mc_m$, где c_1, c_2, \dots, c_m — произвольные элементы.*

Определение. Говорят, что целостное кольцо K — *кольцо с однозначным разложением на простые множители* (или K — *факториальное кольцо*), если любой элемент $a \neq 0$ из K можно представить в виде

$$a = up_1p_2 \dots p_r, \quad (1)$$

где u — обратимый элемент, а p_1, p_2, \dots, p_r — простые элементы (не

обязательно попарно различные), причём из существования другого такого разложения $a = vq_1q_2 \dots q_s$, следует, что $r = s$ и при надлежащей нумерации элементов p_i и q_j будет

$$q_1 = u_1 p_1, \dots, q_r = u_r p_r,$$

где u_1, \dots, u_r — обратимые элементы.

Допуская в равенстве (1) значение $r = 0$, мы принимаем соглашение, что обратимые элементы в K тоже имеют разложение на простые множители. Ясно, что если p — простой, а u — обратимый элемент, то ассоциированный с p элемент up тоже простой. В кольце \mathbb{Z} с обратимыми элементами 1 и -1 отношение порядка ($a < b$) даёт возможность выделить *положительное* простое число p из двух возможных простых элементов $\pm p$. В кольце $P[X]$ удобно рассматривать нормализованные (см. замечание в конце § 2) неприводимые многочлены.

Справедлива следующая общая

Теорема 1. Пусть K — произвольное целостное кольцо с разложением на простые множители. Однозначность разложения в K (факториальность K) имеет место тогда и только тогда, когда любой простой элемент $p \in K$, делящий произведение $ab \in K$, делит по крайней мере один из множителей a, b .

Доказательство. Пусть K факториально, и пусть $ab = pc$. Если

$$a = \prod a_i, \quad b = \prod b_j, \quad c = \prod c_k$$

— разложения a, b, c на простые множители, то из равенства $\prod a_i \times \prod b_j = p \prod c_k$ следует, что элемент p ассоциирован с одним из a_i или b_j , т.е. p делит a или b .

Обратно: установим однозначность разложения в K , где $p|ab \Rightarrow p|a$ или $p|b$. Рассуждая по индукции, допустим, что разложение всех элементов из K с числом $\leq n$ простых множителей единственно (конечно, с точностью до порядка множителей и их ассоциированности). Докажем теперь это для любого элемента $a \neq 0$, который может быть разложен на $n+1$ простых множителей. Именно, пусть

$$a = \prod_{i=1}^{n+1} p_i = \prod_{j=1}^{m+1} r_j \tag{2}$$

— два разложения элемента a с $m \geq n$. Условие теоремы, применённое к $p = p_{n+1}$, даёт нам, что p_{n+1} должен делить один из элементов r_1, \dots, r_{m+1} . Без ограничения общности (ибо это вопрос нумерации) считаем, что $p_{n+1}|r_{m+1}$. Но r_{m+1} — простой элемент, поэтому $r_{m+1} = up_{n+1}$, где u — обратимый элемент. Опираясь на закон сокращения в K (теорема 1 из § 3 гл. 4), получаем из (2) равенство $\prod_{i=1}^n p_i =$

$= u \prod_{j=1}^m r_j$. В левой его части стоит произведение n простых множителей. По предположению индукции $m = n$, и оба разложения отличаются лишь порядком простых элементов, снабжённых, возможно, какими-то обратимыми множителями. \square

В произвольном целостном кольце K элемент $a \neq 0$ вообще не обязан допускать разложение типа (1). Что более интересно, имеются целостные кольца, в которых разложение на простые множители хотя и возможно, но не является однозначным, т.е. условие теоремы 1, кажущееся тривиальным, не всегда выполняется.

Пример 1. Рассмотрим мнимое квадратичное поле $\mathbb{Q}(\sqrt{-5})$ (см. пример в п. 5 из § 1), а в нём целостное кольцо $K = \{a + b\sqrt{-5} \mid a, b \in \mathbb{Z}\}$. Норма $N(a + b\sqrt{-5}) = a^2 + 5b^2$ каждого отличного от нуля элемента $a \in K$ — целое положительное число. Если a обратим в K , то $(N(a))^{-1} = N(a^{-1}) \in \mathbb{Z}$, откуда $N(a) = 1$. Это возможно лишь при $b = 0, a = \pm 1$. Таким образом, в K , как и в \mathbb{Z} , обратимыми элементами являются только ± 1 . Если $a = \varepsilon a_1 a_2 \dots a_r \neq 0$, $\varepsilon = \pm 1$, то $N(a) = N(a_1) \dots N(a_r)$. Так как $1 < N(a_i) \in \mathbb{N}$, то при заданном a число множителей r не может неограниченно расти. Стало быть, разложение на простые множители в K возможно.

Вместе с тем число 9 (да и не только оно) допускает два существенно различных разложения на простые множители:

$$9 = 3 \cdot 3 = (2 + \sqrt{-5})(2 - \sqrt{-5}).$$

Неассоциированность элементов 3 и $2 \pm \sqrt{-5}$ очевидна. Далее, $N(3) = N(2 \pm \sqrt{-5}) = 9$. Поэтому из разложения $a = a_1 a_2$ для $a = 3$ или $2 \pm \sqrt{-5}$ с необратимыми a_1, a_2 следовало бы $9 = N(a) = N(a_1)N(a_2)$, т.е. $N(a_i) = 3$, $i = 1, 2$, что невозможно, поскольку уравнение $x^2 + 5y^2 = 3$ с $x, y \in \mathbb{Z}$ неразрешимо. Этим доказана простота элементов 3 и $2 \pm \sqrt{-5}$.

Рассмотренный пример содержит в зародыше большой круг вопросов, частично остающихся пока нерешёнными, о квадратичных полях $\mathbb{Q}(\sqrt{d})$. Их изучение входит в круг вопросов алгебраической теории чисел.

Прежде чем устанавливать при помощи теоремы 1 факториальность тех или иных колец, мы введём важные вспомогательные понятия, представляющие независимый интерес.

2. НОД и НОК в кольцах. Пусть K — целостное кольцо. Под *наибольшим общим делителем* двух элементов $a, b \in K$ мы будем понимать элемент $d \in K$, обозначаемый $\text{НОД}(a, b)$ и обладающий двумя свойствами:

- i) $d|a, d|b;$
- ii) $c|a, c|b \implies c|d.$

Ясно, что вместе с d свойствами i), ii) обладает любой ассоциированный с ним элемент. Обратно: если c и d — два наибольших делителя элементов a и b , то будем иметь $c|d, d|c$, так что c и d ассоциированы. Обозначение $\text{НОД}(a, b)$ относится к любому из них, т.е. в этой записи ассоциированные элементы не различаются. С учётом такого соглашения к определяющим свойствам i), ii) наибольшего общего делителя добавятся следующие:

- iii) $\text{НОД}(a, b) = a \iff a|b;$

- iv) $\text{НОД}(a, 0) = a;$
- v) $\text{НОД}(ta, tb) = t\text{НОД}(a, b);$
- vi) $\text{НОД}(\text{НОД}(a, b), c) = \text{НОД}(a, \text{НОД}(b, c)).$

Проверка их не вызывает никаких трудностей и оставляется читателю. Свойство vi) позволяет также распространить понятие НОД на произвольное конечное число элементов.

По аналогии с $\text{НОД}(a, b)$ вводится дуальное понятие *наименьшего общего кратного* $m = \text{НОК}(a, b)$ элементов $a, b \in K$, также определённого с точностью до ассоциированности двумя свойствами:

- i') $a|m, b|m;$
- ii') $a|c, b|c \Rightarrow m|c.$

В частности, полагая $c = ab$, получаем $m|ab$.

Теорема 2. Пусть для элементов a, b целостного кольца K существуют $\text{НОД}(a, b)$ и $\text{НОК}(a, b)$.

Тогда:

- a) $\text{НОК}(a, b) = 0 \Leftrightarrow a = 0$ или $b = 0$.
- б) $a, b \neq 0, m = \text{НОК}(a, b), ab = dm \Rightarrow d = \text{НОД}(a, b).$

Доказательство. Утверждение а) вытекает непосредственно из определения $\text{НОК}(a, b)$. Для доказательства б) нам нужно убедиться, что элемент d , определённый равенством $ab = dm$, обладает свойствами i), ii). В самом деле, i') $\Rightarrow m = a'a, m = b'b$. Значит, $ab = dm = da'a$, откуда после сокращения на a , допустимого в любом целостном кольце, имеем $b = da'$, т.е. $d|b$. Аналогично, $ab = dm = db'b \Rightarrow a = db'$, т.е. $d|a$. Мы пришли к i).

Далее, пусть $a = fa'', b = fb''$. Положим $c = fa''b''$. Тогда $c = ab'' = ba''$ — общее кратное a и b . Согласно свойству ii') $c = c'm$ для некоторого $c' \in K$, откуда $fc'm = fc = f^2a''b'' = ab = dm$, т.е. $d = fc'$ и $f|d$. Мы пришли к ii). \square

Определение. Элементы a, b целостного кольца, в котором существует НОД, называются *взаимно простыми*, если $\text{НОД}(a, b) = 1$.

Из свойств i), ii), i'), ii') или из теоремы 2 нельзя извлечь ни способа вычисления, ни доказательства существования $\text{НОД}(a, b)$ и $\text{НОК}(a, b)$. Теоремой 2, б) устанавливается лишь соотношение между ними.

Предположим теперь на время, что K — факториальное кольцо. Обозначим через \mathcal{P} множество простых элементов в K такое, что всякий простой элемент из K ассоциирован с одним и только одним элементом из \mathcal{P} . Рассматривая разложения двух элементов $a, b \in K$, удобно считать, что в них входят одинаковые элементы из \mathcal{P} , но некоторые, возможно, с нулевыми показателями, т.е.

$$a = up_1^{k_1} \dots p_r^{k_r}, \quad b = vp_1^{l_1} \dots p_r^{l_r}, \quad (3)$$

$$u|1, \quad v|1; \quad k_i \geq 0, \quad l_i \geq 0; \quad p_i \in \mathcal{P}; \quad 1 \leq i \leq r.$$

При помощи теоремы 1 получается легко запоминающийся

Признак делимости. Пусть a, b — элементы факториального кольца K , записанные в виде (3).

Справедливы утверждения:

- 1) $a|b$ тогда и только тогда, когда $k_i \leq l_i$, $i = 1, 2, \dots, r$;
- 2) $\text{НОД}(a, b) = p_1^{s_1} \dots p_r^{s_r}$, где $s_i = \min\{k_i, l_i\}$, $i = 1, 2, \dots, r$;
- 3) $\text{НОК}(a, b) = p_1^{t_1} \dots p_r^{t_r}$, где $t_i = \max\{k_i, l_i\}$, $i = 1, 2, \dots, r$.

Таким образом, в качестве s_i нужно брать наименьший из двух показателей k_i, l_i , а в качестве t_i — наибольший. В частности, элементы $a, b \in K$ взаимно просты в точности тогда, когда простые множители, входящие в разложение одного элемента, не входят в разложение другого.

Недостаток этого признака делимости заключается, конечно, в том, что на практике бывает весьма трудно получить разложение вида (3). Даже в случае $K = \mathbb{Z}$ (этим не предвосхищается факториальность \mathbb{Z}) приходится довольствоваться незначительными вариациями метода прямого перебора простых чисел, меньших данного числа n . Тем более приятно, что в факториальных кольцах, о которых пойдёт речь ниже, имеется эффективный способ вычисления $\text{НОД}(a, b)$ и $\text{НОК}(a, b)$.

3. Факториальность евклидовых колец. Алгоритм деления с остатком в \mathbb{Z} и $P[X]$ (см. п. 3 § 9 гл. 1 и п. 3 § 2) делает естественным рассмотрение целостного кольца K , в котором каждому элементу $a \neq 0$ поставлено в соответствие неотрицательное целое число $\delta(a)$, т.е. определено отображение

$$\delta : K \setminus \{0\} \rightarrow \mathbb{N} \cup \{0\}$$

так, что при этом выполняются условия:

E1) $\delta(ab) \geq \delta(a)$ для всех $a, b \neq 0$ из K ;

E2) каковы бы ни были $a, b \in K$, $b \neq 0$, найдутся $q, r \in K$ (q — “частное”, r — “остаток”), для которых

$$a = qb + r; \quad \delta(r) < \delta(b) \text{ или } r = 0. \quad (4)$$

Целостное кольцо K с этими свойствами называется *евклидовым кольцом*. Полагая $\delta(a) = |a|$ для $a \in \mathbb{Z}$ и $\delta(a) = \deg a$ для $a = a(X) \in P[X]$, мы приходим к выводу, что \mathbb{Z} и $P[X]$ — евклидовы кольца.

В евклидовых кольцах существует способ нахождения $\text{НОД}(a, b)$, называемый *алгоритмом последовательного деления* или *алгоритмом Евклида* и заключающийся в следующем. Пусть даны ненулевые элементы a, b евклидова кольца K . Применяя достаточно большое (но конечное) число раз предписание E2), мы получим систему равенств

типа (4) с последним нулевым остатком:

$$\begin{aligned} a &= q_1 b + r_1, & \delta(r_1) &< \delta(b), \\ b &= q_2 r_1 + r_2, & \delta(r_2) &< \delta(r_1), \\ r_1 &= q_3 r_2 + r_3, & \delta(r_3) &< \delta(r_2), \\ &\dots & & \\ r_{k-2} &= q_k r_{k-1} + r_k, & \delta(r_k) &< \delta(r_{k-1}), \\ r_{k-1} &= q_{k+1} r_k, & r_{k+1} &= 0. \end{aligned} \tag{5}$$

Это действительно так, поскольку строго убывающая цепочка неотрицательных целых чисел $\delta(b) > \delta(r_1) > \delta(r_2) > \dots$ должна оборваться, а обрыв может произойти только за счёт обращения в нуль одного из остатков.

Утверждается, что последний отличный от нуля остаток r_k является как раз наибольшим общим делителем элементов a и b в смысле определения, данного в п. 2. В самом деле, по условию $r_k | r_{k-1}$. Двигаясь в системе (5) снизу вверх и используя свойство 4) отношения делимости, сформулированное в п. 1, получим цепочку $r_k | r_{k-1}$, $r_k | r_{k-2}, \dots, r_k | r_2, r_k | r_1$ и, наконец, $r_k | b, r_k | a$. Стало быть, r_k — общий делитель элементов a и b . Обратно: пусть c — любой другой делитель тех же элементов; тогда $c | r_1$, и, двигаясь теперь в системе (5) сверху вниз, мы получим цепочку отношений делимости $c | r_2, c | r_3, \dots, c | r_k$. Последнее из них окончательно убеждает нас в том, что $\text{НОД}(a, b)$ существует, причём имеет место равенство

$$r_k = \text{НОД}(a, b). \tag{6}$$

Обратим, далее, внимание на то обстоятельство, что каждый остаток r_i , в системе (5) выражается в виде линейной комбинации с коэффициентами в K двух предыдущих остатков r_{i-1} и r_{i-2} . При этом r_1 выражается через a и b : $r_1 = a - q_1 b$, а r_2 , выражаясь через b и r_1 , тем самым является опять линейной комбинацией a и b . Последовательная подстановка в r_i выражений r_{i-1} и r_{i-2} через a и b даст нам при $i = k$ выражение

$$r_k = au + bv \tag{7}$$

с какими-то элементами $u, v \in K$.

Сопоставляя (6) и (7) и принимая во внимание теорему 2, б), получаем следующее утверждение.

Теорема 3. В евклидовом кольце K любые два элемента a, b имеют наибольший общий делитель и наименьшее общее кратное. При помощи алгоритма Евклида можно найти такие $u, v \in K$, что будет выполнено соотношение

$$\text{НОД}(a, b) = au + bv.$$

В частности, элементы $a, b \in K$ взаимно просты тогда и только тогда, когда существуют элементы $u, v \in K$, для которых

$$au + bv = 1.$$

Следствие. Пусть a, b, c — элементы евклидова кольца K .

- i) Если $\text{НОД}(a, b) = 1$ и $\text{НОД}(a, c) = 1$, то $\text{НОД}(a, bc) = 1$.
- ii) Если $b|bc$ и $\text{НОД}(a, b) = 1$, то $a|c$.
- iii) Если $b|a$, $c|a$ и $\text{НОД}(b, c) = 1$, то $bc|a$.

Доказательство. i) Согласно теореме 3 имеем равенства $au_1 + bv_1 = 1$, $au_2 + cv_2 = 1$. Перемножая соответственно их левые и правые части, находим $a(au_1u_2 + bv_2v_1 + cu_1v_2) + bc(v_1v_2) = 1$, что и даёт нужное утверждение.

ii) Имеем $au + bv = 1$, откуда $ac \cdot u + (bc)v = c$. Но $bc = aw$, поэтому $c = a(cu + vw)$, т.е. $a|c$.

iii) Согласно свойству ii') НОК

$$b|a, c|a \Rightarrow \text{НОК}(b, c)|a \Rightarrow bc|a,$$

поскольку $bc = \text{НОД}(b, c)\text{НОК}(b, c)$ и $\text{НОД}(b, c) = 1$ по условию. \square

Читатель легко распространит утверждение теоремы 3 на случай произвольного конечного числа элементов евклидова кольца.

Непосредственным шагом к установлению факториальности евклидова кольца служит

Лемма. Всякое евклидово кольцо K является кольцом с разложением (т.е. любой элемент $a \neq 0$ из K записывается в виде (1)).

Доказательство. Пусть элемент $a \in K$ обладает собственным делителем b : $a = bc$, где b и c — необратимые элементы (другими словами, a и b не ассоциированы). Докажем, что $\delta(b) < \delta(a)$.

В самом деле, согласно Е1) непосредственно имеем $\delta(b) \leq \delta(bc) = \delta(a)$. Предположив выполнение равенства $\delta(b) = \delta(a)$, мы воспользуемся условием Е2) и найдем q, r с $b = qa + r$, где $\delta(r) < \delta(a)$ или же $r = 0$. Случай $r = 0$ отпадает ввиду неассоциированности a и b . По той же причине $1 - qc \neq 0$. Стало быть, снова по Е2) (поменять местами a и b) имеем

$$\delta(a) = \delta(b) \leq \delta(b(1 - qc)) = \delta(b - qa) = \delta(r) < \delta(a)$$

— противоречие. Итак, $\delta(b) < \delta(a)$.

Если теперь $a = a_1a_2 \dots a_n$, где все a_i необратимы, то $a_{m+1}a_{m+2} \dots a_n$ — собственный делитель $a_ma_{m+1} \dots a_n$, и по доказанному

$$\delta(a) = \delta(a_1a_2 \dots a_n) > \delta(a_2 \dots a_n) > \dots > \delta(a_n) > \delta(1).$$

Эта строго убывающая цепочка неотрицательных целых чисел имеет длину $n \leq \delta(a)$. Значит, для элемента $a \in K$ имеется разложение максимальной длины, которое и будет разложением на простые множители. \square

Теорема 4. Всякое евклидово кольцо K факториально (т.е. обладает свойством однозначности разложения на простые множители).

Доказательство. С учётом леммы и критерия факториальности, содержащегося в теореме 1, нам остаётся показать, что если p — простой элемент кольца K , делящий произведение bc каких-то элементов $b, c \in K$, то p делит либо b , либо c .

Действительно, при $b = 0$ или $c = 0$ доказывать нечего. Если же $bc \neq 0$ и $d = \text{НОД}(b, p)$, то d , будучи делителем простого элемента p , либо равен 1 (точнее, является делителем 1), либо ассоциирован с p . В первом случае b и p оказываются взаимно простыми, и утверждение ii) следствия теоремы 3 позволяет заключить, что $p|c$. Во втором случае $d = up$, $u \neq 1$ и, значит, $p|b$. \square

Следствие. Кольца \mathbb{Z} и $P[X]$ факториальны (P — произвольное поле).

Доказательство. Как отмечалось непосредственно после определения евклидовости, на каждом из колец $\mathbb{Z}, P[X]$ задана естественная функция δ с нужными свойствами E1), E2), так что остаётся сослаться на теорему 4. \square

Очень рекомендуется провести отдельно для \mathbb{Z} и для $P[X]$ доказательства факториальности, чтобы устраниТЬ всякую видимость какого-либо наукообразия в этом вопросе.

Факториальность кольца многочленов $P[X_1, \dots, X_n]$, $n > 1$, уже не являющегося евклидовым, устанавливается в [ВА III]. Там же приводятся дополнительные примеры евклидовых колец.

4. Неприводимые многочлены. Специализируя данное ранее определение простого элемента, ещё раз подчеркнем, что многочлен f иенулевой степени из кольца $P[X]$ называется **неприводимым** в $P[X]$ (или **неприводимым над полем P**), если он не делится ни на какой многочлен $g \in P[X]$, у которого $0 < \deg g < \deg f$. В частности, всякий многочлен первой степени неприводим. Совершенно очевидно, что неприводимость многочлена степени > 1 или разложение его на неприводимые множители — понятия, тесно связанные с основным полем P , как это показывает уже известный нам по построению комплексных чисел многочлен $X^2 + 1 = (X + i)(X - i)$. Многочлен $X^4 + 4$ приводим над \mathbb{Q} , хотя об этом и нелегко догадаться:

$$X^4 + 4 = (X^2 - 2X + 2)(X^2 + 2X + 2).$$

Оба множителя справа неприводимы не только над \mathbb{Q} , но и над \mathbb{R} , будучи приводимыми, однако, над \mathbb{C} .

Как простых чисел в \mathbb{Z} (см. § 9 гл. 1), так и нормализованных **неприводимых многочленов над произвольным полем P** бесконечно много.

В случае бесконечного поля P это ясно: достаточно рассмотреть неприводимые многочлены вида $X - c$, $c \in P$.

Если же поле P конечно, то годится рассуждение Евклида. Именем, пусть уже найдены n неприводимых многочленов p_1, \dots, p_n . Многочлен $f = p_1 p_2 \dots p_n + 1$ имеет хотя бы один нормализованный простой делитель, поскольку $\deg f \geq n$. Обозначим его через p_{n+1} . Он отличен от p_1, \dots, p_n , поскольку из $p_{n+1} = p_s$ для какого-то $s \leq n$ следовало бы $p_s|(f - p_1 \dots p_n)$, т.е. $p_s|1$. \square

Так как многочленов заданной степени над конечным полем конечное число, то можно сделать следующее полезное заключение.

Над любым конечным полем существуют неприводимые многочлены сколь угодно высокой степени.

Это утверждение качественного характера будет уточнено в [BA III].

Неприводимые многочлены над полем \mathbb{Q} играют особую роль в теории полей алгебраических чисел. Так как умножением на подходящее натуральное число от многочлена из $\mathbb{Q}[X]$ всегда можно перейти к многочлену из $\mathbb{Z}[X]$, то естественно уточнить сначала связь между свойствами приводимости над \mathbb{Q} и над \mathbb{Z} . Имея в виду другие приложения, мы докажем одно общее утверждение о многочленах над факториальным кольцом K .

Назовём *содержанием* многочлена $f = a_0 + a_1 X + \dots + a_n X^n \in K[X]$ наибольший общий делитель $d = d(f)$ всех его коэффициентов. До сих пор мы говорили о НОД(a, b) двух элементов, но свойства i)–vi) НОД позволяют без труда распространить это понятие на любое конечное число элементов целостного кольца.

Если $d(f)$ — обратимый элемент в K , то многочлен f называют *примитивным*.

Лемма Гаусса. Пусть K — факториальное кольцо и $f, g \in K[X]$. Тогда

$$d(fg) \approx d(f) \cdot d(g).$$

В частности, произведение двух примитивных многочленов снова будет примитивным многочленом (здесь и ниже \approx означает равенство с точностью до ассоциативности).

Доказательство. Начнём с последнего утверждения. Пусть

$$f = a_0 + a_1 X + \dots + a_n X^n, \quad g = b_0 + b_1 X + \dots + b_m X^m$$

— примитивные многочлены из $K[X]$, произведение fg которых не является примитивным. Существует, стало быть, простой элемент $p \in K$, делящий $d(fg)$. Выберем наименьшие индексы s, t , для которых $p \nmid a_s$, $p \nmid b_t$. Такие индексы найдутся ввиду примитивности f и g . Коэффициентом при X^{s+t} в fg будет

$$c_{s+t} = a_s b_t + (a_{s+1} b_{t-1} + a_{s+2} b_{t-2} + \dots) + (a_{s-1} b_{t+1} + a_{s-2} b_{t+2} + \dots).$$

Так как a_{s-i} и b_{t-i} при $i > 0$ делятся на p по условию и $p|c_{s+t}$ по предположению, то мы имеем соотношение

$$pu = a_s b_t + pv,$$

из которого следует, что $p|a_s b_t$. Ввиду факториальности K имеем $p|a_s$ или $p|b_t$ — противоречие, доказывающее наше утверждение.

Переходя к общему случаю, запишем произвольные многочлены $f, g \in K[X]$ в виде

$$f = d(f)f_0, \quad g = d(g)g_0,$$

где f_0, g_0 — примитивные многочлены. Так как $fg = d(f)d(g) \cdot f_0g_0$ и по доказанному $d(f_0g_0) \approx 1$, то, стало быть, $d(fg) \approx d(f)d(g)$. \square

Следствие. *Многочлен $f \in \mathbb{Z}[X]$, неприводимый над \mathbb{Z} , продолжает оставаться неприводимым и над \mathbb{Q} ($\deg f > 0$).*

Доказательство. Согласно следствию теоремы 4 \mathbb{Z} — факториальное кольцо, поэтому к $\mathbb{Z}[X]$ применима лемма Гаусса. Предположим, что $f = gh$, где $f \in \mathbb{Z}[X]$, а $g, h \in \mathbb{Q}[X]$. Умножая обе части этого равенства на наименьшее общее кратное знаменателей всех коэффициентов у g и h , мы перепишем его в виде $af = bg_0h_0$, где $a, b \in \mathbb{Z}$ и g_0, h_0 — примитивные многочлены над \mathbb{Z} . По лемме Гаусса $a \cdot d(f) = b$ (в данном случае без ограничения общности ассоциированность заменяется на равенство), так что получается разложение $f = d(f)g_0h_0$ над \mathbb{Z} . Остаётся вспомнить о неприводимости f в $\mathbb{Z}[X]$. \square

Критерий неприводимости (Эйзенштейн). Пусть

$$f(X) = X^n + a_1 X^{n-1} + \dots + a_{n-1} X + a_n$$

— нормализованный многочлен над \mathbb{Z} , все коэффициенты a_1, \dots, a_n которого делятся на некоторое простое число p , но a_n не делится на p^2 . Тогда $f(X)$ неприводим над \mathbb{Q} .

В самом деле, предположив противное и воспользовавшись следствием леммы Гаусса, мы запишем f в виде произведения двух многочленов над \mathbb{Z} :

$$f(X) = (X^s + b_1 X^{s-1} + \dots + b_s)(X^t + c_1 X^{t-1} + \dots + c_t), \quad st > 0.$$

Это разложение сохранится и в кольце $\mathbb{Z}_p[X]$, элементы которого получаются из целочисленных многочленов взятием их коэффициентов по модулю p . По условию $\bar{a}_i = \bar{0}$, где \bar{a}_i — класс вычетов по модулю p , соответствующий целому числу a_i . Но кольцо $\mathbb{Z}_p[X]$ факториально (следствие теоремы 4). Сравнивая два разложения:

$$X^s X^t = (X^s + \bar{b}_1 X^{s-1} + \dots)(X^t + \bar{c}_1 X^{t-1} + \dots), \quad s + t = n,$$

мы неизбежно приходим к заключению, что $\bar{b}_i = \bar{0} = \bar{c}_j$, т.е. все коэффициенты b_i, c_j делятся на p . В таком случае $a_n = b_s c_t$ делится на p^2 — противоречие, устанавливающее справедливость критерия Эйзенштейна. \square

Примечание. Критерий действует и в том случае, когда старший коэффициент a_0 отличен от 1, но не делится на p .

Пример 2. Многочлен $f(X) = X^{p-1} + X^{p-2} + \dots + X + 1$ неприводим над \mathbb{Q} при любом простом p .

Достаточно заметить, что вопрос о неприводимости $f(X)$ эквивалентен вопросу о неприводимости многочлена

$$f(X+1) = \frac{(X+1)^p - 1}{(X+1) - 1} = X^{p-1} + \binom{p}{1} X^{p-2} + \dots + \binom{p}{p-2} X + \binom{p}{p-1},$$

все коэффициенты которого, кроме старшего, делятся на p в первой степени (свойство биномиальных коэффициентов, отмеченное в упр. 6 из § 3 гл. 4) и к которому, следовательно, применим критерий Эйзенштейна.

УПРАЖНЕНИЯ

1. Показать, что

$$n\mathbf{Z} + m\mathbf{Z} = \mathbf{Z} \cdot \text{НОД}(n, m),$$

$$n\mathbf{Z} \cap m\mathbf{Z} = \mathbf{Z} \cdot \text{НОК}(n, m).$$

2. Пусть f, g — нормализованные многочлены из $\mathbf{Z}[X]$. Показать, что в выражении $\text{НОД}(f, g) = fu + gv$ с $u, v \in \mathbf{Z}[X]$ можно считать $\deg u < \deg g$, $\deg v < \deg f$.

3. Являются ли кольца $\mathbf{Z}[\sqrt{-3}]$ и $\mathbf{Z}_5[X]$ факториальными?

4. Разложить на неприводимые множители в $\mathbf{Z}[X]$ многочлены $X^n - 1$ при $5 \leq n \leq 12$.

5. Доказать, что неприводимые множители однородного многочлена

$$f(X, Y) = a_0 X^n + a_1 X^{n-1} Y + \dots + a_{n-1} X Y^{n-1} + a_n Y^n \in \mathbb{Q}[X, Y]$$

однородны и $f(X, Y)$ неприводим тогда и только тогда, когда неприводим многочлен $f(X, 1) = a_0 X^n + a_1 X^{n-1} + \dots + a_{n-1} X + a_n \in \mathbb{Q}[X]$.

6. Пусть P — поле и $f(X) = \sum_{i \geq 0} a_i X^i$ — формальный степенной ряд из $P[[X]]$ (см. упр. 5 из § 2). Условие $a_0 \neq 0$, или, что эквивалентно, $\omega(f) = 0$ необходимо и достаточно для существования степенного ряда $g(X) \in P[[X]]$, обратного к f : $fg = 1$. Например, $(1 - X)^{-1} = \sum_{i \geq 0} X^i$. С точностью до ассоциации X — единственный простой элемент в $P[[X]]$. Кольцо $P[[X]]$ факториально. Обосновать эти утверждения.

7. Показать, что $\det(x_{ij}) = \sum_{\pi \in S_n} \epsilon_\pi x_{\pi(1), 1} \dots x_{\pi(n), n}$ — неприводимый однородный степени n многочлен от n^2 независимых переменных x_{ij} .

Указание. Рассуждая от противного, предположить, что

$$\det(x_{ij}) = g_1(\dots, x_{ij}, \dots) \times g_2(\dots, x_{ij}, \dots).$$

Так как $\det(x_{ij})$ — линейный однородный многочлен от переменных, стоящих в одном фиксированном столбце, то один из множителей g_1, g_2 является линейным однородным многочленом от x_{ij} , $1 \leq i \leq n$, при фиксированном j , в то время как другой совсем не зависит от x_{ij} , $1 \leq i \leq n$. Аналогичные рассуждения сохраняются при замене столбцов на строки. Пусть, скажем, x_{11} входит в g_1 . Тогда g_2 не содержит x_{1j} , $1 \leq j \leq n$, откуда следует что g_2 не содержит x_{ij} , $1 \leq i, j \leq n$, т.е. g_2 — константа.

§ 4. Поле отношений

1. Построение поля отношений целостного кольца. В предыдущих двух параграфах было установлено много свойств, общих для \mathbb{Z} и $P[X]$. Наша ближайшая цель — вложить $P[X]$ в поле, причём сделать это нужно самым экономным способом, образцом для которого может служить вложение \mathbb{Z} в \mathbb{Q} . Фактически нисколько не сложнее решать точно такую же задачу для произвольного целостного кольца A .

Рассмотрим множество $A \times A^*$ ($A^* = A \setminus \{0\}$) всех пар (a, b) элементов $a, b \in A$ с $b \neq 0$. Это множество разобьём на классы, полагая пары (a, b) и (c, d) принадлежащими одному и тому же классу, как только $ad = bc$; в записи: $(a, b) \sim (c, d)$. Ясно, что всегда $(a, b) \sim (a, b)$. Далее, $(a, b) \sim (c, d) \iff (c, d) \sim (a, b)$ и, наконец, $(a, b) \sim (c, d), (c, d) \sim (e, f) \implies (a, b) \sim (e, f)$. Действительно, имеют место равенства $ad = be, cf = de$, откуда $adf = bcf = bde$, т.е. $d(ad - be) = 0$. Но $d \neq 0$, и в силу целостности кольца A получаем $af = be$, что и означает $(a, b) \sim (e, f)$. Итак, отношение \sim рефлексивно, симметрично и транзитивно, т.е. (см. § 6 гл. 1) оно является отношением эквивалентности на множестве $A \times A^*$ и, следовательно, определяет разбиение $A \times A^*$ на непересекающиеся классы.

Пусть $Q(A)$ — множество всех классов эквивалентности, или, что то же самое, $Q(A)$ есть фактормножество $A \times A^*/\sim$ множества $A \times A^*$ по отношению эквивалентности \sim . Будем обозначать символом $[a, b]$ класс, в котором лежит упорядоченная пара (a, b) . По определению

$$[a, b] = [c, d] \iff ad = bc. \quad (1)$$

Если на множестве $A \times A^*$ задать операции сложения и умножения формулами

$$(a, b) + (c, d) = (ad + bc, bd), \quad (a, b)(c, d) = (ac, bd)$$

(а это возможно, поскольку в A из $b \neq 0, d \neq 0$ следует $bd \neq 0$), то эти бинарные операции можно перенести на $Q(A)$. В самом деле, нам нужно показать, что

$$(a', b') \sim (a, b) \implies \begin{cases} (a, b) + (c, d) \sim (a', b') + (c, d), \\ (a, b) \cdot (c, d) \sim (a', b') \cdot (c, d). \end{cases}$$

То же самое выражается соотношениями

$$(ad + bc)b'd = (a'd + b'c)bd,$$

$$ac \cdot b'd = a'c \cdot bd,$$

истинность которых прямо вытекает из условия $a'b = ab'$. Аналогичный результат получим, заменяя (c, d) на (c', d') , где $cd' = c'd$. Мы приходим к заключению, что на $Q(A)$ операциями сложения и умно-

жения, не зависящими от выбора представителей в классах эквивалентности, будут

$$[a, b] + [c, d] = [ad + bc, bd], \quad [a, b][c, d] = [ac, bd]. \quad (2)$$

Здесь следовало бы писать $[a, b] \oplus [c, d]$ и $[a, b] \odot [c, d]$, но без ущерба для ясности \oplus и \odot заменены обычными знаками суммы и произведения.

Убедимся теперь в том, что $Q(A)$, рассматриваемое вместе с операциями (2), есть поле. Действительно, например, из соотношений

$$[a, b] + ([c, d] + [e, f]) = [a, b] + [cf + de, df] = [adf + bcf + bde, bdf],$$

$$([a, b] + [c, d]) + [e, f] = [ad + bc, bd] + [e, f] = [adf + bcf + bde, bdf]$$

вытекает закон ассоциативности для операции сложения. Ассоциативность умножения очевидна. Далее, соотношения

$$([a, b] + [c, d]), [e, f] = [ade + bce, bdf],$$

$$[a, b][e, f] + [c, d][e, f] = [adef + bcef, bfdf] = [(ade + bce)f, (bdf)f]$$

и условия (1) равенства классов эквивалентности показывают, что выполняется закон дистрибутивности.

Столь же просто проверяется коммутативность операций сложения и умножения. Нулем для сложения является класс $[0, 1]$ ($[0, 1] + [a, b] = [a, b]$), а единицей для умножения — класс $[1, 1]$. Далее, $-[a, b] = [-a, b]$, поскольку $[a, b] + [-a, b] = [0, b^2] = [0, 1]$. Всё это вместе взятое означает, что $Q(A)$ — коммутативное кольцо с единицей. Если $[a, b] \neq [0, 1]$, то $a \neq 0$ в A , стало быть, $[b, a] \in Q(A)$ и $[a, b][b, a] = [1, 1]$, так что мультипликативным обратным к $[a, b] \neq [0, 1]$ служит $[b, a]$. Тем самым показано, что $Q(A)$ — поле.

Сопоставление $a \mapsto [a, 1]$ определяет инъективное отображение $f: A \rightarrow Q(A)$, которое на самом деле является морфизмом ((мономорфизмом) колец ($f(a+b) = f(a) + f(b)$, $f(ab) = f(a)f(b)$; $a \neq b \Rightarrow f(a) \neq f(b)$). Для любого элемента $x = [a, b] \in Q(A)$ имеем

$$[b, 1]x = [a, 1],$$

так что x есть “отношение” $f(a)/f(b)$ элементов из $f(A)$. По этой причине $Q(A)$ называется *полем отношений* кольца A .

Удобно отождествить каждый элемент $a \in A$ с его образом $f(a) = [a, 1] \in Q(A)$, т.е. заменить A на $f(A)$. Можно поступить несколько иначе: заменить каждый из элементов $[a, 1] \in Q(A)$ на $a \in A$, оставив без изменения все другие элементы поля $Q(A)$, и произвести надлежащие замены в формулах (2). Именно, следует положить

$$a + [b, c] = [ac + b, c], \quad a[b, c] = [ab, c].$$

В результате целостное кольцо A окажется с самого начала подкольцом поля, изоморфного $Q(A)$ и изображаемого обычно тем же символом $Q(A)$. После такого отождествления разумно называть элементы

$[a, b]$ дробями и писать короче и в привычной форме

$$[a, b] = \frac{a}{b}.$$

Введённые выше правила действий с классами $[a, b]$ повторяют, как нетрудно догадаться, правила действий с дробями в поле (см. (8) в п. 4 § 3 гл. 4). Нами доказана

Теорема 1. Для каждого целостного кольца A существует поле отношений (или поле частных, поле дробей) $Q(A)$, элементы которого имеют вид a/b , $a \in A$, $0 \neq b \in A$. Действия с дробями подчиняются правилам (1), (2), где следует положить $[a, b] = a/b$.

Конструкция полей отношений довольно часто используется в математике. Её естественность оправдывается хотя бы тем, что поле \mathbb{Q} есть не что иное, как поле отношений $Q(\mathbb{Z})$ кольца \mathbb{Z} . Легко видеть (проверьте это), что $Q(A) \cong A$, если A — поле.

З а м е ч а н и е. Можно доказать, что если целостное кольцо A есть подкольцо поля P и каждый элемент $x \in P$ записывается в виде отношения a/b элементов $a \in A$, $0 \neq b \in A$, то $P \cong Q(A)$. Например, $Q(\sqrt{d}) = Q(\mathbb{Z}[\sqrt{d}])$.

2. Поле рациональных дробей. Пусть P — поле, $P[X]$ — кольцо многочленов над P . Поле отношений $Q(P[X])$ кольца $P[X]$ обозначается символом $P(X)$ (смена квадратных скобок на круглые) и называется *полем рациональных дробей* от переменной X с коэффициентами в P .

Следует заметить, что поле рациональных дробей $P(X)$ всегда содержит бесконечное число элементов, а его характеристика совпадает с характеристикой поля P . Поле $\mathbb{F}_p(X)$ доставляет пример бесконечного поля характеристики $p > 0$.

Каждая рациональная дробь поля $P(X)$ записывается (притом многими способами) в виде f/g (или $\frac{f}{g}$, если не стремиться к экономии бумаги), где f, g — многочлены из кольца $P[X]$, $g \neq 0$. По определению $f/g = f_1/g_1 \iff fg_1 = f_1g$. Естественно назвать f *числителем*, а g — *знаменателем* дроби f/g . Дробь не меняется, если её числитель и знаменатель умножаются на один и тот же ненулевой многочлен или сокращаются на любой общий множитель. В частности, целое число (положительное или отрицательное) $\deg f - \deg g$ не зависит от представления ненулевой рациональной дроби в виде отношения (частного) f/g двух многочленов. Это число называется *степенью дроби*. Рациональная дробь от переменной X называется *несократимой*, если её числитель взаимно прост со знаменателем. С точностью до множителя из P , общего для числителя и знаменателя, любая рациональная дробь f/g однозначно определяется некоторой несократимой дробью. В самом деле, деление f и g на НОД(f, g) приводит к несократимой дроби, а равенство $f/g = f_1/g_1$ двух несокра-

тимых дробей, выражение в виде $fg_1 = f_1g$, даёт $f = cf_1$, $c \in P$, $g = cg_1$ (использовать следствие теоремы 4 из § 3).

Если $\deg(f/g) = \deg f - \deg g < 0$, то (несократимая) дробь f/g называется *правильной* (нулевой многочлен причисляется к правильным дробям, поскольку мы условились считать $\deg 0 = -\infty$).

Теорема 2. Каждая рациональная дробь из $P(X)$ однозначно представима в виде суммы многочлена и правильной дроби.

Доказательство. Алгоритм деления с остатком, применённый к числителю и знаменателю дроби f/g , дает равенство $f = qg + r$, где $\deg r < \deg g$. Теперь $f/g = q + r/g$ есть искомая запись, сравнение которой с любой другой записью того же типа $f/g = \bar{q} + \bar{r}/\bar{g}$ ($\bar{q}, \bar{r}, \bar{g} \in P[X]$, $\deg \bar{r} < \deg \bar{g}$) приводит к соотношению

$$\bar{q} - q = \frac{r}{g} - \frac{\bar{r}}{\bar{g}} = \frac{r\bar{g} - \bar{r}g}{g\bar{g}}.$$

Так как $\bar{q} - q \in P[X]$, а

$$\deg\left(\frac{r\bar{g} - \bar{r}g}{g\bar{g}}\right) = \deg(r\bar{g} - \bar{r}g) - \deg g - \deg \bar{g} < 0,$$

то это возможно лишь в случае $\bar{q} - q = 0$ и $r/g = \bar{r}/\bar{g}$. \square

Замечание. Множество $P_0(X)$ всех правильных дробей, рассматриваемое вместе с операциями сложения и умножения в $P(X)$, является кольцом без единицы 1.

Действительно, пусть $f_1/g_1, f_2/g_2 \in P_0(X)$. Так как

$\deg f_1 f_2 = \deg f_1 + \deg f_2 < \deg g_1 + \deg g_2 = \deg g_1 g_2$,
то

$$\left(\frac{f_1}{g_1}\right)\left(\frac{f_2}{g_2}\right) = \frac{f_1 f_2}{g_1 g_2} \in P_0(X).$$

Далее,

$$\frac{f_1}{g_1} \pm \frac{f_2}{g_2} = \frac{f_1 g_2 \pm f_2 g_1}{g_1 g_2} \in P_0(X),$$

поскольку степени каждого из слагаемых $f_1 g_2$ и $f_2 g_1$ строго меньше степени знаменателя $g_1 g_2$. Как мы условились перед формулировкой теоремы 2, $0 \in P_0(X)$. В то же время $1 \notin P_0(X)$. \square

До сих пор мы всё время подчёркивали, насколько похожи кольца Z и $P[X]$. При переходе к их полям отношений появляется существенное различие: правильные дроби в Q не образуют кольца. Например,

$$\frac{2}{3} + \frac{3}{5} = \frac{19}{15}.$$

3. Простейшие дроби. Правильная рациональная дробь $f/g \in P(X)$ называется *простейшей*, если $g = p^n$, $n \geq 1$, где $p = p(X)$ — неприводимый многочлен, причём $\deg f < \deg p$.

Основной теоремой о рациональных дробях является

Теорема 3. Каждая правильная рациональная дробь может быть разложена, и притом единственным образом, в сумму простейших дробей.

Доказательство. Пусть $f/g \in P(X)$ — данная нам правильная рациональная дробь, в которой без ограничения общности многочлен g можно считать нормализованным (если $\lambda \neq 0$ — старший коэффициент многочлена g , то следует перейти от f/g к равной ей дроби $\lambda^{-1}f/\lambda^{-1}g$). Дальнейшие рассуждения распадаются на ряд этапов.

Этап 1. Предположим, что $g = g_1g_2$ — произведение двух взаимно простых нормализованных многочленов. Тогда

$$\frac{f}{g_1g_2} = \frac{f_1}{g_1} + \frac{f_2}{g_2}, \quad (3)$$

причём обе дроби в правой части правильные, а сама запись в виде суммы единственна.

В самом деле, из взаимной простоты g_1, g_2 следует (теорема 3 из § 3), что $1 = u_1g_1 + u_2g_2$ для некоторых $u_1, u_2 \in P[X]$. Умножив обе части этого соотношения на f , получим $f = fu_1g_1 + fu_2g_2$. Если теперь $fu_2 = qg_1 + f_1$, $\deg f_1 < \deg g_1$ (деление fu_2 на g_1 с остатком), то $f = f_1g_2 + f_2g_1$, где $f_2 = fu_1 + qg_2$. Разделив обе части этого соотношения на g_1g_2 , т.е. рассмотрев дроби со знаменателем g_1g_2 , мы придём к исковому разложению (3), поскольку по построению $f_1/g_1 \in P_0(X)$ (множество правильных дробей) и в соответствии с замечанием в конце п. 2 $f_2/g_2 = f/g - f_1/g_1 \in P_0(X)$.

Пусть теперь наряду с (3) имеется ещё разложение $f/g = f'_1/g_1 + f'_2/g_2$ в сумму правильных дробей. Тогда из равенства $f_1/g_1 + f_2/g_2 = f'_1/g_1 + f'_2/g_2$ будем иметь

$$(f_1 - f'_1)g_2 = (f'_2 - f_2)g_1.$$

Из делимости $(f_1 - f'_1)g_2$ на g_1 и из взаимной простоты g_1, g_2 следует, что разность $f_1 - f'_1$ должна делиться на g_1 . Но $\deg(f_1 - f'_1) < \deg g_1$, и, стало быть, $f_1 - f'_1 = 0$. Единственность разложения (3) установлена.

Этап 2. Пусть в правильной рациональной дроби f/g для (нормализованного) знаменателя g имеется каноническое разложение

$$g = p_1^{n_1} p_2^{n_2} \dots p_m^{n_m} \quad (4)$$

в произведение степеней попарно различных нормализованных не-приводимых над P многочленов $p_1(X), p_2(X), \dots, p_m(X)$. Тогда существует однозначно определённое разложение

$$\frac{f}{g} = \sum_{i=1}^m f_i p_i^{n_i}$$

в сумму правильных дробей $f_i/p_i^{n_i}$ (эти дроби называются ещё *примарными*).

Наше утверждение легко получается индукцией по m , базу для которой даёт этап 1:

$$\frac{f}{g} = \frac{f_1}{p_1^{n_1}} + \frac{f_0}{p_2^{n_2} \cdots p_m^{n_m}} = \frac{f_1}{p_1^{n_1}} + \left(\frac{f_2}{p_2^{n_2}} + \cdots + \frac{f_m}{p_m^{n_m}} \right).$$

Так как f_1 и f_0 определены однозначно, то по предположению индукции это верно и относительно f_2, \dots, f_m .

Этап 3. Всякая правильная примарная дробь a/p^n представляется, и притом единственным образом, в виде суммы простейших дробей.

Действительно, так как по условию $\deg a < n \deg p$, то евклидов алгоритм деления с остатком приведёт нас к системе равенств

$$\begin{aligned} a &= q_1 p^{n-1} + r_1, & \deg r_1 &< (n-1) \deg p, \\ r_1 &= q_2 p^{n-2} + r_2, & \deg r_2 &< (n-2) \deg p, \\ &\vdots & &\vdots \\ r_{n-2} &= q_{n-1} p + r_{n-1}, & \deg r_{n-1} &< \deg p, \\ r_{n-1} &= q_n, & & \end{aligned}$$

где $\deg q_i < \deg p$ для всех однозначно определённых частных q_1, \dots, q_n . Мы видим, что

$$a = q_1 p^{n-1} + q_2 p^{n-2} + \cdots + q_{n-1} p + q_n,$$

откуда

$$\frac{a}{p^n} = \frac{q_1}{p} + \frac{q_2}{p^2} + \cdots + \frac{q_{n-1}}{p^{n-1}} + \frac{q_n}{p^n}.$$

Так как $\deg q_i < \deg p$, то дроби q_i/p^i являются простейшими. По построению они однозначно определены (аналог разложения целого числа в 2-адическую или в десятичную дробь).

Этап 4. Рассуждения этапов 1–3, соединённые вместе, дают всё, что нужно.

Из доказательства теоремы 3 видно, что если f/g — правильная рациональная дробь, то знаменателями соответствующих простейших дробей при заданном каноническом разложении (4) для g будут $p_1^{n_1}, p_1^{n_1-1}, \dots, p_1; \dots; p_m^{n_m}, p_m^{n_m-1}, \dots, p_m$.

Тема простейших дробей, не очень актуальная для алгебры (хотя и дающая новые примеры колец), находит важные приложения в анализе. Это обусловлено специальным видом неприводимых многочленов над полями \mathbb{C} и \mathbb{R} , о чём более подробно будет говориться в гл. 6.

УПРАЖНЕНИЯ

1. Построить поле отношений $\mathbb{R}((X))$ кольца $\mathbb{R}[[X]]$ формальных степенных рядов от X с коэффициентами в поле \mathbb{R} . Опираясь на упр. 6 из § 3, показать, что каждый элемент поля $\mathbb{R}((X))$ имеет вид так называемого *мероморфного степенного ряда*

$$\varphi(X) = a_{-m} X^{-m} + a_{-m+1} X^{-m+1} + \dots + a_{-1} X^{-1} + a_0 + a_1 X + a_2 X^2 + \dots, \quad a_i \in \mathbb{R},$$

в котором допускается конечное число отрицательных показателей. Другими словами, $\varphi(X) = X^{-m} f(X)$, где $f(X)$ — обычный степенной ряд из $\mathbb{R}[[X]]$.

2. Пусть бесконечная последовательность вещественных чисел a_0, a_1, a_2, \dots периодична, начиная с некоторого члена. Показать, что степенной ряд $f(X) = a_0 + a_1 X + a_2 X^2 + \dots$ записывается в виде рациональной дроби из $\mathbb{R}(X)$.

3. Показать, что множество простейших дробей вида f/p^n , $n \geq 1$ (и их линейных комбинаций), является подкольцом кольца $P_0(X)$ (см. замечание в конце п. 2).

4. Показать, что в соответствии с упр. 3 простейшие дроби над $P = \mathbb{Z}_3$

$$\frac{a_1 X + b_1}{X^2 + 1} + \frac{a_2 X + b_2}{(X^2 + 1)^2} + \dots, \quad a_i, b_i \in \mathbb{Z}_3,$$

образуют кольцо K с бесконечно убывающей цепью

$$K = K_1 \supset K_2 \supset \dots \supset K_N \supset \dots$$

подкольцо K_N , натянутых на дроби $(aX + b)/(X^2 + 1)^n$, $n \geq N$.

КОРНИ МНОГОЧЛЕНОВ

Займёмся тем, ради чего в прошлом изучали алгебру, — корнями многочленов. Эта область перестала быть доминирующей в алгебре, но её важность никем не оспаривается. Дело в том, что многие задачи математики в конечном счёте сводятся к вычислению отдельных корней конкретных многочленов или к качественному описанию их совокупности. Нам удастся затронуть лишь простейшие свойства корней, но их во всяком случае будет достаточно, чтобы в полной мере оценить особое место, занимаемое полем \mathbb{C} комплексных чисел.

§ 1. Общие свойства корней

1. Корни и линейные множители. Пусть коммутативное кольцо A с единицей содержится в целостном кольце K .

Определение. Элемент $c \in K$ называется **корнем** (или **нулем**) **многочлена** $f \in A[X]$, если $f(c) = 0$. Говорят также, что c — корень уравнения $f(x) = 0$.

Необходимость рассмотрения колец, содержащих A собственным образом, станет понятной, если вспомнить, что многочлен $f(X) = X^2 + 1$ над \mathbb{R} не имеет нулей в \mathbb{R} , но $f(i) = 0$, $i \in \mathbb{C} = \mathbb{R}[i]$. Сначала мы рассмотрим, однако, случай $K = A$.

Теорема 1 (теорема Безу). Элемент $c \in A$ является корнем многочлена $f \in A[X]$ тогда и только тогда, когда $X - c$ делит f в кольце $A[X]$.

Доказательство. Эта теорема — часть более общего утверждения, которое мы могли бы доказать давно. А именно, алгоритм деления с остатком (теорема 5 из § 2 гл. 5) гласит, что $f(X) = (X - c)q(X) + r(X)$, где $\deg r(X) < \deg(X - c) = 1$. Значит, $r(X)$ — константа. Подстановка c вместо X (т.е. применение отображения Π_c из теоремы 2 § 2 гл. 5) даёт $f(c) = r$, так что всегда

$$f(X) = (X - c)q(X) + f(c). \quad (1)$$

В частности, $f(c) = 0 \iff f(X) = (X - c)q(X)$. \square

Деление многочлена $f(X)$ с коэффициентами в целостном кольце A на линейный многочлен $X - c$ удобно осуществлять по так называемой *схеме Горнера*, более простой, чем общий алгоритм деления с остатком. Именно, пусть

$$f(X) = a_0 X^n + a_1 X^{n-1} + \dots + a_n, \quad a_i \in A.$$

Согласно формуле (1)

$$q(X) = b_0 X^{n-1} + b_1 X^{n-2} + \dots + b_{n-1}, \quad b_j \in A.$$

Подставляя теперь в (1) эти выражения для $f(X)$ и $q(X)$ и сравнивая коэффициенты при одинаковых степенях X (начиная со старших), мы после небольшого преобразования получим

$$\boxed{b_0 = a_0 \quad \dots \quad b_k = b_{k-1}c + a_k \quad \dots} \\ \dots \boxed{b_{n-1} = b_{n-2}c + a_{n-1} \quad f(c) = b_{n-1}c + a_n}, \quad (2)$$

так что заодно вычисляется значение f при $X = c$. Рекуррентные формулы (2), в которых и заключается схема Горнера, удобны при счёте.

Ввиду теоремы 1 естественно ввести следующее более общее

Определение. Элемент $c \in A$ называется *k-кратным корнем* (или *k-кратным нулем*) многочлена $f \in A[X]$, если f делится на $(X - c)^k$, но не делится на $(X - c)^{k+1}$. Корень кратности 1 называется *простым корнем* (соответственно при $k = 2$ и $k = 3$ говорят о *двойном* и *тройном корне*).

Итак, $c \in A$ — корень кратности k многочлена $f \in A[X]$ тогда и только тогда, когда $f(X) = (X - c)^k g(X)$, где $\text{НОД}(X - c, g(X)) = 1$. Последнее условие в силу формулы 1 выражается также неравенством $g(c) \neq 0$. Далее, ввиду теоремы 1 § 2 гл. 5 замечаем, что $\deg f = k + \deg g$, откуда $k \leq \deg f$.

Имеет место важная

Теорема 2. Пусть A — целостное кольцо, $f \neq 0$ — многочлен из $A[X]$ и c_1, \dots, c_r — его корни в A кратностей соответственно k_1, \dots, k_r .

Тогда

$$f(X) = (X - c_1)^{k_1} \dots (X - c_r)^{k_r} g(X), \\ g(X) \in A[X], \quad g(c_i) \neq 0, \quad i = 1, \dots, r.$$

В частности, число корней многочлена $f \in A[X]$, рассматривающихся вместе с их кратностями, не превосходит степени многочлена

$$k_1 + k_2 + \dots + k_r \leq \deg f. \quad (3)$$

Доказательство. Достаточно перейти к полю отношений $Q(A)$ (если кольцо A не было полем с самого начала) и воспользоваться однозначностью разложения на простые множители (в данном случае на $X - c_1, \dots, X - c_r$) в кольце $Q(A)[X]$ (результаты § 3 и § 4 гл. 5). Однако необходимости в столь мощном оружии сейчас нет. Будем рассуждать прямо.

Так как $\deg f = (k_1 + \dots + k_r) + \deg g$, то неравенство (3) — следствие делимости f на $(X - c_1)^{k_1} \dots (X - c_r)^{k_r}$, которую мы установим индукцией по r . При $r = 1$ доказывать нечего. Пусть мы уже знаем, что

$$f(X) = (X - c_1)^{k_1} \dots (X - c_{r-1})^{k_{r-1}} h(X).$$

Так как у нас $c_r - c_1 \neq 0, \dots, c_r - c_{r-1} \neq 0$ и A — целостное кольцо, то элемент c_r не является корнем многочлена $(X - c_1)^{k_1} \dots (X - c_{r-1})^{k_{r-1}}$. Но c_r — k_r -кратный корень многочлена f , т.е. $f(X) = (X - c_r)^{k_r} \times u(X)$. Поэтому $h(c_r) = 0$. Соответственно $h(X) = (X - c_r)^s v(X)$, $s \leq k_r$. Имеем

$$(X - c_r)^{k_r} u(X) = f(X) = (X - c_1)^{k_1} \dots (X - c_{r-1})^{k_{r-1}} (X - c_r)^s v(X).$$

Используя закон сокращения в целостном кольце $A[X]$, приходим к заключению, что $s = k_r$. \square

Без предположения о целостности кольца A теорема 2 перестаёт быть верной, как показывает пример многочлена $f(X) = X^3$ над кольцом \mathbb{Z}_8 : $f(0) = f(2) = f(4) = f(6) = 0$. Разложение f на простые множители в $\mathbb{Z}_8(X)$ также неоднозначно: $f = X^3 = X(X - 4)^2 = (X - 2)(X^2 + 2X + 4) = (X - 6)(X^2 - 2X + 4)$.

Из теоремы 2 вытекает

Следствие. Два многочлена $f, g \in A[X]$ степени $\leq n$, принимающие одинаковые значения при подстановке $n+1$ различных элементов из целостного кольца A , равны: $f = g$.

Доказательство. Положим $h = f - g$, так что $\deg h \leq n$. По условию $h(c_1) = \dots = h(c_{n+1}) = 0$ для попарно различных элементов $c_1, \dots, c_{n+1} \in A$, т.е. многочлен h степени $\leq n$ имеет не менее $n+1$ корней. Полученное противоречие с неравенством (3) можно устраниТЬ, лишь признав, что $h = 0$. \square

2. Полиномиальные функции. Следствие теоремы 2 позволяет решить затрагиваемый ранее (см. п. 1 § 2 гл. 5) вопрос о соотношении теоретико-функциональной и алгебраической точек зрения на многочлены. Каждому многочлену $f \in A[X]$ ставится в соответствие функция

$$\tilde{f}: A \rightarrow A, \quad a \mapsto f(a).$$

Множество всех таких функций составляет кольцо A_{pol} полиномиальных (или целых рациональных) функций, являющееся подкольцом кольца функций $A^A = \{A \rightarrow A\}$ с поточечным сложением и умножением (см. пример 3 в п. 1 § 3 гл. 4 и теорему 2 § 2 гл. 5). Совершенно аналогичным образом вводятся полиномиальные функции от нескольких независимых переменных.

Как уже отмечалось ранее, отличный от нуля многочлен $X^2 + X \in F_2[X]$ определяет нулевую функцию. Вообще, если $f(X) = (X^p - X)g(X)$ — многочлен над конечным полем из p элементов, то f — нулевая функция, поскольку $x^p - x = x(x^p - 1) = 0$ для всех $x \in F_p$. Лишь в случае $\deg f \leq p-1$ многочлен $f \in F_p[X]$ определяется своей функцией \tilde{f} . Произвольный многочлен $f \in F_p[X]$ можно заменить однозначно определённым редуцированным многочленом f^* степени $\leq p-1$, взяв в качестве f^* остаток от деления f на $X^p - X$. Тогда, очевидно, $\tilde{f} = \tilde{f}^*$.

В случае бесконечных полей или целостных колец ситуация значительно проще.

Теорема 3. Если A — целостное кольцо с бесконечным числом элементов, то отображение кольца многочленов $A[X]$ на кольцо полиномиальных функций A_{pol} , определяемое соотношением $f \mapsto \tilde{f}$, является изоморфизмом.

Собственно говоря, это есть переформулировка следствия теоремы 2, поскольку речь идёт лишь о том, что многочлену $f \neq 0$ сопоставляется ненулевая функция \tilde{f} , т.е. $\tilde{f}(a) \neq 0$ хотя бы для одного $a \in A$. На самом же деле f имеет не более чем n нулей в A , если $\deg f = n$. \square

На основании теоремы 3 кольцо многочленов над бесконечным полем P отождествляют с кольцом полиномиальных функций (обозначаемых $f(x)$), и остаётся только решить вопрос о том, как по \tilde{f} (а фактически по нескольким значениям многочлена f) восстановить в явном виде сам многочлен.

Точная постановка задачи “интерполяции” заключается в следующем. Пусть b_0, b_1, \dots, b_n — произвольные элементы, а c_0, c_1, \dots, c_n — попарно различные элементы поля P . Требуется найти многочлен $f \in P[X]$ степени $\leq n$ такой, что $f(c_i) = b_i$, $i = 0, 1, \dots, n$. Согласно следствию теоремы 2 решение задачи, если оно существует, единственно. Но один многочлен f с заданными свойствами всегда существует, как показывает *интерполяционная формула Лагранжа*.

$$f(X) = \sum_{i=0}^n b_i \frac{(X - c_0) \dots (X - c_{i-1})(X - c_{i+1}) \dots (X - c_n)}{(c_i - c_0) \dots (c_i - c_{i-1})(c_i - c_{i+1}) \dots (c_i - c_n)}. \quad (4)$$

Впрочем, существование и единственность решения сразу усматриваются из линейной системы

$$\begin{aligned} a_0 c_0^n + a_1 c_0^{n-1} + \dots + a_n &= b_0, \\ \dots &\dots \dots \dots \dots \dots \\ a_0 c_n^n + a_1 c_n^{n-1} + \dots + a_n &= b_n \end{aligned}$$

для коэффициентов a_0, \dots, a_n искомого многочлена f . Определитель этой системы, являющийся определителем Вандермонда, отличен от нуля, и a_i находятся по правилу Крамера. Удобство формулы (4) заключается в её простоте и лёгкости запоминания. Некоторые преимущества иногда имеет *интерполяционная формула Ньютона*

$$f(X) = u_0 + u_1(X - c_0) + \dots + u_n(X - c_0)(X - c_1) \dots (X - c_{n-1}), \quad (5)$$

где коэффициенты u_0, u_1, \dots, u_n определяются путём последовательной подстановки значений $X = c_0, X = c_1, \dots, X = c_n$. Интерполяционные формулы (4), (5) находят практическое применение при вычислении и графическом изображении функции $\varphi: \mathbb{R} \rightarrow \mathbb{R}$, заданной таблично или полученной из опыта. Зная из каких-нибудь косвенных соображений, что функция φ ведёт себя на интервале I вещественной прямой \mathbb{R} “достаточно хорошо”, стараются приблизённо изобразить

изобразить φ на I такой "гладкой" функцией, как полиномиальная (рис. 24). При этом используют в качестве так называемых узлов интерполяции часть точек c_0, c_1, \dots, c_n внутри интервала I , в которых (и только в них) известны значения $\varphi(c_i) = b_i$.

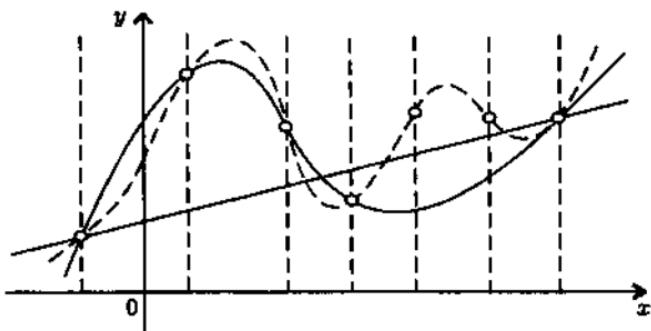


Рис. 24

Тонким вопросам выбора узлов интерполяции и разработки общих методов приближения функций посвящены целые разделы математики. Стоит отметить, что применение интерполяционных процессов сыграло большую роль в развитии теории трансцендентных чисел (определение алгебраических и трансцендентных чисел см. в § 2 гл. 5), так что здесь смыкаются интересы теории функций, теории чисел и алгебры.

Отметим в заключение, что каждой рациональной несократимой дроби $f/g \in P(X)$ (см. § 4 гл. 5) и каждому расширению $F \supset P$ с бесконечным числом элементов сопоставляется *рациональная функция* $f/g : F_{(f/g)} \rightarrow F$ с областью определения $F_{(f/g)}$, получающейся из F удалением конечного числа элементов — нулей многочлена g в F . Можно доказать, что при указанных условиях отображение $f/g \mapsto \widetilde{f/g}$ взаимно однозначно. Нам это утверждение не потребуется. Интуитивно оно ясно. Несмотря на это соответствие, нужно делать чёткое различие между рациональными функциями и рациональными дробями. Рациональная функция $x \mapsto 1/x$ не определена в точке $x = 0$, в то время как вопрос об определимости рациональной дроби $1/X$ вообще не возникает.

3. Дифференцирования кольца многочленов. Функциональная точка зрения на многочлены делает естественным следующее определение. Пусть

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

— многочлен степени n над полем P . Его производной называется многочлен

$$f'(X) = n a_0 X^{n-1} + (n-1) a_1 X^{n-2} + \dots + a_{n-1}. \quad (6)$$

Если $P = \mathbb{R}$ — поле вещественных чисел, а \tilde{f} — связанная с f полиномиальная функция, то определение (6) производной совпадает с обычным её определением как предела

$$\lim_{\Delta x \rightarrow 0} \frac{\tilde{f}(x + \Delta x) - \tilde{f}(x)}{\Delta x}.$$

В случае же произвольного поля P говорить о каких-либо свойствах непрерывности полиномиальной функции бессмысленно (что такое сходящаяся последовательность в \mathbb{Z}_p ?) и нужно исходить из формального определения (6).

Имеют место хорошо известные из анализа соотношения

$$(\alpha f + \beta g)' = \alpha f' + \beta g', \quad \alpha, \beta \in P, \quad (7)$$

$$(fg)' = f'g + fg'. \quad (8)$$

Соотношение (7) прямо вытекает из (6) и из определения суммы многочленов. Используя (7) и определение произведения многочленов, проверку (8) можно свести к тому случаю, когда $f = X^k$, $g = X^l$:

$$\begin{aligned} (X^{k+l})' &= (k+l)X^{k+l-1} = (kX^{k-1})X^l + X^k(lX^{l-1}) = \\ &= (X^k)'X^l + X^k(X^l)' \end{aligned}$$

Обобщением (8) служит легко доказываемая индукцией по k формула

$$(f_1 f_2 \dots f_k)' = \sum_{i=1}^k f_1 \dots f_{i-1} f_i' f_{i+1} \dots f_k.$$

В частности,

$$(f^k)' = k f^{k-1} f'. \quad (9)$$

Соотношения (7), (8), переписанные в терминах отображения $\frac{d}{dX} : f \mapsto f'$ (говорят также, что $\frac{d}{dX}$ — оператор дифференцирования), наводят на мысль ввести в рассмотрение для произвольного кольца K отображение $\mathcal{D} : K \rightarrow K$, обладающее свойствами

$$\mathcal{D}(u + v) = \mathcal{D}u + \mathcal{D}v, \quad (7')$$

$$\mathcal{D}(uv) = (\mathcal{D}u)v + u(\mathcal{D}v). \quad (8')$$

Такого рода отображения кольца K в себя, называемые дифференцированиями, весьма полезны для изучения K , а их множество $\text{Der}(K)$ оказывается интереснейшим объектом, вводящим в обширную область математики (группы и алгебры Ли).

Обобщением (8') служит формула Лейбница

$$\mathcal{D}^m(uv) = \sum_{k=0}^m \binom{m}{k} \mathcal{D}^k u \mathcal{D}^{m-k} v, \quad (8'')$$

получаемая индукцией по $m \geq 1$ (применение \mathcal{D} к $(8'')$, использование $(8')$ и соотношение $\binom{m}{k-1} + \binom{m}{k} = \binom{m+1}{k}$ дадут $(8'')$ при $m+1$).

В случае $K = P[X]$ из соотношений $(7')$, $(8')$, дополненных правилом

$$\mathcal{D}(\lambda f) = \lambda \mathcal{D}f, \quad \lambda \in P,$$

непосредственно вытекает

$$\mathcal{D}f(X) = f'(X)\mathcal{D}X.$$

Стало быть, любое дифференцирование кольца многочленов $P[X]$ определяется заданием единственного многочлена $\mathcal{D}X$. При $\mathcal{D}X = 1$ мы получаем обычный оператор дифференцирования $\frac{d}{dX}$.

4. Кратные множители. Результат m -кратного применения отображения $\frac{d}{dX}$ к $f(X)$ обычно обозначается символом $f^{(m)}(X)$. Очевидно, что

$$f(X) = a_0 X^n + a_1 X^{n-1} + \dots + a_n \implies f^{(n)}(X) = n! a_0, \quad f^{(n+1)}(X) = 0.$$

Если P — поле нулевой характеристики, то

$$\deg f' = \deg f - 1.$$

Однако для полей конечной характеристики p это уже не так, поскольку

$$(X^{kp})' = kpX^{kp-1} = 0.$$

Всё же некоторую пользу из рассмотрения производной можно извлечь и в общем случае. Разделив произвольный многочлен $f \in P[X]$ на $(X - c)^2$, $c \in F$, $F \supset P$, а затем записав (линейный) остаток в виде $(X - c)s + r$, где $s, r \in F$, мы придём к соотношениям $f = (X - c)^2t + (X - c)s + r$, $f' = (X - c)[2t + (X - c)t'] + s$. Подставив в них значение $X = c$, получим $r = f(c)$, $s = f'(c)$, т.е.

$$f(X) = (X - c)^2t(X) + (X - c)f'(c) + f(c).$$

Мы пришли к следующему утверждению.

Теорема 4. Пусть P — произвольное поле и F — любое его расширение. Многочлен $f \in P[X]$ имеет кратный корень $c \in F$ тогда и только тогда, когда $f(c) = f'(c) = 0$.

Пример 1. В любом поле характеристики p многочлен $X^n - 1$ имеет лишь простые корни, если n не делится на p . Действительно, корни производной pX^{n-1} не могут быть корнями $X^n - 1$.

Далее предполагается, что P — поле нулевой характеристики, и без ограничения общности под P можно понимать одно из полей \mathbb{Q} , \mathbb{R} или \mathbb{C} . Нормализованный неприводимый многочлен $p_i(X)$ в разложении

$$f(X) = \lambda p_1(X)^{k_1} \dots p_i(X)^{k_i} \dots p_r(X)^{k_r}, \quad \lambda \in P, \quad (10)$$

многочлена $f(X) \in P[X]$ (по аналогии с определением кратного корня) называется k_i -кратным множителем для f . Ранее уже говорилось о том, что получить разложение (10) на практике довольно сложно. Опишем вкратце метод, основанный на понятии производной и дающий возможность узнать, содержит ли $f(X)$ над данным полем P (или над его расширением) кратные множители.

Теорема 5. Пусть $p(X)$ есть k -кратный неприводимый множитель многочлена $f \in P[X]$ ($k \geq 1$, $\deg p(X) \geq 1$).

Тогда $p(X)$ будет $(k-1)$ -кратным множителем производной $f'(X)$. В частности, при $k=1$ f' не делится на $p(X)$.

Доказательство. По условию имеем $f(X) = p(X)^k g(X)$, где $\text{НОД}(p(X), g(X)) = 1$, т.е. $g(X)$ не делится на $p(X)$. Применяя правила (8) и (9), находим

$$f'(X) = p(X)^{k-1} [kp'(X)g(X) + p(X)g'(X)].$$

Достаточно показать, что многочлен, стоящий в квадратных скобках, не делится на $p(X)$. Если бы это было не так, то на $p(X)$ делился бы многочлен $kp'(X)g(X)$, что, однако, невозможно (см. следствия теорем 3 и 4 § 3 гл. 5), поскольку $g(X)$ не делится на $p(X)$, а $\deg kp'(X) < \deg p(X)$. \square

Понятно, что в ходе доказательства существенно использованы и неприводимость $p(X)$, и условие $\text{char } P = 0$.

Следствие 1. Для многочлена $f(X)$ с коэффициентами в поле P характеристики нуль следующие два условия эквивалентны:

- i) f имеет в некотором расширении $F \supset P$ поля P корень с кратностью k ;
- ii) $f^{(j)}(c) = 0$, $0 \leq j \leq k-1$, но $f^{(k)}(c) \neq 0$.

Для доказательства нужно k раз применить теорему 5, имея в виду линейный множитель $p(X) = X - c$ и с самого начала замечая, в случае необходимости, P на его расширение F , содержащее корень c . \square

Следствие 2. Если многочлен $f \in P[X]$ степени ≥ 1 имеет разложение (10), то разложением для наибольшего общего делителя f и его производной f' будет

$$\text{НОД}(f, f') = p_1(X)^{k_1-1} p_2(X)^{k_2-1} \dots p_r(X)^{k_r-1} \quad (11)$$

(НОД всегда можно считать нормализованным многочленом).

Действительно, по теореме 5 каждый из простых делителей $p_i(X)$ многочлена $f(X)$ с каноническим разложением (10) входит в разложение $f'(X)$ с показателем $k_i - 1$, т.е.

$$f'(X) = p_1(X)^{k_1-1} p_2(X)^{k_2-1} \dots p_r(X)^{k_r-1} \cdot u(X),$$

где $\text{НОД}(u, p_i) = 1$, $1 \leq i \leq r$ (предполагается, что $p_i(X)^0 = 1$). Поэтому по известному нам признаку делимости (см. п. 2 § 3 гл. 5) мы заключаем, что $\text{НОД}(f, f')$ вычисляется по формуле (11). \square

Используя выражение (11) для НОД(f, f'), мы получаем средство освободиться от кратных множителей, входящих в разложение $f(X)$. Именно, многочлен

$$g(X) = \frac{f(X)}{\text{НОД}(f, f')} = p_1(X)p_2(X) \dots p_r(X)$$

содержит те же простые делители, что и $f(X)$, но с единичной кратностью. Важно отметить, что многочлен $g(X)$ можно найти, не зная фактически разложений для f и f' , а лишь используя алгоритм Евклида.

Пример 2. Многочлен $f(X) = X^5 - 3X^4 + 2X^3 + 2X^2 - 3X + 1$ и его производная $f'(X) = 5X^4 - 12X^3 + 6X^2 + 4X - 3$ имеют в качестве НОД нормализованный многочлен $X^3 - 3X^2 + 3X - 1 = (X - 1)^3$. “Свободный от квадратов” многочлен $g(X) = f(X)/(X - 1)^3 = X^2 - 1 = (X - 1)(X + 1)$ имеет два корня: ± 1 . Таким образом, $f(X) = (X - 1)^4(X + 1)$ обладает корнем $+1$ кратности 4 и простым корнем -1 .

5. Формулы Виета. В связи с теорией систем линейных уравнений мы уже имели случай упомянуть о благотворном влиянии на её развитие хорошей системы обозначений, приведшей, в частности, к определителям. Это заслуга математиков XVIII в. и начала XIX в. Но гораздо раньше, когда алгебра ещё отождествлялась с “анализом уравнений”, решающее усовершенствование алгебраических обозначений у Ф. Виета и у Р. Декарта коснулось теории многочленов и алгебраических уравнений. От частных типов уравнений с числовыми коэффициентами, скрывавшими общие закономерности, был совершен смелый переход к уравнениям с буквенными коэффициентами. Новый способ записи нередко порождает новые результаты. У Декарта это завершилось революционным применением алгебры к геометрии. Мы остановимся на более скромном достижении его предшественника Виета.

Предположим, что нормализованный многочлен $f \in P[X]$ степени n имеет в поле P или в некотором его расширении n корней c_1, c_2, \dots, c_n , среди которых, возможно, есть и одинаковые. Тогда в соответствии с теоремой 2 справедливо разложение

$$f(X) = (X - c_1)(X - c_2) \dots (X - c_n).$$

Запишем $f(X)$ в обычном виде по степеням X :

$$f(X) = X^n + a_1 X^{n-1} + \dots + a_k X^{n-k} + \dots + a_n,$$

для чего перемножим все двучлены $X - c_i$ и приведём подобные члены. Тогда для коэффициентов a_1, \dots, a_n получатся выражения через

корни c_1, \dots, c_n :

$$\begin{aligned} a_1 &= -(c_1 + c_2 + \dots + c_n), \\ a_k &= (-1)^k \sum_{i_1 < i_2 < \dots < i_k} c_{i_1} c_{i_2} \dots c_{i_k}, \\ a_n &= (-1)^n c_1 c_2 \dots c_n. \end{aligned} \quad (12)$$

Формулы (12) называются *формулами Виета*.

Если бы многочлен f не был нормализованным, т.е. имел старший коэффициент $a_0 \neq 1$, то формулы, аналогичные (12), давали бы выражения для отношений a_i/a_0 .

Формулы Виета, устанавливающие явную связь между корнями и коэффициентами произвольного многочлена, замечательны тем, что их правые части не меняются при любых перестановках корней c_1, \dots, c_n . Это даёт нам повод ввести понятие *симметрической функции* подобно тому, как в связи с определителями оказалось удобным рассматривать общие кососимметрические функции. Согласно определению, приведённому в п. 4 § 8 гл. 1, элемент π симметрической группы S_n действует на функцию $\tilde{f}(x_1, \dots, x_n)$ от n аргументов по правилу

$$(\widetilde{\pi \circ f})(x_1, \dots, x_n) = \tilde{f}(x_{\pi(1)}, \dots, x_{\pi(n)}).$$

Функция f называется *симметрической*, если $\widetilde{\pi \circ f} = \tilde{f}$ для всех $\pi \in S_n$. Примером симметрических функций служат так называемые *элементарные симметрические функции* s_k :

$$s_k(x_1, \dots, x_n) = \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} x_{i_1} x_{i_2} \dots x_{i_k}. \quad (13)$$

Они позволяют переписать формулы (12) в виде

$$a_k = (-1)^k s_k(c_1, \dots, c_n), \quad k = 1, 2, \dots, n, \quad (12')$$

так что с точностью до знака коэффициент a_k многочлена f есть значение функции s_k на множестве корней многочлена f . Обратим внимание на тот факт, что по определению $a_k \in P$, хотя корни c_1, \dots, c_n , вообще говоря, лежат в некотором расширении $F \supset P$. Вопрос о существовании F мы сейчас не затрагиваем. Но иногда разложение многочлена на линейные множители является прямым следствием свойств поля P .

Пример 3. Рассмотрим многочлен $X^{p-1} - 1$ над конечным полем F_p (см. п. 6 § 4 гл. 4). Мы знаем, что $x^{p-1} = 1$ для всех $x \in F_p^*$, т.е. все ненулевые элементы — корни многочлена $X^{p-1} - 1$. Стало быть, имеет место разложение

$$X^{p-1} - 1 = (X - 1)(X - 2) \dots (X - (p-1)). \quad (14)$$

Предполагается, что мы уже настолько освоились с полем \mathbb{F}_p , что без труда различаем двойственную природу чисел $1, 2, \dots, p-1$ как обычных элементов из \mathbb{Z} и как элементов поля $\mathbb{F}_p \cong \mathbb{Z}/p\mathbb{Z}$ (представителей классов вычетов $\{i\}_p$). Из (12') и (14) при $p > 2$ получаем

$$\begin{aligned}s_k(1, 2, \dots, p-1) &\equiv 0 \pmod{p}, \quad k = 1, 2, \dots, p-2, \\ s_{p-1}(1, 2, \dots, p-1) &\equiv -1 \pmod{p}.\end{aligned}$$

Последнее соотношение, переписанное в виде

$$(p-1)! + 1 \equiv 0 \pmod{p} \quad (15)$$

и как таковое известное по названию *теоремы Вильсона*, выражает фактически необходимый и достаточный признак простоты целого числа p . Действительно, выполнение (15) для простых p мы только что доказали. С другой стороны, $p = p_1 p_2 \implies (p-1)! = p_1! \implies (p-1)! + 1 \not\equiv 0 \pmod{p_1} \implies (p-1)! + 1 \not\equiv 0 \pmod{p}$.

УПРАЖНЕНИЯ

1. Будет ли кольцо полиномиальных функций над полем из p элементов цестным?

2. Пусть P — бесконечное поле и f — ненулевой многочлен из $P[X_1, \dots, X_n]$. Опираясь на теорему 3 и используя индукцию по n , доказать существование $a_1, \dots, a_n \in P$, для которых $f(a_1, \dots, a_n) \neq 0$. Это даёт изоморфизм $P[X_1, \dots, X_n]$ с кольцом полиномиальных функций от n переменных над P .

3. Ненулевой многочлен $f \in \mathbb{Z}_p[X_1, \dots, X_n]$ степени $< p$ по каждой переменной обладает сформулированным в упр. 2 свойством: $f(a_1, \dots, a_n) \neq 0$ для некоторых $a_1, \dots, a_n \in \mathbb{Z}_p$. Показать, что любой многочлен $f \in \mathbb{Z}_p[X_1, \dots, X_n]$ можно записать в виде

$$f(X_1, \dots, X_n) = \sum_{i=1}^n g_i(X_1, \dots, X_n)(X_i^p - X_i) + f^*(X_1, \dots, X_n),$$

где f^* — редуцированный многочлен ($\deg_{X_i} f^* \leq p-1$, $i = 1, 2, \dots, n$) степени $\deg f^* \leq \deg f$. Сделать обоснованное заключение, что отображение $f \mapsto \bar{f} = \bar{f}^*$ является эпиморфизмом кольца $\mathbb{Z}_p[X_1, \dots, X_n]$ на кольцо полиномиальных функций от n переменных над \mathbb{Z}_p с ядром $L = \sum_{i=1}^n (X_i^p - X_i) \mathbb{Z}_p[X_1, \dots, X_n]$.

4. Теорема (Шевалле). Пусть $f(X_1, \dots, X_n)$ — однородный многочлен (форма) над \mathbb{Z}_p степени $r < n$. Тогда уравнение $f(x_1, \dots, x_n) = 0$ имеет хотя бы одно нетривиальное решение.

Указание. Так как f — форма, то, очевидно, $f(0, \dots, 0) = 0$. Рассуждая от противного, предположить, что $(a_1, \dots, a_n) \neq (0, \dots, 0) \implies f(a_1, \dots, a_n) \neq 0$. При помощи упр. 3 и малой теоремы Ферма вывести отсюда, что редуцированным многочленом для $g(X_1, \dots, X_n) = 1 - f(X_1, \dots, X_n)^{p-1}$ будет $g^*(X_1, \dots, X_n) = (1 - X_1^{p-1}) \dots (1 - X_n^{p-1})$. Но

$$\deg g = (p-1) \deg f = (p-1)r < (p-1)n = \deg g^*.$$

Полученное противоречие доказывает теорему.

Несколько изменяв рассуждение (вычислив сумму $\sum_{x_1, \dots, x_n \in \mathbb{Z}_p} g(x_1, \dots, x_n)$ двумя способами), доказать, что общее число решений всегда делится на p .

5. Пусть $f(x_1, \dots, x_n)$ — целочисленная квадратичная форма. Теорема Шевалле (см. упр. 4), сформулированная на языке теории сравнений, утверждает, что при $n \geq 3$ сравнение

$$f(x_1, \dots, x_n) \equiv 0 \pmod{p}$$

имеет некульевое решение. Проверить, что все решения сравнения $x^2 - 2y^2 \equiv 0 \pmod{5}$ тривиальны и, следовательно, условие $r < n$ существенно.

6. Показать, что $\text{НОД}(f', f) = 1$, если $\text{char } P = 0$, f — неприводимый над полем P многочлен и f' — его производная.

7. Доказать, что $f' = 0 \implies f = \text{const}$ для многочлена $f(X)$ над полем нулевой характеристики и $f' = 0 \implies f(X) = g(X^p)$ для многочлена $f(X)$ над полем характеристики $p > 0$ (g — некоторый другой многочлен).

8. Из п. 3 мы знаем, что каждое дифференцирование кольца многочленов $P[X]$ имеет вид

$$T_u: f \mapsto uf', \quad u \in P[X].$$

Установить справедливость утверждений:

i) множество констант (то, что переходит при дифференцировании в нуль) — подкольцо в $P[X]$;

ii) произведение $T_u T_v$, вообще говоря, не является дифференцированием, но если $\text{char } P = p > 0$, то степень $(T_u)^p$ — дифференцирование;

iii) коммутатор $[T_u, T_v] = T_u T_v - T_v T_u$ всегда является дифференцированием вида T_w , где $w = uv' - u'v$.

9. В случае кольца многочленов $P[X_1, \dots, X_n]$ от n переменных естественно ввести оператор частного дифференцирования по k -й переменной

$$\frac{\partial}{\partial X_k}: X_1^{i_1} \dots X_k^{i_k} \dots X_n^{i_n} \mapsto i_k X_1^{i_1} \dots X_k^{i_k-1} \dots X_n^{i_n}.$$

i) Показать, что множеством констант для $\frac{\partial}{\partial X_k}$ служит кольцо многочленов $P[X_1, \dots, \hat{X}_k, \dots, X_n]$ от $n-1$ переменной ($\text{char } P = 0$).

ii) Пусть $f(X_1, \dots, X_n)$ — форма (однородный многочлен) степени m . Убедиться в справедливости тождества Эйлера

$$\sum_{k=1}^n X_k \frac{\partial f}{\partial X_k} = m \cdot f(X_1, \dots, X_n).$$

Обратно: если $\text{char } P = 0$, то тождество Эйлера удовлетворяют только формы степени $m = 1, 2, 3, \dots$

10. Показать, что отсутствие линейных множителей у многочлена

$$X^n + a_1 X^{n-1} + \dots + a_n \in \mathbb{Z}_2[X]$$

равносильно выполнению условия

$$a_n(1 + \sum a_i) \neq 0.$$

При $n \leq 3$ неприводимые многочлены над \mathbb{Z}_2 исчерпываются следующими:

$$X, \quad X+1, \quad X^2+X+1, \quad X^3+X+1, \quad X^3+X^2+1.$$

Выписать все неприводимые многочлены над \mathbb{Z}_2 при $n = 4$ и $n = 5$ (их будет соответственно 3 и 6).

11. Исходя из сравнения

$$X^5 - X - 1 \equiv (X^3 + X^2 + 1)(X^2 + X + 1) \pmod{2},$$

установить неприводимость многочлена $X^5 - X - 1$ над \mathbb{Q} .

Указание. Применить следствие леммы Гаусса (§ 3 гл. 5) и предыдущее упражнение, а также сослаться на факториальность кольца $\mathbb{Z}_2[X]$.

Аналогично, доказать неприводимость многочлена $X^5 - X - 1$ над \mathbb{Q} , перейдя к сравнению по модулю 3 (это гораздо проще).

§ 2. Симметрические многочлены

1. Кольцо симметрических многочленов. Следуя определению симметрических функций¹⁾, которое было дано в конце предыдущего параграфа, мы введём аналогичное понятие в кольце $A[X_1, \dots, X_n]$ многочленов над целостным кольцом A . Теорема 3 из § 1, распространённая на многочлены и функции многих переменных, как будто делает такое перенесение излишним. Но следует учесть, что в этой теореме целостное кольцо A коэффициентов бесконечно, а нам хочется иметь универсальную конструкцию.

Итак, полагаем

$$(\pi \circ f)(X_1, \dots, X_n) = f(X_{\pi(1)}, \dots, X_{\pi(n)}).$$

Многочлен f называется *симметрическим*, если $\pi \circ f = f$ для всех $\pi \in S_n$. Как и для функций, вводятся элементарные симметрические многочлены s_k :

$$s_k(X_1, \dots, X_n) = \sum_{\substack{1 \leq i_1 < i_2 < \dots < i_k \leq n \\ k=1, 2, \dots, n}} X_{i_1} X_{i_2} \dots X_{i_k}, \quad (1)$$

Строго говоря, следовало бы рассмотреть многочлен

$$\begin{aligned} f(Y) &= (Y - X_1)(Y - X_2) \dots (Y - X_n) = \\ &= Y^n - s_1 Y^{n-1} + s_2 Y^{n-2} + \dots + (-1)^n s_n \end{aligned} \quad (2)$$

над $A[X_1, \dots, X_n]$ от новой переменной Y и заметить, что s_k — симметрический многочлен, поскольку левая часть тождества (2) не меняется при любых перестановках линейных множителей $Y - X_1, \dots, Y - X_n$.

Обратим внимание на то обстоятельство, что после подстановки нуля вместо X_n в обе части тождества (2) мы получим

$$(Y - X_1) \dots (Y - X_{n-1})Y = Y^n - (s_1)_0 Y^{n-1} + \dots + (-1)^{n-1} (s_{n-1})_0 Y,$$

где $(s_k)_0$ — результат подстановки $X_n = 0$ в s_k . Сокращая обе части на Y (на основании теоремы 1 из § 3 гл. 4, применённой к $A[X_1, \dots, X_n, Y]$), приходим к тождеству

$$\begin{aligned} (Y - X_1)(Y - X_2) \dots (Y - X_{n-1}) &= \\ &= Y^{n-1} - (s_1)_0 Y^{n-2} + \dots + (-1)^{n-1} (s_{n-1})_0. \end{aligned} \quad (3)$$

¹⁾ В [BA II] группа S_n по-прежнему называется симметрической, но многочлены и функции — *симметричными* или *кососимметричными* в зависимости от того, остаются они при действии S_n на месте или приобретают множитель -1 . Такая терминология лучше отвечает сути дела, но традиции гораздо сильнее, поэтому мы оставляем читателю свободу выбора.

Сравнивая (2) и (3), мы приходим к выводу, что $(s_1)_0, \dots, (s_{n-1})_0$ — элементарные симметрические многочлены от $n - 1$ переменных X_1, \dots, X_{n-1} .

Так как, далее, $\tilde{\pi}: f \mapsto \pi \circ f$ — автоморфизм кольца $A[X_1, \dots, X_n]$, то любые линейные комбинации симметрических многочленов и их произведения будут снова симметрическими многочленами. Это значит, что множество всех симметрических многочленов образует кольцо, являющееся подкольцом кольца $A[X_1, \dots, X_n]$. Наша ближайшая цель — разобраться, как устроено это подкольцо.

2. Основная теорема о симметрических многочленах. Оказывается, что наиболее общим способом получения симметрических многочленов является следующий. Нужно взять произвольный многочлен $g \in A[Y_1, \dots, Y_n]$ и подставить вместо Y_1, \dots, Y_n соответственно s_1, \dots, s_n . Получившийся в результате многочлен будет, конечно, симметрическим.

Заметим еще, что одночлен $Y_1^{i_1} \dots Y_n^{i_n}$, входящий в g , переходит при подстановке $Y_k = s_k(X_1, \dots, X_n)$ в однородный многочлен от X_1, \dots, X_n степени $i_1 + 2i_2 + \dots + ni_n$, поскольку $\deg s_k = k$. Сумму $i_1 + 2i_2 + \dots + ni_n$ называют обычно весом одночлена $Y_1^{i_1} \dots Y_n^{i_n}$. Весом многочлена $g(Y_1, \dots, Y_n)$ естественно считать максимум весов одночленов, входящих в g .

Основное утверждение о симметрических многочленах выражает

Теорема 1. Пусть $f \in A[X_1, \dots, X_n]$ — симметрический многочлен полной степени m над целостным кольцом A .

Тогда существует, и примет единственный, многочлен $g \in A[Y_1, \dots, Y_n]$ веса m , для которого

$$f(X_1, \dots, X_n) = g(s_1, \dots, s_n).$$

Коэффициенты многочлена g являются целочисленными линейными комбинациями коэффициентов исходного многочлена f .

Доказательство. В своё время (см. гл. 5 § 2) мы отмечали, что любой многочлен $f = f(X_1, \dots, X_n)$ можно записать в виде суммы однородных форм f_m различных степеней: $f = f_0 + f_1 + \dots + f_k$. Очевидно, что эта запись единственна. Если теперь f — симметрический многочлен, то симметрическими будут и формы f_m , поскольку $\pi \circ f = \sum \pi \circ f_m$, а действие $\tilde{\pi}: f_m \mapsto \pi \circ f_m$ на степень m формы f_m не влияет. Таким образом, без ограничения общности симметрический многочлен f можно считать однородным. Дальнейшие рассуждения разобьём на несколько частей.

1. Условимся располагать одночлены в f лексикографически (по принципу построения словаря), т.е. таким образом, что одночлен $u = aX_1^{i_1} X_2^{i_2} \dots X_n^{i_n}$ предшествует одночлену $v = bX_1^{j_1} X_2^{j_2} \dots X_n^{j_n}$ (или большему одночлену v : $u > v$) в точности тогда, когда последовательность $i_1 - j_1, i_2 - j_2, \dots, i_n - j_n$ имеет вид $0, \dots, 0, t, \dots$, где $t > 0$.

Справа от t могут стоять и отрицательные разности $i_j - j_1$. Одночлен, входящий в f и занимающий первое место при лексикографическом упорядочении, называется высшим членом многочлена f . Обозначим его $\text{ВЧ}(f)$.

Лемма 1. Высшим членом произведения $h = h_1 h_2 \dots h_r$ является произведение высших членов сомножителей h_1, h_2, \dots, h_r .

Действительно, при $n = 1$ утверждение верно, а если

$$h = h(X_1, X_2, \dots, X_n) = g_0(X_2, \dots, X_n)X_1^s + g_1(X_2, \dots, X_n)X_1^{s-1} + \dots,$$

то $\text{ВЧ}(h) = X_1^s \cdot \text{ВЧ}(g_0)$. Взяв теперь разложение по степеням X_1 каждого сомножителя h_i и обратив внимание на то, как получается коэффициент $g_0(X_2, \dots, X_n)$, мы при помощи естественной математической индукции по n придём к нужному выражению $\text{ВЧ}(h) = \prod_{i=1}^r \text{ВЧ}(h_i)$. \square

2. Одночлен $u = aX_1^{i_1} X_2^{i_2} \dots X_n^{i_n}$ условимся называть монотонным, если $i_1 \geq i_2 \geq \dots \geq i_n$.

Лемма 2. Высший член симметрического многочлена всегда монотонен.

В самом деле, пусть $\text{ВЧ}(f) = u = aX_1^{i_1} X_2^{i_2} \dots X_n^{i_n}$. Допустим, что $i_k < i_{k+1}$ при некотором $k \leq n-1$. Переставив в $u = aX_1^{i_1} \dots X_k^{i_k} \times X_{k+1}^{i_{k+1}} \dots X_n^{i_n}$ местами переменные $X_k^{i_k}$ и X_{k+1} , мы получим одночлен $u' = aX_1^{i_1} \dots X_k^{i_{k+1}} X_{k+1}^{i_k} \dots X_n^{i_n}$, из-за симметричности f тоже входящий в f . Но, очевидно, $u' > u$, поскольку показатели при X_1, \dots, X_{k-1} в u , u' одинаковые, а показатель при X_k в u' больше показателя при X_k в u . Полученное противоречие доказывает лемму. \square

3. Существование многочлена $g(Y_1, \dots, Y_n)$. Предположим снова, что $u = aX_1^{i_1} X_2^{i_2} \dots X_n^{i_n} = \text{ВЧ}(f)$. В силу леммы 2 $i_k \geq i_{k+1}$, $1 \leq k \leq n-1$. Поэтому мы можем ввести в рассмотрение симметрический многочлен

$$f_{(1)}(X_1, \dots, X_n) = f(X_1, \dots, X_n) - as_1^{i_1-i_2}s_2^{i_2-i_3} \dots s_n^{i_n},$$

отвечающий одночлену $aY_1^{i_1-i_2} Y_2^{i_2-i_3} \dots Y_n^{i_n}$ веса $(i_1 - i_2) + 2(i_2 - i_3) + \dots + (n-1)(i_{n-1} - i_n) + ni_n = i_1 + i_2 + \dots + i_n = \deg f$. Так как высшими членами элементарных симметрических многочленов s_1, s_2, \dots, s_n являются, очевидно, $X_1, X_1 X_2, \dots, X_1 X_2 \dots X_n$, то по лемме 1 высшим членом в $as_1^{i_1-i_2}s_2^{i_2-i_3} \dots s_n^{i_n}$ будет

$$\begin{aligned} aX_1^{i_1-i_2}(X_1 X_2)^{i_2-i_3} \dots (X_1 X_2 \dots X_{n-1})^{i_{n-1}-i_n}(X_1 X_2 \dots X_n)^{i_n} = \\ = aX_1^{i_1} X_2^{i_2} \dots X_n^{i_n}, \end{aligned}$$

т.е. в точности $u = \text{ВЧ}(f)$. Стало быть, он сокращается, в $f_{(1)}$ не входит и $\text{ВЧ}(f) > \text{ВЧ}(f_{(1)})$. Отметим ещё, что коэффициенты мно-

многочлена $f_{(1)}$ имеют вид $c - qa$, где c, a — коэффициенты многочлена f , $q \in \mathbb{Z}$.

Пусть $v = bX_1^{j_1}X_2^{j_2}\dots X_n^{j_n} = \text{ВЧ}(f_{(1)})$, $b \in A$. Снова по лемме 2 имеем $j_1 \geq j_2 \geq \dots \geq j_n$ и, по изложенным выше соображениям, для симметрического многочлена

$$f_{(2)}(X_1, \dots, X_n) = f_{(1)}(X_1, \dots, X_n) - bs_1^{j_1-j_2}s_2^{j_2-j_3}\dots s_n^{j_n}$$

получаем $\text{ВЧ}(f_{(1)}) > \text{ВЧ}(f_{(2)})$. Кроме того, коэффициенты многочлена $f_{(2)}$ имеют вид $c_1 - q_1 b$, где $q_1 \in \mathbb{Z}$, а c_1, b — коэффициенты многочлена $f_{(1)}$.

Продолжая этот процесс, мы придём к последовательности однородных симметрических многочленов

$$f_{(k)} = f - as_1^{i_1-i_2}\dots s_n^{i_n} - bs_1^{j_1-j_2}\dots s_n^{j_n} - \dots$$

степени $\deg f_{(k)} = \deg f$, для которых

$$\text{ВЧ}(f) > \text{ВЧ}(f_{(1)}) > \text{ВЧ}(f_{(2)}) > \dots > \text{ВЧ}(f_{(k)}) > \dots, \quad (4)$$

причём коэффициенты у $f_{(k)}$ будут \mathbb{Z} -линейными комбинациями коэффициентов многочлена f . Так как одночленов фиксированной степени (и тем более монотонных) конечное число, то цепочка неравенств (4) должна оборваться ($f_{(k)} = 0$ при некотором k); мы получим требуемое выражение $f(X_1, \dots, X_n) = g(s_1, \dots, s_n)$, где $g(Y_1, \dots, Y_n) = aY_1^{i_1-i_2}Y_2^{i_2-i_3}\dots Y_n^{i_n} + bY_1^{j_1-j_2}Y_2^{j_2-j_3}\dots Y_n^{j_n} + \dots$

4. Единственность. В случае существования двух различных представлений $f = g_1(s_1, \dots, s_n) = g_2(s_1, \dots, s_n)$ мы имели бы отличный от нуля многочлен $g(Y_1, \dots, Y_n) = g_1(Y_1, \dots, Y_n) - g_2(Y_1, \dots, Y_n)$ веса $\deg f$, для которого $g(s_1, \dots, s_n) = 0$. Если $aY_1^{k_1}Y_2^{k_2}\dots Y_n^{k_n}$ — одночлен, входящий в g , то, как мы видели, $\text{ВЧ}(as_1^{k_1}\dots s_n^{k_n}) = aX_1^{k_1}(X_1X_2)\dots(X_1\dots X_n)^{k_n} = aX_1^{k_1+k_2+\dots+k_n}X_2^{k_2+\dots+k_n}\dots X_n^{k_n}$. Ясно поэтому, что различным одночленам, входящим в g , отвечают различные высшие члены. Среди них один будет самым высшим, и, стало быть, $\text{ВЧ}(g(s_1, \dots, s_n)) \neq 0$ вопреки предположению. \square

На другом языке утверждение о единственности означает, что s_1, \dots, s_n алгебраически независимы над A , а кольца $A[s_1, \dots, s_n]$ и $A[X_1, \dots, X_n]$ изоморфны (хотя, разумеется, $A[s_1(X_1, \dots, X_n), \dots, s_n(X_1, \dots, X_n)]$ — собственное подмножество в $A[X_1, \dots, X_n]$). Между прочим, при $A = \mathbb{Z}$ коэффициентами многочленов f и g будут целые числа. Из теоремы 1 вытекает ещё полезное

Следствие. Пусть $f(X) = X^n + a_1X^{n-1} + \dots + a_{n-1}X + a_n$ — нормализованный многочлен степени n от одной переменной X над полем P , имеющий n корней c_1, \dots, c_n в некотором большем поле $F \supset P$. Пусть, далее, $h(X_1, \dots, X_n)$ — произвольный симметрический многочлен из $P[X_1, \dots, X_n]$.

Тогда его значение $h(c_1, \dots, c_n)$, получающееся при подстановке c_i вместо X_i , $i = 1, \dots, n$, будет принадлежать полю P .

Доказательство. В самом деле, по основной теореме о симметрических многочленах найдётся многочлен $g(Y_1, \dots, Y_n) \in P[Y_1, \dots, Y_n]$ такой, что

$$h(X_1, \dots, X_n) = g(s_1(X_1, \dots, X_n), \dots, s_n(X_1, \dots, X_n)).$$

Поэтому $h(c_1, \dots, c_n) = g(s_1(c_1, \dots, c_n), \dots, s_n(c_1, \dots, c_n))$, а так как в соответствии с формулами Виета (12) из § 1 $s_k(c_1, \dots, c_n) = (-1)^k a_k \in P$, то и $g(-a_1, \dots, (-1)^n a_n) \in P$. \square

3. Метод неопределённых коэффициентов. Существует несколько различных доказательств основной теоремы о симметрических многочленах, а соответственно и методов выражения заданного многочлена f через элементарные симметрические многочлены. Чтобы описать один из таких наиболее употребительных методов, введём новый тип симметрических многочленов. Для определённости будем брать в качестве A кольцо \mathbb{Z} или поле \mathbb{R} . Пусть $v = X_1^{i_1} X_2^{i_2} \dots X_n^{i_n}$ — какой-то одночлен. Обозначим через $S(v)$ сумму всех различных одночленов, получающихся из v перестановкой независимых переменных. Например,

$$s_k(X_1, \dots, X_n) = S(X_1 \dots X_k)$$

— k -й элементарный симметрический многочлен. Далее,

$$p_k(X_1, \dots, X_n) = S(X_1^k) = X_1^k + X_2^k + \dots + X_n^k, \quad k \geq 0, \quad (5)$$

есть так называемая k -я степенная сумма. Понятно, что всегда $S(v)$ — однородный симметрический многочлен (называемый ещё моногенным) той же полной степени, что и v . Так как $S(v) = S(\sigma \circ v)$ для всех $\sigma \in S_n$, то естественно рассматривать лишь многочлены $S(v)$, отвечающие монотонным одночленам v . По смыслу ясно также, что любой симметрический многочлен f над A является линейной комбинацией с коэффициентами из A многочленов типа $S(v)$:

$$f = \sum a_v S(v).$$

Обычно такая запись получается моментально ("на глазок"). Таким образом, задача сводится к выражению $S(v)$ через элементарные симметрические многочлены.

С каждым монотонным одночленом $v = X_1^{i_1} X_2^{i_2} \dots X_n^{i_n}$ ассоциируется симметрический многочлен

$$g_v = g_v(X_1, \dots, X_n) = s_1^{i_1-i_2} s_2^{i_2-i_3} \dots s_n^{i_n}, \quad (6)$$

высшим членом которого как раз и является v . В соответствии со схемой доказательства теоремы 1 вырисовывается следующий метод выражения $S(v)$ через элементарные симметрические многочлены.

Пусть $\deg v = m$. Берутся все "монотонные" разбиения

$$m = j_1 + j_2 + \dots + j_n, \quad j_1 \geq j_2 \geq \dots \geq j_n \geq 0,$$

целого числа m такие, что $w = X_1^{j_1} X_2^{j_2} \dots X_n^{j_n} < v$. Рассматривается множество M_v всех таких одночленов w . Для каждого $w \in M_v$ составляется одночлен g_w (см. (6)). Мы уже знаем, что

$$S = g_v + \sum_{w \in M_v} n_w g_w, \quad (7)$$

где n_w — какие-то целые числа. Неопределённые коэффициенты n_w (отсюда и название *метод неопределённых коэффициентов*) находятся путём последовательных подстановок в (7) вместо X_1, \dots, X_n каких-нибудь целых чисел, обычно нулей и единиц. Значения g_v, g_w и $S(v)$ при этом известны, и для n_w получается заведомо совместная система линейных уравнений.

Пример 1. $v = X_1^3, S(v) = p_3(X_1, \dots, X_n), n \geq 3, g_v = s_1^3$.

$$\begin{array}{c|cc} M_v & X_1^2 X_2 & X_1 X_2 X_3 \\ \hline g_w & s_1 s_2 & s_3 \end{array}$$

Уравнение (7) в данном случае имеет вид

$$p_3 = s_1^3 + a s_1 s_2 + b s_3.$$

Если $X_1 = X_2 = 1, X_i = 0$ при $i > 2$, то $p_3 = 2, s_1 = 2, s_2 = 1, s_3 = 0$. Если же $X_1 = X_2 = X_3 = 1, X_i = 0$ при $i > 3$, то $p_3 = 3, s_1 = 3, s_2 = 3, s_3 = 1$. Из получившейся системы

$$\begin{aligned} 2 &= 2^3 + a \cdot 2 \cdot 1 + b \cdot 0, \\ 3 &= 3^3 + a \cdot 3 \cdot 3 + b \cdot 1 \end{aligned}$$

находим $a = -3, b = 3$, т.е. $p_3 = s_1^3 - 3s_1 s_2 + 3s_3$.

Для выражения степенных сумм $p_k(X_1, \dots, X_n)$ в виде многочленов от s_1, s_2, \dots, s_n имеются рекуррентные формулы, называемые *формулами Ньютона*:

$$p_k = p_{k-1}s_1 + p_{k-2} + \dots + (-1)^{k-1} p_1 s_{k-1} + (-1)^k k s_k = 0 \quad (8)$$

при $1 \leq k \leq n$;

$$p_k = p_{k-1}s_1 + p_{k-2} + \dots + (-1)^{n-1} p_{k-n+1} s_{n-1} + (-1)^n p_{k-n} s_n = 0 \quad (9)$$

при $k > n$.

Чтобы их доказать, воспользуемся очевидными соотношениями

$$X_i^n - s_1 X_i^{n-1} + \dots + (-1)^{n-1} s_{n-1} X_i + (-1)^n s_n = 0,$$

получающимися при подстановке $Y = X_i$ в (3). Умножая каждое из этих соотношений на X_i^{k-n} ($k \geq n$):

$$X_i^k - s_1 X_i^{k-1} + \dots + (-1)^{n-1} s_{n-1} X_i^{k-n+1} + (-1)^n s_n X_i^{k-n} = 0,$$

и производя затем суммирование по i от 1 до n , мы получим не только формулу (9), но и формулу (8) при $k = n$ ($p_0 = X_1^0 + \dots + X_n^0 = n$). Рассмотрим, далее, симметрический однородный многочлен $f_{k,n}$ степени $k \leq n$ (или $-\infty$, если $f_{k,n} = 0$):

$$f_{k,n}(X_1, \dots, X_n) = p_k - p_{k-1}s_1 + \dots + (-1)^{k-1}p_1s_{k-1} + (-1)^kks_k.$$

Используя индукцию по $r = n - k$, докажем, что $f_{k,n}$ тождественно равен нулю. Для $r = 0$ этот факт был только что установлен. Полагая $X_n = 0$ и замечая, что получающиеся при этом симметрические многочлены $(s_i)_0$, $(p_i)_0$ совпадают с многочленами s_i и p_i , определёнными для $n - 1$ переменных X_1, \dots, X_{n-1} (см. (3) и (5)), мы приходим к равенству

$$\begin{aligned} f_{k,n}(X_1, \dots, X_{n-1}, 0) &= \\ &= (p_k)_0 - (p_{k-1})_0(s_1)_0 + \dots + (-1)^{k-1}(p_1)_0(s_{k-1})_0 + (-1)^k k(s_k)_0 = \\ &= f_{k,n-1}(X_1, \dots, X_{n-1}) = 0, \end{aligned}$$

ибо $n - 1 - k = r - 1 < r$, и применимо предположение индукции.

Соотношение $f_{k,n}(X_1, \dots, X_{n-1}, 0) = 0$ показывает, что многочлен $f_{k,n}$ делится на X_n : $f_{k,n} = X_n f_1$. Используя симметричность $f_{k,n}$, приходим к выводу, что этот многочлен содержит в качестве множителей X_1, X_2, \dots, X_n , а значит, и их произведение $s_n = X_1 X_2 \dots X_n$. Другими словами,

$$f_{k,n}(X_1, \dots, X_n) = s_n(X_1, \dots, X_n) \cdot h(X_1, \dots, X_n). \quad (10)$$

Разложение (10) возможно, однако, лишь при $h = 0$, поскольку $\deg s_n = n$, а $\deg f_{k,n} = k < n$. Итак, $f_{k,n} = 0$, и доказательство формулы (8) завершено. \square

4. Дискриминант многочлена. Рассмотрим в кольце $P[X_1, \dots, X_n]$ многочлен

$$\Delta_n = \prod_{1 \leq j < i \leq n} (X_i - X_j),$$

который, очевидно, можно представить в виде определителя Вандермонда

$$\Delta_n = \begin{vmatrix} 1 & 1 & \dots & 1 \\ X_1 & X_2 & \dots & X_n \\ \dots & \dots & \dots & \dots \\ X_1^{n-1} & X_2^{n-1} & \dots & X_n^{n-1} \end{vmatrix}. \quad (11)$$

Так как определитель является кососимметрической функцией своих столбцов, то $\pi \circ \Delta_n = \varepsilon_\pi \Delta_n$ — знак перестановки $\pi \in S_n$. Но в таком случае Δ_n^2 — симметрический многочлен, и по основной теореме

его можно выразить в виде многочлена от элементарных симметрических функций

$$\Delta_n^2 = \prod (X_i - X_j)^2 = \text{Dis}(s_1, \dots, s_n).$$

Многочлен Dis от $s_1(X_1, \dots, X_n), \dots, s_n(X_1, \dots, X_n)$ называется *дискриминантом семейства* X_1, \dots, X_n . Его коэффициенты, очевидно, лежат в \mathbb{Z} . При подстановке $x_i \in F$ вместо X_i , $i = 1, 2, \dots, n$ (F — какое-то расширение поля P), можно говорить о *дискриминанте семейства* любых n элементов поля F . Если не все $x_1, \dots, x_n \in F$ различны, то дискриминант этого семейства обращается в нуль, поскольку хотя бы один из множителей $x_i - x_j$ будет равен нулю. Способностью Dis выделять этот случай и объясняется сам термин *дискриминант*.

Удобный способ получения дискриминанта основан на интерпретации Δ_n^2 как произведения определителя (11) на транспонированный определитель: $\Delta_n^2 = \Delta_n \cdot {}^t\Delta_n$ (вспомним, что $\det {}^t A = \det A$ для любой квадратной матрицы A). Действуя по правилу умножения матриц, мы сразу же находим

$$\text{Dis}(s_1, \dots, s_n) = \begin{vmatrix} n & p_1 & p_2 & \cdots & p_{n-1} \\ p_1 & p_2 & p_3 & \cdots & p_n \\ p_2 & p_3 & p_4 & \cdots & p_{n+1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ p_{n-1} & p_n & p_{n+1} & \cdots & p_{2n-2} \end{vmatrix}, \quad (12)$$

где p_k — известные нам степенные суммы (5). Вычислив p_k по рекуррентным формулам (8) и (9), мы придём к явному выражению для $\text{Dis}(s_1, \dots, s_n)$. В частности, $p_1 = s_1$, $p_2 = s_1^2 - 2s_2$, так что

$$\text{Dis}(s_1, s_2) = \begin{vmatrix} 2 & s_1 \\ s_1 & s_1^2 - 2s_2 \end{vmatrix} = s_1^2 - 4s_2. \quad (13)$$

Пусть нам дан теперь нормализованный многочлен

$$f(X) = X^n + a_1 X^{n-1} + \dots + a_{n-1} X + a_n \in P[X],$$

имеющий в P или в некотором его расширении F n корней c_1, \dots, c_n . Как мы знаем из формул Виета, $a_k = (-1)^k s_k(c_1, \dots, c_n)$.

Определение. Дискриминант семейства корней c_1, \dots, c_n многочлена f , или, что равносильно, значение дискриминанта $\text{Dis}(s_1, \dots, s_n)$, получающееся при подстановке $(-1)^k a_k$ вместо s_k , называется *дискриминантом многочлена* f и обозначается $D(f)$. Он называется также *дискриминантом алгебраического уравнения*

$$f(x) = x^n + a_1 x^{n-1} + \dots + a_{n-1} x + a_n = 0. \quad (14)$$

Ясно, что $D(f) \in P$ (вспомним в этой связи следствие теоремы 1).

Как видно из определения дискриминанта, справедливо также

Предложение. $D(f) = 0$ тогда и только тогда, когда уравнение (14) имеет кратные корни (хотя бы один корень кратности $k > 1$).

С учётом следствия 2 теоремы 5 из § 1 мы имеем теперь два способа, не требующих выхода за пределы основного поля P , решить, обладает ли или нет многочлен $f \in P[X]$ кратными корнями. Но значение дискриминанта заключается не только в этом. Скажем, формула (13), применённая к квадратному трёхчлену $f(X) = X^2 + aX + b$ с вещественными коэффициентами a, b , даёт $D(f) = a^2 - 4b$ — выражение, известное из элементарной алгебры. В частности, от знака $D(f)$ зависит вещественность или комплексная сопряжённость корней уравнения $x^2 + ax + b = 0$.

Пример 2. Вычислим дискриминант так называемого неполного кубического уравнения

$$f(x) = x^3 + ax + b = 0. \quad (15)$$

В данном случае $s_4 = 0$, и вычисление p_k по рекуррентным формулам даёт $p_1 = s_1 = 0$, $p_2 = s_1^2 - 2s_2 = -2a$, $p_3 = s_1^3 - 3s_1s_2 + 3s_3 = -3b$, $p_4 = s_1^4 - 4s_1^2s_2 + 4s_1s_3 + 2s_2^2 = 2a^2$. Следовательно, по формуле (12) имеем

$$D(f) = \begin{vmatrix} 3 & 0 & -2a \\ 0 & -2a & -3b \\ -2a & -3b & 2a^2 \end{vmatrix} = -4a^3 - 27b^2. \quad (16)$$

Выражение $D(f)$ приобретает более сложный вид (по сравнению с (16)) в случае полного кубического уравнения $x^3 + a_1x^2 + a_2x + a_3 = 0$, однако от его рассмотрения можно избавиться, как показывает следующее общее рассуждение.

Перейдём от аргумента x к $y = x + a_1/n$. Подставляя $x = y - a_1/n$ в уравнение (14) и используя биномиальную формулу, находим

$$g(y) = f\left(y - \frac{a_1}{n}\right) = y^n + ay^{n-2} + \dots = 0, \quad (17)$$

т.е. в новом уравнении коэффициент при y^{n-1} равен нулю. Зная корень y_0 уравнения (17), мы легко найдём также и корень $x_0 = y_0 - a_1/n$ исходного уравнения (14). Поэтому без ограничения общности можно считать $a_1 = 0$.

Если пытаться найти общую формулу для решения уравнения (15) (в чём преуспели средневековые математики Сципион дель Ферро, Кардано и др.), то неизбежно в игру будет вводиться дискриминант (16) (см. формулы (2) из § 2 гл. 1).

5. Результант. Основное свойство $D(f)$, сформулированное в предложении из предыдущего пункта, интерпретируется также как признак наличия общих корней (или общих множителей) у многочленов f и его производной f' . В основе этого признака лежит в конечном счёте алгоритм Евклида. Это даёт основание полагать, что имеется аналогичный критерий, позволяющий непосредственно по коэффициентам любых двух многочленов $f, g \in P[X]$ решить вопрос о том, обладают они общим множителем или не обладают.

Итак, пусть

$$\begin{aligned}f(X) &= a_0 X^n + a_1 X^{n-1} + \dots + a_{n-1} X + a_n, \\g(X) &= b_0 X^m + b_1 X^{m-1} + \dots + b_{m-1} X + b_m\end{aligned}$$

— два многочлена с коэффициентами в поле P . Здесь $n > 0$, $m > 0$, но не исключается возможность того, что $a_0 = 0$ или $b_0 = 0$.

Определение. Результантом $\text{Res}(f, g)$ многочленов f и g называется однородный многочлен (однородная полиномиальная функция) от их коэффициентов (степени m относительно a_0, \dots, a_n и степени n относительно b_0, \dots, b_m) вида

$$\text{Res}(f, g) = \left| \begin{array}{cccccc|c} a_0 & a_1 & \dots & \dots & a_n & & \\ a_0 & a_1 & \dots & \dots & a_n & & \\ \dots & \dots & \dots & \dots & \dots & & \\ & & a_0 & a_1 & \dots & a_n & \\ b_0 & b_1 & \dots & \dots & b_m & & \\ b_0 & b_1 & \dots & \dots & b_m & & \\ \dots & \dots & \dots & \dots & \dots & & \\ & & b_0 & b_1 & \dots & b_m & \end{array} \right| \quad \left. \begin{array}{l} m \\ \text{строк} \\ \hline n \\ \text{строк} \end{array} \right.$$

В этом определении результанта содержится некое утверждение о его степенях как многочлена. Но оно непосредственно вытекает из свойств определителей: если заменить в первых m строках a_i на ta_i , то $\text{Res}(tf, g) = t^m \text{Res}(f, g)$, после чего остаётся сослаться на упражнение 3 из § 2 гл. 5.

Выведем теперь основные свойства результанта.

R1. $\text{Res}(f, g) = 0$ тогда и только тогда, когда $a_0 = 0 = b_0$ или же f и g имеют общий множитель в $P[X]$ степени > 0 .

Убедимся сначала в том, что условие “ $a_0 = 0 = b_0$ или же f и g имеют общий множитель в $P[X]$ степени > 0 ” выполняется тогда и только тогда, когда найдутся многочлены f_1, g_1 , одновременно не равные нулю, для которых

$$fg_1 + f_1g = 0, \quad \deg f_1 < n_1, \quad \deg g_1 < m. \quad (18)$$

Действительно, пусть $h = \text{НОД}(f, g)$, $\deg h > 0$. Тогда $f = hf_1$, $g = -hg_1$, и, следовательно, $fg_1 + g_1f = 0$. Кроме того, $\deg f_1 < n$, $\deg g_1 < m$, так что (18) имеет место. При $a_0 = 0 = b_0$ мы можем положить $f_1 = f$, $g_1 = -g$.

Обратно: предположив при выполнении (18), что $\text{НОД}(f, g) = 1$, мы ввиду факториальности $P[X]$ (см. § 3 гл. 5) придём к импликации $fg_1 = -g_1f \Rightarrow f|f_1$, $g|g_1$. Стало быть, $\deg f < n$, $\deg g < m$, откуда $a_0 = 0 = b_0$.

Мы докажем теперь эквивалентность условий (18) и $\text{Res}(f, g) = 0$.

Положив

$$f_1 = c_0 X^{n-1} + c_1 X^{n-2} + \dots + c_{n-1},$$

$$g_1 = d_0 X^{m-1} + d_1 X^{m-2} + \dots + d_{m-1}$$

и вычислив по формальным правилам коэффициенты многочлена $f_1 g_1 + f_1 g$ степени $\leq n+m-1$, мы запишем условие (18) в виде квадратной однородной системы линейных уравнений с $n+m$ неизвестными $d_0, d_1, \dots, d_{m-1}, c_0, c_1, \dots, c_{n-1}$:

$$\begin{aligned} a_0 d_0 + & \dots + b_0 c_0 & \dots & = 0, \\ a_1 d_0 + a_0 d_1 + & \dots + b_1 c_0 + b_0 c_1 & \dots & = 0, \\ a_2 d_0 + a_1 d_1 + a_0 d_2 + & \dots + b_2 c_0 + b_1 c_1 + b_0 c_2 = 0, \end{aligned} \quad (19)$$

.....

Определитель матрицы системы (19) (точнее, определитель транспонированной матрицы) совпадает как раз с $\text{Res}(f, g)$. Стало быть, система (19) имеет ненулевое решение в точности тогда, когда $\text{Res}(f, g) = 0$, а всякое ненулевое решение приводит к паре многочленов f_1, g_1 , удовлетворяющих условию (18). \square

R2. Пусть многочлены f и g полностью расщепляются на линейные множители в $P[X]$:

$$f(X) = a_0(X - \alpha_1) \dots (X - \alpha_n),$$

$$g(X) = b_0(X - \beta_1) \dots (X - \beta_m).$$

Тогда

$$\text{Res}(f, g) = a_0^m \prod_{i=1}^n g(\alpha_i) = (-1)^{mn} b_0^n \prod_{j=1}^m f(\beta_j) = a_0^m b_0^n \prod_{i,j} (\alpha_i - \beta_j).$$

Доказательство. Ясно, что указанные здесь формулы, если они верны, должны носить универсальный характер, не зависящий от частных типов многочленов f, g . Эта несложная "философия", в природу которой мы не хотим здесь вдаваться, позволяет нам ограничиться рассмотрением "общего случая", когда, скажем, все $g(\alpha_1), \dots, g(\alpha_n)$ и все $f(\beta_1), \dots, f(\beta_m)$ попарно различны.

Далее, так как $\text{Res}(g, f) = (-1)^{mn} \text{Res}(f, g)$ (см. определение), то достаточно убедиться в справедливости соотношения $\text{Res}(f, g) = a_0^m \prod g(\alpha_i)$. С этой целью введём новую переменную Y и над полем рациональных дробей $P(Y)$ рассмотрим многочлены $f(X), g(X) - Y$. Из определения результанта, где следует заменить b_m на $b_m - Y$, получается, что

$$\text{Res}(f, g - Y) = (-1)^n a_0^m Y^n + \dots + \text{Res}(f, g)$$

— многочлен степени n относительно Y со старшим коэффициентом $(-1)^n a_0^m$ и с постоянным членом $\text{Res}(f, g)$. Многочлены $f(X)$ и $g(X) - g(\alpha_i)$ с общим корнем α_i делятся на $X - \alpha_i$. Ввиду свойства R1 имеем $\text{Res}(f, g - g(\alpha_i)) = 0$.

По теореме Безу многочлен $\text{Res}(f, g - Y)$ должен делиться на $g(\alpha_i) - Y$, $1 \leq i \leq n$. Так как все $g(\alpha_i)$ у нас различны, то $\text{Res}(f, g - Y) = a_0^{2n} \prod_{i=1}^n (g(\alpha_i) - Y)$. При $Y = 0$ получаем нужное выражение. \square

Данное в п. 4 определение дискриминанта перенесём на случай ненормализованных многочленов, полагая

$$D(f) = a_0^{2n-2} \prod_{1 \leq j < i \leq n} (\alpha_i - \alpha_j)^2 = \left[a_0^{n-1} \prod_{j < i} (\alpha_i - \alpha_j) \right]^2, \quad a_0 \neq 0.$$

R3. Имеет место формула

$$D(f) = (-1)^{n(n-1)/2} a_0^{-1} \text{Res}(f, f'). \quad (20)$$

Действительно, согласно R2

$$\text{Res}(f, f') = a_0^{n-1} \prod_{i=1}^n f'(\alpha_i).$$

Но

$$f'(\alpha_i) = a_0 \prod_{j \neq i} (\alpha_i - \alpha_j),$$

что является простым следствием подстановки $X = \alpha_i$ в общее выражение

$$f'(X) = a_0 \sum_{i=1}^n \prod_{j \neq i} (X - \alpha_j),$$

получаемое дифференцированием произведения $f(X) = a_0 \prod_{j=1}^n (X - \alpha_j)$. Таким образом,

$$\begin{aligned} \text{Res}(f, f') &= a_0^{2n-1} \prod_{i=1}^n \prod_{j \neq i} (\alpha_i - \alpha_j) = \\ &= a_0 (-1)^{n(n-1)/2} a_0^{2n-2} \prod_{j < i} (\alpha_i - \alpha_j)^2 = a_0 (-1)^{n(n-1)/2} D(f). \quad \square \end{aligned}$$

Формула (20) даёт явное выражение для дискриминанта.

УПРАЖНЕНИЯ

1. Пусть p — простое число. При помощи формул Ньютона (9), (10) показать, что

$$\sum_{i=1}^{p-1} i^m = \begin{cases} -1 \pmod{p}, & \text{если } m \text{ делится на } p-1, \\ 0 \pmod{p}, & \text{если } m \text{ не делится на } p-1. \end{cases}$$

2. Используя систему рекуррентных формул Ньютона и формулы Крамера, получить следующие явные выражения p_k через s_k и s_k через p_k :

$$p_k = \begin{vmatrix} s_1 & 1 & 0 & 0 & \dots & 0 \\ 2s_2 & s_1 & 1 & 0 & \dots & 0 \\ 3s_3 & s_2 & s_1 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ (k-1)s_{k-1} & s_{k-2} & s_{k-3} & s_{k-4} & \dots & 1 \\ ks_k & s_{k-1} & s_{k-2} & s_{k-3} & \dots & s_1 \end{vmatrix},$$

$$s_k = \frac{1}{k!} \begin{vmatrix} p_1 & 1 & 0 & 0 & \dots & 0 \\ p_2 & p_1 & 1 & 0 & \dots & 0 \\ p_3 & p_2 & p_1 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ p_{k-1} & p_{k-2} & p_{k-3} & p_{k-4} & \dots & 1 \\ p_k & p_{k-1} & p_{k-2} & p_{k-3} & \dots & p_1 \end{vmatrix}.$$

3. Пусть c_1, c_2, c_3 — комплексные корни многочлена $X^3 - X + 1$. Что можно сказать о расширении $\mathbb{Q}(c_1^{99} + c_2^{99} + c_3^{99})$?

4. Многочлен $f(X_1, \dots, X_n)$ над полем P характеристики $\neq 2$ называется **кососимметрическим** (или **знакопеременным**), если $\forall \pi \in S_n$ ($\pi \circ f)(X_1, \dots, X_n) = \epsilon_\pi f(X_1, \dots, X_n)$ (как всегда, ϵ_π — знак перестановки). Примером кососимметрического многочлена может служить $\Delta_n = \prod_{i < j} (X_i - X_j)$. Показать, что любой кососимметрический многочлен $f \in P[X_1, \dots, X_n]$ имеет вид $f = \Delta_n \cdot g$, где g — симметрический многочлен.

Указание. Рассмотреть f как многочлен относительно X_n с коэффициентами в $P[X_1, \dots, X_{n-1}]$. Обратить внимание на то, что в силу кососимметричности $f = 0$ при $X_n = X_{n-1}$ и, стало быть, f делится на $X_n - X_{n-1}$.

5. Используя свойство R2 и факт существования поля разложения многочлена (см. следующий параграф), показать, что

$$\text{Res}(fg, h) = \text{Res}(f, h) \cdot \text{Res}(g, h).$$

6. Из упр. 5 и из R3 вывести формулу

$$D(fg) = D(f)D(g)[\text{Res}(f, g)]^2.$$

7. Чему равен результант $\text{Res}(f(X), X - a)$?

8. Показать, что $D(X^n + a) = (-1)^{n(n-1)/2} n^n a^{n-1}$.

9. Пусть $f(X) = X^{n-1} + X^{n-2} + \dots + 1$. Используя соотношение $X^n - 1 = (X-1)f(X)$ и предыдущие упражнения, показать, что $D(f) = (-1)^{(n-1)(n-2)/2} n^{n-2}$.

§ 3. Алгебраическая замкнутость поля \mathbb{C}

1. **Формулировка основной теоремы.** Пусть P — поле и f — произвольный многочлен над P . Как уже отмечалось в п. 2 § 1, поведение полиномиальной функции $\tilde{f}: P \rightarrow P$, ассоциированной с f , существенно зависит от поля P . В частности, $\text{Im } \tilde{f} = P$, коль скоро $\deg f > 0$, и к P применимо следующее

Определение. Поле P называется **алгебраически замкнутым**, если каждый многочлен из кольца $P[X]$ разлагается на линейные множители.

То же самое можно выразить другими словами: поле P алгебраически замкнуто, если неприводимыми над P являются лишь многочлены степени 1 (линейные многочлены).

Если любой многочлен $f \in P[X]$ обладает в P по крайней мере одним корнем, то поле P алгебраически замкнуто. Действительно, тогда $f(X) = (X - a)h(X)$, $a \in P$, $h \in P[X]$, но по условию для многочлена h в P тоже существует хотя бы один корень, т.е. $h(X) = (X - b)r(X)$, $b \in P$, $r \in P[X]$. Продолжая этот процесс, мы придём в конце концов к полному разложению f на линейные множители. Так как f — произвольный многочлен, то поле P удовлетворяет определению алгебраической замкнутости.

Хотя и справедливо утверждение о том, что для всякого поля P существует расширение $\tilde{P} \supset P$, являющееся алгебраически замкнутым полем (теорема Штейнича), на первых порах всё же трудно воспринять не только конструкцию алгебраически замкнутого расширения, но и саму идею такого расширения. Тем более приятно, что мы фактически располагаем ярким и очень важным примером алгебраически замкнутого поля, как об этом гласит так называемая основная теорема алгебры. Именно, справедлива

Теорема 1. Поле комплексных чисел \mathbb{C} алгебраически замкнуто.

Сформулируем ещё раз это фундаментальное утверждение, теперь уже в терминах корней.

Произвольный многочлен $f(X)$ степени $n \geq 1$ с комплексными (или вещественными) коэффициентами имеет ровно n комплексных корней, считаемых со своими кратностями.

Громкий титул “основной” теорема 1 приобрела ещё в те времена, когда решение алгебраических уравнений было одним из главных занятий алгебристов. В наши дни теорема 1 относится к числу рядовых, хотя и важных утверждений.

Впервые строгое доказательство основной теоремы было предложено Гауссом в 1779 г. С тех пор появилось много вариантов доказательства, различающихся между собой, так сказать, степенью алгебраичности. Необходимость опираться на свойства непрерывности полей \mathbb{R} и \mathbb{C} (иначе, на их топологию) проявляется в той или иной форме; есть даже совсем не алгебраическое и очень короткое доказательство, основанное на довольно глубоком понятии аналитической функции комплексной переменной. Сейчас будет приведено доказательство, основанное на элементарных сведениях из математического анализа и восходящее к идеям Даламбера, Эйлера, Гаусса, Коши, Аргана. Именно последнему (Argand R., 1814 г.) принадлежит наиболее прозрачное изложение, которому с тех пор следуют почти все учебники по алгебре.

2. Доказательство основной теоремы. Его неалгебраичность начинается с двух вспомогательных утверждений, которые можно найти в любом курсе анализа.

1) *Каждый комплексный многочлен*

$$f(z) = a_0 z^n + a_1 z^{n-1} + \dots + a_{n-1} z + a_n, \quad n \geq 1, \quad (1)$$

является непрерывной функцией в любой точке плоскости \mathbb{C} (функция $f : \mathbb{C} \rightarrow \mathbb{C}$ непрерывна в точке $z_0 \in \mathbb{C}$, если $\lim_{z \rightarrow z_0} f(z) = f(z_0)$; другими словами, для любой окрестности $V(f(z_0))$ найдётся окрестность $U(z_0)$ такая, что при любом $z \in U(z_0)$ будет $f(z) \in V(f(z_0))$).

2) *Каждая непрерывная функция $f : K \rightarrow \mathbb{R}$ на компакте $K \subset \mathbb{R}^2$ достигает своего минимума в K (компакт — замкнутое ограниченное множество).*

Заметим, что следовало бы говорить о непрерывности полиномиальной функции $\tilde{f} : \mathbb{C} \rightarrow \mathbb{C}$, но мы следуем упрощённому языку, принятому в анализе. Компактом у нас будет круг $|z| \leq r$ некоторого достаточно большого радиуса r , определённого ниже. Тривиальный случай многочлена f со свободным членом $a_n = 0$ исключается из рассмотрения, поскольку тогда f имеет корень $z_0 = 0$.

Чтобы пояснить геометрически идею доказательства, вообразим себе поверхность, в \mathbb{R}^3 , отвечающую уравнению $w = |f(z)|$: значения z изображаются на горизонтальной плоскости \mathbb{R}^2 , а значения $|f(z)|$ откладываются вверх, в направлении оси w , перпендикулярной к \mathbb{R}^2 . Из непрерывности $f(z)$ следует непрерывность функции $|f(z)|$ на всей плоскости \mathbb{C} . Нужно убедиться в том, что хотя бы одной точкой наша поверхность “опирается” на горизонтальную плоскость \mathbb{R}^2 ($w = 0$). Последующие рассуждения разобьём на несколько шагов.

Лемма 1. *Существует положительное число $r \in \mathbb{R}$ такое, что $|f(z)| > |f(0)|$ для всех $z \in \mathbb{C}$ с $|z| > r$.*

Действительно, для $z \neq 0$ имеем $|f(z)| = |z|^n |a_0 + g(z^{-1})|$, где $g(u) = a_1 u + a_2 u^2 + \dots + a_n u^n \in \mathbb{C}[u]$. Из непрерывности g в точке 0 следует существование такого вещественного $\delta > 0$, что $|g(u)| \leq |a_0|/2$ при $|u| < \delta$. Таким образом,

$$|f(z)| \geq |z|^n (|a_0| - |g(z^{-1})|) \geq \frac{1}{2} |a_0| |z|^n$$

при $|z| > \delta^{-1}$. Следовательно, осталось выбрать любое вещественное число $r > \delta^{-1}$, для которого было бы выполнено неравенство $|a_0|r^n > 2|a_n|$. \square

Следствие (лемма Коши о минимуме). *Для каждого многочлена $f \in \mathbb{C}[z]$ существует $z_0 \in \mathbb{C}$ такое, что $|f(z_0)| = \inf_{z \in \mathbb{C}} |f(z)|$.*

В самом деле, ввиду утверждения 2) непрерывная функция $|f(z)|$ принимает в круге $D_r = \{z \in \mathbb{C} \mid |z| \leq r\}$ минимальное значение, т.е. существует $z_0 \in D_r$ такое, что $|f(z_0)| = \inf_{z \in D_r} |f(z)|$. Но так

как $|f(z_0)| \leq |f(0)|$, а по лемме 1 имеет место неравенство $|f(0)| \leq \inf_{z \in \mathbb{C} \setminus D_r} |f(z)|$, то $|f(z_0)| = \inf_{z \in \mathbb{C}} |f(z)|$. \square

Лемма 2. Пусть k — любое целое число ≥ 1 , и пусть $h \in \mathbb{C}[z]$ — многочлен с $h(0) \neq 0$.

Тогда для каждого $a \in \mathbb{C}^*$ найдётся такое $b \in \mathbb{C}$, что

$$|a + b^k h(b)| < |a|.$$

Доказательство леммы исходит из факта непрерывности многочлена h : существует $\delta > 0$ такое, что при $|z| < \delta$ имеет место неравенство $|h(z) - h(0)| < |h(0)|/2$. Это позволяет нам получить оценку для $a + z^k h(z) = a + h(0)z^k + z^k(h(z) - h(0))$:

$$|a + z^k h(z)| \leq |a + h(0)z^k| + \frac{1}{2}|h(0)| \cdot |z|^k \quad (*)$$

из круга $|z| < \delta$.

Выберем теперь комплексное число $b \in \mathbb{C}$, для которого

$$h(0)b^k = -ta, \quad 0 < t < 1$$

(ниже на вещественное число t будут наложены дополнительные ограничения). В качестве b достаточно взять, следуя теореме 3 из § 1 гл. 5, любой корень степени k из $-tah(0)^{-1} \neq 0$. Получаем $|a + h(0)b^k| = (1-t)|a|$ и $|h(0)| \cdot |b|^k / 2 = t|a|/2$, что в соединении с $(*)$ приведёт к нужному неравенству, коль скоро $|b| < \delta$. Мы обеспечим выполнение этого условия, наложив на $t = -h(0)a^{-1}b^k$ ограничение $t < |h(0)a^{-1}|b^k$. Итак, подставив в $(*)$ значение $z = b$, $|b| < \delta$, получаем окончательно

$$|a + b^k h(b)| \leq (1-t)|a| + \frac{1}{2}t|a| = \left(1 - \frac{1}{2}t\right)|a| < |a|. \quad \square$$

Следствие (лемма Даламбера—Аргана). Пусть $f(z)$ — многочлен положительной степени над \mathbb{C} .

Тогда каждой точке $c \in \mathbb{C}$ такой, что $f(c) \neq 0$, отвечает точка $c' \in \mathbb{C}$, для которой

$$|f(c')| < |f(c)|.$$

Для доказательства многочленов $f(z+c)$, подобно $f(z)$ не являющийся константой, разложим по степеням z :

$$f(z+c) = f(c) + b_k z^k + b_{k+1} z^{k+1} + \dots + b_n z^n, \quad b_k \neq 0.$$

Другими словами,

$$f(z+c) = f(c) + z^k h(z),$$

где

$$h(z) = b_k + b_{k+1}z + \dots + b_n z^{n-k}, \quad h(0) \neq 0.$$

Подставив в формулировку леммы 2 значение $a = f(c) \neq 0$, мы можем утверждать существование такого $b \in \mathbb{C}$, что при $c' = b + c$ будет выполнено требуемое неравенство

$$|f(c')| = |f(b+c)| = |f(c) + b^k h(b)| < f(c). \quad \square$$

Геометрический смысл: если на поверхности $w = f(z)$ взята точка, расположенная строго выше плоскости $w = 0$, то обязательно найдётся другая точка на поверхности с более низким расположением.

Окончание доказательства основной теоремы (теоремы 1). Согласно следствию леммы 1 существует такая точка $z_0 \in \mathbb{C}$, что $|f(z_0)| \leq |f(z)|$ для всех $z \in \mathbb{C}$. Если $f(z_0) \neq 0$, то, как утверждает следствие леммы 2, найдётся такая точка $z'_0 \in \mathbb{C}$, что $|f(z'_0)| < |f(z_0)|$ — противоречие. \square

Воздерживаясь пока от каких-либо комментариев по поводу приведённого доказательства, заметим, что явным аналогом леммы 1 служит очевидно,

Лемма 3 (лемма о модуле старшего члена). *Пусть $f(z)$ — многочлен вида (1) с произвольными комплексными коэффициентами a_0, a_1, \dots, a_n , $n \geq 1$. Положим $A = \max(|a_1|, \dots, |a_n|)$, $r = \frac{A}{|a_0|} + 1$.*

Тогда при $|z| > r$ будет выполнено неравенство

$$|a_0 z^n| > |a_1 z^{n-1} + \dots + a_{n-1} z + a_n|.$$

Доказательство. Если взять $|z| > r$, то получим $|a_0| > \frac{A}{|z| - 1}$, откуда согласно правилам действий с модулями комплексных чисел (см. § 1 гл. 5) будем иметь

$$\begin{aligned} |a_0 z^n| &= |a_0| |z|^n > \frac{A |z|^n}{|z| - 1} > \frac{A (|z|^n - 1)}{|z| - 1} = \\ &= A (|z|^{n-1} + \dots + |z| + 1) \geq |a_1| |z|^{n-1} + \dots + |a_{n-1}| |z| + |a_n| = \\ &= |a_1 z^{n-1}| + \dots + |a_{n-1} z| + |a_n| \geq |a_1 z^{n-1}| + \dots + |a_{n-1} z| + |a_n|. \quad \square \end{aligned}$$

Следствие 1. *Пусть многочлен (1) степени $n \geq 1$ имеет вещественные коэффициенты.*

Тогда для всех $x \in \mathbb{R}$, достаточно больших по абсолютной величине, знак (вещественного числа) $f(x)$ совпадает со знаком старшего члена $a_0 x^n$.

Следствие 2. *Многочлен нечётной степени с вещественными коэффициентами имеет хотя бы один вещественный корень.*

Доказательство. Ввиду нечётности n старший член $a_0 x^n$ полиномиального отображения $\tilde{f} : \mathbb{R} \rightarrow \mathbb{R}$ будет принимать при положительных и отрицательных $x \in \mathbb{R}$ разные знаки. Взяв эти

значения x достаточно большими по абсолютной величине, мы согласно следствию 1 можем утверждать, что и $f(x)$ будет иметь разные знаки. Если, например, $a_0 > 0$, то $f(-r) < 0$, а $f(r) > 0$, где r — вещественное число, взятое из леммы 3. По теореме Больцано-Коши о промежуточном значении функция f , непрерывная на отрезке $[-r, r]$ и принимающая на его концах значения разных знаков, должна обращаться в нуль в некоторой точке рассматриваемого отрезка: $\exists c \in [-r, r], f(c) = 0$ (на самом деле $f(x)$ принимает любое промежуточное значение между $f(-r)$ и $f(r)$). То же рассуждение годится и для $a_0 < 0$. \square

3. Ещё одно доказательство основной теоремы. От геометрически наглядного доказательства теоремы 1, приведённого в п. 2, остаётся чувство неудовлетворённости не только у читателя — поклонника алгебры; это чувство было присуще и математикам прошлого века. Недаром Гаусс неоднократно возвращался к основной теореме и дал для неё целых четыре доказательства. Естественно попытаться свести к минимуму атрибутику математического анализа и максимально алгебраизировать все рассуждения. Такое “алгебраическое” доказательство, восходящее к Эйлеру, Лагранжу, Гауссу и Лапласу, приобрело со временем каноническую форму, согласующуюся с общей теорией Галуа. Ни в коей мере не касаясь последней, мы хотели бы только дать почувствовать аромат известной нам техники. Всё доказательство распадается на две части.

1) Для всякого многочлена $f \in P[z]$ степени $n \geq 1$ существует хотя бы одно поле разложения — такое минимальное расширение F поля P , в котором содержатся все корни многочлена f . Можно записать $F = P(u_1, \dots, u_n)$ и $f(z) = a_0(z - u_1)(z - u_2) \dots (z - u_n)$.

Для удобства считаем далее f нормализованным ($a_0 = 1$). Существование и единственность, с точностью до изоморфизма, поля разложения $F \supset P$ для каждого многочлена $f \in P[z]$ — это следствие общеалгебраической конструкции, на которой мы остановимся в [ВА III]. Эта конструкция, напоминающая построение кольца классов вычетов \mathbb{Z}_m в п. 2 § 3 гл. 4, никак не связана со спецификой основного поля P . Утверждение о единственности нам вообще не понадобится. В качестве примера отметим, что полем разложения многочлена $z^2 + 1$ над \mathbb{R} является поле \mathbb{C} .

2) Если часть 1) мы фактически приняли на веру, чтобы не отягощать изложение, то часть 2), являющуюся замечательной иллюстрацией уже усвоенных нами общих принципов (принципа математической индукции и принципа перехода к симметрическим функциям), мы приведём со всеми деталями.

В соответствии с замечанием, сделанным непосредственно после определения алгебраически замкнутого поля, необходимо установить существование хотя бы одного комплексного корня у многочлена

(1). Предположим сначала, что все его коэффициенты вещественные, причём без ограничения общности будем считать $a_0 = 1$, $a_n \neq 0$. Пусть

$$\deg f = 2^m n_0,$$

где n_0 — нечётное целое число. Если $m = 0$, то по лемме 2 многочлен f имеет корень, даже вещественный. Применяя индукцию по m , будем считать теорему доказанной для всех многочленов с вещественными коэффициентами, степень которых имеет вид $2^{m'} n'_0$ с $m' \leq m - 1$ (на нечётный множитель n'_0 никаких ограничений не накладывается). Заметим, что по следствию 2 леммы 3 основание индукции, отвечающее значению $m = 0$, у нас имеется (единственный фрагмент неалгебраической природы).

Рассмотрим поле разложения F многочлена $(z^2 + 1)f(z)$, существующее в силу 1) и содержащее \mathbb{C} в качестве подполя. Пусть u_1, u_2, \dots, u_n — корни многочлена f в F . Рассмотрим в F элементы

$$v_{ij} = u_i u_j + a(u_i + u_j), \quad 1 \leq i < j \leq n, \quad (2)$$

где a — какое-то фиксированное вещественное число. Следовало бы писать $v_{ij}(a)$, но мы этого делать не будем, чтобы не усложнять обозначения. Число n' элементов вида (2) равно

$$n' = \binom{n}{2} = \frac{n(n-1)}{2} = \frac{2^m n_0 (2^m n_0 - 1)}{2} = 2^{m-1} n'_0, \quad (3)$$

где n'_0 — нечётное целое число.

Многочлен

$$f_a(z) = \prod_{1 \leq i < j \leq n} (z - v_{ij}) = z^{n'} + b_1 z^{n'-1} + \dots + b_{n'}$$

из кольца $F[z]$ имеет степень n' , а его корнями по определению являются все элементы (2). В соответствии с формулами Виета (12) из § 1 коэффициентами $b_1, \dots, b_{n'}$ многочлена $f_a(z)$ будут с точностью до знака элементарные симметрические функции s_k от v_{ij} . Подставив в $s_k(v_{12}, v_{13}, \dots, v_{n-1,n})$ выражения элементов v_{ij} через u_1, \dots, u_n , мы получим функцию

$$h_k(u_1, \dots, u_n) = s_k(\dots, u_i u_j + a(u_i + u_j), \dots), \\ k = 1, \dots, n',$$

которая тоже является симметрической. В самом деле, для любой перестановки $\pi \in S_n$ (S_n — симметрическая группа степени n) имеем

$$\hat{\pi} \circ v_{ij} = u_{\pi(i)} u_{\pi(j)} + a(u_{\pi(i)} + u_{\pi(j)}) = v_{\pi(i), \pi(j)}$$

(или $v_{\pi(j), \pi(i)}$, если $\pi(i) > \pi(j)$), так что π индуцирует перестановку $\hat{\pi}$ на множестве элементов вида (2). В силу симметричности

$s_k(v_{12}, v_{13}, \dots, v_{n-1,n})$ не меняется при перестановке аргументов, поэтому

$$\begin{aligned} (\pi \circ h_k)(u_1, \dots, u_n) &= s_k(\hat{\pi} \circ v_{12}, \hat{\pi} \circ v_{13}, \dots, \hat{\pi} \circ v_{n-1,n}) = \\ &= s_k(v_{12}, v_{13}, \dots, v_{n-1,n}) = h_k(u_1, \dots, u_n). \end{aligned}$$

Заметим, что $h_k(u_1, \dots, u_n)$ есть значение при $X_i = u_i$, $i = 1, \dots, n$, симметрического многочлена $h_k(X_1, \dots, X_n)$ с вещественными коэффициентами, зависящими только от $a \in \mathbb{R}$.

По основной теореме о симметрических многочленах (теорема 1 из § 2) найдётся многочлен $g_k(Y_1, \dots, Y_n)$ с вещественными коэффициентами такой, что $h_k(X_1, \dots, X_n) = g_k(s_1(X_1, \dots, X_n), \dots, s_n(X_1, \dots, X_n))$. Стало быть,

$$\begin{aligned} (-1)^k b_k &= h_k(u_1, \dots, u_n) = g_k(s_1(u_1, \dots, u_n), \dots, s_n(u_1, \dots, u_n)) = \\ &= g_k(-a_1, \dots, (-1)^n a_n) \in \mathbb{R} \end{aligned}$$

(напомним, что a_i — коэффициенты рассматриваемого унитарного многочлена $f \in \mathbb{R}[z]$).

Итак, коэффициенты b_k многочлена $f_a(z)$ оказались вещественными при любом $a \in \mathbb{R}$. Так как $\deg f_a = n' = 2^{m-1} n'_0$ (см. (3)), то по предположению индукции f_a имеет хотя бы один комплексный корень, который, конечно, должен совпадать с одним из v_{ij} . Меняя находящийся в нашем распоряжении параметр $a \in \mathbb{R}$, мы будем получать другие многочлены $f_a(z)$ с вещественными коэффициентами. Каждому из них соответствует пара индексов $i < j$ (зависящая от a) такая, что элемент $v_{ij} = u_i u_j + a(u_i + u_j) \in F$ содержится в подполе \mathbb{C} поля F . Так как различных пар индексов $i < j$ всего $\binom{n}{2}$, а вещественных чисел $a \in \mathbb{R}$ бесконечно много, то найдутся два различных вещественных числа a, a' с одной и той же отвечающей им парой индексов, скажем, $i = 1, j = 2$ (этот вопрос нумерации корней u_1, \dots, u_n), для которых

$$\begin{aligned} u_1 u_2 + a(u_1 + u_2) &= c, \\ u_1 u_2 + a'(u_1 + u_2) &= c', \quad a \neq a', \end{aligned}$$

будут комплексными числами. Из этой системы уравнений следует, что и

$$u_1 + u_2 = \frac{c - c'}{a - a'}, \quad u_1 u_2 = c - a \frac{c - c'}{a - a'}$$

принадлежат полю \mathbb{C} . Коль скоро это так, элементы u_1, u_2 будут корнями квадратного многочлена

$$(z - u_1)(z - u_2) = z^2 - (u_1 + u_2)z + u_1 u_2$$

с комплексными коэффициентами. По известным формулам

$$u_1, u_2 = \frac{u_1 + u_2}{2} \pm \sqrt{\left(\frac{u_1 + u_2}{2}\right)^2 - u_1 u_2},$$

так что u_1, u_2 тоже оказываются комплексными числами. Таким образом, для рассматриваемого многочлена $f(z)$ с вещественными коэффициентами найдены даже два комплексных корня.

Пусть теперь $f(z)$ — многочлен вида (1) с произвольными комплексными коэффициентами (можно считать $a_0 = 1$, но это неважно). Заменив все a_i комплексно сопряжёнными числами, мы получим многочлен

$$\bar{f}(z) = \bar{a}_0 z^n + \bar{a}_1 z^{n-1} + \dots + \bar{a}_{n-1} z + \bar{a}_n.$$

Введём многочлен

$$e(z) = f(z)\bar{f}(z) = e_0 z^{2n} + e_1 z^{2n-1} + \dots + e_{2n}$$

степени $2n$ с коэффициентами

$$e_k = \sum_{i+j=k} a_i \bar{a}_j, \quad k = 0, 1, \dots, 2n.$$

Так как операция сопряжения $z \mapsto \bar{z}$ является автоморфизмом порядка 2 поля \mathbb{C} (теорема 1 из § 1 гл. 5), то $\bar{e}_k = \sum_{i+j=k} \bar{a}_i a_j = e_k$, а это означает, что $e_k \in \mathbb{R}$. По доказанному многочлен $e(z)$ с вещественными коэффициентами имеет хотя бы один комплексный корень с:

$$f(c) \cdot \bar{f}(c) = e(c) = 0.$$

Отсюда вытекает, что либо $f(c) = 0$, и теорема доказана, либо $\bar{f}(c) = 0$, т.е. $\bar{a}_0 c^n + \bar{a}_1 c^{n-1} + \dots + \bar{a}_{n-1} c + \bar{a}_n = 0$. Применяя к обеим частям этого равенства автоморфизм комплексного сопряжения, получим $a_0 \bar{c}^n + a_1 \bar{c}^{n-1} + \dots + a_{n-1} \bar{c} + a_n = 0$, т.е. $f(\bar{c}) = 0$. \square

Алгебраической замкнутостью поля \mathbb{C} (а также фактом существования поля разложения многочлена) удобно пользоваться при решении различных задач.

Пример. Пусть $S_0(f)$ — множество всех различных корней многочлена $f \in \mathbb{C}[X]$, а $S_1(f)$ — множество всех его “единиц”: $d \in S_1(f) \iff f(d) = 1$. Пусть теперь f, g — какие-то многочлены из $\mathbb{C}[X]$. Требуется показать, что

$$S_0(f) = S_0(g), \quad S_1(f) = S_1(g) \implies f(X) = g(X).$$

Так как, очевидно, $S_0(f) \cap S_1(f) = \emptyset$, то согласно результатам § 1 достаточно показать, что $|S_0(f) \cup S_1(f)| \geq n + 1$, где $n = \deg f$. По теореме 1

$$f(X) = a_0 \prod_{j=1}^{\nu} (X - c_j)^{s_j}, \quad f(X) - 1 = a_0 \prod_{j=1}^{\mu} (X - d_j)^{t_j}, \quad c_i, d_j \in \mathbb{C},$$

где

$$\sum s_i = n = \sum t_j, \quad \nu + \mu = |S_0(f) \cup S_1(f)|.$$

В соответствии с теоремой 5 § 1 имеем

$$f(X)' = (f(X) - 1)' = \prod_{i=1}^{\nu} (X - c_i)^{s_i-1} \cdot \prod_{j=1}^{\mu} (X - d_j)^{t_j-1} \cdot h(X),$$

так что $(\nu - \nu) + (\mu - \mu) = \sum(s_i - 1) + \sum(t_j - 1) \leq \deg f(X)' = n - 1$. Стало быть,
 $\nu + \mu \geq n + 1$.

Время от времени появляются новые доказательства основной теоремы алгебры как иллюстрации передовых идей математики. Обратим внимание на топологические доказательства, использующие понятия гомотопии, степени отображения, порядка кривой, критической точки и пр. С ними можно познакомиться по элементарным вводным курсам:

1. Стилрод Н., Чинк У. Первые понятия топологии. — М.: Мир, 1967.
2. Милнор Дж., Уоллес А. Дифференциальная топология. — М.: Мир, 1972.
3. Торп Дж. Начальные главы дифференциальной геометрии. — М.: Мир, 1982.

§ 4. Многочлены с вещественными коэффициентами

1. **Разложение на неприводимые множители в $\mathbb{R}[X]$.** Из теоремы 1 § 3 следует, что каждый многочлен f степени n в $\mathbb{C}[X]$ может быть записан, и притом единственным образом (с точностью до перестановки множителей), в виде

$$f(X) = a(X - c_1)(X - c_2) \dots (X - c_n),$$

где $a \neq 0$, c_1, \dots, c_n — комплексные числа. Пусть теперь $f(X) = X^n + a_1 X^{n-1} + \dots + a_{n-1} X + a_n$ — нормализованный многочлен с вещественными коэффициентами a_1, \dots, a_n и c — какой-то его комплексный корень: $c = u + iv$, $v \neq 0$. Применяя к соотношению $f(c) = 0$ автоморфизм комплексного сопряжения, как мы это делали во втором доказательстве теоремы 1 из § 3, получим, что и $f(\bar{c}) = 0$, поскольку $\bar{a}_i = a_i$. Стало быть, $f(X)$ делится на многочлен второй степени

$g(X) = (X - c)(X - \bar{c}) = X^2 - (c + \bar{c})X + c\bar{c} = X^2 - 2uX + (u^2 + v^2)$ с отрицательным дискриминантом $D(g) = 4u^2 - 4(u^2 + v^2) = -4v^2 < 0$. Условие $D(g) < 0$ необходимо и достаточно для неприводимости над \mathbb{R} квадратного многочлена $g \in \mathbb{R}[X]$.

Если, далее, k — кратность корня c многочлена $f(X)$ и $l \leq k$ — кратность корня \bar{c} , то $f(X)$ делится на l -ю степень многочлена $g(X)$:

$$f(X) = g(X)^l q(X).$$

Частное $q(X)$ двух многочленов из $\mathbb{R}[X]$ будет тоже многочленом из $\mathbb{R}[X]$, причём при $k > l$ элемент $c \in \mathbb{C}$ будет его корнем кратности $k - l$, в то время как \bar{c} корнем не является. Мы видели, однако, что это не так. Значит, $k = l$ (предположение $l \geq k$ рассматривается аналогично), т.е. комплексные корни всякого многочлена из $\mathbb{R}[X]$ попарно сопряжены. Мы приходим к заключению, что для элементов факториального кольца $\mathbb{R}[X]$ справедливо следующее утверждение.

Теорема 1. *Любой нормализованный многочлен $f \in \mathbb{R}[X]$ степени n разлагается единственным образом (с точностью до порядка множителей) в произведение $m \leq n$ линейных многочленов $X - c_i$, соответствующих его вещественным корням c_1, \dots, c_m , и $(n-m)/2$ квадратных многочленов, неприводимых над \mathbb{R} и соответствующих парам комплексно сопряжённых корней.*

Замечания. 1) Неприводимый многочлен из $\mathbb{R}[X]$ либо линеен, либо квадратичен, с отрицательным дискриминантом.

2) В обозначениях теоремы 1 имеет место соотношение

$$D(f) = (-1)^{(n-m)/2} |D(f)|,$$

т.е. знак дискриминанта определяется числом пар комплексно сопряжённых корней. Это соотношение получается либо непосредственно из определения дискриминанта, либо при помощи формулы, содержащейся в упр. 6 из § 2.

Пример 1. $g(X) = X^{2n} + 1$. Так как $g(X)$ вещественных корней не имеет, а комплексные корни $c_k = \cos \frac{(2k-1)\pi}{2n} + i \sin \frac{(2k-1)\pi}{2n}$, $1 \leq k \leq 2n$ (определённые по формуле (16) из § 1 гл. 5), все простые, то вещественные неприводимые множители входят в $g(X)$ с показателями 1. Очевидно, $\bar{c}_k = c_{2n+1-k}$ при $k = 1, 2, \dots, n$ и $(X - c_k)(X - \bar{c}_k) = X^2 - \left(2 \cos \frac{(2k-1)\pi}{2n}\right) X + 1$, так что

$$X^{2n} + 1 = \prod_{k=1}^n \left[X^2 - \left(2 \cos \frac{(2k-1)\pi}{2n}\right) X + 1 \right]. \quad (1)$$

2. Простейшие дроби над \mathbb{C} и \mathbb{R} . Теперь, когда мы знаем общий вид неприводимых многочленов над \mathbb{C} и \mathbb{R} , естественно вернуться к теме простейших дробей (п. 3 § 4 гл. 5), поскольку именно эти случаи важны в теории интегрирования. Мы знаем, что нормализованные неприводимые многочлены имеют вид $X - c$ над \mathbb{C} и $X^2 + aX + b$ или $X - c$ над \mathbb{R} . Поэтому простейшими дробями над \mathbb{C} будут $\gamma/(X - c)^m$, $\gamma \in \mathbb{C}$, а в случае \mathbb{R} к ним добавятся дроби вида $(\alpha X + \beta)/(X^2 + aX + b)^m$. Практически удобным методом разложения правильной дроби f/g в сумму простейших над \mathbb{C} и \mathbb{R} при известном каноническом разложении знаменателя $g(X)$ является метод неопределённых коэффициентов. Проиллюстрируем его парой примеров.

Пример 2. Если $g(X) = (X+1)^2(X^2+1)$ — каноническое разложение над \mathbb{R} , то для дроби $1/g(X)$ имеем

$$\frac{1}{(X+1)^2(X^2+1)} = \frac{\alpha}{(X+1)^2} + \frac{\beta}{X+1} + \frac{\gamma X + \delta}{X^2+1}$$

с не определёнными пока коэффициентами $\alpha, \beta, \gamma, \delta \in \mathbb{R}$. Умножив обе части этого равенства на $g(X)$, получим

$$1 = \alpha(X^2+1) + \beta(X+1)(X^2+1) + (\gamma X + \delta)(X+1)^2. \quad (*)$$

Сравнивая теперь коэффициенты при $1, X, X^2, X^3$, придём к неоднородной линейной системе из четырёх уравнений с четырьмя неизвестными

$$\begin{aligned}\alpha + \beta + & \delta = 1, \\ \beta + \gamma + 2\delta &= 0, \\ \alpha + \beta + 2\gamma + & \delta = 0, \\ \beta + \gamma &= 0,\end{aligned}$$

которая, конечно же, совместна и определённа, как это следует из теоремы 3 из § 4 гл. 5. Решая её, приходим к заключению, что

$$\frac{1}{(X+1)^2(X^2+1)} = \frac{1}{2(X+1)^2} + \frac{1}{2(X+1)} - \frac{X}{2(X^2+1)}.$$

Можно поступить более разумно, подставляя непосредственно в $(*)$ вместо X конкретные числовые значения $-1, i$ (корни неприводимых множителей). Это сразу даст $\alpha = 1/2$, $(i\gamma + \delta)2i = 1$, или $\gamma = -1/2$, $\delta = 0$. При $X = 0$ получится соотношение для β .

Пример 3. Пусть $\deg f(X) < n$, $g(X) = (X - c_1)(X - c_2) \dots (X - c_n)$ с попарно различными элементами c_1, c_2, \dots, c_n из \mathbb{C} или \mathbb{R} . Тогда

$$\frac{f(X)}{(X - c_1)(X - c_2) \dots (X - c_n)} = \frac{\alpha_1}{X - c_1} + \frac{\alpha_2}{X - c_2} + \dots + \frac{\alpha_n}{X - c_n},$$

откуда

$$f(X) = \sum_{k=1}^n \alpha_k (X - c_1) \dots (X - c_{k-1})(X - c_{k+1}) \dots (X - c_n).$$

При $X = c_k$, $1 \leq k \leq n$, имеем

$$f(c_k) = \alpha_k (c_k - c_1) \dots (c_k - c_{k-1})(c_k - c_{k+1}) \dots (c_k - c_n),$$

а так как

$$g'(X) = \sum_{k=1}^n (X - c_1) \dots (X - c_{k-1})(X - c_{k+1}) \dots (X - c_n),$$

то

$$\alpha_k = \frac{f(c_k)}{g'(c_k)}, \quad 1 \leq k \leq n.$$

Получающаяся в результате формула Лагранжа

$$\frac{f(X)}{g(X)} = \sum_{k=1}^n \frac{f(c_k)}{g'(c_k)(X - c_k)} \quad (2)$$

имеет прямое отношение к интерполяционной формуле Лагранжа (4) из § 1. Действительно, если умножить обе части в (2) на $g(X)$ и положить $f(c_k) = b_k$, то получится формула (4) из § 1.

В применении к дроби $1/g(X)$ с $g(X) = X^2 + 1$ рассуждения из примера 1 и формула (2) показывают, что

$$\frac{1}{X^{2n} + 1} = -\frac{1}{2n} \sum_{k=1}^{2n} \frac{c_k}{X - c_k}$$

— разложение на простейшие дроби над \mathbb{C} , поскольку $g'(X) = 2nX^{2n-1}$, $g'(c_k) = 2nc_k^{-1}c_k^{2n} = -2nc_k^{-1}$. Объединение слагаемых с комплексно сопряжёнными коэффициентами даёт нам разложение на простейшие дроби над \mathbb{R} :

$$\frac{1}{X^{2n} + 1} = -\frac{1}{2n} \sum_{k=1}^n \left(\frac{c_k}{X - c_k} + \frac{\bar{c}_k}{X - \bar{c}_k} \right) = \frac{1}{n} \sum_{k=1}^n \frac{1 - \left(\frac{\cos((2k-1)\pi)}{2n} \right) X}{X^2 - \left(2 \cos \frac{(2k-1)\pi}{2n} \right) X + 1}.$$

3. Проблема локализации корней многочлена. Будем смотреть на многочлен $f \in \mathbb{R}[X]$ как на вещественнонозначную функцию $x \mapsto f(x)$ вещественного аргумента x , изображая последнюю графиком на плоскости с прямоугольной системой координат xOy . Вещественным корням многочлена $f(X)$ (или нулям функции $f(x)$) отвечают абсциссы точек пересечения графика с осью x (рис. 25). Следует

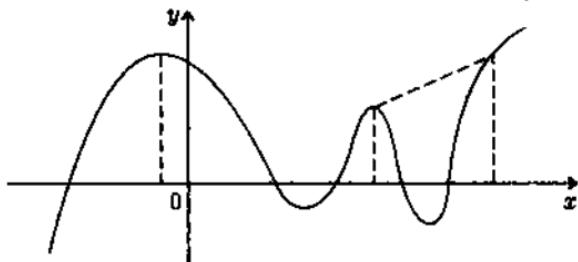


Рис. 25

ожидать, что корни алгебраического уравнения $f(x) = 0$ будут находиться между экстремальными точками (или в экстремальных точках), являющимися в свою очередь корнями алгебраического уравнения $f'(x) = 0$ более низкой степени.

Первый важный вопрос, с которым обычно сталкиваются на практике, — это вопрос о границах вещественных корней, т.е. об интервале $a < x < b$, внутри которого должны содержаться все вещественные корни заданного многочлена f . Собственно говоря, из леммы 3 § 3 мы уже знаем, что при $|x| > A/|a_0| + 1$ (a_0 — старший коэффициент, $A = \max \{|a_1|, \dots, |a_n|\}$) функция $f(x)$ не обращается в нуль, даже если бы мы вышли на комплексную плоскость. Более точные границы корней указаны в упр. 1–4.

Общая проблема локализации (отделения) корней многочлена заключается в том, чтобы для каждого из вещественных корней указать интервал, внутри которого находится только один этот корень, а для каждого интервала указать число находящихся на нём вещественных корней.

Впервые удовлетворительное, хотя и несколько громоздкое решение этой задачи было достигнуто Штурмом в 1829 г. Прежде чем переходить к формулировке соответствующей теоремы и её доказательству, введём необходимые определения.

Определение 1. Пусть $S = \{c_1, c_2, \dots, c_m\}$ — конечная последовательность отличных от нуля вещественных чисел, и пусть $V(S)$ — число индексов i , $1 \leq i \leq m-1$, для которых $c_i c_{i+1} < 0$. Тогда $V(S)$ называется числом перемен знаков в последовательности S . Если S содержит нули, то под $V(S)$ следует понимать число перемен знаков в укороченной последовательности S' , получающейся из S вычёркиванием нулей.

Например, $V(\{1, 0, 2, 0, -3, 4, 0, 0, -2\}) = 3$. В дальнейшем без ограничения общности будем предполагать, что интересующий нас многочлен $f(x)$ с вещественными коэффициентами не имеет кратных корней, чего, как мы знаем (см. конец п. 4 § 1), всегда можно добиться.

Определение 2. Конечная упорядоченная последовательность отличных от нуля многочленов с вещественными коэффициентами

$$f_0(x) = f(x), \quad f_1(x), \quad \dots, \quad f_s(x) \quad (3)$$

называется *системой Штурма* (или *рядом Штурма*) для многочлена $f(x)$ на отрезке $[a, b]$ ($a \leq x \leq b$), если выполнены следующие условия:

- i) последний многочлен $f_s(x)$ не имеет корней на $[a, b]$;
- ii) $f_0(a)f_0(b) \neq 0$;
- iii) если $f_k(c) = 0$ для $c \in [a, b]$ и $1 \leq k \leq s-1$, то $f_{k-1}(c)f_{k+1}(c) < 0$;
- iv) если $f(c) = 0$ для $c \in [a, b]$, то произведение $f_0(x)f_1(x)$ меняет знак с минуса на плюс, когда x , возрастая, проходит через точку c . Другими словами, существует такое $\delta > 0$, что $f_0(x)f_1(x) < 0$ для $x \in]c - \delta, c[$ и $f_0(x)f_1(x) > 0$ для $x \in]c, c + \delta[$.

Заметим, что соседние многочлены системы (3) не имеют на $[a, b]$ общих корней: если $f_{k-1}(c) = f_k(c) = 0$, $k \geq 1$, то $f_{k-1}(c)f_{k+1}(c) = 0$, что противоречит условию iii).

Положим для краткости

$$V_c = V_c(f) = V(\{f_0(c), f_1(c), \dots, f_s(c)\}), \quad c \in [a, b].$$

Теорема 2 (Штурм). Число корней вещественного многочлена $f(x)$ степени $n \geq 1$ на интервале $]a, b[$ равно разности $V_a - V_b$, где величины V_a, V_b отвечают какой-то фиксированной системе Штурма (3).

Доказательство. Совокупность всех различных вещественных корней на $[a, b]$ многочленов системы Штурма (3) разбивает отрезок $[a, b]$ на подинтервалы $]a_j, a_{j+1}[$ с $a = a_0 < a_1 < \dots < a_m = b$, в которых ни один из многочленов f_i , $0 \leq i \leq s$, не имеет корней. Мы собираемся сравнить значения V_c для различных точек $c \in]a_j, a_{j+1}[$.

Для начала пусть $c \in]a_0, a_1[$, так что f_0, \dots, f_s не имеют корней в $]a_0, c[$. По теореме Больцано—Коши о промежуточном значении должно выполняться условие $f_i(a_0)f_i(c) \geq 0$ для $0 \leq i \leq s$. В случае $f_i(c) \neq 0$ для всех i имеем $f_i(a_0)f_i(c) > 0$, откуда $V_{a_0} = V_c$. В случае же $f_k(a_0) = 0$ для некоторого k обязательно $k \neq 0, s$ из-за свойств i), ii) системы Штурма. По свойству iii) имеем $f_{k-1}(a_0)f_{k+1}(a_0) < 0$. В то же время $f_{k-1}(x)$ и $f_{k+1}(x)$ не имеют корней в $]a_0, c[$, так что по теореме Больцано—Коши $f_{k-1}(a_0)f_{k-1}(c) > 0$ и $f_{k+1}(a_0)f_{k+1}(c) > 0$. Значит, что $f_{k-1}(c)f_{k+1}(c) < 0$. Мы приходим к выводу, что при вычислении V_{a_0} и V_c подпоследовательности $f_{k-1}(a_0), 0, f_{k+1}(a_0)$ и $f_{k-1}(c), f_k(c), f_{k+1}(c)$, независимо от значения $f_k(c)$, вносят одинаковый вклад (по одной перемене знака). Это верно для всех k с $f(a_0) = 0$, поэтому $V_{a_0} = V_c$. Аналогичное рассуждение годится для точки из другого крайнего интервала: $c \in]a_{m-1}, a_m[\Rightarrow V_c = V_{a_m}$.

Пусть теперь $c \in]a_{j-1}, a_j[, c' \in]a_j, a_{j+1}[$ — точки из двух соседних интервалов, $1 < j < m-1$ (рис. 26). Действуют те же соображения. Именно, соединение уже проведённых рассуждений показывает, что $V_c = V_{c'}$, если только $f(a_j) \neq 0$:

$$V_c = V_{a_j} = V_{c'}$$

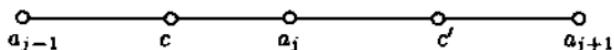


Рис. 26

В случае $f_0(a_j) = f(a_j) = 0$ впервые появляется различие. По условию iv) имеем $f_0(c)f_1(c) < 0$ и $f_0(c')f_1(c') > 0$, т.е. у подпоследовательности $f_0(c), f_1(c)$ будет одно изменение знака, а у $f_0(c'), f_1(c')$ — ни одного. В то же время наши предыдущие рассуждения показывают, что при $k > 1$ у подпоследовательностей $f_{k-1}(c), f_k(c), f_{k+1}(c)$ и $f_{k-1}(c'), f_k(c'), f_{k+1}(c')$ число перемен знаков одинаково. Все это означает, что если $f(a_j) = 0$, то $V_c - V_{c'} = 1$.

Фиксируем точки $c_k \in]a_{k-1}, a_k[, 1 \leq k \leq m$, и записываем тождество

$$V_a - V_b = (V_a - V_{c_1}) + \sum_{k=1}^{m-1} (V_{c_k} - V_{c_{k+1}}) + (V_{c_m} - V_b).$$

Мы знаем, что выражения в крайних скобках равны нулю, в то время как

$$V_{c_k} - V_{c_{k+1}} = \begin{cases} 0, & \text{если } f(a_k) \neq 0, \\ 1, & \text{если } f(a_k) = 0. \end{cases}$$

Других корней на отрезке $[a, b]$ у многочлена $f(x)$ нет (по построению все корни многочленов системы Штурма сосредоточены в точках $a_0, a_1, a_2, \dots, a_m$). Суммируя, получаем окончательно, что разность $V_a - V_b$ равна числу корней многочлена $f(x)$ на интервале $]a, b[$. \square

Чтобы применять доказанную теорему, надо научиться строить системы Штурма для каждого конкретного вещественного многочлена $f(x)$. Чаще всего используется *стандартная система Штурма*, получающаяся небольшим видоизменением известного нам из гл. 5 алгоритма Евклида. Именно, в последовательности (5) из § 3 гл. 5, начинающейся с $f_0(x) = f(x)$, $f_1(x) = f'(x)$ (производная многочлена), остаток, взятый с обратным знаком, принимаем за очередной многочлен строящейся системы. Более точно, полагаем

$$\begin{aligned} f_0(x) &= f(x), \quad f_1(x) = f'(x); \\ f_0(x) &= q_1(x)f_1(x) - f_2(x), \quad \deg f_2 < \deg f_1; \\ &\dots \\ f_{k-1}(x) &= q_k(x)f_k(x) - f_{k+1}(x), \quad \deg f_{k+1} < \deg f_k; \\ &\dots \\ f_{s-1}(x) &= q_s(x)f_s(x). \end{aligned} \tag{4}$$

По определению $f_s(x) = \text{НОД}(f, f')$ — отличная от нуля константа, поскольку мы предполагаем, что $f(x)$ не имеет кратных корней (если мы этого заранее не знали, то, получив систему (4), перешли бы к системе $g_k(x) = f_k(x)/f_s(x)$, $0 \leq k \leq s$).

Теорема 2. *Только что построенная система*

$$f_0(x) = f(x), \quad f_1(x) = f'(x), \quad f_2(x), \quad \dots, \quad f_s(x) \tag{5}$$

является системой Штурма.

Действительно, свойство ii) выполнено по предложению, а свойство i) входит в определение $f_s(x) = \text{const} \neq 0$. Если $f_k(c) = 0$, то из (4) видно, что $f_{k-1}(c)f_{k+1}(c) \leq 0$, причём $f_{k-1}(c) = 0$ в точности тогда, когда $f_{k+1}(c) = 0$. Но если это так, то $0 = f_{k-1}(c) = f_k(c) = f_{k+1}(c) = f_{k+2}(c) = \dots$ вопреки тому, что $f_s(c) \neq 0$. Стало быть, $f_{k-1}(c)f_{k+1}(c) < 0$, а это есть свойство iii). Наконец, предположим, что $f_0(c) = 0$ для некоторой точки $c \in [a, b]$. Тогда $f_0(x) = (x - c)q(x)$, $q(c) \neq 0$ и $f_0(x)f_1(x) = (x - c)[q^2(x) + (x - c)q(x)q'(x)] = (x - c)g(x)$, где $g(x) = q^2(x) + (x - c)q(x)q'(x)$. Имеем $g(c) = q^2(c) > 0$, и, следовательно, $g(x)$ принимает положительные значения в малой окрестности $[c - \delta, c + \delta]$ точки c . Множитель $x - c$, однако, способствует тому, что произведение $f_0(x)f_1(x)$ меняет знак с минуса на плюс при возрастании x и прохождении его через c . Таким образом, система (5) обладает свойством iv). \square

Замечание 1. Система

$$\lambda_0 f_0(x), \quad \lambda_1 f_1(x), \quad \dots, \quad \lambda_s f_s(x), \tag{5'}$$

получающаяся из (5) умножением её членов на положительные константы $\lambda_0, \lambda_1, \dots, \lambda_s$, также будет системой Штурма. Будем на-

зывать её почти стандартной системой Штурма. Это полезно иметь в виду при вычислениях.

Замечание 2. Условие отсутствия кратных корней у $f(x)$ не является существенным при подсчёте числа различных вещественных корней, как показывает конструкция стандартной системы Штурма: следует перейти от $f_k(x)$ к $g_k(x) = f_k(x)/f_s(x)$ и заметить, что $V_c(g) = V_c(f)$.

Замечание 3. Согласно лемме 3 из § 3 для каждого многочлена $f_i(x)$ системы Штурма существует такое число r_i , что вещественные корни этого многочлена лежат между $-r_i$ и r_i . Пусть M — любое достаточно большое число, скажем, $M = \max_{0 \leq i \leq s} r_i$. Тогда вещественные корни всех многочленов $f_i(x) = a_{(i)}x^{k_i} + \dots$ расположены между $-M$ и M . Более того, при $x = M$ знак $f_i(M)$ совпадает со знаком его старшего члена $a_{(i)}M^{k_i}$. Конкретное значение величины M совершенно несущественно для нашей процедуры, поэтому при разыскании общего числа различных вещественных корней многочлена $f(x)$ мы полагаем чисто символически $x = -M$ и $x = M$.

Замечание 4. По преданию, сам Штурм так гордился своим (действительно замечательным) достижением, что обычно, изложив доказательство студентам, добавлял: "Вот теорема, имя которой я нашу".

Рассмотрим несколько примеров.

Пример 4. $f(x) = x^3 + 3x - 1$. Находим $f_1(x) = f'(x) = 3x^2 + 3$; далее, $f(x) = (3x^2 + 3) \cdot \frac{1}{3}x + 2x - 1$ и $f_2(x) = -2x + 1$; $3x^2 + 3 = (-2x + 1) \left(-\frac{3}{2}x - \frac{3}{4}\right) + \frac{15}{4}$ и $f_3(x) = -\frac{15}{4}$. Согласно замечанию 1 в качестве системы Штурма можно взять $x^3 + 3x - 1$, $x^2 + 1$, $-2x + 1$, -1 . Составим таблицу знаков для старших членов:

	x^3	$3x^2$	$-2x$	-1	V
$x = -M$	—	+	+	—	2
$x = M$	+	+	—	—	1

Получаем $V_{-M} - V_M = 1$, т.е. $x^3 + 3x - 1$ имеет один вещественный корень.

Пример 5. $f(x) = x^3 + 3x^2 - 1$. Легко видеть, что стандартная система Штурма для $f(x)$ имеет вид $x^3 + 3x^2 - 1$, $3x^2 + 6x$, $2x + 1$, 1 , а таблицей знаков для старших членов будет

	x^3	$3x^2$	$2x$	1	V
$x = -M$	—	+	—	+	3
$x = M$	+	+	+	+	0

Приходим к выводу, что $f(x)$ имеет три вещественных корня: $V_{-M} - V_M = 3$.

Пример 6. $f(x) = 1 + x + \frac{x^2}{2!} + \dots + \frac{x^n}{n!}$ (срезанная экспонента). Очевидно, что вещественные корни этого многочлена, если они есть, находятся в интервале $[-M, -\delta]$, где $\delta > 0$ — достаточно малое вещественное число (как всегда, M — большое положительное число). В качестве нестандартной системы Штурма на отрезке $[-M, -\delta]$ можно взять тройку $f_0(x) = f(x)$, $f_1(x) = f'(x) = 1 + x + \frac{x^2}{2!} + \dots + \frac{x^{n-1}}{(n-1)!}$ и $-\frac{x^n}{n!} = -f(x) + f'(x)$ (проверить, что все свойства i)-iv)

выполнены). Из таблицы знаков

	f_0	f_1	f_2	V
$-M$	$(-1)^n$	$(-1)^{n-1}$	$(-1)^{n-1}$	1
δ	+	+	$(-1)^{n-1}$	$(1 + (-1)^n)/2$

видно, что $f(x)$ при чётном n не имеет вещественных корней, а при нечётном n имеет один отрицательный корень (как легко понять, стремящийся к $-\infty$ при возрастании $n = 2m+1$).

Пример 7. $f(x) = x^3 + px + q$. Почти стандартной системой Штурма (см. замечание 1) для $f(x)$ на любом отрезке $[a, b]$ может служить $f_0 = f$, $f_1 = 3x^2 + p$, $f_2 = -2px - 3q$, $f_3 = -4p^3 - 27q^2$ с естественным ограничением $f(a)f(b) \neq 0$. Полезно отметить, что $f_3 = D(f)$ — дискриминант многочлена f (см. (16) из § 2, где следует заменить a и b на более традиционные коэффициенты p и q). Из общих соображений ясно, что $f(x)$ имеет либо один вещественный корень, либо три. Если рассмотреть три варианта для пары $(\operatorname{sgn} p, \operatorname{sgn} D(f))$ с учётом импликации $p \geq 0 \Rightarrow D(f) \leq 0$, то из таблицы для знаков

	x^3	x^2	$-2px$	$D(f)$
$-M$	—	+	$\operatorname{sgn} p$	$\operatorname{sgn} D(f)$
M	+	+	$-\operatorname{sgn} p$	$\operatorname{sgn} D(f)$

легко извлекается правило, согласно которому один корень будет при $D(f) < 0$, а три — при $D(f) > 0$.

4. Вещественные многочлены с вещественными корнями. Мы остановимся ещё на практически важном случае, когда из каких-либо соображений известно, что все корни многочлена $f(x) = \sum_{k=0}^n a_k x^{n-k} \in \mathbb{R}[x]$ вещественные. Для удобства введём два обозначения:

$m(f)$ — число положительных корней многочлена f (с учётом «ратностей»);

$W(f) = V(\{a_0, a_1, \dots, a_n\})$ — число перемен знаков в упорядоченной последовательности коэффициентов многочлена f .

Ясно, что всегда $0 \leq W(f) \leq n = \deg f$, причем $W(-f) = W(f)$. Заметим также, что $W(f) = W(aX^k + a_{i_1}X^{n-i_1} + \dots)$, где показатель k удовлетворяет единственному условию $k > n - i_1$ (коэффициенты a_1, \dots, a_{i_1-1} нулевые) и $aa_0 > 0$. Если $W(f) = 0$, то, очевидно, f не имеет положительных корней. С другой стороны, у f может не быть положительных корней и в том случае, когда $W(f) = \deg f$. Пример: $f(X) = X^2 - X + 1$. Всё же, как мы увидим, символ $W(f)$ имеет прямое отношение к числу положительных корней многочлена f . Справедливо, например, следующее *правило знаков Декарта*: $W(f) \geq m(f)$, причём $m(f) = W(f) \pmod{2}$. Не останавливаясь на его доказательстве, перейдём к интересующему нас случаю.

Теорема 3. Пусть все корни многочлена $f \in \mathbb{R}[X]$ вещественные. Тогда $m(f) = W(f)$.

Доказательство. Будем исходить из наглядных соображений. По известной теореме Ролля из анализа (или по теореме о среднем) между корнями a' и b' нашего многочлена $f(X)$ найдётся число $c \in \mathbb{R}$, $a' < c < b'$, для которого $f'(c) = 0$. Отсюда следует,

что все корни производной $f'(X)$ вещественны и $m(f') = m(f)$ или $m(f') = m(f) - 1$.

В самом деле, пусть $c_1 < c_2 < \dots < c_r$ — корни многочлена f кратностей n_1, n_2, \dots, n_r , так что $n_1 + n_2 + \dots + n_r = \deg f = n$. По теореме 5 из § 1 производная f' имеет корни c_1, c_2, \dots, c_r кратностей $n_1 - 1, n_2 - 1, \dots, n_r - 1$, а в промежутках между ними по теореме Ролля ещё хотя бы по одному корню $c'_1, c'_2, \dots, c'_{r-1}$. Всего получается $(n_1 - 1) + \dots + (n_r - 1) + r - 1 = n - 1$ вещественных корней. Так как $\deg f' = n - 1$, то других корней у f' нет.

Пусть, далее, $c_{l-1} < 0$, а c_l, \dots, c_r — все положительные корни кратностей n_l, \dots, n_r : $n_l + \dots + n_r = m = m(f)$. Положительными корнями производной $f(X)$ будут корни c_l, \dots, c_r кратностей $n_l - 1, \dots, n_r - 1$, корни c'_l, \dots, c'_{r-1} и, возможно, ещё корень c'_{l-1} , т.е. число их $m(f') = m(f) - 1$ или $m(f)$, как и утверждалось. Аналитическим выражением этого факта служит почти тавтологическая формула

$$m(f) = m(f') + \varepsilon, \quad \varepsilon = \frac{1}{2}(1 - (-1)^{m(f)+m(f')}). \quad (6)$$

Заметим ещё, что если

$$f(X) = a_0 X^n + \dots + a_{n-\nu} X^\nu, \quad (7)$$

где $a_{n-\nu}$ — последний отличный от нуля коэффициент, то

$$f(X) = (X - c_l)^{n_l} \dots (X - c_r)^{n_r} g(X).$$

Здесь

$$g(X) = a_0 X^{n-m} + \dots + b X^\nu, \quad a_0 > 0, b > 0 (\nu \leq 0).$$

Таким образом $a_{n-\nu} = (-1)^m c_l^{n_l} \dots c_r^{n_r} b$, причём $c_l^{n_l} \dots c_r^{n_r} b > 0$. Другими словами,

$$(-1)^{m(f)} a_{n-\nu} > 0. \quad (8)$$

При $n = 1, 2$ утверждение теоремы очевидно. Рассуждая теперь по индукции относительно $n = \deg f$, допустим, что теорема доказана для всех многочленов степени $< n$. Если в (7) $\nu > 0$, т.е. $a_n = 0$, то $f(X) = X \cdot f_1(X)$, причем $m(f) = m(f_1) = W(f)$ ($m(f_1) = W(f_1)$ по индукции). Остается рассмотреть случаи $a_n \neq 0$. Пусть

$$f'(X) = n a_0 X^{n-1} + \dots + \mu a_{n-\mu} X^{\mu-1}, \quad a_{n-\mu} \neq 0.$$

Тогда

$$W(f) = W(f') + \delta, \quad \delta = \frac{1}{2} \left(1 - \frac{a_n a_{n-\mu}}{|a_n a_{n-\mu}|} \right) = 0 \text{ или } \delta = 1.$$

Но мы знаем (см. (8)), что $(-1)^{m(f)} a_n > 0$ и $(-1)^{m(f')} a_{n-\mu} > 0$. Поэтому $\delta = (1 - (-1)^{m(f)+m(f')})/2$ и, стало быть, $\delta = \varepsilon$. Так как по предположению индукции $W(f') = m(f')$, то окончательно имеем $W(f) = m(f') + \varepsilon$ или, сравнивая с (6), $m(f) = W(f)$. \square

Следствие (частный случай теоремы Бюдана—Фурье). Пусть все корни многочлена f вещественны.

Тогда число его корней, лежащих в интервале $[a, b]$, равно $W(f_a) - W(f_b)$, где

$$f_a(X) = f(X + a) = \sum_{0 \leq k \leq n} \frac{f^{(k)}(a)}{k!} X^k,$$

$$f_b(X) = f(X + b) = \sum_{0 \leq k \leq n} \frac{f^{(k)}(b)}{k!} X^k$$

— разложения в ряд Тейлора (см. упр. 3).

Доказательство. По определению число $m(f_a)$ положительных корней многочлена f_a равно числу корней заданного многочлена f , больших, чем a . То же замечание относится к f_b . Следовательно, число корней многочлена f , заключенных между a и b ($a < b$), равно разности $m(f_a) - m(f_b)$, которая по теореме 3 выражается в виде $W(f_a) - W(f_b)$. \square

5. Устойчивые многочлены.

Нормализованный многочлен

$$f(X) = X^n + a_1 X^{n-1} + \dots + a_{n-1} X + a_n$$

с вещественными коэффициентами называется *устойчивым*, если все его корни лежат в левой полуплоскости (рис. 27):

$$f(\lambda) = 0, \lambda = \alpha + i\beta \Rightarrow \alpha < 0.$$

Терминология ведёт своё происхождение из теории дифференциальных уравнений. Получаемые там критерии асимптотически устойчивого поведения физической (а в более широком смысле — механической, технической или экономической) системы в окрестности положения равновесия требуют, чтобы было

$$\lim_{t \rightarrow +\infty} e^{\lambda t} = 0, \quad (9)$$

где λ — произвольный корень многочлена f , ассоциированного с дифференциальным уравнением порядка n с постоянными коэффициентами. Так как по формуле Эйлера (см. (15) из § 1 гл. 5) $e^{\lambda t} = e^{\alpha t} e^{i\beta t} = e^{\alpha t} (\cos \beta t + i \sin \beta t)$, то доминирующим членом является $e^{\alpha t}$, и условие (9) эквивалентно неравенству $\alpha < 0$.

Возникает своеобразная проблема локализации — *проблема Раяса—Гурвица*²⁾, когда непосредственно по коэффициентам мно-

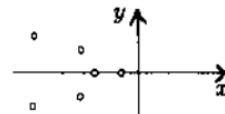


Рис. 27

²⁾ Фактически поставленная гораздо раньше (1868 г.) английским физиком Д.К. Маклевеллом и решённая для небольших степеней русским инженером И.А. Вышнеградским, который занимался задачей устойчивости регулятора (1876 г.).

гочлена f надлежит выяснить, является ли он устойчивым. Эта алгебраическая задача была решена ещё в 1895 г. Критерий Рауса—Гурвица гласит следующее.

Многочлен f устойчив тогда и только тогда, когда выполнены неравенства

$$\Gamma_1 > 0, \quad \Gamma_2 > 0, \quad \dots, \quad \Gamma_n > 0, \quad (10)$$

где

$$\Gamma_k = \begin{vmatrix} a_1 & 1 & 0 & 0 & 0 & 0 & \dots & 0 \\ a_3 & a_2 & a_1 & 1 & 0 & 0 & \dots & 0 \\ a_5 & a_4 & a_3 & a_2 & a_1 & 1 & \dots & 0 \\ a_7 & a_6 & a_5 & a_4 & a_3 & a_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{2k-1} & a_{2k-2} & a_{2k-3} & a_{2k-4} & a_{2k-5} & a_{2k-6} & \dots & a_k \end{vmatrix}$$

(предполагается, что $a_s = 0$ при $s > n$).

Не пытаясь доказать теорему Рауса—Гурвица (это более уместно делать в других курсах), мы обратим внимание на то обстоятельство, что её формулировка своим изяществом целиком обязана теории определителей.

Далее, согласно теореме 1 при выполнении условий (10) многочлен $f(X)$ представляется в виде произведения множителей вида $X + u$, $X^2 + vX + w$ с $u > 0$, $v > 0$, $w > 0$, а это значит, что все коэффициенты устойчивого многочлена $f(X)$ положительны:

$$a_1 > 0, \quad a_2 > 0, \quad \dots, \quad a_n > 0. \quad (11)$$

Таким образом, условия (11) необходимы для устойчивости многочлена $f(X)$. Не являясь в общем случае достаточными, они позволяют, однако, приблизительно вдвое понизить число детерминантных неравенств (10). Это удобно, так как вычисление определителей — трудоёмкое дело.

Пример 8. При $n = 2$ система неравенств $\Gamma_1 > 0$, $\Gamma_2 > 0$ эквивалентна более простой: $a_1 > 0$, $a_2 > 0$, что, между прочим, видно из формул для корней квадратного уравнения.

При $n = 3$ все сводится к неравенствам $a_1 > 0$, $a_2 > 0$, $a_3 > 0$, $a_1 a_2 > a_3$, поскольку $\Gamma_3 = a_3(a_1 a_2 - a_3)$.

Наконец отметим, что критерий Рауса—Гурвица не решает всех вопросов, связанных с устойчивостью, поскольку на практике речь идёт о многочленах и о дифференциальных уравнениях, коэффициенты которых зависят от параметра. В терминах самого параметра должны формулироваться и условия устойчивости, что представляет собой задачу совсем иной природы.

6. Зависимость корней многочлена от коэффициентов. Понятно, что корни многочлена являются функциями его коэффициентов. Мы хотим теперь подчеркнуть, что эти функции непрерывны, т.е. при достаточно малом изменении коэффициентов измен-

нение корней пренебрежимо мало. Кратные корни, впрочем, могут распадаться, и геометрически динамика изменения зачастую приобретает причудливые формы. Достаточно сравнить многочлены z^n и $z^n + \varepsilon$ при $\varepsilon \rightarrow 0$, чтобы почувствовать эту сложность. Качественному и количественному сравнению многочленов способствует

Теорема 4 (Руше). *Если для двух многочленов $f_0(z)$ и $f_1(z)$ имеет место неравенство*

$$|f_1(z)| < |f_0(z)|$$

для всех z на границе замкнутой области $D \subset \mathbb{C}$, то многочлен $f_0(z) + f_1(z)$ имеет внутри D столько же корней, сколько и многочлен $f_0(z)$.

Доказательство этой теоремы, причём в более общем контексте, получается элементарными средствами теории функций комплексной переменной, поэтому мы его опускаем. \square

Пусть теперь z_0 — корень кратности k многочлена $f_0(z) = a_0 z^n + a_1 z^{n-1} + \dots + a_n$. Рассмотрим многочлен

$$f(z) = (a_0 + \delta_0)z^n + (a_1 + \delta_1)z^{n-1} + \dots + (a_n + \delta_n) = f_0(z) + f_1(z)$$

с $|\delta_i| < \delta$, где $\delta > 0$ — сколь угодно малое вещественное число. Рассмотрим круг $D = \{z \in \mathbb{C} \mid |z - z_0| \leq \varepsilon\}$ с центром в z_0 и столь малого радиуса $\varepsilon > 0$, что z_0 — единственный корень многочлена $f_0(z)$ в замкнутой области D . Функция $|f(z)|$ непрерывна и не обращается в нуль на окружности $|z - z_0| = \varepsilon$ — границе круга D , поэтому $\mu = \inf_{|z-z_0|=\varepsilon} |f_0(z)| > 0$. Возьмём δ столь малым, чтобы при $|z - z_0| = \varepsilon$ имело место неравенство $|f_1(z)| < \mu$. Тогда будут выполнены условия теоремы Руше, и мы приходим к выводу, что $f(z)$ и $f_0(z)$ имеют внутри D по одному числу k корней. В частности, простой корень ($k = 1$) при малом шевелении коэффициентов остаётся простым, лишь чуть смешаясь.

Фактически мы обеспечили ту локализацию корней многочлена $f(z)$, которая гарантирует их непрерывную зависимость от коэффициентов многочлена $f_0(z)$.

Доказанный результат формулируется также в следующей форме.

Теорема 5. *Пусть $f_0(z) = z^n + a_1 z^{n-1} + \dots + a_n$ — нормализованный комплексный многочлен, c_1, \dots, c_n — его корни. Для любого $\varepsilon \in \mathbb{R}$, $\varepsilon > 0$, существует $\delta \in \mathbb{R}$, $\delta > 0$, такое, что для каждого нормализованного многочлена $f(z) = z^n + a'_1 z^{n-1} + \dots + a'_n$ такого, что $|a'_j - a_j| < \delta$, $1 \leq j \leq n$, имеет место разложение $f(z) = \prod_{j=1}^n (z - c'_j)$, причём $|c'_j - c_j| < \varepsilon$, $1 \leq j \leq n$.*

Можно привести доказательство этой теоремы, не опирающееся на теорему Руше (см., например: Amer. Math. Monthly. — 1989. — V. 96, № 4), но нам важнее обратить внимание на суть дела.

7. Вычисление корней многочлена. Полное решение проблемы локализации (особенно если учитывать все корни, включая комплексные, когда речь идёт не об интервалах, а об областях на плоскости \mathbb{C}) даётся дорогой ценой. Остановимся вкратце на методах вычисления “локализованного корня” с заданной степенью точности.

Пусть мы уже знаем достаточно узкий интервал $[a, b]$ вещественной оси, содержащий единственный интересующий нас простой корень многочлена $f(x)$. Тогда $f(a)f(b) < 0$. Разделим интервал $[a, b]$ на 10 равных частей. Одна из этих частей $[a_1, b_1] \subset [a, b]$ (и только одна) обладает свойством $f(a_1)f(b_1) < 0$. Значит, $c \in [a_1, b_1]$. Интервал $[a_1, b_1]$ делим снова на 10 равных частей и выбираем ту часть $[a_2, b_2] \subset [a_1, b_1]$, для которой $f(a_2)f(b_2) < 0$. Так как $c \in [a_1, b_1]$, то процесс можно продолжить, получая приближение к истинному значению корня с точностью до 0,1, до 0,01 и т.д. Этот метод испытаний (десятичных, если делят интервал на 10 частей; двоичных, если делят интервал пополам) удобен, если мы не претендуем на вычисления с большой точностью и располагаем простейшими вычислительными средствами.

Универсальным, но весьма трудоёмким является *метод Лобачевского*, позволяющий находить приближённые значения всех корней одновременно, в том числе и комплексных, причём без предварительной процедуры их отделения.

Широкое хождение получил *метод линейной интерполяции* (или метод *ложного положения*). Он заключается в том, что в качестве приближения к корню берётся число $c_{(1)}$, делящее интервал $[a, b]$ на части, пропорциональные $|f(a)|$ и $|f(b)|$. Другими словами,

$$\frac{c_{(1)} - a}{b - c_{(1)}} = -\frac{f(a)}{f(b)}, \quad c_{(1)} = \frac{bf(a) - af(b)}{f(a) - f(b)}.$$

Кусочек кривой $y = f(x)$ при этом заменяется хордой (рис. 28).

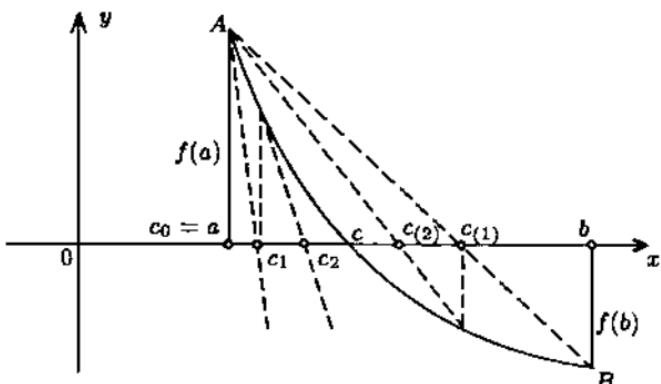


Рис. 28

Процесс можно повторить, аналогичным образом получая приближение $c_{(2)}$ и т.д.

В достаточно малой окрестности $[a, b]$ корня с кусочком той же кривой можно заменить отрезком касательной в одной из точек. Если c_0 — какое-то приближение к корню ($c_0 = a$ на рис. 28), то по теореме Лагранжа о конечном приращении имеем

$$f(x) - f(c_0) = f'(c_0)(x - c_0),$$

откуда при $x = c$ получим $0 = f(c) = f(c_0) + f'(c_0)(c - c_0)$. Поэтому в качестве следующего приближения естественно взять $c_1 = c_0 - f(c_0)/f'(c_0)$. Положим

$$c_{k+1} = c_k - \frac{f(c_k)}{f'(c_k)}, \quad k = 0, 1, 2, \dots \quad (12)$$

Предположив сходимость рекуррентной последовательности (12), а именно $c_k \rightarrow \bar{c}$ при $k \rightarrow \infty$, мы получим $\bar{c} = \bar{c} - f(\bar{c})/f'(\bar{c})$, откуда $f(\bar{c}) = 0$. При правильном выборе исходной точки c_0 все точки нашей последовательности будут лежать в интервале $[a, b]$ и $c = \bar{c}$. На рис. 28 показана лишь одна из четырёх возможных картинок, отвечающих поведению первых двух производных $f'(x)$, $f''(x)$ на интервале $[a, b]$. Детали мы опускаем, предоставив читателю самому рассмотреть оставшиеся случаи.

Только что описанный метод Ньютона относится к числу наиболее употребительных и быстро сходящихся. Элементарными методами анализа показывается, что если

$$|f(x)| \geq M_1, \quad |f''(x)| \leq M_2 \quad \text{при } x \in [a, b]$$

то $|c_1 - c| \leq \frac{M_2}{2M_1} |c_0 - c|^2$. Поэтому выбрав точку c_0 так, что

$$\frac{M_2}{2M_1} |c_0 - c| \leq q < 1,$$

мы придём к оценке $\frac{M_2}{2M_1} |c_k - c| \leq q^{2^k}$. Как говорят, имеет место *квадратичная* (или *сверхэкспоненциальная*) сходимость приближений к корню c . Метод Ньютона хорош тем, что он годится без изменений для вычисления произвольных комплексных корней многочленов из $\mathbb{C}[z]$. В основу кладётся рекуррентная последовательность (12).

Разумеется, ограничившись наброском голой схемы вычислительных методов, мы не коснулись фактической организации вычислений. Современная вычислительная математика располагает для этой цели широким арсеналом средств. Входить в профессиональные тонкости математика-вычислителя у нас нет возможности.

8. Рациональные корни целочисленных многочленов. О многочленах над \mathbb{Q} и над \mathbb{Z} мы имели возможность поговорить в п. 4

§ 3 гл. 5, где обсуждалась проблема разложения данного многочлена над \mathbb{Q} на неприводимые множители. Сейчас мы остановимся на гораздо более простом вопросе о выделении рациональных линейных множителей многочлена $f \in \mathbb{Q}[X]$, т.е. фактически о рациональных корнях. Умножив f на общий знаменатель коэффициентов, мы перейдём к многочлену из $\mathbb{Z}[X]$, поэтому целесообразно с самого начала ограничиться рассмотрением целочисленных многочленов.

Теорема 6. Пусть несократимая дробь p/q является корнем многочлена $f(X) = a_0 X^n + a_1 X^{n-1} + \dots + a_n \in \mathbb{Z}[X]$, $a_0 a_n \neq 0$.

Тогда $p|a_n$ и $q|a_0$.

Доказательство. Действительно, по условию

$$a_0 \left(\frac{p}{q}\right)^n + a_1 \left(\frac{p}{q}\right)^{n-1} + \dots + a_{n-1} \frac{p}{q} + a_n = 0.$$

После умножения обеих частей равенства на q^n получаем

$$\begin{aligned} a_0 p^n + a_1 p^{n-1} q + \dots + a_{n-1} p q^{n-1} + a_n q^n &= 0, \\ a_0 p^n &= q(-a_1 p^{n-1} - \dots - a_{n-1} p q^{n-2} - a_n q^{n-1}). \end{aligned}$$

Таким образом, $q|a_0 p^n$, а так как q и p взаимно прости, то $q|a_0$. Аналогично, из равенства

$$a_n q^n = p(-a_1 p^{n-1} - \dots - a_{n-1} p q^{n-2} - a_n q^{n-1})$$

вытекает, что $p|a_n$.

Следствие. Рациональные корни нормализованного многочлена должны быть целыми числами.

Итак, решение вопроса о наличии рациональных корней многочлена сводится к следующим действиям: 1) перебору всех делителей свободного члена и всех делителей старшего члена; 2) составлению из них несократимых дробей; 3) проверке посредством подстановки дроби в многочлен. На этом этапе можно воспользоваться методом Горнера. Если все испытания приведут к отрицательному результату, то это значит, что у многочлена нет рациональных корней.

Громоздкий перебор всех делителей полезно начинать с ± 1 . Вычисление $f(1)$ и $f(-1)$ не представляет затруднений. Если теперь целое число c является корнем многочлена $f(X)$, то $f(X) = (X - c)q(X)$, где $q(X) = b_0 X^{n-1} + b_1 X^{n-2} + \dots + b_{n-1}$. Из схемы Горнера непосредственно следует, что $b_i \in \mathbb{Z}$, $0 \leq i \leq n-1$. Поэтому частные

$$\frac{f(1)}{c-1} = -q(1), \quad \frac{f(-1)}{c+1} = -q(-1)$$

тоже должны быть целыми числами. А это значит, что если $d \in \mathbb{Z}$ и $d|a_n$, но хотя бы одно из чисел $f(1)/(d-1)$ или $f(-1)/(d+1)$ не является целым, то заведомо $f(d) \neq 0$. Разумеется, целостность $f(1)/(d-1)$ и $f(-1)/(d+1)$ не является гарантией того, что $f(d) = 0$.

Пример 9. $f(X) = X^5 + 2X^4 - 15X^3 - 2X + 6$. Имеем $f(1) = -8$, $f(-1) = 24$. Делители $d = \pm 6$ сразу отпадают, поскольку $d+1$ не делит 24. С другой стороны, для $d = 2$ имеем $f(1)/(2-1) \in \mathbb{Z}$ и $f(-1)/(2+1) \in \mathbb{Z}$, но $f(2) \neq 0$. То же относится и к $d = -3$. Целым корнем на самом деле является делитель $d = 3$.

УПРАЖНЕНИЯ

1. Пусть $f(X) = a_0X^n + a_1X^{n-1} + \dots + a_n$ — вещественный многочлен степени n . Показать, что значение верхних границ положительных корней многочленов $f(X)$, $X^n f(1/X)$, $f(-X)$, $X^n f(-1/X)$ даёт нижние и верхние границы как положительных, так и отрицательных корней многочлена $f(X)$.

2. В обозначениях упр. 1 пусть $a_0 > 0$, m — наименьший индекс, для которого $a_m < 0$, B — максимум среди абсолютных величин отрицательных коэффициентов. Показать, что

$$c \leq 1 + \sqrt[n]{B/a_0}$$

для всякого положительного вещественного корня многочлена $f(X)$.

Указание. При $x > 1$ исходить из оценки $f(x) \geq a_0x^n - B \frac{x^{n-m+1}-1}{x-1} > \frac{x^{n-m+1}}{x-1} [a_0x^{m-1}(x-1) - B]$.

3. Пусть P — поле нулевой характеристики, $a \in P$. Для любого многочлена $f \in P[X]$ степени n имеет место формула (формула Тейлора)

$$f(X) = f(a) + \frac{f'(a)}{1!}(X-a) + \frac{f''(a)}{2!}(X-a)^2 + \dots + \frac{f^{(n)}(a)}{n!}(X-a)^n.$$

Указание. Продифференцировать k раз формальное выражение $f(X) = \sum b_i(X-a)^i$ и положить $X = a$.

4. Показать, что если $f(a) > 0$, $f'(a) > 0, \dots, f^{(n)}(a) > 0$ для вещественного многочлена $f(X)$ степени n с положительным старшим коэффициентом a_0 , то $f(c) = 0$, $c > 0 \Rightarrow c < a$.

Указание. Применить упр. 3.

5. Воспользовавшись правилом знаков Декарта, найти знак дискриминанта многочленов $X^5 - X^2 + 1$, $X^3 - 6X - 9$ (см. замечание в конце п. 1).

6. Могут ли многочлены $X^5 - X - 1$, $X^3 + aX + b \in \mathbb{Q}[X]$ иметь общие комплексные корни? Напомним (см. упр. 11 из § 1), что многочлен $X^5 - X - 1$ неприводим над \mathbb{Q} .

7. Показать, что корни многочлена $f(X) = X^5 + uX^4 + vX^3 + w \in \mathbb{R}[X]$ со свободным членом $w \neq 0$ не могут быть все вещественными.

Указание. Удобно перейти к взаимному многочлену $X^5 f(1/X)$ и далее воспользоваться формулами (12) из § 1 и (8) из § 2.

8. Любой многочлен $f(X)$ с $f(x) \geq 0$ для всех $x \in \mathbb{R}$ можно представить в виде

$$f(X) = g(X)^2 + h(X)^2,$$

где $g, h \in \mathbb{R}[X]$.

Указание. При помощи теоремы 1 разложить $f(X)$ на множители вида $(X+a)^2 + b^2$ и воспользоваться формальным тождеством

$$(p^2 + q^2)(r^2 + s^2) = (pr + qs)^2 + (ps - qr)^2,$$

вытекающим из соотношения $|p + iq|^2 |r + is|^2 = |(p + iq)(r + is)|^2$.

9. Получить самостоятельно критерий устойчивости многочленов степеней 3 и 4. При $n = 4$ записать его в виде неравенств:

$$a_1 > 0, \quad a_4 > 0, \quad a_1 a_2 > 0, \quad a_3(a_1 a_2 - a_3) > a_1^2 a_4.$$

Указание. $f(X) = X^3 + aX^2 + bX + c = (X^2 + \alpha X + \beta)(X + \theta)$, где $a = \alpha + \theta$, $b = \beta + \alpha\theta$, $c = \beta\theta$, причем $\alpha, \beta, \theta \in \mathbb{R}$. Устойчивость $f(X)$ эквивалентна устойчивости пары многочленов $X^2 + \alpha X + \beta$, $X + \theta$, т.е. выполнению неравенств $\alpha > 0$, $\beta > 0$, $\theta > 0$. Легко проверяется, что эта система эквивалентна системе неравенств $a > 0$, $b > 0$, $c > 0$, $ab - c > 0$. Аналогичные соображения применять к вещественному многочлену степени 4.

10. Имеет ли многочлен $f(X)$, стоящий в числителе несократимой рациональной дроби

$$\frac{f(X)}{g(X)} = \frac{3}{X+2} + \frac{1}{(X-1)^2} - \frac{2}{X-1} + \frac{X-3}{X^2+1},$$

вещественные корни?

11. Показать, что все три корня неприводимого над \mathbb{Q} многочлена $f(z) = z^3 - 7z - 7$ — вещественные и лежат в интервале $]-2, 4[$. Вычислить положительный корень методом Ньютона с точностью до третьего десятичного знака.

12. Опираясь на теорему Руше (теорема 4), показать, что многочлен $f(z) = z^5 + 5z^2 - 3$ имеет два корня в единичном круге и три корня в кольце между окружностями $|z| = 1$ и $|z| = 2$.

13. Сколько вещественных корней имеет многочлен $z^4 + 12z^2 + 5z - 9$?

14. Многочлены Лежандра $P_0(X) = 1$, $P_1(X) = X, \dots, P_n(X), \dots$ определяются рекуррентной формулой

$$mP_m(X) - (2m-1)XP_{m-1}(X) + (m-1)P_{m-2}(X) = 0.$$

Показать, что:

а) $P_n(1) = 1$, $P_n(-1) = (-1)^n$;

б) $\{P_n, P_{n-1}, \dots, P_0\}$ — система Штурма для $P_n(X)$ на отрезке $[-1, 1]$;

в) $P_n(X)$ имеет n различных корней на интервале $]-1, 1[$.

НЕРЕШЁННЫЕ ЗАДАЧИ О МНОГОЧЛЕНАХ

Ниже звёздочками отмечены задачи, решений которых в математической литературе (по состоянию на 2000 г.) действительно не существует. Остальные задачи в принципе решены, но либо неалгебраическими, либо незлементарными средствами. Понятно, что для решения открытых проблем все средства хороши.

Формулировки задач сопровождаются небольшими комментариями, призванными способствовать пониманию существа дела и расширению кругозора. Дополняющий их список литературы минимальный.

1*. Проблема якобиана. Пусть $f_1, \dots, f_n \in \mathbb{C}[X_1, \dots, X_n]$. Под $D_j f_i = \partial f_i / \partial X_j$ понимается частная производная многочлена f_i по j -й переменной X_j — результат применения оператора частного дифференцирования (см. упр. 9 из § 1 гл. 6). Будем считать, что $f_i(0, \dots, 0) = 0$, $1 \leq i \leq n$. Введём новые переменные X'_1, \dots, X'_n формулами

$$X'_1 = f_1(X_1, \dots, X_n), \quad \dots, \quad X'_n = f_n(X_1, \dots, X_n).$$

Полиномиальное отображение $F = (f_1, \dots, f_n)$: $X_i \mapsto X'_i$, $1 \leq i \leq n$, определяет эндоморфизм (гомоморфизм в себя) алгебры многочленов $\mathbb{C}[X_1, \dots, X_n]$. Его **матрица Якоби**

$$J(F) = \begin{pmatrix} D_1 f_1 & \dots & D_n f_1 \\ \dots & \dots & \dots \\ D_1 f_n & \dots & D_n f_n \end{pmatrix}$$

обратима в точности тогда, когда определитель

$$\det J(F) = \begin{vmatrix} D_1 f_1 & \dots & D_n f_1 \\ \dots & \dots & \dots \\ D_1 f_n & \dots & D_n f_n \end{vmatrix},$$

называемый **якобианом** и являющимся, вообще говоря, многочленом, будет отличной от нуля константой (элементом из \mathbb{C}^*).

Если F — автоморфизм, т.е. если найдутся многочлены g_1, \dots, g_n такие, что

$$X_1 = g_1(X'_1, \dots, X'_n), \quad \dots, \quad X_n = g_n(X'_1, \dots, X'_n),$$

то, как легко убедиться, якобиан обратим. **Верно ли обратное утверждение?** Другими словами, **верна ли импликация**

$$\det J(F) \in \mathbb{C}^* \implies F \text{ — автоморфизм?} \tag{*}$$

Это и есть *проблема якобиана*, сформулированная О. Келлером в 1939 г. и остающаяся нерешённой при всех $n \geq 2$.

Две специальные группы автоморфизмов алгебры $\mathbb{C}[X_1, \dots, X_n]$ видны сразу. Это группа GL_n всех линейных невырожденных преобразований и группа B_n , элементы которой получаются композицией "треугольных" полиномиальных отображений вида

$$X_i \mapsto X_i + t_i(X_{i+1}, \dots, X_n), \quad 1 \leq i \leq n, \quad t_i \in \mathbb{C}[X_1, \dots, X_n].$$

В том, что эти отображения обратимы, убедиться совсем несложно. Например, при $n = 3$ имеем

$$X'_1 = X_1 + t_1(X_2, X_3), \quad X'_2 = X_2 + t_2(X_3), \quad X'_3 = X_3.$$

Отсюда

$$X_1 = X'_1 - t_1(X'_2 - t_2(X'_3), X'_3),$$

$$X_2 = X'_2 - t_2(X'_3), \quad X_3 = X'_3.$$

Каждый ли автоморфизм, оставляющий скаляры на месте, получается композицией элементов из GL_n и B_n ? Предполагается, что при $n \geq 3$ это не так. Более того, гипотеза Нагаты гласит, что ответ должен быть отрицательным даже применительно к конкретному автоморфизму $F: X_i \mapsto X'_i$ вида

$$X'_1 = X_1 - 2X_2(X_1X_3 + X_2^2) - X_3(X_1X_3 + X_2^2)^2,$$

$$X'_2 = X_2 + X_3(X_1X_3 + X_2^2), \quad X'_3 = X_3.$$

Отметим, что $J(F) = 1$, причём

$$X_1 = X'_1 + 2X'_2(X'_1X'_3 + X'_2^2) - X'_3(X'_1X'_3 + X'_2^2),$$

$$X_2 = X'_2 - X'_3(X'_1X'_3 + X'_2^2), \quad X_3 = X'_3.$$

Существует много разных подходов к проблеме якобиана (алгебро-геометрических и функциональных), комбинирование которых привело к частичному успеху. Положим $\deg F = \max_{1 \leq i \leq n} \deg f_i$. Если $n = 2$ и $\deg F \leq 150$, то ответ на вопрос (*) оказывается положительным. Кроме того, доказано, что с точностью до GL_n достаточно рассматривать многочлены вида

$$f_i = X_i + h_i(X_1, \dots, X_n), \quad 1 \leq i \leq n,$$

где все компоненты h_i — кубические однородные формы ($\deg h_i = 3$), а матрица Якоби $J(H)$ для $H = (h_1, \dots, h_n)$ нильпотентна: $(J(H))^n = 0$. Правда, в процессе редукции число переменных n увеличивается по сравнению с исходным.

Более подробно с имеющимися результатами относительно проблемы якобиана можно познакомиться по обзорной статье: Bass H., Connell E.H., Wright D. Jacobian conjecture// Bull. Amer. Math. Soc. — 1982. — V. 7, № 2. — P. 287–330.

2*. Задача о дискриминанте (Е.А. Горин). Пусть

$$f(X) = X^n + a_2 X^{n-2} + \dots + a_n$$

— нормализованный многочлен, коэффициенты которого a_i , $2 \leq i \leq n$, суть рациональные функции на \mathbb{C} (т.е. рациональные дроби из $\mathbb{C}(z)$). Предположим, что дискриминант $D(f)$ многочлена f тождественно равен 1. Возможно ли при этом, чтобы не все коэффициенты a_i были константами?

Известно следующее.

а) Если особенности (полюса) всех a_i расположены в точках 0 и ∞ (либо знаменатель q несократимой дроби $a_i = p/q$ делится на z , либо $\deg p > \deg q$), то a_i , $2 \leq i \leq n$, — константы.

б) Если $n = 3$ или $n = 4$, то все a_i — константы.

При $n = 3$ дело сводится к рациональным решениям уравнения $u^2 + v^3 = 1$ (см. п. 4 § 2 гл. 6), которое в свою очередь редуцируется к уравнению Ферма степени 3. Но известно, что если $f^n + g^n = h^n$, где $f, g, h \in \mathbb{C}[z]$ и $\text{НОД}(f, g, h) = 1$, то при $n > 2$ многочлены f, g, h являются на самом деле константами. При $n = 4$ используется кубическая резольвента Феррари. Случай $n = 5$ остаётся без ответа.

3. Задача о двух порождающих кольцах многочленов. Известная из математической литературы теорема Абъянкара—Моха утверждает, что если $\mathbb{C}[f(z), g(z)] = \mathbb{C}[z]$, т.е. многочлены f, g порождают всё кольцо многочленов, то:

i) пара f, g разделяет \mathbb{C} , т.е. $z_1 \neq z_2 \implies f(z_1) \neq f(z_2)$ или $g(z_1) \neq g(z_2)$;

ii) производные $f'(z), g'(z)$ не имеют общих нулей.

Для этой теоремы пока нет простых доказательств. Требуется найти элементарные подходы.

Решающее утверждение в одном из вариантов доказательства: если для пары f, g выполнены свойства i), ii), то либо $\deg f \mid \deg g$, либо $\deg g \mid \deg f$.

Теорема Зайденберга—Лина (ДАН СССР. — 1983. — Т. 271, № 5) устанавливает в свою очередь, что если пара $f, g \in \mathbb{C}[z]$ разделяет \mathbb{C} , то

$$\mathbb{C}[f(z), g(z)] = \mathbb{C}[(z - c)^k, (z - c)^l],$$

где $c \in \mathbb{C}$, а $\text{НОД}(k, l) = 1$. В частности, система $f'(z) = 0, g'(z) = 0$ имеет не более одного решения. Это утверждение сильнее, чем теорема Абъянкара—Моха. К теореме Зайденберга—Лина также требуется найти элементарный подход.

Следует отметить, что замена \mathbb{C} на \mathbb{R} невозможна. По-видимому, нет аналогов указанных теорем, где вместо пары f, g рассматривались бы тройки f, g, h и т.д.

4*. Задачи о критических точках и критических значениях (B. Sendov, S. Smale, A.I. Кострикин, Э.Б. Винберг). Пусть $f(z)$ — комплексный многочлен, $\Theta_f = \{\theta \in \mathbb{C} \mid f'(\theta) = 0\}$ — множество его критических точек. Значения $f(\theta)$, отвечающие критическим точкам $\theta \in \Theta_f$, иногда называют критическими.

a) Доказать, что если все нули (корни) многочлена $f(z) = \prod_{k=1}^n (z - c_k)$ степени $n \geq 2$ лежат в единичном круге $D_1 = \{z \in \mathbb{C} \mid |z| \leq 1\}$, то для каждого c_k круг $\{z \in \mathbb{C} \mid |z - c_k| \leq 1\}$ содержит по крайней мере одну критическую точку.

С результатами в этом направлении можно познакомиться по статье: Miller M.J. // Trans. AMS. — 1990. — V. 321, № 1. — P. 285–303.

Общая алгебро-геометрическая концепция, относящаяся к роли критических точек многочленов, изложена в книге: Marden M. Geometry of polynomials. — Providence, R.I.: AMS, 1966.

Проблема Сендова уточняет один известный результат из этой книги, согласно которому любой круг D_r , содержащий все нули многочлена $f(z)$, содержит также все нули его производной. В это очень легко поверить, предположив, что все нули $f(z)$ вещественные, и просмотрев ещё раз доказательство теоремы 3 из § 4 гл. 6.

б) Доказать, что если f — комплексный многочлен степени n такой, что $f(0) = 0, f'(0) \neq 0$, то

$$\min_{\theta \in \Theta_f} \left| \frac{f(\theta)}{\theta} \right| \frac{1}{|f'(0)|} \leq \frac{n-1}{n}.$$

Смейл доказал существование критической точки θ , для которой

$$\frac{|f(\theta)|}{|\theta f'(0)|} \leq 4,$$

т.е. дана требуемая оценка с худшей константой. Для многочленов степени $n \leq 4$ проблема решена. Константа $(n-1)/n$ неулучшаема, как показывает пример многочлена $f(z) = z^n - \frac{n-1}{n}z$.

в) Обозначим символом C_n множество так называемых консервативных многочленов $f(z) = z^n + a_1 z^{n-1} + \dots + a_{n-1} z$, определяемых свойством: $f'(\theta) = 0 \implies f(\theta) = \theta$. Таким образом, консервативный многочлен, рассматриваемый как отображение $f: \mathbb{C} \rightarrow \mathbb{C}$, оставляет начало координат и все свои критические точки на месте.

Известно (Tischler D. // Complexity. — 1989), что $|C_n| = \binom{2n-2}{n-1}$, поэтому для консервативных многочленов степени n проблема Смейла, сводящаяся к доказательству неравенства $|f'(0)| \geq \frac{n-1}{n}$, решается в принципе конечным перебором. Но для этого нужно располагать явным описанием многочленов из C_n , что известно пока лишь при $n \leq 6$. Более слабое неравенство $|f'(0)| > 1$ справедливо для любого $f \in C_n$.

Было бы крайне интересно исследовать геометрию нулей и неподвижных точек консервативных многочленов.

г) Пусть f — многочлен степени n с вещественными коэффициентами и с вещественными критическими точками ($\Theta_f \subset \mathbb{R}$). Множество всех таких многочленов обозначим через \mathcal{P}_n . Каждому $f \in \mathcal{P}_n$ сопоставим вектор $\operatorname{crf} \in \mathbb{R}^{n-1}$ по следующему правилу. Пусть $\theta_1, \dots, \theta_{n-1}$ — критические точки для f (с учётом кратностей), расположенные в порядке неубывания. Тогда $\operatorname{crf} = (f(\theta_1), \dots, f(\theta_{n-1}))$ — вектор критических значений. Очевидно, что вектор $(c_1, \dots, c_{n-1}) \in \mathbb{R}^{n-1}$ может иметь вид crf для некоторого $f \in \mathcal{P}^n$ только в том случае, когда $(-1)^k(c_k - c_{k+1}) \geq 0$ для всех k или, наоборот, $(-1)^k(c_k - c_{k+1}) \leq 0$ для всех k , т.е. последовательность c_1, \dots, c_{n-1} «пилообразна».

Довольно сложным образом доказано, что всякому пилообразному вектору $c = (c_1, \dots, c_{n-1})$ отвечает многочлен $f \in \mathcal{P}_n$ такой, что $\operatorname{crf} = c$. Многочлен f определён однозначно с точностью до линейной замены аргумента $x \mapsto ax + b$, где $a, b \in \mathbb{R}$, $a > 0$.

Существует ли элементарное доказательство этого факта?

5. Задача о глобальной сходимости метода Ньютона (Smale S. // Bull. Amer. Math. Soc. — 1985. — V. 13, № 2). Будем понимать под римановой сферой S множество \mathbb{C} комплексных чисел с присоединённым символом ∞ , а под рациональным эндоморфизмом S в себя — отображение $z \mapsto P(z)/Q(z)$, где P и Q — многочлены.

Рациональный эндоморфизм Ньютона $N_f : S \rightarrow S$, ассоциированный с комплексным многочленом $f(z)$ степени n , определяется известной нам из п. 7 § 4 гл. 6 формулой (12):

$$N_f(z) = z - \frac{f(z)}{f'(z)}.$$

Число $\zeta \in \mathbb{C} \subset S$ является неподвижной точкой для N_f (т.е. $N_f(\zeta) = \zeta$) в точности тогда, когда ζ — нуль многочлена f . В этом случае, кстати, значением производной N'_f будет $N'_f(\zeta) = (n-1)/n$ — число, встретившееся нам в п. 4, б).

Метод Ньютона вычисления корня (нуля) многочлена f может рассматриваться как итерация отображения N_f : $N_f^0(\zeta_0) = \zeta_0$, $\zeta_m = N_f(\zeta_{m-1}) = N_f^m(\zeta_0)$. В математической литературе наших дней пару (N_f, S) считают динамической системой и применяют к её изучению хорошо развитую технику.

Как уже отмечалось, $|N'_f(\zeta)| < 1$, и, значит, существует окрестность U точки ζ такая, что $\lim_{m \rightarrow \infty} N_f^m(z) = \zeta$ для любой точки $z \in U$. При этом ζ называется стоком (или притягивающей неподвижной точкой для N_f), а открытое множество $B = \bigcup_{m \geq 0} N_f^{-m}(U)$ — бассейном стока ζ . Точка $\alpha \in \mathbb{C}$ называется стоком периода k для

N_f , если $N_f^k(\alpha) = \alpha$ и $|N_f^k'(\alpha)| < 1$. При $k > 1$ точки $\alpha, N_f(\alpha), \dots, N_f^{k-1}(\alpha)$ различны, причём в некоторой окрестности $U \ni \alpha$ итерации $N_f^i(z)$ для $z \in U$ будут приводить к асимптотическим циклам вокруг $\alpha, N_f(\alpha), \dots, N_f^{k-1}(\alpha)$, не давая в результате неподвижных точек. А это значит, что если N_f обладает стоком периода $k \geq 2$, то о глобальной сходимости N_f "для почти всех $z \in \mathbb{C}$ " говорить не приходится. Построить многочлен f любой степени $n \geq 3$ с периодическим стоком для N_f довольно легко. Например, при $n = 3$ таковым является многочлен $f_0(z) = \frac{1}{2}z^3 - z + 1$. Многочлены, достаточно близкие к f_0 , будут также обладать этим свойством.

При $n = 2$ ситуация совершенно иная. Именно, пусть многочлен $f(z) = z^2 + az + b$ обладает двумя различными корнями ζ, η , и пусть L — прямая, перпендикулярная отрезку $[\zeta, \eta]$ и проходящая через его середину. Тогда непосредственно проверяется (хорошее упражнение), что для любой точки $z \in \mathbb{C} \setminus L$, т.е. почти всюду, последовательность $\{N_f^m(z)\}$ сходится к ζ или η . Пусть, например, $f(z) = z^2 - d$, $d \in \mathbb{R}$, $d > 0$. В этом случае $L = i\mathbb{R}$ — мнимая ось, проходящая через середину отрезка $[-\sqrt{d}, \sqrt{d}]$. Так как $N_f(t) \in \mathbb{R}$ для любого $t \in \mathbb{R}$ и $N_f(it) = it + \frac{t^2 + d}{2it} = i\left(t - \frac{t^2 + d}{2t}\right) \in \mathbb{R}$, то $N_f^m(it) \in i\mathbb{R}$. Значит, начав с чисто мнимого числа $z = it$, мы никогда не сойдём с прямой L и, естественно, не получим методом Ньютона приближения к \sqrt{d} или $-\sqrt{d}$. Наоборот, единственное условие $\operatorname{Re} z \neq 0$ гарантирует такое приближение, хотя, возможно, и не очень быстрое.

Ещё одно замечание. Преобразованием $z \mapsto az$ с достаточно большим по модулю a многочлен f_0 степени n переводится в многочлен $f_1(z) = b_0 z^n + \dots + b_n$ такой, что $|b_0| \geq |b_k|$ для всех k . Далее, нули и критические точки многочленов f_1 и λf_1 для $\lambda \in \mathbb{C}^*$ одни и те же. Аналогично, $N_{\lambda f_1} = N_{f_1}$, а это значит, что при обсуждении метода Ньютона f можно брать из множества $\mathcal{P}_n(1)$ нормализованных многочленов, коэффициенты которых по модулю не превосходят 1. По лемме 3 § 3 гл. 6 все нули многочлена $f \in \mathcal{P}_n(1)$ лежат в круге $D_2 = \{z \in \mathbb{C} \mid |z| \leq 2\}$. Пусть

$$B_f = \{z \in \mathbb{C} \mid \lim_{m \rightarrow \infty} N_f^m(z) = \zeta(z), f(\zeta(z)) = 0\}$$

— объединение бассейнов всех стоков для N_f . Попросту, B_f — область сходимости метода Ньютона в применении к f .

Специфика множества $\mathcal{P}_n(1)$ даёт нам возможность ограничиться рассмотрением пересечения $B_f \cap D_2$, площадь которого обозначается $v(B_f \cap D_2)$. Тогда отношение

$$A_f = \frac{v(B_f \cap D_2)}{v(D_2)} = \frac{v(B_f \cap D_2)}{4\pi}$$

можно интерпретировать как вероятность того, что метод Ньютона будет сходиться для случайно выбранной точки из D_2 . Мы вплотную подошли к формулировке главной задачи.

Положим

$$A_n = \min_{f \in \mathcal{P}_n(1)} A_f.$$

Разумеется, $0 \leq A_n \leq 1$. Согласно замечаниям, сделанным выше, $A_2 = 1$, в то время как $A_n < 1$ при $n \geq 3$.

Требуется доказать, что при любом n имеет место неравенство $A_n > 0$. Хорошо бы также оценить A_n как функцию от n .

Пока не доказано даже, что $A_3 > 0$.

* * *

*“Ещё многое имею сказать вам,
но вы теперь не можете вместить”.*

Еванг. от Иоанна, 16:12

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

Абелева группа 140
Автоморфизм группы 146
— поля 60
Аддитивная группа кольца 152
Аксиомы векторного (линейного)
пространства 66
— группы 139
— теории определителей 130
Алгебра как наука 11, 15
— матриц 86
Алгебраическая структура (система) 134
Алгебраически замкнутое поле 232
Алгебраический элемент 184
Алгебраическое дополнение 110, 129
Алгоритм деления 63, 187
— Евклида 194
Аргумент комплексного числа 170
Ассоциативная бинарная операция 135
Ассоциативное кольцо 152
Ассоциированная однородная система 20
Ассоциированные элементы 61
Аффинные преобразования вещественной прямой 150

Базис векторного пространства строк 69
— индукции 46
Базисные столбцы 76
Биективное (взаимно однозначное)
отображение 36, 38
Бинарная алгебраическая операция 134

Бинарное отношение 41
Биномиальная формула 48
Биномиальный коэффициент 48
Блоки (клетки) матрицы 100

Вектор критических значений 263
Векторное пространство строк (столбцов) 66
Векторы—строки 66
Вероятностный вектор—столбец 98
Вероятность 265
Верхняя треугольная матрица 25
Вес многочлена 221
Вещественная ось 169
Взаимная (присоединенная) матрица 121
Взаимно простые числа 63
— — элементы целостного кольца 193, 195
Взаимный многочлен 257
Включение множеств 34
Вырожденная матрица 87
Высший член многочлена 222

Гипотеза Нагаты 260
Главная диагональ квадратной матрицы 20
Главные неизвестные 24, 76
Гомоморфизм групп 147
— кольц 156
График функции 41
Группа 139
— внутренних автоморфизмов 147
— симметрий правильного многоугольника 142

- Двойная сумма 73
 Двойное отношение четверок чисел 177
 Декартова степень 35
 Декартово произведение 35
 Делимость в целостных кольцах 190, 192
 Делитель единицы в кольце 158
 — нуля в кольце 158
 — целого числа 61
 Диагональная матрица 20
 Динамическая система 263
 Дискриминант алгебраического уравнения 227
 — многочлена 227
 — семейства элементов 227
 Дистрибутивные законы кольца 83
 Дифференцирование кольца 213
 Дополнение к подмножеству 35
 Дробь 159, 203
 — несократимая 203
- Евклидово кольцо 194
 Единичная матрица 20
 Единичное (тождественное) отображение 36
 Единичный элемент 135
 Естественное отображение 42
- Закон сокращения в целостном кольце 158
 Знак (сигнатура, четность) перестановки 61
 Знакопеременная группа 148
- Изоморфизм групп 144
 — колец 157
 Инверсия относительно перестановки 61
 Интерполяционная формула Лагранжа 211
 — — Ньютона 211
- Каноническая проекция 42
- Квадратичная (сверхэкспоненциальная) сходимость 255
 Квадратичное поле 176
 Квадратная матрица данного порядка 20, 86
 Класс эквивалентности 41
 Классы вычетов 155
 Код, исправляющий ошибки 18, 165
 Кольцо 152
 — классов вычетов 155
 — многочленов 182
 — с делением 159
 — функций 153
 — целых чисел 153
 Коммутативная диаграмма 37
 Коммутативное кольцо 152
 Комплексное сопряжение 170
 Комплексные числа 169
 Композиция отображений 37
 Компонента решения 21
 Консервативный многочлен 262
 Конструктивные числа и поля 178, 179
 Координаты вектора-строки 70
 Корень из единицы 174
 Корень (нуль) многочлена 208
 Кососимметрическая функция 58
 Кососимметрический многочлен 232
 — определитель 116
 Коэффициенты многочлена 183
 Кратные множители многочлена 215
 Кратный корень 209
 Критерий невырожденности матрицы 87, 122
 — Рауса—Гурвица 252
 — совместности линейной системы 77
 — Эйзенштейна неприводимости многочлена 200
 Критические значения 262
 Критические (экстремальные) точки 241, 262
- Левонормированное произведение 137
 Лексикографическое упорядочение одночленов 221

- Лемма Гаусса 198
 Лемма Даламбера—Аргана 235
 — Коши о минимуме 234
 Линейная зависимость векторов-строк 68
 — оболочка 67
 — однородная система 20
 — система 19
 — функция 80
 Линейно зависимая (независимая) система 68
 — упорядоченное множество 44
 Линейное отображение 79
 Линейные комбинации 67
 — уравнения 19
 Локализация корней многочлена 244
- Максимальный элемент 44
 Марковская (стохастическая) матрица 98
 Матрица линейного отображения 79
 Матричное кольцо 86
 Матричные единицы 91
 Метод Гаусса 26
 — испытаний 254
 — линейной интерполяции 254
 — неопределенных коэффициентов 225, 242
 — Ньютона 255
 — окаймляющих миноров 126
 — последовательного исключения неизвестных 26
 — Штрассена 28
 Минимальный элемент 45
 Минор матрицы 110
 Минимая ось 169
 — часть комплексного числа 169
 Многочлен (полином) 182
 — Лежандра 258
 Множество 33
 Модуль комплексного числа 170
 — сравнения 155
 Моногенный симметрический многочлен 224
 Моноид 135
 — преобразований 136
 Мономорфизм 157
- Монотонный одночлен 222
 Морфизм 148
 Мощность множества 35, 40
 Мультиликативная группа поля 159
 — полугруппа кольца 152
- Наибольший общий делитель 63, 192
 — элемент 44
 Наименьшее общее кратное 63, 193
 Наименьший элемент 45
 Невырожденная матрица 87
 Независимые (непересекающиеся) циклы 54
 Нейтральный элемент 139
 Неопределенная линейная система 21
 Неподвижная притягивающая точка 263
 — точка 263
 Неприводимый многочлен 190, 197
 Неравенство Коши—Буняковского—Шварца 176
 Несобственный делитель 61
 Нечетная перестановка 57
 Нижняя треугольная матрица 25
 Норма числа в квадратичном поле 176
 Нормализованный (нормированный, унитарный) многочлен 188
 Нормальная фундаментальная система решений 198
- Область значений отображения 35
 — определения отображения 35
 Обобщенная ассоциативность 136
 Обратимая матрица 87
 Обратимый элемент кольца 158
 — — моноида 138
 Обратная матрица 87
 Обратное отображение 38
 Общий закон дистрибутивности 154
 Объединение множеств 34
 Ограниченнная стрелка 36
 Однородный многочлен 187
 Одночлен (моноид) 187
 Окаймляющий минор 126

- Оператор дифференцирования 213
 — частного дифференцирования 219
- Определенная линейная система 21
- Определитель Вандермонда 115
 — матрицы 29, 104
 — произведения матриц 118
 — с углом нулей 117
- Орбита 53
- Ориентированный объем 103
- Основная теорема алгебры 233
 — арифметики 62
- Основные свойства определителя 107, 112
- Отделение корней (нулей) многочлена 244
- Отношение эквивалентности 41
- Отображение 35
 — биективное 36
 — инъективное 36
 — сюръективное 36
- Пересечение множеств 34
- Перестановка 51
- Плоскость комплексных чисел 169
- Подгруппа 140
- Подкольцо 152
- Подмножество 34
 — замкнутое относительно операций 136
- Подполе 160
- Подстановка в многочлен 184
- Поле 159
 — Галуа 163
 — комплексных чисел 169
 — отношений (частных, дробей) 202
- Поле разложения многочлена 237
 — рациональных дробей 203
- Полилинейная функция 105
- Полиномиальная функция 210
- Полная линейная группа 139, 140
 — степень многочлена 186
- Полное матричное кольцо 153
- Положительная марковская матрица 99
- Полугруппа 135
- Порядок на множестве 44
 — перестановки 53
 — степенного ряда 189
- Порядок элемента 142
- Постоянная функция 153
- Постоянное отображение 38
- Поточечная операция 153
- Правило знаков Декарта 249
- Правильная дробь 204
- Представитель класса 41
- Приведенная линейная система 20
- Признак делимости 194
- Примарная дробь 206
- Примитивный (первообразный) корень из единицы 174
 — многочлен 198
- Принцип двойной индукции 50
 — индукции 46
- Присоединенная (взаимная) матрица 121
- Проблема якобиана 260
- Произведение матриц 82
- Произведение (композиция) отображений 37
- Производная многочлена 212
- Производящая функция 189
- Простейшая дробь 204
- Простое поле 162
- Простой корень 209
 — элемент кольца 190
- Пространство решений 96
 — столбцов (строк) прямоугольной матрицы 74
- Прямоугольная матрица 20
- Пустое множество 34
- Разложение определителя 113
- Размерность линейной оболочки 71
- Разность множеств 34
- Ранг матрицы 75
 — по столбцам (по строкам) 74
 — произведения матриц 84
 — системы векторов 71
- Расширение поля 160
- Расширенная матрица линейной системы 21
- Рациональная функция 212
- Редуцированный многочлен 210, 218
- Результант 229

- Рефлексивность 41
 Решение линейной системы 21
- Свободные неизвестные 24
 Свободный член уравнения 20
 Символ Кронекера 86
 Симметрическая группа 52
 — разность множеств 40
 — функция 217
 Симметрический многочлен 220
 Симметричность 41
 Система линейных уравнений 19
 — (ряд) Штурма 245
 Скалярная матрица 20, 86
 Собственное подмножество 34
 Совместная линейная система 21,
 77
 Содержание многочлена 198
 Специальная линейная группа 140
 Сравнение 155
 Срезания экспонента 248
 Стандартная система Штурма 247
 Стандартный базис 70
 Старший коэффициент 186
 Степенная сумма 224
 Степень многочлена 286
 — элемента 137
 Столбец матрицы 20
 — определителя 105
 Стока матрицы 20
 — определителя 105
 Ступенчатый вид линейной системы 24
 — — матрицы 24
 Суперпозиция отображений 37, 153
 Схема Горнера 208
 Сюръективное отображение 36
- Таблица Кэли 144
 Тело 159
 Теорема Безу 208
 — Бюдана—Фурье 251
 — Вильсона 218
 — Кронекера—Капелли 77
 — Кэли 146
 — Раше 253
 — Ферма (малая) 161
- Теорема Шевалье 218
 — Штейница 233
 — Штурма 245
 Тождество Эйлера 219
 Транзитивность 41
 Транспозиция 55
 Транспонированная матрица 83
 Трансцендентный элемент 184
 Треугольная диаграмма 37
 — матрица 25
 Тригонометрическая форма комплексного числа 171
- Универсальное свойство кольца многочленов 184
 Унимодулярная группа 140
 Упорядочение множества 44
 Устойчивый многочлен 251
- Факториальное кольцо 190, 197
 Факторизация отображений 43
 Фактормножество 42
 Форма 187
 Формальный степенной ряд 188
 Формула (интерполяционная)
 Лагранжа 211
 Формула Лагранжа 243
 — Лейбница 213
 — Муавра 173
 — (интерполяционная) Ньютона
 211
 — “полного развертывания” 104
 — Стирлинга 60
 — Тейлора 251, 257
 — Эйлера 173
 Формулы Виета 217
 — Крамера 123
 — Ньютона 225
 Фундаментальная система решений 98
 Функция 35
 — Эйлера 64
- Характеристика поля 161

- Целая рациональная функция 210
Целостное кольцо 158
Цепь 44
Цикл 53
Циклическая группа 142, 145
- Частичный порядок 44
Частное 159
— дифференцирование 219
— от деления 63, 188
Четверная группа 151
Четная перестановка 57
Число перемен знаков 245
— Ферма 47
— Фибоначчи 27
Чисто мнимое число 169
- Эквивалентность линейных систем 22
Эквивалентные матрицы 92
Элемент кольца нильпотентный 165
— множества 33
Элементарные матрицы 91
— преобразования 21, 74
— симметрические многочлены 221
Эндоморфизм алгебры многочленов 259
Эпиморфизм 148, 157
- Ядро гомоморфизма 147, 156
— линейного отображения 97
Якобиан 259

Учебное пособие

КОСТРИКИН Алексей Иванович

ВВЕДЕНИЕ В АЛГЕБРУ

Часть I

ОСНОВЫ АЛГЕБРЫ

**Редактор Е.Ю. Ходак
Оригинал-макет Н.Н. Андреева**

ЛР № 071930 от 06.07.99. Подписано в печать 05.12.99.
Формат 60x90, 16. Бумага офсетная № 1. Печать офсетная.
Усл. печ. л. 17. Уч.-изд. л. 18,7. Тираж 3000 экз. Заказ № 36

Издательская фирма "Физико-математическая литература"
МАИК "Наука /Интерperiодика"
117864 Москва, ул. Профсоюзная, 90



Отпечатано с оригинал-макета в Чебоксарской типографии № 1
428019, г. Чебоксары, пр. И. Яковleva, 15