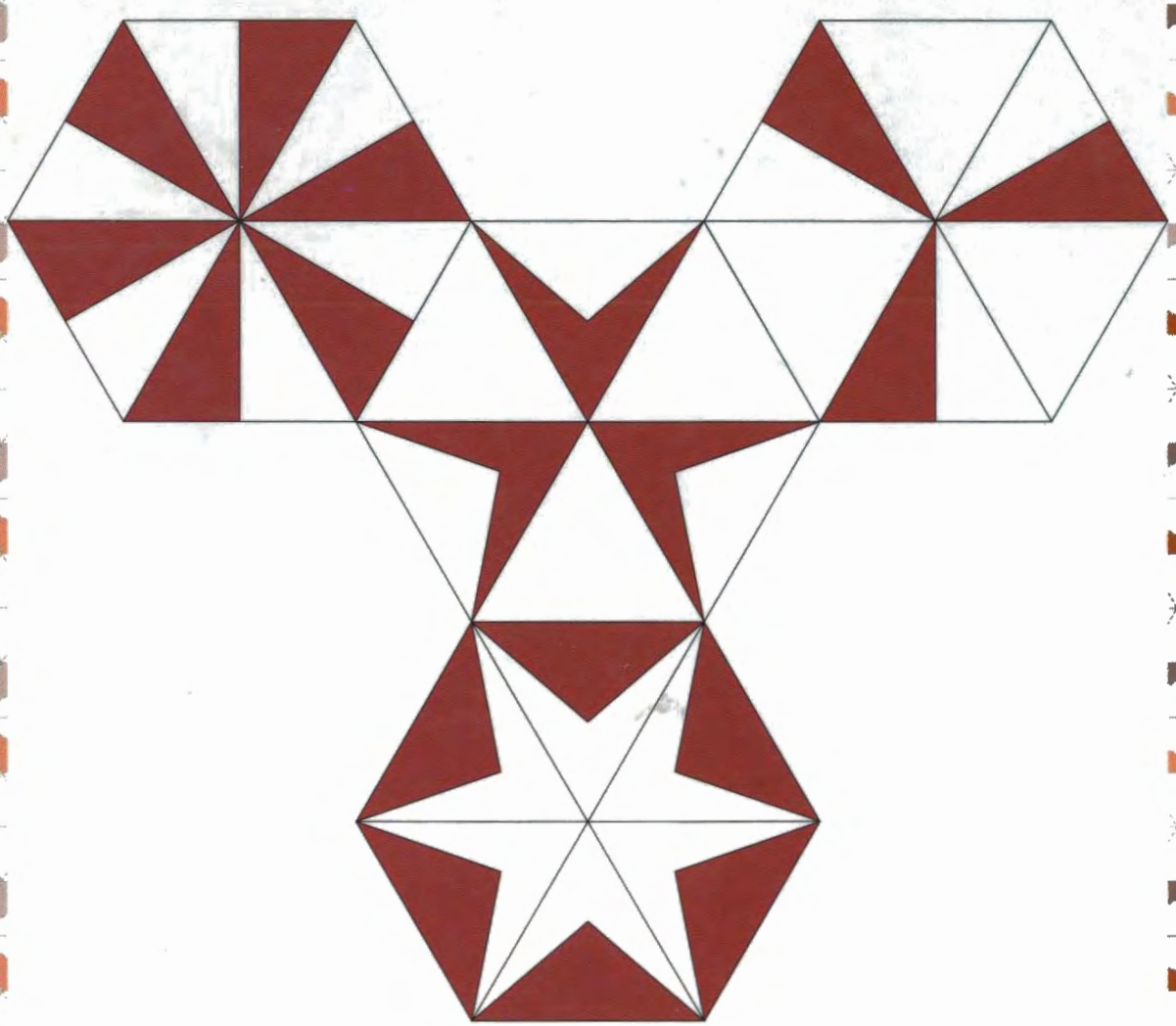
ГРАНИ АЛГЕБРЫ

М. Н. Аршинов Л. Е. Садовский



Факториал

М. Н. Аршинов
Л. Е. Садовский

ГРАНИ АЛГЕБРЫ

Под редакцией
Ю. В. Кузьмина

Москва
ФАКТОРИАЛ ПРЕСС
2008

УДК 512
ББК 22.14
А 89

Аршинов М. Н.

А 89 Грани алгебры / М. Н. Аршинов, Л. Е. Садовский; Под ред. Ю. В. Кузьмина. — М.: Факториал Пресс, 2008. — 328 с.

ISBN 978-5-88688-091-5

Книга является сборником этюдов на различные темы из алгебры и её приложений. Большое внимание уделяется истории и разъяснению мотивировок. Хотя степень трудности глав книги различна, она рассчитана на самый широкий круг читателей: учащихся математических школ, студентов ВУЗов специальности «Прикладная математика», студентов-математиков педагогических ВУЗов и университетов.

УДК 512
ББК 22.14

Научное издание

Михаил Наумович Аршинов
Леонид Ефимович Садовский

ГРАНИ АЛГЕБРЫ

Под редакцией *Ю. В. Кузьмина*

Формат 60 × 90/16. Усл. печ. л. 20. Бумага офсетная №1. Гарнитура литературная. Подписано к печати 18.02.2008. Тираж 1200 экз. Заказ 261

Издательство «Факториал Пресс», 117449, Москва, а/я 331; ЛР ИД № 00316 от 22.10.99. e-mail: press@factorialco.com

Отпечатано с готовых диапозитивов издательства «Факториал Пресс» в ППП типографии «Наука» Академиздатцентра «Наука» РАН. 121099, Москва Г-99, Шубинский пер., 6.

ISBN 978-5-88688-091-5

ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ РЕДАКТОРА

«Грани алгебры» — книга довольно необычная. Это не учебник, это скорее сборник этюдов на темы из алгебры. Стиль изложения доверительный, слышится живая интонация лектора. Всё время подчёркивается историческая перспектива, сообщаются сведения о знаменитых математиках. Однако книгу вряд ли можно назвать научно-популярной. В ней развивается достаточно серьёзный математический аппарат. Полные доказательства приводятся не всегда, но существо дела разъясняется подробно.

Для кого написана эта книга? Честно говоря, и мне — профессиональному алгебраисту — было полезно прочесть эту книгу и некоторые грани алгебры приобрели для меня дополнительные оттенки. Думаю, что и гордый мехматянин найдёт здесь для себя много интересного. Во многих технических ВУЗах открыта специальность «Прикладная математика», где предусмотрено изучение элементов алгебры. Алгебра необходима и для студентов модной теперь специальности «Компьютерная безопасность». Обычно у преподавателя не остаётся времени объяснить, как возникли те или иные алгебраические понятия, какую роль они сыграли в развитии математики. Книга М. Н. Аршинова и Л. Е. Садовского поможет студентам разобраться в этих вопросах. Она также полезна и доступна студентам-математикам педагогических ВУЗов. Часть материала доступна и школьникам старших классов, тем более учащимся математических школ. Приходит на ум такой образ: накрыт праздничный стол, а каждый выберет себе то, что ему по зубам и по вкусу. Мне, скажем, понравился этюд об алгебре в комбинаторике.

Эта книга была написана в середине восьмидесятых минувшего века, но не была издана по причинам нематематического

характера. Одна из причин — смутное время начала девяностых. Авторов уже нет в живых, так что издание является посмертным. Леонид Ефимович Садовский долгое время заведовал кафедрой «Прикладная математика» МИИТа — нынешний Университет путей сообщения. Он был одним из её создателей и строителей. Михаил Наумович Аршинов — доцент кафедры, один из её ведущих преподавателей. Книгу «Грани алгебры» мы обнаружили, разбирая архив Аршинова. Совместными усилиями был сделан её электронный вариант. Кое-что пришлось изменить и сократить. Был пополнен список литературы. Фактическим соредактором стал Ю. С. Семёнов. Компьютерную версию рисунков сделала А. П. Иванова.

Мы искренне надеемся, что книга найдёт своего читателя.

Профессор кафедры «Прикладная математика»
Университета путей сообщения
Ю. В. Кузьмин

ВВЕДЕНИЕ

Всякому, кто взял в руки эту книгу, слово «алгебра» хорошо известно. Это и понятно, ведь алгебра наряду с геометрией — один из двух столпов школьного курса математики. Однако далеко не все знают, сколь велика дистанция от школьной алгебры до той, которая составляет предмет научных интересов современных математиков-алгебраистов. Эта незнакомая читателю обширная область знаний обладает всеми чертами абстрактной математической науки, а тому, кто впервые её изучает, порой нелегко бывает дойти до сути, смысла и происхождения многих её понятий и фактов. Не легче понять и то, в чём их значимость для других разделов математики и какова их практическая ценность.

В то же время без ответа на подобные вопросы едва ли возможно зримо представить себе истинное лицо алгебры. Поэтому авторы стремились показать историю возникновения и развития основных алгебраических идей, неразрывно связанную с именами математиков разных времён и народов. Другая, не менее важная цель книги — рассказать о различных применениях алгебры как внутри самой математики, так и вне её.

Пожалуй, не стоит в нашем коротком введении пытаться сколько-нибудь определённо характеризовать современную алгебру. Будет, вероятно, лучше, если суждение по этому поводу сформируется у читателя постепенно, по мере знакомства с разными гранями непростой, но увлекательной науки, именуемой «Алгебра».

ГЛАВА 1

ПУНКТ ОТПРАВЛЕНИЯ — ВАВИЛОН

В древней математике явно преобладала геометрия. Даже самого слова «алгебра» не было — оно появилось гораздо позднее. Однако задачи, по смыслу алгебраические, хотя и облечённые в геометрическую форму, всё же существовали. Так, в древнем Вавилоне, в XVIII веке до нашей эры (до расцвета античной математики ещё должно было пройти без малого 15 столетий) умели решать задачи, сводящиеся к квадратным уравнениям и даже системам квадратных уравнений. Свидетельством тому — дошедшие до нас из эпохи вавилонской династии царя Хаммурапи многочисленные клинописные таблички. Когда удалось расшифровать их содержимое, перед исследователями открылись любопытнейшие тексты, дающие представление о математических знаниях далёких предков. В этих текстах нет и следа какой бы то ни было алгебраической символики, с помощью которой мы записываем уравнения. Нет, в обычном смысле, и самих уравнений. Таблички содержат лишь словесные формулировки задач и рецепты их решения. Прочтём одну из табличек (шестидесятиричная система счисления заменена на привычную нам десятичную).

«То, на сколько длина превышает ширину, возведи в квадрат, от площади отними, и 500 это есть. Длину и ширину сложи, и 50 это есть. Спрашивает длина и ширина. Твой способ: 50 возведи в квадрат и 2500 это даёт. Прибавь 2500 к 500 и 3000 это даёт. Пятую часть отломай и 600 это даёт. Половину от 50 отломай и 25 это даёт. Имеет 25 своим квадратным корнем 5. К 25 прибавь 5 и 30 длину это даёт. От 25 отними 5 и 20 ширину это даёт.»

В первом абзаце текста формулируется задача, которую мы можем свести к системе двух уравнений с двумя неизвестными. Если x — длина, y — ширина, то система такова

$$\begin{cases} xy - (x-y)^2 = 500 \\ x + y = 50. \end{cases}$$

В этих обозначениях решение вавилонского математика сводится к следующему. Второе уравнение возводится в квадрат

$$(x+y)^2 = 2500,$$

и полученное уравнение складывается с первым из уравнений системы. В итоге получается

$$5xy = 3000, \quad \text{или} \quad xy = 600.$$

Исходная система приводится к равносильному виду

$$x + y = 50, \quad xy = 600.$$

Решение последней получается при помощи правила, которое встречается и во многих других клинописных текстах. В современной алгебраической символике оно выглядит так. Если $x = 25 + a$, то $y = 25 - a$. Тогда $xy = (25+a)(25-a) = 600$, откуда

$$a = \sqrt{625 - 600} = 5, \quad x = 25 + 5 = 30, \quad y = 25 - 5 = 20.$$

Правила, которым следовали вавилонские вычислители в этом и других клинописных текстах, удивительным образом соответствуют хорошо известным нам формулам. Может даже показаться, что известны они были и составителям табличек, но это, конечно, иллюзия. Скорее всего, при поиске своих правил вавилоняне исходили из геометрических соображений. Ван-дер-Варден так пишет об этом в своей книге «Пробуждение науки». «Вавилоняне обычно рассматривали произведение как площадь прямоугольника, квадрат числа — как площадь квадрата, что подтверждается их собственной терминологией. «Однако, — добавляет Ван-дер-Варден, — вавилоняне мыслили прежде всего алгебраически. Хотя они изображали неизвестные числа линиями

и площадями, но последние всегда оставались числами... Даже в задачах геометрического содержания вопрос, который ставили вавилоняне, всегда сводился к вычислению чего-нибудь и никогда к построению или доказательству. Сквозь геометрическую внешность просвечивала алгебраическая суть».

Иное дело греки. Многие свои математические знания они заимствовали из Вавилона, но дальнейшее развитие у них пошло по геометрическому пути. Соотношения между числами они всегда представляли геометрически. Именно в геометрии античные мыслители достигли своего идеала математической строгости, впервые в истории человечества построив теории, основанные на точных, логически выверенных доказательствах. Тогда же возникли и первые трудные математические проблемы. Некоторые из них стали поистине легендарными. Таковы проблемы трисекции угла, квадратуры круга, удвоения куба, построения правильных многоугольников циркулем и линейкой и т. д.

Приведём образец «геометрической алгебры» древних греков ([28], §2). Есть основания считать, что её элементы были известны уже в Вавилоне. Вот как формулирует Евклид в своих «Началах» теорему, равносильную тождеству

$$(a+b)^2 = a^2 + 2ab + b^2. \qquad (1.1)$$

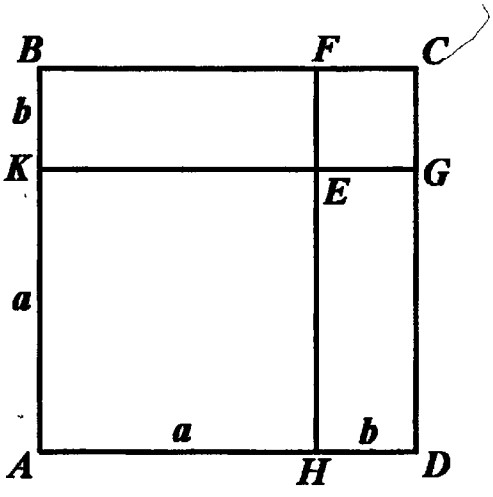


Рис. 1.

«Если прямая линия как-либо рассечена, то квадрат на всей прямой равен квадратам на отрезках вместе с дважды взятым прямоугольником, заключённым между отрезками.»

Суть рассуждений сводится к следующему (рис. 1).

Площадь $(a+b)^2$ квадрата $ABCD$ складывается из площадей a^2 и b^2 квадратов $AKEH$, $EFCG$ и суммы $2ab$ двух площадей двух прямоугольников $KBFE$ и $HEGD$.

Шли века, математика развивалась, решая всё новые задачи и осваивая всё новые области. Старые задачи, казалось, отсту-

пали на второй план. Однако лучшие умы вновь и вновь обращались к проблемам античных времён, пытаясь вскрыть их неразгаданную тайну. Ключом к ответу часто была алгебра.

Термин «алгебра» имеет восточное происхождение. В IX веке н. э. жил аль-Хорезми (родом из Хорезма — древнего района в Средней Азии). История не сохранила биографических сведений об аль-Хорезми. Принято считать годом его рождения 783-й, а годом смерти — 850-й. Он был автором нескольких математических трактатов. Одно из своих сочинений, посвящённое методам решения уравнений, он назвал «Хитаб ал-джебр ал-мукабала», что переводится как «Исчисление восполнения и противопоставления». Арабское «ал-джебр» в латинской транскрипции превратилось в «algebra», и начиная с XII века это слово приобрело права гражданства в Европе. Обозначать оно стало учение о решении уравнений.

Настоящей сенсацией в алгебре стало открытие в начале XVI века (эпоха Возрождения коснулась и математики) формул для решения уравнений третьей, а затем и четвёртой степени. Это событие, окружённое множеством легенд, так или иначе связано с именами итальянских математиков Сципиона дель Ферро (1465–1526), Николо Тарталья (1500–1557), Джероламо Кардано (1501–1576), Луиджи Феррари (1522–1565)[1]. Способы решения уравнений 3-й и 4-й степени впервые были опубликованы в сочинении Дж. Кардано «Ars magna» («Великое искусство»), а доказательство их, по сохранившейся ещё традиции, носило геометрический характер. Формулу для решения кубического уравнения называют формулой Кардано. В несколько модернизированном виде её вывод выглядит так.

Прежде всего, решение полного кубического уравнения

$$y^3 + ay^2 + by + c = 0$$

сводится к решению кубического уравнения вида

$$x^3 + px + q = 0, \tag{1.2}$$

[1] Об этих учёных и любопытных подробностях их открытия читатель может прочесть, например, в книге [10].

не содержащего квадрата неизвестного. Для этого, как может убедиться читатель, достаточно сделать подстановку $y = x - a/3$.

Следующая идея — искать решение уравнения (1.2) в виде $x = u + v$, где u, v — вспомогательные неизвестные. После подстановки в (1.2) будем иметь

$$(u+v)^3 + p(u+v) + q = 0,$$

или

$$u^3 + v^3 + (3uv + p)(u+v) + q = 0. \tag{1.3}$$

Наличие двух неизвестных вместо одного отнюдь не усложняет задачу, напротив, позволяет распорядиться ими наиболее удобным способом. В самом деле, выберем их так, чтобы $3uv + p = 0$, или $uv = -p/3$. Тогда уравнение (1.3) принимает вид $u^3 + v^3 = -q$ и мы приходим к системе

$$u^3 + v^3 = -q, \quad u^3 v^3 = -\frac{p^3}{27}.$$

Согласно теореме Виета числа u^3, v^3 являются корнями квадратного уравнения

$$z^2 + qz - \frac{p^3}{27} = 0.$$

Решая последнее уравнение, находим

$$u^3 = -\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}, \quad v^3 = -\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}},$$

или наоборот, что в данном случае не играет роли. Наконец, мы приходим к формуле Кардано

$$x = \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}. \tag{1.4}$$

Открытие этой формулы явилось первым крупным достижением европейской математики, и Кардано не жалеет слов, чтобы подчеркнуть его значимость. «Так как это искусство превосходит всю человеческую ловкость и всю ясность ума смертного», —

восклицает он, — «то его нужно рассматривать как подарок небесного происхождения... и это настолько славное открытие, что от того, кто мог его достигнуть, можно ожидать, что он достигнет всего.»

Быть может, приведённое высказывание и покажется нам сильным преувеличением, всё же для истории алгебры, её последующего развития значимость формулы Кардано, как и формулы Феррари для уравнения четвёртой степени, действительно велика. Правда, того же нельзя сказать об их практическом значении. Уже первооткрывателями было понято, что при пользовании формулами возникают не только технические, но и принципиальные трудности. О природе этих трудностей рассказывается в дополнении.

Достижения итальянских учёных произвели громадное впечатление на современников и вдохновили математиков на поиски аналогичных формул для решения алгебраических уравнений пятой, шестой, вообще произвольных степеней, то есть уравнений вида

$$a_0x^n + a_1x^{n-1} + \ldots + a_{n-1}x + a_n = 0. \qquad (1.5)$$

Исследование таких уравнений, не только связанное с поиском формулы корней, оказывается вплоть до XIX века в центре внимания алгебры, и в этот период всё более утверждается взгляд на эту науку как на науку об алгебраических уравнениях. Одновременно появляется и ещё одна точка зрения на алгебру, отражающая не столько предмет, сколько складывающийся её общий дух, как учения о формальных действиях, о так называемых алгебраических операциях.

Старая традиция — записывать уравнения и правила их решения словами — давно уже была большой обузой, а при рассмотрении уравнений произвольных степеней стала просто непреодолимым препятствием. Необходимо было создать какую-то новую систему кратких обозначений. Такие попытки делались в разные времена, однако самый решительный шаг в этом направлении сделал французский математик Франсуа Виет (1540–1603) — не зря его называют творцом математической формулы. Но и буквенное исчисление Виета ещё нуждалось в усовершенствова-

нии. Более компактный и, по существу, современный вид придал ему один из создателей новой математики француз Рене Декарт (1596–1660). Лишь после Декарта стали в ходу у математиков выражения, подобные, например, (1.4) и (1.5), а алгебра приняла форму науки о буквенных вычислениях, о преобразовании формул и уравнений, составленных из букв (в отличие от арифметики, оперирующей с конкретными числами).

С созданием удобной алгебраической символики неизмеримо упростились математические рассуждения. То, что прежде было доступно лишь избранным, теперь зачастую выполнялось механически, с помощью цепочки преобразований, проводимых по определённым правилам. Вот что писал об этом сам Виет: «Все математики знали, что под их алгеброй и алмукабалой были скрыты несравненные сокровища, но не умели их найти. Задачи, которые они считали наиболее трудными, совсем легко решаются десятками с помощью нашего искусства, представляющего поэтому самый верный путь для математических изысканий.»

Как бы в подтверждение этих слов, буквенная алгебра Виета и Декарта вскоре активно вторглась в геометрию. Созданные Декартом и его современником Ферма (1601–1665) метод координат и аналитическая геометрия позволили действительно «решать десятками» ранее считавшиеся очень трудными геометрические задачи. В этом методе, как вероятно известно читателю, точки задаются координатами, линии и поверхности — уравнениями, а геометрические рассмотрения заменяются алгебраическими вычислениями с уравнениями и их системами. Однако главные события в алгебре были ещё впереди. Начало им положили великие математики Жозеф Луи Лагранж (1736–1813) и Карл Фридрих Гаусс (1777–1855).

Лагранж был уже зрелым математиком, когда его заинтересовала задача о разрешимости в радикалах алгебраических уравнений. Вместо того, чтобы искать решение вслепую, как это делали очень многие до него, Лагранж попытался проанализировать причины предшествующих неудач. Ему удалось открыть глубокие принципы, лежащие в основе проблемы, и сдвинуть её с мёртвой точки. Вскоре Гаусс, используя сходные идеи, сумел дать алгебраическое решение задачи о построении с помощью циркуля

и линейки правильных многоугольников. Это была его первая, но поистине замечательная математическая работа. Исследования Лагранжа и Гаусса были подхвачены молодыми математиками — норвежцем Нильсом Генрихом Абелем (1802–1829) и французом Эваристом Галуа (1811–1832). Оказалось, что формул в радикалах для решения общего алгебраического уравнения степени большей четырёх просто не существует. Это доказал Абель. Галуа подвёл черту: он выяснил, каким необходимым и достаточным условиям должно удовлетворять каждое конкретное алгебраическое уравнение для того, чтобы его корни выражались в радикалах через коэффициенты этого уравнения.

В период, о котором мы говорим, не просто одна за другой решались казавшиеся неприступными задачи. Одновременно создавался совершенно особый круг понятий и идей, который затем полностью преобразил лицо алгебры. О том, как это происходило, мы и расскажем в последующих главах. Но сначала мы познакомим читателя с несколькими алгебраическими персонажами, которым отведена важная роль в этой книге.

Дополнения

1. Предлагаем читателю ещё один пример клинописных текстов.

«Длина, ширина. Длину и ширину я перемножил и площадь получил. Затем избыток длины над шириной я прибавил к площади: 183 получилось у меня. Затем я длину и ширину сложил: 27. Спрашивает длина, ширина и площадь.

Ты сделаешь так:

$$27 + 183 = 210, \quad 2 + 27 = 29.$$

Возьми половину от 29, это даёт 14,5;

$$14{,}5 \times 14{,}5 = 210{,}25, \quad 210{,}25 - 210 = 0{,}25;$$

0,25 имеет 0,5 квадратный корень,

$$14{,}5 + 0{,}5 = 15 \text{ — длина} \quad 14{,}5 - 0{,}5 = 14 \text{ — ширина.}$$

Отними 2, которые ты прибавил к 27 от ширины, 12 истинная ширина; 15 длину, и 12 ширину я перемножил: 15×12=180 — площадь...»

Задача сходна с приведённой в основном тексте. Читатель без труда разберётся в методе её решения (см. также [28], § 1 б).

2. А вот как греки доказывали тождество

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

Левая его часть выражает площадь прямоугольника $AEFG$, которая, понятно, равна разности площадей квадратов $ABCD$ и $CHFK$ (рис. 2).

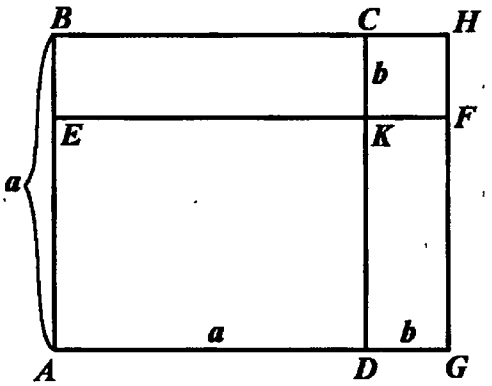


Рис. 2.

Нетрудно указать и геометрическую интерпретацию формулы

$$(a+b)^3=a^3+3a^2b+3ab^2+b^3. \quad (1.6)$$

На языке геометрической алгебры решались и квадратные уравнения. Конечно, возможности такой алгебры ограничены. К примеру, совершенно вне её компетенции находится доказательство при $n>3$ формулы бинома Ньютона

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

обобщающей (1.1) и (1.6).

3. Трудности использования формулы Кардано, о которой упоминалось, проистекают не только от её внешней громоздкости. Чтобы это понять, рассмотрим уравнение

$$x^3-2x+1=0, \quad (1.7)$$

один из корней которого, а именно 1, очевиден. Левую часть можно разложить на множители

$$x^3-2x+1=(x-1)(x^2+x-1),$$

поэтому корнями уравнения (1.7) будут, кроме того, корни квадратного трёхчлена $x^2 + x - 1$, то есть числа $(-1 \pm \sqrt{5})/2$. Итак, наше кубическое уравнение имеет три действительных корня. Ну а что же даёт формула Кардано? Нечто странное. Действительно, выражение, стоящее под знаком квадратного радикала,

$$\frac{q^2}{4} + \frac{p^3}{27} = \frac{1}{4} - \frac{8}{27} = -\frac{5}{108}$$

получается отрицательным, и следовательно, никак нельзя по нашей формуле получить перечисленные корни, если оперировать только с обычными действительными числами. Так впервые в математике возникла необходимость расширения числовых множеств за пределы действительных чисел. С новыми, так называемыми комплексными числами мы познакомимся в следующей главе. Не понятно также и то, как по формуле Кардано может получиться три различных корня, и этот вопрос также прояснится лишь позднее.

Здесь же отметим ещё, что даже в том случае, когда

$$\frac{q^2}{4} + \frac{p^3}{27} > 0,$$

решение может получиться в «нераспознаваемом» виде. Так, легко проверить, что уравнение $x^3 + 3x - 4 = 0$ имеет единственный действительный корень, равный 1. Формула же Кардано даёт для него значение

$$\sqrt[3]{2 + \sqrt{5}} + \sqrt[3]{2 - \sqrt{5}}.$$

Нелегко понять, что это и есть 1.

4. Несколько слов об эволюции алгебраических обозначений. Вот пример своеобразной символики, встречающейся у Кардано ([28], §6б). Выражение

$$\sqrt[3]{\sqrt{5} + 2} - \sqrt[3]{\sqrt{5} - 2},$$

являющееся корнем уравнения $x^3 + 3x = 4$, он записал бы примерно так

$$R_x u.cu.R_x 5p.2 \,|\, m.R_x u.cu.R_x 5m.2.$$

Здесь R_x — знак корня (Radix), $R_x u.cu$ (Radix universalis cubica) означает корень кубический из всего выражения до вертикальной черты, p — плюс, m — минус. Формул, записанных в общем виде, у Кардано не было, так как буквенных обозначений для произвольных числовых данных он не применял. Впервые такие обозначения появились у Виета. В его записи уравнение $A^3 + 3B^2A = 2Z^3$ (A обозначает неизвестную) выглядит так:

$$A\,cubus + B\,plano\,3\,in\,A\,aequari\,Z\,cubus\,2,$$

а формулу Кардано для этого уравнения он записывает в виде

$$\sqrt{C\sqrt{B\,plano - plano - plani + Z\,solido - solido} + Z\,solido} - \sqrt{C\sqrt{B\,plano - plano - plani + Z\,solido - solido} - Z\,solido}.$$

У Декарта запись почти современная. Вид формулы Кардано для уравнения $x^3 + px + q = 0$ в обозначениях Декарта мало отличается от (1.4):

$$\sqrt{C - \frac{1}{2}q + \sqrt{\frac{1}{4}qq + \frac{1}{27}p^3}} + \sqrt{C - \frac{1}{2}q - \sqrt{\frac{1}{4}qq + \frac{1}{27}p^3}}.$$

5. Как узнать, можно ли осуществить то или иное геометрическое построение, пользуясь классическим набором инструментов — циркулем и линейкой? Методы аналитической геометрии позволяют получить простой алгебраический критерий возможности такого построения.

Заметим, что всякое построение циркулем и линейкой сводится к ряду шагов, на каждом из которых отыскивается точка пересечения либо двух прямых, либо прямой и окружности, либо двух окружностей. Фиксируем на плоскости систему координат. Прямая задаётся линейным уравнением

$$Ax + By + C = 0,$$

а окружность радиуса r с центром в точке (a, b) — уравнением вида

$$(x - a)^2 + (y - b)^2 = r^2.$$

Отыскание точки пересечения прямых равносильно решению системы уравнений

$$A_1x + B_1y + C_1 = 0, \quad A_2x + B_2y + C_2 = 0,$$

задающих эти прямые, а после исключения неизвестного — решению линейного уравнения с одним неизвестным.

Точка пересечения прямой и окружности может быть найдена как решение системы

$$Ax + By + C = 0, \quad (x-a)^2 + (y-b)^2 = r^2. \tag{1.8}$$

Соответственно, точка пересечения двух окружностей — это решение системы

$$(x-a_1)^2 + (y-b_1)^2 = r_1^2, \quad (x-a_2)^2 + (y-b_2)^2 = r_2^2. \tag{1.9}$$

Система (1.8) после исключения неизвестного из первого уравнения приводится к решению квадратного уравнения, а система (1.9) после вычитания одного уравнения из другого приводится к виду (1.8). Итак, каждый этап задачи, решаемый циркулем и линейкой, сводится к уравнению первой или второй степени с одним неизвестным, а все величины, дающие решение задачи, выражаются через исходные данные с помощью арифметических операций и операций извлечения квадратного корня. Наоборот, если это так, то задача может быть решена циркулем и линейкой.

Из сказанного вытекает, что при заданной единице длины циркулем и линейкой могут быть построены те и только те отрезки, длины которых выражаются через единицу с помощью конечного числа операций сложения, вычитания, умножения, деления и извлечения квадратного корня. Такие числа называют *квадратичными иррациональностями.*

6. Одна из древних задач, в которой речь идёт о построениях с помощью циркуля и линейки, формулируется так. Построить куб, имеющий объём, вдвое больший, чем данный куб. Её называют задачей об удвоении куба. Легко сформулировать алгебраический вариант. Если x — отношение сторон искомого и заданного куба, то $x^3 = 2$. Если бы было возможно построение циркулем

и линейкой, то это, согласно предыдущему пункту, означало бы, что $\sqrt[3]{2}$ является квадратичной иррациональностью. То, что это неверно может показаться очевидным. Однако внешняя простота здесь обманчива. Доказать, что $\sqrt[3]{2}$ нельзя выразить через квадратные радикалы, далеко не просто. Строгое доказательство этого факта, а значит и неразрешимости задачи удвоения куба, было дано лишь в XIX веке.

Другая задача — разделить данный угол циркулем и линейкой на три равные части (трисекция угла). Она также сводится к аналогичному алгебраическому вопросу. Выяснить это можно, используя тригонометрию. Опишем дугу единичной окружности с центром в вершине заданного угла α (рис. 3).

Для построения угла $\alpha/3$ достаточно построить отрезок AE, длины $|AE| = \cos(\alpha/3)$. При этом длина отрезка AC, построение которого очевидно, равна $\cos\alpha$. Полагая в известном тригонометрическом тождестве

$$\cos\alpha = 4\cos^3(\alpha/3) - 3\cos(\alpha/3)$$

$\cos(\alpha/3) = x$, $\cos\alpha = a$, получаем, что длина x неизвестного отрезка AE является корнем кубического уравнения

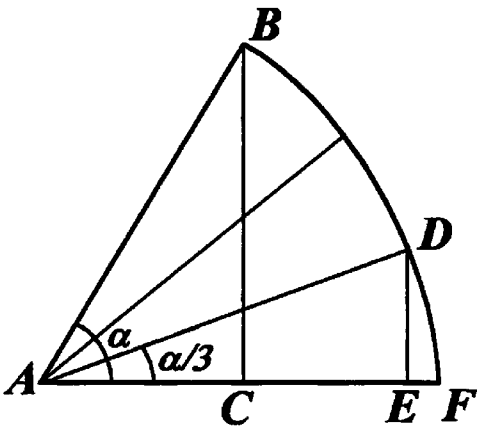


Рис. 3.

$$4x^3 - 3x - a = 0. \qquad (1.10)$$

Итак, геометрическая задача на построение вновь сведена к задаче алгебраической. Надо выяснить, будет ли квадратичной иррациональностью корень уравнения (1.10). Ответ и на этот вопрос отрицательный, откуда вытекает невозможность трисекции угла. Вопросы подобного рода имеют прямое отношение к теории Галуа (о ней в главе 14).

ГЛАВА 2

НЕ ТОЛЬКО ЧИСЛА

Классы вычетов

Из письма П. Ферма: «Всякое простое число непременно измеряет[1] одну из степеней минус 1 любой прогрессии, которая пусть дана; и показатель названной степени является делителем данного простого числа минус 1; и после того, как нашли первую степень, удовлетворяющую искомому, все те степени, показатели которых кратны показателю первой, также удовлетворяют искомому. Пример: пусть дана прогрессия

1	2	3	4	5	6
3	9	27	81	243	729

со своими показателями степени наверху. Возьмите, например, число 13. Оно измеряет третью степень минус 1 [2], показатель которой 3 является делителем числа 12, на единицу меньшего числа 13, и так как показатель степени 729, равный 6, кратен первому показателю, равному 3, то следует, что 13 измеряет также 729 – 1. Это предложение верно для всех прогрессий и всех простых чисел, чему я Вам прислал бы доказательство, если бы не опасался быть слишком многословным.»

В форме, не очень привычной для современного читателя, здесь высказана важная теорема из элементарной теории чисел,

[1] Является делителем.

[2] То есть число $27 - 1 = 26$.

известная как малая теорема Ферма. Мы бы её сформулировали так:

если p — простое число и a взаимно просто с p, то найдётся такой делитель k числа $p-1$, что a^k-1 делится на p, причём для всякого l, кратного числу k, a^l-1 также делится на p.

Ферма иллюстрирует эту теорему примером, в котором $a=3$, $p=13$, и тогда l может принимать значения 3, 6, 12, ... Из сформулированного выше утверждения легко следует более популярная формулировка малой теоремы: *если p — простое число и a взаимно просто с p, то $a^{p-1}-1$ делится на p.*

Ферма не оставил нам доказательства своей теоремы, как впрочем, и многих других высказанных им теоретико-числовых утверждений. Большинство из них, в том числе и малая теорема, были доказаны Леонардом Эйлером (1707–1783), для которого занятия теорией чисел таили в себе такую же неотразимую привлекательность, как и для его предшественника.

Рассуждения Эйлера (по-видимому, таков же был ход мыслей и у Ферма) основывались на рассмотрении остатков, или иначе вычетов, образующихся при делении на p числа a и его степеней, и в обнаружении строгой периодичности в появлении этих остатков. Так, в упомянутом выше примере Ферма остатки при делении на 13 чисел 1, 3, 9, 27, 81, 243, 729, 2178, 6561, ... появляются в следующей последовательности: 1, 3, 9, 1, 3, 9, 1, 3, 9, ... Использование такого приёма, основанного на работе с вычетами, вообще характерно при изучении вопросов делимости чисел и различных числовых арифметических выражений. Наиболее стройную и законченную форму придал этому приёму Гаусс, введя понятие классов вычетов и сравнимости.

Два целых числа a и b называют *сравнимыми по модулю n* (n — натуральное число), если их разность $a-b$ делится на n без остатка. Это выражают следующей записью:

$$a \equiv b \mod n. \tag{2.1}$$

Число n называют модулем сравнения (2.1). Например, $37 \equiv 4 \mod 11$, так как разность $37-4=33$ делится на 11. Аналогично, $50 \equiv -13 \mod 9$, так как $50-(-13)=63$ делится на 9. Запись

$a \equiv 0 \mod n$ означает, что само число a делится на n, то есть $a = kn$.

Фиксируем некоторый модуль сравнения n. Всякое натуральное число c можно единственным образом представить в виде

$$c = sn + r, \tag{2.2}$$

где s — неполное частное от деления на n, а r — остаток (вычет), совпадающий с одним из чисел

$$0, 1, 2, \ldots, n-1. \tag{2.3}$$

Легко видеть, что запись вида (2.2), где $0 \leqslant r \leqslant n-1$, допускают не только натуральные числа, но и любые целые числа. Например, если $c = -35$, $n = 3$, то равенство $-35 = (-3) \cdot 13 + 4$ и есть указанное представление.

Очевидно, из равенства (2.2) следует, что $c \equiv r \mod n$, то есть всякое число сравнимо со своим вычетом по модулю n. Пусть теперь a и b — два произвольных числа, записанные в виде $a = s_1 n + r_1$, $b = s_2 n + r_2$. Каждый из остатков r_1 и r_2 — это одно из чисел (2.3), и потому их разность может делиться на n лишь в одном случае, когда $r_1 = r_2$. Но тогда разность $a - b = (s_1 - s_2)n + (r_1 - r_2)$ может делиться на n в том и только том случае, когда $r_1 = r_2$. Отсюда вывод: *$a \equiv b \mod n$ тогда и только тогда, когда числа a и b имеют одинаковые остатки при делении на n.*

Гаусс систематизировал основные свойства сравнений, во многом аналогичные свойствам обычных равенств. Подобно тому, как это делается с равенствами, сравнения по одинаковому модулю оказалось возможным складывать, перемножать и т. д. Так, перемножив сравнения $17 \equiv 5 \mod 4$ и $7 \equiv 3 \mod 4$, получим, как нетрудно убедиться, верное сравнение $119 \equiv 15 \mod 4$. Вообще, если $a_1 \equiv b_1$, $a_2 \equiv b_2$, то

$$a_1 + a_2 \equiv b_1 + b_2, \quad a_1 a_2 \equiv b_1 b_2. \tag{2.4}$$

(здесь сравнения рассматриваются по одному модулю n, который в записи опущен). Доказательство сводится к проверке определения. Чтобы убедится в справедливости, например, второго

из сравнений (2.4), рассмотрим разность $a_1a_2 - b_1b_2 = a_1a_2 - a_2b_1 + a_2b_1 - b_1b_2 = a_2(a_1 - b_1) + b_1(a_2 - b_2)$. В преобразованном выражении оба слагаемых делятся на модуль без остатка, так как по условию $a_1 \equiv b_1$, $a_2 \equiv b_2$, следовательно, этим свойством обладает и вся сумма. Но это и означает, что $a_1a_2 \equiv b_1b_2$.

Как применяются отмеченные свойства, наглядно иллюстрирует следующая задача. Выяснить, при каких натуральных значениях k число $2005^k + 2006^k$ делится на 3. Замечаем, что $2005 \equiv 1 \mod 3$, $2006 \equiv 2 \mod 3$. Но тогда согласно (2.4) $2005^k \equiv 1$ и $2006^k \equiv 2^k$. Заменяя в исходном выражении числа 2005 и 2006 их вычетами по модулю 3, получим

$$2005^k + 2006^k \equiv 1 + 2^k \mod 3,$$

следовательно, левая часть сравнения делится на 3 тогда и только тогда, когда на 3 делится сумма $1 + 2^k$. Для степеней же двойки имеем: $2^2 \equiv 1$, $2^3 \equiv 2$, $2^4 \equiv 1$, $2^5 \equiv 2$ и т. д. Вообще, применяя индукцию по k убеждаемся, что $2^k \equiv 1 \mod 3$ при чётном k и $2^k \equiv 2 \mod 3$ при k нечётном. Таким образом, в первом случае $1 + 2^k \equiv 1 + 1 \equiv 2 \mod 3$, а во втором $1 + 2^k \equiv 1 + 2 \equiv 0 \mod 3$, и поэтому исходное выражение делится на 3 тогда и только тогда, когда k нечётно.

В разобранной задаче числа 2005 и 2006 можно заменить любыми целыми числами a и b, дающими при делении на 3 остатки соответственно 1 и 2. Ни ответ задачи, ни способ решения от этого не изменились бы. Таким образом, в некоторых вопросах все числа, имеющие один и тот же вычет r по модулю n и, следовательно, сравнимые между собой по этому модулю, оказываются взаимозаменяемыми. Объединим их все в один класс и обозначим через $\overline{r}$:

$$\overline{r} = \{c \mid c \equiv r \mod n\}. \tag{2.5}$$

Иными словами, класс $\overline{r}$ состоит из всех тех чисел, которые записываются в виде (2.2). Класс, определяемый равенством (2.5), называют классом вычетов. Каждому из возможных остатков $0, 1, 2, \ldots, n-1$ отвечает свой класс вычетов, так что имеется ровно n различных классов

$$\overline{0}, \overline{1}, \overline{2}, \ldots, \overline{n-1}. \tag{2.6}$$

Ясно, что эти классы попарно не пересекаются и каждое целое число попадает ровно в один из перечисленных классов. И вот что важно. Оказывается с классами вычетов можно оперировать, как и с числами — их можно складывать и перемножать, используя для этой цели операции сложения и умножения чисел.

Разберёмся с этим подробнее. Рассмотрим какие-нибудь два класса $\overline{r}_1$ и $\overline{r}_2$ и выберем в каждом из них любые числа $a_1 \in \overline{r}_1$, $a_2 \in \overline{r}_2$. Пусть оказалось, что их сумма $a_1 + a_2$ имеет вычет r, а произведение $a_1 a_2$ — вычет s, иными словами,

$$a_1 + a_2 \in \overline{r}, \quad a_1 a_2 \in \overline{s}.$$

Тогда будем считать, что сумма классов $\overline{r}_1$ и $\overline{r}_2$ равна $\overline{r}$, а их произведение равно $\overline{s}$

$$\overline{r}_1 + \overline{r}_2 = \overline{r}, \quad \overline{r}_1 \overline{r}_2 = \overline{s}.$$

Всё здесь довольно необычно. Необычна сама идея оперирования над числовыми классами — слишком велика сила привычки к тому, что арифметические операции являются монополией чисел. Наверное, необычен также и характер наших определений суммы и произведения классов вычетов. Необычно ещё и то, что в этих определениях имеется элемент произвола. Действительно, для того, чтобы найти сумму и произведение классов вычетов, мы сначала оперируем с числами — представителями этих классов, а их, как было сказано, мы вольны выбирать как угодно. Не будет ли тогда сумма и произведение $\overline{r}$ и $\overline{s}$ зависеть от произвола в выборе представителей? Если бы это было так, то операции над классами были бы лишены определённого смысла. К счастью, наши опасения напрасны. В самом деле, если b_1 и b_2 — другие представители классов $\overline{r}_1$ и $\overline{r}_2$, отличные от a_1 и a_2, то

$$b_1 \equiv a_1 \mod n, \quad b_2 \equiv a_2 \mod n.$$

Но согласно (2.4)

$$b_1 + b_2 \equiv a_1 + a_2 \mod n, \quad b_1 b_2 \equiv a_1 a_2 \mod n,$$

а значит, сумма и произведение новых представителей лежат в тех же классах $\overline{r}$ и $\overline{s}$, что сумма и произведение старых.

Самый простой пример, иллюстрирующий сказанное, получится, если за модуль сравнения взять число $n = 2$. В этом случае имеем два класса вычетов $\overline{0}$ и $\overline{1}$, а операции над ними выглядят так:

$$\overline{0}+\overline{0}=\overline{0}, \quad \overline{0}+\overline{1}=\overline{1}+\overline{0}=\overline{1}, \quad \overline{1}+\overline{1}=\overline{0},$$
$$\overline{0}\cdot\overline{0}=\overline{0}\cdot\overline{1}=\overline{1}\cdot\overline{0}=\overline{0}, \quad \overline{1}\cdot\overline{1}=\overline{1}.$$

А вот каковы таблицы сложения и умножения классов вычетов по модулю 4 (табл. 1):

Таблица 1.

$+$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$
$\overline{0}$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$
$\overline{1}$	$\overline{1}$	$\overline{2}$	$\overline{3}$	$\overline{0}$
$\overline{2}$	$\overline{2}$	$\overline{3}$	$\overline{0}$	$\overline{1}$
$\overline{3}$	$\overline{3}$	$\overline{0}$	$\overline{1}$	$\overline{2}$

$\cdot$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$
$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$
$\overline{1}$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$
$\overline{2}$	$\overline{0}$	$\overline{2}$	$\overline{0}$	$\overline{2}$
$\overline{3}$	$\overline{0}$	$\overline{3}$	$\overline{2}$	$\overline{1}$

Читатель легко может составить аналогичные таблицы сложения и умножения классов по другим модулям (например, для $n = 3, 5$ и т. д.). Рассматривая эти таблицы, можно подметить некоторые новые неожиданные свойства, не свойственные числовым операциям. Всё же нельзя не заметить и того, что операции над классами вычетов наследуют от своих «родителей» — сложения и умножения чисел — немало общих с ними свойств. Так, обе операции над классами коммутативны и ассоциативны, а умножение дистрибутивно относительно сложения:

а) коммутативность

$$\overline{r}_1+\overline{r}_2=\overline{r}_2+\overline{r}_1, \qquad \overline{r}_1\overline{r}_2=\overline{r}_2\overline{r}_1,$$

б) ассоциативность

$$(\overline{r}_1+\overline{r}_2)+\overline{r}_3=\overline{r}_1+(\overline{r}_2+\overline{r}_3), \qquad (\overline{r}_1\overline{r}_2)\overline{r}_3=\overline{r}_1(\overline{r}_2\overline{r}_3),$$

в) дистрибутивность

$$(\overline{r}_1 + \overline{r}_2)\overline{r}_3 = \overline{r}_1\overline{r}_2 + \overline{r}_2\overline{r}_3.$$

Всё это так, поскольку операции над классами вычетов сводятся в конечном счёте к операциям над числами из этих классов, а они, разумеется, обладают перечисленными свойствами. Можно продолжать алгебраические аналогии, заметив, что классы $\overline{0}$ и $\overline{1}$ играют такую же роль, какую выполняют 0 и 1 в сложении и умножении чисел, а именно:

$$\overline{0} + \overline{r} = \overline{r}, \quad \overline{0} \cdot \overline{r} = \overline{0}, \quad \overline{1} \cdot \overline{r} = \overline{r}$$

для любого класса вычетов $\overline{r}$.

Множество классов вычетов с определёнными выше операциями сложения и умножения классов обычно обозначают $\mathbb{Z}_n$ и называют *кольцом вычетов* $\bmod n$ (по модулю n).

К сказанному об алгебраических свойствах операций над классами вычетов можно добавить, что в множестве $\mathbb{Z}_n$ операция сложения всегда обратима. Это означает, что для любых двух классов $\overline{r}$ и $\overline{r}_1$ можно определить их разность $\overline{r}_2 = \overline{r}_1 - \overline{r}$, понимая под этой разностью (как и для чисел), такой элемент $\overline{r}_2 \in \mathbb{Z}_n$, что

$$\overline{r} + \overline{r}_2 = \overline{r}_1.$$

Например, легко видеть, что в $\mathbb{Z}_5$

$$\overline{4} - \overline{2} = \overline{2}, \quad \text{а} \quad \overline{2} - \overline{4} = \overline{3}.$$

Вычитание в $\mathbb{Z}_n$ может быть сведено к сложению, если использовать понятие противоположного элемента. Класс $\overline{s}$ называется противоположным к классу $\overline{r}$, если $\overline{r} + \overline{s} = \overline{0}$. Очевидно, противоположным к классу $\overline{0}$ будет сам этот класс, а для любого ненулевого класса $\overline{r}$ противоположный элемент равен $\overline{n - r}$. Действительно, согласно определению сложения

$$\overline{r} + \overline{n - r} = \overline{0}.$$

Элемент, противоположный к $\overline{r}$, естественно обозначить $-\overline{r}$. Тогда разность можно записать в виде

$$\overline{r}_1 - \overline{r} = \overline{r}_1 + (-\overline{r}) = \overline{r}_1 + \overline{n - r}.$$

Так например, в $\mathbb{Z}_7$

$$\overline{3} - \overline{5} = \overline{3} + (-\overline{5}) = \overline{3} + \overline{2} = \overline{5}.$$

Сложнее в $\mathbb{Z}_n$ обстоит дело с умножением классов, которое не во всех случаях обратимо. Иными словами, уравнение

$$\overline{x} \cdot \overline{r} = \overline{r}' \qquad (2.7)$$

может оказаться неразрешимым в $\mathbb{Z}_n$ даже, если $\overline{r} \neq \overline{0}$. Таблица 1 показывает, например, что в $\mathbb{Z}_4$ не имеет решения уравнение $\overline{x} \cdot \overline{2} = \overline{3}$, то есть не определено отношение классов $\overline{3}$ и $\overline{2}$. Однако справедливо такое важное утверждение:
если n — простое число, то в кольце классов вычетов $\mathbb{Z}_n$ уравнение (2.7)*, в котором $\overline{r} \neq \overline{0}$, всегда имеет решение и притом единственное.*

Для доказательства будем в левую часть уравнения (2.7) подставлять по очереди элементы из $\mathbb{Z}_n$, получая при этом классы

$$\overline{0} \cdot \overline{r}, \overline{1} \cdot \overline{r}, \overline{2} \cdot \overline{r}, \ldots, \overline{k} \cdot \overline{r}, \ldots, \overline{n-1} \cdot \overline{r}. \qquad (2.8)$$

Покажем, что все они различны. Если это не так, то найдутся k и m такие, что $0 \leqslant k < m \leqslant n - 1$ и $\overline{k} \cdot \overline{r} = \overline{m} \cdot \overline{r}$. Последнее равенство перепишем в виде $(\overline{m} - \overline{k})\overline{r} = \overline{0}$. Переходя к числам, приходим к выводу, что произведение $(m - k)r$ делится на простое n без остатка. Но это противоречит тому, что сомножители $m - k$ и r — числа меньшие n, и следовательно, с n взаимно простые. Итак, классы (2.8) исчерпывают все элементы $\mathbb{Z}_n$, и поэтому среди них встретится ровно один раз класс $\overline{r}'$ — правая часть уравнения (2.7). Это и доказывает его однозначную разрешимость.

Часто, когда это не вызывает путаницы, в обозначениях классов вычетов опускают черту, записывая их как обычные натуральные числа. Правила сложения и умножения классов по модулю n сводятся тогда к следующему:

$$r_1 + r_2 = r, \quad r_1 \cdot r_2 = s,$$

где r и s — остатки от деления на n обычных суммы и произведения соответственно.

Эти правила можно понимать и буквально, считая, что они определяют две операции на множестве $\{0, 1, 2, \ldots, n-1\}$ — сложение и умножение по модулю n. Например, по модулю 7

$$4 + 5 = 2, \quad 4 \cdot 5 = 6.$$

Преобразования и подстановки

Они являются следующими персонажами нашей книги, рассказывающей об алгебре. С понятием преобразования произвольного множества читатель, видимо, уже знаком. Всё же напомним, что под *преобразованием* f понимается закон или правило, по которому каждому элементу a данного множества M сопоставляется единственный элемент $f(a)$ того же множества, называемый *образом* элемента a. Хорошо известны такие, например, преобразования как поворот плоскости вокруг некоторой точки на заданный угол α или параллельный перенос на данный вектор (рис. 4).

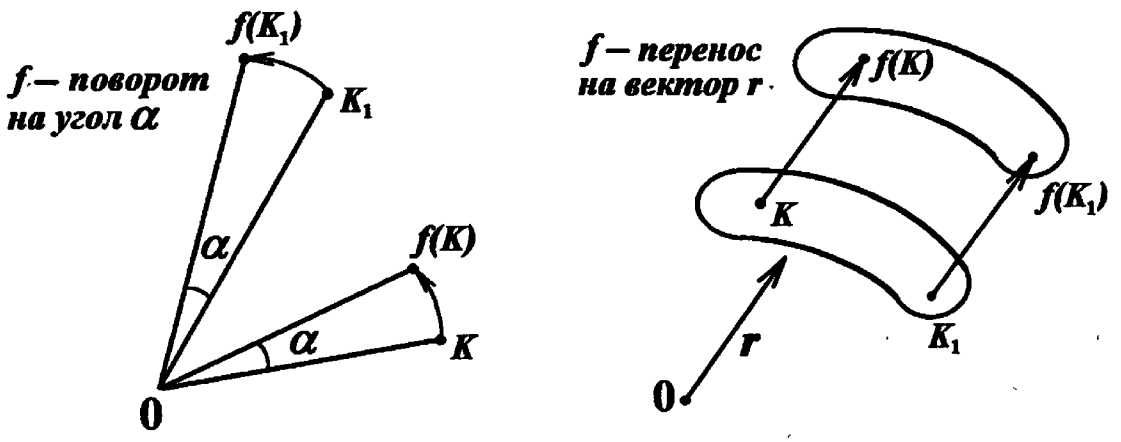
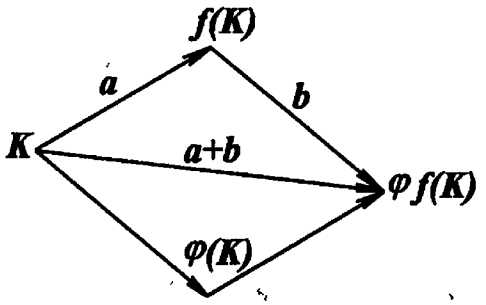


Рис. 4.

Нет недостатка и в других примерах, которые читатель может вспомнить или придумать самостоятельно. Нас будут интересовать прежде всего так называемые взаимно однозначные преобразования. Преобразование множества M называется взаимно однозначным, если каждый элемент этого множества является

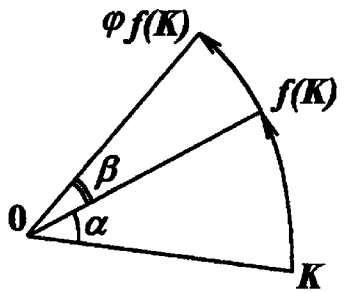
образом некоторого элемента множества M и притом единственного. Упомянутые только что повороты и параллельные переносы плоскости являются как раз примерами таких преобразований.

Если f и φ — два преобразования произвольного множества M, то можно выполнить их одно вслед за другим, например, сначала f, а затем φ. Результатом будет некоторое новое преобразование, которое называется произведением, или композицией преобразований f и φ и обозначается $\varphi \circ f$, или просто φf. Например, если f — параллельный перенос на вектор a, а φ — на вектор b, то их произведение — снова параллельный перенос на вектор, равный сумме $a + b$ (рис. 5 а)). Если же f и φ — повороты на углы α и, соответственно, β вокруг общего центра O, то их композиция, очевидно, будет опять-таки поворотом вокруг того же центра. Углом этого результирующего поворота φf будет, понятно, $\alpha + \beta$ (рис. 5 б)).



φf - композиция двух переносов

а)



φf - композиция двух поворотов

б)

Рис. 5.

Наряду с этими геометрическими примерами особый интерес для нас будут представлять взаимно однозначные преобразования конечного множества, которые принято называть *подстановками*. Если конечное множество M состоит из n элементов, то удобно для их обозначения использовать числа $1, 2, \ldots, n$. Тогда для задания подстановки σ множества M достаточно для каждого из этих чисел (элементов множества M) указать то число, в которое оно отображается этой подстановкой. Можно использовать для этой цели таблицу из двух строк — в первой её строке выписывать элементы $1, 2, \ldots, n$, а во второй под каждым из них

указать соответствующий ему при данной подстановке элемент

$$\begin{pmatrix} 1 & 2 & 3 & \dots & n \\ \sigma(1) & \sigma(2) & \sigma(3) & \dots & \sigma(n) \end{pmatrix}.$$

Например, запись

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$$

означает подстановку множества $M = \{1, 2, 3\}$, которая 1 переводит в $2 = \sigma(1)$, 2 переводит в $3 = \sigma(2)$, а 3 — в $1 = \sigma(3)$. Нетрудно указать все подстановки такого множества, их будет 6:

$$\begin{aligned} &\sigma_1 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}, \\ &\sigma_4 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}, \quad \sigma_5 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}, \quad \sigma_6 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}. \end{aligned} \tag{2.9}$$

Легко понять, как перемножаются эти подстановки. Например,

$$\sigma_3\sigma_4 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} = \sigma_2$$

(σ_4 переводит 1 в 2, после чего σ_3 переводит 2 в 1, так что в результате последовательного выполнения этих двух подстановок 1 перейдёт снова в 1 и т. д.).

Обратим внимание на существенную особенность умножения подстановок и других преобразований — их произведение зависит, вообще говоря, от порядка сомножителей. Так например, выполняя преобразования σ_3 и σ_4 в другом порядке, получим

$$\sigma_4\sigma_3 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}\begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} = \sigma_6.$$

Всё же мы используем термин умножение по отношению к этой не совсем привычной и к тому же некоммутативной операции. Такую традицию можно оправдать тем, что остальные свойства обычного числового умножения сохраняются и для композиции взаимно однозначных преобразований множества M. Прежде

всего, эта операция *ассоциативна*, то есть для любых преобразований f_1, f_2, f_3 множества M

$$f_3(f_2f_1) = (f_3f_2)f_1. \tag{2.10}$$

Действительно, обе части этого равенства совпадают с одним и тем же преобразованием, которое получается при последовательном применении трёх — сначала f_1, затем f_2, и затем f_3.

Среди взаимно однозначных преобразований имеется тождественное e, оставляющее на месте всякий элемент $a \in M$ ($e(a) = a$). Для операции умножения преобразований оно, очевидно, играет роль *единичного элемента*, то есть такого, что

$$ef = fe = f \tag{2.11}$$

для любого f. В приведённом выше списке подстановок (2.9) такую роль играет подстановка σ_1.

И третье, для всякого взаимно однозначного преобразования f множества M существует *обратное* f^{-1}, то есть такое, что

$$ff^{-1} = f^{-1}f = e. \tag{2.12}$$

Например, для подстановки σ_4 обратной будет σ_5, а для σ_2 — сама же эта подстановка.

Совокупность всех взаимно однозначных преобразований множества M с определённой для них операцией умножения принято называть *группой всех преобразований*, а в случае, когда множество M конечно, — группой всех подстановок, или *симметрической группой* множества M. В (2.9) приведён полный список элементов симметрической группы S_3.

Термин «группа подстановок» впервые ввёл в употребление известный уже читателю Э. Галуа, о котором ещё немало будет сказано в этой книге. Галуа использовал этот термин не только для совокупности всех подстановок S_n, но и для подмножеств в S_n, если только для них оказываются выполненными перечисленные выше свойства. Более точно, пусть $G \subseteq S_n$ — такая совокупность подстановок, что для любых $f_1, f_2 \in G$ их композиция f_1f_2 также принадлежат G. Если, кроме того, G содержит

тождественную подстановку и вместе со всякой подстановкой f содержит f^{-1}, то множество G называется *группой подстановок*. Говорят также, что G является *подгруппой* симметрической группы S_n.

Простыми примерами групп подстановок могут служить следующие подмножества в S_4:

$$G_1 = \left\{ e = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{pmatrix}, \sigma_1 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \end{pmatrix}, \right.$$
$$\left. \sigma_2 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \end{pmatrix}, \sigma_3 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 2 & 3 \end{pmatrix} \right\};$$
$$G_2 = \left\{ e = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{pmatrix}, f_1 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{pmatrix}, \right.$$
$$\left. f_2 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \end{pmatrix}, f_3 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{pmatrix} \right\}.$$

Читатель может убедиться в этом непосредственной проверкой. Например, в первом случае

$$\sigma_1\sigma_2 = \sigma_3, \quad \sigma_1^2 = \sigma_1\sigma_1 = \sigma_2, \quad \sigma_1^3 = \sigma_3, \quad \sigma_1 = e$$

и т. д., а во втором

$$f_1 f_2 = f_3, \quad f_1^2 = f_2^2 = f_3^2 = e.$$

Группы подстановок во многих вопросах являются очень полезным инструментом и, что не менее важно, они послужили одним из главных источников понятия группы — одного из важнейших алгебраических понятий. Разговор об этом, и довольно обстоятельный, нам ещё предстоит.

Пока же продолжим знакомство с подстановками. Всякая подстановка $\sigma \in S_n$ может некоторые символы оставлять на месте, другие же действительно перемещать. Первые назовём *неподвижными*, а остальные — *перемещаемыми*. Скажем, для подстановок

$$\sigma_1 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 3 & 2 & 4 & 7 & 5 & 8 & 6 & 1 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 3 & 2 & 4 & 1 & 5 & 8 & 6 & 7 \end{pmatrix}$$

неподвижными являются символы 2 и 5, а все остальные — перемещаемыми. Однако характер перемещения символов этими подстановками различен. Подстановка σ_1 обладает тем свойством, что при повторении её достаточное число раз всякий из перемещаемых символов может быть переведён в любой другой из этих символов. Например,

$$\sigma_1(1) = 3, \qquad \sigma_1^2(1) = \sigma_1(3) = 4, \quad \sigma_1^3(1) = \sigma_1(4) = 7,$$
$$\sigma_1^4(1) = \sigma_1(7) = 6, \quad \sigma_1^5(1) = \sigma_1(6) = 8, \quad \sigma_1^6(1) = \sigma_1(8) = 1.$$

Подстановка, обладающая указанным свойством, называется циклом, или циклической подстановкой. Не такова подстановка σ_2, например,

$$\sigma_2(1) = 3, \quad \sigma_2^2(1) = 4, \quad \sigma_2^3(1) = 1,$$

и поэтому символ 1 не может быть переведён повторным применением σ_2 ни в один из символов 6, 7, 8.

Для циклов, которые в каком-то смысле являются наиболее просто устроенными подстановками, употребляется следующая запись. Перемещаемые символы записываются последовательно в скобках в том порядке, в каком они переходят друг в друга. Запись начинается с любого из перемещаемых символов, а последний символ считается переходящим в первый. Применительно к σ_1 эта запись выглядит следующим образом

$$\sigma_1 = (1, 3, 4, 7, 6, 8).$$

Число символов, перемещаемых циклом, называется его *длиной*, так что длина цикла σ_1 равна 6. Хотя вторая подстановка σ_2 и не является циклом, но можно заметить, что она как бы состоит из двух циклов

$$(1, 3, 4), \quad (7, 6, 8), \tag{2.13}$$

которые действуют на перемещаемые символы независимо один от другого. Конкретнее это выражается в том, что σ_2 является композицией этих циклов, причём порядок их выполнения безразличен

$$\sigma_2 = (1, 3, 4)(7, 6, 8) = (7, 6, 8)(1, 3, 4).$$

Циклы, которые подобно (2.13) не содержат общих перемещаемых элементов, называются *независимыми*, а ситуация, отмеченная нами для подстановки σ_2, носит общий характер:
всякая подстановка представляется в виде произведения коммутирующих между собой независимых циклов. Циклы, входящие в указанное разложение, определяются единственным образом.

Доказательство этого утверждения приводится в дополнении. Читателю же мы рекомендуем найти план такого доказательства экспериментальным путём, разобрав ещё несколько примеров подстановок и их разложений.

Циклы длины 2 называют транспозициями. Другими словами, транспозиция — это подстановка, переставляющая между собой два символа, оставляя прочие неподвижными. Из подстановок, приведённых в (2.9), три являются транспозициями, а именно, $\sigma_2, \sigma_3, \sigma_4$. Утверждение, что всякую подстановку можно получить, выполняя последовательно только транспозиции, может показаться неожиданным. Тем не менее оно справедливо. Например,

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

Вообще,

$$(1,2,3,\ldots,l) = (1,l)(1,l-1)\ldots(1,3)(1,2).$$

Аналогичное разложение можно написать и для произвольного цикла $(i_1, i_2, \ldots, i_l)$. Учитывая теперь, что все подстановки представляются в виде произведения независимых циклов, получим, что любая из них есть произведение транспозиций. Например,

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

Одна и та же подстановка может иметь различные (даже по числу сомножителей) разложения на транспозиции. Например,

$$(1,2,3) = (1,2)(2,3) = (1,3)(1,2).$$

Однако невозможно, чтобы в одно из разложений данной подстановки входило чётное число транспозиций, а в другое — нечётное (доказательство этого — хотя и не очень простая, но в общем доступная задача). По этой причине все подстановки из S_n могут быть разделены на два класса. В один из них входят все *чётные подстановки*, то есть разлагающиеся на чётное число транспозиций, в другой — нечётные. В списке (2.9), исчерпывающем S_3, подстановки $\sigma_1, \sigma_4, \sigma_5$ — чётны, остальные — нечётны.

Множество всех чётных подстановок даёт нам ещё один пример группы. В самом деле, произведение двух чётных подстановок — снова чётная подстановка. Чётной является тождественная подстановка и подстановка обратная к чётной. Эту группу обозначают A_n и называют *знакопеременной группой*. Происхождение терминов «симметрическая» и «знакопеременная» в названиях групп S_n и A_n станет ясным из дальнейшего.

Комплексные числа

Рассматривая классы вычетов и подстановки, мы могли убедиться, что существуют нечисловые объекты, для которых естественно определяются операции, похожие на операции с числами. Можно заподозрить, что в математике имеется немало других примеров такого рода, и это действительно так. Однако новый персонаж, без знакомства с которым никак не обойтись, это всё-таки числа, точнее, так называемые комплексные числа. Впервые они появляются в упоминавшемся уже сочинении «Великое искусство», автор которого Дж. Кардано наделяет их эпитетом «поистине софистические». Позднее (XVII век) Р. Декарт, сетуя на то, что величины эти никак невозможно себе представить, называет их «мнимыми». И в XVIII веке комплексные числа по-прежнему остаются таинственными сущностями. «Это почти амфибии», — говорит о них Лейбниц, — «находящиеся где-то между бытием и небытием». Что же это за «амфибии» и почему они проникли в математику, доставив её служителям немало хлопот?

Математики давно уже столкнулись с тем обстоятельством, что многие алгебраические уравнения оказываются неразрешимыми (не имеющими корней), если их решения разыскивать среди обычных действительных чисел. Таковыми, например, являются квадратные уравнения $x^2 + 1 = 0$, $x^2 + 2x + 2 = 0$. Вообще, всякое уравнение, дискриминант которого отрицателен. Такое положение само по себе не вызывало особого беспокойства, но всё изменилось, когда были открыты формулы для решения уравнений 3-й и 4-й степени. В предыдущей главе мы уже обсуждали те сложности, которые возникли при использовании формулы Кардано в так называемом неприводимом случае, когда кубическое уравнение имеет три различных действительных корня. Тот факт, что формула для их вычисления содержит в этом случае корни квадратные из отрицательных чисел, и послужил истинным толчком для введения «мнимых», или комплексных чисел и оперирования с ними. Выражения $a + b\sqrt{-1}$, где a, b — действительные числа, всё чаще стали встречаться в математических текстах. К ним относились как к полезным «фикциям», действия над которыми позволяли получать результаты при решении различных задач, и не только алгебраических. Позднее для обозначения $\sqrt{-1}$ Эйлером был введён символ i, названный мнимой единицей, а для комплексных чисел стала употребляться запись

$$z = a + bi, \tag{2.14}$$

в которой a называется *действительной частью*, b — *мнимой частью* комплексного числа z. Сложение (вычитание) комплексных чисел означало просто сложение (вычитание) отдельно их действительных и мнимых частей. Например,

$$(2 - 3i) + (1 + 4i) = 3 + i.$$

При умножении же комплексных чисел обычным образом раскрывали скобки и использовали равенство $i^2 = -1$, вытекающее из определения мнимой единицы. Так,

$$(2 + 3i)(3 - 5i) = 6 - 10i + 9i - 15i^2 = 6 - i + 15 = 21 - i.$$

Эти правила в общем виде можно записать тогда следующим образом:

$$(a+bi)\pm(c+di)=(a\pm c)+(b\pm d)i,$$
$$(a+bi)(c+di)=(ac-bd)+(ad+bc)i.$$

Некоторым аргументом в пользу комплексных чисел было то, что законы, которым подчиняются действия над ними, в точности совпадали с законами действий над обыкновенными действительными числами. Четвёртая арифметическая операция — деление — также может быть определена для комплексных чисел. Действительно, пусть требуется найти такое комплексное число $x+yi$, что

$$(x+yi)(a+bi)=c+di. \qquad (2.15)$$

Если оно существует, то как раз и является частным от деления $c+di$ на $a+bi$. С уравнением (2.15) проще всего обойтись следующим образом. Умножим обе части на комплексное число $\overline{z}=a-bi$, называемое *сопряжённым* к числу (2.14). Заметим, что $(a+bi)(a-bi)=a^2+b^2$ является уже действительным числом, поэтому после умножения (2.15) на $a-bi$ получим

$$(x+yi)(a^2+b^2)=(c+di)(a-bi)=(ac+bd)+(ad-bc)i.$$

Из последнего равенства находим частное

$$x+yi=\frac{ac+bd}{a^2+b^2}+\frac{ad-bc}{a^2+b^2}i.$$

Например,

$$\frac{1+2i}{3+i}=\frac{(1+2i)(3-i)}{(3+i)(3-i)}=\frac{5+5i}{10}=\frac{1}{2}+\frac{i}{2}.$$

Лишь в одном случае, когда $a=b=0$, частного не существует. Это и понятно, так как тогда $a+bi=0$.

Прошло немало времени, прежде чем было понято, что имеется способ избежать какого-то мистического толкования комплексных чисел и, в частности, числа i. Надо лишь встать на

несколько иную, правда, более формальную точку зрения, что и было сделано ирландским математиком Гамильтоном (1805–1865). Вместо выражений вида $a + bi$ он рассматривал упорядоченные пары (a, b) действительных чисел a и b. Такую пару можно геометрически интерпретировать как точку плоскости, зафиксировав предварительно систему координат (рис. 6).

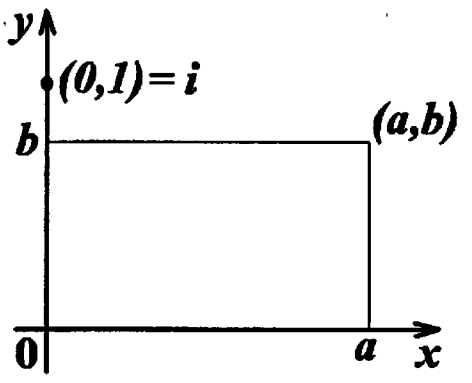


Рис. 6.

На множестве всех пар Гамильтон определил операции сложения и умножения, используя аналогию с операциями, рассмотренными выше:

$$(a, b) + (c, d) = (a + c, b + d), \tag{2.16}$$

$$(a, b)(c, d) = (ac - bd, ad + bc). \tag{2.17}$$

При этом, как и прежде, оказываются выполнимы обратные операции — вычитание и деление. Именно,

$$(a, b) - (c, d) = (a - c, b - d),$$

$$\frac{(c, d)}{(a, b)} = \left(\frac{ac + bd}{a^2 + b^2}, \quad \frac{ad - bc}{a^2 + b^2}\right).$$

Чтобы убедиться в справедливости, например, последнего равенства, достаточно, используя (2.17), проверить, что

$$\left(\frac{ac + bd}{a^2 + b^2}, \quad \frac{ad - bc}{a^2 + b^2}\right)(a, b) = (c, d).$$

Для операций, определённых равенствами (2.16) и (2.17), пары $(0, 0)$ и $(1, 0)$ играют роль нулевого и, соответственно, единичного элемента. Например, $(1, 0)(a, b) = (a, b)$ для любых $a, b \in \mathbb{R}$. Точкам действительной оси отвечают пары вида $(a, 0)$, и операции над ними (опять же в силу (2.16) и (2.17)) не отличаются от операций над действительными числами

$$(a, 0) + (b, 0) = (a + b, 0), \quad (a, 0)(b, 0) = (ab, 0).$$

Это позволяет отождествить каждую пару $(a,0)$ с действительным числом a и считать, что построенная числовая система содержит в себе все действительные числа.

Совершенно другую роль играют пары вида $(0,b)$, заполняющие ось ординат, и в частности, комплексное число $(0,1)$. Оно-то и выполняет роль мнимой единицы. В самом деле, из (2.17) следует. что

$$(0,1)(0,1) = (-1,0). \tag{2.18}$$

Используя обозначение i теперь уже для пары $(0,1)$ и отождествление $(a,0) = a$, мы можем записать равенство (2.18) в виде $i \cdot i = i^2 = -1$. И наконец, последний шаг, возвращающий нас, но уже с иных позиций, к записи (2.14)

$$(a,b) = (a,0) + (0,b) = (a,0) + (b,0)(0,1) = a + bi.$$

Впрочем, это возвращение к записи (2.14) (она называется *алгебраической формой* комплексного числа) вовсе не обязательно. Можно было бы работать с комплексными числами как с парами, оперируя над ними по правилам (2.16),(2.17). Просто алгебраическая форма удобнее, к тому же она освящена традицией.

Справедливости ради отметим, что геометрическое представление комплексных чисел в виде точек плоскости предшествовало описанному выше гамильтонову обоснованию. При этом, что особенно важно, была дана и геометрическая интерпретация действий над комплексными числами. Остановимся на этом подробнее. Рассмотрим радиус-вектор, заканчивающийся в точке (a,b) и изображающий комплексное число $z = a + bi$ (рис. 7).

Длину этого вектора

$$r = \sqrt{a^2 + b^2}$$

называют *модулем* числа z, а угол φ, образованный вектором и положительным направлением оси абсцисс, — *аргументом* числа z. Очевидно, что $a = r\cos\varphi$, $b = r\sin\varphi$, и тогда

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

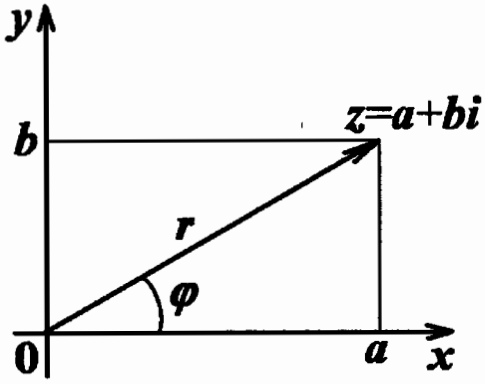


Рис. 7.

Полученное выражение называется *тригонометрической формой* комплексного числа.

Понятно, что сложение и вычитание комплексных чисел сводится к сложению и вычитанию соответствующих им векторов. Для выяснения геометрического смысла остальных операций удобно воспользоваться тригонометрической формой. Так, если

$$z_1 = r_1(\cos\varphi_1 + i\sin\varphi_1), \quad z_2 = r_2(\cos\varphi_2 + i\sin\varphi_2),$$

то по правилу умножения

$$z_1 z_2 = r_1 r_2[(\cos\varphi_1\cos\varphi_2 - \sin\varphi_1\sin\varphi_2) + \\ + i(\sin\varphi_1\cos\varphi_2 + \cos\varphi_1\sin\varphi_2)],$$

и следовательно,

$$z_1 z_2 = r_1 r_2[\cos(\varphi_1 + \varphi_2) + i\sin(\varphi_1 + \varphi_2)]. \qquad (2.19)$$

Таким образом, умножение комплексных чисел сводится к перемножению их модулей и сложению аргументов. Аналогично, при делении комплексных чисел их модули также делятся, а аргументы вычитаются

$$z_1/z_2 = (r_1/r_2)[\cos(\varphi_1 - \varphi_2) + i\sin(\varphi_1 - \varphi_2)].$$

Теперь нетрудно понять, что с комплексными числами связаны хорошо нам знакомые преобразования плоскости — поворот, гомотетия и параллельный перенос. Так например, если z — произвольное комплексное число, а z_0 — фиксированное, то соответствие $z \to z + z_0$ определяет параллельный перенос на вектор, отвечающий числу z_0. При умножении же получаем соответствие $z \to zz_0$, равносильное, как показывает формула (2.19), композиции двух преобразований — поворота на угол φ_0 (аргумент числа z_0) и гомотетии с коэффициентом r_0 равным модулю этого числа. В частности, если $r_0 = 1$, то получаем чистый поворот, а при $\varphi_0 = 2\pi k$, то есть $z_0 = r_0$, получаем чистую гомотетию.

Выходит, что Декарт был не прав — представить комплексные числа всё-таки можно, и названия «мнимые» они, пожалуй, не

заслуживают. Важно, что на этих геометрических представлениях и основаны многие применения комплексных чисел в разных областях математики. Для алгебры наиболее существенно то, что в области комплексных чисел становится разрешимым алгебраическое уравнение любой степени. Это утверждает так называемая основная теорема алгебры, точную формулировку которой читатель найдёт в следующей главе.

Особое значение для дальнейшего имеет уравнение вида $z^n = c$, где c — некоторое комплексное число. Решения этого уравнения (мы покажем, что их ровно n) называют корнями n-й степени из числа c. Для их отыскания удобнее всего использовать тригонометрическую форму комплексных чисел. Пусть

$$c = \rho(\cos\alpha + i\sin\alpha) \tag{2.20}$$

и $z = r(\cos\varphi + i\sin\varphi)$ — некоторое решение уравнения $z^n = c$. Тогда согласно (2.19) получаем

$$z^n = r^n(\cos n\varphi + i\sin n\varphi). \tag{2.21}$$

Сравнивая (2.20) и (2.21), заключаем, что

$$r^n = \rho, \quad n\varphi = \alpha + 2\pi k, \quad (k \in \mathbb{Z}).$$

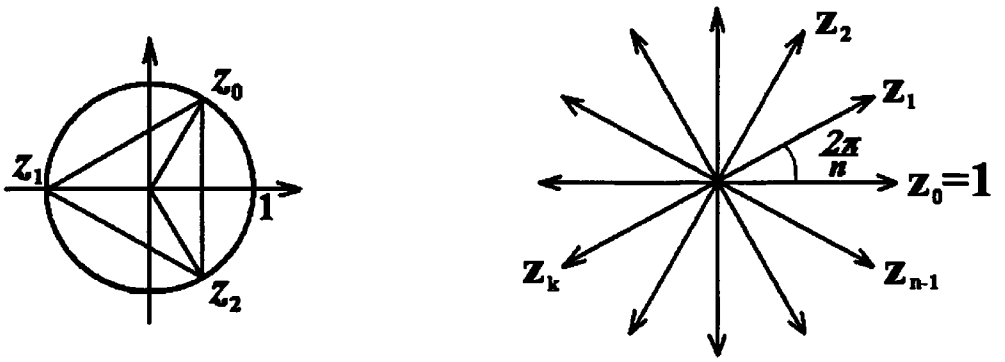
Следовательно,

$$r = \sqrt[n]{\rho}, \quad \varphi = \frac{\alpha}{n} + \frac{2\pi k}{n},$$

и всякий корень n-й степени из c может быть записан в виде

$$z_k = \sqrt[n]{\rho}\left[\cos\left(\frac{\alpha}{n} + \frac{2\pi k}{n}\right) + i\sin\left(\frac{\alpha}{n} + \frac{2\pi k}{n}\right)\right], \quad (k \in \mathbb{Z}). \tag{2.22}$$

Хотя k и пробегает здесь бесконечное множество значений, однако число различных корней, как нетрудно заметить, конечно. Всё дело в том, что двум значениям k сравнимым по модулю n, отвечает в формуле (2.22) одно и то же комплексное число. А значит, для того, чтобы получить все корни уравнения, достаточно рассмотреть значения k от 0 до $n - 1$. Например, корни



а) б)

Рис. 8.

кубические из -1 (решения уравнения $z^3 = -1$) таковы

$$z_0 = \cos\frac{\pi}{3} + i\sin\frac{\pi}{3} = \frac{1}{2} + i\frac{\sqrt{3}}{2}, \qquad z_1 = \cos\pi + i\sin\pi = -1,$$
$$z_2 = \cos\frac{5\pi}{3} + i\sin\frac{5\pi}{3} = \frac{1}{2} - i\frac{\sqrt{3}}{2}.$$

Легко заметить (рис. 8 а)), что точки на плоскости, изображающие числа z_0, z_1, z_2, лежат в вершинах правильного треугольника, вписанного в окружность единичного радиуса. Распространяя это наблюдение на общую ситуацию, можно прийти к выводу, что корни n-й степени из c изображаются точками, лежащими в вершинах правильного n-угольника, вписанного в окружность радиуса $\sqrt[n]{\rho}$.

В дальнейшем нам не раз придётся обращаться к решениям уравнения $z^n = c$ и прежде всего к решениям уравнения $z^n = 1$, то есть к корням n-й степени из 1. Как следует из предыдущего, они имеют вид

$$z_k = \cos\frac{2\pi k}{n} + i\sin\frac{2\pi k}{n}.$$

Рисунок 8 б) даёт их геометрическое изображение. Множество этих корней — будем его обозначать C_n — обладает рядом интересных алгебраических свойств. Можно заметить, что все корни могут быть получены из одного z_1 путём возведения в подходя-

щую степень. В самом деле,

$$z_1^k = \left(\cos\frac{2\pi}{n} + i\sin\frac{2\pi}{n}\right)^k = \\ = \cos\frac{2\pi k}{n} + i\sin\frac{2\pi}{k}n = z_k \quad (k = 0, 1, \ldots, n-1).$$

По этой причине z_1 называют первообразным корнем. И ещё, если перемножить любые два корня из 1 z_k и z_m, то результатом будет снова корень из 1, поскольку

$$(z_k z_m)^n = z_k^n z_m^n = 1.$$

Любопытно также и то, что в правиле перемножения корней мы можем отметить аналогию со сложением классов вычетов. Действительно,

$$z_k z_m = \cos\frac{2\pi(k+m)}{n} + i\sin\frac{2\pi(k+m)}{n}.$$

Если $k + m < n$, то, очевидно, $z_k z_m = z_{k+m}$. Если же $k + m = n + r \geqslant n$, то

$$\cos\frac{2\pi(k+m)}{n} + i\sin\frac{2\pi(k+m)}{n} = \\ = \cos\left(2\pi + \frac{2\pi r}{n}\right) + i\sin\left(2\pi + \frac{2\pi r}{n}\right) = \cos\frac{2\pi r}{n} + i\sin\frac{2\pi r}{n},$$

и следовательно, $z_k z_m = z_r$, где $r \equiv k + m \mod n$ $(0 \leqslant r < n)$.

Всё это означает, что равенство $z_k z_m = z_r$ в C_n выполняется тогда и только тогда, когда $\overline{k} + \overline{m} = \overline{r}$ в кольце вычетов по модулю n, то есть умножению корней из 1 соответствует сложение классов вычетов.

Иллюстрацией сказанного может служить приводимая ниже (табл. 2) таблица умножения корней 5-й степени из 1 (сравните её с таблицей сложения классов вычетов по модулю 5).

Таблица 2.

$\cdot$	z_0	z_1	z_2	z_3	z_4
z_0	z_0	z_1	z_2	z_3	z_4
z_1	z_1	z_2	z_3	z_4	z_0
z_2	z_2	z_3	z_4	z_0	z_1
z_3	z_3	z_4	z_0	z_1	z_2
z_4	z_4	z_0	z_1	z_2	z_3

Дополнения и задачи

1. В $\mathbb{Z}_7$ и в $\mathbb{Z}_{10}$ указать все обратимые элементы и найти к ним обратные.

2. Найти элемент $x \in \mathbb{Z}_5$, являющийся решением уравнения $\overline{2}x = \overline{3}$ и убедиться, что это решение единственно.

3. Определить, сколько решений в $\mathbb{Z}_{12}$ имеет каждое из следующих уравнений, и найти эти решения

$$\text{а) } \overline{3}x = \overline{6}, \qquad \text{б) } \overline{4}x = \overline{5}.$$

4. Показать, что система

$$\begin{cases} \overline{3}x + \overline{2}y = \overline{8} \\ \overline{2}x + \overline{3}y = \overline{5} \end{cases}$$

имеет единственное решение $(x; y)$, где $x, y \in \mathbb{Z}_{11}$, и найти это решение. Имеет ли решение в $\mathbb{Z}_{11}$ система

$$\begin{cases} \overline{3}x + \overline{2}y = \overline{8} \\ \overline{2}x + \overline{5}y = \overline{5} \ ? \end{cases}$$

Можно ли указать какой-нибудь признак, позволяющий понять, когда система линейных уравнений

$$\begin{cases} a_1x + b_1y = c_1 \\ a_2x + b_2y = c_2 \end{cases}$$

с коэффициентами из $\mathbb{Z}_n$ имеет в $\mathbb{Z}_n$ единственное решение, когда решений несколько, когда их нет вовсе? В поисках такого признака полезно рассмотреть ещё какие-нибудь примеры.

5. Вот как доказывается сформулированная в основном тексте теорема о разложении любой подстановки в произведение независимых циклов. Пусть σ — подстановка, действующая на множестве M. Для произвольного элемента $a \in M$ рассмотрим множество

$$A = \{a, \sigma(a), \sigma^2(a), \ldots\}, \tag{2.23}$$

содержащее вместе с a его образы при повторном применении подстановки σ. Множество A называют *орбитой* элемента a. Если $b \in M$ — другой элемент, то для его орбиты

$$B = \{b, \sigma(b), \sigma^2(b), \ldots\}$$

имеются лишь две возможности: либо она совпадает с орбитой (2.23), либо вообще не имеет с ней общих элементов. В самом деле, пусть орбиты A и B имеют непустое пересечение и пусть элемент $c \in A \cap B$. В этом случае

$$c = \sigma^k(a) = \sigma^m(b). \tag{2.24}$$

Пусть для определённости $k \geqslant m$. Применяя к обеим частям равенства (2.24) подстановку σ^{-m}, получим:

$$\sigma^{-m}(\sigma^k(a)) = \sigma^{-m}(\sigma^m(b)) = b, \quad \text{или} \quad b = \sigma^l(a), \quad \text{где } l = k - m. \tag{2.25}$$

Последнее означает, что $b \in A$, а значит, и все образы $\sigma^i(b) = \sigma^{l+i}(a)$ принадлежат орбите A, то есть $B \subseteq A$.

Пусть $\sigma^t = e$. Снова обратимся к равенству (2.25). Применяя к обеим его частям $t - l$ раз подстановку σ, получим

$$\sigma^{t-l}(b) = \sigma^{t-l}(\sigma^l(a)) = \sigma^t(a) = e(a) = a,$$

поэтому $a \in B$, а следовательно, и $A \subseteq B$, то есть $A = B$. Итак, орбиты либо совпадают, либо не имеют общих элементов. Значит, множество M можно представить как объединение непересекающихся орбит

$$M = A \cup B \cup C \cup \ldots$$

Этому разбиению и соответствует разложение подстановки σ в произведение независимых циклов:

$$\sigma = (a, \sigma(a), \sigma^2(a), \ldots)\,(b, \sigma(b), \sigma^2(b), \ldots)\,(c, \sigma(c), \sigma^2(c), \ldots)\ldots$$

Попытайтесь самостоятельно понять, как убедиться в единственности такого разложения.

6. Ясно, что если цикл длины l повторить l раз, то в результате получится тождественная подстановка. Сколько раз (как минимум) надо для этого повторить произвольную подстановку, если она раскладывается на независимые циклы, длины которых $l_1, l_2, \ldots, l_r$? Докажите, что наименьшее число повторений равно наименьшему общему кратному чисел $l_1, l_2, \ldots, l_r$.

7. Докажите, что число чётных подстановок в S_n совпадает с числом нечётных. Для этого достаточно убедиться, что если τ — произвольная нечётная подстановка, а f пробегает знакопеременную группу A_n, то соответствие $f \to \tau f$ является взаимно однозначным соответствием между чётными и нечётными подстановками.

8. Кроме группы подстановок, рассмотрим ещё один важный для дальнейшего пример группы преобразований. Движением плоскости назовём такое её преобразование f, которое сохраняет неизменным расстояние между любыми двумя точками этой плоскости[1]. Другими словами (см. рис. 9), расстояние между образами $A' = f(A)$ и $B' = f(B)$ точек A и B всегда такое же, как и между исходными точками: $|A'B'| = |AB|$. Ясно, что композиция двух движений есть снова движение. Читатель без труда проверит, что свойства (2.10), (2.11), (2.12) также оказываются выполненными и для движений.

Совокупность всех движений с определённой для них операцией умножения (композиции) называют группой движений плоскости. Примерами движений, понятно, являются параллельные переносы и повороты плоскости на произвольный угол. Ещё один тип движений — это так называемые зеркальные отражения

[1] Это понятие имеет, конечно, смысл и для пространства.

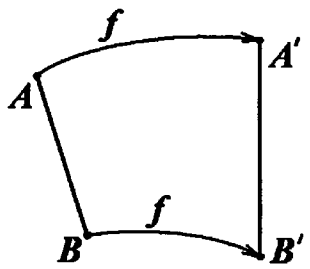


Рис. 9.

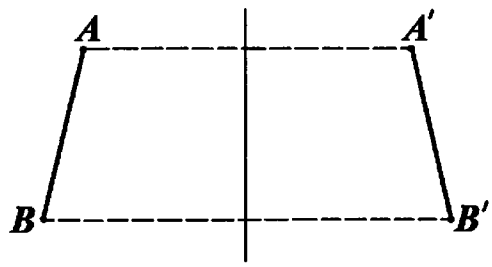


Рис. 10.

(осевые симметрии) относительно любых осей (рис. 10). В следующих пунктах мы покажем, что этими тремя типами преобразований «почти» исчерпываются все движения плоскости.

9. Выше говорилось о том, что имеется тесная связь между операциями над комплексными числами, с одной стороны, и преобразованиями плоскости, с другой. Можно добавить к этому следующее.

а) Пусть f — поворот плоскости на угол α вокруг произвольной точки z_0, точнее, точки, изображающей комплексное число z_0 (рис. 11). Тогда, если $z' = f(z)$ — образ точки z при этом преобразовании, то

$$z' = z_0 + (z - z_0)(\cos\alpha + i\sin\alpha). \qquad (2.26)$$

Действительно, вектор $z' - z_0$ получается из вектора $z - z_0$ поворотом на угол α. Следовательно, $z' = z_0 + (z - z_0)(\cos\alpha + i\sin\alpha)$.

б) Если f — зеркальное отражение относительно оси OA, составляющей с Ox угол α (рис. 12), и $z' = f(z)$, то

$$z' = \overline{z}\,(\cos 2\alpha + i\sin 2\alpha). \qquad (2.27)$$

где $\overline{z}$ — число, комплексно сопряжённое с z. Для доказательства достаточно убедиться, что угол между векторами z' и $\overline{z}$ равен 2α (сделайте это самостоятельно).

Справедливы и утверждения, обратные к а) и б); таким образом, формулы (2.26) и (2.27), связывающие z и z', полностью характеризуют преобразования поворота и зеркального отражения соответственно.

10. Что можно сказать о композиции движений различных типов? Ответ содержится в утверждениях этого пункта.

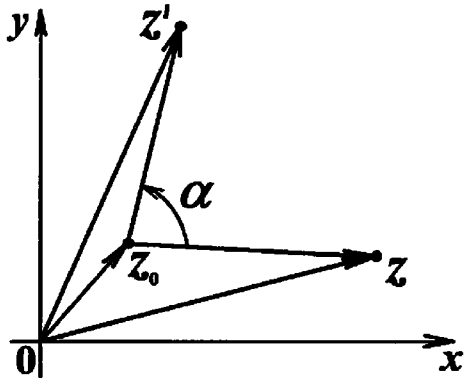


Рис. 11.

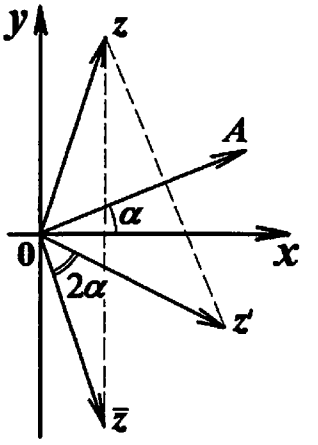


Рис. 12.

а) Произведение двух сдвигов есть сдвиг.

б) Произведение (в любом порядке) поворота, отличного от тождественного, и сдвига есть поворот.

в) Произведение двух поворотов на углы α_1 и α_2 есть поворот на угол $\alpha_1+\alpha_2$, если $\alpha_1+\alpha_2 \neq 2\pi k$, и параллельный перенос в противном случае.

г) Произведение двух зеркальных отражений относительно осей, пересекающихся под углом α, есть поворот на угол 2α вокруг точки пересечения осей.

д) Произведение двух зеркальных отражений относительно параллельных осей есть сдвиг в направлении, перпендикулярном осям, на расстояние, равное удвоенному расстоянию между осями.

е) Произведение (в любом порядке) поворота и зеркального отражения есть зеркальное отражение.

Первое утверждение очевидно, к тому же оно комментировалось в основном тексте. Остальные можно доказать, используя связь движений с комплексными числами[1]. Вот как доказываются, например, утверждения б) и г). Пусть f_1 — сдвиг на вектор z_1, f_2 —

[1] Имеются и чисто геометрические доказательства перечисленных утверждений (см., например, [30]).

нетривиальный поворот с центром в точке z_0 на угол α. Положим для краткости

$$e(\alpha) = \cos\alpha + i\sin\alpha.$$

Тогда, с учетом (2.26),

$$f_1(z) = z + z_1, \qquad f_2(z) = z_0 + (z - z_0)\,e(\alpha).$$

Образ произвольной точки z при действии композиции $f_1 f_2$ находим простым вычислением

$$f_1(f_2(z)) = f_1(z_0 + (z - z_0)\,e(\alpha)) = z_0 + z_1 + (z - z_0)\,e(\alpha). \quad (2.28)$$

Проверим, что полученное равенство можно записать в виде

$$f_1(f_2(z)) = z_0' + (z - z_0')\,e(\alpha). \quad (2.29)$$

где z_0' — подходящим образом подобранное комплексное число. Этим и будет доказано, что $f_1 f_2$ есть поворот. Сравнивая правые части равенств (2.28) и (2.29), приходим к соотношению

$$z_0 + z_1 - z_0\,e(\alpha) = z_0' - z_0'\,e(\alpha).$$

Так как по предположению f_2 — нетривиальный поворот, то $e(\alpha) \neq 1$, и следовательно, искомый центр поворота может быть найден из равенства

$$z_0' = \frac{z_0 + z_1 - z_0\,e(\alpha)}{1 - e(\alpha)}.$$

Мы доказали утверждение б).

Пусть теперь f_1 и f_2 — зеркальные отражения относительно осей OA_1 и OA_2, составляющих неравные углы α_1 и α_2 с осью Ox (рис. 13). Тогда угол между осями OA_1 и OA_2 равен $\alpha_2 - \alpha_1$. Согласно (2.27)

$$f_1(z) = \overline{z}\,e(2\alpha_1), \qquad f_2(z) = \overline{z}\,e(2\alpha_2),$$

$$f_2 f_1(z) = f_2(\overline{z}\,e(2\alpha_1)) = \overline{\overline{z}\,e(2\alpha_1)}\,e(2\alpha_2) = z\,e(-2\alpha_1)\,e(2\alpha_2).$$

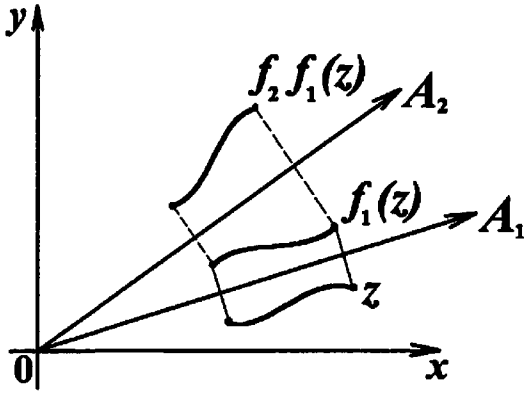


Рис. 13.

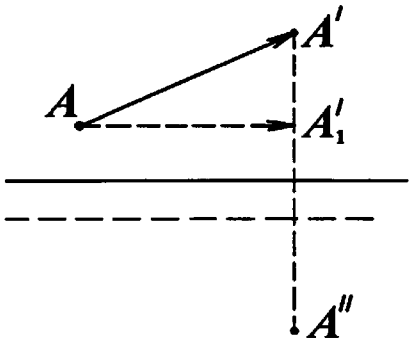


Рис. 14.

Так как всегда

$$e(\alpha)e(\beta) = (\cos\alpha + i\sin\alpha)(\cos\beta + i\sin\beta) = \\ = \cos(\alpha+\beta) + i\sin(\alpha+\beta) = e(\alpha+\beta),$$

то получаем окончательно $f_2(f_1(z)) = z\,e(2(\alpha_2 - \alpha_1))$, и значит, f_2f_1 — поворот вокруг точки O на угол $2(\alpha_2-\alpha_1)$, вдвое больший, чем угол между осями. Утверждение г) доказано.

11. Читатель мог заметить, что лишь один случай — композиция сдвига и зеркального отражения — не охвачен утверждениями а)–е) предыдущего пункта. Возникающие при этом преобразования — они называются *скользящими отражениями* — уже не сводятся к трём предыдущим типам. Можно лишь доказать, что всякое скользящее отражение может быть представлено как композиция сдвига и зеркального отражения, ось которого параллельна вектору сдвига. Комплексные числа позволяют сделать и это, однако удобнее здесь воспользоваться геометрическими рассуждениями (предоставляем читателю провести их, используя рис. 14). Из сделанного здесь замечания и утверждений предыдущего пункта вытекает следующий основной результат.

Всякое движение плоскости является или параллельным переносом, или поворотом, или зеркальным отражением, или скользящим отражением.

Впрочем, зеркальное отражение можно рассматривать как частный случай скользящего отражения, а именно, когда вектор параллельного переноса нулевой.

ГЛАВА 3

«ВЫЗОВ ЧЕЛОВЕЧЕСКОМУ РАЗУМУ»

Как уже отмечалось, с момента появления «Великого искусства» не прекращались настойчивые попытки найти формулы для решения алгебраических уравнений пятой и более высоких степеней и, быть может, общий метод решения таких уравнений. Накопленный опыт как будто подсказывал, что решение уравнений степени n не потребует операций более сложных, чем извлечение корней степени, не превосходящей n. Так обстояло дело для уравнений второй, третьей и четвертой степени. Так, по мнению многих, должно было быть и в общем случае. Математики разных рангов атаковали проблему, однако неудача следовала за неудачей. К XVIII веку прежний энтузиазм в поисках магических формул несколько истощился. Всё же убеждённость в разрешимости всех алгебраических уравнений в радикалах не была ещё поколеблена предшествующими неудачами.

Новая страница в этой затянувшейся на два с лишним века истории открывается с выходом в свет в 1770 году обширного мемуара «Размышления об алгебраическом решении уравнений». Автором его был замечательный французский математик Жозеф Луи Лагранж. Нельзя сказать, что старая проблема, которой занялся среди многих математиков и Лагранж, находилась на столбовых путях развития математики того времени. Прежде всего, конечно, это был период бурного накопления новых знаний в области математического анализа, дифференциальных уравнений, вариационного исчисления, аналитической механики. Здесь же группировались и основные интересы Лагранжа. Математики

понимали также, что поиск формул для решения алгебраических уравнений даже в случае успеха не мог сулить каких-либо выгод для практических приложений. Даже уравнения третьей и четвёртой степени никто не решал по известным формулам Кардано и Феррари — это было слишком неудобно. Основную роль играли различные приближённые методы, которые позволяли отыскивать корни алгебраических уравнений произвольной степени (и не только алгебраических) с любой степенью точности. Но вопрос о возможности выражения корней многочлена через его коэффициенты с помощью алгебраических операций был всё же принципиальным. И Лагранж со свойственным ему всегда стремлением к отысканию общих идей, лежащих в основе той или иной задачи, идёт на приступ проблемы, которая, по собственному его выражению, была «как бы вызовом человеческому разуму». Стоит напомнить, что по сравнению с эпохой Кардано лицо алгебры к этому моменту сильно изменилось. Трудами Виета и Декарта в обиход прочно вошла удобная математическая символика, позволяющая компактно записывать уравнения и неравенства; легко, почти автоматически, производить различные алгебраические преобразования. Был снят покров таинственности с комплексных чисел, и математики достаточно свободно оперировали с ними. За двести лет алгебра обогатилась рядом новых важных фактов, и в этом ряду центральное место занимала так называемая «основная теорема алгебры». Согласно этой теореме всякий многочлен

$$f(x) = x^n + a_1 x^{n-1} + \ldots + a_{n-1}x + a_n$$

степени n с комплексными (в частности, действительными) коэффициентами имеет ровно n комплексных корней, если каждый корень считать столько раз, какова его кратность (кратность корня x_0 считается равной k, если многочлен делится на $(x - x_0)^k$, но не делится на $(x - x_0)^{k+1}$). При этом, если $x_1, x_2, \ldots, x_n$ — корни (среди этих корней возможны одинаковые), то многочлен раскладывается на линейные множители следующим образом

$$f(x) = (x - x_1)(x - x_2)\ldots(x - x_n).$$

Простейшим частным случаем этой формулы является известное

из школьного курса разложение квадратного трёхчлена

$$x^2 + px + q = (x - x_1)(x - x_2).$$

Утверждение основной теоремы алгебры было сначала найдено эмпирическим путём. Первые доказательства (правда, небезупречные) принадлежали Даламберу и Эйлеру, и лишь на рубеже XVIII и XIX веков полное доказательство (и не одно, а целых четыре!) было дано Гауссом. Задолго до доказательства основной теоремы алгебры Виетом были открыты соотношения между корнями алгебраического уравнения и его коэффициентами. Формулы Виета для квадратного уравнения $x^2 + px + q = 0$ хорошо известны читателю. Они имеют вид

$$x_1 + x_2 = -p, \qquad x_1 x_2 = q. \tag{3.1}$$

Метод получения этих формул легко распространяется на уравнения высших степеней. Например, пусть x_1, x_2, x_3 — корни кубического уравнения $x^3 + a_1 x^2 + a_2 x + a_3 = 0$. Тогда $x^3 + a_1 x^2 + a_2 x + a_3 = (x - x_1)(x - x_2)(x - x_3)$. Раскрывая скобки в правой части, получим

$$x^3 + a_1 x^2 + a_2 x + a_3 = \\ = x^3 - (x_1 + x_2 + x_3)x^2 + (x_1 x_2 + x_1 x_3 + x_2 x_3)x - x_1 x_2 x_3.$$

Сравнивая коэффициенты при одинаковых степенях x, приходим к формулам Виета для кубического уравнения:

$$x_1 + x_2 + x_3 = -a_1, \qquad x_1 x_2 + x_1 x_3 + x_2 x_3 = a_2, \qquad x_1 x_2 x_3 = -a_3. \tag{3.2}$$

Подобные вычисления можно проделать и для уравнения произвольной степени

$$x^n + a_1 x^{n-1} + a_2 x^{n-2} + \ldots + a_{n-1} x + a_n = 0. \tag{3.3}$$

В результате получим

$$\begin{aligned} x_1 + x_2 + x_3 + \ldots + x_n &= -a_1, \\ x_1x_2 + x_1x_3 + \ldots + x_{n-1}x_n &= \ \ a_2, \\ x_1x_2x_3 + x_1x_2x_4 + \ldots + x_{n-2}x_{n-1}x_n &= -a_3, \\ &\ldots\ldots\ldots\ldots \\ x_1x_2x_3\ldots x_n &= (-1)^n a_n. \end{aligned} \tag{3.4}$$

Это и есть наиболее общий вид формул Виета. В связи с ними можно рассмотреть понятие симметрического многочлена от нескольких переменных. Многочлен $P(x_1, x_2, x_3, \ldots, x_n)$ называется симметрическим, если он остается неизменным при любой перестановке своих переменных. Многочлены в левых частях равенств (3.1), (3.2) и (3.4) представляют собой как раз симметрические многочлены от двух, трёх и n переменных. Они носят особое название — *элементарные симметрические многочлены*. Разумеется, не всякий симметрический многочлен является элементарным. Не таков, например, симметрический многочлен $x_1^3x_2 + x_1x_2^3$ от двух переменных. Однако оказывается, что его можно выразить через элементарные $x_1 + x_2$ и x_1x_2. Действительно,

$$x_1^3x_2 + x_1x_2^3 = x_1x_2\,(x_1^2 + x_2^2) = x_1x_2\,((x_1 + x_2)^2 - 2x_1x_2). \tag{3.5}$$

Замечательно, что *любой симметрический многочлен можно единственным образом представить в виде многочлена от элементарных симметрических многочленов*. Этот факт, называемый основной теоремой о симметрических многочленах, был уже известен во времена Лагранжа[1].

Однако вернёмся к работе самого Лагранжа. Автор «Размышлений» ставит перед собой задачу выявить и как-то сформулировать общие причины того, что методы, ведущие к успеху в случае уравнений степени не выше четырёх, оказываются бессильными для уравнений более высокой степени. Оказывается, эти причины связаны как раз со свойствами симметрии многочленов от нескольких переменных. Разберемся в этом поподробнее.

[1] Доказательство можно найти, например, в книге [3].

Прежде всего, из основной теоремы о симметрических многочленах и формул Виета непосредственно следует, что при замене переменных $x_1, x_2, \ldots, x_n$ в симметрическом многочлене $\varphi(x_1, x_2, \ldots, x_n)$ корнями уравнения (3.3) он принимает значение, выражающееся через коэффициенты $a_1, a_2, \ldots, a_n$ с помощью обычных арифметических операций. Это означает, что если *коэффициенты уравнения рациональны, то любой симметрический многочлен от его корней также обязательно имеет рациональное значение*, хотя сами корни могут быть и числами иррациональными. Например, корни $x_{1,2} = 1 \pm \sqrt{2}$ уравнения $x^2 - 2x - 1 = 0$ иррациональны. В то же время, значение симметрического многочлена $x_1^3 x_2 + x_1 x_2^3$ от этих корней рационально и равно -6. Конечно, иное дело, если многочлен $\varphi(x_1, x_2, \ldots, x_n)$ не является симметрическим. Однако, имеется немало таких многочленов, которые хотя и не являются симметрическими, но обладают, так сказать, частичной симметрией. Поясним сказанное примером многочлена от трёх переменных

$$\Delta(x_1, x_2, x_3) = (x_1 - x_2)(x_1 - x_3)(x_2 - x_3).$$

Легко заметить, что он не меняет своего вида при следующих подстановках из S_3 (не считая тождественной)

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \quad \text{и} \quad \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}.$$

Например,

$$\Delta(x_2, x_3, x_1) = (x_2 - x_3)(x_2 - x_1)(x_3 - x_1) = \Delta(x_1, x_2, x_3).$$

Остальные три подстановки из S_3

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$$

меняют, как нетрудно видеть, этот многочлен на противоположный. Таким образом, если переменные заменить корнями некоторого кубического уравнения, то при шести возможных подстановках из симметрической группы S_3 многочлен $\Delta(x_1, x_2, x_3)$

принимает не больше двух различных значений, тогда как произвольный многочлен от трёх переменных мог бы принять значений столько, сколько имеется различных подстановок, то есть шесть. Примером может служить хотя бы многочлен $x_1+2x_2+3x_3$. Резюмировать сказанное можно так: частичную симметрию многочлена $\varphi(x_1, x_2, \ldots, x_n)$ можно охарактеризовать числом k различных многочленов $\varphi_1, \varphi_2, \ldots, \varphi_k$, которые могут получиться при всевозможных подстановках аргументов $x_1, x_2, \ldots, x_n$. Чем меньше это число, тем симметричнее многочлен. Симметричные многочлены ни при каких подстановках не меняются, для них $k = 1$. Наоборот, те многочлены, для которых k максимально и равно $n!$, можно считать полностью несимметричными.

Имеется и другая, быть может, даже более естественная возможность охарактеризовать свойство симметрии многочлена. Выше на примере многочлена $\Delta(x_1, x_2, x_3)$ мы видели, что некоторые подстановки аргументов могут оставлять многочлен неизменным (к их числу всегда принадлежит тождественная подстановка). Для фиксированного многочлена $\varphi(x_1, x_2, \ldots, x_n)$ рассмотрим множество H всех таких подстановок. Их число m также может служить характеристикой симметрии многочлена φ: многочлен тем более симметричен, чем больше m. Лагранж выяснил, какова связь между числами k и m. Прежде всего, он обнаруживает, что множество H подстановок, сохраняющих неизменным данный многочлен, образуют подгруппу в симметрической группе S_n. Напомним, что это означает следующее:

1) если $\sigma_1, \sigma_2 \in H$, то есть подстановки σ_2 и σ_2 не меняют многочлена φ, то и их композиция также обладает этим свойством, то есть $\sigma_1\sigma_2 \in H$;

2) тождественная подстановка $e \in H$;

3) если $\sigma \in H$, то и $\sigma^{-1} \in H$.

В истинности этих трёх утверждений читатель легко может убедиться самостоятельно.

Далее Лагранж отвечает на вопрос, каковы те множества подстановок, которые одинаковым образом преобразуют многочлен φ, скажем, один из многочленов φ_i $(i = 1, 2, \ldots, k)$? Если

σ_i — одна из таких подстановок, то любая из таких подстановок имеет вид $\sigma_i h$, где $h \in H$ — произвольная подстановка, не меняющая многочлен φ. Зафиксируем σ_i и обозначим

$$\sigma_i H = \{\sigma_i h \mid h \in H\}.$$

В традиционной терминологии это так называемый *левый смежный класс* по подгруппе H. Один из смежных классов совпадает с самой подгруппой H: если положить $\varphi_1 = \varphi$ и если σ_1 — любая из подстановок, не меняющих многочлен φ, то $\sigma_1 H = H$. Всего же получается k смежных классов

$$\sigma_1 H, \sigma_2 H, \ldots, \sigma_k H,$$

отвечающих вышеописанным образом многочленам $\varphi_1, \varphi_2, \ldots, \varphi_k$ соответственно. В примере с многочленом $\Delta(x_1, x_2, x_3)$ ситуация такова. Один из смежных классов, как отмечено, есть сама подгруппа

$$H = \left\{ \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix} \right\},$$

а другой класс — это множество всех остальных подстановок из S_3

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} H = \left\{ \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} \right\}.$$

Этот простой пример ясно отражает общую ситуацию:

1) любые два смежных класса либо совпадают, либо не содержат общих элементов (две подстановки, принадлежащие разным смежным классам, не могут совпадать просто потому, что они по-разному преобразуют многочлен φ;

2) любые два смежных класса содержат одинаковое число подстановок, равное m — числу подстановок в H (разным подстановкам $h_1, h_2 \in H$ отвечают подстановки $\sigma_i h_1$, $\sigma_i h_2$ смежного класса $\sigma_i H$, которые также различны).

Таким образом, симметрическую группу S_n можно представить в виде объединения k непересекающихся смежных классов

$$S_n = \sigma_1 H \cup \sigma_2 H \cup \ldots \cup \sigma_k H,$$

в каждом из которых m элементов. Отсюда и вытекает искомая связь между числом k различных многочленов, получающихся из многочлена φ подстановками переменных, и числом m подстановок, сохраняющих этот многочлен неизменным:

$$km = n!. \tag{3.6}$$

Мы подходим к важному пункту «Размышлений» Лагранжа, а именно к вопросу о том, как отражается свойство частичной симметрии многочлена $\varphi(x_1, x_2, \ldots, x_n)$ на значениях, которые он принимает при подстановке корней уравнения (3.3). Напомним, что то единственное значение, которое принимает симметрический многочлен при любой подстановке корней уравнения (3.3), рационально, если рациональны коэффициенты уравнения.

Пусть $x_1, x_2, \ldots, x_n$ — корни произвольного уравнения (3.3) с рациональными коэффициентами, и пусть $\varphi(x_1, x_2, \ldots, x_n)$ принимает k различных значений $\varphi_1 = \varphi, \varphi_2, \ldots, \varphi_k$ при всевозможных подстановках аргументов, причём $\varphi_i = \varphi(x_{i_1}, x_{i_2}, \ldots, x_{i_n})$ для некоторой подстановки

$$\sigma_i = \begin{pmatrix} 1 & 2 & \ldots & n \\ i_1 & i_2 & \ldots & i_n \end{pmatrix}.$$

Рассмотрим многочлен k-ой степени

$$(x - \varphi_1)(x - \varphi_2)\ldots(x - \varphi_k) = x^k + b_1 x^{k-1} + b_2 x^{k-2} + \ldots + b_k,$$

корнями которого являются значения $\varphi_1, \varphi_2, \ldots, \varphi_k$. По формулам Виета коэффициенты $b_1, b_2, \ldots, b_k$ равны значениям соответствующих симметрических многочленов от $\varphi_1, \varphi_2, \ldots, \varphi_k$:

$$\begin{aligned} b_1 &= -(\varphi_1 + \varphi_2 + \ldots + \varphi_k), \\ b_2 &= \varphi_1\varphi_2 + \varphi_1\varphi_3 + \ldots + \varphi_{k-1}\varphi_k, \\ &\ldots\ldots\ldots\ldots\ldots \\ b_k &= (-1)^k \varphi_1\varphi_2\ldots\varphi_k. \end{aligned} \tag{3.7}$$

Легко видеть, что любая подстановка корней $x_1, x_2, \ldots, x_n$ приводит лишь к перестановке слагаемых в правых частях равенств (3.7) и, следовательно, оставит значения правых частей без изменения. Это означает, что коэффициенты $b_1, b_2, \ldots, b_k$ являются симметрическими многочленами от корней уравнения (3.3) с рациональными коэффициентами, и значит, сами они также рациональны. Итак, если многочлен $\varphi(x_1, x_2, \ldots, x_n)$ при всех подстановках корней принимает k значений, то эти значения (в том числе и $\varphi_1 = \varphi(x_1, x_2, \ldots, x_n)$) являются корнями уравнения степени k с рациональными коэффициентами. Если многочлен $\varphi(x_1, x_2, \ldots, x_n)$ достаточно симметричен, скажем, настолько, что $k < n$, то для отыскания $\varphi_1, \varphi_2, \ldots, \varphi_k$ придётся решать уравнение меньшей степени, чем первоначальное. Так для рассмотренного выше многочлена $\Delta = \Delta(x_1, x_2, x_3)$ число $k = 2$, при этом $\Delta_1 = \Delta$, $\Delta_2 = -\Delta$. Следовательно,

$$b_1 = -(\Delta_1 + \Delta_2), \qquad b_2 = \Delta_1 \Delta_2 = -\Delta^2,$$

и значения Δ_1, Δ_2 являются корнями квадратного уравнения $x^2 - \Delta^2 = 0$. Коэффициенты получившегося уравнения рациональны, поскольку Δ^2 является уже, очевидно, симметрическим многочленом.

Детально анализируя известные методы решения уравнений, Лагранж обнаруживает общность их в том, что все они сводятся к рассмотрению как раз таких функций от корней, которые обладают описанными выше свойствами симметрии, а именно, при всевозможных подстановках принимают меньше различных значений $\varphi_1, \varphi_2, \ldots, \varphi_k$, чем степень исходного уравнения. Поэтому и отыскание их, как мы говорили выше, представляет задачу более простую, чем отыскание самих корней.

Лагранж замечает также, что каждый из кубических радикалов, входящих в формулу Кардано, можно записать в виде $\frac{1}{3}(x_1 + \alpha x_2 + \alpha^2 x_3)$, где α — один из кубических корней из единицы, отличный от единицы. Это наблюдение приводит Лагранжа к общему по своей идее выводу формул для решения уравнений второй, третьей и четвёртой степени. Для произвольного

уравнения степени n он рассматривает выражения вида

$$t = x_1 + \alpha x_2 + \alpha^2 x_3 + \ldots + \alpha^{n-1} x_n, \tag{3.8}$$

где α — один из корней степени n из единицы, отличный от единицы, а $x_1, x_2, \ldots, x_n$ — искомые корни данного уравнения. Эти выражения называют сейчас *резольвентами Лагранжа*, они и могут быть использованы для построения функций, обладающих частичной симметрией. Продемонстрируем сказанное на примерах уравнений третьей и четвёртой степени.

Рассмотрим резольвенту $t(x_1, x_2, x_3) = x_1 + \alpha x_2 + \alpha^2 x_3$ для приведённого кубического уравнения $x^3 + px + q = 0$. При циклической подстановке корней $\sigma = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$ получаем

$$t(x_2, x_3, x_1) = x_2 + \alpha x_3 + \alpha^2 x_1 = \\ = \alpha^2 (x_1 + \alpha x_2 + \alpha^2 x_3) = \alpha^2 t(x_1, x_2, x_3),$$

то есть при указанной подстановке исходная резольвента умножается на α^2. Возводя в куб обе части предыдущего равенства и используя то, что $\alpha^3 = 1$, мы приходим к соотношению

$$t^3(x_2, x_3, x_1) = t^3(x_1, x_2, x_3).$$

Иными словами, функция t^3 инвариантна (то есть не меняется) при действии циклической подстановки σ. Отсюда следует её инвариантность и относительно подстановки $\sigma^2 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}$. Впрочем, это можно доказать и непосредственно. Итак, имеется три подстановки ($m = 3$), не меняющие значение t^3:

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}.$$

Вспомнив соотношение (3.6), заключаем, что число k различных значений t^3 при всевозможных подстановках корней равно двум, и следовательно, эти значения, равные $\theta_1 = t^3(x_1, x_2, x_3)$ и $\theta_2 = t^3(x_1, x_3, x_2)$, являются корнями квадратного уравнения

$$y^2 - (\theta_1 + \theta_2)y + \theta_1\theta_2 = 0,$$

коэффициенты которого выражаются через p и q. Непосредственным вычислением находим

$$\begin{aligned}\theta_1+\theta_2 &= (x_1+\alpha x_2+\alpha^2 x_3)^3+(x_1+\alpha x_3+\alpha^2 x_2)^3 = \\ &= 2(x_1^3+x_2^3+x_3^3+6x_1x_2x_3)+ \\ &+3(\alpha+\alpha^2)(x_1^2x_2+x_1x_2^2+x_1^2x_3+x_1x_3^2+x_2^2x_3+x_2x_3^2).\end{aligned}$$

Поскольку α — корень уравнения $x^2+x+1=0$, то $\alpha+\alpha^2 = -1$. Дополняя выражение в первых скобках до полного куба, получим с учётом сказанного

$$\theta_1+\theta_2 = 2(x_1+x_2+x_3)^3-9(x_1^2x_2+x_1x_2^2+x_1^2x_3+x_1x_3^2+x_2^2x_3+x_2x_3^2).$$

Согласно формулам Виета,

$$x_1+x_2+x_3=0, \tag{3.9}$$

Поэтому

$$\theta_1+\theta_2 = -9\big[x_1x_2(x_1+x_2)+x_1x_3(x_1+x_3)+x_2x_3(x_2+x_3)\big].$$

Но $x_1+x_2=-x_3$, $x_1+x_3=-x_2$, $x_2+x_3=-x_1$, откуда $\theta_1+\theta_2 = 27x_1x_2x_3 = -27q$. Совершенно аналогично,

$$\begin{aligned}\theta_1\theta_2 &= (x_1+\alpha x_2+\alpha^2x_3)^3(x_1+\alpha x_3+\alpha^2x_2)^3 = \\ &= \big[x_1^2+x_2^2+x_3^2+(\alpha+\alpha^2)(x_1x_2+x_1x_3+x_2x_3)\big]^3 = \\ &= \big[x_1^2+x_2^2+x_3^2-(x_1x_2+x_1x_3+x_2x_3)\big]^3 = \\ &= \big[(x_1+x_2+x_3)^2-3(x_1x_2+x_1x_3+x_2x_3)\big]^3 = \\ &= -27p^3.\end{aligned} \tag{3.10}$$

Таким образом, получаем, что квадратное уравнение, корнями которого являются θ_1 и θ_2, имеет вид $y^2+27qy-27p^3=0$, а значит,

$$\theta_{1,2} = -\frac{27q}{2}\pm 27\sqrt{\frac{q^2}{4}+\frac{p^3}{27}}\,. \tag{3.11}$$

До формулы Кардано теперь лишь один шаг. В самом деле, $t(x_1,x_2,x_3)=\sqrt[3]{\theta_1}$, $t(x_1,x_3,x_2)=\sqrt[3]{\theta_2}$. В качестве значения

$t(x_1, x_2, x_3)$ может быть взят любой из трёх кубических корней из θ_1, тогда при выборе подходящего значения $t(x_1, x_3, x_2)$ нужно исходить из равенства $t(x_1, x_2, x_3)\, t(x_1, x_3, x_2) = -3p$ (сравните с (3.10)). Учитывая, кроме того, равенство (3.9), приходим к системе:

$$\begin{cases} x_1 + \alpha x_2 + \alpha^2 x_3 = \sqrt[3]{\theta_1} \\ x_1 + \alpha^2 x_2 + \alpha x_3 = \sqrt[3]{\theta_2} \\ x_1 + x_2 + x_3 = 0. \end{cases}$$

Сложим все три уравнения, не забывая, что $1 + \alpha + \alpha^2 = 0$. Это даёт $3x_1 = \sqrt[3]{\theta_1} + \sqrt[3]{\theta_2}$, откуда с учётом (3.11) получаем наконец

$$x_1 = \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}.$$

Важно, что проведённые рассуждения позволяют определить все корни кубического уравнения (вспомните трудности с формулой Кардано). Так, умножая первое уравнение системы на α^2, второе — на α и складывая получившиеся уравнения (включая и третье), получим

$$3x_2 = \alpha^2 \sqrt[3]{\theta_1} + \alpha \sqrt[3]{\theta_2},$$

$$x_2 = \alpha^2 \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} + \alpha \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}.$$

Совершенно аналогично,

$$x_3 = \alpha \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} + \alpha^2 \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}.$$

Рассмотрим теперь общее уравнение четвёртой степени

$$x^4 + a_1 x^3 + a_2 x^2 + a_3 x + a_4 = 0.$$

Среди корней четвёртой степени из единицы имеется корень -1, его и удобно взять в качестве α в резольвенте Лагранжа для этого уравнения. Резольвента тогда имеет вид

$$t = x_1 - x_2 + x_3 - x_4. \tag{3.12}$$

Нетрудно сообразить, что её не меняют следующие четыре подстановки (и только они)

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

В группе S_4 в общей сложности $4! = 24$ подстановки. При действии этих подстановок на выражение (3.12) получим шесть различных выражений

$$\begin{aligned}\theta_1 &= \ \ x_1 - x_2 + x_3 - x_4, & -\theta_1 &= -x_1 + x_2 - x_3 + x_4,\\ \theta_2 &= -x_1 + x_2 + x_3 - x_4, & -\theta_2 &= \ \ x_1 - x_2 - x_3 + x_4,\\ \theta_3 &= \ \ x_1 + x_2 - x_3 - x_4, & -\theta_3 &= -x_1 - x_2 + x_3 + x_4.\end{aligned}$$

Они являются корнями многочлена шестой степени $P(y)$, коэффициенты которого рационально выражаются через коэффициенты a_1, a_2, a_3, a_4. К счастью, этот многочлен (читателю предлагается самостоятельно найти его вид) содержит лишь чётные степени y. В самом деле, корни его разбиваются на пары противоположных друг другу чисел, и потому он разлагается на множители следующим образом

$$P(y) = (y-\theta_1)(y+\theta_1)(y-\theta_2)(y+\theta_2)(y-\theta_3)(y+\theta_3) = \\ = (y^2-\theta_1^2)(y^2-\theta_2^2)(y^2-\theta_3^2).$$

Очевидная подстановка $z = y^2$ сводит тогда задачу отыскания θ_1, θ_2, θ_3 к решению уравнения третьей степени. Найдя их, определяем корни x_1, x_2, x_3, x_4 из системы

$$\begin{cases} x_1 + x_2 + x_3 + x_4 = -a_1\\ x_1 - x_2 + x_3 - x_4 = \theta_1\\ -x_1 + x_2 + x_3 - x_4 = \theta_2\\ x_1 + x_2 - x_3 - x_4 = \theta_3. \end{cases}$$

Но вот в случае уравнения более высокой степени ($n \geqslant 5$) резольвенты Лагранжа уже не давали решения задачи. Причину этого понять нетрудно. Ведь, аналогично случаю $n = 3$, n-ая степень

t^n резольвенты (3.8) остаётся инвариантной относительно циклической подстановки $\sigma = (1, 2, 3, \ldots, n)$, а следовательно, и относительно её степеней σ^k, $k = 0, 1, 2, \ldots, n-1$. Всего таких подстановок, очевидно, n. Более детальный анализ показывает, что в большинстве случаев (мы не уточняем, что́ это означает) подстановками σ^k исчерпываются все те подстановки, относительно которых инвариантна степень t^n. Но тогда t^n является корнем многочлена степени $n!/n = (n-1)!$. Уже при $n = 5$ степень этого многочлена получается $4! = 24$, то есть больше, чем степень исходного многочлена. Вывод таков: резольвенты при $n \geqslant 5$, как правило, уже не сводят решение уравнения к более простой задаче. Попытки найти какие-то иные вспомогательные функции от корней, упрощающие исходную задачу, также не приносили успеха. В результате Лагранж приходит к мысли, что едва ли вообще возможно решить «одну из самых знаменитых и самых важных проблем Алгебры» — найти формулы для корней алгебраических уравнений любой степени. Однако он высказывает уверенность в том, что рассмотренные им принципы, использующие симметрию относительно группы всех подстановок или её подгрупп, и есть «истинные принципы решения уравнений». Как это вскоре выяснилось, предвидение Лагранжа оказалось справедливым. В 1824 г. Абель доказал невозможность решения в радикалах общего уравнения степени $n \geqslant 5$, и объяснялась она как раз особенностями строения симметрической группы S_n. Этими же особенностями объяснялось и то, что общее уравнение произвольной степени $n \geqslant 5$ вообще нельзя свести к решению более простых вспомогательных уравнений подобно тому, как это было сделано в случае $n = 3$ и $n = 4$. Решающее слово здесь ещё предстояло сказать Галуа, но разговор об этом впереди.

ГЛАВА 4

ОТ ПОДСТАНОВОК К АБСТРАКТНОЙ ГРУППЕ

Первые ростки теории групп историки математики находят уже у Л. Эйлера в одной из его работ по теории чисел. Эта работа, датированная 1774 годом, называлась «Доказательство, относящееся к вычетам, происходящим от деления степеней на простые числа». Более заметными эти ростки становятся в «Арифметических исследованиях» Гаусса (1801). Изучение сравнений, классов квадратичных форм и законов их композиции, свойств комплексных корней из единицы — все эти вопросы объединены общими групповыми идеями. Подлинный их праздник — в первой математической работе Гаусса о построении правильного 17-угольника. В ней Гаусс (ему тогда не исполнилось ещё 19 лет) решает проблему, возраст которой измеряется двумя тысячелетиями.

Однако главным источником понятий теории групп стали всё-таки исследования по разрешимости алгебраических уравнений в радикалах. В предыдущей главе мы могли проследить, как в «Размышлениях» Лагранжа непроизвольно возникают такие понятия, как группа, подгруппа, смежный класс. Хотя Лагранж свободно оперирует с ними, но соответствующих терминов ни у него, ни у его предшественников не было. Впервые термин «группа» встречается у Галуа, которому удалось полностью решить задачу о корнях алгебраических уравнений. Итогом его глубоких исследований было не только решение очень трудной проблемы, но и, что ещё важнее, рождение новой математиче-

ской теории. Основные работы Галуа не получили признания при жизни автора и были опубликованы только через 14 лет после его трагической смерти. Галуа погиб на дуэли в возрасте двадцати одного года. Об удивительных обстоятельствах его жизни можно прочесть в книгах [11], [13]. Значение работ Галуа было до конца понято много позже. Первое полное и систематическое изложение теории Галуа было дано Жорданом в его монументальном сочинении «Трактат о подстановках» (1870). Вплоть до начала XX века учение о группах так и развивается в виде теории групп подстановок или иных преобразований. Правда, некоторые алгебраисты гораздо раньше начинают осознавать, что понятие группы имеет куда более универсальный характер. Первое упоминание об абстрактной группе содержится в одной из статей крупного английского математика А. Кэли, относящейся ещё к 1854 году. Слово «абстрактная» понимается им в том смысле, что элементами группы являются некие отвлеченные символы, конкретная природа которых не уточняется. Это могут быть числа, векторы, подстановки или какие-то иные объекты. Главное, на чём сосредоточивается внимание, — это те свойства композиции элементов (умножения или сложения), которые являются общими для чисел, векторов, подстановок и т. д. В определении абстрактной группы, которое мы сейчас дадим, для групповой операции используется термин «умножение», да и вся терминология и символика связана с обычным числовым умножением (эту терминологию называют мультипликативной). Однако термин «умножение» в определении абстрактной группы не нужно понимать буквально. Речь идёт о любой операции, удовлетворяющей определенным требованиям. Общее определение группы таково.

Множество G, на котором определено умножение элементов, называется группой, если выполняются три аксиомы:

1. *Умножение ассоциативно:*
   $a(bc) = (ab)c$ *для любых* $a, b, c \in G$.

2. *В множестве G существует единичный элемент, то есть такой элемент e, что* $ea = ae = a$ *для любого* $a \in G$.

3. *Для каждого элемента $a \in G$ существует обратный, то есть такой элемент a^{-1}, что $a^{-1}a = aa^{-1} = e$.*

Содержащиеся в приведённом определении аксиомы полностью воспроизводят свойства композиции подстановок конечного множества или движений плоскости (вообще взаимно однозначных преобразований любого множества). Можно поэтому сказать, что совокупность всех подстановок n-элементного множества (симметрическая группа S_n), равно как и совокупность всех движений плоскости, является группой относительно операции умножения в смысле нашего общего определения. Не надо обладать обширными математическими знаниями, чтобы привести немало других примеров. Так, многие числовые множества являются группами относительно арифметических операций: сложения или умножения. Вот некоторые из них.

1) Множество $\mathbb{Q}^*$ всех ненулевых рациональных чисел относительно операции числового умножения — так называемая мультипликативная группа рациональных чисел[1].

2) Множество $\mathbb{R}^*$ всех ненулевых действительных чисел относительно той же операции — это мультипликативная группа действительных чисел.

3) Более интересный пример числовой группы (точнее, целую серию примеров) мы получим, если вспомним материал из второй главы о комплексных корнях из единицы. Действительно, на множестве C_n таких корней фиксированной степени n определена операция умножения, очевидно, ассоциативная. Среди элементов C_n содержится 1, и наконец, для всякого элемента $z_k \in C_n$ имеется в C_n обратный элемент.

В ряде случаев естественнее называть групповую операцию сложением (например, когда речь идёт об обычном числовом сложении). Тогда и аксиомы группы придётся переформулировать, используя терминологию, связанную со сложением — аддитивную терминологию. Аксиомы в этом случае будут выглядеть так:

[1] Стоит заметить здесь, что ненулевые целые числа не образуют группы по умножению (почему?).

1. *Сложение ассоциативно*: $a+(b+c)=(a+b)+c$ *для любых элементов* $a, b, c \in G$.

2. *В множестве* G *существует нулевой элемент, то есть так элемент* 0, *такой что* $0+a=a+0=a$ *для любого* $a \in G$.

3. *Для каждого элемента* $a \in G$ *существует в* G *противоположный, то есть такой элемент* $(-a)$, *что*

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

Приведём несколько примеров групп, в которых аддитивная терминология оказывается более уместной.

1) Множество $\mathbb{Z}$ всех целых чисел относительно операции сложения — аддитивная группа целых чисел.

2) Множество всех векторов в пространстве относительно операции сложения векторов.

3) Множество $\mathbb{Z}_n$ классов вычетов относительно операции сложения классов — аддитивная группа вычетов по модулю n.

В связи с последним примером может возникнуть вопрос, обладает ли групповыми свойствами другая операция, определённая в $\mathbb{Z}_n$ — умножение классов вычетов. Поскольку нулевой класс заведомо необратим, то вопрос разумно отнести лишь к множеству $\mathbb{Z}_n^*$, состоящему из ненулевых классов $\overline{1}, \overline{2}, \ldots, \overline{n-1}$. Из материала второй главы вытекает, что *множество* $\mathbb{Z}_n^*$ *является группой относительно умножения тогда и только тогда, когда* n *— простое число.*

Среди свойств групповой операции мы не находим столь привычной нам коммутативности ($ab=ba$ или $a+b=b+a$). Это и понятно, если учесть, что главным источником понятия группы были подстановки, а их умножение как раз и не обладает указанным свойством. В то же время имеется немало групп, например, все числовые группы, операция в которых коммутативна. Они так и называются коммутативными или абелевыми группами (в честь норвежского математика Нильса Абеля, впервые отметившего особую их роль в вопросе разрешимости алгебраических уравнений в радикалах). Группы могут содержать

бесконечно много элементов, могут быть и конечными — таковы симметрическая группа, группа корней n-ой степени из единицы, аддитивная группа вычетов. Число элементов конечной группы G называют её *порядком* и обозначают $|G|$. Так, $|S_n| = n!$, $|C_n| = |\mathbb{Z}_n| = n$.

Вполне естественным является желание распространить на произвольные абстрактные группы те понятия, которые знакомы нам по группам подстановок и, в общем, для этого нет никаких помех. Первое из них — понятие подгруппы. Подмножество H группы G назовём подгруппой, если оно само является группой относительно операции, определённой в G. Вот некоторые примеры.

1) Мультипликативная группа $\mathbb{Q}^*$ рациональных чисел — подгруппа мультипликативной группы $\mathbb{R}^*$ действительных чисел.

2) Множество чётных чисел — подгруппа аддитивной группы $\mathbb{Z}$ целых чисел. Этого нельзя сказать о множестве нечётных чисел (почему?)

3) Множество всех параллельных переносов есть подгруппа группы движений плоскости. Этого нельзя сказать о множестве всех поворотов, если не фиксировать центр поворота.

Если H — подгруппа группы G, а g — произвольный (но фиксированный) элемент из G, то, как и раньше, множество всевозможных произведений gh, где h пробегает всю подгруппу H, называется смежным классом (точнее, левым смежным классом) по подгруппе H, а элемент g — его представителем. Обозначая левый смежный класс через gH, получим, следовательно,

$$gH = \{gh \mid h \in H\}.$$

Аналогично определяются правые смежные классы Hg, и в неабелевых группах они могут не совпадать с левыми. Как и в случае групп подстановок, смежные классы по подгруппе (всё равно, левые или правые) обладают двумя основными свойствами.

1. *В случае конечной подгруппы H число элементов в любом смежном классе одно и то же и совпадает с порядком подгруппы.*

2. *Любые два смежных класса либо совпадают, либо не имеют общих элементов.*

Требуется небольшое терпение, чтобы разобраться в доказательстве этих фактов. Для доказательства первого из них рассмотрим отображение подгруппы H на смежный класс gH, сопоставляя всякому элементу $h \in H$ элемент $gh \in gH$. Если h пробегает подгруппу, то соответствующий ему элемент gh пробегает весь смежный класс, при этом различным h_1 и h_2 из подгруппы H будут соответствовать также различные элементы gh_1 и gh_2 из смежного класса gH. В самом деле, предположим, что $gh_1 = gh_2$. Умножив обе части этого равенства слева на элемент g^{-1}, обратный к g, получим $g^{-1}(gh_1) = g^{-1}(gh_2)$, или, используя ассоциативность, $(g^{-1}g)\,h_1 = (g^{-1}g)\,h_2$. Поскольку $g^{-1}g = e$, то (аксиома 2) $h_1 = h_2$, что противоречит предположению. Таким образом, имеется взаимно однозначное соответствие между H и gH. Это и означает, что число элементов в H и в gH одинаково, что и требовалось доказать.

Докажем теперь второе утверждение. Предположим, что смежные классы g_1H и g_2H имеют непустое пересечение, и пусть $a \in$ $\in g_1H \cap g_2H$. Тогда элемент a можно записать двумя способами

$$a = g_1h_1 = g_2h_2, \quad h_1, h_2 \in H.$$

Умножая это равенство справа на элемент h_1^{-1}, получим: $g_1 =$ $= g_2h_2h_1^{-1}$. Заметим, что $h_2h_1^{-1} \in H$ — это вытекает из определения подгруппы. Раз так, то $g_1 \in g_2H$, но тогда произвольный элемент g_1h, принадлежащий g_1H, равен $g_2h_2h_1^{-1}h$. Поскольку $h_2h_1^{-1}h$ снова принадлежит подгруппе H, то $g_1h \in g_2H$. Мы тем самым доказали, что $g_1H \subseteq g_2H$. Аналогично можно доказать обратное включение $g_2H \subseteq g_1H$, поэтому $g_2H = g_1H$. Итак, если два смежных класса имеют хотя бы один общий элемент, то они совпадают. Понятно, что это равносильно второму утверждению.

Характерной чертой наших рассуждений является то, что в них используются лишь свойства групповой операции, перечисленные в определении группы, и какая-либо дополнительная информация о её элементах не играет роли. Именно поэтому доказанные утверждения, как и те, что из них вытекают, справедливы для произвольных групп независимо от природы их элементов.

Для иллюстрации понятия смежного класса рассмотрим следующий пример. Пусть $\mathbb{Z}$ — аддитивная группа целых чисел, а H — её подгруппа, состоящая из чисел кратных пяти. Тогда можно выписать следующие смежные классы:

$$\begin{aligned} H = 0 + H &= \{\ldots, -10, -5, 0, 5, 10, \ldots\}, \\ 1 + H &= \{\ldots, -9, -4, 1, 6, 11, \ldots\}, \\ 2 + H &= \{\ldots, -8, -3, 2, 7, 12, \ldots\}, \\ 3 + H &= \{\ldots, -7, -2, 3, 8, 13, \ldots\}, \\ 4 + H &= \{\ldots, -6, -1, 4, 9, 14, \ldots\}. \end{aligned}$$

Читатель сразу же заметит, что это классы вычетов $\overline{0}$, $\overline{1}$, $\overline{2}$, $\overline{3}$, $\overline{4}$ по модулю 5. Легко видеть также, что всякий смежный класс совпадает с одним из этих пяти и что классы эти не пересекаются. Столь же очевидно, что

$$\mathbb{Z} = H \cup (1 + H) \cup (2 + H) \cup (3 + H) \cup (4 + H).$$

То же самое верно, конечно, и для любой группы: если g_1H, $g_2H, \ldots, g_rH$ — все различные (левые) смежные классы группы G по подгруппе H, то

$$G = g_1H \cup g_2H \cup \ldots \cup g_rH.$$

Множество смежных классов может быть и бесконечным. Тогда в этом равенстве будет объединение бесконечного семейства множеств.

В случае, когда G и H — конечные группы, из последнего равенства и утверждений 1 и 2 вытекает, что $|G| = |H| \cdot r$. Число r различных смежных классов называют *индексом* подгруппы H

в группе G и обозначают $|G : H|$. Мы пришли к следующей теореме Лагранжа.

Порядок любой подгруппы H конечной группы G является делителем порядка группы, при этом $|G| = |G : H| \cdot |H|$.

Для тех, кто разобрался в содержании предыдущей главы, в формулировке этой теоремы нет ничего неожиданного — аналогичное утверждение рассматривалось там для групп подстановок. Теперь оно доказано для произвольных конечных групп. Несмотря на его простоту, оно имеет важное значение как при изучении самих групп, так и в приложениях. Чтобы уже в этой главе хоть немного поговорить об этом, познакомимся ещё с несколькими понятиями.

Прежде всего ответим на такой вопрос. Как должна быть устроена наименьшая из тех подгрупп, которые содержат заданный элемент g некоторой группы G? Если H — искомая подгруппа, то ясно, что вместе с g она должна содержать и произведение $g \cdot g$, и вообще всевозможные произведения $g \cdot g \cdot \ldots \cdot g$ для произвольного k. Как это принято для чисел, будем обозначать такое произведение g^k и называть k-ой степенью элемента g. Кроме того, подгруппе H должны принадлежать обратный элемент g^{-1} и всевозможные его степени. Определим обычным образом целую отрицательную степень, полагая $g^{-k} = (g^{-1})^k$. Наконец, искомая подгруппа должна содержать единичный элемент e. Будем, по определению, считать, что $g^0 = e$. Итак, все целые степени g^m, $m \in \mathbb{Z}$, элемента g должны принадлежать подгруппе H. Но оказывается, уже само множество $\langle g \rangle$ всех целых степеней элемента g

$$\langle g \rangle = \{g^m \mid m \in \mathbb{Z}\} \tag{4.1}$$

является подгруппой группы G.

Проверить это можно, сославшись на равенства

$$g^m g^k = g^{m+k}, \qquad (g^m)^{-1} = (g^{-1})^m = g^{-m},$$

которые устанавливаются так же, как и для числовых степеней. Они показывают, что произведение $g^m g^k$ любых двух элементов из множества $\langle g \rangle$ и элемент $(g^m)^{-1}$, обратный к произвольному элементу из $\langle g \rangle$, снова принадлежат $\langle g \rangle$. Это и означает, что $\langle g \rangle$ —

подгруппа и, очевидно, наименьшая из тех, которые содержат данный элемент g.

Найденная нами подгруппа, определённая равенством (4.1), называется *циклической подгруппой*, порождённой элементом g, а сам этот элемент — *образующим* циклической подгруппы. Циклические подгруппы играют роль строительных блоков, из которых собирается, подобно зданию, вся группа, а её облик в сильной степени зависит от способа соединения этих блоков — циклических подгрупп. Проще всего устроены те группы, для образования которых достаточно всего лишь одного блока. Такие группы сами называются циклическими. Точнее говоря, группа G является циклической, если она совпадает с некоторой из своих циклических подгрупп. Это означает, что найдётся такой элемент $g \in G$ (образующий), что все другие элементы является его степенями

$$G = \langle g \rangle = \{g^m \mid m \in \mathbb{Z}\}.$$

Пример циклической группы даёт известная нам группа C_n корней n-ой степени из единицы. Чтобы понять это, вспомним, что один из этих корней имеет вид

$$z_1 = \cos\frac{2\pi}{n} + i\sin\frac{2\pi}{n},$$

а все другие связаны с ним соотношениями

$$z_k = z_1^k = \cos\frac{2\pi k}{n} + i\sin\frac{2\pi k}{n}, \quad k = 0, 1, 2, \ldots, n-1.$$

Следовательно, $C_n = \langle z_1 \rangle$.

Другим примером циклической группы (уже бесконечной) является аддитивная группа $\mathbb{Z}$ целых чисел. Только здесь, поскольку групповая операция — сложение, придётся говорить не о степенях образующего элемента, а о его кратных. Убедитесь, что в качестве образующего можно взять элемент 1 (а можно и -1).

Несколько ниже мы покажем, что приведённые примеры в некотором смысле исчерпывают всевозможные циклические группы (конечные и бесконечные). Пример группы C_n показывает, что

в группе может существовать такой элемент $g \neq e$, что $g^n = e$ для некоторого $n \neq 0$. Это обстоятельство даёт повод ввести следующее определение. Пусть G — произвольная группа.

Порядком элемента $g \in G$ называется наименьшее натуральное n такое, что $g^n = e$. Если такого n не существует, то g называется элементом бесконечного порядка.

Например, в мультипликативной группе $\mathbb{R}^*$ ненулевых действительных чисел все элементы, кроме чисел 1 и -1, имеют бесконечный порядок. Группа C_n позволяет сделать ещё одно наблюдение. Порядок её образующего элемента z_1 равен n и совпадает с порядком группы. Следующее утверждение показывает, что это так для любой циклической группы конечного порядка n.

Если образующий элемент циклической группы $\langle g \rangle$ имеет порядок n, то все различные элементы группы исчерпываются степенями

$$g^0 = e, g, g^2, \ldots, g^{n-1}. \tag{4.2}$$

Для доказательства нужно, во-первых, убедиться, что любая степень g^k ($k \in \mathbb{Z}$) совпадает с одной из перечисленных в (4.2). Действительно, пусть r — остаток от деления k на n

$$k = qn + r, \quad 0 \leqslant r < n.$$

Тогда

$$g^k = g^{qn+r} = (g^n)^q \cdot g^r = e \cdot g^r = g^r$$

— один из элементов (4.2). Во-вторых, надо проверить, что все элементы (4.2) различны. Действительно, если $g^k = g^p$, где $0 \leqslant p < k \leqslant n-1$, то $g^{k-p} = e$, причём $1 \leqslant k - p \leqslant n - 1$, а этого не может быть, так как порядок n — наименьшее натуральное число такое, что $g^n = e$.

Из только что доказанного утверждения вытекает следующее. В любой конечной группе порядок всякого элемента является делителем порядка группы. Действительно, порядок любого элемента $h \in G$ совпадает с порядком порождённой им циклической подгруппы $\langle h \rangle$, а значит, по теореме Лагранжа, является делителем порядка группы.

Уже эти простые сведения о группах могут быть применены в различных областях математики, например, в теории чисел. Вспомним хотя бы упоминавшуюся во второй главе малую теорему Ферма. Язык групп позволяет дать для неё удивительно компактное доказательство. Вот оно. Рассмотрим кольцо вычетов $\mathbb{Z}_p$ по простому модулю p. Его ненулевые элементы, как мы уже знаем, образуют по умножению группу порядка $p-1$. Поэтому для любого ненулевого класса $\overline{r} \in \mathbb{Z}_p$, согласно последнему утверждению, справедливо равенство $\overline{r}^{p-1} = \overline{1}$, то есть

$$\overline{r}^{p-1} - \overline{1} = \overline{0}. \qquad (4.3)$$

Если число a взаимно просто с p, то содержащий число a класс вычетов $\overline{r}$ ненулевой. Справедливое для него равенство (4.3) означает, что число $a^{p-1} - 1$ содержится в классе $\overline{0}$, то есть делится на p, а это и есть утверждение теоремы.

Поучительны групповые доказательства и других теорем элементарной теории чисел. Две из них — теорема Вильсона и теорема Эйлера — рассматриваются в дополнении.

Эту главу мы завершим обсуждением понятия, важного не только для алгебры, но и для всей математики в целом, — понятия изоморфизма. Этот термин образован соединением греческих слов «isos» (одинаковый, равный) и «morphe» (форма). В энциклопедическом словаре читаем:

> *«Изоморфизм — соответствие между объектами, выражающее тождество их структуры.»*

Хотя в этой характеристике и отражено самое главное, ей всё же недостает полной ясности. Действительно, слова «тождество структуры» (или, иначе, строения) не могут быть во всех случаях расшифрованы однозначно — всё зависит от того, о каких объектах, системах идёт речь. Когда говорят о строении алгебраических систем (в частности, групп), то имеют в виду те их черты, которые определяются свойствами операций. При этом вовсе не принимаются в расчёт индивидуальные свойства элементов, их конкретная природа. Последнее, напомним, как раз и было типично для доказательств почти всех утверждений этой

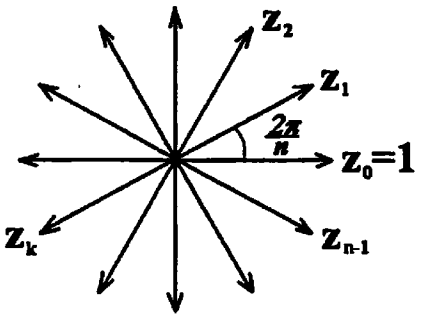


Рис. 15.

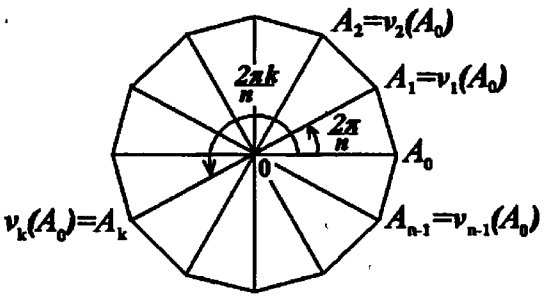


Рис. 16.

главы. Лучше всего разъяснить сказанное примером двух различных по составу элементов и в то же время одинаковых по свойствам операций групп.

Первая группа — это C_n. Её элементами являются корни n-ой степени из единицы (рис. 15). Как известно из второй главы, таких корней имеется ровно n и каждый из них записывается в виде

$$z_k = \cos\frac{2\pi k}{n} + i\sin\frac{2\pi k}{n}, \quad k = 0, 1, 2, \ldots, n-1.$$

Правило их умножения следующее

$$z_k z_m = z_r, \quad \text{где } r = \begin{cases} k+m, & \text{если } k+m < n; \\ k+m-n, & \text{если } k+m \geqslant n. \end{cases} \tag{4.4}$$

Другая группа — V_n — состоит из всех поворотов плоскости вокруг центра O правильного n-угольника, отображающих этот n-угольник на себя. Из рис. 16 ясно, что углы таких поворотов кратны $\dfrac{2\pi}{n}$, то есть имеют вид $\dfrac{2\pi k}{n}$; а для того, чтобы получить все такие повороты, достаточно взять значения k от 0 до $n-1$.

Поворот на угол $\dfrac{2\pi k}{n}$ $(k = 0, 1, 2, \ldots, n-1)$ обозначим через v_k, так что

$$V_n = \{v_0, v_1, v_2, \ldots, v_{n-1}\}.$$

При последовательном выполнении поворотов v_m и v_k получаем, очевидно, поворот на угол $\dfrac{2\pi(k+m)}{n}$. Если $k+m < n$, то получим $v_k v_m = v_{k+m}$. Если же $k+m \geqslant n$, то результирующий

поворот совпадает с поворотом v_{k+m-n} на угол $\dfrac{2\pi(k+m-n)}{n}$. Итак, можно написать

$$v_k v_m = v_r, \quad \text{где } r = \begin{cases} k+m, & \text{если } k+m < n; \\ k+m-n, & \text{если } k+m \geqslant n. \end{cases} \tag{4.5}$$

Выполнимость групповых аксиом в V_n проверяется так же, как и в C_n. Кроме того, сравнение совершенно идентичных правил умножения (4.4) и (4.5) убеждает нас в том, что обе группы, в сущности, одинаковы, отличаясь лишь происхождением элементов и их обозначением. Выразим эту мысль точнее, используя соответствие (очевидно, взаимно однозначное) между элементами двух групп, при котором корню z_k соответствует поворот v_k с тем же номером. Правила (4.4) и (4.5) показывают тогда, что при указанном соответствии произведению элементов $z_k z_m$ отвечает произведение (композиция) их образов $v_k v_m$. Такая согласованность соответствия с операциями и обеспечивает их идентичность. Соответствие, согласованное с групповыми операциями, называются изоморфизмом групп. Точное определение выглядит так:

Изоморфизмом групп G и G' называется такое взаимно однозначное соответствие f между ними, при котором

$$f(g_1 g_2) = f(g_1) f(g_2) \quad \textit{для любых} \quad g_1, g_2 \in G,$$

то есть образ произведения любых двух элементов равен произведению их образов.

Отметим, что обратное соответствие f^{-1}, очевидно, определяет изоморфизм между G' и G.

С точки зрения алгебры изоморфные группы неотличимы. Любой вывод относительно свойств групповой операции в одной из групп распространяется и на любую группу, ей изоморфную. Например, если группа абелева или циклическая, то таким же свойством будет обладать изоморфная ей группа. Конкретная природа элементов, которая может быть совершенно различной в двух группах, не имеет при этом значения. Полное отвлечение от неё,

столь типичное для алгебры и придающее ей абстрактный характер, — существенная, но не единственная черта понятия изоморфизма. Другая его черта, на первый взгляд противоположная первой, а на самом деле её дополняющая, состоит в том, что с помощью понятия изоморфизма можно строить удобные конкретные модели, облегчающие изучение абстрактных алгебраических структур. Так, нередко удается доказать, что абстрактная группа изоморфна, например, некоторой числовой группе или группе преобразований. Абстрактные элементы тогда заменяются вполне конкретными — числами, матрицами, подстановками и т. д. Изучение алгебраических свойств при этом облегчается. В то же время всякое заключение об этих свойствах, полученное в конкретной группе, справедливо и в изоморфной ей абстрактной группе. Вот наиболее простые теоремы, позволяющие сводить изучение операций в абстрактных группах к их изучению в числовых группах и группах преобразований.

1. *Всякая бесконечная циклическая группа изоморфна аддитивной группе $\mathbb{Z}$ целых чисел. Всякая конечная циклическая группа порядка n изоморфна аддитивной группе вычетов $\mathbb{Z}_n$ (или, что то же самое, мультипликативной группе C_n корней n-ой степени из единицы).*

2. *Всякая конечная группа изоморфна группе подстановок некоторого конечного множества (теорема Кэли).*

Докажем первое из этих утверждений. Пусть сначала $G = \langle g \rangle$ — бесконечная циклическая группа. Её элементами, напомним, являются всевозможные степени g^k, $k \in \mathbb{Z}$, причём все они различны. Рассмотрим соответствие $f\colon G \to \mathbb{Z}$, где

$$f(g^k) = k \quad \text{для всякого } k \in \mathbb{Z}.$$

Очевидно, что это соответствие взаимно однозначно. Кроме того,

$$f(g^k g^m) = f(g^{k+m}) = k + m = f(g^k) + f(g^m)$$

для любых элементов g^k и $g^m \in G$. Это и значит, что f — изоморфизм.

В другом случае, когда циклическая группа $G = \langle g \rangle$ имеет конечный порядок n, все её элементы находятся в списке (4.2). Соответствие $f\colon G \to \mathbb{Z}_n$ определим равенством

$$f(g^k) = \overline{k}, \quad k = 0, 1, \ldots, n-1.$$

Заметим, что $g^k g^m = g^{k+m} = g^r$, где $r = 0, 1, \ldots, n-1$ и $r \equiv$ $\equiv k + m \mod n$ (сравните это с формулами (4.4) и (4.5)). Но тогда и $\overline{k} + \overline{m} = \overline{r}$. Окончательно получаем, что

$$f(g^k g^m) = f(g^r) = \overline{r} = \overline{k} + \overline{m} = f(g^k) + f(g^m).$$

Это и значит, что f — изоморфизм. Соответствие $g^k \to z_k$, $k =$ $= 0, 1, \ldots, n-1$, как без труда проверит читатель, даёт изоморфизм между G и C_n.

Есть и ещё одна черта у понятия «изоморфизм». Часто при выяснении особенностей строения той или иной алгебраической структуры (в частности, группы) бывает полезно отобразить её изоморфно саму на себя. Такой изоморфизм называется автоморфизмом. Очевидно, что тождественное преобразование любой группы является её автоморфизмом, но группа может обладать и другими автоморфизмами. То, насколько велик у группы запас автоморфизмов, в большой степени определяет её внутреннюю структуру, а также характеризует присущие ей свойства симметрии. Приведём пример автоморфизма.

В циклической группе $G = \langle g \rangle$ порядка 8 рассмотрим отображение $f\colon G \to G$, задаваемое равенством $f(a) = a^3$ для всякого $a \in G$. Тот факт, что это отображение взаимно однозначно, легко усматривается из следующей таблицы (табл. 3), в которой перечислены элементы группы G и их образы.

Таблица 3.

a	$e = g^0$	g	g^2	g^3	g^4	g^5	g^6	g^7
$f(a) = a^3$	$e = g^0$	g^3	g^6	g	g^4	g^7	g^2	g^5

Очевидно, выполняется и другое требование: $f(ab) = (ab)^3 = a^3b^3 = f(a)f(b)$ для любых $a, b \in G$. Следовательно, f — автоморфизм. Попробуйте самостоятельно определить, является ли автоморфизмом этой группы отображение $a \to a^2$, а также перечислить все автоморфизмы циклической группы порядка 8. Более трудным является тот же вопрос для циклической группы произвольного порядка. Об одном из многих приложений понятия автоморфизма читатель узнает из следующей главы. Здесь обратим лишь внимание на такой факт.

Множество всех автоморфизмов любой группы G само является группой относительно операции умножения преобразований. Она так и называется — группа автоморфизмов группы G и обозначается $\operatorname{Aut} G$.

Чтобы доказать это утверждение, нужно проверить, что композиция двух автоморфизмов — снова автоморфизм, и что отображение, обратное автоморфизму, — автоморфизм. Пусть $f_1, f_2 \in \operatorname{Aut} G$ и $f = f_1 f_2$. Как уже известно, f является взаимно однозначным отображением группы G на себя. С другой стороны,

$$f(ab) = f_1 f_2(ab) = f_1(f_2(ab)) = \\ = f_1(f_2(a)f_2(b)) = f_1 f_2(a) \cdot f_1 f_2(b) = f(a)f(b).$$

Значит, композиция автоморфизмов — автоморфизм. Утверждение о том, что если $f \in \operatorname{Aut} G$, то $f^{-1} \in \operatorname{Aut} G$, проверяется аналогично. Например, для рассмотренной выше циклической группы порядка 8 группа автоморфизмов — абелева (но не циклическая) группа порядка 4. Попробуйте убедиться в этом самостоятельно.

Дополнения и задачи

1. Какими способами можно задать ту или иную группу? Один из них, пригодный для конечных групп, был указан А. Кэли. Он предложил рассматривать так называемую таблицу умножения

группы, которая очень похожа на обычную арифметическую таблицу умножения. Элементы группы располагаются в одном и том же порядке в верхней строке и в левом столбце таблицы, а внутри неё размещаются произведения элементов. Так, таблица умножения, или таблица Кэли для группы S_3 (обозначения элементов те же, что и во второй главе) выглядит так (табл. 4).

Таблица 4.

$\times$	σ_1	σ_2	σ_3	σ_4	σ_5	σ_6
σ_1	σ_1	σ_2	σ_3	σ_4	σ_5	σ_6
σ_2	σ_2	σ_1	σ_5	σ_6	σ_3	σ_4
σ_3	σ_3	σ_4	σ_1	σ_2	σ_6	σ_5
σ_4	σ_4	σ_3	σ_6	σ_5	σ_1	σ_2
σ_5	σ_5	σ_6	σ_2	σ_1	σ_4	σ_3
σ_6	σ_6	σ_5	σ_4	σ_3	σ_2	σ_1

Составьте таблицы умножения для других групп, например, для A_4, V_4. Подумайте над таким вопросом: какими отличительными признаками должна обладать таблица умножения группы? Такой способ задания операции годится, понятно, лишь для конечных групп, но даже и для них в случае больших порядков может оказаться малопригодным. О другом, более эффективном способе (использующем образующие элементы и определяющие соотношения) будет рассказано в главе 11. Для циклических групп этот метод фактически уже реализован.

2. При проверке того, является ли подмножество H группы G подгруппой, удобно пользоваться следующим критерием, доказать который предлагается читателю.

Для того, чтобы подмножество H являлось подгруппой группы G, необходимо и достаточно, чтобы выполнялись два условия:

1) *вместе с любыми двумя элементами h_1, h_2 подмножество H содержит их произведение h_1h_2;*

2) *вместе с любым элементом h подмножество H содержит обратный ему элемент h^{-1}.*

3. *Всякая подгруппа циклической группы сама является циклической.*

Докажите это утверждение, воспользовавшись следующими соображениями. Можно считать, что подгруппа $H \subseteq G$ нетривиальна, то есть не совпадает ни с G, ни с $\langle e \rangle$. Если g — образующий элемент группы G, то некоторые его степени принадлежат H, причём среди них обязательно найдется степень с положительным показателем (почему?). Пусть k — наименьший из положительных показателей со свойством $g \in H$. Тогда $H = \langle g^k \rangle$ (проверьте!).

4. Пусть $G = \langle g \rangle$ — циклическая группа порядка n. Как найти порядок произвольного её элемента g^k? Проделайте эксперимент в конкретной циклической группе, скажем, в C_8, составив таблицу порядков её элементов. Например, порядок элемента z_1^5 равен 8, а порядок элемента z_1^6 равен 4 и т. д. Вообще, справедливо следующее утверждение:

порядок элемента g^k равен n/d, где d — наибольший общий делитель чисел n и k. В частности, если k и n взаимно просты, то порядок элемента g^k равен n.

В последнем случае различные степени элемента g^k (их ровно n) исчерпывают всю группу G, и следовательно, g^k, как и g, служит её образующим элементом. Например, в C_8:

$$(z_1^5)^0 = 1, \quad (z_1^5)^1 = z_1^5, \quad (z_1^5)^2 = z_1^2, \quad (z_1^5)^3 = z_1^7,$$
$$(z_1^5)^4 = z_1^4, \quad (z_1^5)^5 = z_1, \quad (z_1^5)^6 = z_1^6, \quad (z_1^5)^7 = z_1^3.$$

5. В теории чисел рассматривается так называемая функция Эйлера. Её значение $\varphi(n)$ определяется как количество натуральных чисел, которые меньше n и при этом взаимно просты с n. Например, $\varphi(6) = 2$, так как среди чисел, меньших числа 6, и взаимно просты с ним лишь 1 и 5; $\varphi(8) = 4$, так как взаимно просты с 8 числа 1, 3, 5 и 7. Для всякого простого p, очевидно, $\varphi(p) = p - 1$. Обобщением малой теоремы Ферма является теорема Эйлера, которая формулируется так.

Для любого натурального a, взаимно простого с n, $a^{\varphi(n)}-1$ делится на n.

Попробуйте самостоятельно объяснить, почему малая теорема Ферма — частный случай теоремы Эйлера. Например, при $n=8$ и $a=5$ получаем, что $5^4-1=624$ действительно делится на 8. Можно предложить следующий план доказательства теоремы Эйлера (детали — на усмотрение читателя):

а) показать, что множество $\mathbb{Z}_n^*$ обратимых элементов кольца $\mathbb{Z}_n$ образуют группу относительно умножения;

б) убедиться, что порядок этой группы равен $\varphi(n)$;

в) провести те же рассуждения, что и при доказательстве малой теоремы Ферма.

6. Ещё одним примером применения групповых методов в теории чисел может служить доказательство следующей теоремы.

Если p — простое, то $(p-1)!+1$ делится на p (теорема Вильсона).

Это один из наиболее изящных результатов элементарной теории чисел. Утверждение теоремы равносильно следующему равенству для классов вычетов в $\mathbb{Z}_p$

$$\overline{1}\cdot\overline{2}\cdot\overline{3}\cdot\ldots\cdot\overline{(p-1)}+\overline{1}=\overline{0}.$$

Это равенство с учётом того, что $\overline{p-1}=-\overline{1}$, можно переписать в виде

$$\overline{2}\cdot\overline{3}\cdot\ldots\cdot\overline{(p-2)}=\overline{1}. \tag{4.6}$$

Можно считать, что $p>3$, так как при $p=2$ и при $p=3$ утверждение очевидно. Так как в $\mathbb{Z}_p$ есть только два класса, квадрат каждого из которых равен $\overline{1}$ (это $\overline{1}$ и $\overline{p-1}$), то множители в левой части равенства (4.6) разбиваются на пары взаимно обратных классов, что и доказывает равенство (4.6).

7. Опишем все подгруппы циклической группы $G=\langle g\rangle$ порядка n. Согласно теореме Лагранжа, порядок такой подгруппы является делителем числа n. Пусть d делит n. Рассмотрим в G подмножество $H=\{x \mid x^d=e\}$. Очевидно, что H — подгруппа,

и, как было доказано ранее, циклическая. Если h — образующий элемент подгруппы H, то его порядок равен d (почему?). Следовательно, H — циклическая подгруппа порядка d. Если H' — также подгруппа порядка d, то $(h')^d = e$ для любого её элемента h', и значит, $H' \subseteq H$. Но поскольку $|H'| = |H| = d$, то $H' = H$. Мы доказали следующее утверждение.

Для всякого натурального числа d, являющегося делителем порядка n циклической группы G, существует единственная подгруппа H порядка d. Этими подгруппами исчерпываются все подгруппы данной циклической группы.

Используя это утверждение, опишите все подгруппы циклических групп порядка 7, 10, 12, 16 и укажите образующие элементы этих подгрупп.

8. Как уже говорилось, множество $\mathbb{Z}_p^*$ ненулевых классов вычетов по простому модулю p является группой (*мультипликативная группа вычетов по модулю p*). Исчерпывающие сведения о её строении даёт следующая теорема[1].

Мультипликативная группа вычетов $\mathbb{Z}_p^*$ — циклическая группа порядка $p-1$.

Например, непосредственно проверяется, что $\mathbb{Z}_{11}^* = \langle \overline{2} \rangle$.

Для доказательства теоремы нужно убедиться, что в группе $\mathbb{Z}_p^*$ найдется элемент порядка $p-1$, он и будет образующим. План рассуждений таков. Выбираем элемент $a \in \mathbb{Z}_p^*$ наибольшего возможного порядка k. Далее доказываем, что порядок любого другого элемента $b \in \mathbb{Z}_p^*$ является делителем k. На этом этапе понадобится вспомогательное утверждение: если в абелевой группе порядки двух элементов взаимно просты, то порядок их произведения равен произведению их порядков.

Далее, каждый элемент $b \in \mathbb{Z}_p^*$ является корнем уравнения $x^k = \overline{1}$, то есть

$$x^k - \overline{1} = \overline{0}.$$

[1] Этот факт в несколько иной форме был известен уже Гауссу и был использован им при решении задачи о построении правильного n-угольника (см. следующую главу).

Наконец, покажем, что многочлен $x^k - \overline{1}$ можно записать в виде

$$x^k - \overline{1} = \prod_{b \in \mathbb{Z}_p^*} (x - b) = (x - \overline{1})(x - \overline{2}) \dots (x - \overline{(p-1)}),$$

откуда и следует, что $k = p - 1$.

Определите, какова связь этой теоремы с малой теоремой Ферма. Найдите образующие элементы групп $\mathbb{Z}_5^*$, $\mathbb{Z}_7^*$, $\mathbb{Z}_{13}^*$, $\mathbb{Z}_{17}^*$.

9. Зафиксируем некоторое натуральное число k и рассмотрим в группе $\mathbb{Z}_p^*$ все степени с показателем k

$$\overline{1}^k = \overline{1}, \overline{2}^k, \dots, \overline{r}^k, \dots, \overline{(p-1)}^k. \tag{4.7}$$

Классы указанного вида называют k-степенными вычетами. Каковы те элементы группы $\mathbb{Z}_p^*$, которые являются k-степенными вычетами? Этот вопрос равнозначен следующему: какие остатки могут получиться при делении степеней a^k, $a \in \mathbb{N}$, на простое число p? В некоторых случаях ответ понятен. Например, если $k = p - 1$, то все элементы вида (4.7) совпадают с $\overline{1}$, что непосредственно следует из теоремы Ферма. Ситуация прямо противоположна, если k взаимно просто с $p - 1$. Тогда степени (4.7) исчерпывают всю группу $\mathbb{Z}_p^*$ (попробуйте это доказать). Так, в группе $\mathbb{Z}_{11}^*$

$$\begin{array}{lllll} \overline{1}^3 = \overline{1}, & \overline{2}^3 = \overline{8}, & \overline{3}^3 = \overline{5}, & \overline{4}^3 = \overline{9}, & \overline{5}^3 = \overline{4}, \\ \overline{6}^3 = \overline{7}, & \overline{7}^3 = \overline{2}, & \overline{8}^3 = \overline{6}, & \overline{9}^3 = \overline{3}, & \overline{10}^3 = \overline{10}. \end{array}$$

Между этими двумя крайностями масса промежуточных случаев, и самый простой из них — случай квадратичных вычетов — соответствует $k = 2$.

Квадратичные вычеты были предметом интереса многих математиков. Уже Эйлер установил ряд их свойств. Вот некоторые из них.

1) При нечётном p ровно половина элементов мультипликативной группы $\mathbb{Z}_p^*$ (то есть $\dfrac{p-1}{2}$ элементов) является квадратичными вычетами.

2) Элемент $x \in \mathbb{Z}_p^*$ является квадратичным вычетом тогда и только тогда, когда $x^{\frac{p-1}{2}} = \overline{1}$ (критерий Эйлера квадратичности вычета).

Например, в $\mathbb{Z}_{11}^*$ квадратичных вычетов пять. Это классы

$$\overline{1} = \overline{1}^2, \qquad \overline{4} = \overline{2}^2, \qquad \overline{9} = \overline{3}^2, \qquad \overline{5} = \overline{4}^2, \qquad \overline{3} = \overline{5}^2,$$

и для каждого из них $x^5 = \overline{1}$ (проверьте!).

Чтобы доказать свойство 1), надо лишь заметить, что ровно два элемента из $\mathbb{Z}_p^*$, а именно, $\overline{r}$ и $-\overline{r} = \overline{(n-r)}$, служат решениями уравнения $x^2 = \overline{r}^2$, и потому число квадратичных вычетов вдвое меньше числа элементов $\mathbb{Z}_p^*$.

Критерий Эйлера в одну сторону очевиден. Именно, если $x = \overline{r}^2$, то

$$x^{(p-1)/2} = \overline{r}^{p-1} = \overline{1}.$$

Для доказательства обратного утверждения воспользуйтесь цикличностью группы $\mathbb{Z}_p^*$ и правилом отыскания порядка элемента циклической группы, которое сформулировано в дополнении 4. Эйлер открыл и гораздо более глубокие закономерности, которым подчиняются квадратичные вычеты. Особенно знаменита его гипотеза, называемая квадратичным законом взаимности. Её справедливость была установлена Гауссом, им было дано восемь (!) различных доказательств закона взаимности. Чтобы сформулировать этот закон, удобно использовать так называемый символ Лежандра $\left(\frac{a}{p}\right)$. Пусть a и p взаимно просты. Тогда полагают $\left(\frac{a}{p}\right) = 1$, если a является квадратичным вычетом, и $\left(\frac{a}{p}\right) = -1$, если a не является квадратичным вычетом. Квадратичный закон взаимности формулируется так:

Если p и q — различные нечетные простые числа, то

$$\left(\frac{p}{q}\right) \cdot \left(\frac{q}{p}\right) = (-1)^{\frac{p-1}{2} \cdot \frac{q-1}{2}}.$$

Интересный рассказ об истории «золотой теоремы», как именовал квадратичный закон сам Гаусс, вы можете прочесть в книге [10].

10. Убедитесь, что перечисленные ниже группы не являются циклическими:

а) аддитивная группа $\mathbb{Q}$ рациональных чисел;

б) симметрическая группа S_n при $n > 2$;

в) знакопеременная группа A_n при $n > 3$.

Для конечных групп, не являющихся циклическими, утверждение пункта 7 может оказаться неверным. Самым простым контрпримером служит группа A_4 порядка 12. Попробуйте доказать, что в ней нет подгруппы порядка 6.

11. Убедитесь, что при изоморфизме $f\colon G \to G'$

а) единица e группы G переходит в единицу e' группы G';

б) порядок любого элемента $g \in G$ совпадает с порядком его образа $g' = f(g)$;

в) если g — образующий элемент группы G, то его образ $g' = f(g)$ — образующий элемент группы G';

г) если $\varphi(g_1, g_2, \ldots, g_k) = e$ — некоторое соотношение, выполненное для элементов группы G, то такое же соотношение

$$\varphi(f(g_1), f(g_2), \ldots, f(g_k)) = e$$

выполняется для их образов.

ГЛАВА 5

2000 ЛЕТ СПУСТЯ

Одна из задач, приковавших к себе внимание юного Гаусса, имела столь давнее происхождение, что попытки её решения казались совершенно безнадежными. Вот что пишет об этом сам Гаусс:

> *«Всякому начинающему геометру известно, что можно геометрически (то есть циркулем и линейкой) строить разные правильные многоугольники, а именно, треугольник, пятиугольник, пятнадцатиугольник и те, которые получаются из каждого из них путём удвоения числа его сторон. Это было известно во времена Евклида, и, как кажется, с тех пор было распространено убеждение, что дальше область элементарной геометрии не распространяется: по крайней мере, я не знаю удачной попытки распространить её в эту сторону».*

И действительно, какие правильные многоугольники, кроме перечисленных и квадрата, можно строить циркулем и линейкой, было в равной мере загадкой как для современников Евклида, так и для современников Гаусса. Разница, быть может, была лишь в том, что к XIX веку как-то обозначилась возможность алгебраического решения этой древней геометрической задачи. Мы уже упоминали о том, что с помощью циркуля и линейки можно строить кроме рациональных величин лишь так называемые

квадратичные иррациональности, другими словами, такие величины, которые выражаются через единицу масштаба с помощью конечного числа рациональных операций и операций извлечения квадратного корня. С другой стороны, построение правильного n-угольника равносильно делению единичной окружности на n равных частей. Тогда точкам деления соответствуют комплексные числа z_k, являющиеся корнями n-ой степени из 1. Таким образом, вопрос сводится к тому, при каких n все корни n-ой степени из 1 являются квадратичными иррациональностями. Можно сказать и иначе: при каких n корни уравнения $z^n - 1 = 0$ можно получить, последовательно решая некоторые квадратные уравнения?

Но ведь ключи к подобным вопросам уже подбирали современники Гаусса. Как помнит читатель, Лагранж выяснял причины, по которым отыскание корней алгебраического уравнения может быть сведено к решению вспомогательных уравнений меньших степеней. И он впервые заметил, что существование таких вспомогательных уравнений связано с некоторыми свойствами симметрии их корней. Однако Лагранжу не удалось полностью осуществить свою программу. Он рассматривал уравнения с произвольными коэффициентами и каждый раз о свойствах симметрии и частичной симметрии судил по группе такого общего уравнения. Лишь позднее выяснилось, что с каждым конкретным уравнением можно связать его собственную группу, которая учитывает специфику данного уравнения и его корней и не обязательно совпадает с симметрической группой S_n. Суть в том, что между корнями конкретного уравнения кроме соотношений, задаваемых формулами Виета, могут быть и иные, не вытекающие из этих формул, соотношения. Оказалось, что решающую роль в изучении алгебраического уравнения играет группа тех подстановок, которые сохраняют все соотношения, имеющиеся между корнями. Об этом открытии, честь которого принадлежит Эвариcту Галуа, разговор ещё предстоит. Здесь же мы попытаемся понять, как сходную идею реализовал Гаусс применительно к уравнению $z^n - 1 = 0$. Мы уже знаем, что множество его корней относительно умножения образует циклическую группу. Подстановки, сохраняющие все соотношения между корнями, должны

быть автоморфизмами этой группы. Нетрудно описать все такие автоморфизмы, но особенно просто это сделать, если $n = p$ — простое число. Как раз с этим случаем, как объясняется в дополнении, достаточно разобраться в задаче о построении правильного n-угольника.

Группа C_p корней p-ой степени из 1 имеет простой порядок, а значит, любой её неединичный элемент

$$z_1,\ z_2,\ \ldots,\ z_{p-1} \tag{5.1}$$

является образующим этой группы. Всякий автоморфизм φ циклической группы однозначно определяется заданием своего значения на образующем. Действительно, если, например,

$$\varphi(z_1) = z_k, \tag{5.2}$$

то $\varphi(z_1^l) = z_k^l$, а значит, известны и образы всех элементов группы C_p. С другой стороны, для любого $k = 1, 2, \ldots, p-1$, существует автоморфизм φ, определенный формулой (5.2), именно потому, что каждый из элементов (5.1) является образующим. Следовательно, число автоморфизмов равно $p-1$. Мы уже отмечали в главе 2 аналогию между умножением корней из 1 и сложением классов вычетов. Не так заметна, но не менее важна другая нить, связывающая группу C_n и кольцо классов вычетов $\mathbb{Z}_n$. Оказывается, группа автоморфизмов $\operatorname{Aut} C_p$ изоморфна группе ненулевых классов вычетов $\mathbb{Z}_p^*$. Действительно, обозначим через σ_k автоморфизм, переводящий корень z_1 в z_k

$$\sigma_k(z_1) = z_k = z_1^k \qquad (k = 1, 2, \ldots, p-1),$$

и сопоставим этому автоморфизму класс $\overline{k} \in \mathbb{Z}_p^*$. Ясно, что такое соответствие $\sigma_k \to \overline{k}$ определяет взаимно однозначное соответствие между группами $\operatorname{Aut} C_p$ и $\mathbb{Z}_p^*$. Кроме того, для любых двух автоморфизмов σ_k, σ_m

$$\sigma_k\, \sigma_m(z_1) = \sigma_k(z_1^m) = z_1^{km} = \sigma_{km}(z_1), \tag{5.3}$$

поэтому

$$\sigma_k \sigma_m = \sigma_{km}. \tag{5.4}$$

В равенствах (5.3) и (5.4) произведение km понимается, конечно, в смысле умножения по модулю p. Но тогда произведению автоморфизмов σ_k, σ_m будет соответствовать произведение классов вычетов, следовательно, отображение $\sigma_k \to \overline{k}$ является изоморфизмом.

Отметим ещё такое обстоятельство. Многочлен $z^p - 1$ разлагается на множители

$$z^p - 1 = (z-1)(z^{p-1} + z^{p-2} + \ldots + z + 1).$$

Понятно, что все корни, отличные от 1, то есть числа (5.1), являются корнями второго сомножителя.

В общих чертах дальнейший план рассуждений Гаусса таков. Используя свойства полей вычетов $\mathbb{Z}_p$, он устанавливает, что автоморфизмы группы корней p-ой степени сами образуют циклическую группу G порядка $p-1$. Пусть

$$p - 1 = p_1 p_2 \ldots p_{t-1} p_t$$

— разложение порядка этой группы на простые множители (не обязательно различные). Как следует из дополнения 7 к главе 4, в группе G имеется цепочка подгрупп

$$G \supset G_1 \supset G_2 \supset \ldots \supset G_{t-1} \supset G_t = \{e\}$$

таких, что

$$|G : G_1| = p_1; \quad |G_1 : G_2| = p_2; \quad \ldots \quad |G_{t-1} : G_t| = |G_t| = p_t.$$

Гаусс составляет такие выражения $f_1, f_2, \ldots, f_{t-1}$, зависящие от корней, которые симметричны только относительно подстановок из подгрупп $G_1, G_2, \ldots, G_{t-1}$, то есть не меняются при указанных подстановках. Для выражения f_1, симметричного относительно G_1, устанавливается, как и в теории Лагранжа, что при действии всех подстановок из G_1 оно принимает p_1 различных значений, являющихся корнями уравнения степени p_1 с рациональными коэффициентами. В свою очередь, выражение f_2 принимает p_2 различных значений при действии всех подстановок из

подгруппы G_2, и эти значения служат корнями уравнения степени p_2. Правда, коэффициенты последнего уравнения уже не рациональны, но они рационально выражаются через корни предыдущего уравнения. Далее всё аналогично. Гаусс показывает, что последовательно решая указанные уравнения степеней $p_1, p_2, \ldots$ $\ldots, p_{t-1}$, можно дойти в конечном итоге до вычисления корней p-ой степени из 1, выразив их тем самым через корни промежуточных уравнений. В этих рассуждениях немаловажную роль играла применённая Гауссом иная, чем прежде, нумерация корней, вытекающая не из геометрической их симметрии, а из симметрии алгебраической, — симметрии относительно группы автоморфизмов G. Каковы же детали этих рассуждений? Проследим их для частного случая $p = 17$. Именно он прежде всего и интересовал Гаусса, ведь правильный семнадцатиугольник был многоугольником с наименьшим числом сторон, для которого ещё со времён Евклида оставалась неизвестной возможность геометрического построения традиционными инструментами древних — циркулем и линейкой (см. дополнение 2).

При $p = 17$ группа автоморфизмов $G = \operatorname{Aut} C_{17}$ имеет порядок 16 и изоморфна мультипликативной группе $\mathbb{Z}_{17}^*$. Прямой проверкой можно установить, что класс $\overline{3}$ является её образующим. Так как группы $\mathbb{Z}_{17}^*$ и G изоморфны, отсюда следует, что автоморфизм σ такой, что $\sigma(z_1) = z_1^3$, является образующим группы автоморфизмов. Таким образом, все элементы группы G имеют вид σ^k, где $k = 0, 1, \ldots 15$. Так как

$$\sigma^2(z_1) = \sigma(z_1^3) = z_1^{3^2}; \quad \sigma^3(z_1) = \sigma(\sigma^2(z_1)) = \sigma(z_1^{3^2}) = z_1^{3^3}; \quad \ldots$$
$$\sigma^k(z_1) = z_1^{3^k},$$

то мы видим, что элементы группы C_{17}, отличные от 1, имеют вид

$$z_1, z_1^3, z_1^{3^2}, \ldots, z_1^{3^k}, \ldots, z_1^{3^{15}}.$$

Тот факт, что они последовательно переходят друг в друга при повторном применении автоморфизма σ, делает естественной другую нумерацию корней. Гаусс полагает $\zeta_0 = z_1$, $\zeta_1 = z_1^3$, $\zeta_2 = z_1^{3^2}$, и вообще,

$$\zeta_k = z_1^{3^k}, \quad k = 0, 1, \ldots, 15. \tag{5.5}$$

При этом

$$\sigma(\zeta_k) = \zeta_{k+1}, \qquad \sigma^l(\zeta_k) = \zeta_{k+l}, \tag{5.6}$$

где индексы суммируются по модулю 16.

Обратимся теперь к собственным подгруппам группы $G = \operatorname{Aut} C_{17}$. Их в данном случае три:

$$G_1 = \langle \sigma^2 \rangle, \quad G_2 = \langle \sigma^4 \rangle, \quad G_3 = \langle \sigma^8 \rangle.$$

Порядки этих подгрупп соответственно равны 8, 4, 2.

Нетрудно указать функции от корней, симметричные (или инвариантные) относительно перечисленных подгрупп. Например, относительно подстановок из G_1 симметричными будут следующие две суммы

$$\begin{aligned} \eta_0 &= \zeta_0 + \zeta_2 + \zeta_4 + \zeta_6 + \zeta_8 + \zeta_{10} + \zeta_{12} + \zeta_{14}, \\ \eta_1 &= \zeta_1 + \zeta_3 + \zeta_5 + \zeta_7 + \zeta_9 + \zeta_{11} + \zeta_{13} + \zeta_{15}. \end{aligned} \tag{5.7}$$

Понятно, по какому принципу они построены. Так, слагаемыми первой суммы согласно (5.6) являются все те образы элемента ζ_0, которые получаются при повторном применении автоморфизма σ^2. Совершенно аналогично устроена и вторая сумма. Это и обеспечивает инвариантность выражений η_0 и η_1 относительно автоморфизмов из G_1 — их действие сводится просто к циклической перестановке слагаемых. Действие же всех других автоморфизмов из G переводит η_0 в η_1 и обратно (почему?).

Теперь можно, по аналогии с рассуждениями Лагранжа, высказать догадку, что η_0 и η_1 являются корнями квадратного уравнения с рациональными коэффициентами. Это действительно так, и чтобы в этом убедиться, докажем, что $\eta_0 + \eta_1$ и $\eta_0\eta_1$ рациональны. Для суммы $\eta_0 + \eta_1$ это совсем просто. В самом деле,

$$\eta_0 + \eta_1 = \sum_{k=0}^{15} \zeta_k. \tag{5.8}$$

Вспомним, что ζ_k являются корнями уравнения

$$z^{16} + z^{15} + \ldots + z + 1 = 0, \tag{5.9}$$

поэтому по формулам Виета сумма (5.8) равна -1:

$$\eta_0 + \eta_1 = -1. \tag{5.10}$$

Сложнее разобраться с произведением

$$\eta_0\eta_1 = \zeta_0\zeta_1 + \zeta_0\zeta_3 + \ldots + \zeta_{14}\zeta_{15}. \tag{5.11}$$

Сгруппируем слагаемые следующим образом

$$\begin{aligned}\eta_0\eta_1 = {} & (\zeta_0\zeta_1 + \zeta_1\zeta_2 + \zeta_2\zeta_3 + \ldots + \zeta_{14}\zeta_{15} + \zeta_{15}\zeta_0) + \\ & + (\zeta_0\zeta_3 + \zeta_1\zeta_4 + \zeta_2\zeta_5 + \ldots + \zeta_{14}\zeta_1 + \zeta_{15}\zeta_2) + \\ & + (\zeta_0\zeta_5 + \zeta_1\zeta_6 + \zeta_2\zeta_7 + \ldots + \zeta_{14}\zeta_3 + \zeta_{15}\zeta_4) + \\ & + (\zeta_0\zeta_7 + \zeta_1\zeta_8 + \zeta_2\zeta_9 + \ldots + \zeta_{14}\zeta_5 + \zeta_{15}\zeta_6).\end{aligned} \tag{5.12}$$

Принцип группировки достаточно ясен: в каждой группе индекс второго сомножителя отличается по модулю 16 от индекса первого на фиксированное число — в первой группе на 1, далее на 3, 5 и 7. При этом будут охвачены все слагаемые из (5.11). Разберёмся с первой из групп

$$\begin{aligned}&\zeta_0\zeta_1 + \zeta_1\zeta_2 + \zeta_2\zeta_3 + \ldots + \zeta_{14}\zeta_{15} + \zeta_{15}\zeta_0 = \\ &= \zeta_0\zeta_1 + \sigma(\zeta_0\zeta_1) + \ldots + \sigma^{14}(\zeta_0\zeta_1) + \sigma^{15}(\zeta_0\zeta_1) = \sum_{k=0}^{15}\sigma^k(\zeta_0\zeta_1).\end{aligned} \tag{5.13}$$

Так как $\zeta_0\zeta_1 = \zeta_m$ для некоторого m, то сумма (5.13) преобразуется следующим образом

$$\sum_{k=0}^{15}\sigma^k(\zeta_m) = \sum_{k=0}^{15}\zeta_{k+m}.$$

Слагаемые получившейся суммы, очевидно, исчерпывают все корни уравнения (5.9), поэтому она, а вместе с ней рассматриваемое выражение, равны -1. Аналогичные рассуждения можно применить и к остальным суммам в (5.12) — все они также равны -1, но тогда

$$\eta_0\eta_1 = -4. \tag{5.14}$$

Равенства (5.10) и (5.14) приводят нас к желаемому результату — числа η_0, η_1 являются корнями квадратного уравнения $z^2 + z - 4 = 0$. При этом несложно показать, что η_0 равно большему из корней, а η_1 — меньшему

$$\eta_0 = \frac{-1+\sqrt{17}}{2}, \qquad \eta_1 = \frac{-1-\sqrt{17}}{2}. \tag{5.15}$$

На следующем этапе рассматриваются суммы, сохраняющиеся при действии автоморфизмов из группы $G_2 = \langle \sigma^4 \rangle$. Таковыми будут суммы корней

$$\nu_0 = \zeta_0 + \zeta_4 + \zeta_8 + \zeta_{12}, \qquad \nu_1 = \zeta_1 + \zeta_5 + \zeta_9 + \zeta_{13},$$
$$\nu_2 = \zeta_2 + \zeta_6 + \zeta_{10} + \zeta_{14}, \qquad \nu_3 = \zeta_3 + \zeta_7 + \zeta_{11} + \zeta_{15}.$$

Очевидно, $\nu_0 + \nu_2 = \eta_0$, $\nu_1 + \nu_3 = \eta_1$. Можно показать (предоставляем это читателю), что $\nu_0\nu_2 = \nu_1\nu_3 = -1$. Значит, пары чисел (ν_0, ν_2) и (ν_1, ν_3) являются соответственно корнями квадратных уравнений

$$x^2 - \eta_0 x - 1 = 0, \qquad x^2 - \eta_1 x - 1 = 0.$$

Решая эти уравнения с учётом равенств (5.15), получим

$$\nu_0 = \frac{\sqrt{17} - 1 + \sqrt{34 - 2\sqrt{17}}}{4}, \quad \nu_1 = \frac{-\sqrt{17} - 1 + \sqrt{34 + 2\sqrt{17}}}{4},$$
$$\nu_2 = \frac{\sqrt{17} - 1 - \sqrt{34 - 2\sqrt{17}}}{4}, \quad \nu_3 = \frac{-\sqrt{17} - 1 - \sqrt{34 + 2\sqrt{17}}}{4}.$$

На заключительном этапе рассматриваются суммы

$$\mu_0 = \zeta_0 + \zeta_8, \quad \mu_1 = \zeta_1 + \zeta_9, \quad \mu_2 = \zeta_2 + \zeta_{10}, \quad \mu_3 = \zeta_3 + \zeta_{11},$$
$$\mu_4 = \zeta_4 + \zeta_{12}, \quad \mu_5 = \zeta_5 + \zeta_{13}, \quad \mu_6 = \zeta_6 + \zeta_{14}, \quad \mu_7 = \zeta_7 + \zeta_{15},$$

инвариантные относительно автоморфизмов из $G_3 = \langle \sigma^8 \rangle$. Рассмотрим μ_0 и μ_4. Для них выполняются соотношения $\mu_0 + \mu_4 = \nu_0$, $\mu_0\mu_4 = \nu_2$, и уравнение, которому они удовлетворяют, имеет вид $z^2 - \nu_0 z + \nu_2 = 0$. Число μ_0, как можно проверить, равно

большему его корню, так что после преобразований

$$\mu_0 = \frac{\nu_0 + \sqrt{\nu_0^2 - 4\nu_2}}{2} =$$

$$= \frac{\sqrt{17} - 1 + \sqrt{34 - 2\sqrt{17}} + 2\sqrt{17 + 3\sqrt{17} - \sqrt{170 + 38\sqrt{17}}}}{8}.$$

Из подобных формул находятся и все другие μ_i. По ним могут быть найдены и все корни ζ_k из 1. Однако, для нужд самого построения достаточно уже того, что величина μ_0 записана как квадратичная иррациональность.

Действительно, нетрудно проверить, что $\mu_0 = 2\cos\frac{\pi}{17}$. Зная величину $\cos\frac{\pi}{17}$, можно, наконец, построить этот «таинственный» семнадцатиугольник.

Решив задачу для семнадцатиугольника, Гаусс разобрался и с общей ситуацией, описав все правильные многоугольники, доступные для построения циркулем и линейкой. Действительно, подобное описанному выше сведение к квадратичным иррациональностям возможно, если $p - 1$ является степенью двойки. Значит, многоугольник с простым числом сторон p может быть построен тогда и только тогда, когда p есть число вида $2^k + 1$ (подробнее в дополнении 3). Для произвольного правильного n-угольника построение возможно в том и только в том случае, если $n = 2^m p_1 p_2 \ldots p_s$ и все p_i — различные простые числа указанного выше вида: $p_i = 2^{k_i} + 1$. Гаусс высказал этот критерий без доказательства, предостерегая читателей от тщетных попыток искать построение в каких-либо иных случаях. Он добавляет, что может со всей строгостью его доказать, и лишь границы сочинения не позволяют этого сделать.

Дополнения и задачи

1. Следуя рассуждениям Гаусса, указать способ построения правильного пятиугольника.

2. Из критерия Гаусса следует невозможность построения циркулем и линейкой правильного семиугольника. Вообще, для

$n \leqslant 100$ имеются лишь 22 значения n (какие?), для которых правильный n-угольник может быть построен циркулем и линейкой.

3. Доказать, что если число вида $2^k + 1$ является простым, то k является степенью двойки.

Таким образом, простые числа. участвующие в формулировке критерия Гаусса, должны иметь вид $2^{2^r} + 1$. Числа указанного вида рассматривал ещё Ферма, ошибочно полагая, что они всегда простые. Однако Эйлер обнаружил, что получающееся при $r = 5$ число 4 294 967 297 имеет своим делителем число 641. Более того, выяснилось, что и среди последующих чисел Ферма (так называют числа вида $2^{2^r} + 1$) сплошь и рядом встречаются составные. Возникло даже предположение, что множество простых чисел Ферма конечно. Справедливость этой гипотезы означала бы, что имеется лишь конечное число правильных n-угольников с нечётным числом сторон, допускающих построение циркулем и линейкой.

4. Критерий Гаусса означает, что окружность может быть разделена циркулем и линейкой на n равных частей тогда и только тогда, когда $n = 2^m p_1 p_2 \dots p_s$, где все p_i — различные простые числа Ферма. Полное доказательство этого факта опирается на теорию Галуа, представление о которой даёт глава 14. Здесь мы убедимся лишь в том, что если n — число указанного вида, то построение возможно (при этом случай простого n будем считать уже исчерпанным). Доказательство опирается на следующие несложные утверждения.

1) Из возможности деления окружности на h равных частей следует возможность её деления на $2^m h$ равных частей для любого натурального m.

2) Из возможности деления окружности на p_1 равных частей и на p_2 равных частей, где p_1 и p_2 взаимно просты, следует возможность её деления на $p_1 p_2$ равных частей.

Первое утверждение сразу вытекает из того, что любой угол при помощи циркуля и линейки можно разделить пополам. Для проверки второго утверждения нам придётся сослаться на сравнительно элементарный факт, который, впрочем, доказывается в главе 6 (дополнение 6).

Если p_1 и p_2 взаимно просты, то существуют такие целые числа s_1 и s_2, что

$$s_1 p_1 + s_2 p_2 = 1.$$

Пусть теперь c — длина окружности, и c/p_1, c/p_2 — длины дуг, получающихся при делении окружности на p_1 и, соответственно, p_2 равных частей. Понятно, что тогда легко может быть построена дуга длины

$$s_1 \cdot \frac{c}{p_2} + s_2 \cdot \frac{c}{p_1} = \frac{(s_1 p_1 + s_2 p_2)\, c}{p_1 p_2} = \frac{c}{p_1 p_2},$$

и следовательно, окружность можно разделить на $p_1 p_2$ равных частей.

Если $n = 2^m p_1 p_2 \dots p_s$, то ясно, что первое утверждение сводит задачу к делению окружности на $p_1 p_2 \dots p_s$ равных частей. Если $s = 2$, то возможность деления окружности на $p_1 p_2$ равных частей вытекает из второго утверждения. Если же s произвольно ($s > 2$), то дело довершает очевидная индукция.

ГЛАВА 6

ТОЛЬКО ЧИСЛА

Каждый школьник, постепенно знакомящийся с миром чисел, как бы в миниатюре воспроизводит многовековую историю развития представлений о числе. В общих чертах она отражена в следующей последовательности наиболее употребительных в математике и её приложениях числовых систем.

$\mathbb{N}$— множество натуральных чисел,

$\mathbb{Z}$— множество целых чисел,

$\mathbb{Q}$— множество рациональных чисел,

$\mathbb{R}$— множество действительных чисел,

$\mathbb{C}$— множество комплексных чисел.

Греческие математики и философы любили повторять изречение «Всё есть число», но долгое время им не было известно никаких иных чисел, кроме натуральных и их отношений. Отрицательные числа были «узаконены» в математике значительно позднее, чем дроби. Велико же было изумление греков, когда они обнаружили, что этих чисел недостаточно даже для измерения длин. Например, отношение длины диагонали квадрата к длине его стороны, как было выяснено последователями Пифагора, невозможно записать никакой дробью. Мы выражаем этот факт словами, что $\sqrt{2}$ есть число иррациональное. С открытия иррациональностей началась эпоха недоверия к числам в античной математике, что и привело к полному господству в ней геометрии. Любые задачи

в ту пору стремились сформулировать на геометрическом языке и решать геометрически, то есть построением. При этом особая роль приписывалась построениям с помощью циркуля и линейки. Мы уже знаем, что указанными инструментами можно строить лишь квадратичные иррациональности. В то же время, корни произвольных алгебраических уравнений сплошь и рядом не являются таковыми. Отсюда, как уже отмечалось в главе 1, следует неразрешимость многих задач на построение.

Со временем выяснилось, что существуют числа (например, π), которые не являются алгебраическими, то есть корнями уравнений

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

с рациональными коэффициентами a_k. Такие числа называют трансцендентными.

Потребовался длинный путь развития математики, чтобы внести ясность в представления о числе. Достаточно сказать, что учение о квадратичных иррациональностях содержится уже в «Началах» Евклида (III век до н. э.), а первое доказательство существования трансцендентных чисел было дано лишь в 1844 году французским математиком Лиувиллем (1809–1882). Добавим к этому, что само математически корректное понятие действительного числа, не опирающееся на геометрическую интуицию, было введено ещё позднее, в семидесятые годы девятнадцатого столетия.

Какие бы числа мы не рассматривали, мы не мыслим их в отрыве от возможности оперировать ими по законам арифметики. Речь, естественно, идёт об операциях сложения, вычитания, умножения, деления. Можно заметить, однако, что перечисленные выше числовые системы не одинаково ведут себя по отношению к этим четырём арифметическим действиям. Так, в области натуральных чисел мы свободно можем выполнять сложение и умножение, в то время как обратные к ним действия не всегда выполнимы. Не существует, например, натурального числа, являющегося разностью чисел 2 и 3, как нет и натурального числа, которое являлось бы из отношением. В множестве всех целых чисел вычитание выполнимо уже без всяких ограничений.

Расширив множество $\mathbb{Z}$ до множества $\mathbb{Q}$ всех рациональных чисел, мы получаем возможность во всех случаях (кроме деления на 0) выполнять и четвёртую арифметическую операцию. В этом смысле множество $\mathbb{Q}$ обладает замкнутостью относительно четырёх действий арифметики. Подобным свойством обладают и две другие, более обширные числовые системы $\mathbb{R}$ и $\mathbb{C}$. Числовое множество, замкнутое относительно всех арифметических операций и содержащее хотя бы один ненулевой элемент, принято называть *числовым полем*. Особое название имеется и в случае, когда множество замкнуто лишь относительно первых трёх операций. В этом случае его называют *числовым кольцом*. Таким образом, каждое из множеств $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ образует числовое поле: *поле рациональных, действительных и комплексных чисел*. В отличие от них множество $\mathbb{Z}$ поля не образует, но является числовым кольцом — это так называемое *кольцо целых чисел*.

Как мы вскоре увидим, числовые системы, перечисленные выше — это лишь немногие из примеров в большом многообразии числовых колец и полей. Интересно отметить особую роль, которую играют поля $\mathbb{Q}$ и $\mathbb{C}$. Оказывается,

> *поле рациональных чисел является наименьшим из числовых полей, а поле комплексных чисел — наибольшим.*

Смысл второго утверждения прояснится позднее, а первое мы легко можем доказать уже сейчас. Для этого заметим прежде всего, что в любом числовом поле $\mathbb{F}$ обязательно должны содержаться 0 и 1. В самом деле, если f — произвольный ненулевой элемент поля $\mathbb{F}$, то разность $f - f$, равная 0, и отношение f/f, равное 1, по определению числового поля должны ему принадлежать. Но то же самое верно и для любого натурального числа n, поскольку $n = 1 + 1 + \ldots + 1$ (сумма n слагаемых), и даже для любого целого, так как $-n = 0 - n$. Наконец, из того, что целые числа m и n ($n \neq 0$) содержатся в поле $\mathbb{F}$ следует, что этим же свойством обладает и их отношение, то есть всякая дробь m/n. Итак, всякое числовое поле целиком содержит поле рациональных чисел. Это и означает, что $\mathbb{Q}$ — наименьшее из числовых полей.

Другие числовые поля могут быть получены из $\mathbb{Q}$ путём присоединения к нему иррациональных чисел. Характер этого процесса иллюстрирует пример поля, получающегося присоединением к $\mathbb{Q}$ числа $\sqrt{2}$. Это числовое поле, содержащее как $\mathbb{Q}$, так и $\sqrt{2}$, обязано содержать и все числа вида $a + b\sqrt{2}$, где $a, b \in \mathbb{Q}$. Рассмотрим множество $\mathbb{Q}(\sqrt{2})$ всех таких чисел

$$\mathbb{Q}(\sqrt{2}) = \{\, a + b\sqrt{2} \mid a, b \in \mathbb{Q} \,\} \tag{6.1}$$

и убедимся, что оно удовлетворяет всем требованиям, предъявляемым к числовому полю. Действительно, если $a_1 + b_1\sqrt{2}$ и $a_2 + b_2\sqrt{2}$ — произвольные числа из $\mathbb{Q}(\sqrt{2})$, то их сумма, разность и произведение имеют вид

$$\begin{aligned} (a_1 + b_1\sqrt{2}) \pm (a_2 + b_2\sqrt{2}) &= (a_1 \pm a_2) + (b_1 \pm b_2)\sqrt{2}, \\ (a_1 + b_1\sqrt{2})(a_2 + b_2\sqrt{2}) &= (a_1a_2 + 2b_1b_2) + (a_1b_2 + a_2b_1)\sqrt{2}, \end{aligned} \tag{6.2}$$

и поэтому также принадлежат множеству (6.1).

Предполагая, что $a_2 + b_2\sqrt{2} \neq 0$, рассмотрим отношение

$$\begin{aligned} \frac{a_1 + b_1\sqrt{2}}{a_2 + b_2\sqrt{2}} &= \frac{(a_1 + b_1\sqrt{2})(a_2 - b_2\sqrt{2})}{(a_2 + b_2\sqrt{2})(a_2 - b_2\sqrt{2})} = \\ &= \frac{(a_1a_2 - 2b_1b_2) + (a_2b_1 - a_1b_2)\sqrt{2}}{a_2^2 - 2b_2^2} = \\ &= \frac{a_1a_2 - 2b_1b_2}{a_2^2 - 2b_2^2} + \frac{a_2b_1 - a_1b_2}{a_2^2 - 2b_2^2}\sqrt{2}. \end{aligned} \tag{6.3}$$

В полученном выражении знаменатель $a_2^2 - 2b_2^2$ отличен от нуля. Предположив противное, мы сразу придём к противоречию с тем, что число $\sqrt{2}$ иррационально. Значит, и отношение (6.3) содержится в множестве $\mathbb{Q}(\sqrt{2})$. Таким образом, числовая система (6.1), обладая замкнутостью относительно всех арифметических операций, даёт нам новый пример числового поля — одного из простейших после $\mathbb{Q}$ поля алгебраических чисел. Попутно отметим, что если из рассмотренных чисел оставить лишь те, для которых коэффициенты a и b целые, то получающееся множество

$$\mathbb{Z}(\sqrt{2}) = \{\, a + b\sqrt{2} \mid a, b \in \mathbb{Z} \,\}$$

является, подобно $\mathbb{Z}$, кольцом, но не полем. Равенства (6.2) и (6.3) позволяют легко проверить это.

Числовые кольца и поля — это своего рода арены, на которых разворачиваются события при решении многих задач теории чисел. Используя такое сравнение, можно было бы сказать, что ареной для классической теории чисел, связанной прежде всего с именами Ферма и Эйлера, являлись кольцо целых и поле рациональных чисел. Гаусс был первым, кто пришёл к убеждению, что «естественный источник общей теории следует искать в расширении области арифметики» (в кавычки заключены его собственные слова). К необходимости такого расширения его привели глубокие исследования, касающиеся так называемых степенных вычетов. Об их частном случае — квадратичных вычетах и квадратичном законе взаимности — мы упоминали в главе 4. В случае вычетов четвёртой степени (иначе, биквадратичных вычетов) Гауссу пришлось рассмотреть множество $\mathbb{Z}[i]$ комплексных чисел с целыми действительной и мнимой частью

$$\mathbb{Z}[i] = \{a + bi \mid a, b \in \mathbb{Z}\}. \tag{6.4}$$

Это был исторически первый пример кольца *целых алгебраических чисел*, именуемый теперь *кольцом целых гауссовых чисел*. Тот факт, что множество $\mathbb{Z}[i]$ действительно образует числовое кольцо, читатель легко может проверить самостоятельно.

Эпитет «целый» по отношению к гауссовым числам был применён не случайно — они и в самом деле оказались во многом похожи на обычные целые числа. Поясним подробнее, что имеется в виду.

Все хорошо знают, что натуральные числа бывают простые и составные[1]. Основная теорема арифметики утверждает, что всякое составное натуральное число единственным образом разлагается в произведение простых. Единственность понимается с точностью до порядка сомножителей. Хотя этот факт многими воспринимается как нечто само собой разумеющееся, он

[1] Исключение составляет единица, которую не относят ни к простым, ни к составным числам.

нуждается в доказательстве, которое отнюдь не очевидно. Позднее он будет доказан в намного более общей ситуации, а пока лишь отметим, что рассуждения, приводящие к основной теореме арифметики, опираются на известное всем правило деления с остатком:

для любого целого числа m и натурального n существуют и притом единственные целые числа q и r такие, что

$$m = q \cdot n + r, \quad 0 \leqslant r < n \tag{6.5}$$

(q называют неполным частным, а r — остатком при делении m на n).

Любопытно, что ещё Евклид использовал это правило для отыскания наибольшего общего делителя двух натуральных чисел. Придуманный им способ, который называют алгоритмом Евклида, состоит в следующем.

Пусть при делении m на n получено равенство (6.5). Если получившийся остаток r отличен от 0, разделим на него прежний делитель n

$$n = q_1 \cdot r + r_1, \quad 0 \leqslant r_1 < r. \tag{6.6}$$

Если и новый остаток не равен нулю, продолжаем аналогичную процедуру до тех пор, пока в результате последовательного деления не получится нулевой остаток. Такое обязательно произойдёт, так как получающиеся остатки r, r_1, r_2 и т. д. строго убывают. В итоге мы придём к следующей цепочке равенств

$$\begin{aligned}
r &= q_2 \cdot r_1 + r_2, & 0 \leqslant r_2 < r_1,\\
r_1 &= q_3 \cdot r_2 + r_3, & 0 \leqslant r_3 < r_2,\\
&\ldots\ldots\ldots & \ldots\ldots\\
r_{k-2} &= q_k \cdot r_{k-1} + r_k, & 0 \leqslant r_k < r_{k-1},\\
r_{k-1} &= q_{k+1} \cdot r_k, & r_{k+1} = 0.
\end{aligned} \tag{6.7}$$

Оказывается, последний ненулевой остаток r_k совпадает с наибольшим общим делителем чисел m и n. Доказать это можно так. Посмотрим на нашу цепочку (6.5), (6.6), (6.7), начиная с последнего равенства. Из него вытекает, что r_k делит r_{k-1}. Тогда

в правой части предыдущего равенства оба слагаемых делятся на r_k, и значит, r_k делит r_{k-2}. Переходя от равенства к равенству снизу вверх, точно так же последовательно устанавливаем, что r_k делит r_{k-3}, ..., r_1, n, m. Итак, r_k — общий делитель чисел m и n. Чтобы убедиться в том, что он наибольший, пройдём цепочку уже сверху вниз. Предположим, что d — произвольный общий делитель m и n. Из (6.5) следует тогда, что d делит r, а из (6.6) — что d делит и r_1. Спускаясь вниз по равенствам (6.7), убеждаемся в итоге, что d делит r_k. Иными словами, r_k делится без остатка на всякий общий делитель чисел m и n, но это и означает, что он является их наибольшим общим делителем.

Оказалось, — это и было открыто Гауссом, — что арифметические понятия и теоремы, которые всегда связывались с натуральными и целыми числами, имеют гораздо более широкий смысл, распространяющийся, в частности, и на кольцо $\mathbb{Z}[i]$. Правда, чтобы это понять, некоторые из понятий применительно к произвольным числовым кольцам придётся несколько подкорректировать. Прежде всего, это относится к понятию простого числа. Вспомним, что натуральное число n называется простым, если оно делится без остатка только на 1 и на самого себя. Даже в случае целых чисел такое определение теряет смысл. Объясняется это тем, что в кольце $\mathbb{Z}$ делителем любого числа наряду с 1 является ещё и число -1, и поэтому вместе с тривиальным разложением $n = 1 \cdot n$ есть ещё и такое: $n = (-1) \cdot (-n)$.

Это тем более относится к рассмотренной Гауссом числовой области. Здесь имеется четыре числа, на которые делится без остатка любое целое гауссово число. Это

$$1,\ -1,\ i,\ -i. \tag{6.8}$$

Действительно, ясно во-первых, что каждое из них делит 1, а значит, и всякое число из $\mathbb{Z}[i]$ (например, $1 = i(-i)$). С другой стороны, пусть $a + bi$ — делитель 1. Тогда

$$\frac{1}{a+bi} = \frac{a}{a^2+b^2} - \frac{b}{a^2+b^2}\,i$$

— целое гауссово число. Понятно, что это возможно лишь, если $a^2 + b^2 = 1$. Ровно четыре целочисленные решения последнего

уравнения $a = \pm 1;\ b = 0$ и $a = 0;\ b = \pm 1$ показывают. что числами (6.8) исчерпываются все делители 1 кольца $\mathbb{Z}[i]$.

Присутствие в произвольных числовых кольцах таких особых элементов — делителей 1 (их ещё называют *обратимыми элементами*) вынуждает внести корректировку, о которой говорилось выше. Гаусс даёт следующее определение простого числа, которое, в сущности, годится для любого числового кольца с единицей. Число z он называет простым, если оно не является делителем 1, но в любом его разложении $z = z_1 z_2$ один из сомножителей — обязательно делитель 1.

Нетрудно понять. что простыми целыми числами будут простые натуральные и противоположные к ним числа. Полная информация о простых элементах гауссова кольца $\mathbb{Z}[i]$ имеется в дополнении. Всё же обратим внимание на два примера, которые хорошо отражают имеющуюся здесь специфику.

Пример 1. Числа $1 \pm i$ являются простыми в кольце $\mathbb{Z}[i]$. Пусть, например,

$$1 + i = (a_1 + b_1 i)(a_2 + b_2 i). \tag{6.9}$$

Тогда аналогичное соотношение справедливо для модулей этих чисел

$$|1 + i| = |a_1 + b_1 i|\,|a_2 + b_2 i|.$$

Возводя обе части последнего равенства в квадрат, получаем

$$2 = (a_1^2 + b_1^2)(a_2^2 + b_2^2). \tag{6.10}$$

Из (6.10) вытекает, что либо $a_1^2 + b_1^2 = 1$, $a_2^2 + b_2^2 = 2$, либо наоборот. В обоих случаях в равенстве (6.9) один из сомножителей является обратимым элементом.

Пример 2. Число 2, являющееся простым в $\mathbb{Z}$, уже не будет таковым в $\mathbb{Z}[i]$. В самом деле, равенство

$$2 = (1 + i)(1 - i)$$

служит разложением 2 на множители, ни один из которых не обратим.

Рассматривая задачу о биквадратичных вычетах, Гаусс столкнулся со следующим вопросом: справедлива ли основная теорема арифметики в расширенной им числовой области? Правда, сам этот вопрос требует корректной постановки. Ведь единственность разложения на простые сомножители в буквальном смысле невозможна, — этому препятствуют обратимые элементы. Даже в кольце целых чисел $\mathbb{Z}$ уже имеется такая неоднозначность, например,

$$6 = 2 \cdot 3 = (-2) \cdot (-3). \tag{6.11}$$

Значит, единственность можно требовать лишь с точностью до множителей, являющихся делителями единицы. Таким образом, неизбежно следующее соглашение.

Два разложения числа z на простые множители

$$z = p_1 p_2 \dots p_m, \quad z = q_1 q_2 \dots q_n$$

считаются эквивалентными (одинаковыми), если, во-первых, число простых сомножителей в обоих разложениях совпадает, то есть $m = n$, и во-вторых, при подходящей нумерации сомножителей соответствующие элементы p_k и q_k отличаются лишь обратимым множителем (ассоциированы).

Очевидно, два разложения из (6.11) эквивалентны в $\mathbb{Z}$. Также эквивалентными, но уже в кольце $\mathbb{Z}[i]$, являются разложения

$$2 = (1+i)(1-i) = (-1+i)(-1-i).$$

Кольцо, в котором любые два разложения данного числа на простые множители эквивалентны, называется *кольцом с однозначным разложением на множители*. Говорят, что в нём справедлива *основная теорема арифметики*.

Гауссу удаётся доказать, что рассматриваемая им область $\mathbb{Z}[i]$ обладает однозначным разложением в указанном выше смысле, и опираясь на это, прийти к решению исходной задачи. Доказательство он проводит по той же схеме, что и для натуральных чисел, используя способ деления с остатком целых гауссовых чисел, вполне аналогичный правилу (6.5). Но и тут не обходится

без затруднений. В самом деле, в формулировке этого правила существенным является условие $0 \leqslant r < n$. В то же время, аналогичные неравенства теряют смысл для комплексных чисел (их нельзя сравнивать подобно действительным числам). Гаусс преодолевает и эту трудность, рассматривая для своих чисел величину, называемую их нормой.

Под *нормой* комплексного числа $z = a + bi$ понимается число $N(z)$, равное квадрату его модуля

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

Очевидно, что норма целого гауссова числа является числом натуральным или нулём[1]. Легко заключить, что норма, как и модуль, следующим образом согласована с умножением

$$N(z_1 z_2) = N(z_1)\, N(z_2). \tag{6.12}$$

Наконец, главное. Для любых целых гауссовых чисел z и w можно указать такой способ деления с остатком, чтобы норма остатка была меньше нормы делителя. Рассмотрим с этой целью обычное частное комплексных чисел z и w

$$zw^{-1} = \alpha + \beta i. \tag{6.13}$$

Конечно, в последнем равенстве α и β уже не обязательно целые числа. Выберем тогда такие целые $k, l \in \mathbb{Z}$, которые являются ближайшими соответственно к α и β. Получим

$$\alpha = k + \alpha', \quad \beta = l + \beta', \tag{6.14}$$

где $|\alpha'| \leqslant 1/2$, $|\beta'| \leqslant 1/2$.

Из (6.13) и (6.14) вытекает, что

$$z = w(\alpha + \beta i) = w(k + li) + w(\alpha' + \beta' i).$$

Вводя обозначения $q = k + li$, $r = w(\alpha' + \beta' i)$, запишем $z = qw + r$. Так как q, а следовательно, и qw принадлежат кольцу $\mathbb{Z}[i]$, то r,

[1] Модуль же, в отличие от нормы, может оказаться даже числом иррациональным. Этим и объясняется удобство нового понятия.

как разность двух целых гауссовых чисел, также принадлежит $\mathbb{Z}[i]$. При этом

$$N(r) = N(w(\alpha' + \beta' i)) = N(w)N(\alpha' + \beta' i) =$$
$$= N(w)\,(\alpha'^2 + \beta'^2) \leqslant N(w)\left(\frac{1}{4} + \frac{1}{4}\right) < N(w).$$

Итак, для любой пары чисел $z, w \in \mathbb{Z}[i]$ $(w \neq 0)$ существуют такие $q, r \in \mathbb{Z}[i]$, что возможна следующая запись, аналогичная (6.5)

$$z = qw + r, \quad \text{где } N(r) < N(w). \tag{6.15}$$

Пример. Пусть $z = 1 + 4i$, $w = 2 + i$. Тогда

$$\frac{1+4i}{2+i} = \frac{6}{5} + \frac{7}{5}i = (1+i) + \left(\frac{1}{5} + \frac{2}{5}i\right),$$

откуда после перемножения $1 + 4i = (1+i)(2+i) + i$. В этом равенстве число i играет роль остатка, для которого, как и в (6.15), $N(i) = 1 < N(1+i) = 2$.

Теперь можно распространить на кольцо $\mathbb{Z}[i]$ и алгоритм Евклида. Действительно, последовательно проводя для целых гауссовых чисел деление с остатком, получим цепочку соотношений

$$\begin{aligned}
z &= q \cdot w + r, && N(r) < N(w),\\
w &= q_1 \cdot r + r_1, && N(r_1) < N(r),\\
r &= q_2 \cdot r_1 + r_2, && N(r_2) < N(r_1),\\
&\cdots\cdots\cdots && \cdots\cdots\cdots\\
r_{k-2} &= q_k \cdot r_{k-1} + r_k, && N(r_k) < N(r_{k-1}),\\
r_{k-1} &= q_{k+1} \cdot r_k.
\end{aligned} \tag{6.16}$$

Как и прежде, можно гарантировать, что процесс оборвётся через конечное число шагов, так как последовательность норм $N(r)$, $N(r_1)$, $N(r_2)$, ..., $N(r_k)$ — это строго убывающая последовательность натуральных чисел. Тем же способом, что и раньше, доказывается, что последний ненулевой остаток совпадает с наибольшим общим делителем чисел z и w, то есть делит z и w и делится на любой их общий делитель.

Пример. Если $z = 3 + 5i$, $w = 1 + 3i$, то

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

Следовательно, $(1-i)$ — наибольший общий делитель чисел $3+5i$ и $1 + 3i$.

Дальнейшее развитие алгебры и теории чисел показало, что существует много других числовых колец, и не только числовых, в которых возможен алгоритм Евклида. Эти кольца были названы евклидовыми. В любом евклидовом кольце справедлива основная теорема арифметики. С другой стороны, обнаружилось, что не для всех числовых колец вопрос о единственности разложения на простые элементы решается положительно. Это тонкое обстоятельство порой упускали из виду даже самые маститые математики. Так Эйлер, доказывая Великую теорему Ферма[1] для показателя $n = 3$, использовал комплексные числа вида

$$a + b\sqrt{-3}, \quad \text{где } a, b \in \mathbb{Z},$$

видимо, не испытывая сомнений в их «приличном поведении». Лишь позже было осознано, что в этой числовой области основная теорема арифметики не верна.

Например, в равенстве $4 = 2 \cdot 2 = (1 + \sqrt{-3}) \cdot (1 - \sqrt{-3})$ сомножители 2 и $1 \pm \sqrt{-3}$ являются простыми числами. В то же время, можно показать, что эти два разложения не эквивалентны.

Докажем простоту чисел $1 \pm \sqrt{-3}$, предоставляя остальное читателю. Предположим, что

$$1 + \sqrt{-3} = (a + b\sqrt{-3})(c + d\sqrt{-3}). \tag{6.17}$$

Тогда

$$1 - \sqrt{-3} = (a - b\sqrt{-3})(c - d\sqrt{-3}). \tag{6.18}$$

Перемножая эти два равенства, получим

$$4 = (a^2 + 3b^2)(c^2 + 3d^2).$$

[1] Об этом читатель узнает в следующей главе.

Так как здесь каждый сомножитель — число натуральное, то либо оба они равны 2, либо один из них равен 4, а другой равен 1. Первое невозможно, так как уравнение $a^2+3b^2=2$, очевидно, не имеет решений в целых числах. Во втором случае

$$a^2+3b^2=4, \quad c^2+3d^2=1$$

или наоборот. Второе уравнение имеет лишь решения $c=\pm 1$, $d=0$. Но тогда в разложениях (6.17) и (6.18) сомножители $c\pm$ $\pm d\sqrt{-3}=\pm 1$ обратимы, а значит, числа $1\pm\sqrt{-3}$ простые.

К счастью, нашлась возможность устранить некорректность в рассуждениях Эйлера, и смелая идея его доказательства была спасена. К этому мы ещё вернёмся в следующей главе.

В заключение этой главы познакомим читателя с достаточно общим способом построения полей и колец алгебраических чисел. Многие из примеров, рассмотренных в основном тексте и в дополнении, являются частными случаями этого способа.

Рассмотрим произвольное алгебраическое число α, то есть такое, которое является корнем некоторого многочлена с рациональными коэффициентами. Ясно, что такой многочлен не один. Имеется даже бесконечное множество многочленов, корнем которых является данное число α. Среди них выберем тот, который имеет наименьшую степень и старший коэффициент 1

$$\mu(x)=x^m+a_1x^{m-1}+\ldots+a_{m-1}x+a_m. \tag{6.19}$$

Он называется *минимальным многочленом* числа α. Например, минимальным многочленом для числа $\sqrt{2}$ служит многочлен x^2-2, а для $\sqrt[3]{2}$ — многочлен x^3-2.

Для каждого алгебраического числа α минимальный многочлен определяется уже единственным образом. В самом деле, пусть

$$\nu(x)=x^m+b_1x^{m-1}+\ldots+b_{m-1}x+b_m \tag{6.20}$$

— другой минимальный многочлен для α, отличный от $\mu(x)$. Рассмотрим разность многочленов (6.19) и (6.20)

$$\delta(x)=\mu(x)-\nu(x)=$$
$$=(a_1-b_1)\,x^{m-1}+\ldots+(a_{m-1}-b_{m-1})\,x+(a_m-b_m).$$

Так как $\mu(\alpha) = \nu(\alpha) = 0$, то и $\delta(\alpha) = 0$. Но последнее равенство противоречит определению минимального многочлена, поскольку степень $\delta(x)$ меньше степени $\mu(x)$.

Ясно, что всякое числовое поле, содержащее $\mathbb{Q}$ и число α, обязано содержать и все числа вида

$$q_0 + q_1\,\alpha + q_2\,\alpha^2 + \ldots + q_n\alpha^n, \quad \text{где } q_k \in \mathbb{Q},\ n \in \mathbb{N}. \tag{6.21}$$

Это вытекает из замкнутости любого поля относительно сложения вычитания и умножения. Вспоминая, что α — корень многочлена (6.19), приходим к равенству $\mu(\alpha) = 0$, или

$$\alpha^m = -a_1\alpha^{m-1} - \ldots - a_{m-1}\alpha - a_m.$$

Значит α^m, а тогда и любую бóльшую степень, можно записать как комбинацию степеней $1 = \alpha^0$, $\alpha = \alpha^1$, ..., α^{m-1} с рациональными коэффициентами. То же самое, понятно, справедливо и для любого числа вида (6.21), так как каждое из них допускает запись вида

$$p_0 + p_1\,\alpha + p_2\,\alpha^2 + \ldots + p_{m-1}\alpha^{m-1} \quad (p_k \in \mathbb{Q}). \tag{6.22}$$

В отличие от (6.21), последняя запись единственна. Действительно, предположим, что для некоторого элемента z имеются две различные записи

$$z = p_0 + p_1\,\alpha + p_2\,\alpha^2 + \ldots + p_{m-1}\alpha^{m-1} = \\ = p'_0 + p'_1\,\alpha + p'_2\,\alpha^2 + \ldots + p'_{m-1}\alpha^{m-1}. \tag{6.23}$$

Тогда получим

$$r_0 + r_1\,\alpha + r_2\,\alpha^2 + \ldots + r_{m-1}\alpha^{m-1} = 0, \quad r_k = p_k - p'_k. \tag{6.24}$$

Иными словами, число α является корнем ненулевого многочлена $r_0 + r_1\,x + r_2\,x^2 + \ldots + r_{m-1}x^{m-1}$ степени, меньшей m. Мы вновь приходим к противоречию с определением минимального многочлена. Отсюда заключаем, что $p_k = p'_k$ в равенстве (6.23), то есть обе записи совпадают.

Оказывается, множество всех чисел, записываемых в форме (6.22), уже само образует поле, причём наименьшее из тех, которые содержат α. Для доказательства необходимо было бы проверить, что указанное множество (будем его обозначать $\mathbb{Q}(\alpha)$) замкнуто относительно всех арифметических операций. Замкнутость относительно сложения, вычитания и умножения доказывается без труда — предоставляем сделать это читателю. Доказательство замкнутости относительно деления имеется в дополнении 10. Поле $\mathbb{Q}(\alpha)$ называют *простым алгебраическим расширением* поля $\mathbb{Q}$ с помощью элемента α. Интересно. что многие поля алгебраических чисел сводятся к подобным расширениям (см. дополнение 12).

Отметим, наконец, что в произвольных полях алгебраических чисел, так же, как и в $\mathbb{Q}$, имеет смысл понятие целого числа. Целое алгебраическое число определяется как корень уравнения

$$x^m + a_1 x^{m-1} + \ldots + a_{m-1} x + a_m = 0 \qquad (6.25)$$

с целыми коэффициентами ($a_k \in \mathbb{Z}$). Их роль в различных алгебраических полях во многом сходна с той, которую играют обычные целые числа в поле $\mathbb{Q}$ или рассмотренные выше целые гауссовы числа в поле

$$\mathbb{Q}(i) = \{\, a + bi \mid a, b \in \mathbb{Q} \,\}.$$

Так, если $\mathbb{F}$ — произвольное поле алгебраических чисел, то, как можно показать, множество всех его целых элементов, то есть корней уравнений вида (6.25), принадлежащих этому полю, образует числовое кольцо. При этом всякий элемент из $\mathbb{F}$ записывается в виде отношения двух целых элементов. Добавим, что часто, хотя и не всегда, в кольцах целых алгебраических чисел справедлива основная теорема арифметики.

Дополнения и задачи

1. Каковы простые элементы кольца целых гауссовых чисел? Это не такой уж легкий вопрос. Любопытно, что простые целые числа могут не являться простыми элементами в кольце $\mathbb{Z}[i]$. Мы

уже столкнулись ранее с равенством $2 = (1+i)(1-i)$, подтверждающим сказанное. Примеры нетрудно продолжить. Так, $5 = (2+i)(2-i)$, $13 = (3+2i)(3-2i)$ и т. д. Вообще, можно доказать, что всякое простое целое число вида $4n+1$ в гауссовом кольце разлагается на множители. Это следует (объясните, как именно) из теоретико-числовой теоремы, принадлежащей Эйлеру: *всякое простое число $q = 4n+1$ единственным образом представляется в виде суммы квадратов двух натуральных чисел*

$$q = a^2 + b^2.$$

Наоборот, все остальные нечётные простые числа (они имеют вид $4n+3$) являются простыми элементами и в кольце $\mathbb{Z}[i]$. «*Действительно,* — читаем у Гаусса, — *если бы для некоторого такого числа имело место $q = (a+bi)(\alpha+\beta i)$, то было бы также $q = (a-bi)(\alpha-\beta i)$, и потому $q^2 = (a+bi)(a-bi) \times \times(\alpha+\beta i)(\alpha-\beta i) = (a^2+b^2)(\alpha^2+\beta^2)$; но q^2 может быть только одним способом разложено на положительные сомножители, бо́льшие, чем 1, именно $q^2 = q \cdot q$, так что должно было быть $q = a^2+b^2 = \alpha^2+\beta^2$. Это, однако, невозможно, так как сумма двух квадратов не может иметь вид $4n+3$.*»

Итак, *действительное целое число q является простым элементом гауссова кольца тогда и только тогда. когда q — простое число вида $4n+3$. Очевидно, что то же самое верно и для чисто мнимых чисел qi.*

Со смешанными же гауссовыми числами $a+bi$, то есть такими, для которых и действительная и мнимая части отличны от нуля, дело обстоит так: *целое гауссово число $z = a+bi$ ($a \neq 0$, $b \neq 0$) является простым в $\mathbb{Z}[i]$ тогда и только тогда. когда его норма $N(z) = a^2+b^2$ есть простое число в $\mathbb{Z}$.*

Читатель без труда выяснит, почему из простоты нормы вытекает простота гауссова числа. Доказательство обратного утверждения (мы его не приводим) вновь опирается на некоторые факты из теории чисел.

2. Найти разложение на простые множители следующих гауссовых чисел:

$$3-2i, \quad 3+4i, \quad -3+11i.$$

Для каждой пары найти наибольший общий делитель и наименьшее общее кратное.

3. Доказать, что

а) в поле $\mathbb{Q}(i)$ целые алгебраические числа исчерпываются числами из $\mathbb{Z}[i]$;

б) в поле $\mathbb{Q}(\sqrt{2})$ все целые элементы исчерпываются числами вида $a+b\sqrt{2}$ $(a, b \in \mathbb{Z})$.

4. Рассмотрим простое расширение $\mathbb{Q}(\sqrt{-3})$. Покажите, что

$$\mathbb{Q}(\sqrt{-3}) = \{\, a+b\sqrt{-3} \mid a, b \in \mathbb{Q} \,\}.$$

Можно ли утверждать (сравните с задачей 3), что все целые алгебраические элементы этого поля имеют вид $a+b\sqrt{-3}$, где $a, b \in \mathbb{Z}$? Ответ на этот вопрос имеется в следующей главе.

5. В основном тексте мы выяснили, что кроме кольца целых чисел есть и другие числовые кольца, в которых возможен алгоритм Евклида деления с остатком. Мы назвали такие кольца евклидовыми.

Более точно, числовое кольцо R называется евклидовым, если каждому его ненулевому элементу z сопоставлено целое неотрицательное число $N(z)$, называемое нормой, которое обладает следующими свойствами

1) $N(zw) \geqslant N(z)$ для любых ненулевых элементов $z, w \in R$.

2) для любых элементов $z, w \in R$, $w \neq 0$, возможна запись вида

$$z = qw + r, \quad \text{где } q, r \in R, \text{ и либо } r = 0, \text{ либо } N(r) < N(w). \tag{6.26}$$

Кольцо целых чисел $\mathbb{Z}$ и кольцо целых гауссовых чисел $\mathbb{Z}[i]$, очевидно, охватываются этим определением. В первом случае норма числа — это его модуль, а во втором — квадрат модуля.

В равенстве (6.26), как обычно, q называют *частным*, а r — *остатком* при делении z на w. Если $r = 0$, то говорят, что w является делителем z (обозначение: $w \mid z$). Два элемента, каждый из которых является делителем другого, называются *ассоциированными*.

Отметим следующие простые свойства евклидовых колец:

1) *во всяком евклидовом кольце содержится* 1;

2) *если элементы евклидова кольца ассоциированы, то их нормы равны*;

3) *элементы* z_1 *и* z_2 *ассоциированы тогда и только тогда, когда* $z_1 = z_2\alpha$, *где* α — *обратимый элемент кольца*;

4) *для любой пары ненулевых элементов* $z, w \in R$ *существует единственный с точностью до обратимого множителя наибольший общий делитель этих элементов.*

Докажем свойства 1) и 4), предоставляя читателю доказательство остальных.

Из определения евклидова кольца следует, что для всякого $z \neq 0$ в кольце R существуют q и r такие, что

$$z = qz + r. \tag{6.27}$$

Если $r \neq 0$, то $N(r) < N(z)$. С другой стороны, из (6.27) следует равенство $z(1-q) = r$, поэтому $N(r) \geqslant N(z)$. Полученное противоречие показывает, что $r = 0$, и тогда $q = 1$.

Существование наибольшего общего делителя элементов z и w может быть доказано тем же методом, что и для колец $\mathbb{Z}$ и $\mathbb{Z}[i]$. Действительно, цепочка равенств (6.16), отражающая алгоритм последовательного деления с остатком, имеет смысл в произвольном евклидовом кольце, и все рассуждения относительно этой цепочки переносятся в общую ситуацию практически без изменений. Единственность устанавливается так. Если d_1 и d_2 — наибольшие общие делители элементов z и w, то $d_1 \mid d_2$ и наоборот, $d_2 \mid d_1$, и следовательно, d_1 и d_2 ассоциированы. Но тогда по свойству 3) $d_1 = d_2\alpha$, где α — обратимый элемент.

6. Докажем важное утверждение, часто используемое в дальнейшем.

Если d — *наибольший общий делитель элементов* a *и* b *евклидова кольца* R, *то существуют такие элементы* $t, s \in R$, *что*

$$at + bs = d.$$

В частности, если a и b взаимно просты, то найдутся t и s такие, что

$$at + bs = 1.$$

Например, НОД чисел 18 и 33 равен 3 и $2 \cdot 18 - 1 \cdot 33 = 3$.

Доказательство вновь опирается на алгоритм Евклида. Пусть

$$\begin{aligned} a &= q \cdot b + r, & N(r) &< N(a), \\ b &= q_1 \cdot r + r_1, & N(r_1) &< N(r), \\ r &= q_2 \cdot r_1 + r_2, & N(r_2) &< N(r_1), \\ &\ldots\ldots\ldots\ldots & &\ldots\ldots\ldots \\ r_{k-2} &= q_k \cdot r_{k-1} + r_k, & N(r_k) &< N(r_{k-1}). \end{aligned}$$

где r_k — последний ненулевой остаток (он же и наибольший общий делитель). Запишем эти равенства в виде

$$\begin{aligned} r &= a - q \cdot b, \\ r_1 &= b - q_1 \cdot r, \\ r_2 &= r - q_2 \cdot r_1, \\ &\ldots\ldots\ldots\ldots \\ r_k &= r_{k-2} - q_k \cdot r_{k-1}. \end{aligned}$$

Заменим во втором равенстве r на равное ему выражение $a - qb$. Получится равенство $r_1 = b - q_1(a - qb) = (1 + q_1 q)b - q_1 a$. В третьем равенстве вместо r вновь подставим $a - qb$, а вместо r_1 — получившуюся после первой подстановки правую часть второго равенства, то есть $(1 + q_1 q)b - q_1 a$. Продолжая этот процесс, мы в конечном итоге выразим r_k через a и b в виде $r_k = at_1 + bs_1$, где t_1 и s_1 — некоторые элементы кольца R. Известно, что всякий наибольший общий делитель $d = r_k \alpha$, где α — обратимый элемент R, поэтому

$$d = (at_1 + bs_1)\alpha = at + bs; \quad t, s \in R.$$

7. Доказать, что кольцо $\mathbb{Z}[\sqrt{-2}]$ евклидово, а кольца $\mathbb{Z}[\sqrt{-3}]$, $\mathbb{Z}[\sqrt{-5}]$ таковыми не являются.

8. Множество всех многочленов

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

с коэффициентами из данного числового поля $\mathbb{F}$ по своим алгебраическим свойствам похоже на рассмотренные выше числовые кольца. В самом деле, знакомые читателю операции сложения и умножения многочленов, так же, как и числовые операции, коммутативны, ассоциативны и умножение многочленов дистрибутивно относительно сложения. Их множество, обозначаемое $\mathbb{F}[x]$, называют *кольцом многочленов над полем* $\mathbb{F}$.

Аналогию этого кольца с кольцом целых чисел и другими евклидовыми кольцами можно продолжить. Читателю, например, хорошо известно, что на многочлены может быть распространён алгоритм деления с остатком. Иными словами, для любых многочленов $f(x)$ и $\varphi(x) \neq 0$ возможна запись

$$f(x) = q(x)\varphi(x) + r(x),$$

где $q(x)$ — частное, а $r(x)$ — остаток при делении многочлена $f(x)$ на $\varphi(x)$. При этом степень остатка $r(x)$ меньше, чем степень делителя $\varphi(x)$.

Приняв за норму многочлена его степень, $N(f(x)) = n$, мы без труда убеждаемся, что кольцо многочленов удовлетворяет обоим требованиям, предъявляемым к евклидову кольцу. Таким образом, все понятия и факты, установленные в пунктах 5, 6 для евклидовых колец, распространяются и на кольцо $\mathbb{F}[x]$.

В частности, *если* $d(x)$ *— наибольший общий делитель многочленов* $f(x)$ *и* $\varphi(x)$, *то найдутся такие* $t(x), s(x) \in \mathbb{F}[x]$, *что* $f(x)t(x) + \varphi(x)s(x) = d(x)$.

Обратимыми элементами кольца многочленов, как легко видеть, являются многочлены нулевой степени. Простые элементы кольца $\mathbb{F}[x]$ называют обычно *неприводимыми* или *неразложимыми* многочленами над полем $\mathbb{F}$. По определению, неприводимый многочлен $p(x)$ — это многочлен ненулевой степени, который нельзя разложить в произведение многочленов из $\mathbb{F}[x]$, степень каждого из которых меньше, чем степень $p(x)$. Например, многочлены $x^3 - 2$ и $x^4 - x^2 + 1$ неприводимы над полем рациональных чисел.

Обратим внимание читателя на то, что при переходе к новому, более широкому полю неприводимый многочлен может стать приводимым. Например, над полем действительных чисел $\mathbb{R}$ вышеназванные многочлены уже разлагаются на множители с коэффициентами из $\mathbb{R}$

$$x^3 - 2 = (x - \sqrt[3]{2})\,(x^2 + x\sqrt[3]{2} + \sqrt[3]{4}),$$
$$x^4 - x^2 + 1 = (x^2 - x\sqrt{3} + 1)\,(x^2 + x\sqrt{3} + 1).$$

9. Пусть α — алгебраическое число и $\mu(x) \in \mathbb{Q}[x]$ — его минимальный многочлен. Убедитесь, что $\mu(x)$ неприводим над полем $\mathbb{Q}$. Найдите минимальные многочлены для элементов

$$\sqrt[3]{2}, \quad 1 + \sqrt{3}, \quad \frac{1+i}{\sqrt{2}}.$$

Опишите строение соответствующих простых расширений

$$\mathbb{Q}(\sqrt[3]{2}), \quad \mathbb{Q}(1 + \sqrt{3}), \quad \mathbb{Q}\left(\frac{1+i}{\sqrt{2}}\right).$$

10. Ранее осталось не до конца доказанным, что простое расширение $\mathbb{Q}(\alpha)$, где α — алгебраический элемент, действительно образует поле. Нуждался в обосновании тот факт, что всякий ненулевой элемент в $\mathbb{Q}(\alpha)$ обратим. Теперь мы можем восполнить этот пробел. Пусть

$$p(\alpha) = p_0 + p_1\,\alpha + p_2\,\alpha^2 + \ldots + p_{m-1}\alpha^{m-1}, \quad p_k \in \mathbb{Q},$$

— произвольный отличный от нуля элемент, принадлежащий расширению $\mathbb{Q}(\alpha)$. Рассмотрим тогда соответствующий ему многочлен $p(x) = p_0 + p_1\,x + p_2\,x^2 + \ldots + p_{m-1}x^{m-1}$. Очевидно, что $p(x)$ взаимно прост с минимальным для α многочленом $\mu(x)$, поскольку последний неприводим. Из сказанного в пункте 6 вытекает, что найдутся многочлены $t(x)$ и $s(x)$, удовлетворяющие равенству

$$p(x)t(x) + \mu(x)s(x) = 1.$$

Заменяя в нём переменную x числом α и учитывая, что $\mu(\alpha)=0$, получаем

$$p(\alpha)t(\alpha)=1.$$

Но это и означает, что элемент $p(\alpha)$ обратим и обратным для него является $t(\alpha)$.

В качестве примера найдем обратный элемент для $3+2\sqrt[3]{3}+\sqrt[3]{9}$ в поле $\mathbb{Q}(\sqrt[3]{3})$. Здесь $\alpha=\sqrt[3]{3}$ и минимальным многочленом является $\mu(x)=x^3-3$. Роль $p(x)$ играет многочлен $3+2x+x^2$. Для отыскания многочленов $t(x)$ и $s(x)$ используем алгоритм Евклида:

$$x^3-3=(x-2)(x^2+2x+3)+(x+3),$$
$$x^2+2x+3=(x-1)(x+3)+6.$$

Перепишем полученные равенства иначе

$$x+3=(x^3-3)-(x-2)(x^2+2x+3),$$
$$6=(x^2+2x+3)-(x-1)(x+3).$$

Исключая из данных соотношений $x+3$, получим

$$6=(x^2+2x+3)-(x-1)\big[(x^3-3)-(x-2)(x^2+2x+3)\big],$$

или, наконец,

$$1=\frac{1}{6}(x^2-3x+3)(x^2+2x+3)-\frac{1}{6}(x-1)(x^3-3).$$

При $x=\sqrt[3]{3}$ будем иметь

$$1=\frac{1}{6}(\sqrt[3]{9}-3\sqrt[3]{3}+3)(\sqrt[3]{9}+2\sqrt[3]{3}+3),$$

и следовательно, искомый обратный элемент равен

$$\frac{1}{6}(\sqrt[3]{9}-3\sqrt[3]{3}+3)=\frac{1}{2}-\frac{1}{2}\sqrt[3]{3}+\frac{1}{6}\sqrt[3]{9}.$$

11. В качестве исходного поля при построении расширений может быть взято любое числовое поле $\mathbb{F}$ (не обязательно совпадающее с полем рациональных чисел). Все понятия и результаты, отмеченные выше, переносятся на эту более общую ситуацию.

Так, число α называется алгебраическим над полем $\mathbb{F}$, если оно является корнем некоторого многочлена с коэффициентами из $\mathbb{F}$, а многочлен наименьшей степени с указанными свойствами и старшим коэффициентом 1 называется минимальным многочленом для α над полем $\mathbb{F}$.

Неизменной остаётся и конструкция простого расширения $\mathbb{F}(\alpha)$ основного поля $\mathbb{F}$. Если m — степень минимального для α многочлена над полем $\mathbb{F}$, то элементы простого расширения исчерпываются числами вида

$$p_0 + p_1\alpha + p_2\alpha^2 + \ldots + p_{m-1}\alpha^{m-1}, \quad p_k \in \mathbb{F}.$$

Процесс присоединения алгебраических элементов можно повторять многократно. Пусть $\alpha_1, \alpha_2, \ldots, \alpha_k$ — элементы, алгебраические над $\mathbb{F}$. Рассмотрим цепочку простых расширений

$$\mathbb{F}_1 = \mathbb{F}(\alpha_1),\ \mathbb{F}_2 = \mathbb{F}_1(\alpha_2),\ \ldots,\ \mathbb{F}_k = \mathbb{F}_{k-1}(\alpha_k).$$

Полученное в результате расширение $\mathbb{F}_k$ назовём составным алгебраическим расширением. Можно показать, что составное расширение не зависит от того, в какой последовательности присоединяются $\alpha_1, \alpha_2, \ldots, \alpha_k$, и всякий раз получающееся при этом поле $\mathbb{F}(\alpha_1, \alpha_2, \ldots, \alpha_k)$ есть наименьшее из числовых полей, содержащих поле $\mathbb{F}$ и указанные числа. В том случае, когда элементы $\alpha_1, \alpha_2, \ldots, \alpha_k$ исчерпывают множество корней заданного многочлена $f(x)$ с коэффициентами из исходного поля $\mathbb{F}$, описанное составное расширение $\mathbb{F}(\alpha_1, \alpha_2, \ldots, \alpha_k)$ даёт наименьшее из полей, содержащих все эти корни. Его называют *полем разложения* многочлена $f(x)$.

12. Оказывается, что процесс образования составного расширения можно осуществить за один этап, то есть присоединяя единственный подходящий элемент. Иными словами, всякое такое расширение является простым. Это утверждение называют теоремой о примитивном элементе:

пусть $\mathbb{F}_k = \mathbb{F}(\alpha_1, \alpha_2, \ldots, \alpha_k)$ *— составное алгебраическое расширение поля* $\mathbb{F}$, *получающееся присоединением элементов* $\alpha_1, \alpha_2, \ldots, \alpha_k$. *Тогда найдётся такое число* $\theta \in \mathbb{F}_k$ *(оно и называется примитивным элементом), что* $\mathbb{F}_k = \mathbb{F}(\theta)$.

Докажем теорему для $k = 2$. Случай произвольного k сводится к рассматриваемому индукцией по k. Пусть $\mathbb{F}_1 = \mathbb{F}(\alpha_1)$, $\mathbb{F}_2 = \mathbb{F}(\alpha_2)$, а многочлены $\mu_1(x)$ и $\mu_2(x)$ являются минимальными над $\mathbb{F}$ для элементов α_1 и α_2 соответственно. Предположим, что

$$\beta_1 = \alpha_1, \beta_2, \ldots, \beta_n \tag{6.28}$$

— все корни $\mu_1(x)$, а

$$\gamma_1 = \alpha_2, \gamma_2, \ldots, \gamma_m \tag{6.29}$$

— все корни $\mu_2(x)$. Неприводимый многочлен не может иметь кратных корней. В противном случае НОД этого многочлена и его производной имел бы степень > 0 (продумайте детали этого рассуждения). Из отсутствия кратных корней следует, что среди чисел (6.28), так же, как среди чисел (6.29), нет одинаковых. Рассмотрим $n(m-1)$ элементов следующего вида

$$\frac{\beta_i - \beta_1}{\gamma_1 - \gamma_j}; \quad i = 1, 2, \ldots, n; \ j = 2, \ldots, m. \tag{6.30}$$

Возьмём в поле $\mathbb{F}$ какой-нибудь элемент c, не равный ни одному из элементов (6.30). Он существует ввиду бесконечности поля $\mathbb{F}$. Положим

$$\theta = \alpha_1 + c\alpha_2. \tag{6.31}$$

Очевидно, что $\mathbb{F}(\theta) \subseteq \mathbb{F}_2$. Далее заметим, что

$$\theta \neq \beta_i + c\gamma_j \quad (i = 1, 2, \ldots, n; \ j = 2, \ldots, m), \tag{6.32}$$

иначе бы число c совпадало с одним из чисел (6.30).

Рассмотрим над полем $\mathbb{F}(\theta)$ многочлен $g(x) = \mu_1(\theta - cx)$. Из (6.31) следует, что $g(x)$ имеет корень α_2, общий с многочленом $\mu_2(x)$, а из (6.32) вытекает, что это их единственный общий корень. Тогда $x - \alpha_2$ — наибольший общий делитель многочленов $g(x)$ и $\mu_2(x)$, и существуют такие многочлены $t(x)$ и $s(x)$ над полем $\mathbb{P}(\theta)$, что $x - \alpha_2 = g(x)\,t(x) + \mu_2(x)\,s(x)$, поэтому $x - \alpha_2$ — также многочлен над $\mathbb{F}(\theta)$, и значит, $\alpha_2 \in \mathbb{F}(\theta)$. Но тогда $\alpha_1 = \theta - c\alpha_2 \in \mathbb{F}(\theta)$. Из сказанного вытекает включение $\mathbb{F}_2 \subseteq \mathbb{F}(\theta)$. Так как ранее было установлено обратное включение, отсюда следует, что $\mathbb{F}_2 = \mathbb{F}(\theta)$.

ГЛАВА 7

ВОКРУГ ВЕЛИКОЙ ТЕОРЕМЫ

Греческий математик Диофант жил в III веке нашей эры. Годы жизни великого французского математика Пьера Ферма — с 1601 по 1665. Но так случилось, что исследования Ферма по теории чисел стимулировались именно трудами Диофанта. Имя этого греческого ученого известно нам благодаря его многотомной «Арифметике», состоящей из 13 книг (уцелело только 6) и представляющей, по мнению историков математики, одно из загадочных явлений в науке. «Арифметика» Диофанта резко отличалась от произведений классиков античной математики — Евклида, Архимеда, Аполлония, в которых, как мы уже говорили, доминировал геометрический метод даже при решении алгебраических вопросов.

У Диофанта уже встречается алгебраическая символика, правда, совсем не похожая на современную. Он вводит обозначения для неизвестных и их степеней, особые знаки для операций вычитания, для обратной величины. Всё это помогает ему успешно решать алгебраические задачи, куда более сложные, чем те, что рассматривались его предшественниками. Основная тема «Арифметики» — отыскание целых и рациональных решений неопределённых, то есть имеющих бесконечно много решений, уравнений. Одна из задач, упомянутых Диофантом, связана с описанием так называемых пифагоровых чисел, являющихся целыми положительными решениями неопределённого уравнения

$$x^2 + y^2 = z^2. \tag{7.1}$$

Диофант приводит весьма остроумный метод для отыскания всех положительных рациональных решений этого уравнения. Из его рассуждений легко можно вывести (см. дополнение 1) общие формулы, описывающие все пифагоровы числа, а именно

$$x = 2kpq, \quad y = k(p^2 - q^2), \quad z = k(p^2 + q^2), \tag{7.2}$$

где k, p, q — произвольные натуральные числа. Если же пифагоровы числа x, y, z попарно взаимно просты, то они записываются в виде

$$x = 2pq, \quad y = p^2 - q^2, \quad z = p^2 + q^2, \tag{7.3}$$

где p, q — взаимно простые числа, имеющие разную чётность. В формулах (7.2), (7.3) можно поменять ролями x и y, так как они в рассматриваемое уравнение входят симметрично.

Сочинения Диофанта, полные хитроумных методов и приёмов, оставались неизвестными европейским математикам вплоть до XVI века, когда принадлежащая ему рукопись случайно была обнаружена в библиотеке Ватикана. В 1621 году был издан перевод «Арифметики» на латинский язык (язык науки того времени), выполненный французским математиком Баше де Мезириаком. Владельцем одного из экземпляров и стал Пьер Ферма. Значительная часть наследия Ферма в области теории чисел — мысли, идеи, формулировки теорем, гипотезы — дошла до нас в форме записей и замечаний на полях этого экземпляра «Арифметики». Напротив восьмой задачи второй книги Диофанта «разбить квадратное число на два других квадратных числа» (это и есть упомянутая выше задача) Ферма сделал самую знаменитую свою запись: «*Наоборот, невозможно разложить куб на два куба, биквадрат на два биквадрата и вообще никакую степень, бóльшую биквадрата, на две степени с тем же показателем. Я открыл этому поистине чудесное доказательство, но поля слишком узки, чтобы поместить его*». Иными словами, здесь утверждается, что *при $n > 2$ не существует отличных от нуля целых чисел x, y, z, являющихся решением уравнения*

$$x^n + y^n = z^n. \tag{7.4}$$

Так формулируется «Великая теорема Ферма». Теорема Ферма вызвала необычайный резонанс как среди математиков, так и среди людей, далёких от математики, главным образом благодаря простоте формулировки и интригующему воздействию слов о существовании «чудесного доказательства». Подробности об этом можно прочесть в других книгах (например, в [23]). Мы же особо хотим подчеркнуть то, что с Великой теоремой связана ещё одна важная нить исторического развития алгебры. Хотя в формулировке теоремы Ферма участвуют лишь целые числа x, y, z, в ходе её доказательства пришлось обратиться к числам более общего вида — алгебраическим и целым алгебраическим числам (мы познакомились с ними в предыдущей главе). Глубокое проникновение в арифметику этих чисел, зачастую отличную от обычной арифметики, потребовало введения новых понятий — таких, как поле, кольцо, идеал, дивизор и других, одинаково важных для алгебры, теории чисел, алгебраической геометрии.

Единственный случай, для которого известно элементарное (то есть не использующее алгебраических чисел) доказательство Великой теоремы, — это случай показателя $n = 4$. Указанное доказательство, обнаруженное в записях самого Ферма, опирается на упомянутые выше формулы для пифагоровых чисел. Ферма доказывает даже чуть более общее утверждение: *уравнение*

$$x^4 + y^4 = z^2 \tag{7.5}$$

не имеет решений в целых числах, отличных от нуля. В своих рассуждениях Ферма использует изобретённый им метод так называемого бесконечного спуска, которым часто пользовались также и другие математики, в частности, Эйлер. Доказательство, предлагаемое читателю, использует этот метод в несколько модифицированном виде.

Нам достаточно убедиться, что не существует решений, в которых x, y, z попарно взаимно просты. Если, например, $x = x_1 d$, $y = y_1 d$, где d — их наибольший общий делитель, то $z^2 = (x_1^4 + y_1^4)\, d^4$. Полагая $z_1 = z/d^2$, получим равенство $x_1^4 + y_1^4 = z_1^2$, в котором x_1 и y_1 уже взаимно просты. Если x_1 и z_1 или y_1 и z_1 не являются взаимно простыми, то, производя сокращения, придём к ситуации, когда все три переменные попарно

взаимно просты. Кроме того, можно ограничиться рассмотрением натуральных x, y, z. Предположим противное, что такие решения существуют. Выберем среди них то решение (x, y, z), для которого z принимает наименьшее значение. Числа x^2, y^2 и z являются попарно взаимно простыми пифагоровыми числами, и потому на основании (7.3) можно записать

$$x^2 = 2pq, \quad y^2 = p^2 - q^2, \quad z = p^2 + q^2, \tag{7.6}$$

где p и q, как уже говорилось, взаимно простые числа, одно из которых чётно, другое — нечётно.

Пусть $p = 2k$, $q = 2l + 1$, тогда

$$y^2 = (2k)^2 - (2l + 1)^2 = 4(k^2 - l^2 - l) - 1,$$

что невозможно, так как любой нечётный квадрат должен иметь вид $4m + 1$.

Наоборот, если $q = 2l$, то $x^2 = 4pl$, и тогда $pl = (x/2)^2$. Поскольку числа взаимно просты, а их произведение является квадратом, то и каждое из них также квадрат, поэтому $p = w^2$, $l = t^2$ для некоторых натуральных w и t. Согласно второму из равенств (7.6), получим $y^2 = (w^2)^2 - (2t^2)^2$ или $(2t^2)^2 + y^2 = (w^2)^2$, иными словами, числа $2t^2$, y, w^2 — пифагоровы и также взаимно просты. Опять можно записать

$$2t^2 = 2ab, \quad y = a^2 - b^2, \quad w^2 = a^2 + b^2, \tag{7.7}$$

где a и b обладают теми же свойствами, что p и q. Из первого равенства, равносильного $t^2 = ab$, вытекает, что сомножители a и b сами являются квадратами: $a = u^2$, $b = v^2$ для некоторых натуральных u и v. Тогда третье равенство из (7.7) даёт

$$u^4 + v^4 = w^2,$$

следовательно, числа u, v, w составляют решение уравнения (7.5). Но очевидно, что $w < z$, а это противоречит выбору первоначального решения (x, y, z). Полученное противоречие и доказывает наше утверждение, а следовательно, и теорему Ферма для показателя $n = 4$.

Следующий после Ферма шаг делает Эйлер — он доказывает Великую теорему для случая $n = 3$, то есть доказывает, что уравнение $x^3 + y^3 = z^3$ не имеет нетривиальных решений в целых числах. Принципиально новым в доказательстве Эйлера было привлечение алгебраических чисел вида $a + b\sqrt{-3}$ $(a, b \in \mathbb{Z})$ и смелое оперирование с ними. Интересно, что никакого элементарного доказательства для $n = 3$ неизвестно и поныне. Как уже отмечалось в предыдущей главе, рассуждения Эйлера содержали некорректность. Доказательство, которое Эйлер, подобно Ферма, проводит методом бесконечного спуска, опирается на следующую лемму.

Если взаимно простые целые числа a и b таковы, что $a^2 + 3b^2$ является кубом целого числа, то существуют такие целые числа s и t, что

$$a = s(s^2 - 9t^2), \quad b = 3t(s^2 - t^2).$$

Приведём доказательство Эйлера. Он исходит из равенства

$$a^2 + 3b^2 = (a + b\sqrt{-3})(a - b\sqrt{-3}), \tag{7.8}$$

утверждая, что поскольку левая его часть является кубом, то это же справедливо для обоих сомножителей правой части (!?). В частности,

$$a + b\sqrt{-3} = (s + t\sqrt{-3})^3, \tag{7.9}$$

где s и t — некоторые целые числа. Раскрывая скобки, получаем

$$a + b\sqrt{-3} = s^3 + 3s^2t\sqrt{-3} - 9st^2 - 3t^3\sqrt{-3} = \\ = (s^3 - 9st^2) + (3s^2t - 3t^3)\sqrt{-3},$$

откуда следует, что

$$a = s^3 - 9st^2 = s(s^2 - 9t^2), \quad b = 3s^2t - 3t^3 = 3t(s^2 - t^2).$$

В связи с приведённым рассуждением возникает немало вопросов. Прежде всего, вывод относительно сомножителей в правой части равенства (7.8) был бы оправдан лишь при условии их

взаимной простоты, которую Эйлер считает, по-видимому, очевидно вытекающей из взаимной простоты чисел a и b. Однако это не только нуждается в доказательстве, но ещё и сам термин «взаимно простые» требует разъяснения применительно к числам вида $a + b\sqrt{-3}$. В предыдущей главе мы, правда, уже позаботились об определении этого и других арифметических понятий применительно к произвольным евклидовым кольцам. Но дело в том, что множество эйлеровых чисел, как выяснено там же, евклидова кольца как раз и не образует.

Кроме того, появление чисел $a + b\sqrt{-3}$ в случае $n = 3$ кажется довольно странным и неожиданным. Но это лишь на первый взгляд. Действительно, рассмотрим левую часть уравнения Ферма и перепишем её в несколько ином виде

$$x^3 + y^3 = y^3\left(\left(\frac{x}{y}\right)^3 + 1\right) = y^3(t^3 + 1),$$

полагая $t = x/y$. Поскольку корнями многочлена $t^3 + 1$ являются числа $-1, -\xi, -\xi^2$, где

$$\xi = \cos\frac{2\pi}{3} + i\sin\frac{2\pi}{3}$$

— первообразный корень третьей степени из 1, то справедливо разложение на множители

$$t^3 + 1 = (t + 1)(t + \xi)(t + \xi^2).$$

Но тогда

$$x^3 + y^3 = y^3(t+1)(t+\xi)(t+\xi^2) = (x+y)(x+\xi y)(x+\xi^2 y). \quad (7.10)$$

Полученное разложение имеет смысл, конечно, лишь в расширенной по сравнению с $\mathbb{Z}$ числовой области, а именно, в кольце $\mathbb{Z}[\xi]$, которое, как легко можно понять, совпадает с множеством чисел вида

$$a + b\xi + c\xi^2, \quad \text{где } a, b, c \in \mathbb{Z}. \quad (7.11)$$

Вспомним теперь, что

$$\xi = \cos\frac{2\pi}{3} + i\sin\frac{2\pi}{3} = \frac{-1+i\sqrt{3}}{2} = \frac{-1+\sqrt{-3}}{2},$$
$$\xi^2 = \cos\frac{4\pi}{3} + i\sin\frac{4\pi}{3} = \frac{-1-i\sqrt{3}}{2} = \frac{-1-\sqrt{-3}}{2} = -1-\xi.$$

Тогда произвольный элемент из $\mathbb{Z}[\xi]$ запишется в виде

$$a + b\frac{-1+\sqrt{-3}}{2} + c\frac{-1-\sqrt{-3}}{2} = \frac{p+q\sqrt{-3}}{2}, \qquad (7.12)$$

где $p = 2a - b - c$, $q = b - c$ — целые числа. Нетрудно заметить также, что они имеют одинаковую чётность. В частности, если оба они чётные, то мы приходим к числам, которые рассматривал Эйлер.

Итак, появление у Эйлера чисел $a + b\sqrt{-3}$ вовсе не случайно. Теперь ещё ясно и то, что естественнее рассматривать числа вида (7.12).

Заметим, что элементы из $\mathbb{Z}[\xi]$ могут быть записаны также в виде $A + B\xi$ $(A, B \in \mathbb{Z})$. Последняя запись получится, если в (7.11) заменить ξ^2 на равный ему элемент $-1-\xi$. Можно доказать, что кольцо $\mathbb{Z}[\xi]$ является кольцом целых алгебраических чисел и совпадает с множествои всех целых элементов поля $\mathbb{Q}(\xi) = \{\alpha + \beta\xi \mid \alpha, \beta \in \mathbb{Q}\}$. Самое же главное то, что в $\mathbb{Z}[\xi]$, в отличие от исходного кольца $\mathbb{Z}[\sqrt{-3}]$, уже справедлива основная теорема арифметики. Мы докажем это тем же методом, что и для кольца целых гауссовых чисел, убедившись, что кольцо $\mathbb{Z}[\xi]$, подобно $\mathbb{Z}[i]$, является евклидовым относительно некоторой нормы.

Для всякого $z = A + B\xi \in \mathbb{Z}[\xi]$ норма по определению равна $N(z) = z \cdot \overline{z} = |z|^2$. Выразим её через коэффициенты A и B, проделав следующие вычисления

$$N(z) = z \cdot \overline{z} = (A + B\xi)(A + B\overline{\xi}) =$$
$$= A^2 + AB(\xi + \overline{\xi}) + B^2(\xi\overline{\xi}) = A^2 - AB + B^2.$$

Если же для z использована запись вида $a + b\sqrt{-3}$, то очевидно, $N(z) = a^2 + 3b^2$. Как и для гауссовой нормы, в этом случае верно

основное для нас свойство[1]:

$$N(z_1 z_2) = N(z_1)N(z_2).$$

Пусть теперь z и w — произвольные элементы из $\mathbb{Z}[\xi]$. Их частное $zw^{-1} = \alpha + \beta\xi \in \mathbb{Q}(\xi)$. Положим $\alpha = k + \alpha'$, $\beta = l + \beta'$, где k и l — ближайшие соответственно к α и β целые числа, а значит

$$|\alpha'| \leqslant \frac{1}{2} \quad \text{и} \quad |\beta'| \leqslant \frac{1}{2}.$$

Как и в случае гауссова кольца, z записывается в виде $z = qw + r$, где $q = k + l\xi$ и $r = w(\alpha' + \beta'\xi)$ принадлежат $\mathbb{Z}[\xi]$. При этом

$$N(r) = N(w)N(\alpha' + \beta'\xi) = N(w)(\alpha'^2 - \alpha'\beta' + \beta'^2) \leqslant$$
$$\leqslant N(w)\left(|\alpha'|^2 + |\alpha'|\,|\beta'| + |\beta'|^2\right) \leqslant N(w)\left(\frac{1}{4} + \frac{1}{4} + \frac{1}{4}\right) < N(w).$$

Вернёмся теперь к доказательству леммы Эйлера, пытаясь восполнить имеющиеся в нём пробелы. Прежде всего убедимся, что сомножители $a + b\sqrt{-3}$ и $a - b\sqrt{-3}$ из (7.8) действительно взаимно просты. Для этого, как мы уже знаем, нужно доказать, что их наибольший общий делитель $\gamma \in \mathbb{Z}[\xi]$ является в $\mathbb{Z}[\xi]$ обратимым элементом, или иначе, что его норма $N(\gamma) = 1$.

Заметим сначала, что γ является делителем суммы и разности сомножителей, то есть γ делит $2a = (a + b\sqrt{-3}) + (a - b\sqrt{-3})$ и $2b\sqrt{-3} = (a + b\sqrt{-3}) - (a - b\sqrt{-3})$, а следовательно, его норма $N(\gamma)$ делит числа $N(2a) = 4a^2$ и $N(2b\sqrt{-3}) = 12b^2$. Поскольку a и b взаимно просты, то $N(\gamma)$ может быть лишь делителем числа 4, если a не делится на 3, и делителем числа 12, если a кратно 3.

Легко проверить, что последний случай в условиях леммы Эйлера невозможен. Действительно, если $a = 3k$, то норма $a^2 + 3b^2 = 9k^2 + 3b^2$ также делится на 3, но, являясь кубом, должна,

[1] Отметим, что приведённое определение нормы имеет смысл не только для целых алгебраических чисел, но и для произвольных элементов из $\mathbb{Q}(\xi)$; естественно, что и для них сохраняется рассматриваемое здесь свойство.

следовательно, делиться на 27. Можно записать, что $9k^2 + 3b^2 = 27n$ или $b^2 = 9n - 3k^2$. Последнее означает, что b^2, а значит, и b кратно 3, но это противоречит взаимной простоте чисел a и b.

Значит, норма $N(\gamma)$ является делителем числа 4, и для неё имеется три возможности: $N(\gamma) = 4$, $N(\gamma) = 2$, $N(\gamma) = 1$. Если $N(\gamma) = 4$, то $N(\gamma/2) = 1$, и тогда $\gamma = 2\varepsilon$, где ε — обратимый элемент. В этом случае каждое из чисел $a + b\sqrt{-3}$, $a - b\sqrt{-3}$ должно делиться на 2, но это возможно лишь тогда, когда оба числа a и b делятся на 2, что противоречит их взаимной простоте.

Если же $N(\gamma) = 2$, то, полагая $\gamma = x + y\xi$, приходим к уравнению $x^2 - xy + y^2 = 2$, которое, как нетрудно понять, вообще не имеет решений в целых числах. Остаётся лишь третья возможность $N(\gamma) = 1$. Она и означает, что сомножители в (7.8) взаимно просты. Теперь полностью обосновано заключение о том, что каждый сомножитель в (7.8) является кубом. Однако возникает другое, правда, не очень значительное препятствие. Ведь справедливость равенства (7.9) доказана нами не для кольца эйлеровых чисел, а для содержащего его кольца $\mathbb{Z}[\xi]$, а поскольку элементы из $\mathbb{Z}[\xi]$, как мы помним, имеют вид (7.12), то в равенстве (7.9) числа s и t могут оказаться нецелыми. Выручает здесь такое обстоятельство. Если $s + t\sqrt{-3}$ есть решение уравнения (7.9), то его решениями будут также числа $(s + t\sqrt{-3})\xi$, $(s + t\sqrt{-3})\xi^2$. Это следует из того, что $\xi^3 = 1$. Оказывается, что хотя бы одно из перечисленных трёх решений является эйлеровым числом, то есть имеет целые коэффициенты (несложное доказательство этого факта предоставляем читателю). Теперь лемму Эйлера можно считать полностью доказанной.

Доказательство теоремы Ферма для показателя $n = 3$ было опубликовано Эйлером в его «Универсальной арифметике» в 1768 г. Дальнейшее продвижение было медленным. Лишь в 1825 г. был разобран случай $n = 5$ — доказательство теоремы Ферма для этого случая почти одновременно предложили французские математики Дирихле (1805–1859) и Лежандр (1752–1833). В 1839 г. другой французский математик Ламе (1795–1870) сумел доказать теорему для показателя $n = 7$.

На всех этих работах так или иначе отразилась руководящая мысль Эйлера: для серьёзного рассмотрения проблемы Ферма

необходимо привлекать алгебраические числа, подобные числам (7.12). Нетрудно понять, что достаточно решить её для простых показателей n. Постепенно формируется общий подход к проблеме, состоящий в следующем.

Заметим, что левая часть уравнения (7.4) при простом $n \geqslant 3$ разлагается на множители

$$x^n + y^n = (x+y)(x+\xi y)(x+\xi^2 y) \ldots (x+\xi^{n-1} y),$$

где ξ — первообразный корень n-ой степени из 1. Это разложение, справедливое для расширенной числовой области $\mathbb{Z}[\xi]$, аналогично равенству (7.10) и получается тем же способом. Уравнение Ферма переписывается тогда в виде

$$(x+y)(x+\xi y)(x+\xi^2 y) \ldots (x+\xi^{n-1} y) = z^n. \qquad (7.13)$$

Рассмотрим кольцо $\mathbb{Z}[\xi]$, получающееся из $\mathbb{Z}$ в результате присоединения корня n-ой степени из 1.

Равенство (7.13) означает тогда, что в этом кольце n-ая степень числа z может быть записана также и в виде произведения n различных сомножителей, при этом, как отмечалось выше, числа x, y и z можно считать попарно взаимно простыми. Возможно ли такое разложение — зависит от арифметических свойств $\mathbb{Z}[\xi]$. Во всяком случае, ясно, что ответ сильно зависит от того, справедлива ли в кольце $\mathbb{Z}[\xi]$ основная теорема арифметики.

Первые авторы общих «доказательств» Великой теоремы молчаливо предполагали, что это так. Подобную ошибку совершили уже упоминавшийся выше Ламе и немецкий математик Куммер (1810–1893). Существует также мнение, что и «чудесное доказательство», о котором сообщал Ферма, скрывало в себе такого же рода предположение.

Как же всё-таки обстоит дело фактически? Верно то, что всякий необратимый элемент в $\mathbb{Z}[\xi]$ разлагается в произведение простых элементов. Однако такое разложение в общем случае не единственно. Это было обнаружено Куммером после упорных его попыток доказать противоположное утверждение. Самый простой пример кольца $\mathbb{Z}[\xi]$, в котором не выполняется основная теорема арифметики, получается только при $n = 23$. Вот почему

его нелегко было заметить. Оказалось всё-таки, что положение можно исправить, если к числам кольца $\mathbb{Z}[\xi]$ присоединить некие новые объекты, которые Куммер называет идеальными числами. В расширенной этими идеальными числами области восстанавливается однозначность разложения на простые множители. Хотя это изобретение Куммера и привело к очень сильному прогрессу в исследовании проблемы Ферма (Куммеру удалось её решить для всех простых $n < 100$), всё же полного её решения получено не было. Главное, однако, что своими исследованиями он заложил основы замечательного по своей глубине и красоте раздела математики — алгебраической теории чисел.

Эта теория развивалась далее трудами таких крупнейших математиков, как Л. Кронекер, Р. Дедекинд, Д. Гильберт, А. Вейль, и ряда других. Большой вклад в учение об алгебраических числах внесли русские и советские математики: Е. И. Золотарёв, Г. Ф. Вороной, А. А. Марков, И. Р. Шафаревич.

Что же касается самой проблемы Ферма, то она была окончательно решена в 1995 г. английским математиком Э. Уайлсом, который устранил пробелы в своем же доказательстве 1993 г. На самом деле Уайлс доказал более сильное утверждение, чем теорема Ферма, — так называемую гипотезу Таниямы—Шимуры—Вейля [25]. Мы не будем здесь приводить точную формулировку этой гипотезы — для этого необходимы специальные знания из алгебраической геометрии. Отметим лишь, что к тому времени Великая теорема была доказана для всех $n < 100\,000$. Это было сделано не без участия компьютеров, которым, правда, отводилась только техническая роль.

Дополнения

1. Докажем формулы (7.3) для пифагоровых чисел. Точнее, убедимся в следующем.

Если x, y, z — попарно взаимно простые числа, удовлетворяющие уравнению (7.1), *то одно из чисел x или y чётно, а другое нечётно. При чётном x указанная тройка x, y, z является*

решением уравнения (7.1) *тогда и только тогда, когда*

$$x = 2pq, \quad y = p^2 - q^2, \quad z = p^2 + q^2, \tag{7.14}$$

где p и q взаимно просты, $p > q$ и одно из них чётно, а другое нечётно.

Доказательство. Очевидно, что ровно два из трёх чисел x, y, z обязаны быть нечётными. Если нечётны x и y (то есть $x = 2m + 1$, $y = 2n + 1$), то

$$z^2 = x^2 + y^2 = 4(m^2 + m + n^2 + n) + 2$$

— чётное число, не делящееся на 4. Последнее невозможно, поэтому чётным должно являться одно из чисел x или y. Если чётно x, то перепишем равенство (7.1) в виде

$$\left(\frac{x}{2}\right)^2 = \left(\frac{z+y}{2}\right)\left(\frac{z-y}{2}\right). \tag{7.15}$$

В нём числа $\frac{x}{2}$, $\frac{z+y}{2}$, $\frac{z-y}{2}$ — целые, причём $\frac{z+y}{2}$ и $\frac{z-y}{2}$ взаимно просты (почему?).

Но если произведение двух взаимно простых чисел равно квадрату, то и каждый сомножитель в отдельности является квадратом. Пусть

$$\frac{z+y}{2} = p^2, \quad \frac{z-y}{2} = q^2. \tag{7.16}$$

Тогда из (7.15) и (7.16) легко получаем формулы для пифагоровых чисел.

Наоборот, непосредственной проверкой нетрудно убедиться, что числа (7.14) удовлетворяют уравнению (7.1); при этом из условий, наложенных на числа (7.14), вытекает взаимная простота x, y, z.

2. Как уже отмечалось, Эйлер доказывает Великую теорему для $n = 3$ методом бесконечного спуска, заимствованным у Ферма. Однако доказательство существенно сложнее, чем в случае

$n = 4$. Напомним, что числа x, y и z можно считать попарно взаимно простыми, и потому одно из них обязательно чётно, а два другие нечётны.

Далее, в уравнении

$$x^3 + y^3 = z^3, \quad \text{или иначе} \quad x^3 + y^3 + (-z)^3 = 0 \tag{7.17}$$

переменные x, y и $-z$ равноправны, и следовательно, можно считать, что, например, z чётно, а x и y нечётны. Среди всех таких троек x, y, z, являющихся решениями уравнения (7.17), выберем ту, для которой величина $|z|$ наименьшая.

Предоставим слово самому Эйлеру:

> *«Если, следовательно, x и y нечётны, ясно, что как их сумма, так и разность будут числами чётными. Пусть, следовательно,* $\dfrac{x+y}{2} = p$ *и* $\dfrac{x-y}{2} = q$, *тогда мы будем иметь* $x = p + q$ *и* $y = p - q$, *откуда следует, что одно из двух чисел p и q должно быть чётным и другое нечётным*[1]. *Итак, мы имеем*
>
> $$x^3 + y^3 = (p+q)^3 + (p-q)^3 = 2p^3 + 6pq^2 = 2p\,(p^2 + 3q^2),$$
>
> *так что речь идёт о том, чтобы доказать, что это произведение* $2p\,(p^2 + 3q^2)$ *не может стать кубом; ... если* $2p\,(p^2 + 3q^2)$ *было бы кубом, этот куб был бы чётным и, следовательно, делился бы на 8, следовательно, необходимо, чтобы восьмая часть нашей формулы, или* $\dfrac{1}{4}p\,(p^2 + 3q^2)$, *была бы числом целым и, сверх того, кубом. Однако мы знаем, что одно из чисел p и q чётное, а другое нечётное; так как* $p^2 + 3q^2$ *должно быть числом нечётным, вовсе не делящимся на 4, нужно, чтобы таковым было p или чтобы $p/4$ было числом целым.»*

Продолжим далее сами, не отступая, впрочем, от хода доказательства Эйлера, но лишь несколько сокращая его изложение.

[1] Кроме того, очевидно, что p и q взаимно просты.

Если бы $p/4$ и p^2+3q^2 были бы взаимно простыми, то из того, что их произведение $\frac{1}{4}p\,(p^2+3q^2)$ является кубом, следовало бы, что и каждый из сомножителей также куб. Другой случай (когда $p/4$ и p^2+3q^2 имеют общий делитель) требует особого рассмотрения. Поскольку p и q взаимно просты, то взаимно простыми будут p и $3q^2$, если только p не делится на 3. Но тогда, очевидно, $p/4$ будет взаимно просто с p^2+3q^2, а потому сами эти числа. как сказано выше, являются кубами. Согласно лемме Эйлера (см. основной текст) числа p и q записываются в виде

$$p=s(s^2-9t^2),\quad q=3t(s^2-t^2),$$

где s и t — некоторые целые числа (взаимно простые, так же, как p и q). При этом, так как q нечётно, то t нечётно, а s чётно.

Далее, вспомним, что кубом является и $p/4$, но тогда то же самое верно и для $2p$. Иначе говоря, число

$$2s(s^2-9t^2)=2s(s-3t)(s+3t)$$

является кубом. Нетрудно проверить (читатель это сделает сам), что сомножители $2s$, $s-3t$, $s+3t$ попарно взаимно просты, и потому каждый из них, так же, как и всё произведение, является кубом. Полагая

$$x_1^3=s-3t,\qquad y_1^3=s+3t,\qquad z_1^3=2s,$$

приходим к равенству $x_1^3+y_1^3=z_1^3$, означающему, что тройка (x_1,y_1,z_1) есть решение уравнения (7.17). Наконец,

$$|z_1^3|=|2s|<|2s(s^2-9t^2)|=|2p|<|2p(p^2+3q^2)|=|z^3|,$$

и значит, в решении (x_1,y_1,z_1) имеем $|z_1|<|z|$. Последнее находится в противоречии с первоначальным выбором решения (x,y,z).

Осталось рассмотреть случай, когда p делится на 3, то есть имеет вид $p=3r$ (где r — целое число, делящееся на 4). Тогда

$$\frac{z^3}{8}=\frac{1}{4}p\,(p^2+3q^2)=\frac{3}{4}r\,(9r^2+3q^2)=\frac{9}{4}r\,(3r^2+q^2).$$

Теперь уже взаимно простыми (проверьте это) будут сомножители

$$\frac{9}{4}r \quad \text{и} \quad 3r^2+q^2,$$

и следовательно, они являются кубами. По лемме Эйлера

$$q=s(s^2-9t^2), \quad r=3t\,(s^2-t^2),$$

где s и t — некоторые целые числа (снова взаимно простые), причём t чётно, а s нечётно.

Далее, целое число

$$\frac{8}{27}\cdot\frac{9}{4}r=2t\,(s^2-t^2)=2t\,(s-t)\,(s+t)$$

является кубом. Кубами, ввиду их взаимно простоты, являются и сомножители $2t$, $s-t$, $s+t$. Полагая

$$x_1^3=s-t, \quad y_1^3=2t, \quad z_1^3=s+t,$$

получаем, что тройка (x_1,y_1,z_1) есть решение уравнения (7.17). При этом снова нетрудно заключить, что, вопреки предположению, $|z_1|<|z|$. Полученное противоречие окончательно доказывает, что уравнение (7.17) не имеет нетривиальных решений в целых числах.

ГЛАВА 8

ГИПЕРЧИСЛА

Вы помните, вероятно, что Гамильтон придумал способ описания комплексных чисел, свободный от всякого рода мистических их толкований. По Гамильтону комплексные числа есть пары действительных чисел с определёнными правилами действий над этими парами. И важно то, что полученная числовая система, во-первых, сохраняет все алгебраические свойства обычных чисел и, во-вторых, содержит их в себе. Сделав этот шаг, Гамильтон задумался над тем, как сделать и следующий. А что если рассматривать не пары, а скажем, тройки, четвёрки действительных чисел, вообще произвольные их наборы

$$(\alpha_1, \alpha_2, \ldots, \alpha_n), \quad \alpha_i \in \mathbb{R},$$

любой длины n? Нельзя ли, используя их, получить новые, гиперкомплексные числа, которые по своим свойствам также были бы похожи на числа уже известные? Всё, разумеется, зависит от того, как определить операции над этими гиперкомплексными числами. Путеводной нитью в долгих размышлениях Гамильтона на эту тему были уже известные нам геометрические соображения. Ведь комплексные числа (пары) можно использовать для описания преобразований плоскости — сдвигов, вращений, гомотетий. Быть может, тогда тройки (Гамильтон называет их триплетами) окажутся пригодными для описания аналогичных преобразований трёхмерного пространства. Собственно, со сложением их не было проблем — триплеты надо складывать, как и векторы, и понятно, что такой операции будут соответствовать сдвиги

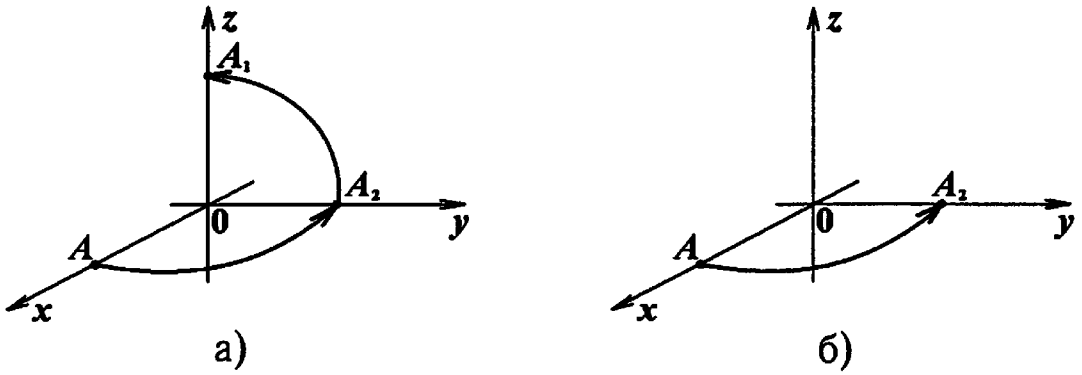


Рис. 17.

пространства. А вот загадка умножения не давалась в руки, и на традиционный утренний вопрос сыновей: «Папа, умеешь ли ты уже умножать триплеты?», сэр Гамильтон каждый раз вынужден был отвечать: «Нет, я их умею только складывать и вычитать». Впрочем, как постепенно выяснялось, причины для этого были достаточно веские. Целый клубок трудностей вырос на пути решения, казалось бы, ясной задачи. Обескураживало, например, то, что умножение вращений в пространстве, в отличие от умножения плоских вращений вокруг данной точки, уже не обладает свойством коммутативности. Это иллюстрируют рисунки.

На первом из них (рис. 17 а)) показано, во что переходит точка A при композиции двух вращений: сначала вокруг оси Oz, затем вокруг Ox. При выполнении этих вращений в ином порядке, как показывает рис. 17 б), и результат оказывается иным — образом точки A является тогда точка A_2, а не A_1, как в первом случае. Но ведь композиции преобразований должно как-то отвечать произведение соответствующих им триплетов, значит, скорее всего, тщетны попытки искать такое правило их умножения, которое сохраняло бы все свойства числового умножения. Во всяком случае свойством коммутативности наверняка придётся пожертвовать. Однако не выручал и отказ от коммутативности — хорошего умножения триплетов всё равно не получалось[1]. Объяснялось это довольно просто — дело в том, что триплетов недостаточно, чтобы охарактеризовать произвольные вра-

[1] Позднее было доказано, что нельзя определить умножение троек, обладающее всеми свойствами числового умножения, исключая коммутативность.

щения пространства с растяжениями, точнее, композиции вращений с гомотетиями. В самом деле, для задания вращения вокруг точки нужно указать направление оси этого вращения, или иначе направляющие косинусы $\cos\alpha$, $\cos\beta$, $\cos\gamma$ такой оси (косинусы углов, которые ось вращения образует с осями координат). Легко видеть, что $\cos^2\alpha+\cos^2\beta+\cos^2\gamma=1$, поэтому два угла из трёх и знак косинуса третьего угла определяют её положение. Третий параметр определяет угол поворота вокруг этой оси. Наконец, четвёртый необходим для задания коэффициента гомотетии. Выходит, что учиться перемножать надо не тройки, а четвёрки действительных чисел. Для них Гамильтон также придумал название — *кватернионы*. Трудность была ещё и в том, что связь кватернионов с преобразованиями была не столь непосредственной, как в случае комплексных чисел.

В множестве всех кватернионов[1]

$$\{(a,b,c,d) \mid a,b,c,d\in\mathbb{R}\}$$

Гамильтон по аналогии с комплексными числами выделяет четвёрки

$$(1,0,0,0);\ (0,1,0,0);\ (0,0,1,0);\ (0,0,0,1).$$

Первая из них должна играть при умножении роль единицы, остальным же, которые Гамильтон обозначает

$$i=(0,1,0,0);\ j=(0,0,1,0);\ k=(0,0,0,1),$$

отводится роль мнимых единиц в том смысле, что

$$i^2=j^2=k^2=(-1,0,0,0)=-1. \tag{8.1}$$

Произвольный кватернион записывается тогда в виде

$$a+bi+cj+dk,\quad a,b,c,d\in\mathbb{R}, \tag{8.2}$$

Число a Гамильтон называет скалярной частью кватерниона, а триплет $bi+cj+dk$ — векторной его частью[2]. Однако формулы (8.1) ещё недостаточны для того, чтобы можно было перемножать любые кватернионы. Надо ещё понять (и это главное), как

[1] В честь Гамильтона (Hamilton) это множество теперь обозначается $\mathbb{H}$.

[2] Теперь скалярная часть называется действительной частью, а векторная — мнимой частью кватерниона.

следует определить перемножение мнимых единиц. Пробы и поиски наконец-то приводят Гамильтона к следующим (в сущности, единственно возможным) формулам

$$ij = k, \quad jk = i, \quad ki = j. \tag{8.3}$$

Из них и формул (8.1) вытекает, что при изменении порядка сомножителей результат должен получиться иным. В самом деле, например, $(ij)^2 = k^2 = -1$. Умножая справа получившееся равенство сначала на j, затем на i, получаем $(ij)^2 ji = -ji$, или

$$(ij)(ij)ji = -ji.$$

Так как $i^2 = j^2 = -1$, то, с учётом ассоциативности умножения, из последнего равенства вытекает, что $ij = -ji$. Совершенно так же обстоит дело с остальными произведениями мнимых единиц. Следовательно,

$$ji = -k, \quad kj = -i, \quad ik = -j. \tag{8.4}$$

Теперь, предполагая выполненным закон дистрибутивности, можно найти произведение любых двух кватернионов

$$\begin{aligned}(a + bi + cj + dk)(a_1 + b_1 i + c_1 j + d_1 k) = \\ = (aa_1 - bb_1 - cc_1 - dd_1) + (ab_1 + ba_1 + cd_1 - dc_1)\, i + \\ + (ac_1 + ca_1 + db_1 - bd_1) j + (ad_1 + da_1 + bc_1 - cb_1)\, k.\end{aligned} \tag{8.5}$$

Итак, на множестве кватернионов определены операции сложения и умножения, которые, как можно показать, обладают всеми алгебраическими свойствами числовых операций, за исключением коммутативности умножения. Собственно говоря, в проверке нуждаются лишь свойства ассоциативности и обратимости кватернионного умножения. Чтобы убедиться в выполнимости первого из них, достаточно проверить ассоциативность умножения мнимых единиц i, j, k (читатель может сделать это простым перебором всех возможностей). Остановимся на менее очевидном втором свойстве, используя для этой цели аналогию с комплексными числами.

Для произвольного кватерниона $\xi = a+bi+cj+dk$ рассмотрим *сопряжённый* ему кватернион $\overline{\xi} = a-bi-cj-dk$. Их произведение в любом порядке согласно (8.5) равно

$$\xi\overline{\xi} = \overline{\xi}\xi = a^2 + b^2 + c^2 + d^2$$

и является, следовательно, действительным числом, которое называют *нормой* кватерниона

$$N(\xi) = a^2 + b^2 + c^2 + d^2.$$

Если $\xi \neq 0$ (равенство $\xi = 0$ по определению означает, что $a = b = c = d = 0$), то $N(\xi) > 0$, и тогда кватернион

$$\eta = \frac{\overline{\xi}}{N(\xi)} = \frac{a - bi - cj - dk}{a^2 + b^2 + c^2 + d^2},$$

очевидно, удовлетворяет соотношениям

$$\eta\xi = \xi\eta = 1$$

и значит, является элементом, обратным к ξ.

Заметим попутно, что норма кватерниона обладает тем же свойством, что и норма комплексного числа: для любых двух кватернионов ξ_1 и ξ_2

$$N(\xi_1\xi_2) = N(\xi_1)N(\xi_2), \tag{8.6}$$

то есть норма произведения равна произведению норм. Это доказывает следующая цепочка равенств:

$$N(\xi_1\xi_2) = (\xi_1\xi_2)(\overline{\xi_1\xi_2}) = \xi_1\xi_2\,(\overline{\xi}_2\overline{\xi}_1) = \xi_1\,(\xi_2\overline{\xi}_2)\overline{\xi}_1 =$$
$$= \xi_1 N(\xi_2)\overline{\xi}_1 = \xi_1\overline{\xi}_1 N(\xi_2) = N(\xi_1)N(\xi_2).$$

Соотношение (8.6) имеет связь с одной задачей теории чисел. Из него следует, что произведение суммы четырёх квадратов целых чисел на подобную же сумму есть снова сумма четырёх квадратов целых чисел.

Читатель вправе, конечно, спросить, какова же всё-таки связь кватернионов с преобразованиями трёхмерного пространства.

Здесь, как уже отмечалось, нет прямой аналогии с комплексными числами. Дело в том, что соответствия $\xi \to \xi_0\xi$ или $\xi \to \xi\xi_0$, где $\xi_0 \neq 0$ — фиксированный кватернион, определяют, вообще говоря, преобразования в четырёхмерном пространстве $\mathbb{H}$, а не в трёхмерном пространстве. Однако Гамильтон заметил, что если ξ имеет нулевую скалярную часть, то есть $\xi = xi + yj + zk$, то выражение $\xi_0\xi\xi_0^{-1}$ также обладает этим свойством для любого кватерниона ξ_0. Проверить это читатель может непосредственным вычислением. Таким образом, соответствие

$$\xi \to \xi_0\xi\xi_0^{-1} \qquad (8.7)$$

определяет уже преобразование в пространстве трёхмерных векторов (рис. 18).

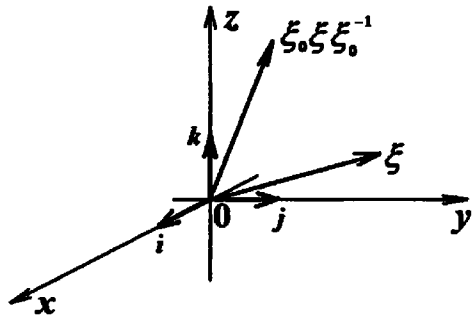


Рис. 18.

При этом согласно (8.6)

$$N(\xi_0\xi\xi_0^{-1}) = N(\xi_0)N(\xi)N(\xi_0^{-1}) = N(\xi_0)N(\xi)N(\xi_0)^{-1} = N(\xi),$$

и значит, преобразование (8.7) сохраняет величину

$$N(\xi) = x^2 + y^2 + z^2,$$

которая в данном случае равна квадрату длины вектора $\xi = xi + yj + zk$. Можно также показать, что оно сохраняет и ориентацию трёхмерного пространства векторов ξ. Линейное преобразование, сохраняющее длину векторов и ориентацию пространства, является, как известно из геометрии, вращением вокруг оси, проходящей через начало координат. Таким образом, всякому кватерниону $\xi_0 \neq 0$ соответствует вращение трёхмерного пространства, при котором образом вектора ξ является вектор $\xi_0\xi\xi_0^{-1}$, причём произведению кватернионов ξ_0 и ξ_0' будет отвечать последовательное выполнение (произведение, композиция) соответствующих им вращений:

$$\xi \to \xi_0\xi\xi_0^{-1} \to \xi_0'(\xi_0\xi\xi_0^{-1})\xi_0'^{-1} = (\xi_0'\xi_0)\xi(\xi_0'\xi_0)^{-1}.$$

Можно совершенно точно описать вращение, задаваемое соответствием (8.7), то есть указать ось и угол поворота. Если $\xi_0 = a_0 + b_0 i + c_0 j + d_0 k$, то векторная часть $\widetilde{\xi_0} = b_0 i + c_0 j + d_0 k$ как раз и является осью поворота, а угол поворота φ вокруг этой оси определяется из равенства

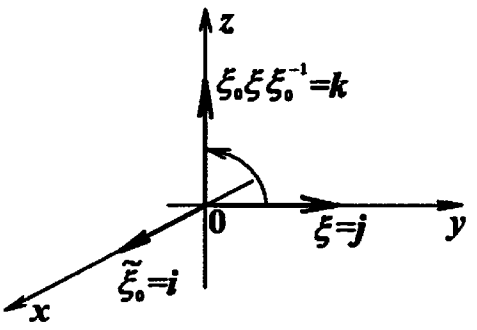


Рис. 19.

$$\cos\frac{\varphi}{2} = \frac{a_0}{\sqrt{a_0^2 + b_0^2 + c_0^2 + d_0^2}}.$$

Например, если $\xi = j$, а $\xi_0 = 1 + i$, то $\cos\dfrac{\varphi}{2} = \dfrac{1}{\sqrt{2}}$, и вектор

$$\xi_0\xi\xi_0^{-1} = (1+i)j\,\frac{1-i}{2} = \frac{j - ji + ij - iji}{2} = k$$

получается, как и должно быть, из вектора ξ поворотом на угол $\varphi = \pi/2$ (рис. 19).

Доказательство этого утверждения и другие сведения о кватернионах читатель может найти в книге [15].

Гамильтон работал преимущественно в области механики, и его интерес к кватернионам объяснялся во многом тем, что они обещали быть полезным инструментом при решении ряда механических задач. Так, в общем-то, и оказалось, и одним из примеров, подтверждающих это, является так называемая задача о сложении поворотов. Формулируется она следующим образом.

Пусть тело поворачивается на угол φ вокруг оси OA, проходящей через заданную точку O, и вслед за этим вновь поворачивается на угол φ' вокруг другой оси OA'. Спрашивается, вокруг какой оси и на какой угол надо повернуть тело, чтобы оно из первого положения сразу перешло в третье.

Эта задача, правда, была решена ещё Эйлером средствами аналитической геометрии, однако её решение с помощью кватернионов особенно естественно и прозрачно. В самом деле, пусть кватернион $\xi_0 = a_0 + b_0 i + c_0 j + d_0 k$ описывает указанным вы-

ше образом первое вращение[1], и пусть ξ_0' — кватернион, соответствующий второму из них. Но тогда, как следует из предыдущих рассуждений, результирующий поворот соответствует произведению $\xi_0'\xi_0$ этих двух кватернионов. Зная произведение, мы сумеем найти ось и угол результирующего поворота.

Кватернионы — важный, в некотором смысле исключительный, но в то же время не единственный пример гиперкомплексной системы. О других примерах (дуальных, или двойных, числах, октавах Кэли) и об общем методе построения гиперкомплексных чисел читатель может также прочесть в книге [15]. Но, пожалуй, самыми известными в обширном многообразии гиперкомплексных систем является так называемые *алгебры матриц*. Эта известность объясняется чрезвычайно большим значением матриц во многих областях математики и её приложениях.

Матрицей называется прямоугольная таблица чисел вида

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}. \tag{8.8}$$

Каждый элемент матрицы снабжен двумя индексами, первый из которых указывает номер строки, а второй — номер столбца, на пересечении которых находится соответствующий элемент. Про матрицу, у которой m строк и n столбцов, говорят, что она имеет размеры $m \times n$. В случае, когда число строк и столбцов одинаково ($m = n$), матрица называется *квадратной* порядка n. Если матрица состоит из одной строки ($m = 1$), её называют *вектор-строкой*. Точно так же, в случае одностолбцовой матрицы пользуются термином *вектор-столбец*. Часто матрицу обозначают одной буквой (A, B и т. д.), а иногда используется обозначение (a_{ij}).

Для матриц довольно понятным образом (так же, как и в любой гиперкомплексной системе) определяются операции сложения и умножения на число, а именно, для любых двух матриц

[1] Хотя таких кватернионов много, можно показать, что любые два из них отличаются друг от друга лишь ненулевым скалярным множителем.

одного и того же размера $m \times n$

$$\begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} + \begin{pmatrix} b_{11} & \dots & b_{1n} \\ \vdots & & \vdots \\ b_{m1} & \dots & b_{mn} \end{pmatrix} =$$

$$= \begin{pmatrix} a_{11}+b_{11} & \dots & a_{1n}+b_{1n} \\ \vdots & & \vdots \\ a_{m1}+b_{m1} & \dots & a_{mn}+b_{mn} \end{pmatrix}$$

или $(a_{ij}) + (b_{ij}) = (a_{ij} + b_{ij})$, и для любой матрицы и любого числа α

$$\alpha \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} = \begin{pmatrix} \alpha a_{11} & \dots & \alpha a_{1n} \\ \vdots & & \vdots \\ \alpha a_{m1} & \dots & \alpha a_{mn} \end{pmatrix}$$

или $\alpha(a_{ij}) = (\alpha a_{ij})$. Другими словами, матрицы складываются и умножаются на число *поэлементно*.

Но вот операция умножения матриц отличается бóльшим своеобразием. Рассмотрим сначала эту операцию для случая квадратных матриц одного порядка n. Если $A = (a_{ij})$ и $B = (b_{ij})$ — две такие матрицы, то их произведением

$$A \cdot B = C = (c_{ij})$$

называется квадратная матрица C порядка n, произвольный элемент c_{ij} которой вычисляется по правилу

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \ldots + a_{in}b_{nj} = \sum_{k=1}^{n} a_{ik}b_{kj}. \qquad (8.9)$$

Иначе говоря, чтобы получить элемент i-ой строки и j-ого столбца матрицы $C = AB$, нужно взять сумму произведений элементов i-ой строки матрицы A на соответствующие элементы j-ого столбца матрицы B. Например,

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

Имеется немало причин для рассмотрения столь странной на первый взгляд матричной операции — читатель о них ещё узнает.

Отметим пока лишь, что умножение матриц (как, впрочем, и умножение кватернионов) некоммутативно, то есть, вообще говоря, $AB \neq BA$. Так, перемножив две предыдущие матрицы в обратном порядке, получим другую матрицу:

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

Вместе с этим, нетрудно доказать (предоставляем это читателю), что умножение матриц ассоциативно:

$$(A \cdot B) \cdot C = A \cdot (B \cdot C).$$

Имеется и нейтральный элемент (единица) для нашего умножения. Такую роль играет матрица

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

называемая *единичной матрицей* порядка n. Действительно, из формулы (8.9) следует, что

$$EA = AE = A$$

для любой квадратной матрицы A порядка n.

Одним из существенных отличий построенной алгебры квадратных матриц[1] порядка n от алгебры кватернионов является то, что в первой из них не всегда возможно деление на ненулевой элемент (нулем является матрица, все элементы которой нулевые). Так, например, можно убедиться, что никакая матрица X порядка 2 не может служить решением уравнения

$$X \cdot \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

[1] Эта алгебра обозначается $\mathrm{Mat}_n(\mathbb{R})$. В общем случае, если матричные элементы принадлежат полю $\mathbb{F}$, алгебра матриц порядка n обозначается $\mathrm{Mat}_n(\mathbb{F})$.

Это обстоятельство носит отнюдь не случайный характер. Теорема, принадлежащая знаменитому немецкому алгебраисту Г. Фробениусу (1849–1917), утверждает, что среди всех гиперкомплексных систем с ассоциативным умножением имеется лишь три, в которых возможно деление на любой ненулевой элемент, — это поле действительных чисел $\mathbb{R}$, поле комплексных чисел $\mathbb{C}$ и алгебра кватернионов $\mathbb{H}$.

Ещё несколько задержимся на операции умножения матриц. Заметим, что сформулированное выше правило можно распространить и на прямоугольные матрицы, не являющиеся квадратными. Формула (8.9) позволяет это сделать, если число столбцов первой матрицы A совпадает с числом строк второй матрицы B. Например,

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

При этом получающаяся матрица имеет столько же строк, сколько первый сомножитель A, и столько же столбцов, сколько второй сомножитель B.

Матрицы очень удобны для компактной записи различных соотношений в математике, и это — одна из многих причин, определяющих полезность матричного исчисления. Важным примером такого рода являются системы линейных уравнений, с которыми часто приходится иметь дело при решении самых разнообразных математических и прикладных задач.

Системой линейных уравнений с n неизвестными $x_1, x_2, \ldots, x_n$ называется совокупность уравнений вида

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m, \end{cases} \tag{8.10}$$

где a_{ij} и b_i — произвольные действительные числа. Количество уравнений в системе может быть любым и в общем случае никак не связано с числом неизвестных. Коэффициенты при неизвест-

ных a_{ij} (подобно элементам матрицы) имеют двойную нумерацию: первый индекс i соответствует номеру уравнения, а второй индекс j — номеру неизвестного, при котором стоит данный коэффициент. Всякое решение системы понимается как набор значений неизвестных $(\alpha_1, \alpha_2, \ldots, \alpha_n)$, обращающих каждое из уравнений системы в верное равенство.

Рассмотрим теперь матрицу

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

составленную из коэффициентов при неизвестных системы (8.10) (кратко, матрицу системы). Пусть также

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \ldots \\ x_n \end{pmatrix}, \quad B = \begin{pmatrix} b_1 \\ b_2 \\ \ldots \\ b_m \end{pmatrix}$$

— столбцы неизвестных и свободных членов. Тогда произведение AX есть матрица с m строками и одним столбцом, то есть вектор-столбец, элементы которого вычисляются согласно формуле (8.9). Таким образом,

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

Каждое из уравнений системы (8.10) означает равенство соответствующих координат векторов AX и B, и вся система в целом означает тогда равенство $AX = B$. Полученная нами компактная запись называется матричной формой системы линейных уравнений. Подобная матричная запись встречается во множестве других ситуаций и широко используется в математической, физической, технической и экономической литературе.

ГЛАВА 9

ЭТИ СТРАННЫЕ КОЛЬЦА

Если основным источником понятия группы были преобразования и подстановки, то к другому не менее важному понятию кольца привели исследования (о них говорилось в предыдущих главах) различных числовых систем и поиски новых алгебраических образований, подобных числам. Термин кольцо впервые появился в алгебраических работах Д. Гильберта (1862–1943), одного из крупнейших математиков нового времени. Трудно объяснить, почему для наименования алгебраических структур с двумя операциями им было выбрано такое слово. Не исключено, что это было связано с особенностями строения того объекта, который мы сейчас называем кольцом классов вычетов. В любом случае введение Гильбертом нового понятия было в полном согласии со всегдашним его стремлением к аксиоматизации математики.

Если выделить и собрать вместе все те свойства операций, которые выполняются в каждой из уже рассмотренных нами структур со сложением и умножением их элементов (начиная с кольца целых чисел и заканчивая алгебрами матриц), то получится такой список:

1. Сложение коммутативно: $a + b = b + a$.
2. Сложение ассоциативно: $(a + b) + c = a + (b + c)$.
3. Существует нулевой элемент 0 такой, что $a + 0 = 0 + a = a$.
4. Для всякого элемента a существует противоположный элемент $(-a)$ такой, что $(-a) + a = a + (-a) = 0$.

5. Умножение ассоциативно: $(ab)c = a(bc)$.

6. Умножение дистрибутивно относительно сложения: $a(b+c) = ab + ac$, $(b+c)a = ba + ca$.

Всякая алгебраическая структура с операциями сложения и умножения, обладающими перечисленными свойствами, называется *кольцом*.

Таким образом, в данное нами определение вписываются многие из рассмотренных ранее числовых систем: $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, и вообще кольца и поля алгебраических чисел. Кольцами в смысле данного определения являются также кольца классов вычетов $\mathbb{Z}_n$, кольца многочленов $\mathbb{F}[x]$, равно как и алгебры квадратных матриц $\mathrm{Mat}_n(\mathbb{R})$ относительно операций сложения и умножения матриц. Позднее мы познакомимся и с другими примерами, но уже сейчас ясно, что понятие кольца, как и группы, обладает очень большой степенью общности.

У читателя может возникнуть (или уже возник раньше) вопрос, что побуждает математиков к образованию столь общих понятий и связанных с ними довольно абстрактных теорий. Не является ли это их простой прихотью, пустой игрой воображения? В нескольких словах ответить можно было бы так. Наверное, прежде всего математиками руководит стремление объединить общей идеей многие из накопившихся в их науке фактов, ранее казавшихся разрозненными, попытаться найти им аналогию и в иных, ранее не изучавшихся объектах. Безусловно, немалую роль играет поддерживаемая интуицией убеждённость, что развитие общей теории позволит применить её затем к разгадке трудных, нерешённых ещё математических (и не только математических) проблем. Мы надеемся, что на страницах этой книги читатель найдёт подтверждение этим словам. Пожалуй, процесс развития общих понятий и теорий движется также и желанием узнать, какое место занимает в них уже известное и изученное, и что ещё новое, необычное они в себе скрывают.

Возвращаясь от этих философских отступлений к понятию кольца, мы как раз сможем заметить, сколь неожиданными могут быть свойства определённых в нём операций. Возможная некоммутативность умножения в кольцах (кольцо матриц) нас едва ли

уже удивит — с подобным обстоятельством мы уже сталкивались прежде. Однако в кольцах матриц, как, впрочем, и в кольцах классов вычетов и в других, мы обнаруживаем и более удивительные вещи. Это, прежде всего, присутствие в них делителей нуля и нильпотентных элементов. Определяются они следующим образом.

Ненулевой элемент a кольца K называется *делителем нуля*, если существует такой элемент $b \in K$, также отличный от нуля, что

$$ab = 0 \quad \text{или} \quad ba = 0. \tag{9.1}$$

Ненулевой элемент a кольца K называется *нильпотентным элементом*, если некоторая его натуральная степень равна нулю:

$$a^n = 0. \tag{9.2}$$

Заметьте, что нильпотентный элемент обязательно является делителем нуля. Обратное, вообще говоря, неверно. Ясно, что в числовых кольцах равенства (9.1) и (9.2) исключены, если числа a и b отличны от 0. В то же время в кольцах матриц и классов вычетов нет недостатка как в делителях нуля, так и в нильпотентных элементах. К примеру, простое вычисление показывает, что ненулевая матрица

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

такова, что её квадрат равен нулю:

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

Достаточно общим примером нильпотентной матрицы (а следовательно, и делителя нуля) порядка n служат матрицы вида

$$A = \begin{pmatrix} 0 & a_{12} & a_{13} & \dots & a_{1n} \\ 0 & 0 & a_{23} & \dots & a_{2n} \\ 0 & 0 & 0 & \dots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}.$$

Читатель может самостоятельно убедиться, что $A^n = 0$.

Что касается кольца классов вычетов $\mathbb{Z}_n$, то оно содержит делители нуля тогда и только тогда, когда n — число составное: $n = n_1 n_2$. Если последнее выполнено, то, очевидно, $\overline{n}_1 \cdot \overline{n}_2 = \overline{n} = \overline{0}$, хотя $\overline{n}_1, \overline{n}_2 \neq \overline{0}$. Справедливость обратного утверждения будет ясна из дальнейшего. Впрочем, читатель в состоянии доказать его уже и сейчас, заодно выяснив, какому условию должно удовлетворять n, чтобы кольцо $\mathbb{Z}_n$ содержало нильпотентные элементы (простейшим примером такого кольца является $\mathbb{Z}_4$ — в нём выполнено равенство $\overline{2}^2 = \overline{0}$).

Следствием этих особенностей являются и другие необычные свойства. Например, числовое уравнение $x^2 = x$ имеет лишь два решения: 0 и 1. То же самое уравнение, рассматриваемое в произвольном кольце, может обладать и иными решениями, их число может быть даже бесконечным. Примеры тому можно найти и в кольцах матриц, и в кольцах классов вычетов. Всё это требует соблюдать известную осторожность при оперировании с элементами произвольных колец.

Подчеркнём ещё раз — аксиомы кольца отобраны таким образом, чтобы объединить этим понятием большое разнообразие важнейших для алгебры объектов. Не удивительно поэтому, что из этих аксиом не вытекают такие привычные для числовых колец и колец многочленов свойства, как коммутативность умножения и отсутствие делителей нуля. Добавляя эти свойства к списку аксиом 1–6, мы приходим к понятию *области целостности* или иначе *целостного кольца* — коммутативного кольца без делителей нуля[1]. Названием «область целостности» подчёркивается сходство таких колец с кольцом целых чисел, иногда весьма отдалённое.

Тем не менее, наше замечание о сходстве произвольных областей целостности с кольцом целых чисел было небезосновательным. Оно подчёркивается тем, что для любой области целостности возможна процедура, подобная расширению кольца целых чисел $\mathbb{Z}$ до поля рациональных чисел $\mathbb{Q}$. Но перед разговором

[1] Часто в определение области целостности включают ещё требование существования единицы, то есть такого элемента $1 \in K$, что $1 \cdot a = a \cdot 1 = a$ для любого $a \in K$.

об этом нужно познакомиться с общеалгебраическим понятием поля, которое, конечно, имеет своими непосредственными предшественниками рассматривавшиеся нами числовые поля.

Полем $\mathbb{F}$ называется коммутативное кольцо с единицей (при этом $1 \neq 0$), в котором каждый ненулевой элемент обратим, то есть для всякого $a \neq 0$ из $\mathbb{F}$ существует элемент $a^{-1} \in \mathbb{F}$ такой, что $a \cdot a^{-1} = a^{-1} \cdot a = 1$.

Сразу заметим, что в поле отсутствуют делители нуля, так как последние не могут быть обратимыми элементами (почему?). Иными словами, поле — это всегда область целостности.

Содержащиеся в определении поля дополнительные по сравнению с кольцом требования достаточны (и необходимы) для того, чтобы обеспечить в этой алгебраической структуре возможность деления на любой ненулевой элемент. Действительно, всякое уравнение вида

$$ax = b, \quad a \neq 0,$$

в поле однозначно разрешимо. Решением его, очевидно, является элемент $x = a^{-1}b$.

Нам уже известны примеры числовых полей. Они, разумеется, удовлетворяют нашему общему определению. Иной и очень важный пример поля доставляют кольца классов вычетов по простому модулю. Так, каждое из колец $\mathbb{Z}_2$, $\mathbb{Z}_3$, $\mathbb{Z}_5$ образует поле — обратимость любого ненулевого элемента в них видна из таблиц умножения в этих кольцах:

$\mathbb{Z}_2$

$\times$	$\overline{0}$	$\overline{1}$
$\overline{0}$	$\overline{0}$	$\overline{0}$
$\overline{1}$	$\overline{0}$	$\overline{1}$

$\mathbb{Z}_3$

$\times$	$\overline{0}$	$\overline{1}$	$\overline{2}$
$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$
$\overline{1}$	$\overline{0}$	$\overline{1}$	$\overline{2}$
$\overline{2}$	$\overline{0}$	$\overline{2}$	$\overline{1}$

$\mathbb{Z}_5$

$\times$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$	$\overline{4}$
$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$
$\overline{1}$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$	$\overline{4}$
$\overline{2}$	$\overline{0}$	$\overline{2}$	$\overline{4}$	$\overline{1}$	$\overline{3}$
$\overline{3}$	$\overline{0}$	$\overline{3}$	$\overline{1}$	$\overline{4}$	$\overline{2}$
$\overline{4}$	$\overline{0}$	$\overline{4}$	$\overline{3}$	$\overline{2}$	$\overline{1}$

Вообще, справедливо следующее утверждение.

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

Понятна необходимость последнего условия. Действительно, если n — число составное, то, как было выяснено выше, $\mathbb{Z}_n$ обязательно содержит делители нуля, а в поле это исключено.

Достаточность тоже можно считать уже доказанной. Ведь в главе 2 мы убедились, что в кольцах вычетов по простому модулю возможно однозначное деление на любой ненулевой элемент. Всё же мы дадим ещё одно доказательство, которое одновременно содержит в себе алгоритм отыскания обратного элемента.

Пусть $n = p$ — простое число и $\overline{k}$ — ненулевой класс вычетов в $\mathbb{Z}_p$. Числа p и k взаимно просты, поэтому (см. дополнение к главе 6) существуют такие целые s и t, что

$$pt + ks = 1. \tag{9.3}$$

Напомним, что отыскание указанных чисел опиралось на алгоритм Евклида. При переходе к классам вычетов равенство (9.3) превратится в следующее

$$\overline{p}\,\overline{t} + \overline{k}\,\overline{s} = \overline{1}.$$

Однако $\overline{p} = \overline{0}$, поэтому мы получаем

$$\overline{k} \cdot \overline{s} = \overline{1}.$$

Последнее и означает, что класс $\overline{s} \in \mathbb{Z}_p$ является элементом, обратным к классу $\overline{k}$.

Итак, в случае простого модуля p мы имеем дело с полем классов вычетов $\mathbb{Z}_p$.

Пример. В поле $\mathbb{Z}_{17}$ найдем класс, обратный к $\overline{8}$. Так как $17 = 2 \cdot 8 + 1$, то $17 - 2 \cdot 8 = 1$. Для классов вычетов получаем $\overline{(-2)} \cdot \overline{8} = \overline{1}$. Очевидно, $\overline{(-2)} = \overline{15}$, поэтому $\overline{15}$ — искомый обратный.

Поля вычетов по модулю p являются простейшими примерами конечных полей, называемых ещё полями Галуа. С другими их примерами мы встретимся позднее. Теория таких полей обладает глубиной, прозрачностью и красотой. Добавим к этому, что конечные поля имеют много применений. Одно из них (для нужд теории кодирования) рассматривается в последней главе этой книги.

Любое поле всегда образует целостное кольцо, обратное, вообще говоря, неверно (вспомните хотя бы кольцо $\mathbb{Z}$). Интересно, однако, что всякая область целостности вкладывается в поле. Что это означает, объясняет конструкция *поля частных*, часто используемая в алгебре.

Пусть K — произвольное целостное кольцо (если оно совпадает с $\mathbb{Z}$, конструкция даёт корректное определение поля рациональных чисел $\mathbb{Q}$). Рассмотрим множество пар (a, b), где $a, b \in K$, причём $b \neq 0$. Каждой такой паре мы хотим уготовить роль дроби, у которой a — числитель, а b — знаменатель. Поступая так, мы исходим, разумеется, из аналогии с обычными рациональными дробями: ведь каждая такая дробь m/n определяется парой целых чисел m и n, где $n \neq 0$. Однако две различные пары (m, n) и (m', n') целых чисел могут быть эквивалентны в том смысле, что определяемые ими рациональные дроби m/n и m'/n' дают одно и то же число (например, $3/9 = 2/6 = 1/3$).

Выполняется это, как хорошо известно, тогда и только тогда, когда $mn' = m'n$. Указанное обстоятельство должно быть, конечно, учтено и в нашей конструкции. Две пары элементов (a, b) и (a', b'), где $b, b' \neq 0$, кольца K будем поэтому называть эквивалентными, отражая это в записи $(a, b) \sim (a', b')$, если в K выполняется равенство $ab' = a'b$. Для каждой пары (a, b) рассмотрим множество всех эквивалентных ей пар (x, y), обозначая его символом a/b

$$a/b = \{(x, y) \mid (x, y) \sim (a, b)\}. \tag{9.4}$$

Заметим, что любые две пары (x_1, y_1) и (x_2, y_2), принадлежащие множеству (9.4), иначе, эквивалентные (a, b), будут эквивалентны и между собой. Действительно, из равенств $x_1 b = y_1 a$ и $x_2 b = y_2 a$ вытекает, что

$$x_1 y_2\, ab = x_2 y_1\, ab, \quad \text{или} \quad ab(x_1 y_2 - x_2 y_1) = 0.$$

Отсутствие в области целостности делителей нуля влечёт тогда

$$x_1 y_2 - x_2 y_1 = 0, \quad \text{или} \quad x_1 y_2 = x_2 y_1,$$

то есть $(x_1, y_1) \sim (x_2, y_2)$.

Далее, если две пары (a,b) и (c,d) не эквивалентны, то определяемые ими множества a/b и c/d не имеют общих элементов. Если предположить противное, то будем иметь, что некоторая пара (x,y) одновременно эквивалентна и (a,b), и (c,d). Но тогда, как и выше, отсюда бы следовало вопреки предположению, что $(a,b) \sim (c,d)$.

Таким образом, множество всех пар оказывается разбитым на непересекающиеся классы эквивалентных между собой пар. Пусть

$$\overline{K} = \{\, a/b \mid a,b \in K;\ b \neq 0 \,\}$$

— множество, элементами которого являются эти классы. Мы будем называть их дробями.

Правила сложения и умножения классов (дробей) скопируем с аналогичных действий над обычными дробями. Пусть r и s — два класса из $\overline{K}$. Возьмём пару (a,b), представляющую r, и пару (c,d), представляющую s. По определению,

$$(a,b) + (c,d) = (ad + bc, bd) \quad \text{и} \quad (a,b) \cdot (c,d) = (ac, bd).$$

Точно так же, для соответствующих классов $r = a/b$ и $s = c/d$ положим

$$a/b + c/d = (ad + bc)/bd \quad \text{и} \quad a/b \cdot c/d = ac/bd. \tag{9.5}$$

Конечно, операции (9.5) над классами могут иметь определённый смысл лишь в том случае, если их результат не зависит от выбора представителей в этих классах[1]. Убедиться, что это так, нетрудно. Действительно, если $(a,b) \sim (a',b')$ и $(c,d) \sim (c',d')$ (иными словами, $a/b = a'/b'$ и $c/d = c'/d'$), то $ab' = a'b$ и $cd' = c'd$, откуда $ab'dd' = a'bdd'$, $cd'bb' = c'dbb'$. Сложив эти равенства, получим, что $(ad + bc)b'd' = (a'd' + b'c')bd$, то есть $(a,b) + (c,d) \sim (a',b') + (c',d')$, $a/b + c/d = a'/b' + c'/d'$. Аналогично, $a/b \cdot c/d = a'/b' \cdot c'/d'$.

Операции, определённые на множестве $\overline{K}$, превращают его в поле. Коммутативность сложения и умножения понятна. Легко проверяется ассоциативность этих операций. Роль нулевого

[1] Положение здесь такое же, как при оперировании с классами вычетов.

элемента играет дробь $0/b$, а единичного — дробь b/b (в качестве b может быть взят любой ненулевой элемент из K). Для дроби a/b противоположной является дробь $-a/b$, а обратной (если $a \neq 0$) — дробь b/a. Выполнен, наконец, и закон дистрибутивности:

$$(a/b + c/d) \cdot e/f = (ad + bc)/bd \cdot e/f = (ade + bce)/bdf = \\ = (adef + bcef)/bdf^2 = ae/bf + ce/df = a/b \cdot e/f + c/d \cdot e/f.$$

Построенное поле $\overline{K}$ называется *полем частных* или *полем отношений* для исходного целостного кольца K, и, что важно, это поле содержит подкольцо, изоморфное K.

В главе 6 мы рассматривали различные примеры колец алгебраических чисел. Все они, разумеется, являются целостными кольцами. Эти примеры показывают, что арифметика в целостных кольцах может очень сильно отличаться от арифметики обычных целых чисел. Главное, в чём это проявляется, — возможная неоднозначность разложения на простые множители. При рассмотрении вопросов делимости в областях целостности важная роль принадлежит понятию *идеала*. Заметим, что это — одно из тех немногих понятий, которое имеет смысл для произвольных колец, а его значение, как мы увидим позднее, далеко не исчерпывается сказанным здесь.

Подмножество I кольца K называется двусторонним *идеалом*, если оно само является кольцом относительно операций на K (короче говоря, подкольцом) и если для любых элементов $a \in K$ и $b \in I$ оба произведения ab и ba принадлежат I.

Так, множество чётных чисел есть идеал кольца $\mathbb{Z}$. Читатель легко проверит, что и вообще всякое множество целых чисел, кратных некоторому числу k, является идеалом кольца $\mathbb{Z}$. Эти примеры в случае коммутативного кольца K обобщаются следующим образом.

Пусть $b \in K$ — произвольный элемент такого кольца. Рассмотрим множество I всех элементов K, кратных b, то есть получающихся из b умножением на любое $a \in K$

$$I = \{ba \mid a \in K\}. \tag{9.6}$$

Для указанного множества все требования, входящие в определение идеала, очевидно, окажутся выполненными. Идеал вида (9.6) называется главным идеалом, точнее, главным идеалом, порождённым элементом b, и обозначается символом (b) или bK. В произвольном кольце вовсе не любой идеал является главным. Неглавные идеалы могут существовать даже в числовых кольцах (примеры читатель найдёт в дополнении).

Понятие главного идеала существенно потому, что оно является удобным средством описания отношения делимости элементов, имеющего смысл в любом коммутативном кольце.

В коммутативном кольце K элемент c называется *делителем* b, если существует такой элемент $d \in K$, что

$$b = cd. \tag{9.7}$$

Иначе говорят, что c делит b, обозначая это так: $c \mid b$. Если каждый из ненулевых элементов b и c делит другой, то такие элементы называются *ассоциированными*.

Ясно, что делители единицы (если она имеется в кольце), являются в точности обратимыми элементами кольца, а связь между ассоциированными элементами может быть выражена следующим простым утверждением.

Ненулевые элементы b и c области целостности с единицей ассоциированы тогда и только тогда, когда в равенстве (9.7) *элемент d обратим*[1].

Действительно, если элемент d обратим, то, умножая обе части равенства (9.7) на обратный d^{-1}, получим $c = bd^{-1}$. Значит, не только $c \mid b$, но и $b \mid c$, то есть элементы b и c ассоциированы.

Наоборот, если это так, то имеем

$$b = cd_1 \quad \text{и} \quad c = bd_2 \tag{9.8}$$

для некоторых элементов d_1 и d_2 кольца. Из равенств (9.8) следует, что

$$b = bd_2d_1, \quad \text{или} \quad b(d_2d_1 - 1) = 0.$$

[1] Полезно сравнить это утверждение с рассмотрениями главы 6, где говорилось о делимости в кольце целых алгебраических чисел, в частности, в кольце $\mathbb{Z}$ и в кольце целых гауссовых чисел $\mathbb{Z}[i]$.

Отсутствие делителей нуля позволяет сократить на b, поэтому

$$d_2 d_1 = 1.$$

Следовательно, элементы d_1 и d_2 в (9.8) обратимы.

Какое отношение ко всему сказанному о делимости имеют идеалы, точнее, главные идеалы? Ответ такой:

1. *Элемент c делит b тогда и только тогда, когда* $(c) \supseteq (b)$.
2. *Элементы b и c ассоциированы тогда и только тогда, когда* $(b) = (c)$.
3. *Элемент d кольца K обратим тогда и только тогда, когда* $(d) = (1) = K$.

Понятно, что утверждение 2 следует из утверждения 1. Кроме того, и утверждение 3 вытекает из утверждения 2. В самом деле, обратимость элемента d эквивалентна его ассоциированности с 1, а главный идеал (1), ею порождённый, очевидно, совпадает со всем кольцом K. Достаточно поэтому доказать лишь первое из трёх утверждений. Сделать это несложно. Если $(c) \supseteq (b)$, то ясно, что $b \in (c)$, но тогда согласно определению главного идеала этот элемент представляется в виде (9.7). Обратно, пусть $c \mid b$, то есть выполняется равенство (9.7). Это означает, что $b \in (c)$, и следовательно, всякое произведение ba, где a — любой элемент кольца, также содержится в идеале (c), поэтому $(b) \subseteq (c)$.

Отмеченные обстоятельства делают понятным интерес к так называемым *кольцам главных идеалов*. Этим термином обозначают целостные кольца с единицей, каждый идеал которых является главным. Давно уже было отмечено, что указанным свойством обладают такие классические объекты, как кольцо целых чисел $\mathbb{Z}$ и кольцо многочленов. Вот, например, как доказывается, что *кольцо $\mathbb{Z}$ является кольцом главных идеалов*.

Пусть I — произвольный идеал в $\mathbb{Z}$. Если I состоит только из нуля, то $I = (0)$. В противном случае выберем наименьшее положительное число a, принадлежащее идеалу I. Произвольный элемент b из I разделим на a с остатком

$$b = s \cdot a + r, \quad 0 \leqslant r < a.$$

Так как b и a принадлежат идеалу, это же верно и для $r = b - sa$. Поскольку $0 \leqslant r < a$, а элемент a — наименьшее положительное число в идеале I, то обязательно $r = 0$. Следовательно, $b = sa$, то есть все числа идеала I кратны a. Это и означает, что $I = (a)$ — главный идеал.

По той же схеме можно провести доказательство аналогичного утверждения и для кольца многочленов $\mathbb{F}[x]$ с коэффициентами из произвольного поля. Фактически же (см. дополнение 5) — для любого евклидова кольца. Понятие евклидового кольца, введённое прежде для числовых колец (см. дополнение 5 к главе 6), легко обобщается на случай произвольной области целостности. Достаточно лишь в данном там определении слова «числовое кольцо» заменить на «целостное кольцо».

Значение и место колец главных идеалов (в частности, и евклидовых колец) в обширном многообразии колец определяется прежде всего тем, что для них справедлива следующая основная теорема.

Всякое кольцо главных идеалов K является кольцом с однозначным разложением на простые множители.

Заметим, что понятия простого элемента и однозначного разложения, уже рассмотренные в главе 6, без изменений переносятся и на произвольные области целостности.

Доказательство основной теоремы, а им мы и закончим эту главу, даёт яркий пример одного из тех алгебраических рассуждений, с помощью которых получаются утверждения большой степени общности.

Убедимся сначала, что всякий ненулевой элемент либо сам является простым, либо обладает хотя бы одним разложением на простые элементы. Это справедливо не во всякой области целостности и потому нуждается в доказательстве. Будем рассуждать от противного: предположим, что имеется хотя бы один элемент, не обладающий этим свойством. Все такие элементы для краткости назовём «плохими» элементами, а порождённые ими идеалы — «плохими» идеалами. Рассмотрим теперь произвольную строго возрастающую цепочку «плохих» идеалов,

то есть такую, что

$$(a_1) \subset (a_2) \subset \ldots \subset (a_n) \subset \ldots, \tag{9.9}$$

и покажем, что всякая такая цепочка может иметь лишь конечное число членов. С этой целью рассмотрим объединение $A = \bigcup_k (a_k)$ всех идеалов, входящих в (9.9). Легко видеть, что A также образует идеал в K, причём главный, так как других в кольце K нет. Пусть a — его порождающий элемент, то есть $A = (a)$. Ясно, что элемент a содержится в некотором идеале (a_n) из цепочки (9.9). Тогда, с одной стороны, $A = (a) \subseteq (a_n)$, а с другой, согласно определению идеала A, верно и противоположное включение $A \supseteq (a_n)$. Следовательно, для некоторого n выполнено равенство $A = (a_n)$, а это и означает, что цепочка (9.9) обрывается на конечном шаге

$$(a_1) \subset (a_2) \subset \ldots \subset (a_n) = A. \tag{9.10}$$

Очевидно, что всякий «плохой» элемент (в частности, и a_n) не может быть простым. Поэтому для a_n существует нетривиальное разложение $a_n = bc$, в котором сомножители b и c не являются обратимыми. Но тогда $(b) \supset (a_n)$ и $(c) \supset (a_n)$, откуда следует, что ни b, ни c уже не могут быть «плохими» элементами. Иначе цепочка (9.10) могла бы быть продолжена за счёт одного из «плохих» идеалов (b) или (c). Итак, элементы b и c обладают разложениями на простые множители, произведение же этих разложений само будет являться разложением на простые множители для $a_n = bc$. Из полученного противоречия вытекает, что в кольце K нет «плохих» элементов.

Осталось теперь доказать однозначность разложения на простые множители. Для этого понадобится следующее свойство простых элементов кольца главных идеалов (для простых чисел в $\mathbb{Z}$ оно хорошо известно читателю).

Если простой элемент p кольца главных идеалов делит произведение ab, то он делит хотя бы один из сомножителей, то есть $p \mid a$ или $p \mid b$.

Действительно, пусть, например, p не делит a. Докажем, что в этом случае $p \mid b$. Рассмотрим наименьший идеал, содержащий

элементы p и a, обозначая его (p, a). Нетрудно понять, что он состоит из всех элементов вида

$$xp + ya, \tag{9.11}$$

где x и y — произвольные элементы кольца. С другой стороны, идеал (p, a) является главным, но тогда его порождающий элемент c, понятно, должен быть делителем всякого элемента из (p, a), в частности, элементов p и a. Из простоты p и из того, что p не делит a, сразу следует, что c обязан быть обратимым элементом – делителем единицы. Как известно, в этом случае

$$(p, a) = (c) = (1) = K,$$

поэтому существуют такие x_0 и y_0, для которых элемент вида (9.11) совпадает с единицей

$$x_0 p + y_0 a = 1.$$

Умножая это равенство на b, получаем

$$bx_0 p + y_0 ab = b.$$

Оба слагаемых в левой части делятся на p (второе потому, что по условию $p \mid ab$), значит, и b делится на p.

Переходим к заключительной части доказательства основной теоремы. Предположим, что некоторый элемент a имеет два разложения на простые множители:

$$a = p_1 p_2 \dots p_r = q_1 q_2 \dots q_s. \tag{9.12}$$

Так как p_1 делит произведение, стоящее справа, то по доказанному оно делит хотя бы один из входящих в него сомножителей. Изменив при необходимости их нумерацию, можем считать, что $p_1 \mid q_1$. Тогда $q_1 = u_1 p_1$, где u_1 обязан быть обратимым элементом, и равенство (9.12) запишется в виде

$$p_1 p_2 \dots p_r = p_1 u_1 q_2 \dots q_s.$$

Сокращая обе части равенства на общий множитель p_1 (в области целостности это допустимо), получаем

$$p_2 \dots p_r = u_1 q_2 \dots q_s. \tag{9.13}$$

Повторив рассуждения теперь уже применительно к равенству (9.13), аналогично будем иметь $q_2 = u_2 p_2$, где множитель u_2 снова обратим. В конечном итоге придём к выводу, что число сомножителей в обоих разложениях (9.12) одинаково (то есть $r = s$), и при этом после подходящей перенумерации каждый элемент q_k отличается от соответствующего элемента p_k на обратимый множитель: $q_k = u_k p_k$. Единственность разложения, а вместе с ней и утверждение нашей теоремы доказаны.

Дополнения и задачи

1. Во всяком кольце K подмножество $\{0\}$, состоящее из одного нуля, и само кольцо K, очевидно, являются идеалами. Это так называемые *тривиальные* идеалы. Докажите следующее утверждение: *коммутативное кольцо является полем тогда и только тогда, когда оно не содержит идеалов, отличных от тривиальных.*

2. Доказать, что в конечном кольце с единицей всякий элемент является или делителем нуля, или обратимым элементом. Отсюда, в частности, вытекает, что конечное коммутативное кольцо с единицей и без делителей нуля является полем.

3. *Характеристика поля.* Символом $n \circ 1$ обозначим сумму n единиц ($n \in \mathbb{N}$) произвольного поля $\mathbb{F}$:

$$n \circ 1 = 1 + 1 + \dots + 1 \quad (n \text{ единиц}).$$

Логически возможны два случая:

1) для различных $m, n \in \mathbb{N}$ элементы $m \circ 1$ и $n \circ 1$ также различны: $m \circ 1 \neq n \circ 1$;

2) существуют различные $m, n \in \mathbb{N}$ такие, что $m \circ 1 = n \circ 1$.

Понятно, что второй случай неизбежен во всяком конечном поле. Рассмотрим его подробнее. Пусть в равенстве $m \circ 1 = n \circ 1$ для определенности $m > n$. Обозначая $k = m - n$, будем иметь

$$k \circ 1 = 0, \tag{9.14}$$

где k — натуральное число. Наименьшее натуральное k с указанным свойством называют *характеристикой* поля $\mathbb{F}$. В первом случае говорят, что *характеристика поля равна нулю.*

Очевидно, что характеристика поля вычетов $\mathbb{Z}_p$ равна простому числу p. Оказывается, что *ненулевая характеристика поля всегда есть некоторое простое число* p. Докажите это утверждение, используя равенство $(n_1 n_2) \circ 1 = (n_1 \circ 1)(n_2 \circ 1)$ и тот факт, что в поле отсутствуют делители нуля.

4. *Простые поля.* Если одно поле $\mathbb{F}_1$ содержится в другом поле $\mathbb{F}$, то говорят, что $\mathbb{F}_1$ — подполе поля $\mathbb{F}$. Поле $\mathbb{F}$ называется простым полем, если оно не содержит подполей, отличных от самого $\mathbb{F}$. Примерами простых полей (проверьте это!) являются поля вычетов $\mathbb{Z}_p$ для всевозможных простых p и поле рациональных чисел $\mathbb{Q}$. Фактически, этими примерами исчерпываются все простые поля. Точнее, справедливо следующее утверждение:

> *всякое простое поле характеристики* 0 *изоморфно полю рациональных чисел* $\mathbb{Q}$, *всякое простое поле характеристики* p *изоморфно полю вычетов* $\mathbb{Z}_p$.

Пусть сначала $\mathbb{F}$ — простое поле нулевой характеристики. Вместе с 1 в нём должны содержаться элементы вида $n \circ 1$, где $n \in \mathbb{N}$, причём все они различны и не равны нулю. Для каждого элемента вида $n \circ 1$ в $\mathbb{F}$ содержится и противоположный элемент $-(n \circ 1)$. Обозначим его $(-n) \circ 1$. Следовательно, в записи $n \circ 1$ множитель n можно считать произвольным целым числом, при этом различным $n \neq 0$ отвечают различные ненулевые элементы в $\mathbb{F}$ (при $n = 0$ выражение $n \circ 1$ по определению считается равным нулю). Рассмотрим отображение f поля $\mathbb{Q}$ в поле $\mathbb{F}$, полагая

$$f(m/n) = (m \circ 1)(n \circ 1)^{-1}. \tag{9.15}$$

Здесь $m \in \mathbb{Z}$, $n \in \mathbb{N}$. Читателю рекомендуется проверить, что это отображение корректно определено, то есть при всех записях одного и того же рационального числа в виде дроби его образ будет один и тот же. Выражение в правой части также определено, поскольку ненулевой элемент $n \circ 1$ имеет в поле $\mathbb{F}$ обратный. Обозначим через $\mathbb{F}_1$ множество всех элементов вида (9.15).

При соответствии f различным дробям будут отвечать также различные элементы в $\mathbb{F}_1$, кроме того, сумма и произведение дробей будут переходить соответственно в сумму и произведение их образов в $\mathbb{F}_1$ (проверка предоставляется читателю). Сказанное означает, что $\mathbb{F}_1$ — подполе в $\mathbb{F}$, изоморфное полю $\mathbb{Q}$. Из простоты $\mathbb{F}$ следует, что $\mathbb{F} = \mathbb{F}_1$, и следовательно, оно само изоморфно полю $\mathbb{Q}$.

В случае поля ненулевой характеристики p ситуация иная. Для целых m и n равенство

$$m \circ 1 = n \circ 1 \tag{9.16}$$

возможно теперь тогда и только тогда, когда $m \equiv n \pmod{p}$. Действительно, равенство (9.16) эквивалентно

$$k \circ 1 = 0, \quad \text{где } k = m - n. \tag{9.17}$$

Пусть $k = sp + r$, где r — остаток при делении k на p ($0 \leqslant r < p$). Тогда

$$k \circ 1 = (sp + r) \circ 1 = sp \circ 1 + r \circ 1 = r \circ 1,$$

и равенство (9.17) равносильно $r \circ 1 = 0$. Так как $0 \leqslant r < p$, то из определения характеристики следует, что $r = 0$. Значит, $k = m - n$ делится на p без остатка и $m \equiv n \pmod{p}$. Обратно, если $m \equiv n \pmod{p}$, то $k = sp$ и $k \circ 1 = sp \circ 1 = 0$.

Следовательно, в поле характеристики p элементами

$$m \circ 1 = 0, \quad \text{где } m = 0, 1, \ldots, p - 1, \tag{9.18}$$

будут исчерпаны все элементы вида $k \circ 1$. Читатель может легко проверить, что все действия над элементами вида (9.18) аналогичны операциям над классами вычетов по модулю p. Поэтому

эти элементы образуют поле, изоморфное $\mathbb{Z}_p$, а изоморфизмом служит соответствие $f(\overline{m}) = m \circ 1$, $m = 0, 1, \ldots, p-1$. Из простоты $\mathbb{F}$ тогда следует, что оно само изоморфно $\mathbb{Z}_p$.

Чуть видоизменяя рассуждения, можно доказать (рекомендуем сделать это самостоятельно) более общий факт:

> *всякое поле характеристики 0 содержит единственное простое поле, изоморфное $\mathbb{Q}$, а всякое поле характеристики p содержит единственное простое поле, изоморфное $\mathbb{Z}_p$.*

5. Докажем утверждение, о котором говорилось в основном тексте главы: *всякое евклидово кольцо K является кольцом главных идеалов.*

Пусть I — идеал такого кольца. Из всех элементов этого идеала выберем тот ненулевой элемент a, норма которого наименьшая. Если b — произвольный элемент в I, то по определению евклидова кольца существуют такие q и r, что $b = qa + r$, и либо $r = 0$, либо $N(r) < N(a)$. Так как $a, b \in I$, то и элемент $r = b - qa \in I$. Тогда обязательно $r = 0$, ибо в противном случае мы входим в противоречие с выбором элемента a. Поэтому $b = qa$, и следовательно, идеал $I = (a)$ является главным.

Осталось показать, что K содержит единицу. Заметим для этого, что само K, в частности, образует главный идеал $K = (c)$, порождённый некоторым элементом c. Тогда существует $e \in K$ такой, что $c = ce$. Элемент e и служит единицей кольца. Действительно, для любого $d \in K$ имеем $d = fc$ для некоторого $f \in K$ (поскольку всякий элемент K кратен c). Поэтому $de = fce = fc = d$, что и требуется.

6. Убедитесь, что

а) кольцо эйлеровых чисел $\mathbb{Z}[\sqrt{-3}]$,

б) кольцо многочленов с целыми коэффициентами $\mathbb{Z}[x]$,

в) кольцо многочленов $\mathbb{F}[x, y]$ от двух переменных над полем $\mathbb{F}$

не являются кольцами главных идеалов.

ГЛАВА 10

ОПЕРИРУЕМ С КЛАССАМИ

Идея разбиения множества на классы и оперирования над полученными классами для нас уже не нова. Наглядной её демонстрацией послужило кольцо классов вычетов с определёнными в нём сложением и умножением классов. Теперь, когда границы наших познаний и нашего опыта в алгебре стали шире, мы можем взглянуть на пример с вычетами иными глазами. Пожалуй даже, мы можем предположить, что этот пример является проявлением некоего общего принципа. Скажем сразу, что такое предположение не лишено оснований — общий принцип действительно существует.

Для начала вспомним, что в разделе, посвящённом группам, встретилось понятие, очень похожее на классы вычетов. Мы имеем в виду, конечно, смежные классы группы по подгруппе, которые в частном случае аддитивной группы целых чисел $\mathbb{Z}$ в точности совпадали с классами вычетов[1]. Вопрос о том, можно ли со смежными классами произвольной группы оперировать так же, как и с классами вычетов, выглядит поэтому вполне естественным. Правда, в произвольной группе определена лишь одна операция (умножение или сложение), следовательно, и говорить в этом случае разумно также об одной операции над её смежными классами, как раз и связанной с заданным групповым законом умножения.

[1] Более точно, смежные классы группы $\mathbb{Z}$ по подгруппе $n\mathbb{Z}$ совпадают с классами вычетов по модулю n (элементами $\mathbb{Z}_n$).

И действительно, можно привести немало примеров, когда ответ на поставленный вопрос положителен. Вот один из них.

В аддитивной группе $\mathbb{R}$ действительных чисел рассмотрим подгруппу $\mathbb{Z}$, состоящую из целых чисел. Каждый смежный класс $r+\mathbb{Z}$ в этом случае есть множество точек числовой оси, обычно называемое решёткой, удалённых от точки r на расстояние, равное целому числу единиц (рис. 20).

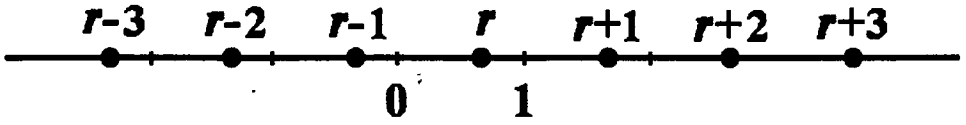


Рис. 20.

Сумму двух таких классов (решёток) $r_1+\mathbb{Z}$ и $r_2+\mathbb{Z}$ мы можем определить совершенно так же, как и сумму классов вычетов. Найдём с этой целью сумму двух произвольных чисел $r_1' \in r_1+\mathbb{Z}$ и $r_2' \in r_2+\mathbb{Z}$, принадлежащих указанным решёткам. Поскольку $r_1' = r_1 + k_1$, $r_2' = r_2 + k_2$ для некоторых целых чисел k_1 и k_2, то эта сумма $r_1'+r_2' = (r_1+r_2)+(k_1+k_2)$, отличается от r_1+r_2 снова на целое число. По этой причине, независимо от выбора r_1' и r_2' она всегда окажется в одном и том же классе $(r_1+r_2)+\mathbb{Z}$, его-то мы и будем считать суммой смежных классов $r_1+\mathbb{Z}$ и $r_2+\mathbb{Z}$. Таким образом,

$$(r_1+\mathbb{Z})+(r_2+\mathbb{Z}) = (r_1+r_2)+\mathbb{Z}.$$

На рис. 21 элементы решёток $r_1+\mathbb{Z}$, $r_2+\mathbb{Z}$ и их суммы $r_1+r_2+\mathbb{Z}$ отмечены точками, крестиками и кружками соответственно.

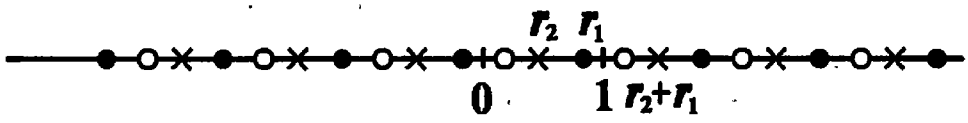


Рис. 21.

Легко видеть, что подобным образом можно поступить и в случае любой абелевой группы. Действительно, пусть H — подгруппа абелевой группы G и g_1H, g_2H — два её смежных класса по этой подгруппе. Если $g_1h_1 \in g_1H$, $g_2h_2 \in g_2H$ — некоторые их элементы, то их произведение (в силу коммутативности) равно

$$(g_1h_1)(g_2h_2) = (g_1g_2)(h_1h_2). \qquad (10.1)$$

Понятно, что элемент h_1h_2 принадлежит подгруппе H, но тогда равенство (10.1) означает, что, независимо от способа выбора элементов g_1h_1 и g_2h_2 в классах g_1H и g_2H их произведение всегда оказывается элементом смежного класса $(g_1g_2)H$. Последний и назовём произведением классов g_1H и g_2H, так что

$$(g_1H)(g_2H) = (g_1g_2)H. \tag{10.2}$$

Самым замечательным является то, что это новое умножение, определённое для классов, наследует все свойства групповой операции исходной группы.

Так, очевидно, что умножение смежных классов ассоциативно (это прямо следует из ассоциативности операции в G). Особую роль играет смежный класс $eH = H$, совпадающий с самой подгруппой. Из равенства (10.2) вытекает, что

$$(eH)(gH) = egH = gH$$

для любого смежного класса gH. Таким образом, класс H является нейтральным элементом (единицей) для нового умножения. Ну и наконец, каждый смежный класс gH обратим и обратным для него, как нетрудно понять, является класс $g^{-1}H$. Действительно, их произведение даёт нейтральный элемент eH.

Всё вместе сказанное означает, что множество смежных классов по подгруппе H само образует группу относительно умножения, определённого равенством (10.2). Построенную группу (она, подобно исходной, также абелева) называют *факторгруппой* группы G по подгруппе H, обозначая её обычно символом G/H.

В примере, рассмотренном выше, каждый элемент факторгруппы $\mathbb{R}/\mathbb{Z}$ представляет собой некоторую решётку точек на числовой оси, а операция в этой группе — это описанная там же операция сложения решёток. Заметим, что нейтральным элементом относительно этого сложения является решётка, составленная из точек с целыми координатами, так называемая целочисленная решётка.

Как же всё-таки обстоит дело в общей ситуации? Если попытаться провести прежние рассуждения для произвольной, уже

не обязательно абелевой группы, то мы сразу же натолкнемся на существенное препятствие. Ведь теперь равенство (10.1) может и не выполняться, поэтому не исключено, что произведение $(g_1h_1)(g_2h_2)$ окажется в смежном классе, отличном от g_1g_2H. Если так, то и попытка определить умножение смежных классов тщетна, поскольку результат такого умножения зависел бы от выбора представителей в этих классах. С другой стороны, требование коммутативности едва ли может показаться совершенно необходимым для успеха нашего предприятия. И действительно, для этого нам нужно нечто меньшее, а именно то, чтобы

$$(g_1h_1)(g_2h_2) \in g_1g_2H \tag{10.3}$$

при любом выборе элементов $g_1h_1 \in g_1H$ и $g_2h_2 \in g_2H$ (иначе, при любом выборе элементов h_1 и h_2 из подгруппы H). Включение (10.3) можно записать по-другому:

$$(g_1h_1)(g_2h_2) = g_1g_2h_3$$

для некоторого элемента $h_3 \in H$. Умножая обе части последнего равенства на g_1^{-1} слева и на h_2^{-1} справа, получим

$$h_1g_2 = g_2h, \quad \text{где } h = h_3h_2^{-1} \in H.$$

Элемент h_1g_2 находится в правом смежном классе Hg_2 но, как показывает равенство, он же принадлежит и левому смежному классу g_2H. Очевидно, что в наших рассуждениях g_2 может быть произвольным элементом группы G, а h_1 — произвольным элементом подгруппы H. Мы приходим, следовательно, к выводу: для корректности умножения классов необходимо, чтобы каждый правый смежный класс Hg целиком содержался в левом gH. Можно убедиться также, что необходимым является и обратное включение, следовательно,

$$gH = Hg \quad \text{для любого } g \in G.$$

Подгруппы, обладающие указанным свойством, играют в теории групп особую роль и называются нормальными подгруппами или

нормальными делителями[1]. Подчеркнём ещё раз, что нормальный делитель — это такая подгруппа, разложение в левые и правые смежные классы по которой совпадают. Очевидно, что в абелевой группе любая подгруппа является нормальной. Другие примеры нормальных подгрупп мы приведём несколько позже. Читатель может найти их также в дополнениях к этой главе. Свойство подгруппы быть нормальной является на самом деле не только необходимым, но и достаточным условием возможности оперирования над классами (всё равно, левыми или правыми) и построения факторгруппы по данной подгруппе. Чтобы понять это, надо, по существу, повторить с небольшими изменениями предыдущие рассуждения. Так, если g_1H и g_2H — смежные классы по нормальному делителю H, то при любом выборе представителей $g_1h_1 \in g_1H$ и $g_2h_2 \in g_2H$ будем иметь

$$(g_1h_1)(g_2h_2) = g_1(h_1g_2)h_2.$$

Элемент h_1g_2, принадлежащий правому смежному классу Hg_2 принадлежит также и левому g_2H, и следовательно, может быть записан в виде

$$h_1g_2 = g_2h_1', \quad h_1' \in H.$$

Но тогда

$$(g_1h_1)(g_2h_2) = g_1(g_2h_1')h_2 = g_1g_2h, \tag{10.4}$$

где $h = h_1'h_2 \in H$. Окончательно получаем, что независимо от выбора представителей их произведение содержится в классе $(g_1g_2)H$. Поэтому (как и прежде, в случае абелевой группы) равенство (10.4) можно и в этой более общей ситуации считать корректным определением умножения смежных классов. При этом снова множество смежных классов образует группу относительно определённой над ними операции, правда, уже не обязательно коммутативную. Тем самым определена факторгруппа G/H группы G по произвольному нормальному делителю H. Приведём пример.

[1] Впервые понятие нормальной подгруппы появилось в исследованиях Галуа об алгебраических уравнениях (см. главу 14).

В разложении симметричной группы S_n по знакопеременной подгруппе A_n, как уже отмечалось, имеется ровно два левых смежных класса. Один из них совпадает с самой подгруппой A_n и состоит, следовательно, из всевозможных чётных подстановок, другой имеет вид τA_n (τ — любая нечётная подстановка) и его элементами являются всевозможные нечётные подстановки. Но таковы же, как нетрудно понять, и правые смежные классы A_n и τA_n. Значит подгруппа A_n является нормальным делителем, а факторгруппа S_n/A_n состоит из двух элементов (классов) A_n и $\tau A_n = A_n\tau$.

Вот как выглядит таблица умножения в этой факторгруппе (табл. 5).

Таблица 5.

$\cdot$	A_n	τA_n
A_n	A_n	τA_n
τA_n	τA_n	A_n

Понятно, таблица лишь отражает то, что произведение двух чётных подстановок всегда снова чётная подстановка (отсюда и равенство $A_n \cdot A_n = A_n$), произведение чётной и нечётной подстановки — подстановка нечётная ($A_n \cdot \tau A_n = \tau A_n \cdot A_n = \tau A_n$), и наконец, подстановка, являющаяся результатом перемножения двух нечётных подстановок, чётная ($\tau A_n \cdot \tau A_n = A_n$).

В связи с понятием факторгруппы возникает вопрос. Какие имеются возможности оперирования над классами в других алгебраических структурах, например, в кольцах? К тому же его нам подсказывает пример кольца класса вычетов. Ответ в случае произвольного кольца приведёт нас опять к уже знакомому понятию идеала.

Вспомним, что в любом кольце мы имеем дело с двумя операциями — сложением и умножением. С ними мы хотим связать также две операции над его классами. Половину нашей задачи можно считать решённой. Ведь относительно сложения кольцо образует абелеву группу — аддитивную группу кольца, поэтому и классы следует строить по старому образцу. А именно, они

должны являться смежными классами по некоторой подгруппе J этой аддитивной группы[1]. Если a и b — произвольные элементы кольца K, то сумма соответствующих смежных классов $a+J$ и $b+J$ определяется по известному нам правилу

$$(a+J)+(b+J)=(a+b)+J. \tag{10.5}$$

Можно показать, что корректное определение умножения смежных классов возможно тогда и только тогда, когда указанная подгруппа является идеалом.

Операцию умножения смежных классов по идеалу определяем по обычной уже для нас схеме. Выбираем произвольные представители в смежных классах

$$a+J \quad \text{и} \quad b+J, \tag{10.6}$$

перемножаем их как элементы кольца K, а в качестве произведения классов берём тот смежный класс, в который попало произведение выбранных представителей. Корректность операции (независимость её от выбора представителей) проверяется просто. Представители классов (10.6), как бы мы их не выбирали, обязательно имеют вид

$$a+j_1 \quad \text{и} \quad b+j_2,$$

где j_1 и j_2 — некоторые элементы идеала J. Тогда их произведение равно

$$(a+j_1)(b+j_2)=ab+aj_2+j_1b+j_1j_2. \tag{10.7}$$

Из определения идеала непосредственно следует, что второе, третье и четвёртое слагаемое в (10.7), а значит, и их сумма принадлежит J. Но тогда при любом выборе представителей их произведение, как и требуется, непременно принадлежит одному и тому же классу $ab+J$, сама формула перемножения смежных классов приобретает вид

$$(a+J)(b+J)=ab+J. \tag{10.8}$$

[1] В силу коммутативности сложения такая подгруппа автоматически является нормальным делителем.

На множестве K/J смежных классов по идеалу J равенства (10.5) и (10.8) определяют, таким образом, операции сложения и умножения. Оставляем читателю проверку того, что выполнены все аксиомы кольца. В проверке, впрочем, очевидной, нуждаются лишь свойства ассоциативности и дистрибутивности. Полученное кольцо K/J по аналогии с прежней терминологией называют факторкольцом кольца K по идеалу J.

Видимо, не покажется удивительным, что кольцо классов вычетов $\mathbb{Z}_n$ легко вписывается в общее понятие факторкольца. В самом деле, рассмотрим в кольце целых чисел $\mathbb{Z}$ идеал $n\mathbb{Z}$ и факторкольцо $\mathbb{Z}/n\mathbb{Z}$ по этому идеалу. Элементами $\mathbb{Z}/n\mathbb{Z}$ являются смежные классы, которые записываются в виде $r + n\mathbb{Z}$ $(r = 0, 1, \ldots, n-1)$.

Понятно, что они в точности совпадают с классами вычетов по модулю n, а правила сложения и умножения смежных классов в этом случае превращаются в уже известные правила действия с классами вычетов. Следовательно, кольцо классов вычетов $\mathbb{Z}_n$ совпадает с факторкольцом $\mathbb{Z}/n\mathbb{Z}$.

Конструкция факторкольца позволяет по-новому взглянуть на другие факты. Используя её, можно, например, получить иное, чем прежде, описание поля комплексных чисел. Поступим так: в кольце многочленов $\mathbb{R}[x]$ с действительными коэффициентами рассмотрим многочлен x^2+1 и идеал $J = (x^2+1)$, порождённый этим многочленом. Конечно, выбор именно этого многочлена не случаен. Он, как мы знаем, является минимальным многочленом для мнимой единицы. Оказывается, факторкольцо $\mathbb{R}[x]/J$ по указанному идеалу, по существу, совпадает с полем комплексных чисел, точнее сказать, ему изоморфно. Чтобы убедиться в этом, заметим сначала, что в качестве представителя любого смежного класса $f(x) + J$ можно взять многочлен вида $a + bx$ степени не выше первой. В самом деле, при делении $f(x)$ на $x^2 + 1$ всегда будем иметь в остатке многочлен степени $\leqslant 1$, то есть

$$f(x) = s(x)(x^2 + 1) + a + bx. \tag{10.9}$$

Равенство (10.9) показывает, что $f(x)$ и остаток принадлежат одному и тому же смежному классу или, иначе, $f(x) + J = (a + bx) + J$.

Пусть теперь

$$(a_1 + b_1x) + J, \quad (a_2 + b_2x) + J \tag{10.10}$$

— произвольные смежные классы.

Очевидно тогда, что

$$(a_1+b_1x)+J+(a_2+b_2x)+J = ((a_1+a_2)+(b_1+b_2)x)+J. \tag{10.11}$$

Разберёмся с умножением классов (10.10), перемножая для этого их представители:

$$(a_1 + b_1x)(a_2 + b_2x) = a_1a_2 + (a_1b_2 + a_2b_1)x + b_1b_2x^2.$$

Вычитая и прибавляя в правой части число b_1b_2 получим

$$(a_1 + b_1x)(a_2 + b_2x) = (a_1a_2 - b_1b_2) + (a_1b_2 + a_2b_1)x + b_1b_2(x^2 + 1).$$

Последнее равенство показывает, что

$$((a_1+b_1x)+J)((a_2+b_2x)+J) = ((a_1a_2-b_1b_2)+(a_1b_2+a_2b_1)x)+J. \tag{10.12}$$

Из формул (10.11) и (10.12) видно, что правила сложения и умножения смежных классов (элементов кольца $\mathbb{R}[x]/J$) полностью идентичны правилам действий над комплексными числами. Но тогда соответствие $(a + bx) + J \to a + bi$ и будет требуемым изоморфизмом факторкольца $\mathbb{R}[x]/J$ и поля $\mathbb{C}$ комплексных чисел.

Остановимся ещё на одном важном примере, во многом аналогичном двум предыдущим. Будем исходить из кольца $\mathbb{F}[x]$ многочленов над произвольным полем $\mathbb{F}$. Читатель на первых порах может подразумевать под $\mathbb{F}$ любое числовое поле. Мы уже отмечали, что идеалы в кольце $\mathbb{F}[x]$ исчерпываются множествами вида $g(x)\,\mathbb{F}[x]$, состоящими из всех многочленов, кратных некоторому фиксированному многочлену $g(x)$. Пусть $I = g(x)\,\mathbb{F}[x]$ — такой идеал. Попытаемся описать строение факторкольца $\mathbb{F}[x]/I$, используя для этой цели известные нам сведения о делимости многочленов. По аналогии с предыдущим примером покажем, что, если m — степень $g(x)$, то в качестве представителя в каждом смежном классе может быть выбран (причём единственным

образом) многочлен, степень которого не превосходит $m-1$. Читатель, видимо, уже понял, как следует для этого поступить. Если $f(x)+I$ — произвольный смежный класс, то рассмотрим остаток $r(x)$, получающийся при делении представителя $f(x)$ на многочлен $g(x)$, порождающий идеал

$$f(x)=s(x)g(x)+r(x).$$

Это равенство, как и 10.9, показывает, что многочлен $f(x)$ и остаток $r(x)$ принадлежат одному смежному классу

$$f(x)+I=r(x)+I, \tag{10.13}$$

причём степень $r(x)$, как это и требуется, не превосходит $m-1$. Найдётся ли в смежном классе (10.13) другой многочлен $r'(x)$, степень которого также $\leqslant m-1$? Предположив, что ответ утвердительный, мы бы сразу пришли к противоречию — ведь тогда разность $r'(x)-r(x)$, принадлежащая I, должна была бы делиться без остатка на порождающий многочлен $g(x)$. Очевидно, такое невозможно, так как степень разности меньше степени $g(x)$.

Итак, запись вида $r(x)+I$, в которой степень $r(x)\leqslant m-1$, для всякого смежного класса является единственной. Назовём её канонической, и для краткости обозначим $r(x)+I=\overline{r(x)}$.

Остаётся понять, что происходит с каноническими записями при сложении и умножении элементов (смежных классов) факторкольца $F[x]/I$. Если $\overline{r_1(x)}=r_1(x)+I$ и $\overline{r_2(x)}=r_2(x)+I$ — канонические записи двух смежных классов, то, чтобы получить запись, также каноническую, для суммы и произведения, требуется в классах

$$(r_1(x)+r_2(x))+I \quad \text{и} \quad r_1(x)\,r_2(x)+I$$

выбрать по представителю степени $\leqslant m-1$. Ясно, что в первом случае уже сам многочлен $r_1(x)+r_2(x)$ обладает этим свойством, значит

$$\overline{r_1(x)}+\overline{r_2(x)}=\overline{r_1(x)+r_2(x)}. \tag{10.14}$$

С другой стороны, степень произведения $r_1(x)r_2(x)$ может оказаться большей $m-1$, поэтому в качестве представителя канонической записи нужно взять, как ясно из предыдущего, остаток

$r(x)$ при делении $r_1(x)r_2(x)$ на $g(x)$. То есть, если $r_1(x)r_2(x) = s(x)g(x) + r(x)$, и степень $r(x) \leqslant m-1$, то

$$\overline{r_1(x)} \cdot \overline{r_2(x)} = \overline{r(x)}. \tag{10.15}$$

Описанное здесь умножение называется *умножением по модулю многочлена* $g(x)$, а факторкольцо K/I с правилами операций (10.14) и (10.15) *кольцом вычетов по модулю многочлена* $g(x)$.

Кольцо $\mathbb{R}[x]/(x^2+1)$, изоморфное, как мы установили выше, полю $\mathbb{C}$, является, понятно, одним из примеров применения только что описанной конструкции. Рассмотрим другой пример, в котором в качестве исходного берется уже нечисловое поле. Он важен тем, что указывает способ получения конечных полей, отличных от известных нам полей вычетов.

Пример. Пусть $\mathbb{Z}_2[x]$ — кольцо многочленов над полем $\mathbb{Z}_2$ (с коэффициентами 0 и 1). Выясним, как устроено кольцо вычетов по модулю многочлена x^2+x+1, или, иначе $\mathbb{Z}_2[x]/(x^2+x+1)$. Все его элементы, как мы знаем, находятся во взаимно однозначном соответствии с многочленами, степень которых меньше двух. Таких многочленов ровно четыре: 0, 1, x, $x+1$. Столько же элементов и в рассматриваемом кольце вычетов: $\overline{0}$, $\overline{1}$, $\overline{x}$, $\overline{x+1}$. Таблицы их сложения и умножения по модулю x^2+x+1 (табл. 6) получаются в соответствии с правилами (10.14) и (10.15) (черта в записи классов опущена).

Например, $x(x+1) = x^2+x = (x^2+x+1)+1$, поэтому $\overline{x} \cdot \overline{x+1} = \overline{1}$. Из таблицы умножения легко заключить, что каждый ненулевой элемент нашего кольца обратим, поэтому оно образует поле, состоящее из четырёх элементов. Всегда ли, действуя подобным образом, мы будем получать обязательно поле? Ответ на этот вопрос отрицательный. И чтобы убедиться в этом, достаточно рассмотреть факторкольцо $\mathbb{Z}_2[x]/(x^2+1)$. В нём снова 4 элемента, и записываются они так же, как и элементы $\mathbb{Z}_2[x]/(x^2+x+1)$, однако умножение здесь, конечно, иное. Таблица 7 показывает, что в этом случае кольцо вычетов содержит делители нуля и полем не является.

Ситуация полностью разъясняется следующей теоремой.

Таблица 6.

$+$	0	1	x	$x+1$
0	0	1	x	$x+1$
1	1	0	$x+1$	x
x	x	$x+1$	0	1
$x+1$	$x+1$	x	1	0

$\times$	0	1	x	$x+1$
0	0	0	0	0
1	0	1	x	$x+1$
x	0	x	$x+1$	1
$x+1$	0	$x+1$	1	x

Таблица 7.

$\times$	0	1	x	$x+1$
0	0	0	0	0
1	0	1	x	$x+1$
x	0	x	1	$x+1$
$x+1$	0	$x+1$	$x+1$	0

Кольцо вычетов $\mathbb{F}[x]/(g(x))$ *по модулю многочлена* $g(x)$ *является полем тогда и только тогда, когда* $g(x)$ *неприводим над полем* $\mathbb{F}$.

Понятно, что теорема подтверждается нашими примерами. Так, многочлен x^2+x+1 неприводим над $\mathbb{Z}_2$ (почему?). Уже не будет таковым над $\mathbb{Z}_2$ многочлен $x^2+1=(x+1)(x+1)$. Именно поэтому кольца вычетов $\mathbb{R}[x]/(x^2+1)$ и $\mathbb{Z}_2[x]/(x^2+x+1)$ являются полями, а $\mathbb{Z}_2[x]/(x^2+1)$ — нет.

Нетрудно заметить аналогию сформулированной теоремы с подобным утверждением для колец вычетов по модулю n (глава 9). Эта аналогия распространяется и на метод доказательства. Предлагаем читателю самому убедиться в этом.

Что касается конечных полей (иначе, полей Галуа), то картина здесь вкратце такова.

1. *Число элементов всякого конечного поля есть степень p^n некоторого простого числа p. Какова бы ни была эта степень p^n, существует поле, содержащее p^n элементов.*

2. *Любые два поля Галуа, содержащие одинаковое число элементов, изоморфны между собой.*

Первый пункт можно конкретизировать.

3. *Всякое поле Галуа с p^n элементами изоморфно полю вычетов $\mathbb{Z}_p[x]/(m(x))$ по модулю многочлена $m(x)$ степени n, неприводимого над полем $\mathbb{Z}_p$ (можно показать, что для любого n такой неприводимый над $\mathbb{Z}_p$ многочлен всегда существует).*

Поле Галуа с p^n элементами, единственное с точностью до изоморфизма, обозначают обычно[1] $\mathrm{GF}(p^n)$. Перечисленные факты дополняет следующий красивый результат.

4. *Мультипликативная группа всякого поля $\mathrm{GF}(p^n)$ (т. е. группа по умножению его ненулевых элементов) является циклической группой порядка $p^n - 1$.*

Образующий этой циклической группы называют *примитивным элементом* поля Галуа.

Доказательства сформулированных здесь теорем мы вынуждены опустить, почти все они не простые. Ограничимся ещё одним примером, который иллюстрирует утверждения 3 и 4. К этому же этот пример нам будет полезен в главе, посвященной кодированию.

Пример. Поле $\mathrm{GF}(2^4)$. Для построения этого поля, как следует из 3, можно использовать неприводимый над $\mathbb{Z}_2$ многочлен степени 4. Покажем, что таким многочленом является, например, $x^4 + x + 1$. Если бы он разлагался на множители меньшей степени, то среди них обязательно нашлись бы многочлен первой или второй степени. Простой перебор показывает[2], что это невозможно. Таким образом, $\mathbb{Z}_2[x]/(x^4 + x + 1)$ — искомое поле

[1] G и F — начальные буквы слов «Галуа» (Galois) и «поле» (field).

[2] Достаточно проверить, что $x + 1$ и $x^2 + x + 1$ не являются делителями $x^4 + x + 1$.

из 16 элементов. Удобно ввести новый символ α, и вместо $\overline{r(x)} = r(x) + I$, где $I = (x^4 + x + 1)$ писать $r(\alpha)$. Это фактически означает, что через α обозначен смежный класс $x + I$. Перечислив все многочлены от α степени $\leqslant 3$, мы получим список всех элементов поля

$$0;\ 1;\ 2;\ \alpha+1;\ \alpha^2+\alpha;\ \alpha^2+1;\ \alpha^3;\ \alpha^3+\alpha^2;\ \alpha^3+\alpha;\ \alpha^3+1;$$
$$\alpha^3+\alpha^2+\alpha;\ \alpha^3+\alpha^2+1;\ \alpha^3+\alpha+1;\ \alpha^3+\alpha^2+\alpha+1.$$

Полная таблица умножения в таком поле состояла бы из 256 клеток. Понятно, что мы её не приводим. Однако достаточно научиться умножать одночлены 1, α, α^2, α^3 — ведь всякий другой многочлен есть их сумма.

Таблица 8.

$\times$	α	α^2	α^3
α	α^2	α^3	$\alpha+1$
α^2	α^3	$\alpha+1$	$\alpha^2+\alpha$
α^3	$\alpha+1$	$\alpha^2+\alpha$	$\alpha^3+\alpha^2$

Таблица 8 и служит этой цели (единица из таблицы исключена, поскольку умножение на неё ничего не меняет). Например, для отыскания $\alpha^3 \cdot \alpha^2$ нужно найти остаток от деления x^5 на $x^4 + x + 1$. Имеем

$$x^5 = x(x^4 + x + 1) + x^2 + x,$$

поэтому $\alpha^3 \cdot \alpha^2 = \alpha^2 + \alpha$. С учетом таблицы 8 мы легко можем перемножить любые два многочлена от α. Например,

$$(\alpha^2+\alpha)(\alpha^3+\alpha^2+1) = \alpha^5+\alpha^3+\alpha^2+\alpha^4+\alpha^2+\alpha =$$
$$= (\alpha^2+\alpha)+\alpha^3+(\alpha+1)+\alpha = \alpha^3+\alpha^2+\alpha+1.$$

Заметим, наконец, что элемент α служит примитивным элементом нашего поля. Действительно, таблица 8 показывает, что его степени α^i $(i = 0, 1, \ldots, 14)$ исчерпывают все ненулевые элементы поля $\mathrm{GF}(2^4)$.

Дополнения и задачи.

1. Во многих случаях удобен следующий признак того, что подгруппа нормальна.

Подгруппа $H \subseteq G$ является нормальной тогда и только тогда, когда для любых $g \in G$ и $h \in H$ элемент $g^{-1}hg$, называемый сопряжённым к h, также принадлежит H.

Докажите это.

2. Убедиться, что в группе движений плоскости множество всех сдвигов образует нормальный делитель. Покажите также, что если τ — сдвиг, а v — поворот на угол α, то сопряжённый элемент $v^{-1}\tau v$ есть сдвиг, направление которого составляет угол α с направлением τ. Аналогично, если σ — зеркальная симметрия относительно некоторой оси, то $\sigma^{-1}\tau\sigma$ — сдвиг, направление которого симметрично направлению τ относительно этой оси. Для доказательства удобно использовать комплексные числа (см. дополнение к главе 2).

3. Доказать, что в знакопеременной группе A_4 имеется единственный нетривиальный нормальный делитель. Он имеет порядок 4 и порождён двумя коммутирующими подстановками (12)(34) и (13)(24).

4. Проверьте, что в любой группе G подгруппа H индекса 2 является нормальным делителем и факторгруппа G/H — циклическая группа порядка 2.

5. Доказать, что если H_1 и H_2 — нормальные делители группы G, имеющие тривиальное пересечение ($H_1 \cap H_2 = \{e\}$), то всякий элемент из H_1 коммутирует с любым элементом из H_2. Указание: рассмотрите элемент ${h_1}^{-1}{h_2}^{-1}h_1h_2$, где $h_1 \in H_1, h_2 \in H_2$ и покажите, что он равен e.

6. В теории групп имеется целый ряд конструкций, с помощью которых можно исходя из данных групп строить новые группы. Самыми употребительными из них являются так называемые прямые и полупрямые произведения. Пусть подгруппы H_1 и H_2 такие же, как в предыдущем пункте. Через $H_1 \times H_2$ обозначим множество всех элементов вида h_1h_2, где $h_1 \in H_1$, $h_2 \in H_2$. Покажите, что $H_1 \times H_2$ — подгруппа (даже нормальный делитель)

группы G. В том случае, если группа $G = H_1 \times H_2$, она называется прямым произведением подгрупп H_1 и H_2, а подгруппы — прямыми сомножителями. Таким образом, всякий элемент g прямого произведения записывается в виде

$$g = h_1 h_2, \quad h_1 \in H_1,\ h_2 \in H_2.$$

Убедитесь, что для всякого g такая запись единственна. Покажите что строение группы G с точностью до изоморфизма определяется строением её прямых сомножителей.

7. Докажите, что $\mathbb{Q}^* = \langle -1 \rangle \times \mathbb{Q}^+$, где $\mathbb{Q}^+$ — мультипликативная группа положительных рациональных чисел, $\langle -1 \rangle = \{1, -1\}$ — циклическая группа, порождённая числом -1.

8. Сохраним все предположения относительно H_1, $H_2 \subseteq G$ пункта 5, отказавшись, быть может, лишь от нормальности одной из подгрупп, например, H_1. В этом случае множество

$$\{\, h_1 h_2 \mid h_1 \in H_1,\ h_2 \in H_2 \,\}$$

образует подгруппу в группе G. Если это множество исчерпывает всю группу G, то последняя называется *полупрямым произведением* H_1 и H_2 и обозначается $G = H_1 \leftthreetimes H_2$ (верхний «хвостик» у символа $\leftthreetimes$ обращён в сторону нормального делителя). В случае полупрямого произведения элементы h_1 и h_2 уже не обязаны коммутировать и наряду с записью $g = h_1 h_2$ возможна и иная: $g = h_2' h_1$, где $h_2' \in H_2$. Для полного описания полупрямого произведения уже не достаточно знать устройство сомножителей H_1 и H_2. Нужна ещё дополнительная информация иного сорта (о характере её можно судить по дополнениям к следующей главе). Проверьте, что $S_n = \langle \tau \rangle \leftthreetimes A_n$, где τ — любая из транспозиций.

Ещё один пример полупрямого произведения мы получим, если в качестве группы G рассмотрим совокупность всех движений плоскости первого рода (переносов и поворотов). Как уже говорилось, множество T всех переносов образует в G нормальный делитель. Пусть далее U множество всех поворотов вокруг некоторого фиксированного центра. Докажите, что $G = U \leftthreetimes T$.

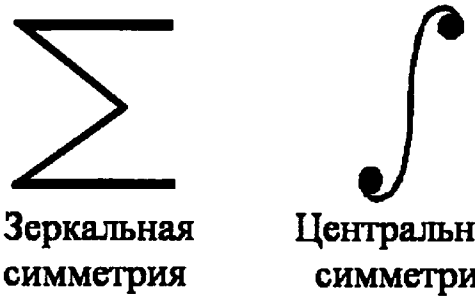
9. Пусть $G = H_1 \leftthreetimes H_2$. Докажите, что $G/H_2 \cong H_1$.

ГЛАВА 11

ЯЗЫК СИММЕТРИИ

Проявления симметрии в окружающем нас мире поистине необъятны. Мы можем наблюдать её в строении тела человека и животных, в архитектурных формах, в разнообразнейших орнаментах и узорах. Симметричные правильные геометрические фигуры, распустившийся цветок, кристалл поваренной соли, снежинка... Распознать симметричное может всякий, не так легко объяснить, однако, что такое симметрия. Ещё сложнее понять, что основные представления, касающиеся симметрии, имеют тесную связь с алгеброй. Не исключено, что даже само упоминание о такой связи иному читателю кажется неожиданным.

Буквальный перевод с греческого — соразмерность — позволяет толковать слово «симметрия» весьма широко. Так симметрию иногда понимают как гармонию пропорций, и в этом смысле можно говорить о симметрии также и в живописи, музыке, поэзии. Правда, такой слишком широкий взгляд делает понятие симметрии довольно расплывчатым, и почти не поддающимся изучению (по крайней мере, точными науками). Более узкое, но зато и более конкретное, осязаемое представление о симметрии даёт, например, такое определение (нечто подобное можно найти в энциклопедических справочниках): симметрия есть правильное, пропорциональное расположение частей в одном предмете. Но и такое объяснение оставляет вопросы — действительно, можно лишь догадываться, что означают в нём слова «правильное», «пропорциональное». Вскоре мы увидим, что средствами алгебры можно придать этим словам вполне отчётливый смысл, но



Зеркальная симметрия

Центральная симметрия

Рис. 22.

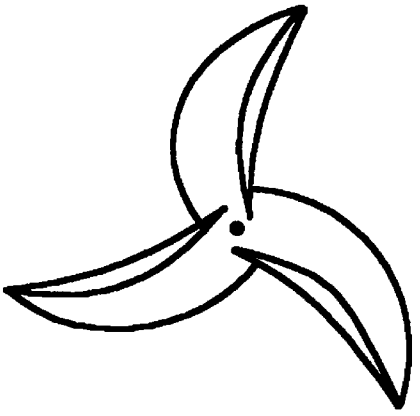


Рис. 23.

сначала вспомним о вещах, хорошо всем известных.

Из школьного курса геометрии читателю знакомы понятия зеркальной или осевой симметрий. Этими видами симметрии обладают, например, фигуры (это знаки суммы и интеграла), представленные на рис. 22.

На другом рисунке (рис. 23) изображена детская вертушка, которая не обладает ни осевой, ни центральной симметрией; однако этой фигуре присущ другой её вид, так называемая поворотная симметрия. Вертушка совмещается сама с собой при повороте на 120° и 240°. Поворотной симметрией обладают также правильные многоугольники. Каждый такой многоугольник самосовмещается (переходит в себя) при поворотах вокруг своего центра на углы, кратные $2\pi/n$ где n — число сторон многоугольника. В отличие от вертушки, правильный n-угольник имеет и оси симметрии — их столько же, сколько и сторон. Пожалуй, «самой симметричной» из плоских фигур следует признать круг — он переходит в себя при повороте на любой угол, и всякая прямая, проходящая через его центр является его осью симметрии. Мы пришли к тому, что «количество симметрии» (позволим себе такое выражение), которым обладает фигура, естественно измерять числом преобразований плоскости, самосовмещающих эту фигуру. При этом рассматривается, конечно, только такие преобразования, которые не нарушают целостность фигуры, т. е. сохраняют неизменными расстояния между любыми её точками. Мы и будем поэтому говорить здесь лишь о тех преобразованиях плоскости, которые обладают этим свойством по отношению к любым двум

её точкам. Такие преобразования были названы нами движениями (см. дополнения к главе 2). Множество G всех движений, самосовмещающих фигуру, может служить характеристикой не только «количества», но и типа симметрии этой фигуры. Чтобы это понять, мы должны обратиться к хорошо нам знакомой операции композиции преобразований. Здесь мы вновь выходим на алгебраическую тропу. Дело в том, что множество G самосовмещений любой фигуры образует группу — строение этой группы как раз и определяет тип симметрии фигуры. Проверка групповых аксиом не вызовет трудностей (предоставляем её читателю). Группа G движений, самосовмещающих данную фигуру, называется группой симметрии этой фигуры. Так, группа симметрии вертушки (рис. 23) состоит из трёх элементов: тождественного преобразования (поворота на 0°) и поворотов на 120° и 240°.

В главе 2 было фактически доказано, что всякое движение плоскости есть либо поворот вокруг некоторого центра, либо зеркальное отражение, либо сдвиг, либо скользящее отражение (комбинация сдвига и отражения). Значит и группа симметрии любой плоской фигуры может содержать лишь эти четыре вида преобразований. Если же фигура вдобавок имеет ограниченные размеры, то остаются лишь первые два: поворот и зеркальная симметрия. Причина понятна. Ведь, если, например, сдвиг f содержится в группе симметрии, то и преобразования f^n (итерации этого сдвига) также должны в ней содержаться. Однако, если f сдвигает все точки фигуры на расстояние a, то преобразование f^n — уже на расстояние na (рис. 24).

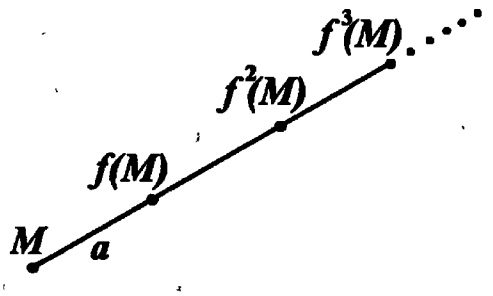


Рис. 24.

Получается, что при многократном повторении преобразования f мы можем сдвинуть фигуру на сколь угодно большое расстояние, оставляя её при этом самосовмещённой саму с собой. Очевидно, что это невозможно, если размеры фигуры ограничены. По той же причине в группе симметрии такой фигуры не могут содержаться и скользящие симметрии.

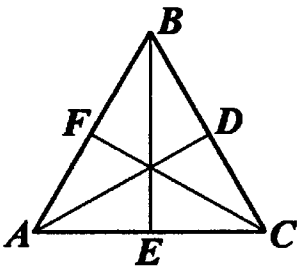


Рис. 25.

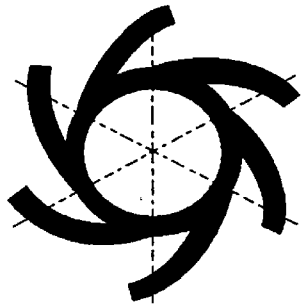


Рис. 26.

Сказанное позволяет легко находить группы симметрий плоских фигур. Возьмём в качестве простого, но типичного примера правильный треугольник (рис. 25).

Из всевозможных поворотов плоскости имеется (как и у вертушки) только три, совмещающих треугольник с самим собой, — это повороты на углы 0° (тождественное преобразование), 120°, 240° вокруг центра треугольника. Обозначим их соответственно v_0, v_1, v_2. Точно так же, лишь три зеркальных отражения относительно осей симметрии треугольника AD, BE, CF приводят его к самосовмещению, — обозначим их σ_1, σ_2, σ_3. Таким образом, группа симметрии правильного треугольника (обозначим её D_3) содержит 6 преобразований, иначе, её порядок равен 6. Ясно, что совершенно иной тип симметрии у фигуры, изображённой на рис. 26, однако и её группа симметрии также состоит из шести элементов — вращений на углы 0°, 60°, 120°, 180°, 240°, 300°.

Не следует ли отсюда, что группа симметрии не всегда «улавливает» тип симметрии данной фигуры? К счастью, это не так. Действительно, хотя порядки обеих групп одинаковы, но группы эти не изоморфны. Группа симметрии второй фигуры циклическая, и вращение на угол 60° является её образующим. Не такова группа правильного треугольника: каждое из его нетривиальных вращений v_1, v_2, очевидно, имеет порядок 3, что же касается элементов $\sigma_1, \sigma_2, \sigma_3$, то их порядок равен 2 (дважды выполненное зеркальное отражение приводит к тождественному преобразованию). Следовательно, ни один из перечисленных элементов не может служить образующим группы D_3.

Группа D_3 несёт в себе черты, присущие группам симметрии плоских фигур, поэтому любопытно остановиться на некоторых

подробностях её строения. Можно убедиться непосредственно (или используя одно из утверждений дополнения 10 к главе 2), что

$$\sigma_1 v_1 = \sigma_2, \quad \sigma_1 v_2 = \sigma_3, \tag{11.1}$$

т. е. последовательное выполнение сначала поворота, затем зеркального отражения приводит также к зеркальному отражению, но относительно оси, повёрнутой на соответствующий угол.

Произведение же двух зеркальных симметрий есть уже поворот, например,

$$\sigma_1 \sigma_2 = v_1, \quad \sigma_2 \sigma_1 = v_2, \tag{11.2}$$

Пользуясь указанными соотношениями и им аналогичными, читатель легко составит умножения для группы D_3 (заметим, что она не абелева). Однако возможно и более компактное её описание. В самом деле, из сказанного выше следует, что все 6 элементов группы симметрии могут быть записаны в виде произведений

$$\sigma_1^{\varepsilon} v_1^{\alpha}, \tag{11.3}$$

где $\varepsilon = 0, 1; \quad \alpha = 0, 1, 2$. Так, при $\varepsilon = 0$ получаем

$$v_1^0 = v_0 = e, \quad v_1, \quad v_1^2 = v_2, \tag{11.4}$$

а при $\varepsilon = 1$

$$\sigma_1 v_0, \quad \sigma_1 v_1 = \sigma_2, \quad \sigma_1 v_1^2 = \sigma_1 v_2 = \sigma_3. \tag{11.5}$$

На языке теории групп отмеченное обстоятельство выражают словами: группа D_3 порождена двумя элементами σ_1 и v_1, а сами эти элементы являются её образующими[1].

Изучая симметрию, мы заодно познакомились с таким способом задания группы (он возможен для многих более сложно, чем D_3, устроенных групп), при котором для оперирования над её элементами вовсе нет необходимости прибегать к достаточно

[1] Таким образом, в отличие от циклической группы у D_3 не один, а два образующих.

громоздкой таблице Кэли. Обойтись без неё в нашем конкретном примере мы сумеем, если заметим, что между образующими имеются простые соотношения

$$v_1\sigma_1 = \sigma_1 v_1^2 = \sigma_1 v_1^{-1}, \quad v_1^2\sigma_1 = \sigma_1 v_1. \tag{11.6}$$

Читатель без труда сумеет проверить эти равенства и может также убедиться (это чуть сложнее), что каждое из них является следствием другого. С учётом соотношений (11.6) и равенств $\sigma_i^2 = v_i^3 = e$, умножение любых двух элементов группы D_3 производится без затруднений. Например:

$$\sigma_2\sigma_3 = (\sigma_1 v_1)(\sigma_1 v_1^2) = \sigma_1(v_1\sigma_1)v_1^2 = \sigma_1(\sigma_1 v_1^2)v_1^2 = \sigma_1^2 v_1^4 = v_1, \tag{11.7}$$

$$v_2\sigma_2 = v_1^2(\sigma_1 v_1) = (v_1^2\sigma_1)v_1 = (\sigma_1 v_1)v_1 = \sigma_1 v_1^2 = \sigma_3. \tag{11.8}$$

Очень много общего с группой D_3 имеет группа симметрии D_n правильного многоугольника с произвольным числом сторон n. Эта группа (она называется группой диэдра степени n) содержит ровно n поворотов $v_0, v_1, \ldots$ $\ldots, v_k, \ldots, v_{n-1}$ на углы

$$0, \frac{2\pi}{n}, \ldots, \frac{2\pi k}{n}, \ldots, \frac{2\pi(n-1)}{n},$$

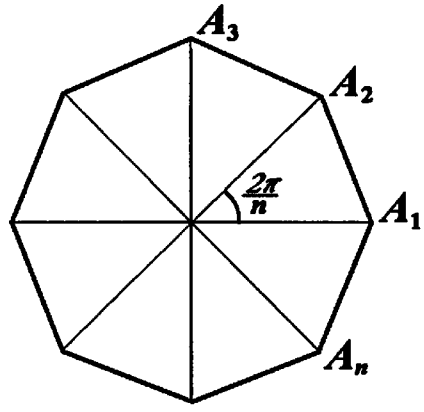


Рис. 27.

и столько же зеркальных симметрий $\sigma_1, \sigma_2, \ldots, \sigma_n$ относительно осей, проходящих соответственно через вершины $A_1, A_2, \ldots, A_n$ (рис. 27).

Таким образом, порядок группы диэдра равен $2n$. Как и прежде (сравните с равенствами (11.3)), все зеркальные отражения получаются как комбинации σ_1 и соответствующих поворотов:

$$\begin{aligned} \sigma_2 &= \sigma_1 v_1, \\ \sigma_3 &= \sigma_1 v_2 = \sigma_1 v_1^2, \\ &\ldots\ldots\ldots\ldots \\ \sigma_n &= \sigma_1 v_{n-1} = \sigma_1 v_1^{n-1}. \end{aligned} \tag{11.9}$$

Поэтому, как и выше, каждое преобразование из D_n записывается в виде[1]

$$\sigma_1^{\varepsilon} v_1^{\alpha}, \quad \text{где } \varepsilon = 0, 1;\ \alpha = 0, 1, 2, \ldots, n-1. \tag{11.10}$$

Нелишне отметить, что все повороты образуют подгруппу V_n индекса 2 в группе диэдра, а следовательно (см. дополнение к главе 10), и нормальный делитель симметрии же, как показывают равенства (11.9), — смежный класс по этой подгруппе.

Итак группа D_n для любого n имеет, как и для $n = 3$, два образующих σ_1 и v_1, которые по-прежнему связаны соотношением

$$v_1\sigma_1 = \sigma_1 v_1^{-1} = \sigma_1 v_1^{n-1} \tag{11.11}$$

и следствиями из него

$$v_1^{\alpha}\sigma_1 = \sigma_1 v_1^{-\alpha} = \sigma_1 v_1^{n-\alpha}, \quad \alpha = 1, 2, \ldots, n-1. \tag{11.12}$$

Равенства

$$\sigma_1^2 = e, \quad v_1^n = e. \tag{11.13}$$

вместе с (11.12) образуют, как можно показать, так называемую систему определяющих соотношений группы D_n, т. е. таких, из которых следуют все прочие соотношения, выполненные в группе. Используя (11.12) и (11.13), можно производить вычисления группе D_n при любом n столь же просто, как и в случае $n = 3$.

Тот факт, что D_n имеет образующие σ_1 и v_1, обозначим по аналогии с циклической группой записью $D_n = \langle \sigma_1, v_1 \rangle$. Подобная запись используется для указания образующих произвольных групп.

Интересно, что группами диэдра D_n и их подгруппами V_n, состоящими из поворотов, исчерпываются любые мыслимые группы симметрии плоских фигур ограниченных размеров[2] (доказательство см. в дополнении). Это на первый взгляд может показаться странным, потому что даже на глазок многообразие

[1] Нетрудно убедиться, что запись указанного вида для каждого элемента группы D_n единственна.

[2] Тут нужна оговорка — речь идёт о фигурах, группа симметрии которых конечна.

таких симметричных фигур на плоскости гораздо «обширнее», чем список упомянутых групп D_n и V_n. Всё же ошибки здесь нет по той простой причине, что многие из внешне различных фигур (вспомните снежинки и шестиугольники) обладают одинаковым типом симметрии. Алгебраическим эквивалентом этого типа и является как раз строение соответствующей группы D_n или V_n. Добавим к этому, что любую ограниченную симметричную фигуру (узор) с конечной группой симметрии можно «сконструировать» подобно тому, как это делается в калейдоскопе, с помощью групп вращений или диэдра. чтобы пояснить это, нам будет удобен термин «мотив». Им часто пользуются специалисты, изучающие проявления симметрии и её законы в самых различных областях (математике, физике, химии, кристаллографии и т. д.). Мотив — это и есть та часть (сама не обязательно симметричная) симметричного объекта, за счёт правильного повторения которой и создаётся ощущение симметрии. Неодинаковые «одежды» фигур или иных предметов, сходных по типу симметрии, мы можем объяснить теперь тем, что различие в исходном мотиве сочетается у них с одинаковым законом (регулируемым группой) правильного чередования мотива.

Выбирая тот или иной мотив (здесь всё зависит от нашей фантазии) и желаемую группу V_n или D_n, мы можем создать любой доступный воображению узор. Для этого нужно лишь применить преобразования из группы к избранному мотиву, получающаяся при этом фигура и будет иметь исходную группу своей группой симметрии. На рис. 28 изображён мотив («аккуратно» разбитый жбан) и ниже — узор, полученный из него с помощью группы D_{10} и не лишенный некоторой прихотливости.

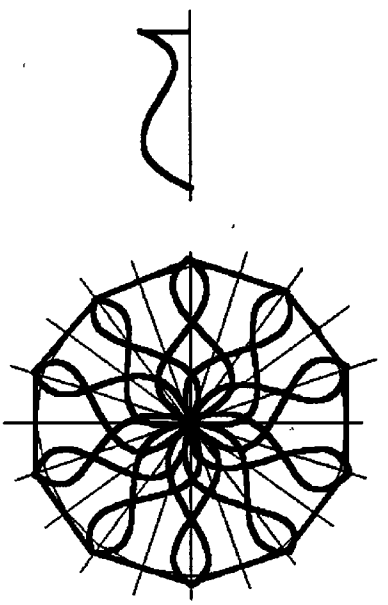


Рис. 28.

От симметричных фигур ограниченных размеров перейдём теперь к потенциально бесконечным узорам (их назы-

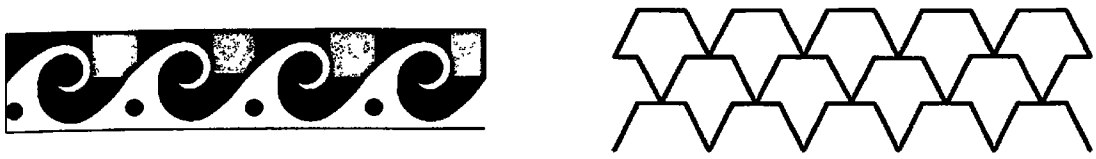


Рис. 29.

вают орнаментами), в которых повторение мотива достигается не только с помощью поворотов и отражений, но также посредством сдвигов и скользящих отражений. Ясно видно различие между двумя орнаментами, изображёнными на рис. 29. Группа симметрии первого из них содержит сдвиги лишь в одном направлении, такие орнаменты мы назовём *линейными* или короче *бордюрами*.

Другой же орнамент допускает параллельные переносы в двух различных (неколлинеарных) направлениях, орнаменты такого рода называются *плоскими* или *сетчатыми*. И в том и другом случае возможна полная классификация, подобная той, которую мы уже указали для узоров ограниченных размеров. Удивительно, что множество различных типов (т. е. имеющих различные группы симметрии) орнаментов конечно. Можно указать даже точные числа. Всего имеется 7 различных типов бордюров, и несколько богаче, как, впрочем, и следует ожидать, многообразие типов плоских орнаментов — их 17. С некоторыми оговорками это означает, что исходя из заданного мотива мы можем построить ровно 7 различных бордюров и ровно 17 различных плоских орнаментов с тем или иным законом повторения исходного мотива. Хотя искусство орнамента и является одним из древнейших искусств, но выяснение упомянутых фактов относится лишь ко второй половине XIX века. Не менее поражает то, что все типы орнаментов (7 и 17) были экспериментально открыты уже в Древнем Египте фантазией тамошних художников-декораторов. Работа выдающегося русского геометра и кристаллографа Е. С. Фёдорова (1853–1919), в которой впервые была дана полная классификация орнаментов, подтвердила, что египтяне в этой области знали действительно всё.

Попытаемся теперь представить себе, каким путём может быть получена эта классификация. В дальнейшем речь будет идти

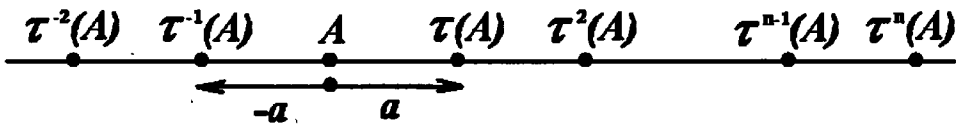


Рис. 30.

только о плоских орнаментах[1], поэтому эпитет «плоский» мы часто будем опускать.

Между тем другой эпитет, «сетчатый», указывает на связь орнаментов этого типа с простым геометрическим понятием сетки или решётки точек на плоскости. Такую решётку мы можем получить следующим образом. Возьмём произвольную точку A на плоскости и будем действовать на неё всевозможными переносами, содержащимися в группе симметрии орнамента (напомним, что они образуют подгруппу). Структура образующегося при этом множества Σ очень проста, и чтобы разобраться в ней, поступим так. Среди всех переносов выберем тот, который сдвигает исходную точку на вектор a наименьшей возможной длины (обозначим этот сдвиг через τ). Множество точек, получающихся при итерациях сдвига τ и обратного к нему преобразования τ^{-1} (т. е. преобразованиях τ^n, где $n \in \mathbb{Z}$), состоит, очевидно, из равноотстоящих точек прямой, параллельной вектору a (рис. 30).

Среди сдвигов, имеющих иное направление (они обязательно имеются), вновь выберем такой перенос σ, который сдвигает точку A на вектор b наименьшей длины. Множество точек $\sigma^m(A)$ ($m \in \mathbb{Z}$) имеет ту же структуру, что и предыдущее. Ясно, что в подгруппе переносов наряду с преобразованиями τ^n и σ^m содержатся и их комбинации $\tau^n\sigma^m$. Порядок перемножения здесь не важен, поскольку как преобразование $\tau^n\sigma^m$, так и $\sigma^m\tau^n$ представляют из себя сдвиг на вектор $na + mb$ (рис. 31).

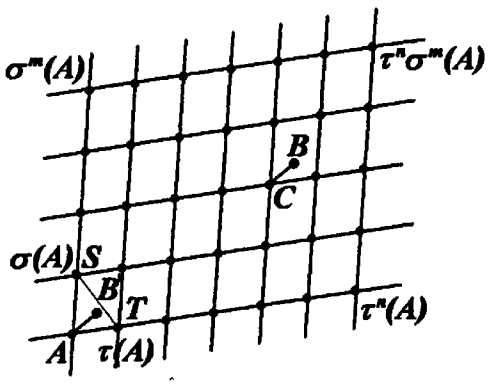


Рис. 31.

[1] Краткая информация о бордюрах имеется в дополнении.

Таким образом, действуя на точку A всевозможными преобразованиями указанного вида, мы получим множество точек, расположенных на векторах a и b и заполняющих всю плоскость так, как это показано на том же рис. 31. Это множество и называется плоской решёткой или сеткой (термин «сетка» лучше, впрочем, относится к множеству прямых, соединяющих точки решётки). Покажем, что полученная решётка исчерпывает всё множество Σ, здесь-то мы и воспользуемся специальным выборов переносов τ^n и σ^m. Одновременно мы покажем, что подгруппа переносов не содержит никаких иных преобразований, кроме $\tau^n\sigma^m$ $(n, m \in \mathbb{Z})$.

Предположим, что в исходной группе симметрии имеется сдвиг g, переводящий точку A в некоторую другую точку B, не принадлежащую решётке. Среди всех её точек рассмотрим ту точку C, которая является ближайшей к B. Тогда для некоторых $\alpha, \beta \in \mathbb{Z}$, точка $C = \tau^\alpha\sigma^\beta(A)$. Применим в таком случае к исходной точке A преобразование $f = (\tau^\alpha\sigma^\beta)^{-1}g$. Получающаяся при этом точка $B' = f(A)$, окажется, как нетрудно понять, в параллелограмме с вершиной A и по отношению к последней займёт такое же положение, каково положение точки B относительно C. Из неравенства $|AS| \geqslant |AT|$, вытекающего из определения сдвигов τ^n и σ^m, очевидно следует тогда, что

$$|AB'| < |AS|. \tag{11.14}$$

С другой стороны, перенос f как произведение преобразований из группы симметрии также принадлежит этой группе, в то же время неравенство (11.14) показывает, что он сдвигает точку на меньшее расстояние, чем σ. Это противоречит выбору σ, а полученное противоречие доказывает, что множество Σ совпадает с решёткой на рис. 31.

Для всякого орнамента решётка является своего рода каркасом, на котором нанизывается повторяющийся мотив. К тому же она, как понятно из предыдущего, позволяет указать все переносы, самосовмещающие заданный орнамент. Ими являются сдвиги на любой из векторов, соединяющих две произвольные точки решётки.

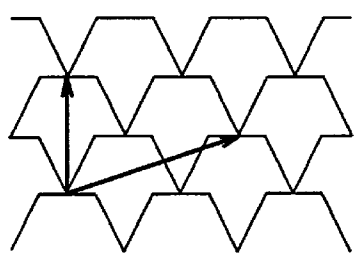


Рис. 32.

В группе симметрии орнамента, как уже отмечалось, переносы могут соседствовать с поворотами и зеркальными отражениями. Оказывается, решётка содержит в себе немалую информацию и о том, какими могут быть эти «соседи». В самом деле, если некоторый поворот или отражение самосовмещают орнамент, то и решётку этого орнамента они должны переводить саму в себя. Это накладывает довольно жёсткие ограничения на возможные в группе симметрии повороты и отражения. Разглядывая орнаменты на рис. 40 (с. 206), мы не обнаружим ни в одном из них центров поворота, порядок которых был бы отличен от 2, 3, 4, 6. Это обстоятельство отнюдь не является случайным, — то же самое, оказывается, справедливо и для любого мыслимого орнамента. Чтобы это понять, предположим, что некоторая точка A его решётки является центром поворота порядка n. Поворот на угол $2\pi/n$ вокруг точки A приведёт тогда к самосовмещению решётки, поэтому всякая её точка B перейдёт в точку B', снова ей принадлежащую (рис. 33).

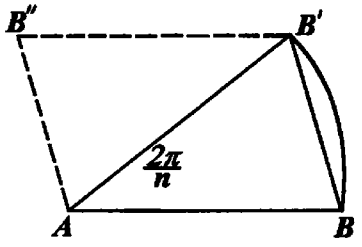


Рис. 33.

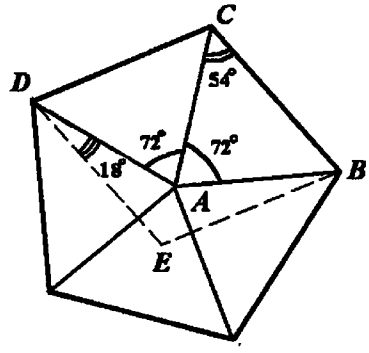


Рис. 34.

Если в качестве B выбрана ближайшая к A точка решётки, то при $n > 6$ мы сразу получим противоречие этому выбору. В самом деле,

$$|B'B| = 2|AB|\sin\frac{\pi}{n} < \frac{2\pi}{n}|AB| \leqslant \frac{2\pi}{7}|AB| < |AB|.$$

Но тогда $|AB''| = |BB'| < |AB|$, и следовательно, точка B'' (очевидно, принадлежащая решётке) вопреки предположению окажется к точке A ближе, чем B. Легко исключается и единственный оставшийся случай $n = 5$. На рис. 34 изображён центр по-

ворота A указанного порядка. Если существует решётка, содержащая точку A, то она обязана вместе с ней и близлежащей к A точкой B содержать и повёрнутые точки C, D, а тогда и точку E — четвёртую вершину параллелограмма, построенного на отрезках BC и CD. Вполне понятно, что последняя снова, вопреки предположению, отстоит от центра поворота A на меньшее расстояние, чем точка B.

Итак, возможный список групп, образованных поворотами того или иного орнамента вокруг фиксированного центра, невелик, он исчерпывается следующими пятью группами:

$$V_1,\ V_2,\ V_3,\ V_4,\ V_6. \tag{11.15}$$

V_1 состоит лишь из тождественного преобразования — поворота на нулевой угол). Если же точечная группа орнамента (так называют группы, оставляющие неподвижной некоторую точку) кроме поворотов содержит ещё и отражения относительно осей, проходящих через центр поворота, то она может быть лишь одной из следующих групп диэдра:

$$D_1,\ D_2,\ D_3,\ D_4,\ D_6. \tag{11.16}$$

Мы увидим, что каждый из десяти случаев (11.15) и (11.16) можно воплотить в плоском орнаменте. Иначе, для любой из десяти групп существует орнамент, точечная группа которого для подходящей точки совпадает с заданной. Но сначала вернёмся к решёткам. Каждая из них представляет из себя в некотором роде простейший орнамент, в качестве исходного мотива которого выбрана точка. Естественно поэтому, что они поддаются и более простой классификации, чем произвольные орнаменты.

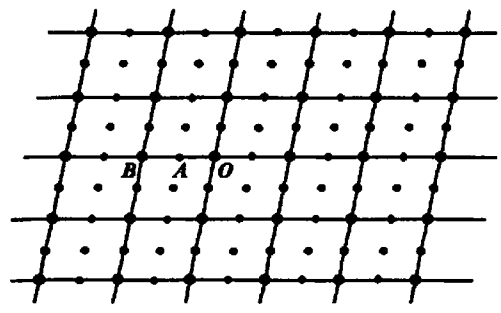
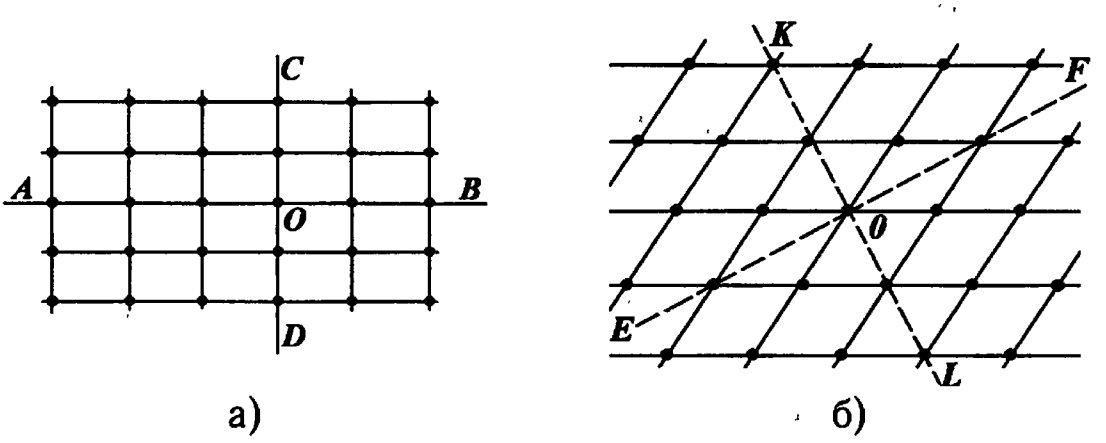


Рис. 35.

Нетрудно увидеть, что любая данная решётка обладает центральной симметрией относительно середины отрезка, соединяющего любые две точки. На рис. 35 соответствующие центры симметрии выделены кружками. Решётка называется общей, если

её самосовмещения исчерпываются параллельными переносами и только что указанными центральными симметриями. Можно показать, что так будет в том и только том случае, когда величины двух её минимальных сдвигов τ_1 и τ_2 различны и угол между ними отличен от 60° и 90°.

Предположим теперь, что решётка имеет ось симметрии и преобразование σ_1 — отражение относительно этой оси. Здесь возможны два случая — оба они представлены на рис. 36.



а) б)

Рис. 36.

В первом из них какая-то из осей симметрии (прямая AB) совпадает с одной из прямых параллелограммной сетки. Ясно, что в этом случае решётка должна быть прямоугольной. При этом, если υ — центральная симметрия относительно точки O, то

$$\sigma_2 = \sigma_1 \upsilon$$

является зеркальным отражением относительно прямой CD, перпендикулярной AB.

Во втором случае (рис. 36 б)) ни одна из осей симметрии не лежит на прямых параллелограммной сетки. Следовательно, каждая из осей должна разбивать всякий пересекаемый ею параллелограмм на две зеркально симметричные части. Нетрудно увидеть, что такое возможно лишь, если каждый параллелограмм решётки является ромбом, а ось симметрии проходит через его диагональ. Мы приходим к так называемой ромбической решётке. Заметим, что и этом случае группа симметрии содержит зеркальные симметрии σ_1 и σ_2 относительно перпендикулярных осей

(например, EF и KL), причём снова

$$\sigma_2 = \sigma_1 \upsilon.$$

Группы прямоугольной и ромбической решёток отличны от группы симметрии общей решётки. Последняя в качестве своих точечных подгрупп содержит лишь группы V_2, в то время как группа симметрии решёток, изображённых на рис. 36, содержит группы D_2.

Предположив существование центров поворота 4-го порядка, мы придём к квадратной решётке, которая сочетает в себе свойства прямоугольной и ромбических решёток. Она содержит сдвиги в двух взаимно перпендикулярных направлениях, и в то же время величины минимальных из этих сдвигов равны. Действительно, если υ — вращение на угол $\pi/2$ вокруг центра поворота, а τ_1 — минимальный сдвиг в одном из направлений, то

$$\tau_2 = \tau_1 \upsilon$$

— сдвиг той же величины в перпендикулярном направлении. Применяя к произвольной точке O решётки преобразования

$$\tau_1^\alpha \tau_2^\beta \quad (\alpha, \beta \in \mathbb{Z})$$

мы, как обычно, получим все другие точки решётки, которая и окажется квадратной (рис. 37). Отметим, что решётка инвариантна не только относительно группы V_4, но и относительно более широкой группы диэдра D_4, состоящей из четырёх вращений (на углы 0, $\pi/2$, π, $3\pi/2$) вокруг произвольной точки O решётки и четырёх отражений (оси этих отражений, проходящие через точку O, показаны на рис. 37).

Ещё только один случай не охвачен предыдущими рассмотрениями. В самом деле, пусть группа симметрии решётки содержит группу

$$V_3 = \langle \upsilon \rangle,$$

где υ — поворот на $2\pi/3$ вокруг некоторого центра поворота. Вновь, если τ_1 — минимальный сдвиг в одном из направлений, то

$$\tau_2 = \tau_1 \upsilon$$

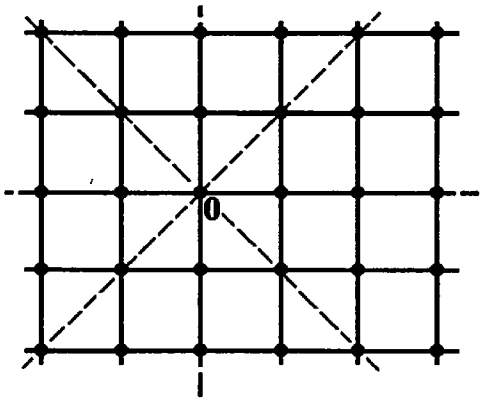


Рис. 37.

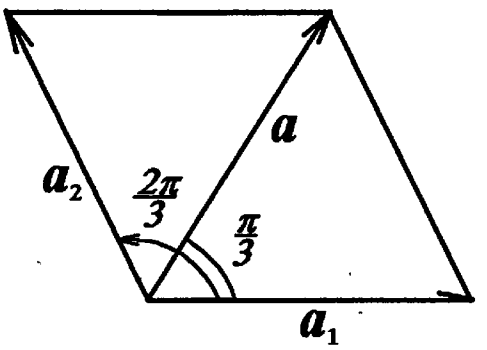


Рис. 38.

— такой же величины сдвиг в направлении, составляющем с исходным угол $2\pi/3$ (см. рис. 38, на котором a_1 и a_2 — векторы, соответствующие параллельным переносам τ_1 и τ_2).

Но тогда группа симметрии решётки должна содержать перенос

$$\tau = \tau_1\tau_2$$

— его вектором является

$$a = a_1 + a_2,$$

составляющий угол $\pi/3$ со слагаемыми. Действуя на любой узел преобразованиями $\tau_1^\alpha\tau_2^\beta$ ($\alpha, \beta \in \mathbb{Z}$) мы получим так называемую шестиугольную решётку (рис. 39).

Решётка содержит центры поворотов 6-го порядка (узлы решётки), 3-го порядка (центры правильных треугольников) и 2-го порядка (середины отрезков, соединяющих соседние узлы решётки). Точечная группа, самосовмещающая решётку и оставляющая неподвижным её узел, — это группа D_6. Группы D_3 и D_2 совпадают с теми точечными группами, которые оставляют неподвижными центры поворотов соответственно 3-го и 2-го порядка.

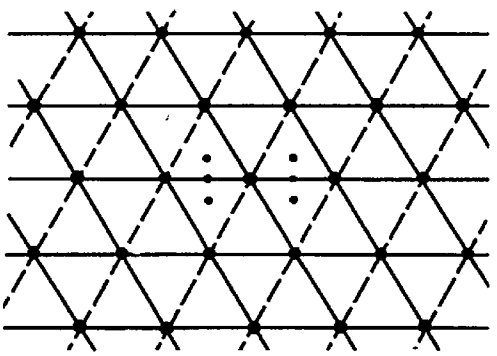


Рис. 39.

Наша классификация решёток и соответствующих групп симметрии закончена. Её итог — 5 типов решёток и 5 групп, являю-

щихся их группами симметрии. Абстрактное строение этих групп рассматривается в дополнении.

Напомним теперь, что группа симметрии произвольного орнамента должна сохранять его решётку. Иными словами, всякая такая группа непременно является подгруппой в группе симметрии решётки того или иного типа, причём такой, которая содержит сдвиги в обоих направлениях. Детальный анализ (он выходит за рамки нашей книги) показывает, что в перечисленных 5 типах групп имеется в общей сложности 17 неизоморфных подгрупп, содержащих неколлинеарные сдвиги. При этом каждая из них может быть реализована в некотором орнаменте (т. е. является его группой симметрии). Схемы всех 17 плоских орнаментов приведены на рис. 40. При построении их использован однотипный мотив — окрашенный в чёрное треугольник.

Орнаменты 1 и 5 на рис. 40 не содержат ни осей симметрии, ни центров поворота, иначе говоря их точечные группы тривиальны (совпадают с V_1). Но второй, в отличие от первого, выдерживает скользящую симметрию, это обстоятельство вызывает отличие и абстрактного строения их групп. Орнаменты 4 и 6 на рис. 40 уже обладают осями симметрии, и их группы содержат подгруппы D_1 по одной для каждой оси, но не содержат никаких иных подгрупп из списка (11.15), исключая, конечно, тривиальный случай группы V_1.

Парой взаимно перпендикулярных осей симметрии обладают орнаменты 8, 10, а их группы симметрии содержат подгруппы, совпадающие с D_2. В таблице 9 приводится полная информация о том, какие подгруппы из (11.15) и (11.16) содержатся в группах симметрии каждого из 17 орнаментов[1].

Ещё больший интерес (не только математический, но и практический) представляет задача классификации «пространственных орнаментов». Хотя такие орнаменты и не используются в декоративном искусстве, но они встречаются в природе. С давних времён воображение людей поражали строгие геометрические

[1] При рассмотрении этой таблицы читатель должен иметь в виду, что некоторые из групп (11.15) и (11.16) содержатся в других группах этого списка. Так что, если группа симметрии обладает, например, подгруппой D_4, то автоматически она содержит и группы V_4, D_2, V_2.

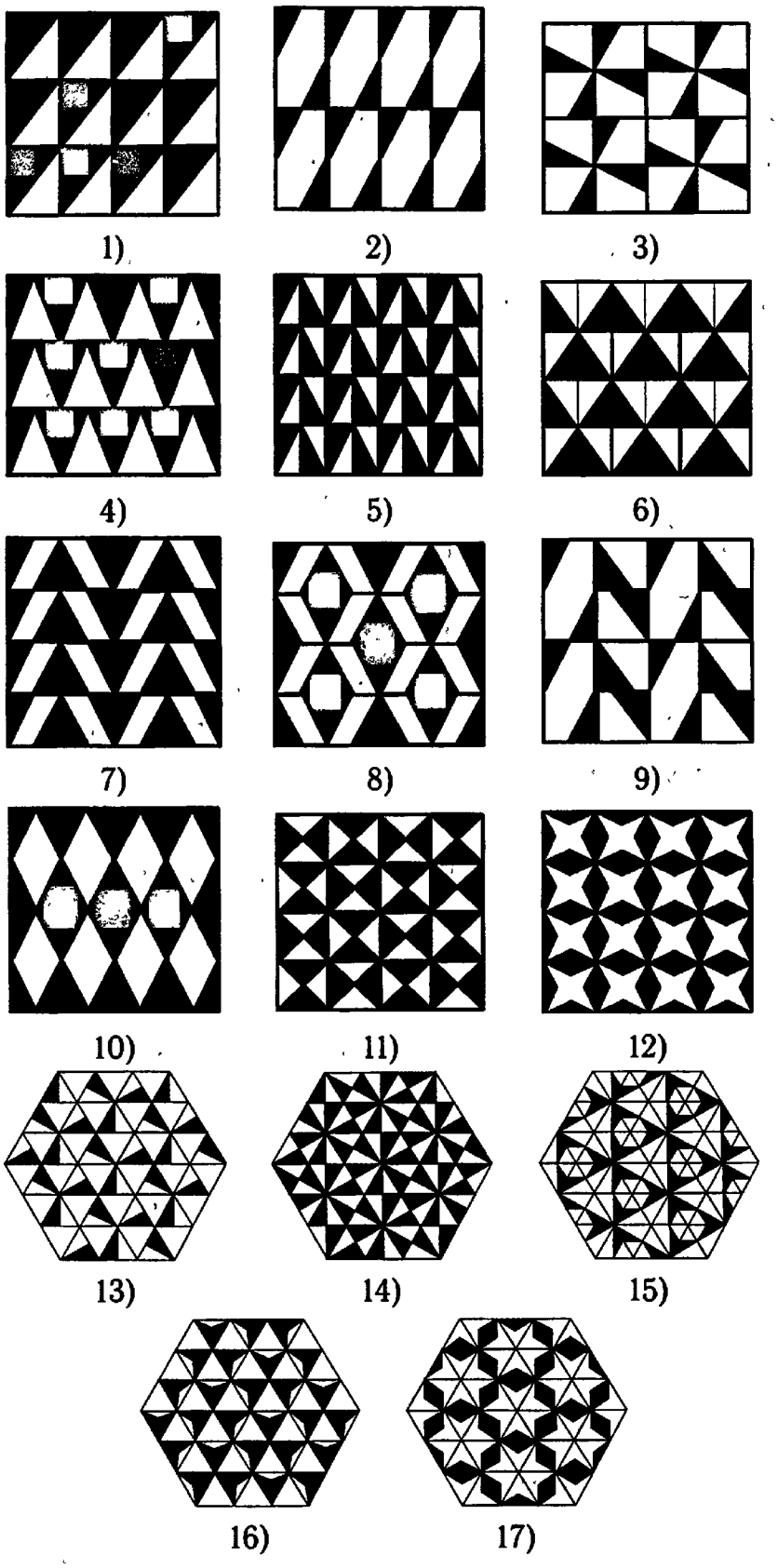


Рис. 40.

Таблица 9.

Тип орнамента	Точечные группы
1, 5	V_1
2, 9	V_2
4, 6	D_1
7	D_1, V_2
3	V_4
8, 10	D_2

Тип орнамента	Точечные группы
11	D_2, V_4
12	D_4
13	V_3
14	V_6
15, 16	D_3
17	D_6

формы кристаллов, в которых воплощались самые разнообразные и причудливые виды симметрии (рис. 41). Чем объясняются эти столь совершенные формы, что управляет их многообразием и как велико это многообразие — такими вопросами задавались многие учёные. Лишь постепенно было выяснено, что внешняя симметрия кристаллов вызвана особенностями внутреннего строения кристаллического вещества, правильностью пространственного расположения составляющих его молекул и атомов.

Оказалось, что атомная структура кристаллов сродни решётке, только уже не плоской, а пространственной. Устройство таких решёток для кристаллов поваренной соли и алмаза показано на рис. 42.

Кристаллическая решётка образует для вещества кристалла каркас, подобно тому, каким плоская решётка является для построенного на ней орнамента. Важнейшие свойства кристаллических решёток, от которых и зависит симметрия кристалла, могут быть охарактеризованы чисто математически.

Множество точек пространства называется правильной пространственной системой точек, если выполняются следующие свойства:

1. Любую точку системы можно перевести в любую другую её точку с помощью движения, совмещающего систему с собой.

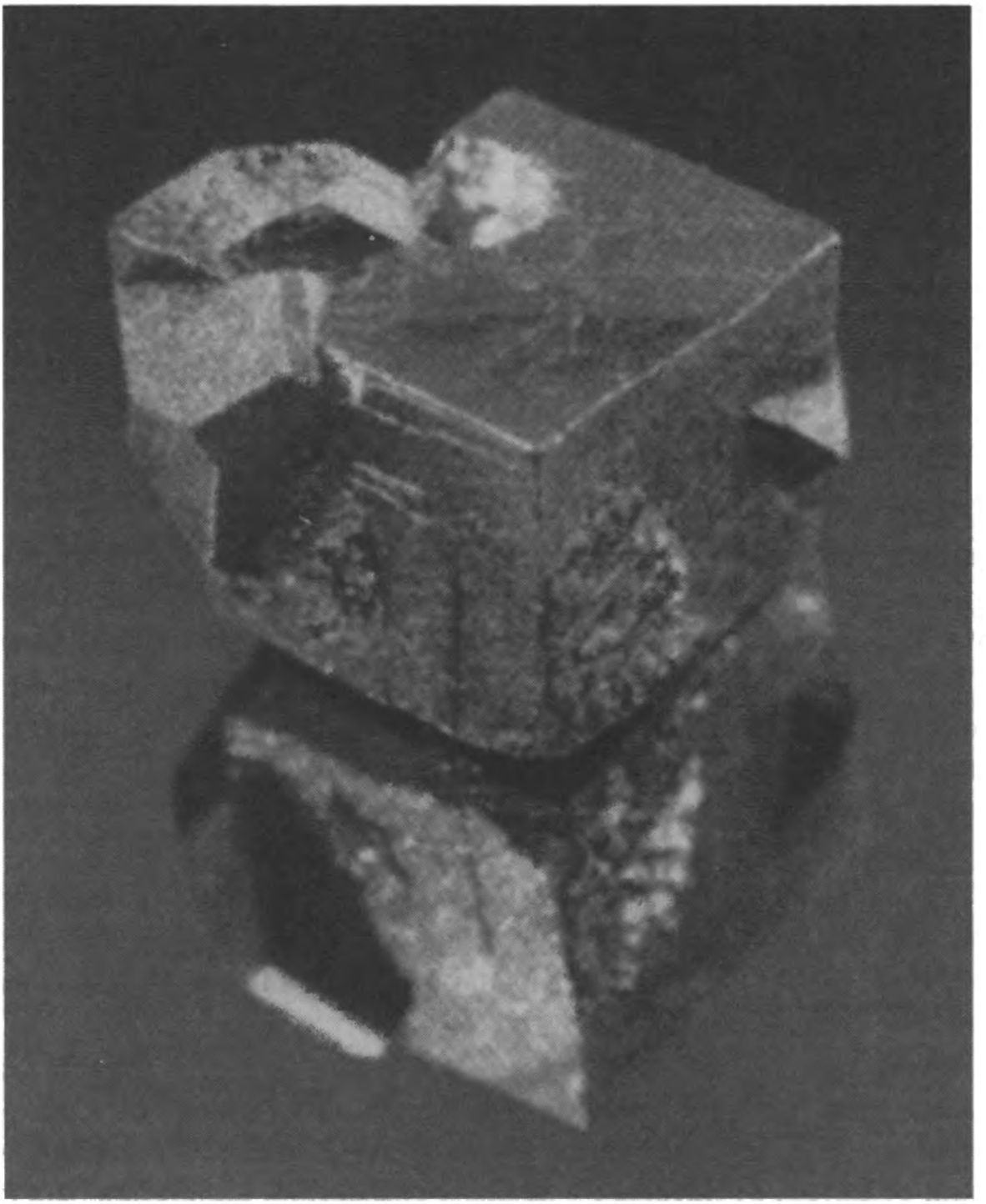


Рис. 41.

2. Всякий шар конечного радиуса содержит лишь конечное множество точек системы.

3. Существует такое положительное число r, что всякий шар радиуса r содержит хотя бы одну точку системы.

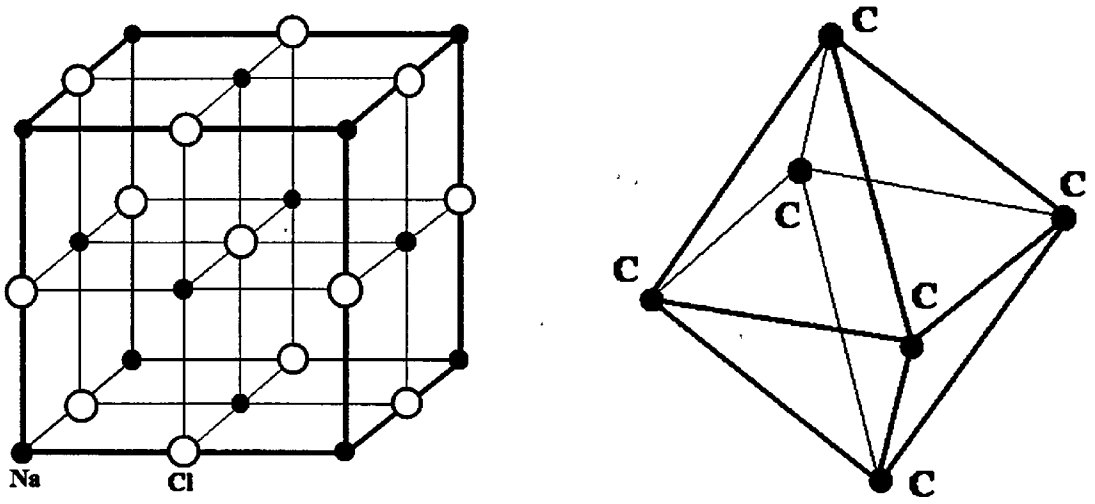


Рис. 42.

Нетрудно понять, что рассматривавшиеся выше решётки плоских орнаментов удовлетворяют всем этим требованиям, если только в пунктах 2 и 3 слово «шар» заменить словом «круг». Иными словами, плоские решётки образуют правильные системы точек на плоскости.

На первый взгляд может показаться, что список свойств 1, 2, 3 слишком скромен для того, чтобы из него можно было бы вывести какие-то суждения о возможных формах симметрии кристаллов и тем более полностью обозреть все эти формы. Однако это не так. Детальный анализ, который впервые полностью провёл русский ученый Е. С. Фёдоров, показал, что в пространстве (как и на плоскости) имеется лишь конечное число существенно различных правильных систем точек, и каждая из них отвечает реально встречающейся в природе кристаллической структуре, полностью исчерпывая всё многообразие таких структур[1]. Выяснилось, что имеется 230 возможностей, в которых реализуются все природные формы симметрии кристаллов. Обозрение и классификация этих возможностей основывались, как и в случае орнаментов, на рассмотрении групп симметрии кристаллических решёток[2], или иначе, правильных систем точек в пространстве.

[1] Результаты Е. С. Фёдорова были опубликованы в 1890г. в книге «Симметрия правильных систем фигур». К этому же открытию несколько позже (1891) пришёл немецкий математик А. Шёнфлис.

[2] Т. е. групп движений пространства, совмещающих решётку саму с собой.

Эти группы называются фёдоровскими или кристаллографическими группами. Их полное описание и позволило окончательно разобраться со всеми видами симметрии, присущими кристаллам. Подробный вывод и перечисление всех фёдоровских групп даже и сейчас требует не одного десятка страниц. Подобно тому, как для описания плоских движений применяются комплексные числа, здесь могут быть использованы кватернионы. Полезен также и аппарат линейной алгебры (краткие сведения о ней — в главе 13).

Дополнения и задачи.

1. Докажем утверждение, упомянутое в основном тексте: если группа симметрий G плоской фигуры конечна, то для некоторого n она совпадает с одной из групп V_n или D_n.

Во-первых, как уже отмечалось, такая группа G может содержать лишь повороты и зеркальные отражения. Установим последовательно несколько простых фактов.

1) Если v — произвольный поворот из G, то его угол α обязательно имеет вид

$$\alpha = 2\pi r, \quad \text{где } r \in \mathbb{Q}.$$

Действительно, для некоторого n $v^n = e$ — тождественное преобразование, тогда угол $n\alpha$ обязан быть кратным 2π:

$$n\alpha = 2\pi r \cdot n = 2\pi m, \quad \text{для некоторого } m \in \mathbb{Z}.$$

Из последнего равенства и следует, что $r = m/n$ является числом рациональным.

2) В группе G должен существовать поворот на наименьший положительный угол.

Если это не так, то можно указать бесконечную убывающую последовательность положительных чисел

$$\alpha_1, \alpha_2, \ldots, \alpha_k, \ldots,$$

члены которой являются углами поворотов

$$v_1, v_2, \ldots, v_k, \ldots,$$

очевидно, различных. Понятно, что это противоречит конечности группы G.

3) Если v — поворот на наименьший положительный угол α, то угол любого другого поворота кратен α.

Пусть u — произвольный поворот на угол β. Для чисел α и β непременно найдется такое целое k, что

$$k\alpha \leqslant \beta < (k+1)\alpha$$

Рассмотрим тогда преобразование из G, равное uv^{-k}. Если $\beta > k\alpha$, то это преобразование есть поворот (см. дополнение к главе 2) на угол $\beta - k\alpha$, очевидно, меньший, чем α, что невозможно. Поэтому $\beta = k\alpha$.

Заметим теперь, что поворот u должен иметь тот же центр, что и v. В противном случае uv^{-k} являлось бы сдвигом на ненулевой вектор, а преобразований такого типа группа G не содержит. Итак, u и v — повороты вокруг общего центра и тогда $uv^{-k} = e$ — тождественное преобразование, а значит $u = v^k$.

В итоге доказано, что множество всех поворотов группы G образуют циклическую группу $\langle v \rangle$ состоящую из вращений v^k вокруг фиксированного центра.

Если порядок поворота v равен n, то ясно, что угол $\alpha = 2\pi/n$, множество всех поворотов, содержащихся в G совпадает с V_n.

В том случае, когда группа G не содержит зеркальных отражений, справедливо равенство $G = V_n$.

Пусть, наконец, σ — зеркальная симметрия относительно оси, содержащейся в G. Тогда преобразования

$$\sigma v^k, \quad k = 0, 1, \ldots, n-1.$$

— также зеркальные симметрии относительно осей, повернутых на соответствующие углы. Нетрудно убедиться, что других зеркальных отражений группа G не содержит. Отсюда следует, что $G = D_n$.

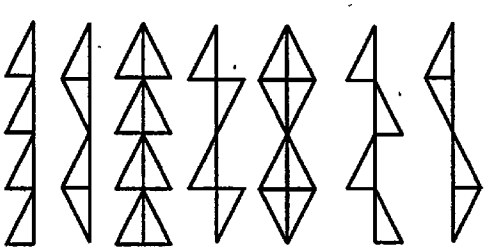


Рис. 43.

2. На рис. 43 приведены всевозможные типы (их семь) одномерных орнаментов — бордюров. Выясните, какие движения входят

в группу симметрии каждого из них. Попробуйте описать абстрактное строение группы симметрии каждого из них. Например, группа бордюра 1 содержит лишь переносы и изоморфна бесконечной циклической группе $\langle\tau\rangle$, где τ — наименьший перенос на величину a. А группа симметрии бордюра 2 кроме переносов содержит ещё отражения относительно осей, перпендикулярных направлению переноса. Если σ — одно из таких отражений, τ, как и выше, наименьший перенос, то группа симметрии порождена этими двумя элементами,

$$G = \langle\tau, \sigma\rangle\,, \tag{11.17}$$

а её определяющие соотношения таковы:

$$\sigma^2 = e; \quad \sigma^{-1}\tau\sigma = \tau^{-1}. \tag{11.18}$$

Можно сказать также, что G есть полупрямое произведение (см. дополнение к главе 10) $\langle\sigma\rangle \leftthreetimes \langle\tau\rangle$ циклической группы $\langle\sigma\rangle$ порядка 2 и бесконечной циклической группы $\langle\tau\rangle$, причём $\sigma^{-1}\tau\sigma = \tau^{-1}$.

3. Рассмотрим более подробно устройство групп симметрии некоторых из рассмотренных в основном тексте плоских решёток.

Сначала общая решётка. Её группа G, кроме двух неколлинеарных переносов τ_1 и τ_2 и всевозможных произведений $\tau_1^{\alpha_1}\tau_2^{\alpha_2}$ содержит ещё, как упоминалось, центральные симметрии (повороты на 180°). Пусть v — один из таких поворотов вокруг некоторого узла решётки (рис. 35, с. 201). Тогда для любого сдвига τ комбинация $v\tau$ — снова поворот, на рисунке это поворот вокруг точки A, являющейся серединой отрезка, соединяющего узлы O и $B = \tau^{-1}(0)$ решётки. Таким путём можно получить повороты вокруг любого из центров симметрии общей решётки, т. е. все они записываются в виде

$$v\tau = v\tau_1^{\alpha_1}\tau_2^{\alpha_2}.$$

Остальные элементы группы G, очевидно, являются переносами $\tau = \tau_1^{\alpha_1}\tau_2^{\alpha_2}$. Сказанное означает, что преобразования v, τ_1, τ_2 являются образующими рассматриваемой группы:

$$G = \langle v, \tau_1, \tau_2\rangle\,.$$

Для полного задания G необходимо ещё указать её определяющие соотношения. Приведём их список:

$$\begin{aligned} v^2 &= e & \tau_1\tau_2 &= \tau_2\tau_1 \\ v^{-1}\tau_1 v &= \tau_1^{-1} & v^{-1}\tau_2 v &= \tau_2^{-1}. \end{aligned} \tag{11.19}$$

Первые два соотношения понятны, проверку остальных предоставляем читателю. Равенства (11.19) позволяют находить произведения любых элементов группы G, каждый из которых, как мы уже выяснили, допускает запись

$$v^\epsilon \tau_1^{\alpha_1}\tau_2^{\alpha_2}, \quad \text{где} \quad \varepsilon = 0 \text{ или } 1; \quad \alpha_1, \alpha_2 \in \mathbb{Z}. \tag{11.20}$$

Например,

$$(\tau_1^2\tau_2)(v\tau_1\tau_2^2) = v(\tau_1^{-2}\tau_2^{-1})(\tau_1\tau_2^2) = v\tau_1^{-1}\tau_2.$$

Нетрудно убедиться, что запись (11.20) для всякого преобразования решётки единственна, а группа G есть полупрямое произведение

$$G = \langle v \rangle \leftthreetimes \langle \tau_1, \tau_2 \rangle,$$

в котором $\langle v \rangle$ — циклическая группа порядка 2, а $\langle \tau_1, \tau_2 \rangle$ — прямое произведение двух бесконечных циклических групп. Группа симметрий прямоугольной решётки устроена несколько сложнее. Докажите, что она порождена четырьмя преобразованиями: переносами τ_1 и τ_2 в двух взаимно-перпендикулярных направлениях и двумя зеркальными симметриями σ_1 и σ_2 относительно осей, перпендикулярных направлениям τ_1 и τ_2 соответственно. То есть в этом случае

$$G = \langle \sigma_1, \sigma_2, \tau_1, \tau_2 \rangle.$$

Докажите также, что определяющие соотношения группы G следующие:

$$\sigma_1^2 = \sigma_2^2 = e, \quad \sigma_1\sigma_2 = \sigma_2\sigma_1, \quad \tau_1\tau_2 = \tau_2\tau_1,$$
$$\sigma_1\tau_1\sigma_1 = \tau_1^{-1}, \quad \sigma_2\tau_2\sigma_2 = \tau_2^{-1}, \quad \sigma_1\tau_2 = \tau_2\sigma_1, \quad \sigma_2\tau_1 = \tau_1\sigma_2.$$

Выведите отсюда, что группа прямоугольной решётки разлагается в прямое произведение двух групп

$$\langle \sigma_1, \tau_1 \rangle \times \langle \sigma_2, \tau_2 \rangle ,$$

каждая из которых изоморфна группе (11.17) с определяющими соотношениями (11.18).

Опишите группы симметрий остальных решёток в терминах образующих и определяющих соотношений.

4. Подобным образом могут быть описаны и группы симметрий каждого из 17 орнаментов, изображённых на рис. 40 (с. 206). Сделайте это.

ГЛАВА 12

АЛГЕБРА В КОМБИНАТОРИКЕ

Комбинаторика как самостоятельный раздел математики возникла где-то во второй половине XVII века, и к её рождению причастны великие математики, такие как Ферма, Паскаль, Лейбниц, Я. Бернулли. В «Диссертации о комбинаторном искусстве» (1666 г.), принадлежащей перу Лейбница, впервые употребляется и сам термин «комбинаторный» примерно в том же смысле, в котором мы используем его сегодня. Не так просто в двух словах объяснить, чем занимается комбинаторика (или, как чаще теперь говорят, комбинаторный анализ). Круг её вопросов достаточно велик (см., например, [7], [8], [27]). Но одна из её областей, которой мы и коснёмся в этой главе, связана с подсчётом числа (перечислением) тех или иных комбинаций, составляемых по разным правилам из элементов конечных множеств. Найти число возможных исходов (распределений мест) футбольного чемпионата, или число способов разменять рубль, или число химических соединений того или иного вида и т. д. и т. п. — всё это вопросы, подвластные искусству комбинаторики.

Мы предполагаем, что читатель знаком с простейшими из комбинаций, рассматриваемых в комбинаторике, — *сочетаниями*, *размещениями* и *перестановками*[1] и что ему известны также формулы для их подсчёта:

$$C_n^m = \frac{n!}{m!(n-m)!} = \frac{n(n-1)\dots(n-m+1)}{m!}$$

[1] Если это не так, советуем обратиться к книге [8].

— число всех сочетаний без повторения из n элементов по m;

$$A_n^m = \frac{n!}{(n-m)!}$$

— число всех размещений без повторения из n элементов по m;

$$P_n = n! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot n \quad \text{(читается «}n\text{ факториал»)}$$

— число всех перестановок из n элементов.

Середина прошлого века оказалась периодом второго рождения комбинаторики. Бурное развитие вычислительной техники и прикладной математики способствовало активному применению комбинаторных методов для решения самых различных практических задач. Спектр этих применений широк: химия и биология, статистическая физика и кристаллография, теория кодирования и планирование эксперимента, экономика и исследование операций... Теперь уже и сама комбинаторика — не изолированная, как прежде, наука, — она черпает свои методы из всего арсенала современной математики, и не последнюю роль здесь играет алгебра.

Имеется ряд путей, связывающих алгебру с комбинаторикой. Один из них проходит через так называемые производящие функции. Об этом мы и поговорим подробнее.

Большинству читателей известна, наверное, формула бинома Ньютона[1]

$$(a+b)^n = a^n + C_n^1 a^{n-1}b + \ldots + C_n^k a^{n-k}b^k + \ldots + b^n = \sum_{k=0}^{n} C_n^k a^{n-k}b^k,$$

или в других обозначениях

$$(1+x)^n = 1 + C_n^1 x + \ldots + C_n^k x^k + \ldots + x^n = \sum_{k=0}^{n} C_n^k x^k. \quad (12.1)$$

[1] Её частными случаями являются формулы для квадрата и куба суммы: $(a+b)^2 = a^2 + 2ab + b^2$, $(a+b)^3 = a^3 + 3a^2b + 3ab^2 + b^3$.

Формулу (12.1) можно получить из таких соображений (они будут полезны для дальнейшего). Рассмотрим сначала произведение

$$(1+a_1x)(1+a_2x)(1+a_3x). \tag{12.2}$$

Раскрывая в нём скобки, получим выражение

$$1+(a_1+a_2+a_3)\,x+(a_1a_2+a_1a_3+a_2a_3)\,x^2+(a_1a_2a_3)\,x^3,$$

коэффициенты которого «перечисляют» всевозможные сочетания без повторений из 3-х элементов соответственно по одному, по два, по три. Например, слагаемым a_1a_2, a_1a_3, a_2a_3 коэффициента при x^2 соответствуют всевозможные такие сочетания из 3-х элементов по 2. Совершенно аналогично

$$\begin{aligned}(1+a_1x)(1+a_2x)(1+a_3x)\ldots(1+a_nx) = \\ = 1+(a_1+a_2+a_3+\ldots+a_n)\,x+ \\ +(a_1a_2+a_1a_3+\ldots+a_{n-1}a_n)\,x^2+ \\ +(a_1a_2a_3+a_1a_2a_4+\ldots+a_{n-2}a_{n-1}a_n)\,x^3+\ldots \\ \ldots+(a_1a_2\,a_3\ldots a_n)\,x^n.\end{aligned} \tag{12.3}$$

Здесь снова каждый коэффициент при произвольной степени x^k перечисляет в указанном выше смысле всевозможные сочетания из n элементов по k, и следовательно, число слагаемых в этом коэффициенте совпадает с числом C_n^k всех таких сочетаний. Полагая в равенстве $a_1=a_2=\ldots=a_n=1$, приходим к формуле (12.1).

Оказывается, с помощью алгебраических выражений, подобных (12.2) и (12.3), можно также получать решения многих других задач на перечисление. Так, если по аналогии с (12.2) рассмотреть многочлен

$$(1+a_1x+a_1^2x^2)(1+a_2x+a_2^2x^2)(1+a_3x+a_3^2x^2), \tag{12.4}$$

равный

$$\begin{aligned}1+(a_1+a_2+a_3)\,x+(a_1a_2+a_1a_3+a_2a_3+a_1^2+a_2^2+a_3^2)\,x^2+ \\ +(a_1a_2a_3+a_1a_2^2+a_1^2a_2+a_1\,a_3^2+a_1^2a_3+a_2^2a_3+a_2a_3^2)\,x^3+\ldots \\ \ldots+({a_1}^2{a_2}^2{a_3}^2)\,x^6,\end{aligned} \tag{12.5}$$

то можно заметить, что его коэффициенты, так же, как и в формуле (12.2), перечисляют сочетания из 3-х элементов, теперь, правда, уже такие, в которых элементы могут повторяться, а число повторений не превышает двух. Например, слагаемым коэффициента при x^3 соответствуют все такие сочетания по 3 элемента

$$a_1a_2a_3,\ a_1a_2a_2,\ a_1a_3a_3,\ a_1a_1a_2,\ a_1a_1a_3,\ a_2a_2a_3,\ a_2a_3a_3.$$

Значит, при $a_1 = a_2 = a_3 = 1$ мы получим многочлен

$$(1+x+x^2)^3 = 1+3x+6x^2+7x^3+6x^4+3x^5+x^6, \qquad (12.6)$$

коэффициенты которого дают число сочетаний для 3-х элементов с не более, чем двукратными повторениями. К примеру, число таких сочетаний из 3-х элементов по 4 совпадает с коэффициентом при x^4 и равно 6. Вот эти сочетания:

$$a_1a_1a_2a_3,\ a_1a_2a_2a_3,\ a_1a_2a_3a_3,\ a_1a_1a_2a_2,\ a_1a_1a_3a_3,\ a_2a_2a_3a_3.$$

Теперь ясно, что для перечисления сочетаний из n элементов с не более, чем двукратными повторениями, следует рассмотреть произведение

$$(1+a_1x+a_1^2x^2)(1+a_2x+a_2^2x^2)\dots(1+a_nx+a_n^2x^2).$$

Тогда коэффициенты при x, x^2, x^3, ..., x^{2n} многочлена

$$(1+x+x^2)^n \qquad (12.7)$$

дают число таких сочетаний по одному, по два, по три, ..., по $2n$ элементов. Подобные многочлены (они называются производящими функциями) можно использовать для подсчёта числа сочетаний с самыми разнообразными ограничениями на вхождение элементов. Скажем, произведение

$$(1+a_1x+a_1^2x^2)(a_2x+a_2^2x^2+a_2^3x^3)\,a_3x$$

перечисляет все сочетания из 3-х элементов, в которые a_1 входит не более 2-х раз, a_2 — не менее одного и не более 3-х раз, a_3 —

ровно один раз. Однако далеко не всегда многочлены оказываются достаточным орудием для построения производящих функций. Так будет, например, в задаче отыскания числа таких сочетаний из n элементов по m, в которых каждый из элементов может повторяться любое число раз. По аналогии с предыдущим можно понять, что в качестве производящей функции в этом случае следовало бы взять выражение

$$(1 + x + x^2 + \ldots + x^m + \ldots)^n, \tag{12.8}$$

где в скобках стоит бесконечная сумма. Такого типа бесконечные суммы

$$a_0 + a_1 x + a_2 x^2 + \ldots + a_m x_m + \ldots \quad (a_i \in \mathbb{R})$$

в математике называются *рядами*, точнее *степенны́ми рядами*. Некоторым, быть может, известно, что ряды часто служат удобным средством представления различных функций, рассматриваемых в математическом анализе, очень полезны они для целей приближённых вычислений и т. д. В таком контексте ряды требуют довольно аккуратного обращения с собой (например, нуждается в точном определении понятие суммы ряда, — ведь речь идёт о суммировании бесконечного множества слагаемых). Однако для наших целей будет достаточным ввести алгебраические операции над степенными рядами, понимаемыми как формальные выражения. Мы определим эти операции так, чтобы в случае конечный сумм (т. е. многочленов) они приводили к обычному сложению и умножению многочленов. Суммой двух степенных рядов

$$a_0 + a_1 x + a_2 x^2 + \ldots + a_m x^m + \ldots, \tag{12.9}$$

$$b_0 + b_1 x + b_2 x^2 + \ldots + b_m x^m + \ldots, \tag{12.10}$$

назовём ряд

$$(a_0 + b_0) + (a_1 + b_1)\, x + (a_2 + b_2)\, x^2 + \ldots + (a_m + b_m)\, x^m + \ldots,$$

получающийся сложением коэффициентов при одинаковых степенях. Роль нулевого элемента играет тогда ряд, все коэффициенты которого равны 0, а противоположным для ряда (12.9)

будет, ясно, ряд

$$-a_0 - a_1x - a_2x^2 - \ldots - a_mx_m - \ldots$$

Произведением рядов (12.9) и (12.10) назовём степенной ряд

$$c_0 + c_1x + c_2x^2 + \ldots + c_mx_m + \ldots$$

коэффициенты которого вычисляются по формуле

$$c_m = a_0b_m + a_1b_{m-1} + \ldots + a_{m-1}b_1 + a_mb_0 = \sum_{i=0}^{m} a_ib_{m-i}. \quad (12.11)$$

Эта формула полностью соответствует обычному правилу умножения многочленов. Такое определение операций превращает множество степенных рядов в коммутативное кольцо с единицей (читатель может без труда это проверить), которое называется кольцом формальных степенных рядов[1] и обозначается $\mathbb{R}[[x]]$. Поскольку всякий многочлен принадлежит этому кольцу, то можно сказать, что кольцо многочленов $\mathbb{R}[x]$ образует подкольцо в $\mathbb{R}[[x]]$. Элементы кольца формальных степенных рядов и используются в качестве производящих функций для решения различных задач перечисления. Более точно, если $a_0, a_1, \ldots$ $\ldots, a_m, \ldots$ — последовательность чисел (дающих решение некоторой комбинаторной задачи), то производящей функцией этой последовательности называется ряд

$$a_0 + a_1x + a_2x^2 + \ldots + a_mx^m + \ldots,$$

принадлежащий $\mathbb{R}[[x]]$. Как обстоит дело с делением степенных рядов — разобраться с этим очень полезно для дальнейшего. Вспомним, что в кольце многочленов обратимые элементы исчерпывались многочленами нулевой степени. В кольце $\mathbb{R}[[x]]$

[1] К рассмотренным операциям можно добавить и другие, также определяемые формально: суперпозиция (подстановка ряда в ряд), дифференцирование, интегрирование. Множество степенных рядов, на котором определены все перечисленные в тексте и в сноске операции, называют обычно алгеброй Коши, по имени французского математика О. Коши(1798–1857), одного из создателей современного математического анализа.

запас обратимых элементов, оказывается, намного богаче. Выясним, какому условию (необходимому и достаточному) должен удовлетворять произвольный ряд (12.9), чтобы для него существовал обратный элемент, также, разумеется, некоторый ряд. Пусть таковым является ряд (12.10), тогда в кольце $\mathbb{R}[[x]]$ должно выполняться равенство

$$(a_0 + a_1x + \ldots + a_mx^m + \ldots)(b_0 + b_1x + \ldots + b_mx^m + \ldots) = 1.$$

Используем правило (12.11) перемножения рядов и приравняем коэффициенты при одинаковых степенях в обеих частях последнего равенства. Это приводит к следующей системе бесконечного числа уравнений:

$$\begin{aligned} a_0b_0 &= 1, \\ a_0b_1 + a_1b_0 &= 0, \\ a_0b_2 + a_1b_1 + a_2b_0 &= 0, \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ a_0b_m + a_1b_{m-1} + \ldots + a_{m-1}b_1 + a_mb_0 &= 0, \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \end{aligned} \tag{12.12}$$

с неизвестными $b_0, b_1, \ldots, b_m, \ldots$.

Первое из этих уравнений имеет решение $b_0 = 1/a_0$ тогда и только тогда, когда $a_0 \neq 0$. Это же условие необходимо и достаточно для разрешимости второго и всех последующих уравнений. Действительно, из второго и третьего уравнений получаем

$$b_1 = -\frac{a_1b_0}{a_0} = \frac{a_1}{a_0^2}, \quad b_2 = -\frac{a_1b_1 + a_2b_0}{a_0} = \frac{a_1^2a_2a_0}{a_0^3},$$

аналогично находятся и все последующие коэффициенты.

Итак, условие $a_0 \neq 0$ является необходимым и достаточным для того, чтобы система (12.12) имела решение $(b_0, b_1, \ldots, b_m, \ldots)$ и притом единственное, следовательно, и для того, чтобы ряд (12.9) являлся обратимым в $\mathbb{R}[[x]]$. Иллюстрируя сказанное, найдём ряд, обратный элементу $1 - x \in \mathbb{R}[[x]]$. В этом случае $a_0 = 1$,

$a_1 = -1$, все прочие коэффициенты равны 0, и система (12.12) принимает вид

$$b_0 = 1, \quad b_1 - b_0 = 0, \quad b_2 - b_1 = 0, \quad \ldots, \quad b_m - b_{m-1} = 0, \quad \ldots,$$

откуда следует, что $b_m = 1$ для всех m. Таким образом, приходим к равенству[1]

$$\frac{1}{1-x} = 1 + x + x^2 + \ldots + x^m + \ldots \tag{12.13}$$

Вернёмся теперь к задаче отыскания числа сочетаний с неограниченными повторениями. Обозначим через $\overline{C}_n^m$ число таких сочетаний из n элементов по m. Здесь n фиксировано, а m пробегает все неотрицательные целые значения. Производящей функцией для чисел $\overline{C}_n^m$ будет, как это уже отмечалось, выражение (12.8). Учитывая равенство (12.13), мы можем написать

$$(1 + x + x^2 + \ldots + x^m + \ldots)^n = \frac{1}{(1-x)^n} = (1-x)^{-n}.$$

В дополнении к этой главе доказывается, что формула бинома Ньютона распространяется и на целые отрицательные показатели, поэтому

$$(1-x)^{-n} = 1 - C_{-n}^1 x + C_{-n}^2 x^2 + \ldots + (-1)^n C_{-n}^m x^m + \ldots, \tag{12.14}$$

где $C_{-n}^m = \dfrac{(-n)(-n-1)\ldots(-n-m+1)}{m!}$.

Таким образом, число сочетаний с неограниченными повторениями из n элементов по m равно

$$\overline{C}_n^m = (-1)^m C_{-n}^m = \frac{(-1)^m(-n)(-n-1)\ldots(-n-m-1)}{m!} = \frac{(n+m-1)\ldots(n+1)n}{m!}.$$

[1] Сравните его с формулой суммы бесконечно убывающей геометрической прогрессии.

Окончательно,

$$\overline{C}_n^m = C_{n+m-1}^m. \tag{12.15}$$

Так, при $n = 2$, $m = 4$ получаем $\overline{C}_2^4 = C_5^4 = 5$ сочетаний с повторениями:

$$a_1a_1a_1a_1,\ a_1a_1a_1a_2,\ a_1a_1a_2a_2,\ a_1a_2a_2a_2,\ a_2a_2a_2a_2.$$

Отметим, что, хотя формулу (12.15) можно доказать более коротким и элементарным путём (см., например, [7]), всё же наш труд не был напрасным. На примере задачи о сочетаниях с повторениями мы продемонстрировали действенность метода производящих функций и общую схему его применения, годную для решения многих других задач.

К их числу относятся задачи, связанные с разбиениями чисел на слагаемые (они имеют различные применения, например, в теории чисел и теории кодирования). Вот пример такой задачи, облечённой в занимательную форму.

> Сколькими способами можно разменять рубль монетами достоинством 5, 10 и 20 коп.[1]?

Обозначив через i, j, k число монет достоинством соответственно 5, 10, 20 коп., участвующих в некотором способе размена, мы можем сформулировать задачу по-иному: сколько различных решений в целых неотрицательных числах имеет уравнение

$$5i + 10j + 20k = 100,$$

или, что то же самое, уравнение

$$i + 2j + 4k = 20.$$

Рассмотрим аналогичное соотношение

$$i + 2j + 4k = l \tag{12.16}$$

[1] На момент написания этой книги все такие монеты имели хождение. — *Прим. ред.*

с произвольной правой частью l, и через a_l обозначим число различных его целых неотрицательных решений. Убедимся, что для последовательности a_l производящей функцией является следующее выражение

$$f(x) = \\ = (1+x+x^2+\ldots)(1+x^2+x^4+\ldots)(1+x^4+x^8+\ldots) = \\ = \sum_{i=0}^{\infty} x^i \sum_{j=0}^{\infty} x^{2j} \sum_{k=0}^{\infty} x^{4k}. \qquad (12.17)$$

В самом деле, произведение i-го, j-го и k-го слагаемых из первого, второго и третьего сомножителей соответственно даёт одночлен $x^{i+2j+4k}$. Поскольку a_l — число решений уравнения (12.16), то ряде, получившемся после перемножения, появится ровно a_l слагаемых вида x^l для каждого l. Поэтому после приведения подобных членов выражение (12.17) примет вид

$$f(x) = \sum_{l=0}^{\infty} a_l x^l,$$

а это и означает, что оно является нужной нам производящей функцией. Вспоминая теперь равенство (12.13), выпишем совершенно ему аналогичное

$$\frac{1}{1-x^t} = 1 + x^t + x^{2t} + \ldots + x^{nt} + \ldots, \qquad (12.18)$$

справедливое для любого натурального t. Используем (12.18) и запишем производящую функцию (12.17) в ином, более удобном виде

$$f(x) = \frac{1}{(1-x)(1-x^2)(1-x^4)}.$$

Ещё несколько тождественных преобразований довершают дело. Умножим числитель и знаменатель на $(1+x)(1+x^2)^2$ и применим

формулу (12.14), заменяя в ней x на x^4. Тогда

$$f(x) = \frac{(1+x)(1+x^2)^2}{(1-x^4)^3} = (1+x)(1+x^2)^2(1-x^4)^{-3} =$$
$$= (1+x)(1+2x^2+x^4)(1+3x^4+6x^8+\ldots$$
$$\ldots+\frac{(m+1)(m+2)}{2}x^{4m}+\ldots) =$$
$$= (1+x+2x^2+2x^3+x^4+x^5)\sum_{m=0}^{\infty}\frac{(m+1)(m+2)}{2}x^{4m}.$$

Теперь уже нетрудно сообразить, чему равны интересующие нас числа a_l. Если $l = 4s$, то

$$a_{4s} = \frac{(s+1)(s+2)}{2} + \frac{s(s+1)}{2} = (s+1)^2.$$

Аналогично, при $l = 4s+1$

$$a_{4s+1} = (s+1)^2.$$

Если же $l = 4s+2$ или $l = 4s+3$, то

$$a_{4s+2} = a_{4s+3} = (s+1)(s+2). \qquad (12.19)$$

В частности, $a_{20} = 36$, это и есть число способов разменять рубль. Одновременно получено решение и многих других задач о размене. К примеру, a_{10} указывает сколькими способами можно разменять полтинник. Согласно формуле (12.19) таких способов имеется 12. Нетрудно их перечислить:

$$50 = 10\cdot 5 = 8\cdot 5+10 = 6\cdot 5+2\cdot 10 = 6\cdot 5+20 =$$
$$= 4\cdot 5+3\cdot 10 = 4\cdot 5+10+20 = 2\cdot 5+4\cdot 10 =$$
$$= 2\cdot 5+2\cdot 10+20 = 2\cdot 5+2\cdot 20 = 5\cdot 10 =$$
$$= 3\cdot 10+20 = 10+2\cdot 20.$$

Подобным же образом решаются и другие задачи на разбиения, правда, вычисление коэффициентов производящей функции для них может потребовать больших усилий.

Достаточно общая формулировка, в которую укладываются эти задачи, выглядит так.

Задано некоторое множество

$$A = \{n_1, n_2, \ldots, n_l, \ldots\}$$

натуральных чисел, не обязательно конечное. Сколькими способами можно разбить заданное число m на сумму слагаемых из множества A (при этом разбиения, отличающиеся лишь порядком слагаемых, считаются одинаковыми)?

По-разному уточняя эту общую формулировку, мы будем получать различные варианты задач на разбиения. Так, если ни на число слагаемых, ни на число повторений одинаковых слагаемых не накладывается никаких ограничений, то производящая функция для числа таких разбиений по аналогии с (12.17) запишется в виде

$$(1 + x^{n_1} + x^{2n_1} + \ldots)(1 + x^{n_2} + x^{2n_2} + \ldots)(1 + x^{n_3} + x^{2n_3} + \ldots)\ldots$$

Если же повторения одинаковых слагаемых недопустимы, то придём к производящей функции иного вида

$$(1 + x^{n_1})(1 + x^{n_2})(1 + x^{n_3})\ldots$$

Завершая на время разговор об алгебре производящих функций, хотим ещё отметить, что оперирование над ними по установленным алгебраическим законам позволяет часто устанавливать неожиданные связи между различными комбинаторными задачами (см., например, дополнение 3).

Перечисление порой осложняется тем, что объекты, формально различные, по смыслу задачи приходится считать одинаковыми, эквивалентными. Иллюстрацией тому может служить известная комбинаторная задача, т. н. «задача об ожерельях»[1]. Формулировка её такова.

Пусть ожерелье состоит из n бусинок нескольких цветов. Спрашивается, сколько существует различных ожерелий, если

[1] Она является наглядной моделью для многих практически важных задач на перечисление.

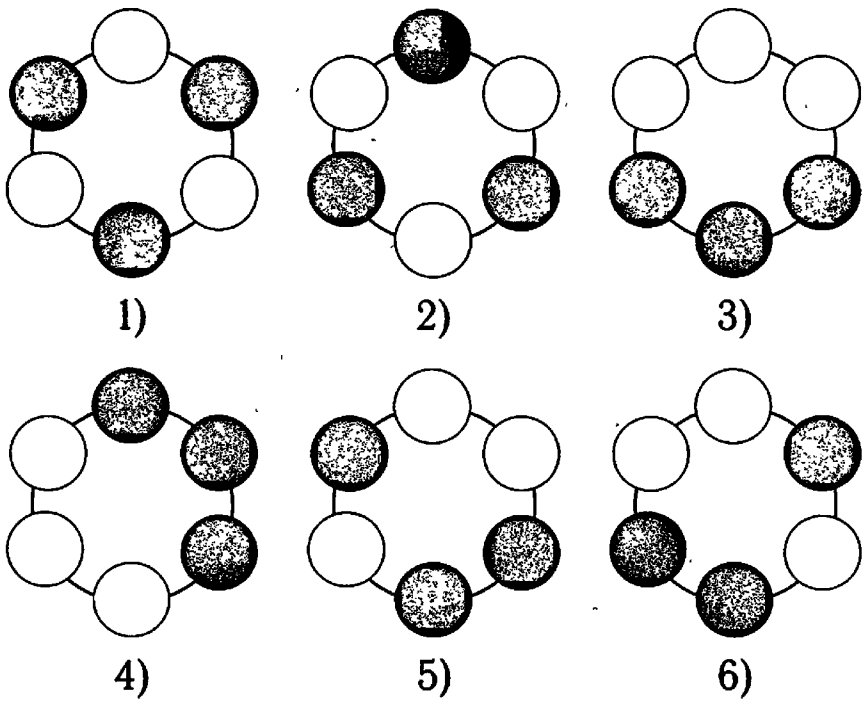


Рис. 44.

задано число бусинок каждого цвета. Рассмотрим, например, все ожерелья, состоящие из трёх белых и трёх чёрных бусинок, некоторые из них изображены на рис. 44.

На первый взгляд может показаться, что общее их число $C_6^3 = 20$, однако многие ожерелья из этого числа по существу одинаковы, отличаются лишь своим положением в пространстве. Так после поворота на некоторый угол первое из ожерелий становится неотличимым от второго, а третье от четвертого. Если считать одинаковыми (эквивалентными) ожерелья, получающиеся одно из другого поворотом, то число различных ожерелий в нашем примере намного меньше 20, и нетрудно убедиться, что их только четыре. Слова «одинаковые», «различные» в этой и других комбинаторных задачах получают совершенно ясное и однозначное истолкование с помощью понятия группы.

Так, если считать, что бусинки расположены в вершинах правильного n-угольника, то эквивалентность ожерелий означает, что одно из них может быть переведено в другое некоторым преобразованием из группы вращений V_n этого n-угольника.

Заметим, что слову «одинаковый» здесь можно было бы придать и чуть иной смысл, допуская не только повороты, но и пе-

реворачивание ожерелий на другую сторону (симметрию относительно оси). В этом случае эквивалентными становятся также пятое и шестое ожерелье нашего рисунка, а число различных ожерелий станет равным трём.

Но эквивалентность ожерелий и в новом смысле может также выражена на групповом языке. Достаточно лишь рассмотреть вместо V_n группу всех самосовмещений правильного n-угольника или иначе группу диэдра D_n. Теперь два ожерелья эквивалентны, если они могут быть переведены одно в другое преобразованием из группы D_n (поворотом или симметрией).

Важно, что подобным образом можно говорить об эквивалентности объектов любого множества, если только задана некоторая группа его преобразований.

В самом деле, пусть M — произвольное множество и G — группа взаимно однозначных преобразований этого множества. Элементы a и b, принадлежащие M, назовём *эквивалентными*[1], если существует преобразование $g \in G$, переводящее элемент a в b: $g(a) = b$. Понятно, что имеется тогда и преобразование (а именно, g^{-1}), осуществляющее обратный переход: $g^{-1}(b) = a$.

Ситуация, с которой мы здесь сталкиваемся, в сущности, не нова. В том или ином виде она уже не раз встречалась на страницах этой книги. Так, в теории алгебраического уравнения рассматривалось множество многочленов от нескольких переменных вместе с группой его преобразований, возникающих при перестановках аргументов (корней изучаемого уравнения). Свойства симметрии или частичной симметрии таких многочленов, столь важные в задаче о разрешимости в радикалах, как раз и определялись тем, насколько велик запас эквивалентных элементов у этих многочленов. Группы преобразований (точнее, движений) плоскости и пространства были в центре внимания главы «Язык симметрии». Эти группы явились удобным средством классификации типов симметрии геометрических фигур, орнаментов, кристаллов. Можно добавить к этому, что способ образования симметричной фигуры (с данной группой симметрии G) из произ-

[1] Точнее, эквивалентными относительно действия группы G на множестве M.

вольного мотива состоял фактически в присоединении к исходному мотиву всех эквивалентных ему относительно группы G элементов.

Возвращаясь к общей ситуации, назовём орбитой элемента $a \in M$ совокупность $\mathrm{Or}(a)$ всех элементов множества M, эквивалентных этому элементу. Оказывается, что любые два элемента b и c, принадлежащие $\mathrm{Or}(a)$, не только эквивалентны a, но и эквивалентны также между собой. Действительно, поскольку $b, c \in \mathrm{Or}(a)$, то существуют такие преобразования $g_1, g_2 \in G$, что $b = g_1(a)$ и $c = g_2(a)$. Тогда преобразование $g_2 g_1^{-1}$ переводит элемент b в c, в самом деле

$$g_2 g_1^{-1}(b) = g_2(a) = c$$

и, значит, b и c эквивалентны. Отсюда, кстати сказать, следует, что $\mathrm{Or}(b) = \mathrm{Or}(c) = \mathrm{Or}(a)$, или иначе, орбиты любой пары эквивалентных элементов совпадают. С другой стороны, если элементы a_1 и a_2 неэквивалентны, то их орбиты вообще не имеют общих элементов. Предполагая противное, т. е. что некоторый элемент $a \in \mathrm{Or}(a_1) \cap \mathrm{Or}(a_2)$, получаем $a = g_1(a_1)$ и $a = g_2(a_2)$ для некоторых преобразований $g_1, g_2 \in G$. Но тогда $a_2 = g_2 g_1^{-1}(a_1)$, что противоречит предположению о неэквивалентности a_1 и a_2.

Резюмируя сказанное, приходим к следующим выводам:

> Множество M под действием группы G распадается на непересекающиеся классы эквивалентных между собой элементов. Каждый такой класс представляет из себя орбиту для любого из входящих в него эквивалентных элементов. Неэквивалентные элементы принадлежат различным орбитам.

На рис. 45 приведены две из четырёх орбит множества ожерелий, состоящих из 3 белых и 3 чёрных бусинок, относительно группы V_6.

Мы подошли к тому, что возникающая в различных задачах комбинаторики проблема перечисления неэквивалентных объектов сводится к отысканию числа орбит той или иной группы, действующей на множестве всех объектов. Полезным инструментом

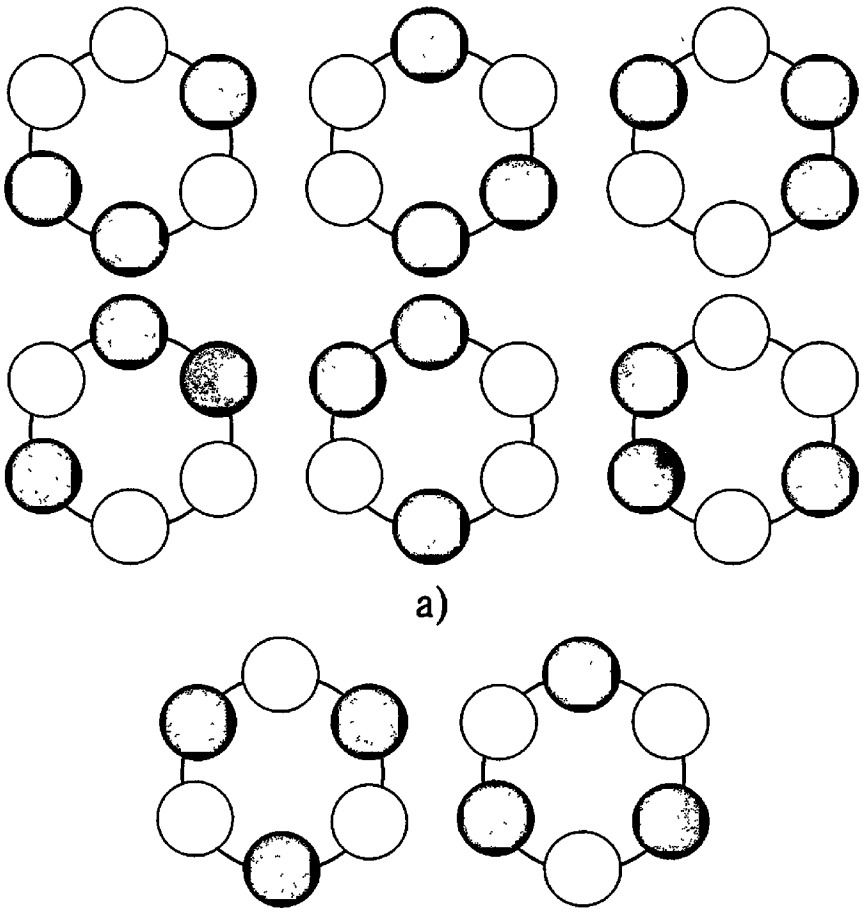


Рис. 45.

для этого является утверждение, известное в теории групп как лемма Бернсайда. Пусть M — конечное множество и G — группа его подстановок. Число орбит относительно действия группы G вычисляется тогда по следующей формуле:

$$t(G) = \frac{1}{|G|} \sum_{g \in G} \chi(g), \tag{12.20}$$

где $\chi(g)$ — число элементов множества M, неподвижных относительно подстановки g (знак $\sum_{g \in G}$ означает, что сумма распространяется на все элементы группы G). Хотя доказательство формулы (12.20) для числа орбит впервые было опубликовано в 1911 году, известна она была, по-видимому, гораздо раньше, а основные соображения, которые к ней приводят, по существу восходят к Лагранжу. Трудности с перечислением орбит происходят от то-

го, что число элементов в различных орбитах, как мы могли заметить, не обязательно одинаково. Сначала разберёмся, от чего зависит это число, называемое числом орбиты. Чтобы ответить, возьмём произвольный элемент $a \in M$ и выпишем элементы его орбиты $\mathrm{Or}(a)$. Они, очевидно, следующие:

$$a = g_0(a),\ g_1(a),\ g_2(a),\ \ldots,\ g_{n-1}(a). \tag{12.21}$$

Здесь $g_0, g_1, g_2, \ldots, g_{n-1}$ — всевозможные подстановки из группы G и g_0 — тождественная подстановка. Однако в ряду (12.21) возможны повторения одинаковых элементов, и связано это как раз с тем, что некоторые из подстановок могут оставлять элемент a неподвижным. Ведь если, к примеру, $g_i(a) = g_j(a)$, $i \neq j$, то $g_j^{-1}g_i(a) = a$, а это и означает, что подстановка $g_j^{-1}g_i$ оставляет элемент a неподвижным. Очевидно, что длина орбиты тем меньше, чем больше таких подстановок. Нельзя ли описать эту связь поточнее? Оказывается, можно, и цель достигается теми же рассуждениями, которыми пользовался ещё Лагранж (читатель легко это поймёт, если вернётся к содержанию главы 3). В самом деле, рассмотрим множество всех подстановок G_a, для которых элемент a является неподвижным. Указанное множество, называемое *стабилизатором* элемента a, образует подгруппу в группе G (проверьте). Рассмотрим смежные классы $gG_a = \{\, gh \mid h \in G_a \,\}$ по этой подгруппе и заметим, что все подстановки, принадлежащие одному и тому же смежному классу gG_a, одинаково преобразуют элемент a, действуя на него так же, как подстановка g:

$$(gh)(a) = g(h(a)) = g(a).$$

Понятно, что верно и обратное: если $g_1(a) = g_2(a)$, то подстановки принадлежат одному смежному классу по стабилизатору. Вывод, к которому мы приходим, таков: число различных элементов в (12.21) совпадает с числом различных смежных классов, иными словами, длина орбиты $|\mathrm{Or}(a)|$ равна индексу стабилизатора элемента a, т. е.

$$|\mathrm{Or}(a)| = |G : G_a| = |G|/|G_a|. \tag{12.22}$$

Теперь уже нетрудно доказать саму лемму Бернсайда. Рассмотрим для этого все пары вида (g, a), где $g \in G$ и $a \in M$ — элемент, неподвижный относительно g. Наши рассуждения будут основываться на подсчёте числа таких «стабильных» пар двумя различными способами. С одной стороны, их общее число, очевидно равно

$$\sum_{g \in G} \chi(g). \tag{12.23}$$

Но с другой стороны, для каждого $a \in M$ число подстановок со свойством $g(a) = a$ совпадает с порядком стабилизатора G_a, т. е. равно, согласно (12.22),

$$|G_a| = |G|/|\operatorname{Or}(a)|.$$

Таким же будет это число для всех элементов $b \in \operatorname{Or}(a)$, так как в этом случае $\operatorname{Or}(b) = \operatorname{Or}(a)$. Тогда количество стабильных пар (g, b) для всех элементов одной орбиты $\operatorname{Or}(a)$ окажется равным

$$|\operatorname{Or}(a)| \cdot \frac{|G|}{|\operatorname{Or}(a)|} = |G|$$

и следовательно, одним и тем же для любой из орбит. Умножив полученное число на число орбит, получим общее число стабильных пар

$$t(G)|G|. \tag{12.24}$$

Наконец, сравнивая (12.23) и (12.24) приходим к утверждению леммы Бернсайда.

Мы можем теперь проиллюстрировать применение доказанной формулы для решения некоторых комбинаторных задач, пока не очень сложных. Взять хотя бы пример с чёрно-белыми ожерельями, изображёнными на рис. 44 и 45. Очевидно, что для тождественного преобразования v_0 все 20 ожерелий являются неподвижными элементами: $\chi(v_0) = 20$, а для поворота v_1 на угол $\pi/3$ неподвижные ожерелья отсутствуют совсем: $\chi(v_1) = 0$. Нетрудно сообразить также, что, например, $\chi(v_2) = 2$, таким же будет это число и для любой симметрии σ_i. Всю информацию о неподвижных элементах сведем в таблицу 10.

Таблица 10.

элементы g_i	v_0	v_1	v_2	v_3	v_4	v_5	σ_0	σ_1	σ_2	σ_3	σ_4	σ_5
$\chi(g_i)$	20	0	2	0	2	0	2	2	2	2	2	2

Тогда по формуле Бернсайда число орбит относительно группы V_6 действительно равно

$$t(V_6) = \frac{1}{6}(20 + 2 \cdot 2) = 4,$$

а относительно D_6 —

$$t(D_6) = \frac{1}{12}(20 + 2 \cdot 8) = 3.$$

Ещё один пример, чуть посложней, — задача о раскраске куба. Сколькими неэквивалентными способами можно раскрасить грани куба, если каждая грань может быть раскрашена одним из двух цветов, скажем, зелёным и жёлтым? При этом два способа окраски считаются эквивалентными, если от одного можно перейти к другому путём подходящего поворота в пространстве. Задача, как мы теперь уже понимаем, сводится к подсчёту числа орбит группы вращений куба на множестве всевозможных способов окрасок этого куба (их общее число очевидно $2^6 = 64$). Можно понять, что группа вращений куба состоит из 24 преобразований, которые можно разбить на 5 типов (рис. 46):

1. Тождественное вращение.
2. Три поворота на 180° вокруг прямых, соединяющих центры противоположных граней.
3. Шесть поворотов на 90° вокруг тех же прямых.
4. Шесть поворотов на 180° вокруг прямых, соединяющих середины противоположных рёбер.

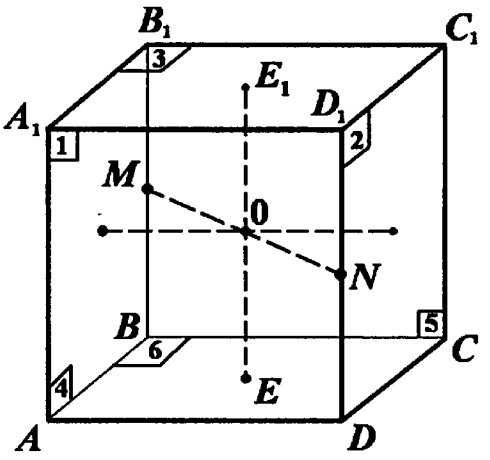


Рис. 46.

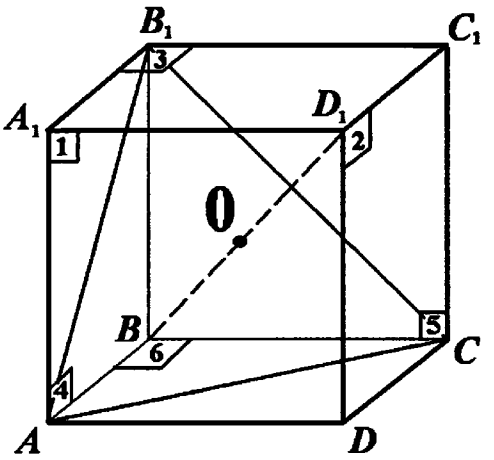


Рис. 47.

5. Восемь поворотов на 120° вокруг прямых, соединяющих противоположные вершины (рис. 47).

Занумеруем грани куба так, как показано на рис. 46 (номера проставлены в углах граней[1]). С каждым из перечисленных преобразований связана, очевидно, некоторая подстановка на множестве граней, или, что то же самое, на множестве их номеров. Выполняя эти преобразования, мы и переходим от одних способов окраски куба к другим, им эквивалентным. При этом, конечно, может случиться так, что в результате поворота окрашенный куб переходит в куб, неотличимый от исходного. Например, если грани с номерами 3, 6 окрашены в зелёный цвет, а остальные — в жёлтый, то любое вращение вокруг оси EE_1 не меняет такой окраски, т. е. она является неподвижным элементом относительно указанных вращений.

Понятно далее, что преобразования, принадлежащие к одному типу, имеют одинаковое число неподвижных элементов и потому достаточно посчитать это число для одного из преобразований каждого типа.

Сделаем ещё одно полезное наблюдение. Пусть соответствующая преобразованию подстановка разложена в произведение независимых циклов (см. главу 2). Ясно, что раскраска куба яв-

[1] Заметим для ясности, что эти номера относятся не к граням, как таковым, а к их положению в пространстве, т. е. приведённая нумерация означает, что передняя грань всегда имеет номер 1, правая боковая — номер 2 и т. д.

ляется неподвижным элементом относительно этой подстановки тогда и только тогда, когда грани, номера которых входят в один и тот же цикл, окрашены одинаково.

Так, повороту на 180° вокруг оси EE_1 соответствует подстановка

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 5 & 4 & 3 & 2 & 1 & 6 \end{pmatrix} = (1,5)(2,4),$$

и для неподвижных относительно неё раскрасок одинаковый цвет должны иметь грани с номерами 1 и 5, точно так же, как и грани с номерами 2 и 4. Грани с номерами 3 и 6 могут быть раскрашены произвольно. Число таких раскрасок равно $2^4 = 16$.

Подстановка, отвечающая повороту против часовой стрелки на 90° вокруг той же оси, сама является циклом (1, 2, 5, 4) длины 4, так что для неподвижной раскраски все четыре грани с этими номерами должны быть окрашены одинаково, и таких раскрасок $2^3 = 8$. В качестве преобразования четвёртого типа рассмотрим поворот вокруг прямой MN. Соответствующая подстановка имеет вид (1, 2)(3, 6)(4, 5), и неподвижных для неё элементов также 8. Для поворота вокруг оси BD_1 получаем подстановку (1, 3, 2)(4, 5, 6) с четырьмя неподвижными раскрасками (рис. 47).

Учитывая, что для тождественного преобразования все 64 раскраски неподвижные элементы, находим по формуле (12.20) искомое число орбит

$$t(G) = \frac{64 + 3 \cdot 16 + 6 \cdot 8 + 6 \cdot 8 + 8 \cdot 4}{24} = 10.$$

Это и есть число различных (точнее, неэквивалентных) раскрасок куба.

Мы рассказали о двух важных способах перечисления: методе производящих функций и групповом методе, основанном на лемме Бернсайда. Эти два метода гармонично соединяются в теории перечисления, созданной известным математиком Д. Пойа[1].

[1] Читателю этот математик, быть может, известен по его замечательным книгам, переведенным на русский язык, таким, например, как [20], [21].

Основная теорема этой теории связывает перечисляющую производящую функцию для числа орбит с более простыми производящими функциями, одна из которых — так называемый цикловой индекс группы подстановок. Краткое знакомство с теорией Пойа мы и начнём с понятия циклового индекса.

Пусть, как и выше, G — группа подстановок конечного множества M из m элементов. Пусть g — некоторая подстановка и $g = \pi_1\pi_2\ldots\pi_r$ — её разложение на независимые циклы. Если в указанном разложении имеется b_1 циклов длины 1, b_2 циклов длины 2 и т. д., b_m циклов длины m, то будем говорить, что подстановка имеет тип $(b_1, b_2, \ldots, b_m)$. Для каждой такой подстановки образуем одночлен $x_1^{b_1}x_2^{b_2}\ldots x_m^{b_m}$ от m переменных. Их сумму, взятую по всем подстановкам из группы G, разделим на число элементов в группе. Полученный многочлен

$$P_G(x_1, x_2, \ldots, x_m) = \frac{1}{|G|}\sum_{g\in G} x_1^{b_1}x_2^{b_2}\ldots x_m^{b_m}$$

и называют *цикловым индексом* группы G.

Посчитаем цикловой индекс для некоторых из рассмотренных выше групп.

В таблицу 11 сведены типы подстановок из группы V_6 вращений правильного шестиугольника.

Таблица 11.

v_0	v_1	v_2
(6, 0, 0, 0, 0, 0)	(0, 0, 0, 0, 0, 1)	(0, 0, 2, 0, 0, 0)
v_3	v_4	v_5
(0, 3, 0, 0, 0, 0)	(0, 0, 2, 0, 0, 0)	(0, 0, 0, 0, 0, 1)

Из таблицы следует, что цикловой индекс этой группы имеет вид

$$P_G = \frac{1}{6}\left(x_1^6 + x_2^3 + 2x_3^2 + 2x_6\right). \qquad (12.25)$$

Каждое из 24 преобразований группы G вращений куба индуцирует подстановку на множестве его граней. Пять категорий, на которые выше были разбиты вращения, дают соответственно подстановки типов $(6,0,0,0,0,0)$, $(2,2,0,0,0,0)$, $(2,0,0,1,0,0)$, $(0,3,0,0,0,0)$, $(0,0,2,0,0,0)$. И значит, цикловой индекс группы

$$P_G = \frac{1}{6}(x_1^6 + 3x_1^2x_2^2 + 6x_1^2x_4 + 6x_2^3 + 8x_3^2). \qquad (12.26)$$

На примере задачи о раскраске граней куба мы могли понять, что цикловая структура подстановок довольно тесно связана с перечислением орбит. Только что введённое понятие циклового индекса позволяет выразить эту связь в более явной и обозримой форме. Этой же цели служат вводимые далее функции и их веса.

Функции, которые нам будут нужны, могут иметь достаточно произвольные области определения и значений (обозначим их соответственно D и E). В задачах перечисления эти множества всегда являются конечными. Такие функции $f\colon D \to E$ оказываются удачным средством для описания и представления многих перечисляемых объектов. Скажем, в задаче о раскраске граней куба в качестве D можно взять множество его граней или их номеров, а в качестве E — множество используемых цветов. Понятно, что тогда каждому способу раскраски граней отвечает некоторая функция $f\colon D \to E$ и обратно. Так функции, принимающей значения

$$f(1) = f(2) = f(3) = f(4) = \text{«ж»}, \quad f(5) = f(6) = \text{«з»}, \qquad (12.27)$$

соответствует способ окраски, при котором первые четыре грани окрашены в жёлтый цвет, а остальные — в зелёный.

Если на множестве D задана некоторая группа подстановок, то можно пойти и дальше, определив отношение эквивалентности на множестве рассматриваемых функций. Конечно, это нужно сделать так, чтобы эквивалентность функций всегда означала эквивалентность соответствующих им перечисляемых объектов. Определение, которое мы приведём, вполне удовлетворяет указанному требованию.

Две функции f_1 и f_2 будем называть *эквивалентными*, если существует подстановка $g \in G$ такая, что

$$f_1(x) = f_2(g(x)) \quad \text{для всякого } x \in D. \tag{12.28}$$

Понятно, что тогда

$$f_2(x) = f_1(g^{-1}(x)) \quad \text{для всякого } x \in D.$$

Хорошую иллюстрацию данному определению вновь даёт задача о раскраске граней.

В самом деле, пусть g — подстановка граней, соответствующая некоторому вращению v куба и f_1 и f_2 — две эквивалентные функции, удовлетворяющие соотношению (12.28). Как связаны тогда те два способа его окраски, которые отвечают этим эквивалентным функциям? Ответ даёт равенство (12.27). Оно показывает, что цвет грани x при первой окраске таков же, как и цвет грани $g(x)$ при окрашивании вторым способом. Но это значит, что вторую окраску мы получим из первой, выполняя заданное вращение v.

Обратим внимание читателя, на то, что эквивалентность функций не является чем-то существенно новым по сравнению с уже рассмотренным ранее понятием эквивалентности (см. с. 228). Дело в том, что со всякой подстановкой g множества D можно связать преобразование на множестве функций, имеющих D своей областью определения. Сделаем это так: произвольной функции f сопоставим функцию φ такую, что

$$\varphi(x) = f(g(x)), \quad x \in D.$$

После этого группу G можно рассматривать также как группу преобразований (и притом взаимно однозначных) множества M всех функций из D в E. Ясно тогда, что эквивалентность функций относительно группы G, действующей описанным способом на множестве M, это то же самое, что и эквивалентность, определённая равенством (12.28).

Раз так, то множество функций распределяется на орбиты, в каждой из которых собраны эквивалентные между собой функ-

ции. Теперь ясно, что решение многих комбинаторных задач сводится к подсчёту числа орбит, на которые некоторое множество функций разбивается группой G.

Понятие веса, к которому мы переходим, позволяет придать схеме перечисления Пойа необычайную гибкость. Веса приписывают каждому элементу области значений E, и они могут быть числами или переменными, или вообще, элементами некоторого коммутативного кольца K. Следовательно, можно образовывать суммы и произведения весов, и эти операции удовлетворяют обычным законам ассоциативности, коммутативности, дистрибутивности. Вес, приписываемый элементу $r \in E$, будем обозначать $w(r)$.

Можно определить тогда и вес произвольной функции $f\colon D \to E$. Его полагают равным произведению весов значений этой функции, принимаемых для всевозможных элементов d из области определения D,

$$W(f) = \prod_{d \in D} w(f(d)). \tag{12.29}$$

В задаче о раскраске куба припишем жёлтому и зелёному цветам веса, равные соответственно переменным x и y из кольца многочленов от двух переменных. Очевидно, что вес функции, принимающей значения (12.27), окажется равным

$$W(f) = x^4 y^2,$$

а показатели степеней, с которыми входят в него переменные, совпадут с числом граней каждого из двух цветов.

Заметим, что как в примере с кубом, так и в общей ситуации, веса эквивалентных функций совпадают. Всё дело в том (убедитесь самостоятельно), что замена функций f на ей эквивалентную приводит лишь к перестановке сомножителей в произведении (12.29). Если F — произвольный класс попарно эквивалентных функций (орбита), то, пользуясь отмеченным обстоятельством, можно определить вес $W(F)$ каждого такого класса. Положим его равным весу $W(f)$ любой из функций $f \in F$. Роль

перечисляющей производящей функции для орбит, расклассифицированных по весам, будет играть тогда сумма

$$\sum_F W(F), \tag{12.30}$$

взятая по всем орбитам. Действительно, если в указанной сумме сгруппировать слагаемые одинакового веса $W(F) = w \in K$, то (12.30) примет вид $\sum_F W(F) = \sum_{w_i \in K} m_i w_i$, где m_i — число классов веса w_i. Числа m_i и дают искомое решение той или иной задачи на перечисление. Как уже отмечалось, это решение зависит от циклового индекса заданной группы подстановок, однако оно не вполне им определяется. Недостающую информацию несёт в себе перечисляющая производящая функция элементов множества E, также расклассифицированных по весам

$$\sum_{e \in E} w(e) = \sum_{w_i \in K} n_i w_i \tag{12.31}$$

(n_i — число элементов $e \in E$, имеющих вес w_i). В примере с раскраской граней куба множество E состоит из двух элементов «ж» и «з», которым приписаны веса x и y. Функция (12.31) в этом случае имеет простой вид:

$$\sum_{e \in E} w(e) = x + y. \tag{12.32}$$

Мы можем теперь сформулировать основную теорему теории перечисления Пойа:

Производящая функция $\sum_F W(F)$ *для числа орбит различных весов равна*

$$\sum_F W(F) = P_G\left(\sum_{e \in E} w(e), \sum_{e \in E} w^2(e), \ldots, \sum_{e \in E} w^n(e)\right),$$

где P_G *— цикловой индекс группы подстановок множества* D, n *— число его элементов.*

Доказательство мы не приводим, скажем только, что центральным его пунктом является использование известной читателю леммы Бернсайда. Вооружившись теоремой Пойа, вернёмся к уже рассмотренным выше задачам.

Сначала раскраска граней. Цикловой индекс в этом случае даётся формулой (12.25), а производящая функция множества красок — формулой (12.32). Согласно теореме имеем тогда:

$$\sum_F W(F) = \frac{1}{24}\left[(x+y)^6 + 3(x+y)^2(x^2+y^2)^2 + \right.$$
$$\left. + 6(x+y)^2(x^4+y^4) + 6(x^2+y^2) + 8(x^3+y^3)^2\right]. \quad (12.33)$$

После раскрытия скобок и приведения подобных членов (мы этого делать не будем) сумма приняла бы вид:

$$\sum_{i=1}^{6}\sum_{j=1}^{6} m_{ij}x^i y^j. \quad (12.34)$$

Каждый из коэффициентов m_{ij} даёт тогда число классов, имеющих все $x^i y^j$, другими словами, это есть число неэквивалентных раскрасок куба, при которых i граней окрашены в жёлтый цвет, а j — в зелёный. Так, коэффициент при x^4y^2 равен

$$\frac{1}{24}(15 + 3\cdot 3 + 0 + 6\cdot 3 + 0) = 2,$$

т. е. имеется две различные окраски с 4-мя жёлтыми и 2-мя зелёными гранями (рис. 48).

Ясно, что число всех неэквивалентных окрасок равно

$$\sum_i \sum_j m_{ij}.$$

Чтобы найти это число, достаточно в формуле (12.34), и следовательно, в (12.33) положить веса x и y равными 1. Тогда получим

$$\frac{1}{24}(2^6 + 3\cdot 2^2\cdot 2^2 + 6\cdot 2^2\cdot 2 + 6\cdot 2^3 + 8\cdot 2^2) = 10,$$

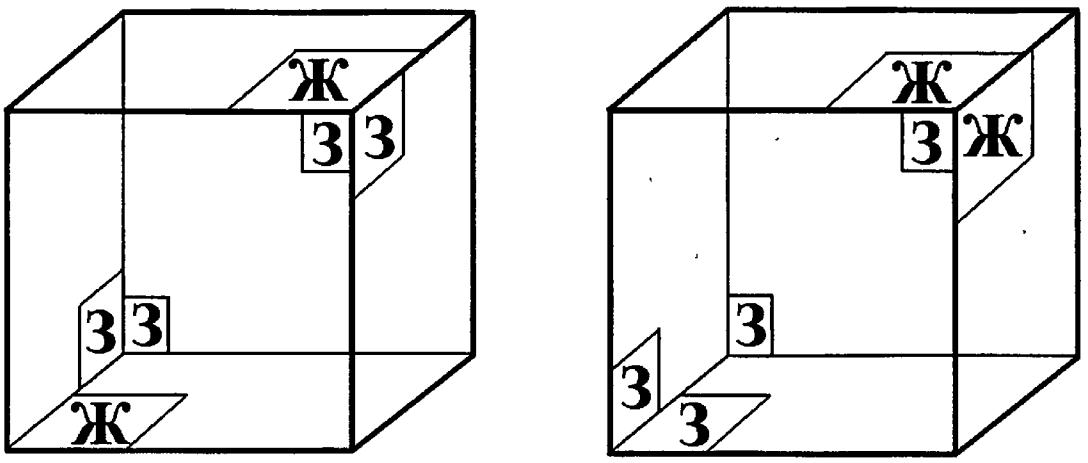


Рис. 48.

что, естественно, совпадает с ранее уже найденным числом $t(G)$ на с. 235.

Этим же способом решается и задача об ожерельях, только теперь необходимо знать цикловые индексы групп V_n и D_n. Найти их несложно.

Пусть v_k — произвольное вращение из группы V_n. Если наибольший общий делитель $(k,n) = d$, то порядок $|v_k| = n/d = m$, а элементу v_k отвечает подстановка на множестве вершин n-угольника, разлагающаяся в произведение d циклов длины m. Все элементы порядка m, как известно, содержатся в подгруппе $\langle v_k \rangle$, а их число равно $\varphi(m)$, где φ — знакомая нам[1] функция Эйлера. Итак, каждой подстановке в этом случае отвечает в цикловом индексе одночлен вида $x_m^d = x_m^{n/m}$, а число таких одночленов для каждого m совпадает с $\varphi(m)$. Следовательно, цикловой индекс группы V_n даётся формулой:

$$P_{V_n}(x_1, x_2, \ldots, x_n) = \frac{1}{n} \sum_{m|n} \varphi(m)\, x_m^{n/m},$$

где запись $m \mid n$ означает, что суммирование распространяется на все различные делители числа n. При $n = 6$ придём к ранее полученной формуле (12.25).

При вычислении циклового индекса группы диэдра D_n (формула приводится в дополнении) нужно учесть, что помимо n

[1] Её определение было дано в дополнении 5 к главе 4 (с. 84).

вращений в D_n содержится такое же количество зеркальных симметрий. Всем симметриям отвечают подстановки одинаковой цикловой структуры и в разложении их на независимые циклы встречаются лишь транспозиции и циклы длины 1. Если при составлении ожерелья используются бусинки s различных цветов, то каждому из цветов (они и образуют в данном случае множество E) сопоставим переменную y_i $(i = 1, 2, \ldots, s)$ — вес этого цвета. Производящая функция множества E, очевидно, равна

$$y_1 + y_2 + \ldots + y_s = \sum_{i=1}^{s} y_i.$$

Весом каждого из ожерелий является тогда одночлен

$$y_1^{r_1} y_2^{r_2} y_3^{r_3} \ldots y_s^{r_s},$$

где r_i — число бусинок i-го цвета, содержащихся в нём.

Согласно теореме Пойа производящая функция для числа неэквивалентных (относительно V_n) ожерелий того или иного веса выглядит так:

$$\frac{1}{n} \sum_{m|n} \varphi(m)(y_1^m + y_2^m + \ldots + y_s^m)^{n/m}. \qquad (12.35)$$

Число $t(V_n)$ всех неэквивалентных ожерелий (как и в задаче с кубом) получается из формулы (12.35) при $y_1 = y_2 = \ldots = y_s = 1$, т. е.

$$t(V_n) = \frac{1}{n} \sum_{m|n} \varphi(m) s^{n/m}.$$

Одна из практических задач (ею занимался ещё А. Кэли), которая может быть с успехом решена методами Пойа, относится к области химии. Давно известно, что разные по составу химические соединения могут сильно отличаться своими химическими свойствами. Это явление, получившее название *изомерия*, объясняется тем, что атомы углерода способны по разному соединяться между собой и с другими атомами. Простейшим примером, на

котором это можно понять, является бутан, имеющий химическую формулу C_4H_{10} и стоящий четвёртым в ряду так называемых парафинов, органических соединений с молекулярной формулой C_nH_{2n+2}. На рис. 49 представлены структурные формулы двух изомеров бутана, первая из них соответствует нормальному бутану, вторая — изобутану. Было замечено, что с ростом n число изомеров парафина C_nH_{2n+2} быстро растёт. Если бутан имеет 2 изомера, а пентан — три, то у гептана их 9, а у декана — 159. Как теоретически определить число возможных изомеров любого парафина C_nH_{2n+2}, как тоже самое сделать для других химических соединений? Эти вопросы очень интересовали химиков. А ответы были даны математиками.

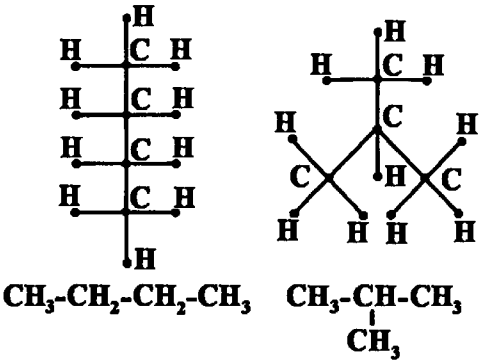


Рис. 49.

С точки зрения математики структурные схемы молекул органических веществ представляют из себя графы. так называют множество точек (вершин) на плоскости, соединённых линиями (рёбрами) по определённому правилу[1]. В графах (структурных схемах) химических соединений вершины — это атомы, рёбра — валентные связи между ними.

В случае парафинов изомерия полностью определяется способом соединения между собой атомов углерода. При $n = 4$ различных способов (т. е. различных графов) имеется ровно два, а при $n = 5$ — три, таково же и число изомеров (рис. 50, 51).

Мы должны уточнить здесь, когда графы считаются различными, а когда — одинаковыми. Два графа будем считать одинаковыми (*изоморфными*), если между вершинами можно установить взаимно однозначное соответствие, при котором каждой паре вершин одного графа, соединённых ребром, соответствует пара вершин другого, также соединённых ребром. Иначе говоря,

[1] Графы используются как наглядная геометрическая модель очень многих (отнюдь не только химических) задач возникающих на практике, а теория графов зачастую помогает в их решении.

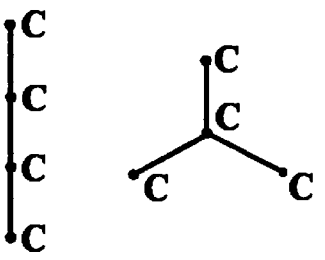


Рис. 50.

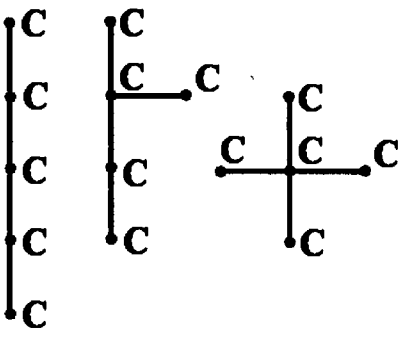


Рис. 51.

при рассмотрении графа не имеет значения, как он изображён на плоскости, существенно лишь то, какие вершины оказываются соединёнными, а какие нет. Графы на рис. 52 в согласии с нашей точкой зрения одинаковы, а все графы, изображённые выше, попарно различны. Для графов, изображающих соединение атомов углерода в любом парафине, в частности и для тех, что представлены на рисунках, характерно то, что они не содержат замкнутых контуров (циклов). Такие графы называются *деревьями*.

Задача перечисления изомеров с химической формулой C_nH_{2n+2} сводится, таким образом, к перечислению различных деревьев с n вершинами. Правда, речь идёт только о таких деревьях (назовём их C-деревьями), к каждой вершине которых примыкает не более четырёх рёбер — валентных связей, ведь валентность атома углерода равна 4. К указанной задаче близка другая — перечислить все изомеры, имеющие химическую формулу $C_nH_{2n+1}OH$. Такие соединения, называемые спиртами, получаются из парафинов C_nH_{2n+2} путём замещения одного из атомов водорода одновалентным радикалом OH. И эта задача легко переводится на язык теории графов. В C-дереве парафина особо выделим ту вершину (атом C), которая соединена ребром (валентной связью) с замещаемым радикалом атомом водорода. На рис. 53 выделенная вершина обведена кружком.

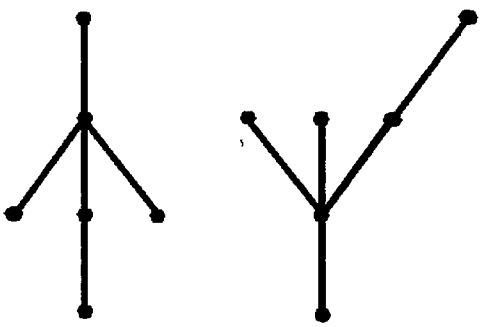


Рис. 52.

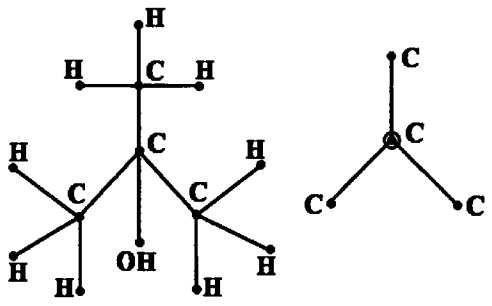


Рис. 53.

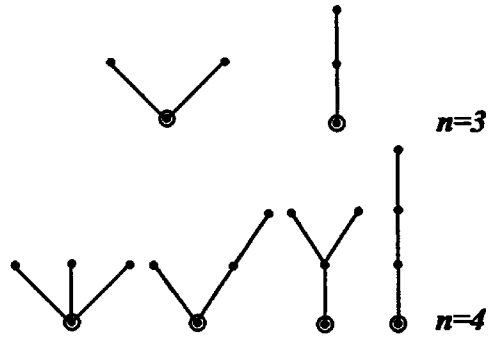


Рис. 54.

Выделенную вершину дерева называют *корнем*, а дерево, в котором имеется корень, называется *корневым*. В нашем случае корневыми вершинами могут, понятно, служить лишь те вершины, к которым примыкает не более трёх рёбер. С-деревья с корневыми вершинами указанного типа для краткости назовём корневыми С-деревьями. Рис. 54 показывает, что при $n = 3$ имеется 2 различных корневых С-дерева, а при $n = 4$ их 4. Смысл слова «различный» в этом случае предлагаем уточнить самим читателям.

Теперь мы можем сказать, что отыскание числа изомеров спирта $C_nH_{2n+1}OH$ сводится к нахождению числа различных корневых С-деревьев с n вершинами. Обе названные задачи (перечисление парафинов и перечисление спиртов) и целый ряд других задач были решены Пойа с помощью изобретённого им метода. Оказалось, что проще начать с решения второй из них — перечисления корневых С-деревьев.

С этой целью рассмотрим производящую функцию

$$r(x) = \sum_{n=0}^{\infty} r_n x^n, \tag{12.36}$$

в которой r_n совпадает с числом различных корневых С-деревьев с n вершинами, а вес каждого такого дерева равен x^n. В случае $n = 0$, который соответствует дереву, вовсе не имеющему вершин, — «пустому» дереву, полагаем $r_0 = 1$. Несколько следующих коэффициентов находятся без труда, например,

$$r_1 = r_2 = 1, \quad r_3 = 2.$$

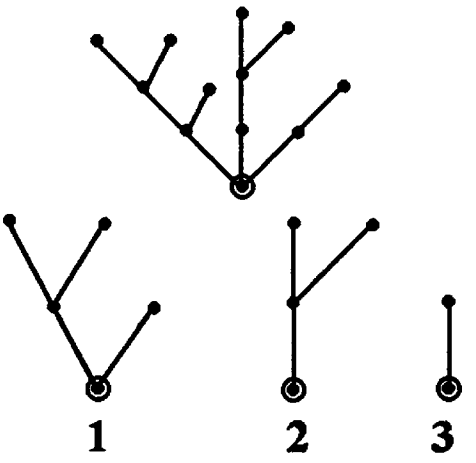


Рис. 55.

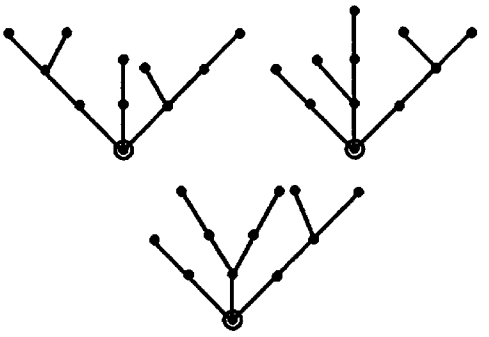


Рис. 56.

Далее рассуждаем так. Можно считать, что из корня любого непустого С-дерева исходят три других С-дерева, называемых главными ветвями (рис. 55) не исключено, что какие-то из них являются пустыми. Так будет, если к корню примыкает менее трёх рёбер. Главные ветви вновь можно считать корневыми С-деревьями, если за корень принять ту вершину ветви, которая является смежной с корнем исходного дерева.

Почти очевидно, что исходное дерево вполне определяется строением своих главных ветвей. Более точно, два корневых С-дерева одинаковы (изоморфны) тогда и только тогда, когда существует такое соответствие между главными ветвями деревьев, при котором каждая ветвь одного дерева переходит в изоморфную ей ветвь другого. На рис. 56 первые два дерева изоморфны, второе и третье различны.

Если ветви занумеровать числами 1, 2, 3, то соответствию, устанавливающему изоморфизм, отвечает подстановка из группы S_3. Перечисление корневых С-деревьев равносильно, таким образом, перечислению неэквивалентных относительно группы S_3 троек главных ветвей. Может показаться, что исходную задачу мы заменили более сложной. На самом же деле эта замена позволяет получить алгебраическое уравнение, которому обязана удовлетворять производящая функция $r(x)$. Обратимся для этого к схеме Пойа, взяв в качестве множества E множество всех корневых С деревьев, перечисляющей функцией которого и является (12.36). Весом каждого его элемента — дерева — будет при

этом x^n, где n — число вершин. Положив далее $D = \{1, 2, 3\}$, мы заметим, что каждая функция,

$$f\colon D \to E$$

отвечает в этом случае тройке корневых деревьев (главных ветвей), некоторым образом занумерованных. Эквивалентность функций относительно группы S_3, понятно, равносильна перенумерации главных ветвей. Значит, эквивалентным функциям отвечают эквивалентные тройки. Если $P_{S_3}(x_1, x_2, x_3)$ — цикловой индекс S_3, то по теореме Пойа производящая функция множества неэквивалентных троек равна

$$P_{S_3}\left(\sum_{n=0}^{\infty} r_n x^n, \sum_{n=0}^{\infty} r_n (x^n)^2, \sum_{n=0}^{\infty} r_n (x^n)^3\right) = \\ = P_{S_3}\left(r(x), r(x^2), r(x^3)\right).$$

Коэффициент при x^k этого перечисляющего ряда должен дать число неэквивалентных троек веса x^k, иначе таких, что суммарное число их вершин равно k (вспомните определение веса функции $f\colon D \to E$). Но их столько же, сколько различных корневых C-деревьев с $k+1$ вершиной: это потому, что суммарное число вершин главных ветвей на единицу меньше числа вершин исходного C-дерева. Тогда перечисляющим рядом для различных корневых C-деревьев с $n \geqslant 1$ вершинами должно служить выражение

$$x\, P_{S_3}\left(r(x), r(x^2), r(x^3)\right).$$

Чтобы учесть ещё пустое дерево, нужно только к полученному добавить 1. Так мы приходим к соотношению

$$r(x) = 1 + x\, P_{S_3}\left(r(x), r(x^2), r(x^3)\right). \qquad (12.37)$$

Остается лишь простая техническая деталь — выписать цикловой индекс группы S_3 и использовать его в полученном равенстве. Так как

$$P_{S_3}(x_1, x_2, x_3) = \frac{1}{6}(x_1^3 + 3x_1x_2 + 2x_3),$$

то соотношение (12.37) даёт

$$r(x) = 1 + \frac{1}{6}\, x \left(r(x)^3 + 3r(x)r(x^2) + 2r(x^3) \right). \qquad (12.38)$$

Полученное алгебраическое уравнение, которому удовлетворяет $r(x)$, полностью решает задачу, так как позволяет последовательно вычислять коэффициенты ряда (12.36). Сделать это можно, подставляя в (12.38) указанный ряд и приравнивая коэффициенты при одинаковых степенях в левой и правой части. Поступив так, получим, например,

$$r_0 = 1; \quad r_1 = \frac{r_0{}^3 + 3r_0^2 + 2r_0}{6} = 1; \quad r_2 = \frac{3r_0^2 r_1 + 3r_0 r_1}{6} = 1;$$

$$r_3 = \frac{3r_0 r_1^2 + 3r_0^2 r_2 + 3r_0 r_1 + 3r_0 r_2}{6} = 2;$$

$$r_4 = \frac{3r_0^3 r_3 + 6r_0 r_1 r_2 + r_1^3 + 3r_1^2 + 3r_0 r_3 + 2r_1}{6} = 4.$$

и т. д. Дальнейшие вычисления показывают, что

$$r(x) = 1 + x + x^2 + 2x^3 + 4x^4 + 8x^5 + 17x^6 + \ldots$$

Решение задачи для корневых С-деревьев позволяет перечислить и неизоморфные С-деревья, т. е. парафины. Опуская метод, которым это может быть достигнуто, укажем лишь на окончательный результат, приведённый в таблице 12. Теорема Пойа пригодна для перечисления изомеров многих других органических соединений с данной молекулярной формулой. Всякий раз

Таблица 12.

число вершин n	1	2	3	4	5	6	7	8	9	10	11	12	13
число С-деревьев (парафинов C_nH_{2n+2})	1	1	1	2	3	5	9	18	35	75	159	357	799

решение такой задачи сводится к отысканию числа неизоморфных графов того или иного типа. Интересующихся этим вопросом можно отослать к оригинальной работе Пойа [22]. Некоторая информация имеется в дополнении.

Новые комбинаторные задачи, связанные с теми или иными предложениями, породили целый ряд обобщений теоремы Пойа. Их обзору посвящена, например, статья [12].

Дополнения и задачи.

1. Докажем формулу (12.14) (обобщение бинома), использованную для подсчёта числа сочетаний с повторением. С этой целью проведём индукцию по n. При $n = 1$ формула верна, так как тогда получается уже установленное ранее равенство (12.13). Проведём шаг индукции. Предполагая, что верно (12.14), покажем, что из него следует равенство

$$(1-x)^{-(n+1)} = \\ = 1 - C^1_{-(n+1)}x + C^2_{-(n+1)}x^2 + \ldots + (-1)^m C^m_{-(n+1)}x^m + \ldots,$$

или, с учётом $(-1)^m C^m_{-(n+1)} = C^m_{n+m}$, равенство

$$(1-x)^{-(n+1)} = 1 + C^1_{n+1}x + C^2_{n+2}x^2 + \ldots + C^m_{n+m}x^m + \ldots \quad (12.39)$$

Умножим обе части доказываемого равенства на $(1-x)$ и докажем равенство, равносильное (12.39):

$$(1-x)^{-n} = (1 + C^1_{n+1}x + C^2_{n+2}x^2 + \ldots + C^m_{n+m}x^m + \ldots)(1-x). \quad (12.40)$$

После перемножения коэффициент при x^m в правой части оказывается

$$C^m_{n+m} - C^{m-1}_{n+m-1} = \frac{(n+m)\ldots(n+1)}{m!} - \frac{(n+m-1)\ldots(n+1)}{(m-1)!} = \\ = \frac{(n+m-1)\ldots(n+1)}{(m-1)!}\left(\frac{n+m}{m} - 1\right) = \\ = \frac{(n+m-1)\ldots(n+1)n}{m!} = C^m_{n+m-1} = (-1)^m C^m_{-n}. \quad (12.41)$$

Следовательно, (12.40) переходит в равенство

$$(1-x)^{-n} = 1 - C^1_{-n}\,x + C^2_{-n}\,x^2 + \ldots + (-1)^n C^m_{-n}\,x^m + \ldots,$$

верное по индуктивному предположению. Это и доказывает (12.14).

2. Указать производящую функцию для числа сочетаний из n элементов по m, в которых k-й элемент встречается не менее a_k и не более b_k раз $(k = 1, 2, \ldots, n)$.

3. Доказать, что число r_m разбиений числа m на различные натуральные слагаемые совпадает с числом его разбиений на нечётные слагаемые (без ограничения на повторения).

Указание: производящая функция для чисел r_m имеет вид:

$$f(x) = (1+x)(1+x^2)(1+x^3)\ldots = \prod_{i=1}^{\infty}(1+x^i).$$

Умножьте и разделите $f(x)$ на выражение

$$(1-x)(1-x^2)(1-x^3)\ldots = \prod_{i=1}^{\infty}(1-x^i),$$

приведя её к виду

$$f(x) = \frac{1}{(1-x)(1-x^3)(1-x^5)\ldots} = \\ = (1+x+x^2+\ldots)(1+x^3+x^6+\ldots)(1+x^5+x^{10}+\ldots)\ldots$$

Последнее и даёт перечисляющий ряд разбиений на нечётные слагаемые.

4. Найдите, используя теорему Пойа, число неэквивалентных окрасок вершин куба в два и три цвета. Та же задача для рёбер куба.

5. Докажите, что цикловой индекс группы диэдра D_n имеет вид

$$P_{D_n}(x_1, x_2, \ldots, x_n) = \begin{cases} \dfrac{1}{2n}\sum\limits_{m|n} \varphi(m) x_m^{n/m} + \dfrac{1}{2} x_1 x_2^{\frac{n-1}{2}} & \text{при } n \text{ нечётном;} \\ \dfrac{1}{2n}\sum\limits_{m|n} \varphi(m) x_m^{n/m} + \dfrac{1}{4}\left(x_2^{\frac{n}{2}} + x_1^2 x_2^{\frac{n-2}{2}}\right) & \text{при } n \text{ чётном,} \end{cases} \tag{12.42}$$

где $\varphi(m)$ — функция Эйлера. Используя (12.42), легко можно найти производящую функцию, подобную (12.35), для числа ожерелий, неэквивалентных относительно D_n (сделайте это).

6. Ещё о химии. Простейшие из циклических углеводородов имеют молекулярную формулу C_nH_{2n}, а в их структурной формуле атомы углерода образуют цикл. Рис. 57 соответствует так называемому циклопропану C_3H_6.

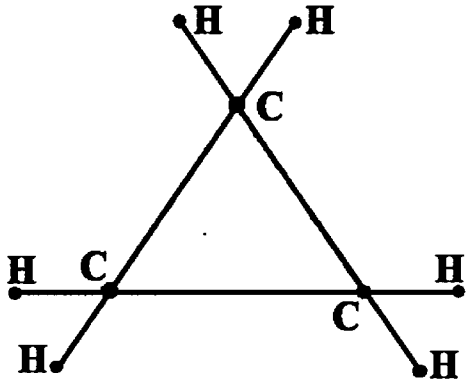


Рис. 57.

При замещении в C_nH_{2n} атомов водорода различными одновалентными радикалами (например, OH, Cl, Br и т. д.) получаются гомологи циклических углеводородов. Теорема Пойа позволяет найти число изомерных гомологов данного углеводорода C_nH_{2n}. Задача близка к перечислению ожерелий. Роль бусинок в этом случае играют атомы углерода, а красок — различные одновалентные радикалы. Правда, в отличие от обычных ожерелий, бусинки (атомы) могут быть окрашены в два цвета, это соответствует тому, что каждый атом углерода может быть соединён с двумя различными радикалами.

В качестве примера найдём число изомерных гомологов циклопропана с химической формулой $C_3H_3(OH)Br_2$. Используя терминологию ожерелий, можно сказать, что для раскраски бусинок используется три цвета, один из них соответствует атому водорода, два других — радикалам OH и Br. Группа G, относительно

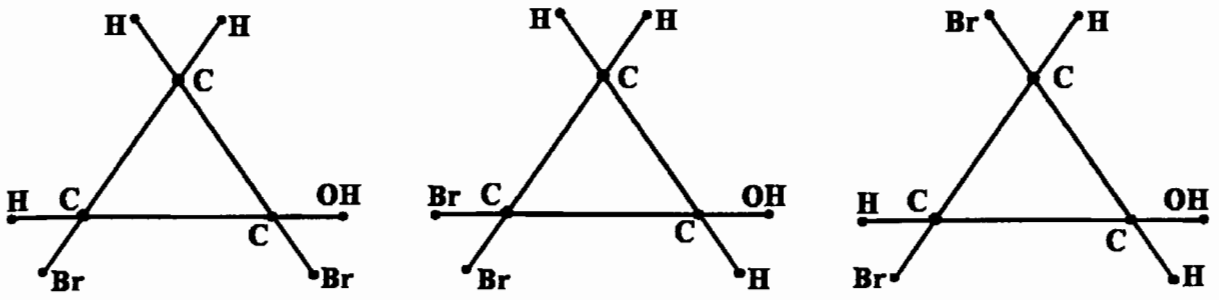


Рис. 58.

которой рассматривается эквивалентность, содержит 6 подстановок из группы диэдра D_3, 8 подстановок, переставляющих цвета отдельных бусинок и их комбинации, всего, как можно убедиться, 48 подстановок. Специальная структура этой группы позволяет записать её цикловой индекс в виде

$$P_6 = \frac{1}{6}\left[\left(\frac{x_1^2+x_2}{2}\right)^3 + 3\left(\frac{x_1^2+x_2}{2}\right)\left(\frac{x_2^2+x_4}{2}\right) + 2\left(\frac{x_3^2+x_6}{2}\right)\right] \tag{12.43}$$

(формула, вывод которой мы опускаем, некоторым специальным приёмом получается из циклового индекса группы диэдра, в данном случае D_3). Подставляя в (12.43), согласно теореме Пойа, производящую функцию цветов $y_1+y_2+y_3$, где y_1 соответствует H, y_2 — OH, y_3 — Br, получим перечисляющий ряд для изомерных гомологов циклопропана

$$\frac{1}{6}\left[\left(\frac{(y_1+y_2+y_3)^2+(y_1^2+y_2^2+y_3^2)}{2}\right)^3 + \right.$$
$$+ 3\frac{(y_1+y_2+y_3)^2+(y_1^2+y_2^2+y_3^2)}{2}\cdot\frac{(y_1^2+y_2^2+y_3^2)^2+(y_1^4+y_2^4+y_3^4)}{2} +$$
$$\left. + (y_1^3+y_2^3+y_3^3)^2 + (y_1^6+y_2^6+y_3^6)\right].$$

Интересующее нас число даётся коэффициентом при члене $y_1^3y_2y_3^2$, который равен 3. Три различных изомера $C_3H_3(OH)Br_2$ представлены на рис. 58.

ГЛАВА 13

ВЕКТОРЫ И ПРОСТРАНСТВА

Из всех разделов алгебраической науки, может быть, наиболее разработанным и эффектным является раздел, называемый линейной алгеброй. К тому же он и особенно важен для решения многочисленных прикладных задач. Спектр приложений линейной алгебры очень широк. Её аппаратом, не говоря уже о самой математике, пользуются естественные, технические, экономические, нередко и гуманитарные науки.

Что ещё характерно для линейной алгебры? Пожалуй, то, что она свободно пользуется хорошо знакомым геометрическим языком. Здесь мы встретим такие термины, как вектор, векторное пространство, скалярное произведение, евклидово пространство, ортогональность и т. д., происхождение которых из геометрии очевидно. Однако у тех, кто впервые знакомится с линейной алгеброй, употребление этих терминов часто вызывает недоумение. Дело в том, что объекты, к которым они применяются, могут «внешне» совсем не походить на свои геометрические прототипы. Достаточно сказать, что роль векторов (элементов векторного пространства) то и дело играют такие, скажем, объекты как многочлены, матрицы, функции и т. д. Позже мы сумеем понять, что именно объединяет эти столь различные вещи, позволяя говорить о них на едином языке. Тот факт, что этот язык именно геометрический, отчасти можно объяснить тем, что отправной точкой для развития линейной алгебры была аналитическая геометрия.

Читатель знает, что при решении геометрических задач методом координат сплошь и рядом возникают системы линейных уравнений с двумя или тремя неизвестными. Например, для отыскания точки пересечения двух прямых на плоскости приходится решать систему двух линейных уравнений с двумя неизвестными

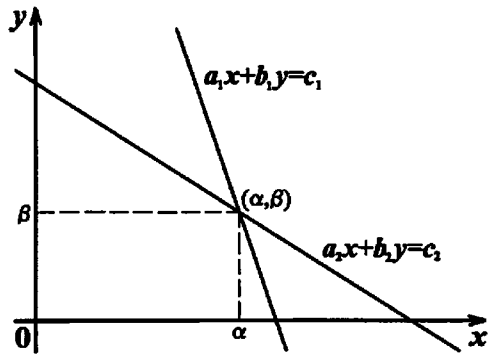


Рис. 59.

$$\begin{cases} a_1x + b_1y = c_1 \\ a_2x + b_2y = c_2, \end{cases}$$

где каждое из уравнений определяет прямую на плоскости (рис. 59), а решение (α, β) даёт координаты точки пересечения прямых, или радиуса-вектора этой точки. Рисунок соответствует случаю, когда система имеет единственное решение. Аналогично и в системе линейных уравнений с тремя неизвестными каждое уравнение можно интерпретировать как уравнение плоскости в пространстве, а всякое решение (α, β, γ) такой системы как точку или вектор в пространстве с указанными координатами.

В главе 8 мы уже говорили о том, что в математике и различных её приложениях приходится иметь дело с системами линейных уравнений, содержащих более трёх неизвестных. Напомним что всякая такая система с n неизвестными $x_1, x_2, \ldots, x_n$ имеет вид

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m, \end{cases} \tag{13.1}$$

а каждое её решение, как обычно, представляет из себя набор значений неизвестных $(\alpha_1, \alpha_2, \ldots, \alpha_n)$, обращающих каждое уравнение в верное равенство.

Хотя непосредственное геометрическое истолкование системы (13.1) при $n > 3$ уже невозможно, однако прежний геометриче-

ский язык двух и трёх измерений удобно сохранить для произвольного n. Этой цели и служат дальнейшие определения.

Любой упорядоченный набор из n действительных чисел $(\alpha_1, \alpha_2, \ldots, \alpha_n)$ назовём *n-мерным вектором*, а сами числа $\alpha_1, \alpha_2, \ldots, \alpha_n$ — координатами этого вектора.

Для таких векторов будем использовать обычную форму записи:

$$a = (\alpha_1, \alpha_2, \ldots, \alpha_n).$$

Нетрудно определить простейшие операции над n-мерными векторами по тем же правилам, что и для обычных векторов. А именно, если

$$a = (\alpha_1, \alpha_2, \ldots, \alpha_n), \quad b = (\beta_1, \beta_2, \ldots, \beta_n)$$

— два n-мерных вектора, то их суммой называется вектор

$$a + b = (\alpha_1 + \beta_1,\ \alpha_2 + \beta_2,\ \ldots,\ \alpha_n + \beta_n).$$

Произведением вектора a на число λ называется вектор

$$\lambda a = (\lambda\alpha_1, \lambda\alpha_2, \ldots, \lambda\alpha_n).$$

Множество всех n-мерных векторов с операциями сложения и умножения на число называется *n-мерным координатным, или арифметическим, пространством*[1].

По аналогии с обычной плоскостью множество всех n-мерных векторов, удовлетворяющих одному линейному уравнению с n неизвестными, называют *гиперплоскостью* в n-мерном пространстве. При таком определении множество всех решений системы (13.1) есть не что иное, как пересечение нескольких гиперплоскостей. В сходном смысле можно говорить и о произвольных поверхностях или областях в n-мерном пространстве, понимая под ними множество векторов, задаваемых одним или несколькими уравнениями, или неравенствами, не обязательно

[1] Векторы n-мерного пространства часто именуют также точками, имея в виду ту связь, которая существует между точками и соответствующими радиус-векторами в случае плоскости и трёхмерного пространства.

линейными. Так, *гиперповерхностью* называется множество точек, координаты которых удовлетворяют одному уравнению

$$F(x_1, x_2, \ldots, x_n) = 0$$

с n неизвестными. Простейшим после плоскости примером гиперповерхности является *сфера* радиуса R в n-мерном пространстве — множество точек, координаты которых удовлетворяют уравнению

$$x_1^2 + x_2^2 + \ldots + x_n^2 = R^2.$$

Эти и другие геометрические термины вовсе не для того, чтобы внушить читателю какие-то «таинственные» представления о столь же «таинственных» многомерных пространствах. Повторим, что такая геометрическая терминология образует удобный и образный язык, который, хотя и не может служить средством доказательств, но очень удачно подчёркивает общность закономерностей, присущих и обычным геометрическим образам и их многомерным аналогам. Здесь мы встретим немало подтверждений сказанному. Поучительным в этом смысле будет наше небольшое отступление в область линейного программирования.

Линейное программирование (кратко ЛП) — сравнительно недавно возникший раздел прикладной математики[1], теоретической базой которого является как раз линейная алгебра. Линейное программирование применимо к решению многих задач оптимизации, часто возникающих на практике. В одной из таких задач, которую мы рассмотрим, речь идёт о рациональном снабжении завода железной рудой. Руда может доставляться на завод из n пунктов добычи. Исходные данные таковы:

c_i — стоимость добычи одной тонны руды на i-й шахте и её доставки на завод,

b_i — объем добычи на i-ой шахте,

b — потребность завода в руде.

[1] Первая работа академика Л. В. Канторовича относится к 1939 году. Методы линейного программирования получили своё дальнейшее развитие в сороковых-пятидесятых годах прошлого века.

Потребитель стремится уменьшить расход на приобретение сырья. Если x_i — количество руды (в тоннах), которое будет доставляться с i-ой шахты, то этот расход в сумме составит

$$L = c_1x_1 + c_2x_2 + \ldots + c_nx_n. \tag{13.2}$$

Нужно учесть при этом, что неизвестные x_i обязаны удовлетворять ряду ограничений. Прежде всего, необходимо полностью удовлетворить потребность завода в сырье. Это означает, что должно выполняться условие

$$x_1 + x_2 + \ldots + x_n = b. \tag{13.3}$$

Кроме того, очевидно, что

$$x_i \leqslant b_i, \quad i = 1, 2, \ldots, n. \tag{13.4}$$

Наконец, по смыслу задачи величины x_i должны быть неотрицательны:

$$x_i \geqslant 0, \quad i = 1, 2, \ldots, n. \tag{13.5}$$

Задачу можно теперь сформулировать так: требуется найти наименьшее значение функции (13.2), или, как говорят, минимизировать её, при соблюдении условий (13.3), (13.4), (13.5). И сама функция (её обычно называют целевой), и левые части ограничений (13.3)–(13.5) линейно зависят от неизвестных x_i — это и является отличительным признаком задач линейного программирования.

В нашем конкретном случае решение может быть найдено совсем элементарно, однако для нас гораздо важнее на этом примере выявить общие черты, свойственные всякой задаче ЛП.

Можно заметить, во-первых, что при небольшом числе переменных задача ЛП имеет простой геометрический смысл. Пусть в нашем примере $n = 3$. Возьмём для определённости конкретные значения исходных данных

$$L = 2x_1 + 3x_2 + 4x_3,$$
$$x_1 + x_2 + x_3 = 10,$$
$$0 \leqslant x_1 \leqslant 3, \quad 0 \leqslant x_2 \leqslant 8, \quad 0 \leqslant x_3 \leqslant 5.$$

Из первого условия находим $x_3 = 10 - x_1 - x_2$, подставляя это в другие ограничения, приходим к равносильной задаче: минимизировать целевую функцию

$$L = 40 - 2x_1 - x_2 \tag{13.6}$$

при ограничениях

$$\begin{gathered} 5 \leqslant x_1 + x_2 \leqslant 10, \\ 0 \leqslant x_1 \leqslant 3, \quad 0 \leqslant x_2 \leqslant 8. \end{gathered} \tag{13.7}$$

Для минимизации функции L изобразим на плоскости многоугольную область S, которая определяется неравенствами (13.7) (рис. 60).

Перепишем равенство (13.6) для целевой функции в виде

$$2x_1 + x_2 = 40 - L. \tag{13.8}$$

При каждом значении L оно определяет прямую на плоскости. Две такие прямые, соответствующие значениям $L = 40$ и $L = 30$, проведены на нашем чертеже. Понятно, что уменьшению L соответствует параллельный перенос прямой (13.8) вправо (в направлении стрелки). Через точку D пройдёт тогда как раз та прямая, на которой достигается интересующий нас минимум. В указанной точке $x_1 = 3$, $x_2 = 7$, $L = 27$, при этом $x_3 = 0$.

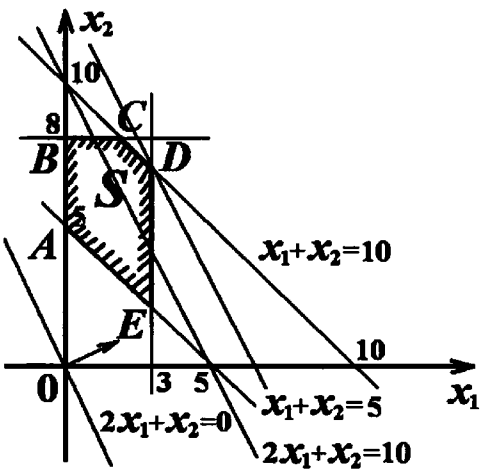


Рис. 60.

Произвольная задача ЛП может быть сформулирована подобно рассмотренной в следующем виде: минимизировать целевую функцию

$$L = c_1x_1 + c_2x_2 + \cdots + c_mx_m \tag{13.9}$$

при соблюдении ограничений

$$\begin{gathered} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1m}x_m \leqslant b_1, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nm}x_m \leqslant b_n, \\ x_1, x_2, \ldots, x_n \geqslant 0. \end{gathered} \tag{13.10}$$

При $m = 2$ ограничения (13.10), как и в примере, определяют выпуклую многоугольную область[1] на плоскости, а при $m = 3$ — выпуклую многогранную область в пространстве, а решение задачи, если оно существует, достигается на границе этой области, более точно — в некоторой её вершине.

В случае произвольного n также говорят, что ограничения (13.10) определяют выпуклую многогранную область в m-мерном координатном пространстве. Такие понятия, как «выпуклость», «граница области», «вершина», могут быть определены и в этом случае. И замечательно то, что утверждение, высказанное нами для $m = 2$ или 3, остаётся в силе и для любой размерности. Если задача ЛП (13.9) и (13.10) имеет решение, то оно достигается в вершине многогранной области, определённой ограничениями (13.10).

На этом общем утверждении основываются различные методы минимизации, в частности, и наиболее известный из них — симплекс-метод.

В упомянутом методе производится направленный перебор вершин, такой, что при переходе от одной вершины к следующей значение целевой функции уменьшается. Перебор заканчивается, когда будет достигнут минимум целевой функции.

Но вернёмся к векторам и точкам многомерных пространств. Мы хотим теперь в этих пространствах придать смысл таким геометрическим понятиям, как коллинеарность и компланарность векторов, длина вектора и расстояние между точками, ортогональность векторов и т. д. Всего этого можно достичь не только в n-мерном координатном пространстве, но и в более общей ситуации. Переход к ней связан с разными обстоятельствами. Главное из них то, что координатное пространство связано исключительно с идеей задания векторов и точек своими координатами. Однако, в случае двух или трёх измерений мы владеем и другим смыслом этих понятий, который, являясь исходным, вовсе не зависит от выбора системы координат. Такой бескоординатной точке зрения отвечает аксиоматическое определение векторного пространства произвольной размерности. Система его аксиом

[1] Или пустое множество, если ограничения несовместны.

является итогом пристального анализа алгебраических свойств операций над геометрическими векторами и над многими другими объектами, имеющими с ними алгебраическое сходство.

Множество V называется *векторным* (иначе, *линейным*) *пространством*, а его элементы называются *векторами*, если на этом множестве определены операции сложения векторов и умножения вектора на произвольный скаляр[1], удовлетворяющие следующим свойствам ($a, b, c \in V$, λ, λ_1, λ_2 — произвольные скаляры):

1. *сложение коммутативно*: $a + b = b + a$;
2. *сложение ассоциативно*: $(a + b) + c = a + (b + c)$;
3. *существует нулевой вектор* 0 *такой, что*
$$0 + a = a + 0 = a;$$
4. *для всякого вектора* $a \in V$ *существует противоположный вектор* $(-a)$ *такой, что* $a + (-a) = 0$;
5. *выполняются два закона дистрибутивности:*
$$\lambda(a + b) = \lambda a + \lambda b,$$
$$(\lambda_1 + \lambda_2)a = \lambda_1 a + \lambda_2 a;$$
6. *умножение на скаляр ассоциативно:* $\lambda_2(\lambda_1 a) = (\lambda_2 \lambda_1)\, a$;
7. $1 \cdot a = a$.

Первые четыре аксиомы устанавливают свойства операции сложения, и их можно было бы выразить короче, сказав, что относительно этой операции векторы образуют абелеву группу. В формулировке остальных свойств участвует и вторая операция — умножение на скаляр. Понятно, что множество всех геометрических векторов трёхмерного пространства является линейным пространством в смысле этого определения. Совершенно

[1] В качестве множества скаляров чаще всего рассматривают либо поле действительных, либо поле комплексных чисел. В первом случае мы имеем дело с действительным векторным пространством, во втором — с комплексным. Однако определение имеет смысл и тогда, когда множество скаляров есть произвольное поле (даже нечисловое).

очевидно также, что все перечисленные в определении свойства оказываются выполненными и в случае n-мерного координатного пространства. Отметим лишь, что роль нулевого элемента в нём играет вектор, у которого все координаты равны 0, тогда вектором, противоположным к $a = (\alpha_1, \alpha_2, \ldots, \alpha_n)$, служит вектор $-a = (-\alpha_1, -\alpha_2, \ldots, -\alpha_n)$. Следовательно, и оно может рассматриваться как один из примеров линейного пространства. Приведём ещё несколько примеров векторных пространств[1].

1. Множество всех действительных (или комплексных) матриц данного порядка $m \times n$ относительно операций сложения матриц и умножения матрицы на число.

2. Множество всех многочленов $\mathbb{F}[x]$ с коэффициентами из данного поля $\mathbb{F}$ (поля скаляров) относительно операций сложения многочленов и умножения многочлена на скаляр.

3. Множество всех действительных функций, непрерывных на заданном отрезке $[a, b]$, относительно сложения функций и умножения функции на число.

К списку примеров векторных пространств (над полем действительных чисел) можно присоединить также различные гиперкомплексные системы, и в частности, кватернионы, рассматривавшиеся в главе 8.

Отметим, что всё это — лишь небольшая часть из всего многообразия объектов, которые могут быть охвачены общим понятием векторного пространства.

Обратим внимание читателей на то, что в определении линейного пространства нет никакого упоминания о размерности. Однако представление о ней вполне может быть сюда привнесено, и этой цели служат понятия линейной зависимости и независимости векторов. Их мы и обсудим.

Во-первых, заметим, что операции, определённые в векторном пространстве, позволяют из произвольной системы векторов

[1] В каждом случае читатель может самостоятельно проверить справедливость аксиом.

$a_1, a_2, \ldots, a_k \in V$ образовывать выражения вида

$$\lambda_1 a_1 + \lambda_2 a_2 + \ldots + \lambda_k a_k, \quad (\lambda_i \text{ — скаляры}), \tag{13.11}$$

называемые *линейными комбинациями* этих векторов. Если все коэффициенты $\lambda_1, \lambda_2, \ldots, \lambda_k$ равны нулю, то и значением линейной комбинации будет нулевой вектор[1]. Не исключено, правда, что такое может случиться и тогда, когда не все коэффициенты λ_i нулевые.

Если существует равная нулевому вектору линейная комбинация

$$\lambda_1 a_1 + \lambda_2 a_2 + \ldots + \lambda_k a_k = 0,$$

в которой хотя бы один из коэффициентов отличен от нуля, то система векторов $a_1, a_2, \ldots, a_k \in V$ называется *линейно зависимой*. В противном случае система называется *линейно независимой*. Так, векторы 4-х мерного координатного пространства

$$a_1 = (2; 0; 1; -1), \quad a_2 = (2; -2; 1; 3), \quad a_3 = (2; -1; 1; 1)$$

линейно зависимы, поскольку

$$a_1 + a_2 - 2a_3 = (0; 0; 0; 0) = 0.$$

Линейная зависимость, как видно из определения, равносильна при $k \geqslant 2$ тому, что хотя бы один из векторов системы является линейной комбинацией остальных. Если система состоит из двух векторов a_1, a_2, то линейная зависимость её означает, что один из векторов пропорционален другому, скажем, $a_1 = \alpha \cdot a_2$; в двумерном и трёхмерном случае это равносильно *коллинеарности* геометрических векторов a_1 и a_2.

Точно так же, линейная зависимость системы из трёх векторов в обычном трёхмерном пространстве означает *компланарность* этих векторов. Понятие линейной зависимости является,

[1] Говоря так, мы подразумеваем, что умножение вектора на нулевой скаляр даёт нулевой вектор: $0 \cdot a = 0$. Как ни странно, но это нуждается в доказательстве — в системе аксиом нет такого свойства. Однако оно является простым их следствием. В самом деле, $a = 1 \cdot a = (1 + 0) \cdot a = 1 \cdot a + 0 \cdot a = a + 0 \cdot a$ для любого $a \in V$. Отсюда и следует, что $0 \cdot a = 0$.

таким образом, естественным обобщением понятий коллинеарности и компланарности геометрических векторов. Мы знаем, что в пространстве таких векторов особая роль принадлежит тройке координатных ортов i, j, k (понятно, линейно независимой), а именно, всякий вектор a единственным образом разлагается по этой тройке

$$a = xi + yj + zk,$$

где x, y, z — действительные числа (координаты вектора a).

Естественно возникает вопрос, нет ли в произвольном векторном пространстве V систем векторов, обладающих подобными свойствами. Речь, следовательно, идёт о такой системе

$$a_1, a_2, \ldots, a_n, \tag{13.12}$$

называемой обычно *базисом*, которая сама является линейно независимой, а любой вектор $a \in V$ оказывается линейной комбинацией её векторов

$$a = \alpha_1 a_1 + \alpha_2 a_2 + \ldots + \alpha_n a_n. \tag{13.13}$$

Обратимся к примерам. В первом из них — n-мерном координатном пространстве — указать такую систему совсем нетрудно. Достаточно рассмотреть векторы

$$\begin{aligned} a_1 &= (1, 0, 0, \ldots, 0), \\ a_2 &= (0, 1, 0, \ldots, 0), \\ a_3 &= (0, 0, 1, \ldots, 0), \\ &\ldots\ldots\ldots\ldots \\ a_n &= (0, 0, 0, \ldots, 1). \end{aligned} \tag{13.14}$$

Всякий n-мерный вектор $a = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ очевидно связан с ними соотношениями вида (13.13), коэффициенты в котором в точности совпадают с координатами этого вектора. Ясно также, что система (13.14) линейно независима, ведь линейная комбинация (13.13) может равняться нулевой строке лишь тогда, когда все коэффициенты $\alpha_i = 0$.

Однако в другом примере — векторном пространстве всех многочленов над данным полем — ситуация иная. Читатель легко поймёт, что всевозможные линейные комбинации конечного множества многочленов (каким бы оно ни было) не могут исчерпать всего этого пространства.

Это означает, что существование конечного базиса логически не следует из аксиом векторного пространства. Если такой базис существует, то пространство называется *конечномерным*, в противном случае — *бесконечномерным*[1]. И те, и другие важны для математики и её приложений. Здесь мы поговорим лишь о конечномерных пространствах.

Возвращаясь к трёхмерному пространству, заметим, что в нём роль базиса может играть любая тройка некомпланарных векторов a_1, a_2, a_3. Действительно, любой вектор a, как показано на рис. 61, разлагается по этой тройке

$$a = \lambda_1 a_1 + \lambda_2 a_2 + \lambda_3 a_3.$$

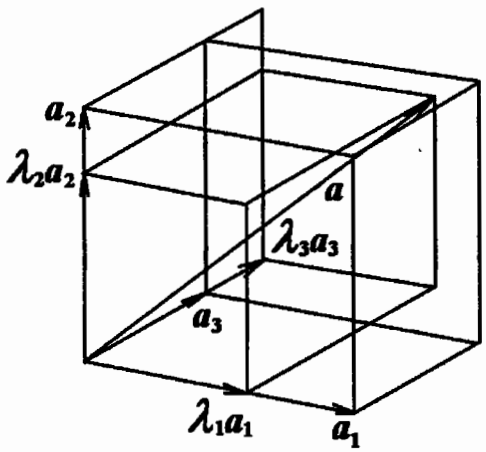


Рис. 61.

Подобно этому и в произвольном конечномерном пространстве базис можно выбрать многими различными способами. Так например, в n-мерном координатном пространстве система векторов

$$(1,0,0,\ldots,0),\ (1,1,0,\ldots,0),\ (1,1,1,\ldots,0),\ \ldots,\ (1,1,1,\ldots,1)$$

также образует базис (проверьте это). Однако не очень сложно доказать, что любые два базиса данного пространства V состоят всегда из одного и того же числа векторов. Это число n и называют размерностью векторного пространства, а само пространство — n-мерным. Выбрав в V какой-нибудь из базисов $a_1, a_2, \ldots, a_n$, мы увидим картину, которая очень напоминает n-мерное

[1] Класс конечномерных пространств может быть выделен дополнительной аксиомой, так называемой, *аксиомой размерности*: имеются n линейно независимых векторов, но любая система из большего числа векторов линейно зависима.

координатное пространство. В самом деле, по определению базиса, всякий вектор $a \in V$ разлагается по базисным векторам

$$a = \alpha_1 a_1 + \alpha_2 a_2 + \ldots + \alpha_n a_n. \qquad (13.15)$$

Заметим также, что для каждого a это разложение единственно. Предположив, что это не так, рассмотрим другое разложение

$$a = \alpha'_1 a_1 + \alpha'_2 a_2 + \ldots + \alpha'_n a_n. \qquad (13.16)$$

Вычитая (13.16) из (13.15), приходим к равенству

$$(\alpha_1 - \alpha'_1)\, a_1 + (\alpha_2 - \alpha'_2)\, a_2 + \ldots + (\alpha_n - \alpha'_n)\, a_n = 0,$$

в котором хотя бы один из коэффициентов $\alpha_i - \alpha'_i \neq 0$. Но последнее противоречит линейной независимости базисных векторов, следовательно, запись (13.15) единственна. Коэффициенты $\alpha_1, \alpha_2, \ldots, \alpha_n$ в разложении по базису называются, по аналогии с прежним, *координатами* вектора a в базисе $a_1, a_2, \ldots, a_n$.

Сказанное позволяет установить очень простую связь между произвольными n-мерными линейными и n-мерными координатными пространствами. Сопоставляя каждому вектору (13.15) строку его координат

$$a = \alpha_1 a_1 + \alpha_2 a_2 + \ldots + \alpha_n a_n \longrightarrow (\alpha_1, \alpha_2, \ldots, \alpha_n),$$

мы получаем взаимно однозначное соответствие между указанными пространствами. Очевидно при этом, что сумме векторов из V будет отвечать сумма соответствующих им координатных строк, а произведению вектора на скаляр — произведение соответствующей ему строки на тот же скаляр. Распространяя понятие изоморфизма на линейные пространства (как?), мы могли бы сказать, что всякое *n-мерное линейное пространство изоморфно n-мерному координатному пространству.*

Следующий шаг связан с введением в пространствах любой размерности скалярного произведения, оно позволит говорить в этой общей ситуации о таких (присущих евклидовой геометрии) понятиях, как длина, угол, ортогональность.

Будем говорить, что в *действительном* линейном пространстве V задано *скалярное произведение*, если каждой паре его векторов $a, b \in V$ соответствует действительное число, обозначаемое (a, b), и при любом выборе векторов выполнены следующие свойства:

$$\begin{aligned}
&1.\ (a, b) = (b, a);\\
&2.\ (a' + a'', b) = (a', b) + (a'', b);\\
&3.\ (\lambda a, b) = \lambda(a, b) \text{ для любого числа } \lambda;\\
&4.\ (a, a) \geqslant 0, \text{ причём } (a, a) = 0\\
&\quad \text{тогда и только тогда, когда } a = 0.
\end{aligned} \tag{13.17}$$

В этом определении (оно вновь аксиоматическое) выделены важнейшие свойства обычного скалярного произведения геометрических векторов, из которых в качестве следствия можно получить и все другие его свойства.

Линейное пространство, в котором задано скалярное произведение, называют обычно *евклидовым пространством*.

Нетрудно привести пример евклидова пространства любой конечной размерности. Для этого нужно лишь обобщить на n-мерное координатное пространство формулу, хорошо известную из аналитической геометрии. А именно, для любой пары векторов $a = (\alpha_1, \alpha_2, \ldots, \alpha_n)$, $b = (\beta_1, \beta_2, \ldots, \beta_n)$ координатного пространства положим

$$(a, b) = \alpha_1\beta_1 + \alpha_2\beta_2 + \ldots + \alpha_n\beta_n. \tag{13.18}$$

Достаточно ясно, что так определённое скалярное произведение удовлетворяет требованиям (13.17). Например,

$$\begin{aligned}
(\lambda a, b) &= (\lambda\alpha_1)\beta_1 + \ldots + (\lambda\alpha_n)\beta_n =\\
&= \lambda(\alpha_1\beta_1 + \ldots + \alpha_n\beta_n) = \lambda(a, b),\\
(a, a) &= \lambda_1^2 + \lambda_2^2 + \ldots + \lambda_n^2 \geqslant 0,
\end{aligned}$$

при этом $(a, a) = 0$ тогда и только тогда, когда $\alpha_1 = \alpha_2 = \ldots = \alpha_n = 0$, то есть вектор a нулевой. Таким образом, координатное пространство с введённым в нём по формуле (13.18) скалярным

произведением является евклидовым пространством. С другими примерами читатель познакомится ниже.

В любом евклидовом пространстве естественным образом определяется *длина* $|a|$ произвольного вектора a. Делается это по формуле, также заимствованной из аналитической геометрии

$$|a| = \sqrt{(a, a)}.$$

Можно говорить поэтому, что скалярное произведение определяет «метрику» в линейном пространстве, то есть способ измерения длин. В евклидовом координатном пространстве со скалярным произведением (13.18) действует знакомое правило для вычисления длины

$$|a| = \sqrt{\alpha_1^2 + \alpha_2^2 + \ldots + \alpha_n^2},$$

— длина вектора есть корень квадратный из суммы квадратов его координат.

Далее, два вектора a, b евклидова пространства назовём ортогональными, если их *скалярное произведение* равно нулю: $(a, b) = 0$ (напомним, что для геометрических векторов это равенство служит признаком ортогональности). Нередко бывает полезным рассматривать и угол φ между векторами a, b. Формула для его отыскания

$$\cos\varphi = \frac{(a, b)}{|a|\,|b|}$$

также скопирована с обычной[1].

Введённые понятия позволяют выделить в евклидовом пространстве базисы особого вида — так называемые *ортонормированные базисы*. Базис

$$e_1, e_2, \ldots, e_n \tag{13.19}$$

назовём ортонормированным, если он составлен из попарно ортогональных векторов единичной длины. Это значит, что векторы (13.19) удовлетворяют соотношениям

$$(e_i, e_j) = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases} \tag{13.20}$$

[1] Можно показать, что дробь справа по модулю не превосходит единицы.

Специальным методом (он называется процессом ортогонализации) в любом конечномерном евклидовом пространстве можно от произвольного базиса перейти к ортонормированному. Так, в трёхмерном пространстве роль ортонормированного базиса играет всякая тройка координатных ортов i, j, k. В то же время, система векторов (13.14) даёт простейший пример ортонормированного базиса в координатном евклидовом пространстве, скалярное произведение в котором определено по формуле (13.18).

Оказывается, та же формула будет действовать и в любом n-мерном евклидовом пространстве, если только в нём зафиксирован ортонормированный базис (13.20), а координаты всех векторов вычисляются именно в этом базисе. В самом деле, пусть любые два вектора

$$a = \alpha_1 e_1 + \alpha_2 e_2 + \ldots + \alpha_n e_n, \quad b = \beta_1 e_1 + \beta_2 e_2 + \ldots + \beta_n e_n$$

разложены по ортонормированному базису. Найдём их скалярное произведение.

$$\begin{aligned}(a,b) &= (\alpha_1 e_1 + \alpha_2 e_2 + \ldots + \alpha_n e_n, \beta_1 e_1 + \beta_2 e_2 + \ldots + \beta_n e_n) = \\ &= \alpha_1\beta_1(e_1,e_1) + \alpha_1\beta_2(e_1,e_2) + \ldots + \alpha_1\beta_n(e_1,e_n) + \\ &\quad + \alpha_2\beta_1(e_1,e_1) + \alpha_2\beta_2(e_2,e_2) + \ldots + \alpha_2\beta_n(e_2,e_n) + \ldots \\ &\quad \ldots + \alpha_n\beta_1(e_n,e_1) + \alpha_n\beta_2(e_n,e_2) + \ldots + \alpha_n\beta_n(e_n,e_n).\end{aligned}$$

Из соотношения (13.20) теперь моментально следует прежняя формула

$$(a,b) = \alpha_1\beta_1 + \alpha_2\beta_2 + \ldots + \alpha_n\beta_n.$$

Для того, чтобы по настоящему понять всю пользу введённых понятий, нужно изучить многие из разделов математики, такие как математический анализ, теорию дифференциальных уравнений, функциональный анализ, теорию вероятностей, теорию приближений и пр. В этих разных областях очень часто возникают ситуации, укладывающиеся в одну и ту же общую схему, которая отражена в понятиях линейного и евклидова пространства и других, связанных с ними.

Так, многие методы теории приближений основываются на факте, который для обычных пространств нам хорошо известен:

длина перпендикуляра, опущенного из точки на прямую или плоскость, меньше длины любой наклонной. В известном смысле он сохраняется и в многомерных евклидовых пространствах. Чтобы понять это, придётся ввести ещё одно важное понятие, обобщающее понятия прямой и плоскости в трёхмерном пространстве.

Подмножество L линейного пространства V назовём подпространством этого пространства, если

1) для любых векторов a и b, принадлежащих L, их сумма $a+b$ также принадлежит L;

2) для любого вектора a, принадлежащего L, и для любого скаляра λ вектор λa также принадлежит L.

Понятно, что указанные свойства выполнены, например, для множества всех векторов, принадлежащих некоторой фиксированной прямой или плоскости обычного пространства (см. рис. 62). Заметим также, что всякое подпространство L само образует линейное пространство относительно операций, определенных в V. Действительно, свойства 1) и 2) показывают, что на множестве L (как и на V) определены операции сложения и умножения на скаляр и, очевидно, для них выполнены все требования, содержащиеся в аксиомах векторного пространства.

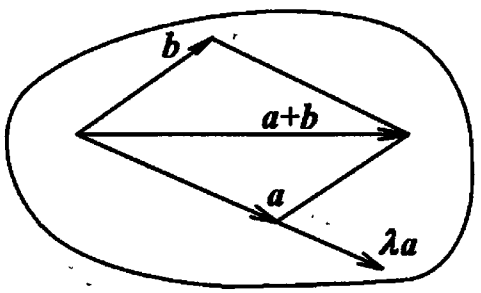


Рис. 62.

Всякое подпространство может быть задано в виде так называемой линейной оболочки подходящей системы векторов. Если $a_1, a_2, \ldots, a_k$ — произвольные векторы пространства V, то их линейной оболочкой называется множество L всевозможных линейных комбинаций этих векторов

$$L = \{\lambda_1 a_1 + \lambda_2 a_2 + \ldots + \lambda_k a_k\},$$

где $\lambda_1, \lambda_2, \ldots, \lambda_k$ — произвольные скаляры. Проверьте, что линейная оболочка удовлетворяет требованиям 1) и 2), входящим

в определение подпространства. Множество L называют также подпространством, порождённым векторами $a_1, a_2, \ldots, a_k$, (оно является наименьшим из подпространств, содержащих эти векторы).

В случае, если L — подпространство евклидова пространства V, можно говорить о величине расстояния от любого вектора b до подпространства. Будем для этого рассматривать произвольные векторы $x \in L$, каждый раз измеряя расстояние между b и x длиной разности $b - x$. Это условно показано на рис. 63.

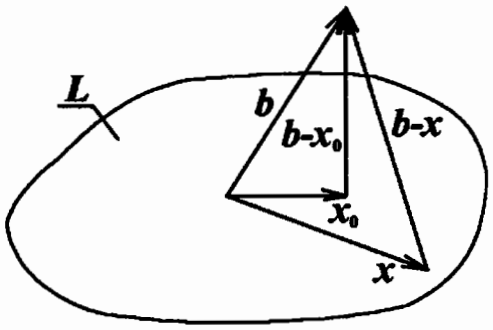


Рис. 63.

За *расстояние d от вектора до подпространства* примем минимальную из указанных длин, т. е.

$$d = \min_{x \in L} |b - x| .$$

Каким же должен быть вектор из L, для которого достигается этот минимум? Ответ следующий: в качестве ближайшего к b нужно взять такой вектор $x_0 \in L$, чтобы разность $b - x_0$ была ортогональна (в смысле данного выше определения) любому вектору из подпространства. Можно доказать, что такой вектор x_0 всегда существует и определён однозначно. Его называют *ортогональной проекцией вектора b*. Простая выкладка показывает, что длина перпендикуляра $b - x_0$ действительно наименьшая:

$$\begin{aligned} |b - x|^2 &= (b - x,\, b - x) = \\ &= \big((b - x_0) + (x_0 - x),\, (b - x_0) + (x_0 - x)\big) = \\ &= (b - x_0,\, b - x_0) + 2(x - x_0,\, b - x_0) + (x - x_0, x - x_0). \end{aligned}$$

Согласно выбору вектора x_0 мы имеем, что $(x - x_0,\, b - x_0) = 0$. Кроме того $(x - x_0,\, x - x_0) \geqslant 0$. Поэтому

$$|b - x|^2 \geqslant (b - x_0,\, b - x_0) = |b - x_0|^2,$$

что и требуется.

Высказанные здесь соображения лежат в основе многих разновидностей известного в математике «метода наименьших квадратов». Проиллюстрируем его на примере, относящемся к серии популярных задач на смеси и сплавы.

Таблица 13.

металл	латунь	бронза	мельхиор	новый сплав
медь	60	80	80	70
цинк	40	0	0	10
олово	0	20	0	10
никель	0	0	20	10

Плавильщик располагает тремя сплавами: латунью, бронзой и мельхиором. Процентное содержание меди, цинка, олова и никеля в этих сплавах указано в таблице 13. Цель плавильщика — получить новый сплав, в котором перечисленные четыре металла составляют соответственно 70, 10, 10, 10 процентов. Осуществима ли такая цель, а если нет, то насколько близко можно подойти к желаемому составу нового сплава?

Пусть x_1, x_2, x_3 — доли, которые должны составить латунь, бронза и мельхиор в предполагаемом сплаве. Баланс по каждому из четырёх металлов даётся тогда уравнениями

$$\begin{aligned}
&\text{медь:} && 0,6x_1+0,8x_2+0,8x_3 = 0,7\\
&\text{цинк:} && 0,4x_1 = 0,1\\
&\text{олово:} && 0,2x_2 = 0,1\\
&\text{никель:} && 0,2x_3 = 0,1
\end{aligned}$$

или иначе,

$$\begin{aligned}
6x_1+8x_2+8x_3 &= 7\\
4x_1 &= 1\\
2x_2 &= 1\\
2x_3 &= 1
\end{aligned} \qquad (13.21)$$

Ясно, что получившаяся система несовместна, и потому ответ на первый вопрос задачи отрицательный. Наш результат можно

выразить и иначе, если рассмотреть векторы-столбцы

$$a_1 = \begin{pmatrix} 6 \\ 4 \\ 0 \\ 0 \end{pmatrix}, \quad a_2 = \begin{pmatrix} 8 \\ 0 \\ 2 \\ 0 \end{pmatrix}, \quad a_3 = \begin{pmatrix} 8 \\ 0 \\ 0 \\ 2 \end{pmatrix}, \quad b = \begin{pmatrix} 7 \\ 1 \\ 1 \\ 1 \end{pmatrix},$$

составленные из коэффициентов при неизвестных и свободных членов. Несовместность системы (13.21) означает тогда, что вектор b нельзя записать в виде линейной комбинации $x_1a_1 + x_2a_2 + x_3a_3$ трёх других векторов, или (это то же самое) вектор b не содержится в их линейной оболочке

$$L = \langle\, x_1a_1 + x_2a_2 + x_3a_3 \mid x_1, x_2, x_3 \text{ — скаляры} \,\rangle\,.$$

В то же время векторы из подпространства L с точностью до пропорциональности соответствуют тем сплавам, которые могут быть получены при смешивании в той или иной пропорции латуни, бронзы и мельхиора. Будем рассматривать наши векторы как элементы четырёхмерного координатного евклидова пространства со скалярным произведением, определённым по формуле (13.18), где $n = 4$. Теперь второй вопрос можно уточнить так: в подпространстве L найти вектор, ближайший к вектору b, иными словами ортогональную проекцию x_0 вектора b на подпространство L. Соответствующий этому вектору сплав и будем считать наилучшим приближением к требуемому в задаче. Для отыскания ортогональной проекции

$$x_0 = x_1a_1 + x_2a_2 + x_3a_3 \tag{13.22}$$

заметим, что для всех $j = 1, 2, 3$

$$(b, a_j) = (x_0 + (b - x_0), a_j) = (x_0, a_j) + (b - x_0, a_j) = (x_0, a_j),$$

так как вектор $b - x_0$ ортогонален любому вектору из L. Подставляя в полученные равенства $(b, a_j) = (x_0, a_j)$ запись вектора x_0 в виде (13.22), приходим к системе линейных уравнений для определения x_1, x_2, x_3

$$\begin{cases} x_1(a_1, a_1) + x_2(a_2, a_1) + x_3(a_3, a_1) = (b, a_1), \\ x_1(a_1, a_2) + x_2(a_2, a_2) + x_3(a_3, a_2) = (b, a_2), \\ x_1(a_1, a_3) + x_2(a_2, a_3) + x_3(a_3, a_3) = (b, a_3). \end{cases}$$

Коэффициенты системы (скалярные произведения) подсчитываем по упомянутой выше формуле, например,

$$(a_1, a_1) = 6\cdot 6 + 4\cdot 4 = 52, \quad (a_1, a_2) = (a_2, a_1) = 6\cdot 8 = 48, \quad \text{и т. д.}$$

В результате получим

$$\begin{cases} 52x_1 + 48x_2 + 48x_3 = 46, \\ 48x_1 + 68x_2 + 64x_3 = 58, \\ 48x_1 + 64x_2 + 68x_3 = 58. \end{cases}$$

Решая систему, находим

$$x_1 \approx 0,22, \qquad x_2 = x_3 \approx 0,39$$

и вектор $x_0 = (7,56;\ 0,88;\ 0,78;\ 0,78)$. Соответствующий ему сплав, которым может удовлетвориться плавильщик, содержит 75,6% меди, 8,8% цинка и по 7,8% олова и никеля.

На многомерные пространства могут быть перенесены идеи не только евклидовой, но и неевклидовых геометрий. Одним из замечательных примеров этого может служить четырёхмерное «пространство-время», служащее математической моделью специальной теории относительности Эйнштейна. В общем случае многомерные неевклидовы пространства получаются, если видоизменить определение скалярного произведения, отказавшись от свойства положительности скалярного квадрата (аксиома 4) и заменив его более слабым свойством.

В нашем беглом экскурсе в линейную алгебру мы остановимся ещё на одном из центральных её понятий. Речь идёт о линейных преобразованиях векторного пространства. Линейным называется такое преобразование A векторного пространства V, для которого выполнены следующие свойства:

$$\begin{aligned} A(a+b) &= A(a) + A(b) \quad \text{для любых } a, b \in V, \\ A(\lambda a) &= \lambda A(a) \quad \text{для любого } a \in V \text{ и любого скаляра } \lambda. \end{aligned} \tag{13.23}$$

Поворот плоскости вокруг начала координат, проектирование трёхмерного пространства на координатную плоскость дают нам

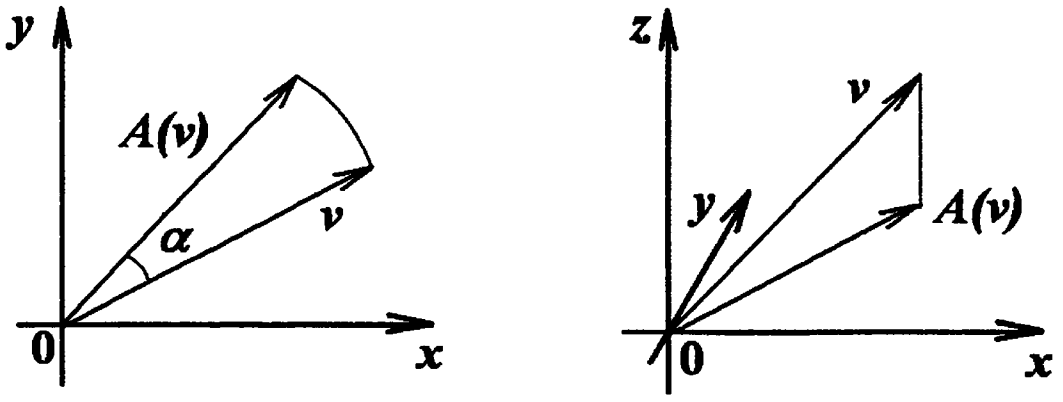


Рис. 64.

простейшие примеры линейных преобразований (рис. 64). Линейные преобразования имеют самую тесную связь с матрицами. Чтобы это понять, выберем в n-мерном пространстве некоторый базис $e_1, e_2, \ldots, e_n$ и рассмотрим образы базисных векторов $A(e_1), A(e_2), \ldots, A(e_n)$. Каждый из них, как и всякий вектор пространства, является линейной комбинацией векторов базиса, поэтому мы можем написать

$$\begin{aligned} A(e_1) &= a_{11}e_1 + a_{21}e_2 + \ldots + a_{n1}e_n, \\ A(e_2) &= a_{12}e_1 + a_{22}e_2 + \ldots + a_{n2}e_n, \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ A(e_n) &= a_{1n}e_1 + a_{2n}e_2 + \ldots + a_{nn}e_n. \end{aligned} \tag{13.24}$$

Сопоставим теперь преобразованию A матрицу

$$A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{pmatrix},$$

поместив в каждый её столбец координаты образов

$$A(e_1), A(e_2), \ldots, A(e_n)$$

базисных векторов в самом этом базисе. Рекомендуем убедиться, что для упомянутых выше примеров линейных преобразований

(поворота на угол α и проектирования на плоскость xOy) матрицы имеют вид соответственно

$$\begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \quad \text{и} \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

Оказывается, матрицей A линейное преобразование определено однозначно — зная её, мы можем найти образ любого вектора. Действительно, если

$$a = \alpha_1 e_1 + \alpha_2 e_2 + \ldots + \alpha_n e_n,$$

то, используя свойства (13.23) и равенства (13.24), находим

$$\begin{aligned} A(a) &= A(\alpha_1 e_1 + \alpha_2 e_2 + \ldots + \alpha_n e_n) = \\ &= \alpha_1 A(e_1) + \alpha_2 A(e_2) + \ldots + \alpha_n A(e_n) = \\ &= \alpha_1(a_{11}e_1 + a_{21}e_2 + \ldots + a_{n1}e_n) + \\ &\quad + \alpha_2(a_{12}e_1 + a_{22}e_2 + \ldots + a_{n2}e_n) + \\ &\quad + \ldots + \alpha_n(a_{1n}e_1 + a_{2n}e_2 + \ldots + a_{nn}e_n) = \\ &= (\alpha_1 a_{11} + \alpha_2 a_{12} + \ldots + \alpha_n a_{1n})\, e_1 + \\ &\quad + \ldots + (\alpha_1 a_{n1} + \alpha_2 a_{n2} + \ldots + \alpha_n a_{nn})\, e_n. \end{aligned}$$

В матричной форме это можно выразить так: чтобы получить координаты образа, нужно матрицу преобразования умножить на столбец вектора a:

$$A \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix} = \begin{pmatrix} \alpha_1 a_{11} + \alpha_2 a_{12} + \ldots + \alpha_n a_{1n} \\ \alpha_1 a_{21} + \alpha_2 a_{22} + \ldots + \alpha_n a_{2n} \\ \vdots \\ \alpha_1 a_{n1} + \alpha_2 a_{n2} + \ldots + \alpha_n a_{nn} \end{pmatrix}.$$

Кроме умножения преобразований, определяемого обычным образом как композиция, имеют смысл операции их сложения и умножения преобразования на число. Суммой преобразований A и B называется такое преобразование $A + B$, что

$$(A + B)(a) = A(a) + B(a) \quad \text{для любого } a \in V.$$

Соответственным образом определяется и умножение на скаляр

$$(\lambda A)(a) = \lambda A(a) \quad \text{для любого } a \in V.$$

Читатель может убедиться, что имеется полная параллель между указанными операциями над линейными преобразованиями и соответствующими операциями над их матрицами. Так, произведение матриц служит матрицей произведения соответствующих преобразований, аналогично и для двух других операций.

При изучении линейных преобразований обычно выделяют различные их классы, обладающие некоторыми дополнительными свойствами. Так, один из важнейших классов составляют так называемые *ортогональные* преобразования евклидова пространства, то есть такие, которые сохраняют длину всякого вектора:

$$|A(a)| = |a|.$$

В случае плоскости и трёхмерного пространства преобразования этого типа исчерпываются поворотами и зеркальными симметриями. Эти и другие типы линейных преобразований широко используются во многих математических и прикладных вопросах.

Дополнения и задачи

1. Мы отмечали, что определение векторного пространства со всеми входящими в него аксиомами имеет смысл и тогда, когда в качестве скаляров используются элементы произвольного поля $\mathbb{F}$. Это поле может быть даже и конечным. Такие понятия как линейная зависимость и независимость векторов, базис, размерность, подпространство и т. д. переносятся без изменений и в эту более общую ситуацию[1]. Дело, конечно, не только в логической возможности подобных обобщений. Важнее то, что во многих теоретических и практических задачах векторные пространства над теми или иными полями (не обязательно $\mathbb{R}$ или $\mathbb{C}$)

[1] Сказанное не относится, вообще говоря, к тем понятиям, которые используют скалярное произведение и расстояние.

оказываются полезной математической моделью изучаемого круга вопросов, а аппарат линейной алгебры — ценным орудием для его изучения. Позднее (в главе 15) мы познакомимся с тем, как векторные пространства над конечным полем могут применяться в теории кодирования, занимающейся вопросами экономной и надёжной передачи информации.

2. Пусть V — n-мерное пространство над конечным полем $\mathbb{F}$, содержащим q элементов. Докажите, что число векторов такого пространства V равно q^n.

Вы это легко можете понять, если вспомните, что каждому вектору соответствует строка его координат $(a_1, a_2, \ldots, a_n)$, причём в рассматриваемом случае все a_i пробегают элементы конечного поля $\mathbb{F}$.

3. Пусть поле $\mathbb{F}$ является расширением поля $\mathbb{P}$ ($\mathbb{F} \supseteq \mathbb{P}$). Взгляд со стороны линейной алгебры оказывается удобным и в этой ситуации. Элементы расширения $\mathbb{F}$ будем считать «векторами», элементы $\mathbb{P}$ — «скалярами». Векторы будем складывать в соответствии со сложением, определенным в поле $\mathbb{F}$. Другая операция — умножения позволяет, в частности, рассматривать произведения «вектора» на «скаляр», то есть выражение вида λa, где $a \in \mathbb{F}$, $\lambda \in \mathbb{P}$. Все аксиомы векторного пространства (проверьте) будут при этом выполнены, они являются непосредственным следствием определения поля. Значит, расширение можно рассматривать как векторное пространство над своим подполем.

В качестве простейшего приложения такой точки зрения можно получить следующий результат:

число элементов конечного поля есть степень простого числа.

Для доказательства достаточно, во-первых, вспомнить, что любое поле $\mathbb{F}$ содержит простое подполе (дополнение 4 к главе 9), если $\mathbb{F}$ к тому же и конечно, то это простое подполе $\mathbb{P}$ изоморфно $\mathbb{Z}_p$ и содержит p элементов. Но тогда $\mathbb{F}$ есть векторное пространство над $\mathbb{Z}_p$, к тому же и конечномерное (в противном случае $\mathbb{F}$ содержало бы бесконечно много элементов). Если n — размерность $\mathbb{F}$ над $\mathbb{Z}_p$ то, согласно пункту 2, число элементов в $\mathbb{F}$ равно p^n.

4. Если $\mathbb{F} \supseteq \mathbb{P}$ и $\mathbb{F}$ имеет конечную размерность n как векторное пространство над $\mathbb{P}$, то $\mathbb{F}$ называется *конечным расширением* поля $\mathbb{P}$, а n называется *степенью расширения*.

Например, поле $\mathbb{C}$ — конечное расширение поля $\mathbb{R}$ степени 2. В качестве базиса можно взять элементы 1 и i, так как всякое комплексное число $a + bi$ есть их линейная комбинация с коэффициентами $a, b \in \mathbb{R}$. С другой стороны, поле $\mathbb{R}$ не является конечным расширением поля $\mathbb{Q}$.

Проверьте, что поле $\mathbb{Q}(\xi)$, где $\xi \neq 1$ и является корнем кубическим из 1, есть конечное расширение поля $\mathbb{Q}$ степени 2.

Можно доказать — это сложнее — аналогичное утверждение, если $\xi = \cos\dfrac{2\pi}{p} + i\sin\dfrac{2\pi}{p}$ — корень p-ой степени из 1 (p — простое). В этом случае $\mathbb{Q}(\xi)$ имеет степень $p - 1$, а в качестве его базиса над $\mathbb{Q}$ могут быть взяты элементы $1, \xi, \xi^2, \ldots, \xi^{p-2}$.

Такие поля, получающиеся в результате присоединения корней из 1, называют *круговыми*. Их первоисточником, как можно понять, является работа Гаусса о построении правильных многоугольников (см. главу 5).

5. Пусть вновь $\mathbb{F}$ — расширение поля $\mathbb{P}$. Напомним, что элемент $\alpha \in \mathbb{F}$ называется алгебраическим над $\mathbb{P}$, если он является корнем некоторого многочлена с коэффициентами из $\mathbb{P}$. В дополнении 2 к главе 5 это понятие рассматривалось для случая числовых полей. Теперь мы можем, если угодно, не стеснять себя таким ограничением.

Если всякий элемент поля $\mathbb{F}$ алгебраичен над $\mathbb{P}$, то само $\mathbb{F}$ называется *алгебраическим расширением* поля $\mathbb{P}$. Если $\mathbb{F} = \mathbb{P}(\theta)$, где θ — алгебраический элемент, то $\mathbb{F}$ называется простым алгебраическим расширением поля $\mathbb{P}$.

Существует интересная связь между конечными и алгебраическими расширениями:

любое конечное расширение является алгебраическим; любое простое алгебраическое расширение является конечным.

ГЛАВА 14

ВНОВЬ К ИСТОКАМ (О ГАЛУА И ЕГО ТЕОРИИ)

«Я сделал в анализе несколько открытий... В теории уравнений я исследовал, в каких случаях они разрешимы в радикалах, что мне дало повод углубить эту теорию и описать все преобразования над уравнением, допустимые, даже когда оно и не решается в радикалах».

Это выдержка из письма, написанного Э. Галуа накануне своей трагической гибели. В письме, адресованном близкому другу, Галуа кратко излагает содержание своих исследований. Он заканчивает письмо такими словами:

«Часто в своей жизни я решался высказывать предложения, в которых я не был уверен, но всё, что я написал здесь, уже около года в моей голове, и слишком в моих интересах не ошибиться, чтобы можно было заподозрить меня в том, что я объявляю теоремы, для которых не имел бы полного доказательства.

Ты попросишь публично Якоби и Гаусса дать свое заключение не об истинности, а о значении теорем.

Надеюсь, после этого появятся люди, которые найдут свою выгоду в расшифровке всей этой неразберихи».

Надеждам Галуа на признание со стороны Гаусса и Якоби, как, впрочем, и других именитых современников, не суждено было, однако, исполниться. И причиной тому были не только исключительная трудность темы и совершенно необычный стиль рассуждений Галуа. Позднее (1846 г.) знаменитый французский математик Жозеф Лиувилль, первый, кто серьёзно разобрался в работах Галуа, скажет:

«Референтам показались неясными формулировки молодого математика ... Преувеличенное стремление к краткости породило этот недостаток, которого нужно в первую очередь стараться избегать, когда имеешь дело с отвлечёнными и таинственными категориями чистой алгебры. Тому, кто намерен вести читателя к неизведанной земле, далеко от проторенной дороги, воистину необходима ясность».

Однако сам Галуа ничуть не заблуждался ни в оценке истинности, ни в оценке значимости своих результатов. Он разве лишь не мог во всей полноте представить, насколько сильно новизна и глубина его идей повлияют на дальнейшее развитие алгебры, вызвав появление в ней многих новых понятий и направлений. Группа Галуа, соответствия Галуа, поля Галуа, когомологии Галуа — это лишь немногие из словосочетаний, которыми увековечено в алгебре и за её пределами имя гениального французского математика.

Бесспорно, что на формировании математических интересов юного Галуа сказались прежде всего алгебраические работы Лагранжа и Гаусса. Лагранж, как мы знаем, впервые связал задачу разрешимости алгебраических уравнений с совершенно новыми принципами, а Гаусс блестяще их реализовал для уравнений деления круга.

Вспомним, что одним из исходных пунктов мемуара Лагранжа было то, что симметрические (или иначе, инвариантные относительно группы S_n) функции от корней уравнения принимают значения в основном поле $\mathbb{P}$. В качестве $\mathbb{P}$ можно взять любое поле, содержащее коэффициенты уравнения (у Лагранжа это поле рациональных чисел). Нетрудно убедиться, что в случае общего алгебраического уравнения, то есть уравнения с произвольными буквенными коэффициентами, верно и обратное. Именно, всякая рациональная функция $f(x_1, x_2, \ldots, x_n)$, которая на корнях любого уравнения $a_0x^n + a_1x^{n-1} + \ldots + a_n = 0$ с коэффициентами из $\mathbb{P}$ принимает значения, также принадлежащие $\mathbb{P}$, является симметрической (инвариантной относительно группы S_n).

Этими обстоятельствами в сущности и объясняется, почему симметрическая группа оказалась удобным инструментом для анализа общего алгебраического уравнения.

Однако в случае конкретного уравнения с заданными числовыми коэффициентами ситуация может оказаться иной — вполне могут найтись рациональные функции $f(x_1, x_2, \ldots, x_n)$, которые на корнях данного уравнения принимают значения в основном поле, вовсе не являясь при этом симметрическими. Нетрудно привести примеры. Так, уравнение

$$x^4 - 4x^2 + 2 = 0 \tag{14.1}$$

имеет, как показывает вычисление, корни

$$\alpha_1 = \sqrt{2+\sqrt{2}},\ \alpha_2 = -\sqrt{2+\sqrt{2}},\ \alpha_3 = \sqrt{2-\sqrt{2}},\ \alpha_4 = -\sqrt{2-\sqrt{2}},$$

а, например, функции $x_1x_2 + x_3x_4$, $x_1^2 + x_3^2$, $x_2^2 + x_4^2$ (несимметрические!) принимают на этих корнях рациональные значения:

$$\alpha_1\alpha_2 + \alpha_3\alpha_4 = -4, \quad \alpha_1^2 + \alpha_3^2 = \alpha_2^2 + \alpha_4^2 = 4. \tag{14.2}$$

В главе 5 мы уже отмечали, что между корнями конкретного уравнения могут быть соотношения, не являющиеся следствием формул Виета. Именно это обстоятельство и связано с существованием несимметрических функций $\varphi(x_1, x_2, \ldots, x_n)$, принимающих на корнях уравнения значения в основном поле $\mathbb{P}$. Если

$\alpha_1, \alpha_2, \ldots, \alpha_n$ — корни уравнения и

$$\varphi(\alpha_1, \alpha_2, \ldots, \alpha_n) = a \in \mathbb{P}, \tag{14.3}$$

то равенство (14.3) и будет соотношением между корнями, которое не вытекает из формул Виета. В только что рассмотренном примере подобными соотношениями (не всеми) являются равенства (14.2).

Рассуждения Лагранжа, справедливые для общего уравнения, очевидно, никак не учитывали специфику конкретных уравнений. «Нельзя ли тогда, — задаётся вопросом Галуа, — заменить в этих рассуждениях группу всех подстановок S_n другой группой, которая, с одной стороны, отражала бы особенности данного уравнения, и в то же время, соотносилась бы с ним в каком-то смысле так же, как группа S_n с общим уравнением». Основная мысль, в которой после упорных размышлений утверждается Галуа, состояла в следующем. Для каждого конкретного уравнения нужно рассматривать не только симметрические рациональные функции $f(x_1, x_2, \ldots, x_n)$, но и все такие, которые на его корнях принимают значения из основного поля. А тогда из группы S_n следует оставить множество G лишь тех подстановок, которые не меняют значений указанных рациональных функций. Точнее, если $\alpha_1, \alpha_2, \ldots, \alpha_n$ — корни, а $f(\alpha_1, \alpha_2, \ldots, \alpha_n) = a \in \mathbb{P}$, то для всякой подстановки $\sigma \in G$ $f(\alpha_{\sigma(1)}, \alpha_{\sigma(2)}, \ldots, \alpha_{\sigma(n)}) = a$.

Например, для уравнения (14.1) такими подстановками будут следующие:

$$\begin{aligned}
\sigma_1 &= \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{pmatrix} = e, \\
\sigma_2 &= \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{pmatrix} = (1,2)(3,4), \\
\sigma_3 &= \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \end{pmatrix} = (1,3)(2,4), \\
\sigma_4 &= \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{pmatrix} = (1,4)(2,3),
\end{aligned} \tag{14.4}$$

и никакая другая подстановка указанным выше свойством не обладает (достаточно в этом убедиться для транспозиций).

Из предыдущего ясно, что введённое Галуа множество подстановок (он называет его *группой уравнения*, а мы будем называть её также *группой Галуа*) можно охарактеризовать иначе.

> *Это множество G состоит в точности из всех таких подстановок σ, которые всякое соотношение вида* (14.3) *между корнями уравнения переводят вновь в верное соотношение*
>
> $$\varphi(\alpha_{\sigma(1)}, \alpha_{\sigma(2)}, \ldots, \alpha_{\sigma(n)}) = a$$
>
> *между этими корнями*[1].

Подобным свойством, как можно вспомнить, обладала и циклическая группа подстановок, рассмотренная Гауссом в случае уравнения деления круга и оказавшаяся ключом к анализу этого уравнения.

Скорее всего, успешный анализ Гаусса и придал исследованиям Галуа то направление, которое привело его к понятию группы уравнения. По мысли Галуа эта группа должна была стать тем объектом, который бы вбирал в себя всю информацию о трудностях и способах решения каждого данного уравнения. И от того, как она устроена, всецело должен зависеть ответ на вопрос, можно ли свести решение данного уравнения к решению более простых вспомогательных уравнений.

Здесь уместно заметить, что строение группы уравнения зависит не только от самого этого уравнения, но и от того, над каким полем его рассматривают. И это связано с тем, что «свойства и трудности уравнения могут быть сделаны», как отмечает Галуа, «совершенно разными сообразно количествам, которые к нему присоединены. Например, присоединение некоторого количества может сделать приводимым неприводимое уравнение». Простой иллюстрацией служит рассмотренное выше уравнение (14.1).

---

[1] Читатель легко может убедиться, что относительно операции умножения подстановок множество G образует группу в смысле данного в главе 4 определения. Для множества подстановок (14.4) это можно сделать непосредственным вычислением.

Над полем $\mathbb{Q}(\sqrt{2})$, получающемся присоединением $\sqrt{2}$ (корня многочлена x^2-2) к основному полю $\mathbb{Q}$, левая часть этого уравнения и в самом деле разлагается на множители

$$x^4-4x^2+2=(x^2-2+\sqrt{2})(x^2-2-\sqrt{2}),$$

а к прежним соотношениям между корнями добавляются ещё и новые, например,

$$\alpha_1\alpha_2=-(2+\sqrt{2}),\quad \alpha_3\alpha_4=-(2-\sqrt{2}),\quad \alpha_1\alpha_3=\sqrt{2}.$$

Это отражается и на группе уравнения, теперь она состоит, как можно понять, всего лишь из двух подстановок:

$$\sigma_1=e,\quad \sigma_2=(1,2)(3,4).$$

Всякий раз в результате решения вспомогательного уравнения мы можем к исходному основному полю присоединить новые величины θ_1, θ_2, ..., θ_k — корни этого уравнения, а заданное уравнение рассматривать уже над более широким полем $\mathbb{P}_1=$ $=\mathbb{P}(\theta_1,\theta_2,\ldots,\theta_k)$ — расширением основного поля. Какими же должны быть эти вспомогательные уравнения? Ясно одно: выбирать их надо так, чтобы во-первых, решение их являлось бы задачей более простой, чем задача решения исходного уравнения, и во-вторых, при переходе к расширению $\mathbb{P}_1$ основного поля упростилось бы и решение заданного уравнения. Каким-то образом его группа должна «сводиться» тогда к более простым: группе вспомогательного уравнения над основным полем $\mathbb{P}$ и группе заданного уравнения над расширением $\mathbb{P}_1$.

Теперь, когда мы в основных чертах обрисовали замысел Галуа, попытаемся изложить некоторые из его результатов в переводе на язык современной алгебры. Пусть над основным полем $\mathbb{P}$ задано уравнение

$$p(x)=a_0x^n+a_1x^{n-1}+\cdots+a_{n-1}x+a_n=0. \tag{14.5}$$

Рассмотрим расширение $\mathbb{F}=\mathbb{P}(\alpha_1,\alpha_2,\ldots,\alpha_n)$, являющееся полем разложения этого уравнения (определение в дополнении 11 к главе 6). Оказывается, с каждой подстановкой σ из группы

Галуа можно связать отображение поля $\mathbb{F}$ на себя (будем его обозначать той же буквой, надеясь, что это не вызовет путаницы). Поступим так. Возьмём произвольный элемент $\alpha \in \mathbb{F}$. Если $\alpha \in \mathbb{P}$, то будем считать по определению, что $\sigma(\alpha) = \alpha$, то есть отображение, которое мы строим, оставляет элементы основного поля неподвижными. Если же элемент $\alpha \notin \mathbb{P}$, то он, очевидно, выражается через корни $\alpha_1, \alpha_2, \ldots, \alpha_n$ в виде рациональной функции с коэффициентами из основного поля

$$\alpha = f(\alpha_1, \alpha_2, \ldots, \alpha_n). \tag{14.6}$$

Сопоставим ему элемент $\sigma(\alpha)$, определённый следующим образом:

$$\sigma(\alpha) = f(\alpha_{\sigma(1)}, \alpha_{\sigma(2)}, \ldots, \alpha_{\sigma(n)}).$$

Иными словами, $\sigma(\alpha)$ получается, если в формуле (14.6) переставить корни $\alpha_1, \alpha_2, \ldots, \alpha_n$ согласно подстановке σ. Данное определение нуждается, правда, в пояснении. Дело в том, что запись вида (14.6) для элемента α может оказаться не единственной. Нужно проверить поэтому, что образ $\sigma(\alpha)$ не зависит от того, в какой форме элемент α выражен через корни. Если

$$\alpha = f_1(\alpha_1, \alpha_2, \ldots, \alpha_n)$$

— другая запись, выражающая элемент α в виде рациональной функции от $\alpha_1, \alpha_2, \ldots, \alpha_n$, то для проверки «корректности» нашего определения мы должны убедиться, что из равенства

$$f(\alpha_1, \alpha_2, \ldots, \alpha_n) = f_1(\alpha_1, \alpha_2, \ldots, \alpha_n) \tag{14.7}$$

следует также, что и

$$f(\alpha_{\sigma(1)}, \alpha_{\sigma(2)}, \ldots, \alpha_{\sigma(n)}) = f_1(\alpha_{\sigma(1)}, \alpha_{\sigma(2)}, \ldots, \alpha_{\sigma(n)}). \tag{14.8}$$

Однако равенство (14.7) является соотношением между корнями. Умножая, если угодно, обе его части на $f_1(\alpha_1, \alpha_2, \ldots, \alpha_n)^{-1}$, можно привести это соотношение к виду (14.3)

$$f_1(\alpha_1, \alpha_2, \ldots, \alpha_n)^{-1} \cdot f(\alpha_1, \alpha_2, \ldots, \alpha_n) = 1. \tag{14.9}$$

Поскольку подстановки из группы Галуа всякое верное соотношение между корнями переводят в верное соотношение, то справедливо равенство

$$f_1(\alpha_{\sigma(1)}, \alpha_{\sigma(2)}, \ldots, \alpha_{\sigma(n)})^{-1} \cdot f_1(\alpha_{\sigma(1)}, \alpha_{\sigma(2)}, \ldots, \alpha_{\sigma(n)}) = 1,$$

а значит, и (14.8).

Ясно, что отображения поля $\mathbb{F}$, порождённые двумя взаимно обратными подстановками σ и σ^{-1} будут также взаимно обратными друг другу. Раз так, то всякое отображение поля $\mathbb{F}$, полученное указанным выше образом, является взаимно однозначным.

И последнее, если

$$\alpha = f(\alpha_1, \alpha_2, \ldots, \alpha_n) \quad \text{и} \quad \beta = \varphi(\alpha_1, \alpha_2, \ldots, \alpha_n),$$

то, очевидно,

$$\begin{aligned} \sigma(\alpha + \beta) &= f(\alpha_{\sigma(1)}, \alpha_{\sigma(2)}, \ldots, \alpha_{\sigma(n)}) + \varphi(\alpha_{\sigma(1)}, \alpha_{\sigma(2)}, \ldots, \alpha_{\sigma(n)}) = \\ &= \sigma(\alpha) + \sigma(\beta), \\ \sigma(\alpha\beta) &= f(\alpha_{\sigma(1)}, \alpha_{\sigma(2)}, \ldots, \alpha_{\sigma(n)})\, \varphi(\alpha_{\sigma(1)}, \alpha_{\sigma(2)}, \ldots, \alpha_{\sigma(n)}) = \\ &= \sigma(\alpha)\, \sigma(\beta). \end{aligned}$$

Всё сказанное означает, что отображение $\sigma\colon \mathbb{F} \to \mathbb{F}$ является автоморфизмом поля $\mathbb{F}$. Напомним,что автоморфизм поля — это его изоморфизм на себя. Итак,

> *каждой подстановке группы Галуа уравнения* (14.5) *соответствует автоморфизм поля разложения* $\mathbb{F}$ *этого уравнения, оставляющий на месте всякий элемент основного поля.*

Не составляет труда установить, что верно и обратное утверждение, а именно:

> *всякому автоморфизму поля разложения*
>
> $$\mathbb{F} = \mathbb{P}(\alpha_1, \alpha_2, \ldots, \alpha_n)$$
>
> *данного уравнения, оставляющему неподвижными элементы основного поля* $\mathbb{P}$, *отвечает подстановка из группы Галуа этого уравнения.*

Действительно, пусть α_i — корень уравнения (14.5), то есть справедливо равенство

$$p(\alpha_i) = a_0\alpha_i^n + a_1\alpha_i^{n-1} + \cdots + a_{n-1}\alpha_i + a_n = 0,$$

и $\alpha_i' = \sigma(\alpha_i)$ — образ α_i при автоморфизме σ. Тогда, с одной стороны,

$$\sigma(p(\alpha_i)) = \sigma(0) = 0,$$

а с другой —

$$\sigma(p(\alpha_i)) = a_0\alpha_i'^n + a_1\alpha_i'^{n-1} + \cdots + a_{n-1}\alpha_i' + a_n = 0.$$

Следовательно,

$$a_0\alpha_i'^n + a_1\alpha_i'^{n-1} + \cdots + a_{n-1}\alpha_i' + a_n = 0,$$

и значит, α_i' — также корень уравнения (14.5). Таким образом, всякому автоморфизму σ отвечает подстановка на множестве корней уравнения, и очевидно, такая, которая сохраняет соотношения между корнями. В итоге мы получили несколько иное и часто более удобное описание группы уравнения:

> *группа Галуа уравнения* (14.5) *с коэффициентами из основного поля* $\mathbb{P}$ *состоит из всех автоморфизмов поля разложения* $\mathbb{F}$ *этого уравнения, которые оставляют неподвижными элементы основного поля.*

Будем далее её обозначать $\operatorname{Gal}\mathbb{F}/\mathbb{P}$, называя её также группой Галуа поля разложения или группой Галуа $\mathbb{F}$ над $\mathbb{P}$.

Вообще говоря, не всякое расширение $\mathbb{F}$ основного поля $\mathbb{P}$ является полем разложения некоторого многочлена (примером может служить $\mathbb{Q}(\sqrt[3]{2})$). Особая важность полей разложения и их групп Галуа определяется прежде всего следующими двумя обстоятельствами.

А. *Если* $t(x)$ *— произвольный многочлен, неприводимый над полем* $\mathbb{P}$*, и поле разложения* $\mathbb{F}$ *содержит хотя бы один его корень, то это поле содержит и все другие корни многочлена* $t(x)$.

B. *Если ξ — один из корней неприводимого над $\mathbb{P}$ многочлена $t(x)$, то все другие его корни можно получить, действуя на ξ автоморфизмами из группы Галуа* $\operatorname{Gal}\mathbb{F}/\mathbb{P}$. *Иначе говоря, если η — любой корень $t(x)$, то существует такой автоморфизм $\sigma \in \operatorname{Gal}\mathbb{F}/\mathbb{P}$, что $\eta = \sigma(\xi)$.*

Рассуждения, которые устанавливают справедливость утверждений A и B, составляют суть теории Галуа. Вот эти рассуждения.

Пусть $\mathbb{F} = \mathbb{P}(\alpha_1, \alpha_2, \ldots, \alpha_n)$ — поле разложения многочлена $p(x)$ и ξ — один из корней неприводимого над $\mathbb{P}$ многочлена $t(x)$. Тогда, как известно, $\xi = g(\alpha_1, \alpha_2, \ldots, \alpha_n)$, где $g(x_1, x_2, \ldots, x_n)$ — рациональная функция от переменных $x_1, x_2, \ldots, x_n$. Пусть $\sigma \in$ $\in S_n$ — произвольная подстановка

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & \ldots & n \\ i_1 & i_2 & i_3 & \ldots & i_n \end{pmatrix}.$$

Определим функцию $g_\sigma(x_1, x_2, \ldots, x_n)$, полагая

$$g_\sigma(x_1, x_2, \ldots, x_n) = g(x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(n)}) = g(x_{i_1}, x_{i_2}, \ldots, x_{i_n}),$$

иными словами, производя в исходной функции $g(x_1, x_2, \ldots, x_n)$ перестановку аргументов, соответствующую подстановке σ. Для удобства занумеруем все подстановки из S_n:

$$e = \sigma_1, \sigma_2, \sigma_3, \ldots, \sigma_k, \ldots, \sigma_{n!}.$$

Для каждой из них будем иметь тогда функцию

$$g_{\sigma_k}(x_1, x_2, \ldots, x_n), \quad k = 1, 2, \ldots, n!,$$

определённую выше (при этом заметим, что $g_{\sigma_1}(x_1, x_2, \ldots, x_n) =$ $= g(x_1, x_2, \ldots, x_n)$). Обозначим, наконец, $g_{\sigma_k}(\alpha_1, \alpha_2, \ldots, \alpha_n) =$ $= \xi_k$. Рассмотрим теперь многочлен

$$G(x) = \prod_{k=1}^{n!}(x - \xi_k) = \prod_{k=1}^{n!}(x - g_{\sigma_k}(\alpha_1, \alpha_2, \ldots, \alpha_n)).$$

Его коэффициенты являются элементарными симметрическими многочленами от ξ_k. Нетрудно понять (вспомните аналогичные рассуждения Лагранжа из главы 3), что тогда эти коэффициенты будут также симметрическими рациональными функциями от корней $(\alpha_1, \alpha_2, \ldots, \alpha_n)$ многочлена $p(x)$. Отсюда следует, что коэффициенты многочлена $G(x)$ выражаются рационально через коэффициенты $p(x)$ и поэтому принадлежат основному полю $\mathbb{P}$.

Построенный многочлен имеет общий корень ξ с неприводимым многочленом $t(x)$, но такое возможно, в силу неприводимости, лишь тогда, когда $t(x)$ является делителем $G(x)$. Значит, все корни $t(x)$ содержатся среди элементов ξ_k, и следовательно, принадлежат полю разложения $\mathbb{F}$. Этим доказано утверждение А.

Для доказательства предложения В удобней всего воспользоваться теоремой о примитивном элементе (дополнение 12 к главе 6). Попутно мы получим очень простое описание автоморфизмов группы $\operatorname{Gal}\mathbb{F}/\mathbb{P}$. Согласно упомянутой теореме всякое конечное расширение $\mathbb{F} \supseteq \mathbb{P}$ содержит такой элемент θ, что $\mathbb{F} = \mathbb{P}(\theta)$. Если m — степень его минимального многочлена $\mu(x)$, то, как отмечалось в дополнении 9 главы 6, всякий элемент поля $\mathbb{F}$ имеет однозначную запись вида

$$\sum_{i=0}^{m-1} c_i \theta^i, \quad c_i \in \mathbb{P}. \tag{14.10}$$

Рассмотрим наряду с θ и все другие корни многочлена $\mu(x)$: $\theta_1 = \theta, \theta_2, \ldots, \theta_m$ (напомним, что все они согласно А принадлежат полю $\mathbb{F}$).

Понятно, что $\mathbb{P}(\theta_k) \subseteq \mathbb{P}(\theta)$, но, кроме того, степени обоих полей над $\mathbb{P}$ совпадают. Поэтому $\mathbb{P}(\theta_k) = \mathbb{P}(\theta)$. Значит, для любого фиксированного $k = 1, 2, \ldots, m$ множество элементов вида

$$\sum_{i=0}^{m-1} c_i \theta_k^i, \quad c_i \in \mathbb{P}, \tag{14.11}$$

так же, как и (14.10), исчерпывает всё поле $\mathbb{F}$.

Рассмотрим для каждого $k = 1, 2, \ldots, m$ отображение $\sigma_k \colon \mathbb{F} \to \mathbb{F}$, полагая

$$\sigma_k \left(\sum_{i=0}^{m-1} c_i \theta^i \right) = \sum_{i=0}^{m-1} c_i \theta_k^i. \tag{14.12}$$

Очевидно, отображение σ_k взаимно однозначно. Легко видеть также и то, что оно является автоморфизмом. Последнее объясняется тем, что правила сложения и умножения элементов вида (14.11) такие же, как и для элементов (14.10). Оказывается, построенные отображения σ_k исчерпывают все автоморфизмы группы Галуа $\operatorname{Gal} \mathbb{F}/\mathbb{P}$. В самом деле, если $\sigma \in \operatorname{Gal} \mathbb{F}/\mathbb{P}$, то $\sigma(\theta)$ является корнем многочлена $\mu(x)$, а потому совпадает с одним из элементов θ_k. Но тогда из определения автоморфизма следует равенство (14.12). Это и значит, что произвольный автоморфизм σ совпадает с одним из описанных выше автоморфизмов σ_k и

$$\operatorname{Gal} \mathbb{F}/\mathbb{P} = \{\sigma_1, \sigma_2, \ldots, \sigma_m\}.$$

Заметим, что попутно мы установили такой факт.

С. *Порядок группы Галуа* $\operatorname{Gal} \mathbb{F}/\mathbb{P}$ *совпадает со степенью расширения* $\mathbb{F}$ *над* $\mathbb{P}$. *Он равен также степени многочлена* $\mu(x)$, *минимального для примитивного элемента* θ.

Утверждение В доказывается теперь так. Подействовав на корень ξ произвольного неприводимого многочлена $t(x)$ автоморфизмами из группы Галуа, получим элементы

$$\xi_1 = \xi = \sigma_1(\xi), \quad \xi_2 = \sigma_2(\xi), \ldots, \quad \xi_m = \sigma_m(\xi). \tag{14.13}$$

Среди всех перечисленных элементов не все обязательно различны. Все они являются корнями $t(x)$, при этом, если $\xi = \sum_{i=0}^{m-1} a_i \theta^i$, то $\xi_k = \sum_{i=0}^{m-1} a_i \theta_k^i$. Нам нужно убедиться, что элементами ξ_k исчерпываются все корни $t(x)$. Рассмотрим с этой целью многочлен

$$f(x) = (x - \xi_1)(x - \xi_2) \ldots (x - \xi_m),$$

также имеющий указанные элементы своими корнями. Повторив рассуждения, проведённые ранее для многочлена $G(x)$, мы убеждаемся, что коэффициенты $f(x)$ являются симметрическими функциями от θ_1, θ_2, ..., θ_m, а значит, принадлежат основному полю. Неприводимый многочлен $t(x)$ является тогда делителем $f(x)$, и потому не имеет корней, отличных элементов (14.13). Утверждение B доказано.

Установленные факты ощутимо приближают нас к цели. Так, утверждение A показывает, что если вспомогательный многочлен $t(x)$ имеет хотя бы один корень из поля $\mathbb{F}$, то тогда в $\mathbb{F}$ целиком содержится и всё поле разложения $\mathbb{K}$ этого многочлена. Из утверждений A и B также следует, что всякий автоморфизм $\sigma \in \in \operatorname{Gal}\mathbb{F}/\mathbb{P}$ является одновременно автоморфизмом любого другого поля разложения $\mathbb{K} \subseteq \mathbb{F}$.

В самом деле, пусть $a \in \mathbb{K}$ и $\mu_a(x)$ — его минимальный многочлен. Согласно B элемент $\sigma(a)$, так же, как и a, является корнем этого многочлена, но тогда из A вытекает, что $\sigma(a) \in \mathbb{K}$. Это и нужно было установить. Рассмотрим теперь цепочку полей

$$\mathbb{F} \supseteq \mathbb{K} \supseteq \mathbb{P}$$

и соответствующие группы Галуа $\operatorname{Gal}\mathbb{F}/\mathbb{K}$ и $\operatorname{Gal}\mathbb{K}/\mathbb{P}$.

Возникает вопрос, как связаны эти группы с группой $\operatorname{Gal}\mathbb{F}/\mathbb{P}$ исходного уравнения? Ясно, во-первых, что всякий автоморфизм h, принадлежащий $\operatorname{Gal}\mathbb{F}/\mathbb{K}$, также принадлежит и $\operatorname{Gal}\mathbb{F}/\mathbb{P}$ (действительно, h оставляет неподвижным все элементы поля $\mathbb{K}$, а значит, и поля $\mathbb{P}$). Следовательно, $\operatorname{Gal}\mathbb{F}/\mathbb{K}$ является подгруппой в группе $\operatorname{Gal}\mathbb{F}/\mathbb{P}$. Заметим, что левые смежные классы по этой подгруппе состоят из всех подстановок, одинаково преобразующих каждый элемент $a \in \mathbb{K}$ (сравните с главой 3). Действительно, $\sigma_k h(a) = \sigma_k(a)$ для всякого автоморфизма $h \in \operatorname{Gal}\mathbb{F}/\mathbb{K}$. Легко видеть, что верно и обратное: всякий автоморфизм σ такой, что $\sigma(a) = \sigma_k(a)$, содержится в смежном классе $\{\sigma_k h \mid h \in \in \operatorname{Gal}\mathbb{F}/\mathbb{K}\}$ (проверьте). Но оказывается, что таким же свойством обладают и правые смежные классы $\{h\sigma_k \mid h \in \operatorname{Gal}\mathbb{F}/\mathbb{K}\}$. Ведь $\sigma(a) \in \mathbb{K}$ (это было установлено чуть выше), и следовательно, $h\sigma_k(a) = \sigma_k(a)$. Таким образом, каждый левый смежный

класс $\{\sigma_k h \mid h \in \operatorname{Gal}\mathbb{F}/\mathbb{K}\}$ совпадает с соответствующим правым смежным классом $\{h\sigma_k \mid h \in \operatorname{Gal}\mathbb{F}/\mathbb{K}\}$, — оба они состоят из всех автоморфизмов, преобразующих всякий элемент $a \in \mathbb{K}$ в элемент $\sigma_k(a)$. Иными словами, мы установили следующее.

Утверждение 1. *Если $\mathbb{K} \subseteq \mathbb{F}$ — поле разложения некоторого вспомогательного многочлена $t(x)$ с коэффициентами из $\mathbb{P}$, то группа Галуа $\operatorname{Gal}\mathbb{F}/\mathbb{K}$ является нормальным делителем в группе $\operatorname{Gal}\mathbb{F}/\mathbb{P}$ исходного уравнения.*

Вместе с тем выяснено и нечто большее. Ранее мы уже отмечали, что всякому элементу $\sigma \in \operatorname{Gal}\mathbb{F}/\mathbb{P}$ отвечает автоморфизм промежуточного поля $\mathbb{K}$, то есть элемент группы $\operatorname{Gal}\mathbb{K}/\mathbb{P}$. Правда, разным элементам σ и σ' из группы $\operatorname{Gal}\mathbb{F}/\mathbb{P}$ вполне может соответствовать один и тот же автоморфизм этого промежуточного поля. Предыдущие рассуждения показывают, что так будет в том и только том случае, когда σ и σ' принадлежат одному и тому же смежному классу по подгруппе $\operatorname{Gal}\mathbb{F}/\mathbb{K}$. Следовательно, между указанными смежными классами (они, как мы знаем, являются элементами факторгруппы $\operatorname{Gal}\mathbb{F}/\mathbb{P}/\operatorname{Gal}\mathbb{F}/\mathbb{K}$) и автоморфизмами поля $\mathbb{K}$ можно установить взаимно однозначное соответствие. Смежному классу $\{\sigma h \mid h \in \operatorname{Gal}\mathbb{F}/\mathbb{K}\}$ соответствует автоморфизм $\overline{\sigma}$ поля $\mathbb{K}$, определённый так: $\overline{\sigma}(a) = \sigma(a)$ для $a \in \mathbb{K}$. Из этого определения ясно, что произведению смежных классов $\{\sigma h \mid h \in \operatorname{Gal}\mathbb{F}/\mathbb{K}\}$ и $\{\tau h \mid h \in \operatorname{Gal}\mathbb{F}/\mathbb{K}\}$ будет отвечать произведение соответствующих им автоморфизмов $\overline{\sigma}$ и $\overline{\tau}$. Значит, описанное соответствие является изоморфизмом. Окончательный вывод таков:

Утверждение 2. *Группа Галуа $\operatorname{Gal}\mathbb{K}/\mathbb{P}$ промежуточного поля разложения $\mathbb{K}$ изоморфна факторгруппе $\operatorname{Gal}\mathbb{F}/\mathbb{P}/\operatorname{Gal}\mathbb{F}/\mathbb{K}$.*

Утверждения 1 и 2 дают нам точное описание того, каким образом группа Галуа $\operatorname{Gal}\mathbb{F}/\mathbb{P}$ исходного уравнения сводится к более простым группам $\operatorname{Gal}\mathbb{F}/\mathbb{K}$ и $\operatorname{Gal}\mathbb{K}/\mathbb{P}$ в том случае, когда существует вспомогательное уравнение $t(x) = 0$, поле разложения $\mathbb{K}$ которого является промежуточным между $\mathbb{P}$ и $\mathbb{F}$.

Можно доказать и факт, в некотором смысле обратный к утверждениям 1 и 2, а именно:

Утверждение 3. *Пусть* $\mathbb{F} \supseteq \mathbb{P}$ *— поле разложения многочлена* $f(x)$ *и* $G = \operatorname{Gal}\mathbb{F}/\mathbb{P}$ *— его группа Галуа. Тогда, если* H *— нормальный делитель в* G*, то существует такое промежуточное подполе* $\mathbb{K}$ $(\mathbb{P} \subseteq \subseteq \mathbb{K} \subseteq \mathbb{F})$*, также являющееся полем разложения некоторого многочлена, что*

$$H = \operatorname{Gal}\mathbb{F}/\mathbb{K} \quad \textit{и} \quad G/H \cong \operatorname{Gal}\mathbb{K}/\mathbb{P}.$$

С доказательством этого факта, а также с обстоятельным и одновременно доступным изложением теории Галуа рекомендуем познакомиться по книге [24]. Ограничимся лишь замечанием, что в качестве поля $\mathbb{K}$, о котором говорится в утверждении, следует взять множество всех элементов из $\mathbb{F}$, которые остаются неподвижными при автоморфизмах из подгруппы H:

$$\mathbb{K} = \{\, k \in \mathbb{F} \mid h(k) = k \text{ для всякого } h \in H \,\}.$$

Как могут быть использованы только что сформулированные утверждения для ответа на вопрос о разрешимости уравнений в радикалах? Чтобы это понять, вспомним следующее. Решение вспомогательного уравнения $t_1(x) = 0$ позволяет расширить запас чисел, присоединяя его корни θ_i к исходному (основному) полю $\mathbb{P}$ и переходя к новому основному полю $\mathbb{P}_1 = \mathbb{P}(\theta_1, \ldots, \theta_k)$ — полю разложения многочлена $t_1(x)$. Причем, если хотя бы один из корней $\theta_i \in \mathbb{F}$, то и всё поле $\mathbb{P}_1 \subseteq \mathbb{F}$. Последовательное решение нескольких вспомогательных уравнений (корни которых принадлежат $\mathbb{F}$)

$$t_1(x) = 0; \quad t_2(x) = 0; \quad \ldots; \quad t_r(x) = 0$$

равносильно последовательному присоединению корней, сначала многочлена $t_1(x)$, затем $t_2(x)$ и т. д. Это приводит к цепочке расширений

$$\mathbb{P} = \mathbb{P}_0 \subseteq \mathbb{P}_1 \subseteq \mathbb{P}_2 \subseteq \ldots \subseteq \mathbb{P}_r. \qquad (14.14)$$

В ней каждое поле $\mathbb{P}_i$ является полем разложения соответствующего многочлена $t_i(x)$, для которого основным полем является предыдущий член цепочки $\mathbb{P}_{i-1}$. При этом все поля $\mathbb{P}_i \subseteq \mathbb{F}$.

Если оказалось так, что все корни уравнения $f(x) = 0$ можно рационально выразить через корни вспомогательных уравнений, то это означает, что последний член цепи (14.14) совпадает с полем разложения многочлена $f(x)$, то есть

$$\mathbb{P} = \mathbb{P}_0 \subseteq \mathbb{P}_1 \subseteq \mathbb{P}_2 \subseteq \ldots \subseteq \mathbb{P}_r = \mathbb{F}. \tag{14.15}$$

Из утверждений 1, 2, 3 непосредственно вытекает, что *наличие такой цепи равносильно существованию в группе Галуа* $G = \operatorname{Gal}\mathbb{F}/\mathbb{P}$ *убывающей последовательности подгрупп*

$$G = G_0 \supseteq G_1 \supseteq G_2 \supseteq \ldots \supseteq G_i \supseteq \ldots \supseteq G_r = \{e\}. \tag{14.16}$$

При этом G_i — нормальный делитель в G_{i-1}, $G_i = \operatorname{Gal}\mathbb{F}/\mathbb{P}_i$ и $G_{i-1}/G_i \cong \operatorname{Gal}\mathbb{P}_i/\mathbb{P}_{i-1}$.

Этот общий результат даёт решение исходной задачи о разрешимости уравнения в радикалах. Действительно, разрешимость в радикалах равносильна тому, что корни заданного уравнения могут быть выражены рационально через корни последовательно решаемых двучленных уравнений

$$x^{n_1} - a_1 = 0, \quad x^{n_2} - a_2 = 0, \ldots, \quad x^{n_r} - a_r = 0, \tag{14.17}$$

где a_1 принадлежит основному полю $\mathbb{P}$, a_2 — полю разложения многочлена $x^{n_1} - a_1$ и т. д. С устройством группы Галуа H двучленного уравнения $x^n - a$ разобраться уже не так сложно. Она, как можно показать, либо циклическая, либо содержит циклическую нормальную подгруппу H_1, факторгруппа по которой H/H_1 также циклическая. Значит, если цепи (14.14) и (14.15) соответствуют решаемым двучленным уравнениям (14.17), то каждая из факторгрупп G_{i-1}/G_i устроена так, как было только что сказано. Дело техники теперь доказать, что для разрешимости уравнения $f(x) = 0$ в радикалах необходимо и достаточно, чтобы в группе Галуа $\operatorname{Gal}\mathbb{F}/\mathbb{P}$ этого уравнения существовала убывающая цепь подгрупп

$$G = G_0 \supseteq G_1 \supseteq G_2 \supseteq \ldots \supseteq G_i \supseteq \ldots \supseteq G_s = \{e\}$$

такая, что G_i — нормальный делитель в G_{i-1}, и каждая факторгруппа G_{i-1}/G_i является циклической. Группы указанного строения так и называются *разрешимыми*, а высказанное здесь утверждение (это и есть знаменитый критерий разрешимости в радикалах) короче можно сформулировать так:

уравнение $f(x) = 0$ разрешимо в радикалах тогда и только тогда, когда его группа Галуа разрешима.

Указав этот критерий, Галуа дал также принципиальный способ для отыскания группы любого уравнения. Он выяснил кроме того, что для каждого n существует такое уравнение степени n, группа которого совпадает с симметрической группой S_n (т. е. такова же, что и для общего уравнения). Уже для $n = 5$, как и для всех бóльших значений n, группа S_n оказывается неразрешимой[1]. Отсюда и следует, что для каждого $n \geqslant 5$ существуют алгебраические уравнения, не разрешимые в радикалах.

Теория Галуа уже при своём создании имела намного более общий характер, чем та задача, ради решения которой она была построена. Это отмечал и сам автор. Вспомните его высказывание, приведённое в начале этой главы. Действительно, для любого уравнения, быть может и не разрешимого в радикалах, можно ставить вопрос, к цепи каких наиболее простых уравнений сводится его решение. Соответствие между подполями поля разложения и подгруппами группы Галуа, о которых (оно называется соответствием Галуа) говорится в утверждениях 1, 2 и 3, позволяет отвечать на вопросы подобного типа.

Мы рассказали здесь о теории Галуа числовых полей, однако основные её положения могут быть распространены на произвольные поля и другие алгебраические объекты. В наиболее широком смысле теория Галуа — так её понимают современные алгебраисты — это теория, изучающая те или иные математические структуры на основе их групп автоморфизмов. Хотим подчеркнуть также, что наряду с другими учёными, весомый вклад в развитие этой теории внесли российские математики. Среди них должны быть особо отмечены замечательные алгебраисты

[1] Доказательство этого можно найти, например, в [24] или в [5].

Н. Г. Чеботарёв (1895–1947) и И. Р. Шафаревич (род. в 1923 г.). Так, один из глубоких результатов, полученных Шафаревичем, связан с так называемой обратной задачей теории Галуа, и поныне до конца не решённой: всякая ли группа подстановок служит группой Галуа некоторого алгебраического уравнения? И. Р. Шафаревич доказал, что это, во всяком случае, верно для разрешимой группы подстановок.

ГЛАВА 15

ПОЛЯ ГАЛУА И КОДИРОВАНИЕ

Кодирование — одна из тех сравнительно новых прикладных областей, где алгебра заявляет о себе в полный голос. В решении земных (порой и космических) проблем сегодняшнего дня, связанных с передачей информации, не обойтись без синтеза различных разделов алгебры, в том числе теории групп, полей, линейной алгебры. Читатель, знакомый с популярной книгой [2], видимо, может составить об этом пускай и неполное, но всё же достаточно широкое представление. Избегая излишних повторений, мы хотим здесь несколько глубже вникнуть в задачи построения и реализации помехоустойчивых кодов и более выпукло осветить роль алгебры в решении этих задач. Известно, что при передаче информации по линиям связи символам или буквам передаваемого текста, а иногда и целым словам или фразам (сообщениям) сопоставляются определённые комбинации сигналов (электрических, магнитных и т. д.), называемые кодами, или кодовыми словами.

Читателю, видимо, знаком код, который связан с именем изобретателя телеграфного аппарата Сэмюэля Морзе — азбука Морзе. В этом коде каждой букве или цифре сопоставляется своя последовательность из кратковременных (называемых точками) и длительных (тире) импульсов тока, разделяемых паузами. Другой код, столь же широко распространённый в телеграфии (код Бодо), использует для кодирования два элементарных сигнала — импульс и паузу, при этом сопоставляемые буквам кодовые слова состоят из пяти таких сигналов. Коды, использующие два раз-

личных элементарных сигнала, называются двоичными. Удобно бывает, отвлекаясь от их физического смысла, обозначать эти два сигнала символами 0 и 1. Тогда кодовые слова можно представлять как последовательности из нулей и единиц.

Нередко наряду с двоичными кодами применяют и коды, использующие не два, а большее число элементарных сигналов, или иначе, кодовых символов. Их число q называют основанием кода, а множество кодовых символов — алфавитом.

Операцию перевода сообщений в кодовые слова называют кодированием, а обратную операцию, восстанавливающую по принятым сигналам передаваемые сообщения, — декодированием. Казалось бы, ничего особенно сложного тут нет, однако это не совсем так. Дело в том, что в реальных каналах связи, по которым происходит передача информации, всегда присутствуют помехи или, как говорят иначе, «шумы», природа которых может быть самой различной.

Накладываясь на передаваемые сигналы, они вызывают их искажения, и принятая адресатом информация может поэтому содержать ошибки. Действие шумов, конечно, можно ослабить, но устранить его полностью нельзя. Как же в таком случае обеспечить надёжность передачи? Ясно, что необходим какой-то контроль передаваемой информации. Самый незатейливый способ его организовать состоит в том, что каждый информационный символ повторяется несколько раз, скажем, символ 0 заменяется блоком из n нулей, а символ 1 — блоком из n единиц. При декодировании n-буквенного блока, содержащего, быть может, ошибочные символы, решение нужно принимать «большинством голосов». Если в принятом блоке нулей больше, чем единиц, то он декодируется как $00\ldots0$ (т. е. считается, что был послан нулевой символ), в противном случае — как $11\ldots1$. Такое правило декодирования позволяет верно восстановить посланные символы, если помехи в канале искажают менее половины символов в каждом блоке. Указанный код (его называют *кодом с повторением*) не имеет большого практического значения. Для эффективного контроля может потребоваться слишком много лишних символов. Однако он содержит в себе основной принцип построения помехоустойчивых кодов: для того, чтобы выявлять возможные

при передаче ошибки, надо вводить дополнительные избыточные символы. Эти новые символы уже не несут информации о передаваемых сообщениях, но могут дать информацию о происшедших ошибках, иными словами, они как раз и контролируют правильность передачи кодового слова. Вводимые дополнительные символы в отличие от информационных называют проверочными (или контрольными).

Рассмотрим два примера, иллюстрирующих сказанное.

1. Код с проверкой на чётность. Пусть $\alpha_1\alpha_2\ldots\alpha_n$ — двоичное кодовое слово ($\alpha_i = 0; 1$). Добавим к нему всего лишь один проверочный символ α_{n+1}, положив

$$\alpha_{n+1} = \alpha_1 + \alpha_2 + \ldots + \alpha_n \pmod 2.$$

Проверочный символ α_{n+1} будет равен нулю, если в слове $\alpha_1\alpha_2\ldots\alpha_n$ содержится чётное число единиц, и единице — в противном случае. Например, присоединяя таким способом проверочный символ к слову 1010, получаем слово 10100, а из слова 1110 получим слово 11101.

Очевидно, все удлинённые кодовые слова $\alpha_1\alpha_2\ldots\alpha_n\alpha_{n+1}$ содержат чётное число единиц, то есть всегда

$$\alpha_1 + \alpha_2 + \ldots + \alpha_n + \alpha_{n+1} = 0 \pmod 2. \qquad (15.1)$$

Допустим, что в процессе передачи в удлинённое слово $\alpha_1\alpha_2\ldots$ $\ldots\alpha_n\alpha_{n+1}$ вкралась одна ошибка (или даже любое нечётное число ошибок). Тогда в искажённом слове $\alpha'_1\alpha'_2\ldots\alpha'_n\alpha'_{n+1}$ число единиц станет нечётным. Это и служит указанием на искажение в передаче слова. В конечном итоге всё сводится к проверке соотношения (15.1) для символов принятого слова, что легко сделает простейшее вычислительное устройство — сумматор по модулю 2. Итак, правило приёма следующее: если равенство (15.1) выполняется, считаем, что сообщение передано правильно, в противном случае отмечаем, что произошла ошибка, и, когда это возможно, требуем повторить передачу кодового слова. Понятно, что иначе ошибки не исправить. Например, если принять «неправильное слово» 11100, то одинаково возможно, что было послано любое

из кодовых слов

$$01100,\ 10100,\ 11000,\ 11110,\ 11101.$$

Каждый из перечисленных случаев соответствует одной ошибке, а ведь ошибки могли произойти даже и в большем числе символов. Ещё большая неприятность подстерегает нас в случае двойной ошибки или вообще чётного числа ошибок. Ведь тогда соотношение (15.1) не нарушится, и мы воспримем искажённое слово как верное.

Описанный код, который и называют кодом с проверкой на чётность, позволяет, следовательно, обнаружить любое нечётное число ошибок, но «пропускает» искажения, если число ошибок чётно.

В реальных каналах связи, как правило, приходится считаться с возможностью ошибок более чем в одном символе, поэтому в чистом виде код с проверкой на чётность применяется редко. Гораздо чаще применяют коды с несколькими проверочными символами. Они позволяют не только обнаруживать, но и исправлять ошибки, и не только одиночные, но и кратные.

2. Код Хемминга — важный пример кода, исправляющего любые одиночные ошибки. Для простоты рассмотрим случай, когда для кодирования используются двоичные слова $\alpha_1\alpha_2\alpha_3\alpha_4$ длины 4 (понятно, что таких слов имеется 16). Мы уже представляем, что одного проверочного символа для исправления ошибок (даже одиночных) недостаточно. Нетрудно убедиться, что не хватит для этой цели и двух проверочных символов. Попробуем обойтись тремя, действуя следующим образом.

Добавим к каждому слову $\alpha_1\alpha_2\alpha_3\alpha_4$ три символа α_5, α_6, α_7 так, чтобы выполнялись следующие соотношения:

$$\begin{cases} \alpha_4 + \alpha_5 + \alpha_6 + \alpha_7 = 0 \\ \alpha_2 + \alpha_3 + \alpha_6 + \alpha_7 = 0 \\ \alpha_1 + \alpha_3 + \alpha_5 + \alpha_7 = 0 \end{cases} \tag{15.2}$$

(здесь и далее все равенства берутся по модулю 2). Можно сообразить, что этими соотношениями дополнительные символы

определены однозначно, причём,

$$\begin{cases} \alpha_5 = \alpha_2 + \alpha_3 + \alpha_4 \\ \alpha_6 = \alpha_1 + \alpha_3 + \alpha_4 \\ \alpha_7 = \alpha_1 + \alpha_2 + \alpha_4. \end{cases}$$

Следовательно, кодовых слов $\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5\alpha_6\alpha_7$ длины 7, символы которых удовлетворяют соотношениям (15.2), также 16. Они и образуют рассматриваемый здесь код Хемминга с 4-мя информационными символами.

Предположим теперь, что в кодовом слове кода Хемминга произошла одиночная ошибка. Выясним, как это отразится на равенствах (15.2), их левые части обозначим соответственно S_1, S_2, S_3. Понятно,что первое из них сохранится лишь в том случае, если символы α_4,α_5,α_6,α_7 переданы верно, а ошибка содержится в одном из символов α_1, α_2, α_3. Тогда $S_1 = \alpha_4 + \alpha_5 + \alpha_6 + \alpha_7 = 0$, в противном же случае $S_1 = 1$. Как бы то ни было, значение $S_2 = \alpha_2 + \alpha_3 + \alpha_6 + \alpha_7$ снова позволяет уточнить местоположение ошибки. Так, если $S_1 = 1$, а $S_2 = 0$, то ошибка содержится либо в α_4 либо в α_5. Всего имеется четыре комбинации значений S_1 и S_2. Они приведены в таблице 14 (в ней же указаны возможности для положения ошибочного символа)

Таблица 14.

S_1	S_2	Место ошибки
1	1	α_6 или α_7
1	0	α_4 или α_5
0	1	α_2 или α_3
0	0	нет ошибки или α_1

В каждом из четырёх случаев нужно выбрать одну из двух возможностей. Сделать это позволяет значение суммы

$$S_3 = \alpha_1 + \alpha_3 + \alpha_5 + \alpha_7.$$

Итак, значения S_1, S_2, S_3 позволяют либо установить, что ошибки нет (если $S_1 = S_2 = S_3 = 0$, и значит, принятое слово удо-

влетворяет всем проверочным соотношениям), либо однозначно указать её место. Отметим особо, что если произошла одиночная ошибка, то её положение указывается числом с двоичной записью $S_1S_2S_3$. Пусть, например, $S_1 = 1$, $S_2 = 1$, $S_3 = 0$. Согласно таблице 14 ошибка допущена в шестом или седьмом символе; поскольку $S_3 = 0$, она — в шестом символе, но $S_1S_2S_3 = 110$ как раз и есть двоичная запись числа 6. В общем случае кодовое слово двоичного кода Хемминга, позволяющего исправить любую одиночную ошибку, имеют длину $2^m - 1$ (m — натуральное). Для определения положения ошибки тогда уже нужно m проверок, то есть m проверочных символов. Оставшиеся $2^m - 1 - m$ символов являются информационными. Проверки строятся по аналогии с рассмотренным случаем, а их значения, как и выше, образуют номер положения ошибки. Информацию об этом, а также о расширенном коде Хемминга, не только исправляющем одиночные, но и обнаруживающем двойные ошибки можно найти в уже упомянутой книге [2].

Приведённые примеры обозначают общий алгебраический подход к построению помехоустойчивых двоичных кодов. Каждое двоичное слово $x_1x_2\ldots x_n$ можно отождествить с вектором $(x_1, x_2, \ldots, x_n)$ n-мерного координатного пространства L_n над полем $\mathbb{Z}_2 = \{0, 1\}$. Те из векторов, которые соответствуют кодовым словам, будем называть кодовыми векторами. Кодовые векторы выделяются среди прочих тем, что их координаты обязаны удовлетворять некоторой системе проверочных соотношений, которая, как и в рассмотренных примерах, представляет из себя систему линейных уравнений

$$\begin{cases} b_{11}x_1 + b_{12}x_2 + \ldots + b_{1n}x_n = 0 \\ b_{21}x_1 + b_{22}x_2 + \ldots + b_{2n}x_n = 0 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ b_{m1}x_1 + b_{m2}x_2 + \ldots + b_{mn}x_n = 0 \end{cases} \tag{15.3}$$

с коэффициентами b_{ij} из $\mathbb{Z}_2$. Соотношения (15.3) называют проверочными.

Введённые понятия имеют смысл не только для двоичных кодов. Они легко обобщаются и на тот случай, когда кодовый

алфавит является произвольным конечным полем $\mathbb{F}$. Только тогда кодовые векторы являются, понятно, элементами векторного пространства над полем $\mathbb{F}$, точно так же, над этим полем рассматриваются и проверочные соотношения (15.3), которым эти векторы удовлетворяют[1]. Для кода, удовлетворяющего равенствам (15.3) выполняется следующее свойство: если векторы $(a_1, a_2, \ldots, a_n)$ и $(b_1, b_2, \ldots, b_n)$ являются кодовыми, то и их сумма $(a_1 + b_1, a_2 + b_2, \ldots, a_n + b_n)$ также является кодовым вектором. Справедливо и другое свойство решений системы (15.3): если α — элемент поля и $(a_1, a_2, \ldots, a_n)$ — решение этой системы, то и вектор $(\alpha a_1, \alpha a_2, \ldots, \alpha a_n)$ также является её решением. Оба отмеченных свойства проверяются непосредственной подстановкой в систему (15.3) векторов $(a_1 + b_1, a_2 + b_2, \ldots, a_n + b_n)$ и $(\alpha a_1, \alpha a_2, \ldots, \alpha a_n)$. Вместе эти свойства означают на языке линейной алгебры, что кодовые слова образуют линейное подпространство в пространстве L_n всех n-буквенных слов (векторов). По этой причине такие коды называют *линейными* (двоичные линейные коды называют также *групповыми*). Если кодовое пространство в пространстве L_n имеет размерность k то употребляют для большей определённости термин линейный (n, k)-код.

Конечно, для кодирования сообщений не всегда применяют линейные коды, но всё же они являются важнейшим классом помехоустойчивых кодов и на это есть немало причин. Одна из них связана с удобствами в обнаружении и исправлении ошибок, что хорошо видно из рассмотренных выше примеров. Другая причина — это возможность компактного задания кода. Действительно, в случае линейного кода нет необходимости указывать полный список кодовых слов, ведь код вполне определён системой линейных уравнений (15.3) или матрицей этой системы (*проверочной матрицей*):

$$H = \begin{pmatrix} b_{11} & b_{12} & \ldots & b_{1n} \\ b_{21} & b_{22} & \ldots & b_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ b_{n1} & b_{n2} & \ldots & b_{nn} \end{pmatrix}.$$

[1] Последнее означает, что $b_{ij} \in \mathbb{F}$, и все действия выполняются в соответствии с «арифметикой» поля $\mathbb{F}$.

В дальнейшем мы будем предполагать, что строки этой матрицы линейно независимы.

К числу других достоинств линейных кодов, которые связаны с предыдущими, относятся простые алгоритмы кодирования и декодирования, легко реализуемые электронными схемами. Вообще, можно сказать, что бурное развитие теории кодирования объясняется главным образом тем, что к линейным кодам приложим хорошо развитый аппарат линейной алгебры и теории конечных полей.

Читатель легко сообразит, как устроены проверочные матрицы ранее рассмотренных кодов. Так, для кода с проверкой на чётность имеем одно проверочное соотношение $x_1 + x_2 + \ldots + x_n = 0$, соответственно этому проверочная матрица состоит из одной строки и имеет вид

$$H = \begin{pmatrix} 1 & 1 & 1 & \ldots & 1 \end{pmatrix}.$$

Проверочная же матрица двоичного кода Хемминга длины 7 выглядит так (сравните с (15.2)):

$$H = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}. \tag{15.4}$$

Другой способ матричного задания линейного кода основывается на том, что во всяком подпространстве линейного пространства можно выбрать базис, то есть такую линейно независимую систему векторов, через которые линейно выражаются все векторы подпространства. Пусть

$$\begin{aligned} a_1 &= (a_{11}, a_{12}, \ldots, a_{1n}), \\ &\ldots\ldots\ldots\ldots\ldots \\ a_k &= (a_{k1}, a_{k2}, \ldots, a_{kn}) \end{aligned} \tag{15.5}$$

— базисные векторы линейного кода. Тогда всевозможные кодовые векторы исчерпываются линейными комбинациями

$$\alpha_1 a_1 + \alpha_2 a_2 + \ldots + \alpha_k a_k, \tag{15.6}$$

где коэффициенты α_i — любые элементы исходного поля.

Таким образом, система базисных векторов (15.5) полностью определяет линейный (n,k)-код. Матрица G, составленная из них,

$$G=\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \dots & a_{kn} \end{pmatrix} \tag{15.7}$$

называется порождающей матрицей кода.

Например, в качестве базисных векторов двоичного кода с проверкой на чётность могут быть взяты следующие $n-1$ векторов

$$\begin{aligned} a_1 &= (1,1,0,0,\dots,0), \\ a_2 &= (1,0,1,0,\dots,0), \\ a_3 &= (1,0,0,1,\dots,0), \\ a_{n-1} &= (1,0,0,0,\dots,1). \end{aligned}$$

Поэтому матрица

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

является порождающей матрицей этого кода.

В общем случае для отыскания порождающей матрицы кода, заданного системой проверочных соотношений (15.3), нужно фактически найти $k=n-m$ линейно независимых решений этой системы. Найденные решения, являющиеся базисными кодовыми векторами, как раз и будут строками порождающей матрицы. Пользуясь этим методом, найдём порождающую матрицу кода Хемминга длины 7 с проверочной матрицей (15.4). В данном случае требуется найти 4 независимых решения следующей системы

$$\begin{cases} x_4+x_5+x_6+x_7=0 \\ x_2+x_3+x_6+x_7=0 \\ x_1+x_3+x_5+x_7=0. \end{cases}$$

Разрешаем систему относительно неизвестных x_1, x_2, x_4

$$\begin{aligned} x_1 &= x_3 + x_5 + x_7, \\ x_2 &= x_3 + x_6 + x_7, \\ x_4 &= x_5 + x_6 + x_7. \end{aligned} \tag{15.8}$$

Неизвестным x_3, x_5, x_6, x_7 можно придавать любое из значений (0 или 1), тогда из равенств (15.8) могут быть определены остальные неизвестные. Придавая поочерёдно одному из неизвестных x_3, x_5, x_6, x_7 значение 1, а остальным 0, получим 4 решения $a_1 = (1110000)$, $a_2 = (1001100)$, $a_3 = (0101010)$, $a_4 = (1101001)$, которые, как читатель может убедиться, линейно независимы. Составленная из этих решений матрица

$$G = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix} \tag{15.9}$$

и является искомой порождающей матрицей.

При использовании линейного кода для передачи сообщений полезно знать и порождающую, и проверочную матрицу. С помощью порождающей матрицы удобно кодировать сообщения. Заметим сначала, что линейный (n, k)-код с порождающей матрицей (15.7) имеет q^k кодовых слов (любой из k коэффициентов в (15.6) можно выбрать q способами, где q — число элементов поля $\mathbb{F}$). Поэтому он позволяет закодировать все сообщения длины k в алфавите из q символов. Если $A = \alpha_1\alpha_2\ldots\alpha_k$ — такое сообщение (α_i — информационные символы), то самое удобное — сопоставить ему кодовый вектор a совпадающий с линейной комбинацией (15.6) строк порождающей матрицы. Вектор a нетрудно записать в матричном виде, используя правило умножения матриц

$$a = A \cdot G = (\alpha_1\, \alpha_2\, \ldots\, \alpha_k) \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \ldots & a_{kn} \end{pmatrix}. \tag{15.10}$$

Например, в случае (7,4)-кода Хемминга с порождающей матрицей (15.9) сообщение $A = (1011)$ кодируется кодовым словом

$$a = (1011) \cdot \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix} = (0110011).$$

Наиболее простой вид это правило приобретает для так называемых систематических линейных кодов. Линейный (n, k)-код называется систематическим, если его порождающая матрица имеет вид

$$G = \begin{pmatrix} 1 & 0 & \dots & 0 & p_{11} & \dots & p_{1m} \\ 0 & 1 & \dots & 0 & p_{21} & \dots & p_{2m} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 & p_{k1} & \dots & p_{km} \end{pmatrix},$$

то есть если она получается приписыванием к единичной матрице порядка k некоторой матрицы порядка $k \times m$. В этом случае равенство (15.10) принимает вид

$$a = (\alpha_1 \alpha_2 \dots \alpha_k) \begin{pmatrix} 1 & 0 & \dots & 0 & p_{11} & \dots & p_{1m} \\ 0 & 1 & \dots & 0 & p_{21} & \dots & p_{2m} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 & p_{k1} & \dots & p_{km} \end{pmatrix} =$$

$$= (\alpha_1 \alpha_2 \dots \alpha_k a_{k+1} \dots a_n).$$

Таким образом, первые k символов любого кодового слова как раз и оказываются информационными, а остальные — проверочными. Они связаны с информационными символами соотношениями

$$\begin{aligned} a_{k+1} &= p_{11}\alpha_1 + p_{21}\alpha_2 + \dots + p_{k1}\alpha_k \\ a_{k+2} &= p_{12}\alpha_1 + p_{22}\alpha_2 + \dots + p_{k2}\alpha_k \\ &\dots\dots\dots\dots\dots\dots\dots\dots \\ a_n &= p_{1m}\alpha_1 + p_{2m}\alpha_2 + \dots + p_{km}\alpha_k. \end{aligned}$$

Очень важно, что всякий линейный код в некотором смысле эквивалентен систематическому [2]. Что касается проверочной матрицы, то она используется в основном при декодировании полученных слов и исправлении ошибок. Об этом подробно рассказано в той же книге [2].

Уже в первый период развития теории кодирования среди линейных кодов были обнаружены и такие коды, которые позволяли исправлять не только одиночные, но и кратные ошибки. Правда это были лишь отдельные примеры, и долгое время никак не удавалось найти какой-то достаточно общий метод построения линейных кодов, исправляющих любое наперёд заданное количество ошибок в кодовом слове. Для решения этой задачи очень полезной оказалась теория конечных полей. К тому же хорошо разработанная техника вычислений в конечных полях позволила не только строить такие коды (они были названы циклическими), но и указывать удобные алгоритмы их декодирования. Особенно удачным оказалось ещё и то, что по своей структуре циклические коды оказались идеально приспособленными к реализации в современных технических устройствах.

Наше знакомство с циклическими кодами мы начнём с понятия циклического сдвига вектора. Пусть задан произвольный n-мерный вектор $a = (a_0, a_1, a_2, \ldots, a_{n-1})$ с координатами из любого поля $\mathbb{F}$ (нумерацию координат в случае циклических кодов удобно начинать с нуля). Циклическим сдвигом этого вектора назовём вектор

$$a' = (a_{n-1}, a_0, a_1, a_2, \ldots, a_{n-2}).$$

Например, для вектора (01101) последовательными циклическими сдвигами являются такие векторы:

$$(10110),\ (01011),\ (10101),\ (11010).$$

Циклическим кодом называется линейный код, который вместе с любым своим вектором содержит также и его циклический сдвиг. Иными словами, циклический код обладает следующим свойством: циклический сдвиг любого кодового вектора снова

является кодовым вектором. Циклическим является, например, код с проверкой на чётность.

Менее тривиальным примером циклического кода (проверка предоставляется читателю) является код с проверочной матрицей

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

в некотором смысле эквивалентный (7,4) — коду Хемминга.

При рассмотрении циклических кодов кроме обычных действий над векторами мы имеем дело фактически ещё с одной операцией, сопоставляющей каждому вектору его циклический сдвиг. Удобным алгебраическим средством для её описания являются многочлены, точнее же, многочлены как элементы кольца вычетов по модулю многочлена $x^n - 1$ (см. главу 10). С каждым вектором $a = (a_0, a_1, \ldots, a_{n-1})$ свяжем многочлен

$$a(x) = a_0 + a_1 x + \ldots + a_{n-1} x^{n-1}, \qquad (15.11)$$

коэффициенты которого совпадают с соответствующими координатами вектора. Именно в такой форме может быть записан всякий элемент кольца вычетов по модулю любого многочлена степени n, в частности, многочлена $x^n - 1$. В дальнейшем будем отождествлять вектор a с соответствующим ему многочленом $a(x)$, свободно переходя от векторной записи к записи в виде многочлена. Циклическому сдвигу

$$a' = (a_{n-1}, a_0, a_1, a_2, \ldots, a_{n-2})$$

вектора a соответствует тогда многочлен

$$a'(x) = a_{n-1} + a_0 x + a_1 x^2 + \ldots + a_{n-2} x^{n-1}. \qquad (15.12)$$

Связь между многочленом (15.11) и его циклическим сдвигом (15.12) может быть выражена теперь в терминах операции умножения в кольце вычетов[1] $\mathbb{F}_n = \mathbb{Z}_2[x]/(x^n - 1)$.

[1] В дальнейшем многочлены по модулю $x^n - 1$ будут обозначаться, как и обычные многочлены. Всюду по контексту ясно, в каком именно смысле понимать тот или иной многочлен.

Эта связь (поскольку в F_n справедливо равенство $x^n = 1$) такова:

$$a'(x) = xa(x).$$

Отсюда один шаг до утверждения, дающего полезное алгебраическое описание циклических кодов:

линейный код V является циклическим тогда и только тогда, когда V является идеалом в кольце F_n.

В самом деле, если V — идеал, то для всякого кодового вектора (многочлена) $a(x) \in V$ имеем $xa(x) \in V$, т. е. циклический сдвиг снова является кодовым вектором.

Обратно, если V — циклический код, то для всякого кодового вектора $a(x)$ его последовательные циклические сдвиги $xa(x)$, $x^2a(x)$, ..., $x^{n-1}a(x)$ также являются кодовыми векторами.

Остается показать, что для всякого многочлена

$$b(x) = b_0 + b_1x + b_2x^2 + \ldots + b_{n-1}x^{n-1}$$

произведение $b(x)\,a(x)$ является кодовым вектором. Мы имеем:

$$b(x)\,a(x) = b_0a(x) + b_1x\,a(x) + b_2x^2\,a(x) + \ldots + b_{n-1}x^{n-1}\,a(x). \tag{15.13}$$

Согласно сказанному выше, каждое слагаемое в (15.13) принадлежит кодовому пространству, но тогда и вся сумма обладает этим свойством. Итак, V — идеал.

Полезность полученного нами описания циклического кода связана с тем, что всякий идеал в кольце вычетов $\mathbb{F}_n$ является главным идеалом. Доказать это можно по известной нам схеме, которой мы придерживались ранее для кольца целых чисел и произвольного евклидова кольца.

Многочлен $g(x)$, порождающий идеал V, называется порождающем многочленом соответствующего циклического кода.

Таким образом, если известен порождающий многочлен кода, то, тем самым, известны и все кодовые многочлены — их список исчерпывается всевозможными произведениями $g(x)s(x)$, где $s(x)$ — произвольный многочлен степени, меньшей n.

Вспомним, что для задания произвольного кода нужно указать полный список кодовых слов, а для задания линейного кода —

только список базисных кодовых векторов. Циклический же код может быть задан ещё более компактно. Действительно, как мы выяснили, для этого достаточно указать всего лишь один кодовый многочлен, а именно порождающий многочлен $g(x)$.

По порождающему многочлену легко найти порождающую матрицу циклического кода. Пусть

$$g(x) = g_0 + g_1 x + \ldots + g_m x^m$$

— многочлен степени m и $k = n - m$. Рассмотрим следующие многочлены

$$g(x),\ xg(x),\ x^2 g(x),\ \ldots,\ x^{k-1} g(x). \tag{15.14}$$

Все они являются кодовыми, степень их, очевидно, не превосходит $n-1$. Нетрудно убедиться, что рассматриваемые как векторы, они образуют линейно независимую систему, и всякий кодовый вектор является их линейной комбинацией. Следовательно, матрица G, составленная из векторов (15.14), является порождающей матрицей циклического кода. Порядок её равен $k \times n$, и она имеет следующий вид:

$$G = \begin{pmatrix} g_0 & g_1 & \ldots & g_m & 0 & \ldots & 0 \\ 0 & g_0 & g_1 & \ldots & g_m & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \ldots & 0 & g_0 & g_1 & \ldots & g_m \end{pmatrix} = \begin{pmatrix} g(x) \\ xg(x) \\ \vdots \\ x^{k-1} g(x) \end{pmatrix}. \tag{15.15}$$

Так, для двоичного циклического кода длины 7 с порождающим многочленом $g(x) = 1 + x^2 + x^3 = (1011000)$ имеем[1]

$$\begin{aligned} xg(x) &= x + x^3 + x^4 = (0101100), \\ x^2 g(x) &= x^2 + x^4 + x^5 = (0010110), \\ x^3 g(x) &= x^3 + x^5 + x^6 = (0001011), \end{aligned}$$

и потому его порождающей будет следующая матрица

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

[1] Многочлен $g(x) = 1 + x^2 + x^3$ является делителем $x^7 - 1$.

Особая структура порождающей матрицы циклического кода позволяет придать правилу кодирования (15.10) иную, гораздо более удобную форму.

Если $a(x)$ — многочлен, соответствующий вектору a, и $A(x) = \alpha_0 + \alpha_1 x + \ldots + \alpha_{k-1} x^{k-1}$ — многочлен, соответствующий сообщению A, то равенство (15.10) перейдёт в равенство многочленов

$$a(x) = A(x)g(x), \tag{15.16}$$

где $g(x)$ — порождающий многочлен циклического кода.

Для того, чтобы убедиться в этом, достаточно подставить в (15.10) матрицу (15.15) и проверить, что координаты получающегося вектора служат коэффициентами многочлена $A(x)g(x)$.

Правило кодирования (15.16) особенно ценно тем, что оно естественно и красиво реализуется в несложных электронных схемах, состоящих из сумматоров и сдвигающих регистрах [2]. Подобные же схемы используются и для целей декодирования, и для исправления ошибок.

Стоит особо поговорить ещё об одном, наиболее интересном способе задания циклического кода. С этой целью рассмотрим поле разложения $\overline{\mathbb{F}}$ порождающего многочлена $g(x)$, а в этом поле — все элементы $\alpha_1, \alpha_2, \ldots, \alpha_m$, являющиеся его корнями: $g(\alpha_i) = 0$, $i = 1, 2, \ldots, m$.

Очевидно, заданием этих корней многочлен $g(x)$, а следовательно, и порождённый им циклический код определены однозначно. Действительно, $g(x) = (x - \alpha_1)(x - \alpha_2) \ldots (x - \alpha_m)$. Заметим также, что элементы α_i являются корнями и любого кодового многочлена $a(x)$ (это следует из равенства (15.16)).

Нетрудно показать и обратное, а именно: всякий многочлен $a(x)$, корнями которого являются элементы α_i ($i = 1, 2, \ldots, m$), будет кодовым многочленом (доказательство предоставляем читателю). Таким образом, если $\alpha_1, \alpha_2, \ldots, \alpha_m$ — все корни порождающего многочлена $g(x)$, то соответствующий циклический код может быть охарактеризован так: он состоит из всех тех многочленов $a(x)$ степени не выше $n - 1$, для которых элементы $\alpha_1, \alpha_2, \ldots, \alpha_m$ также являются корнями. Иными словами, кодовые многочлены $a(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_{n-1} x^{n-1}$, и только

они, удовлетворяют соотношениям

$$\begin{aligned} a(\alpha_1) &= a_0 + a_1\,\alpha_1 + a_2\,\alpha_1^2 + \ldots + a_{n-1}\,\alpha_1^{n-1} = 0, \\ a(\alpha_2) &= a_0 + a_1\,\alpha_2 + a_2\,\alpha_2^2 + \ldots + a_{n-1}\,\alpha_1^{n-1} = 0, \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a(\alpha_m) &= a_0 + a_1\,\alpha_m + a_2\,\alpha_m^2 + \ldots + a_{n-1}\,\alpha_m^{n-1} = 0. \end{aligned} \tag{15.17}$$

Равенства (15.17) играют роль проверочных соотношений циклического кода. Особенность их, правда, состоит в том, что коэффициенты этих соотношений (элементы α_i и их степени) принадлежат не основному полю, а его расширению. Отметим, однако, что представляя элементы расширения $\overline{\mathbb{F}}$ векторами над полем $\mathbb{F}$ (см. дополнение 3 к главе 11), можно, в случае необходимости, прийти к проверочным соотношениям обычного вида.

Равенства (15.17) служат также исходным пунктом построения алгебраических методов исправления ошибок, пригодных для циклических кодов. Общая схема такова.

Пусть принятый вектор u содержит ошибочные символы, тогда его можно представить в виде суммы посланного кодового вектора a (который пока неизвестен) и вектора-ошибки e:

$$u = a + e. \tag{15.18}$$

Ясно, что вектор $e = (e_1, e_2, \ldots, e_n)$ содержит ненулевые символы как раз в тех позициях, в которых символы вектора a искажены. Равенству (15.18), понятно, соответствует аналогичное равенство для многочленов

$$u(x) = a(x) + e(x). \tag{15.19}$$

По принятому вектору мы легко можем, выполняя вычисления в конечном поле $\overline{\mathbb{F}}$, найти величины

$$S_i = u(\alpha_i) = u_0 + u_1\alpha_i + a_2\alpha_i^2 + \ldots + a_{n-1}\alpha_i^{n-1}, \quad i = 1, 2, \ldots, m.$$

С другой стороны, из равенства (15.19) следует, что

$$e(\alpha_i) = u(\alpha_i) = S_i, \tag{15.20}$$

поскольку для всякого кодового многочлена $a(\alpha_i) = 0$. Записывая (15.20) в развёрнутой форме, мы приходим к системе линейных уравнений

$$\begin{aligned} e(\alpha_1) &= e_0 + e_1\alpha_1 + e_2\alpha_1^2 + \ldots + e_{n-1}\alpha_1^{n-1} = S_1, \\ e(\alpha_2) &= e_0 + e_1\alpha_2 + e_2\alpha_2^2 + \ldots + e_{n-1}\alpha_1^{n-1} = S_2, \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ e(\alpha_m) &= e_0 + e_1\alpha_m + e_2\alpha_m^2 + \ldots + e_{n-1}\alpha_m^{n-1} = S_m \end{aligned} \tag{15.21}$$

относительно $e_0, e_1, \ldots, e_{n-1}$ — координат вектора-ошибки.

Конечно, всё не так просто. Дело в том, что система (15.21) имеет не единственное решение (можно даже показать, что число её решений таково же, что и системы (15.17), то есть совпадает с числом кодовых векторов). Всё же, если число ошибочных символов в векторе u и, значит, ненулевых координат в векторе e не слишком велико, система (15.21) позволяет найти их единственным образом. Проиллюстрируем сказанное примером циклического кода, исправляющего одиночные и двойные ошибки.

Пример. Пусть α — примитивный элемент поля $\mathbb{F} = \mathrm{GF}(2^4)$, порядок этого элемента, понятно, равен 15.

Как было показано в главе 10, в качестве $\mathrm{GF}(2^4)$ можно взять поле вычетов $\mathbb{Z}_2[x]/(x^4+x+1)$ по модулю неприводимого над $\mathbb{Z}_2$ многочлена x^4+x+1, а за элемент α взять его корень, содержащийся в этом поле. Элементы 1, α, α^2, α^3 образуют базис поля $\mathrm{GF}(2^4)$ как векторного пространства над $\mathbb{Z}_2$, то есть все степени α^i, $i = 0, 1, 2, \ldots, 14$ записываются единственным образом в виде линейной комбинации

$$a_0 + a_1\alpha + a_2\alpha^2 + a_3\alpha^3, \quad a_i \in \mathbb{Z}_2. \tag{15.22}$$

Например, $\alpha^4 = 1+\alpha$, так как $\alpha^4+\alpha+1 = 0$, $\alpha^5 = \alpha\cdot\alpha^4 = \alpha+\alpha^2$, $\alpha^6 = \alpha\cdot\alpha^5 = \alpha^2+\alpha^3$, $\alpha^7 = \alpha\cdot\alpha^6 = \alpha^3+\alpha^4 = 1+\alpha+\alpha^3$ и т. д.

Удобно для дальнейшего все степени элемента α свести в таблицу 15, отождествляя запись вида (15.22) с вектором (a_0, a_1, a_2, a_3), записываемым в таблице как столбец.

Рассмотрим двоичный циклический код длины 15, состоящий из всех многочленов степени < 15, корнями которых являются

Таблица 15.

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	0	0	0	1	0	0	1	1	0	1	0	1	1	1
0	1	0	0	1	1	0	1	0	1	1	1	1	0	0
0	0	1	0	0	1	1	0	1	0	1	1	1	1	1
0	0	0	1	0	0	1	1	0	1	0	1	1	1	0

элементы α и α^3. Очевидно, что порождающий многочлен $g(x)$ является произведением многочленов $m_1(x)$ и $m_2(x)$, минимальных для α и α^3 соответственно. Как уже было отмечено,

$$m_1(x) = 1 + x + x^4.$$

Далее, поскольку элемент α^3 имеет порядок 5, он является отличным от 1 корнем многочлена

$$x^5 - 1 = (x - 1)(x^4 + x^3 + x^2 + x + 1). \tag{15.23}$$

Второй сомножитель в разложении (15.23) неприводим и потому является минимальным для α^3, так что

$$m_2(x) = 1 + x + x^2 + x^3 + x^4.$$

Следовательно, порождающий многочлен равен

$$g(x) = (1 + x + x^4)(1 + x + x^2 + x^3 + x^4) = 1 + x^4 + x^6 + x^7 + x^8,$$

а порождённый им идеал образует $(15, 7)$-циклический код.

Покажем теперь, что построенный двоичный код позволяет исправить любые одиночные и двойные ошибки. Предположим сначала, что ошибочными являются два символа с номерами i и j, которые и предстоит определить. Тогда $e_i = e_j = 1$, и система (15.21) приобретает вид

$$\begin{aligned} e(\alpha) &= \alpha^i + \alpha^j = S_1 \\ e(\alpha^3) &= \alpha^{3i} + \alpha^{3j} = S_2. \end{aligned} \tag{15.24}$$

Если же ошибка одиночная (и i — номер искажённого символа), то

$$\begin{aligned} e(\alpha) &= \alpha^i = S_1 \\ e(\alpha^3) &= \alpha^{3i} = S_2 = S_1^3. \end{aligned} \tag{15.25}$$

Можно заметить, что отличительным признаком одиночной ошибки является условие $S_2 = S_1^3$. Если оно выполняется, то первое равенство в (15.25) сразу позволяет найти номер ошибочного символа.

В противном случае рассматриваем систему (15.24), поступая в ней следующим образом. Возведём первое равенство в куб

$$(\alpha^i + \alpha^j)^3 = \alpha^{3i} + \alpha^{3j} + \alpha^i\alpha^j(\alpha^i + \alpha^j) = S_1^3.$$

С учётом обоих уравнений системы (15.24) будем иметь

$$S_2 + \alpha^i\alpha^j S_1 = S_1^3,$$

откуда $\alpha^i\alpha^j = (S_1^3 - S_2)/S_1$. Итак, приходим к системе

$$\begin{cases} \alpha^i + \alpha^j = S_1 \\ \alpha^i\alpha^j = S_1^2 - S_2S_1^{-1}, \end{cases} \tag{15.26}$$

из которой, как нетрудно показать, находится не более, чем одна пара элементов α^i, α^j $(i < j)$[1].

Подтвердим сказанное конкретным вычислением. Пусть к адресату поступило следующее слово $u = (100010010111000)$, представимое, как мы знаем, многочленом $u(x) = 1 + x^4 + x^7 + x^9 + + x^{10} + x^{11}$. Тогда (с учётом таблицы 15)

$$\begin{aligned} u(\alpha) &= 1 + \alpha^4 + \alpha^7 + \alpha^9 + \alpha^{10} + \alpha^{11} = \\ &= 1 + (1 + \alpha) + (1 + \alpha + \alpha^3) + (\alpha + \alpha^3) + \\ &\quad + (1 + \alpha + \alpha^2) + (\alpha + \alpha^2 + \alpha^3) = \alpha + \alpha^3 = \alpha^9, \\ u(\alpha^3) &= 1 + \alpha^{12} + \alpha^6 + \alpha^{12} + 1 + \alpha^3 = \alpha^2 + \alpha^3 + \alpha^3 = \alpha^2. \end{aligned}$$

[1] Отметим, что система (15.26) может вовсе не иметь решений. Это служит указанием на то, что ошибочными являются более, чем два символа. Данный код, однако, не позволяет однозначно исправить подобные кратные ошибки.

Следовательно,

$$S_1 = \alpha^9, \quad S_2 = \alpha^2, \quad S_1^2 - S_2 S_1^{-1} = \alpha^3\alpha^8 = \alpha^3 + 1 + \alpha^2 + \alpha^{13}.$$

Так как $S_1^3 \neq S_2$, то ошибка не является одиночной. В системе (15.26) для удобства обозначим $y = \alpha^i$, $z = \alpha^j$ (если $i < j$), и в данном случае она запишется в виде

$$\begin{cases} y + z = \alpha^9, \\ yz = \alpha^{13}. \end{cases}$$

Исключая z из последней системы, приходим к квадратному уравнению

$$y^2 + \alpha^9 y + \alpha^{13} = 0. \tag{15.27}$$

Правда, в случае поля характеристики 2 обычная формула для корней квадратного уравнения непригодна (почему?), но есть другая возможность его решения. Воспользуемся тем, что всякий элемент $y \in \mathrm{GF}(2^4)$ имеет единственную запись вида

$$y = y_0 + y_1\alpha + y_2\alpha^2 + y_3\alpha^3, \quad y_i \in \mathbb{Z}_2,$$

и подставим её в уравнение (15.27)[1]:

$$(y_0 + y_1\alpha^2 + y_2\alpha^4 + y_3\alpha^6) + (y_0 + y_1\alpha + y_2\alpha^2 + y_3\alpha^3)\alpha^9 + \alpha^{13} = 0.$$

После преобразований с использованием таблицы 15 придём к равенству

$$(y_0+y_1+y_2+y_3+1)+(y_0+y_1+y_3)\alpha+(y_2+1)\alpha^2+(y_0+y_2+1)\alpha^3 = 0,$$

из которого следует

$$\begin{cases} y_0 + y_1 + y_2 + y_3 + 1 = 0 \\ y_0 + y_1 + y_3 = 0 \\ y_2 + 1 = 0 \\ y_0 + y_2 + 1 = 0. \end{cases}$$

[1] Мы при этом пользуемся тем, что $(a + b)^2 = a^2 + b^2$ в поле характеристики 2.

Очевидно, имеется два решения этой системы

$$(0,0,1,0), \quad (0,1,1,1).$$

В первом случае $y = \alpha^2$, $z = \alpha^{11}$, во втором, наоборот, $y = \alpha^{11}$, $z = \alpha^2$, $z = \alpha^{11}$. Вспоминая, что $i < j$, окончательно получаем $y = \alpha^2$, но тогда $y = \alpha^{11}$.

Итак, ошибочными являются 3-я и 12-я позиции (напомним, что нумерация y_i начинается с $i = 0$). исправляя эти позиции, получим кодовый вектор $a = (101010010110000)$. В самом деле, соответствующий многочлен $a(x)$ кратен порождающему многочлену

$$a(x) = 1 + x^2 + x^4 + x^7 + x^9 + x^{10} = (1 + x^2)(1 + x^4 + x^6 + x^7 + x^8)$$

и потому является кодовым.

Из сказанного выше можно усмотреть, что все другие кодовые векторы отличаются от принятого вектора u более чем в двух позициях, так что a — «ближайший»[1] к u кодовый вектор. Рассмотренный выше пример циклического кода принадлежит к классу кодов, имеющих особую практическую важность. Это так называемые БЧХ-коды, впервые предложенные американскими математиками Боузом, Чоудхури и Хоквингемом, и получившие такое наименование по начальным буквам их имён.

Кратко опишем общую схему построения циклических кодов БЧХ над произвольным основным полем $\mathbb{F}$, содержащим q элементов. Пусть $\mathrm{GF}(q^m)$ — поле Галуа порядка q^m и α — элемент этого поля (не обязательно примитивный). Зафиксируем натуральные s и l и рассмотрим элементы

$$\alpha^s, \alpha^{s+1}, \alpha^{s+1}, \alpha^{s+2}, \ldots, \alpha^{s+l}. \tag{15.28}$$

Пусть n — наименьшее общее кратное порядков этих элементов. Тогда все они являются корнями многочлена $x^n - 1$ и число n минимально с указанным свойством.

[1] Слову «ближайший» можно придать точный смысл, если ввести соответствующую метрику на множестве двоичных векторов, так называемую метрику Хемминга (см. [2]).

Кодом БЧХ, соответствующим последовательности (15.28) называется код, образованный всеми такими многочленами $f(x) \in \in \mathbb{F}_n$, корнями которых являются все элементы этой последовательности:

$$f(\alpha^{s+i}) = 0, \quad i = 0, 1, \ldots, l.$$

Можно доказать, что такой код (он, конечно, является циклическим) исправляет любые t ошибок, если $2t \leqslant l + 1$.

Код из предыдущего примера состоял из всех двоичных многочленов $f(x) \in \mathbb{Z}_2[x]/(x^{15} - 1)$, корнями которых являются элементы α и α^3, Однако автоматически корнями этих многочленов будут ещё элементы α^2 и α^4, поэтому указанный код является кодом БЧХ, соответствующим последовательности

$$\alpha, \alpha^2, \alpha^3, \alpha^4.$$

Продолжить её далее нельзя, так как α^5 уже не является корнем порождающего многочлена. Здесь $s = 1$, $l = 3$ и условие $2t \leqslant l + 1 = 4$ означает, что код исправляет одиночные и двойные ошибки.

Варьируя в описанной выше схеме m, s, l, мы имеем возможность получать коды БЧХ различных длин, исправляющих то или иное число ошибок. Ясно при этом, что с увеличением l растёт и корректирующая способность кода (число исправляемых ошибок), однако это, естественно, не даётся даром, а лишь, как можно показать, ценой роста числа избыточных символов.

Рассмотрим, например, двоичный БЧХ-код длины 15, состоящий из всех многочленов $f(x) \in \mathbb{Z}_2[x]/(x^{15}-1)$, корнями которых являются элементы

$$\alpha, \alpha^2, \alpha^3, \alpha^4, \alpha^5, \alpha^6,$$

где α (корень многочлена $x^4 + x + 1$) — примитивный элемент $\mathrm{GF}(2^4)$. Как и в ранее рассмотренном примере, многочлены

$$m_1 = 1 + x + x^4, \quad m_2 = 1 + x + x^2 + x^3 + x^4$$

являются минимальными для элементов α и α^3 соответственно. Элемент α^5 имеет порядок 3 и является корнем многочлена

$$x^3 - 1 = (x - 1)(x^2 + x + 1),$$

следовательно, для α^5 минимальным является многочлен

$$m_3(x) = 1 + x + x^2.$$

Но тогда порождающий многочлен рассматриваемого БЧХ-кода равен

$$g(x) = m_1(x)m_2(x)m_3(x) = 1 + x + x^4 + x^5 + x^8 + x^{10}.$$

Мы получили $(15, 5)$-код, исправляющий любые ошибки в трёх или меньшем числе символов, но по сравнению с предыдущим примером размерность кода уменьшилась (и значит наоборот, число избыточных символов увеличилось).

ЗАКЛЮЧЕНИЕ

Вот и завершился наш не слишком длинный, но и не такой уж короткий путь по дорогам алгебры. Вы кое-что узнали о некоторых алгебраических понятиях и идеях, об истории их возникновения, о том, каково их значение для разных областей знания. Естественно, что очень многое осталось за пределами прочитанной Вами книжки. Это немудрено, ведь алгебре и отдельным её разделам посвящены объёмистые учебники и солидные монографии. Один только список изучаемых в ней алгебраических структур содержит не один десяток названий. Не мал также и перечень тех разделов науки, в которых алгебра находит различные применения. Кроме тех, о которых уже говорилось, это квантовая механика и теория элементарных частиц, экономика и методы оптимизации, планирование эксперимента, языки программирования и многое другое. Нужно, правда, оговориться, что чаще всего алгебра применяется в них не сама по себе, а в тесном контакте с другими разделами математики или даже через их посредство.

Мы видели, что ряд приложений алгебры связан с группами тех или иных преобразований. С одной стороны, групповой язык способен выразить разнообразные проявления симметрии, идёт ли речь о симметрии кристаллических решёток или многочленов, а с другой, даёт возможность классифицировать и перечислять различные объекты, от орнаментов до комбинаторных конфигураций.

На тех же общих принципах основывается и ряд других применений алгебры. Так, например, в квантовой механике состояние физической системы изображается точкой (вектором) бесконечномерного векторного пространства. Переход физической си-

стемы из одного состояния в другое на математическом языке означает, что на точку действует некоторое линейное преобразование этого пространства. Группы допустимых преобразований (их строение) могут служить для описания свойств симметрии квантовомеханических систем и во многим определяют физические законы, которым они подчиняются.

Соображения симметрии являются одним из средств «наведения порядка» в знаниях об обширном мире элементарных частиц. Анализ математических свойств возможных групп симметрий в этом мире позволял даже предсказывать существование новых частиц. Так были предсказаны и лишь затем экспериментально найдены частицы, названные пи-мезонами и омега-минус барионами. История этих открытий в микромире заслуживает сравнения с замечательным достижением астрономии XIX века, когда Адамсом и Леверье было теоретически предсказано существование новой планеты, известной сейчас как Нептун.

Группы преобразований и связанные с ними соображения симметрии играют немаловажную роль и в таких физических теориях, как специальная и общая теории относительности. Их математическим обеспечением служит своеобразный сплав алгебры, геометрии и высших разделов анализа. Возникновению этих новых теорий в физике предшествовал коренной пересмотр многих геометрических представлений. Пионерами в этом деле были создатели неевклидовой геометрии — знаменитый русский математик Н. И. Лобачевский и венгр Я. Бойяи. В конце прошлого века крупный немецкий математик Ф. Клейн выдвинул носящую его имя программу, согласно которой понятие группы кладётся в основу классификации тех или иных геометрических теорий. С этой точки зрения традиционная евклидова геометрия есть учение о тех геометрических свойствах, которые сохраняются преобразованиями, принадлежащими группе движений пространства. Другие группы преобразований приводят, например, к аффинной и проективной геометриям. Пройдя через ряд промежуточных ступеней мы можем прийти к такой геометрии, которую интересуют лишь свойства, сохраняющиеся при всевозможных непрерывных и взаимно-однозначных преобразованиях (последние также образуют группу). Этот раздел геометрии называется тополо-

гией. С популярным изложением некоторых фактов топологии можно познакомиться по книге [4]. Из этой же книги можно узнать и о том, что с каждой геометрической фигурой связаны особого типа группы — фундаментальная группа и группы гомологий. Они служат эффективным средством для характеризации многих важных топологических свойств. Развившаяся в связи с проблемами топологии так называемая гомологическая алгебра опирается на уже сложившийся алгебраический арсенал — группы, кольца, линейные пространства и их обобщения.

Значительна роль алгебры в проблемах компьютерной математики и программирования. Уже исходные идеи создателей ЭВМ были связаны с алгебрами Буля, своеобразными алгебраическими системами, рассмотренными английским математиком Дж. Булем почти столетием раньше появления первых ЭВМ. И в настоящее время алгебраические методы плодотворно используются в теории автоматов, в исследованиях по разработке формальных языков и языков программирования.

Развитие вычислительной техники позволило решать сложные практические и научные задачи, казавшиеся ранее недоступными и усилило роль дискретной математики в решении этих задач. Стали бурно развиваться такие разделы дискретного анализа как комбинаторика и теория графов, что отразилось и на алгебре. Объектом изучения в ней стали новые алгебраические структуры (алгебры инцидентности, матроиды, квазигруппы и другие). Они удачно моделируют многие ситуации, типичные для проблем дискретной математики.

Конечно, наивно было бы думать, что всё происходящее в алгебре связанно с решением каких-то практических задач. Это далеко не так. Как и всякая математическая наука, она является сложным организмом со своей логикой развития и со своими внутренними проблемами. К числу этих проблем относятся задачи описания математических структур того или иного типа, поиски новых алгебраических конструкций, выявление связей между различными алгебраическими теориями, углубление и обобщение этих теорий и т. д.

Тем, кто после чтения «Граней алгебры» пожелает углубить свои алгебраические знания, пролить свет на многое нами недо-

казанное и недосказанное, можем посоветовать для дальнейшего изучения книги [9], [17], [18], [19], [29]. Нам остаётся теперь лишь пожелать успеха всякому, кто рискнёт последовать этому совету.

СПИСОК ЛИТЕРАТУРЫ

[1] *Александров П. С.* Введение в теорию групп. — М.: УРСС, 2004.

[2] *Аршинов М. Н., Садовский Л. Е.* Коды и математика. — М.: Наука, 1983.

[3] *Болтянский В. Г., Виленкин Н. Я.* Симметрия в алгебре. — М.: МЦНМО, 2002.

[4] *Болтянский В. Г., Ефремович В. А.* Наглядная топология. — М.: Наука, 1983.

[5] *Ван-дер-Варден Б. Л.* Алгебра. — М.: Наука, 1976.

[6] *Вейль Г.* Симметрия. — М.: Наука, 1968.

[7] *Виленкин Н. Я.* Комбинаторика. — М.: Наука, 1969.

[8] *Виленкин Н. Я.* Популярная комбинаторика. — М.: Наука, 1975.

[9] *Винберг Э. Б.* Курс алгебры. — М.: Факториал, 2001.

[10] *Гиндикин С. Г.* Рассказы о физиках и математиках. — М.: МЦНМО, 2001.

[11] *Дальма А.* Эварист Галуа: революционер и математик. — РХД, 2000.

[12] *Де Брейн Н. Дж.* В кн.: Прикладная комбинаторная математика. — М.: Мир, 1968.

[13] *Инфельд Л.* Эварист Галуа — избранник богов. (Серия ЖЗЛ) — Молодая гвардия, 1965.

[14] *Калужнин Л. А.* Введение в общую алгебру. — М.: Наука, 1973.

[15] *Кантор И. Л., Солодовников А. С.* Гиперкомплексные числа. — М.: Наука, 1973.

[16] *Клейн Ф.* Лекции о развитии математики в XIX столетии. — М.: Наука, 1989.

[17] *Кострикин А. И.* Введение в алгебру. — М.: Наука, 1973.

[18] *Кострикин А. И.* Введение в алгебру. Т. 1. Основы алгебры. Т. 2. Линейная алгебра. Т. 3. Основные структуры алгебры. — М.: Физматлит, 2000.

[19] *Ленг С.* Алгебра. — М.: Мир, 1968.

[20] *Пойя Д.* Математика и правдоподобные рассуждения. — М.: Наука, 1975.

[21] *Пойя Д.* Математическое открытие. — М.: Наука, 1978.

[22] *Пойя Д.* (*Polya D.*) Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen. Acta Math., v.68, 1937, 145–254.

[23] *Постников М. М.* Введение в теорию алгебраических чисел. — М.: Наука, 1982.

[24] *Постников М. М.* Теория Галуа. — М.: Факториал, 2003.

[25] *Соловьёв Ю. П.* Гипотеза Таниямы и последняя теорема Ферма. — СОЖ, 1998, № 2, 135–138.

[26] *Харрари Ф., Палмер Э.* Перечисление графов. — М.: Мир, 1977.

[27] *Холл М.* Комбинаторика. — М.: Наука, 1970.

[28] Христоматия по истории математики, под редакцией Юшкевича А. П. — М.: Просвещение, 1976.

[29] *Шафаревич И. Р.* Основные понятия алгебры. — Ижевск: РХД, 1999.

[30] *Яглом И. М.* Геометрические преобразования (в двух частях). — М.: Гостехиздат, 1955.

«Грани алгебры» — книга довольно необычная. Это не учебник, это скорее сборник этюдов на темы из алгебры. Стиль изложения доверительный, слышится живая интонация лектора. Всё время подчёркивается историческая перспектива, сообщаются сведения о знаменитых математиках. Однако книгу вряд ли можно назвать научно-популярной. В ней развивается достаточно серьёзный математический аппарат. Полные доказательства приводятся не всегда, но существо дела разъясняется подробно.

9 785886 880915
ISBN 978-5-88688-091-5
ФАКТОРИАЛ ПРЕСС