MODERN APPLIED ALGEBRA

GARRETT BIRKHOFF
Department of Mathematics
Harvard University
and
THOMAS C. BARTEE
Division of Engineering
and Applied Physics
Harvard University

McGRAW-HILL BOOK COMPANY
NEW YORK, St. LOUIS, SAN FRANCISCO, DÜSSELDORF, LONDON,
MEXICO, PANAMA, SYDNEY, TORONTO

Г. БИРКГОФ
Т. БАРТИ

СОВРЕМЕННАЯ ПРИКЛАДНАЯ АЛГЕБРА

Перевод с английского
Ю. И. МАНИНА

ИЗДАТЕЛЬСТВО «МИР»
· МОСКВА ·
1976

УДК 519.7

Книга написана двумя американскими учеными, один из которых — Гаррет Биркгоф — известен широтой своих научных интересов, простирающихся от абстрактной алгебры до гидродинамики, а другой — Томас Бартн — является директором вычнслительной лаборатории Гарвардского университета. На русском языке выходила «Теория структур» Биркгофа, два издания его знаменитой «Гидродинамнки», а также совместная с Э. Сарантонелло монография по теории струй.

Книга восполняет существенный пробел в нашей учебной литературе. В отлнчие от другнх математических дисциплин, по которым имеются превосходные руководства, специально ориентированные на приложения, по алгебре таких книг до сих пор не было. Это связано с тем, что алгебра (за исключением теории уравнений) приобрела черты прикладной науки лишь за последние десятилетия.

В книге нзлагаются те идеи и методы современной алгебры, которые нашли широкие применения в таких областях, как теория автоматов и вычислительных машин, передача сообщений и кодирование, языки программирования н математическая лингвистика. Она будет полезна тем, кто работает в этнх областях, а также математикам всех специальностей, занимающимся прикладными вопросамн.

Особый интерес она представляет для студентов университетов и высших технических учебных заведений, связанных с прнкладной математикой.

Редакция литературы по математическим наукам

Б $\frac{20203\text{-}004}{041\ (01)\text{-}76}$ 04-76

ПРЕДИСЛОВИЕ

Термин «современная алгебра» принято относить к теории таких алгебраических систем (групп, колец, булевых алгебр и т. д.), элементы которых, вообще говоря, *не являются числами*. Напротив, «классическая» алгебра занимается в основном алгебраическими уравнениями или системами уравнений, в которых символы обозначают вещественные или комплексные числа. За последние сорок лет «современная» алгебра постепенно вытеснила «классическую» в программах американских университетов.

Прошедшее двадцатилетие было отмечено бурным развитием новых отраслей техники. К ним относятся электронные вычислительные машины, средства передачи, хранения и переработки информации, системы обнаружения типа радара и сонара. Работа в каждой из этих отраслей требует довольно глубокого знания современной алгебры. Изучение последней, таким образом, стало необходимым для специалистов по прикладной математике, инженеров и всех исследователей, которые занимаются перечисленными вопросами и пользуются ЭВМ.

В этой книге мы попытались изложить те идеи и методы современной алгебры, которые оказались наиболее полезными в этих областях. Хотя в тексте неизбежно некоторое разделение между математическими принципами и их приложениями, мы старались добиться единства в изложении фундаментальных идей и основанных на них алгоритмов.

Книга начинается с обсуждения и иллюстрации примерами понятий множества, функции, математической индукции, бинарного отношения и графа. Рассматриваются также булевы алгебры, моноиды, морфизмы и другие основные алгебраические понятия. Все это занимает первые две главы.

В следующей, третьей главе вводятся понятия автомата и машины Тьюринга. Описаны также способы представления этих объектов и методика уменьшения числа состояний при синтезе конечных автоматов. Глава 4 вводит читателя в синтаксис и семантику языка программирования АЛГОЛ.

В пятой главе принят аксиоматический подход, столь характерный для современной алгебры. Булевы алгебры определяются формально системой аксиом; из этих аксиом выводятся их свойства. Объясняются связи булевых алгебр с логикой, теорией вентильных схем и программированием на АЛГОЛе, после чего выводится каноническая форма для булевых многочленов. Основная тема шестой главы — *оптимизация*. Здесь сначала описываются способы отыскания минимальных путей в графах. После этого вводятся методы описания вентильных схем, важных для проектирования ЭВМ, и способы их упрощения с помощью булевых многочленов. Наконец, объясняется, как синтезировать любой автомат из вентилей и триггеров (элементов памяти).

Седьмая глава содержит аксиоматическое изучение моноидов и групп. О моноидах в ней сказано гораздо больше, а о группах — несколько меньше, чем в обычных курсах современной алгебры. В гл. 8 некоторые из этих идей прилагаются к системам связи, которые передают информацию при наличии шума, искажающего сообщение. Для вычисления вероятности ошибки используется стандартная «двоичная симметричная модель» теории передачи информации. Далее, исходя из этого описывается методика порождения, кодирования и декодирования групповых кодов, позволяющая добиться эффективного обнаружения и исправления ошибок. Затем следует гл. 9 о структурах. В ней показывается, что изучение отношений частичного порядка ведет к далеко идущим обобщениям булевых алгебр.

Десятая глава посвящена кольцам и полям. В ней описываются различные типы колец, возникающие в приложениях, объясняется связь между морфизмами и идеалами, а также обсуждаются понятия однозначного разложения и метод исключения по Гауссу. В гл. 11 изучены кольца многочленов. Результаты применяются затем к построению и анализу кодов с исправлением и обнаружением ошибок.

Конечные поля (называемые также полями Галуа) изучены в гл. 12. Они используются затем для описания специального класса кодов — кодов Боуза — Чоудхури — Хоккенгема. В гл. 13 введены разностные уравнения и еще один основанный на них класс кодов. Последний используется совместно с автокорреляционными функциями в системах обнаружения и передачи информации типа радара.

В заключительной, четырнадцатой, главе впервые в этой книге вводятся бесконечные системы. Показано, что вещественные числа образуют несчетное множество. Определяется понятие *вычислимости*, объясняется его связь с машинами Тьюринга. Наконец, понятие машинной вычислимости связывается с некоторыми идеями математической лингвистики и проблемой классификации языков программирования.

Изложенный в этой книге материал может служить основой нескольких разных лекционных курсов. Параграфы, отмеченные звездочками, при недостатке времени можно опустить. В Гарварде этот материал излагается в двух полугодовых курсах: «Прикладная математика 106» (специальный курс для начинающих) и «Прикладная математика 206» (для более подготовленных). Для ПМ 106 не требуется предварительной подготовки. Поскольку эти курсы содержат много фундаментального материала, на них основан и ряд других дисциплин. Курс ПМ 206 включает также краткое введение в теорию вещественных и комплексных матриц; особое внимание в нем уделяется свойствам, важным для приложений. Этим вопросам посвящается примерно треть семестра.

Авторы многим обязаны Джону Липсону, который тщательно прочитал два предварительных варианта рукописи, помог написать раздел об АЛГОЛе и составил содержащиеся в нем программы. Мы признательны также Аспи Вадья, который также прочитал рукопись и тщательно прорешал ряд задач. Наконец, нам очень помогли советами и критикой многие друзья и коллеги, особенно Дональд Андерсон, Маршалл Холл, Сесумо Куно, Дональд Макларен, Альберт Мейер, Вернер Рейнболдт, Хартли Роджерс, Давид Шнейдер и Хао Ван.

Кембридж, Массачусетс *Гарретт Биркгоф*
Томас К. Барти

1

МНОЖЕСТВА И ФУНКЦИИ

1.1. МНОЖЕСТВА И ПОДМНОЖЕСТВА

Понятие алгебраической системы A является *основным* в современной алгебре. Такая система состоит из *множества элементов* $a_1, a_2, a_3, \ldots$, над которыми можно производить *операции* типа сложения или умножения.

Понятие множества так часто употребляется в повседневной речи, что для него имеется много общеизвестных синонимов (класс, совокупность, толпа, стая, ...), некоторые из которых обладают специальными оттенками. В математике это понятие принадлежит к числу самых фундаментальных. Вот некоторые важные математические множества: все целые числа (положительные, отрицательные и нуль) $\mathbf{Z}$, все рациональные числа $\mathbf{Q}$, все вещественные числа $\mathbf{R}$, все комплексные числа $\mathbf{C}$[1]). Каждое из этих множеств бесконечно, т. е. содержит бесконечно много элементов. Кроме того, каждое образует важную алгебраическую систему относительно операций сложения и умножения.

«Нематематические» множества обычно конечны. Любое конечное множество можно задать с помощью списка всех его элементов. Такой список часто заключается в фигурные скобки $\{\quad\}$. Так, множество всех простых чисел между 20 и 40 задается списком $\{23, 29, 31, 37\}$, а множество всех целых положительных делителей 8 — списком $\{1, 2, 4, 8\}$. Перестановка элементов списка не меняет множества его членов: так, множества $\{1, 2, 4, 8\}$ и $\{8, 4, 2, 1\}$ совпадают.

Некоторые бесконечные множества (математический термин «счетные») можно описывать таким же способом, указывая, как перенумеровать все их элементы и представлять их в виде бесконечной последовательности. Например, множество $\mathbf{P}$ всех целых положительных чисел удобно записывать в виде

$$\mathbf{P} = \{1, 2, 3, \ldots\}.$$

---

[1]) Эти обозначения сохраняются на протяжении всей книги.

Аналогично, множество **N** всех неотрицательных целых чисел и множество **Z** всех целых чисел можно записать в виде

$$\mathbf{N}=\{0,\ 1,\ 2,\ 3,\ \ldots\},$$
$$\mathbf{Z}=\{0,\ \pm 1,\ \pm 2,\ \pm 3,\ \ldots\}.$$

Общее правило состоит в выписывании настолько большого количества членов последовательности, чтобы принцип ее бесконечного продолжения по «рекурсии» стал очевиден (см. дальнейшие подробности в гл. 14). Однако элементы из **R** или **C** нельзя представить в виде такой бесконечной последовательности.

Чтобы определить некоторое множество S, мы должны объяснить, как отвечать на следующий вопрос: принадлежит данный объект a множеству S или нет? Вместо «a принадлежит S» мы можем говорить «a является элементом S» и записывать этот факт символически: $a \in S$. Если a не принадлежит S, мы пишем $a \notin S$.

Подмножеством S множества T называется любое множество, все элементы которого принадлежат T. Иными словами, из $a \in S$ следует, что $a \in T$. Это отношение между S и T символически записывается так: $S \subset T$ (или $T \supset S$). Мы говорим также «S *содержится* (включается) в T». Определенное таким образом отношение включения $\subset$ обладает рядом очевидных свойств:

$$S \subset S \text{ для любого множества } S, \tag{1}$$
$$\text{из } S \subset T \text{ и } T \subset U \text{ следует, что } S \subset U. \tag{2}$$

Свойство (1) называется *рефлексивным свойством* включения, а (2) — *транзитивным свойством*.

Всякое множество однозначно определяется своими элементами. Иными словами, два множества S и T совпадают, когда они обладают в точности одними и теми же элементами. Символически:

$S=T$ означает, что $x \in S$ тогда и только тогда, когда $x \in T$.

Так как включение $S \subset T$ имеет место в том и только том случае, когда из $x \in S$ следует, что $x \in T$, мы имеем

$$S=T \text{ тогда и только тогда, когда } S \subset T \text{ и } T \subset S. \tag{3}$$

Множества и подмножества можно указывать явно разными способами. Например, подмножество S множества U часто определяется как множество всех тех элементов $x \in U$, которые обладают некоторым определенным свойством. Если это свойство (т. е. некоторое утверждение относительно x) обозначить через $P(x)$, то определение S можно записать символически

$$S=\{x \in U \mid P(x)\} \quad \text{или (просто)} \quad S=\{x \mid P(x)\};$$

читаются эти формулы так: «S есть множество всех элементов x множества U, для которых справедливо утверждение $P(x)$».

Например, формулы

$$E=\{x \mid x\in \mathbf{Z} \text{ и } x=2y \text{ для некоторого } y\in \mathbf{Z}\},$$
$$\mathbf{P}=\{x \mid x\in \mathbf{Z} \text{ и } x>0\}$$

описывают множества всех четных и всех положительных целых чисел соответственно. Разумеется, вместо первой формулы мы можем написать также

$$E=\{0,\ \pm 2,\ \pm 4,\ \ldots\}.$$

Аксиома, постулирующая, что каждое разумное свойство $P(x)$ определяет некоторое множество S_P элементов, обладающих этим свойством, часто называется *аксиомой выделения* и принадлежит к числу фундаментальных аксиом логики[1]). Обратно, любому множеству S отвечает свойство $x\in S$ «быть элементом S», или «принадлежать S».

Множество всех подмножеств (частей) данного множества U обозначается через $\mathcal{P}(U)$. Оно содержит в качестве элементов само множество U, пустое множество $\varnothing$, а также (если U состоит из $n\geqslant 1$ элементов) 2^n-2 «собственных» подмножеств S, удовлетворяющих условиям $\varnothing\subset S\subset U$, $S\neq\varnothing$ и $S\neq U$.

Например, множество частей $\mathcal{P}(\{a, b, c\})$ множества $U=\{a, b, c\}$ содержит шесть собственных подмножеств: $\{a\}$, $\{b\}$, $\{c\}$, $\{a, b\}$, $\{a, c\}$ и $\{b, c\}$. Вообще если U состоит из n различных элементов $a_1, \ldots, a_n$ (n — положительное целое число), то $\mathcal{P}(U)$ состоит из 2^n различных подмножеств S, включая, кроме собственных подмножеств, пустое множество $\varnothing$, не имеющее элементов, и все множество U. Этот подсчет основан на соображении, что для каждого элемента $p\in U$ имеет место ровно одна из двух возможностей: $p\in S$ или $p\notin S$.

1.2. БУЛЕВЫ АЛГЕБРЫ

В математике термин *бинарная операция на множестве* S означает правило, которое каждой *упорядоченной паре* (a, b) элементов S ставит в соответствие третий элемент S — значение этой операции на паре (a, b). Сложение $+$ в арифметике доставляет пример такой операции на множестве $\mathbf{Z}$ всех целых чисел; значение суммы a и b обычно записывается в виде $a+b$; знак операции $+$ ставится между двумя операндами (или аргументами). Другие известные бинарные операции в $\mathbf{Z}$ — это вычита-

---

[1]) Разумеется, в более формальных изложениях понятие «разумного» свойства уточняется.— *Прим. перев.*

ние и умножение. Действительно, по любым двум целым числам m, $n \in \mathbf{Z}$ однозначно определяются целые числа $m-n$ и $m \times n$.

Бинарная операция — , знак которой фигурирует в выражениях 4 — 2 или 3 — 1, строго говоря, отличается от *унарной* операции — , знак которой фигурирует в выражениях —1 и —2, ибо первая операция совершается над двумя операндами, а вторая — лишь над одним. Вообще унарной операцией на множестве S называется всякое правило f, которое любому элементу $a \in S$ ставит в соответствие однозначно определенный элемент $f(a) \in S$, значение операции f на a. (Таким образом, унарная операция на S есть просто функция $f\colon S \to S$ в обычном смысле слова: см. § 1.3.)

На множестве подмножеств любого данного множества U определены две фундаментальные бинарные операции и одна унарная. Это — операции теоретико-множественного пересечения, объединения и дополнения; они называются также булевыми. Мы покажем, что бинарные операции пересечения и объединения в $\mathcal{P}(U)$, множестве всех частей множества U, аналогичны операциям умножения и сложения в $\mathbf{Z}$.

Пусть R и S — подмножества множества U. Их *пересечением* $R \cap S$ называется множество всех элементов, принадлежащих как R, так и S:

$$R \cap S = \{x \mid x \in R \text{ и } x \in S\}.$$

Объединением R и S называется множество всех элементов, принадлежащих либо R, либо S (либо и R, и S):

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

Пусть, например, $E = \{0, \pm 2, \pm 4, \ldots\}$ — множество всех четных целых чисел, а $\mathbf{P}$ — множество всех положительных целых чисел 1, 2, 3 ... . Тогда $E \cap \mathbf{P}$ состоит из всех положительных четных целых чисел ($E \cap \mathbf{P} = \{2, 4, 6, \ldots\}$), а $E \cup \mathbf{P}$ — из всех целых чисел, не являющихся одновременно отрицательными и нечетными (т. е. не принадлежащих множеству $\{-1, -3, -5, \ldots\}$).

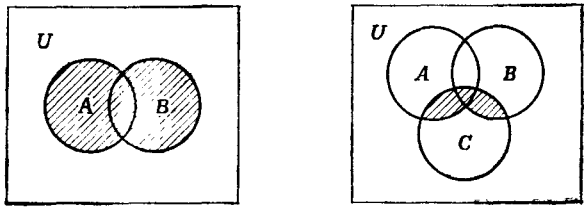
На диаграммах Венна на рис. 1.1 наглядно представлены операции $\cap$ и $\cup$. На них U изображается площадью четырехугольника, а подмножества A и B — кругами. На рис. 1.1,*б* третий круг C также изображает подмножество U, а заштрихованная площадь — множество $(A \cap C) \cup (B \cap C) = (A \cup B) \cap C$.

Обозначим через $n(S)$ число элементов множества S. Если R и S — конечные множества, не имеющие общих элементов (т. е. $R \cap S = \varnothing$), то число элементов объединения $R \cup S$ равно арифметической сумме чисел элементов R и S соответственно: $n(R \cup S) = n(R) + n(S)$. Для любых конечных множеств R и S справедливо более общее равенство:

$$n(R \cup S) = n(R) + n(S) - n(R \cap S). \tag{4}$$

Действительно, $n(R \cap S)$ — это число тех элементов (объемлющего множества U), которые «считаются дважды» (значит, на один раз больше, чем нужно) при последовательном пересчете сначала всех элементов R, а затем всех элементов S.

Рассмотрим в качестве примера подмножества $M = \{mk\}$ и $N = \{nk\}$ всех целых кратных фиксированных целых чисел m, $n \in \mathbf{Z}$; k пробегает все целые числа. Тогда $M \cap N$ совпадает



(а) (б)

Рис. 1.1. Диаграмма Венна.

с множеством всех *общих кратных* чисел m и n. (Евклид доказал важную теорему о том, что это пересечение $M \cap N = \{[m, n]k\}$ состоит из всех целых кратных $[m, n]k$ наименьшего общего кратного $[m, n]$ чисел m и n. См. по этому поводу гл. 10.)

Наконец, рассмотрим множество $\mathbf{Q}^+$ всех положительных рациональных чисел и множество $\mathbf{Q}^-$ всех отрицательных рациональных чисел. Тогда объединение $\mathbf{Q}^+ \cup \mathbf{Q}^-$ является множеством всех ненулевых рациональных чисел.

Когда $R \cap S = \varnothing$, мы говорим, что множества R и S *не пересекаются*. Представление множества U в виде объединения попарно не пересекающихся подмножеств называется *разбиением* U.

Пусть S — любое подмножество U. Множество тех элементов $x \in U$, которые не принадлежат S, называется *дополнением* S (в U) и обычно обозначается через S'. Символически: $S' = \{x \in U \mid x \notin S\}$. Ясно, что если S' — дополнение S, то S — дополнение S'. Подмножества S и T называются *дополнительными*, если $T = S'$ или, что то же самое, $S = T'$. Дополнение S' множества S в U можно также охарактеризовать булевыми равенствами

$$S \cap S' = \varnothing \text{ и } S \cup S' = U. \tag{5}$$

Таким образом, два множества $\{S, S'\}$ образуют разбиение U на два непересекающихся подмножества, объединение которых исчерпывает все U. Так, на 1.1,*а* незаштрихованная площадь изображает $(A \cup B)' = A' \cap B'$; на рис. 1.1,*б* она изображает

$$[(A \cap C) \cup (B \cap C)]' = (A' \cap B' \cap C) \cup C'.$$

Три булевых операции $\cap$, $\cup$ и $'$ на множестве всех частей фиксированного множества U удовлетворяют ряду основных алгебраических законов. Вот некоторые из них:

L1. $S \cap S = S$, $S \cup S = S$ (законы идемпотентности).

L2. $S \cap T = T \cap S$, $S \cup T = T \cup S$ (законы коммутативности).

L3. $R \cap (S \cap T) = (R \cap S) \cap T$, $R \cup (S \cup T) = (R \cup S) \cup T$ (законы ассоциативности).

L4. $S \cap (S \cup T) = S \cup (S \cap T) = S$ (закон поглощения).

L5. Если $R \subset T$, то $R \cup (S \cap T) = (R \cup S) \cap T$ (модулярный закон).

L6. $R \cap (S \cup T) = (R \cap S) \cup (R \cap T)$,
$R \cup (S \cap T) = (R \cup S) \cap (R \cup T)$ (законы дистрибутивности).

L7. $R \cap \varnothing = \varnothing$, $R \cup \varnothing = R$,
$R \cap U = R$, $R \cup U = U$ (универсальные границы, нижняя и верхняя).

L8. $R \cap R' = \varnothing$, $R \cup R' = U$ (дополняемость).

L9. $(S')' = S$ (инволютивный закон).

L10. $(S \cap T)' = S' \cup T'$ и $(S \cup T)' = S' \cap T'$ (законы де Моргана).

Докажем, например, первый закон дистрибутивности. Если $x \in R \cap (S \cup T)$, то $x \in R$ и либо $x \in S$, либо $x \in T$. В первом случае $x \in R$ и $x \in S$, поэтому $x \in R \cap S$. Аналогично, во втором случае $x \in R \cap T$. Следовательно, в любом случае либо $x \in R \cap S$, либо $x \in R \cap T$, так что $x \in (R \cap S) \cup (R \cap T)$. Это доказывает, что

$$R \cap (S \cup T) \subset (R \cap S) \cup (R \cap T).$$

Обратно, пусть $x \in (R \cap S) \cup (R \cap T)$, т. е. либо $x \in R \cap S$, либо $x \in R \cap T$. Значит, в любом случае $x \in R$, а также в любом случае либо $x \in S$, либо $x \in T$. Таким образом, $x \in R$ и $x \in S \cup T$, а потому $x \in R \cap (S \cup T)$, т. е.

$$(R \cap S) \cup (R \cap T) \subset R \cap (S \cup T).$$

Объединяя этот результат с последней формулой предыдущего абзаца и используя (3), получаем первый из двух законов дистрибутивности L6.

Подобно тому как целые числа образуют *алгебраическую систему* $[\mathbf{Z}, +, \times]$ с двумя бинарными арифметическими операциями сложения и умножения, множество всех частей $\mathcal{P}(U)$ любого множества U образует алгебраическую систему $[\mathcal{P}(U), \cap, \cup, ']$ с тремя теоретико-множественными операциями, которые были описаны выше. Свойства сложения и умножения в $\mathbf{Z}$ позволяют отнести систему $[\mathbf{Z}, +, \times]$ к классу *коммутативных колец* (см. гл. 10). Аналогично, свойства L1—L10 алгебраической системы

$[\mathcal{P}(U), \cap, \cup, ']$ позволяют отнести ее к классу *булевых алгебр*. Это — булева алгебра всех подмножеств множества U. Таким образом, считая истинными свойства L1—L10, мы получаем по определению следующий результат.

Теорема 1. *Множество всех частей $\mathcal{P}(U)$ любого множества U является булевой алгеброй.*

Мы вернемся к доказательствам свойств L1—L10 и к объяснению их значения в гл. 5 (и еще раз в гл. 9). А пока формулировку теоремы 1 можно просто воспринимать как удобную сводку списка L1—L10 некоторых основных свойств алгебры множеств. Проиллюстрируем их следующими простыми примерами.

Пример 1. Пусть $U=\{a\}$ — одноэлементное множество. Переобозначив U через 1, а $\varnothing$ через 0, мы получим простейшую нетривиальную булеву алгебру $[\mathcal{P}(\{a\}), \cap, \cup, ']$, операции в которой задаются следующими таблицами:

$\cap$	0	1
0	0	0
1	0	1

$\cup$	0	1
0	0	1
1	1	1

$'$	
0	1
1	0

Читать такую таблицу для бинарной операции нужно так: на пересечении строки a и столбца b стоит результат указанной операции, произведенной над упорядоченной парой (a, b). Так, из наших таблиц видно, что $0 \cup 0=0$, $0 \cup 1=1$, $1 \cap 0=0$ и т. д.

Следующий по сложности пример булевой алгебры есть описание в других обозначениях алгебры $[\mathcal{P}(\{a, b\}), \cap, \cup, ']$. Здесь вместо $\varnothing$, $\{a\}$, $\{b\}$, $\{a, b\}$ мы будем писать символы O, S, S' и I соответственно (S' есть дополнение S, т. е. $\{a\}$, в $\{a, b\}$). В этих обозначениях мы получаем

Пример 2. Четырехэлементная булева алгебра $\{O, S, S', I\}$ определена следующими таблицами:

$\cap$	O	S	S'	I
O	O	O	O	O
S	O	S	O	S
S'	O	O	S'	S'
I	O	S	S'	I

$\cup$	O	S	S'	I
O	O	S	S'	I
S	S	S	I	I
S'	S'	I	S'	I
I	I	I	I	I

x	x'
O	I
S	S'
S'	S
I	O

УПРАЖНЕНИЯ А

1. Положим $S+T=(S \cap T') \cup (S' \cap T)$.
Найти необходимые и достаточные условия для того, чтобы $S+T=S \cup T$.

2. а) Доказать, что $S \subset S \cap (S \cup T)$ и $S \supset S \cap (S \cup T)$.
б) Доказать, что $S=S \cup (S \cap T)$.

3. Доказать, что $S \subset T$ в том и только том случае, если $S' \cup T=U$.

4. а) Доказать, что любое из трех соотношений $S \subset T$, $S \cap T=S$ и $S \cup T=T$ между подмножествами данного миожества U влечет два других (закон согласованиости).
б) Доказать, что для любых элементов S, T любой булевой алгебры свойства $S \cap T=S$ и $S \cup T=T$ равносильны.

5. Используя непосредственно существование универсальиых границ и закон согласованности, доказать, что $S \cup U=U$, $S \cap \varnothing=\varnothing$, $S \cap U=S$, $S \cup \varnothing=S$.

6. Используя подходящие диаграммы Венна, проверить следующие факты:
а) законы дистрибутивности L 6;
б) $(R \cap S) \cup (S \cap T) \cup (T \cap R)=(R \cup S) \cap (S \cup T) \cap (T \cup R)$;
в) показать, что в предыдущем равенстве обе его части состоят в точности из тех точек, которые содержатся в большинстве (по крайней мере в двух) множеств R, S, T.

7. Показать, что из трех кругов A, B, C на диаграмме Венна рис. 1.1,*б* можно образовать в точности $256=2^8=2^{2^3}$ различных булевых комбинаций.

В упр. 8 *и* 9 *следует пользоваться только тождествами* L1—L10.

8. а) Положив $R+S=(R \cap S') \cup (R' \cap S)$, доказать тождества $R+S=S+R$, $R+(S+T)=(R+S)+T$ и $R+R=\varnothing$.
б) Положив $S-T=S \cap T'$, доказать, что $S+T=(S \cup T)-(S \cap T)$.

9. Исходя из определения булевой алгебры, доказать тождества:
а) $S+S=\varnothing$, $S+I=S'$, $S+\varnothing=S$, $S+S'=I$,
б) $S+T=S'+T'$, $(R+S) \cap T=(R \cap T)+(S \cap T)$,
в) $S-(S \cap T)=S-T=T'-S'$,
г) $R \cap (S-T)=(R \cap S)-(R \cap T)$.

*10. Вывести тождество из упр. 6б, несколько раз применяя законы L1—L4 и L6.

* Упражнения, отмеченные звездочкой, можно опускать, не теряя нити дальиейшего изложения.

1.3. ФУНКЦИИ

Пусть S и T — множества. *Функцией* f из *области* S в *кообласть*[1]) T называется правило, которое сопоставляет каждому элементу $s \in S$ единственный элемент из T, называемый *значением* f в s и обозначаемый через $f(s)$. Мы будем также говорить, что f является функцией (или отображением, преобразованием) из S в T, и писать $f\colon S \to T$. *Образом* $\operatorname{Im} f$ отображения $f\colon S \to T$ называется множество $f(S)$ всех значений $f(s)$, которые оно

[1]) В русской литературе T часто называют множеством значений f, но это лишает терминологию симметрии и порождает опасность смешения множества T с образом f. —*Прим. перев.*

принимает при всевозможных $s \in S$. Образ f является подмножеством кообласти T.

Самый прямой способ задать функцию — это задать с помощью списка значения, которые она принимает на элементах области. Например, одна из функций с областью $S = \{a, b, c\}$, кообластью $T = \{a, b, c, d\}$ и образом $f(S) = \{a, b\}$ задается *предписанием*

$$f\colon a \mapsto a, \quad b \mapsto a, \quad c \mapsto b.$$

Другая функция $g\colon S \longrightarrow T$ с образом $g(S) = \{b, c, d\}$ задается предписанием

$$g\colon a \mapsto d, \quad b \mapsto c, \quad c \mapsto b.$$

Заметим, что к образам этих функций можно применять булевы операции: в наших примерах $f(S) \cup g(S) = \{a, b, c, d\}$ и $f(S) \cap g(S) = \{b\}$.

В качестве другого примера определим *функцию следования Пеано* σ, для которой множество $\mathbf{P}$ всех положительных целых чисел является и областью, и кообластью. Каждому целому положительному числу n она ставит в соответствие число $n+1$: $\sigma(n) = n+1$, $\sigma\colon \mathbf{P} \longrightarrow \mathbf{P}$ — функция из $\mathbf{P}$ в $\mathbf{P}$. Функцию σ также можно задать (бесконечным) списком предписаний:

$$\sigma\colon 1 \mapsto 2, \quad 2 \mapsto 3, \quad 3 \mapsto 4, \ldots, \quad n \mapsto n+1, \ldots.$$

Очевидно, образом $\operatorname{Im}\sigma$ функции $\sigma\colon \mathbf{P} \longrightarrow \mathbf{P}$ является множество

$$\sigma(\mathbf{P}) = \{2, 3, 4, \ldots\},$$

которое мы обозначим символом $\mathbf{P}_2$.

Чтобы задать функцию, мы должны определить ее область, кообласть и значения, которые она сопоставляет каждому элементу области. Символически: если $f\colon S \longrightarrow T$ и $g\colon S_1 \longrightarrow T_1$, то $f = g$ в том и только том случае, когда

$$f(x) = g(x) \text{ для всех } x \in S, \tag{6а}$$

$$S = S_1 \text{ и } T = T_1. \tag{6б}$$

Согласно этому определению, функции могут быть строго одинаковыми, только если их области и кообласти совпадают. Например, предписание $n \mapsto n+1$ определяет также функцию $\sigma'\colon \mathbf{P} \longrightarrow \mathbf{P}_2$. Мы не будем считать, что она совпадает с введенной выше функцией следования $\sigma\colon \mathbf{P} \longrightarrow \mathbf{P}$, поскольку кообласть у σ' другая: она не содержит 1. (Заметим, однако, что образы σ и σ' совпадают: $\sigma(\mathbf{P}) = \sigma'(\mathbf{P})$.)

Для любого множества S *тождественная* функция $1_S\colon S \longrightarrow S$ отображает любой элемент S в себя: $1_S(s) = s$ для всех $s \in S$. В силу сказанного выше тождественные функции разных множеств различны.

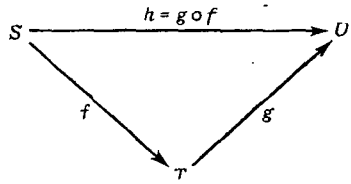
Композиция. *Левой композицией* $g \circ f$ любых двух функций называется функция, полученная в результате их применения в порядке, обратном написанному. Сначала применяется f, а затем g при условии, что область g совпадает с кообластью f. Формально: пусть

$$f\colon S \longrightarrow T \text{ и } g\colon T \longrightarrow U.$$

Тогда левая композиция $g \circ f$ есть функция $g \circ f\colon S \longrightarrow U$, определенная правилом

$$(g \circ f)(s) = g(f(s)) \quad \text{для всех} \quad s \in S. \tag{7}$$

Это соотношение между тремя функциями f, g и $h = g \circ f$ наглядно изображается следующей *диаграммой отображений*:



Она иллюстрирует то обстоятельство, что мы можем перейти из S в U либо непосредственно применяя h, либо в два шага, применяя сначала f, а затем g.

Правая композиция. Операция *правой* композиции $f \diamond g$ получается из описанной выше операции левой композиции перестановкой символов: $f \diamond g = g \circ f$ [1]). Пусть, например, $\varphi_m\colon \mathbf{R} \longrightarrow \mathbf{R}$ — операция возведения в степень m, $\varphi_m(x) = x^m$. Подобно показателю степени m, символ функции φ_m можно записать справа от аргумента: $x^m = x\varphi_m$. Если условиться писать символы функций φ, ψ, ... справа, то естественно записывать их композицию также в правой форме, ибо тогда выполнено правило $(x\varphi)\psi = x(\varphi \diamond \psi)$. Так, в предыдущем примере $x(\varphi_m \diamond \varphi_n) = (x^m)^n = x^{mn} = x\varphi_{mn}$; следовательно, $\varphi_m \diamond \varphi_n = \varphi_{mn}$. Интуитивно преимущество правой композиции заключается в том, что функции пишутся в том же порядке, в каком они действуют.

Лемма 1. *Композиция функций подчиняется ассоциативному закону:*

$$(h \circ g) \circ f = h \circ (g \circ f) = f \diamond (g \diamond h) = (f \diamond g) \diamond h. \tag{8}$$

Предполагается, что все записанные композиции определены.

---

[1]) Едва ли можно считать это определение удачным. Ясно, что это та же композиция — изменены только форма ее записи, а также (см. ниже) форма записи значений, которые принимают функции. Только в случае $S = U$ есть две существенно разные композицни $f \circ g$ и $g \circ f$. — *Прим. перев.*

Интуитивно это очевидно. Как $h \circ (g \circ f)$, так и $(h \circ g) \circ f$ получаются последовательным применением f, g и h именно в таком порядке. То же можно сказать относительно $f \diamond (g \diamond h)$ и $(f \diamond g) \diamond h$.

Формально, пусть $f\colon S \longrightarrow T$, $g\colon T \longrightarrow U$ и $h\colon U \longrightarrow V$. Любому элементу $x \in S$ обе композиции приписывают значения

$$[(hg)f]x = \underset{(hg)f}{(hg)(fx)} = \underset{hg}{h(g(fx))} = \underset{gf}{h((gf)x)} = \underset{h(gf)}{[h(gf)]x}.$$

Проверка каждого равенства состоит в применении определения композиции (7) к той композиции, которая указана под знаком равенства. После этого определение равенства функций (6а) и (6б) доказывает ассоциативный закон $(hg)f = h(gf)\colon S \longrightarrow V$.

Заметим, что для краткости символ композиции $\circ$ в этом доказательстве опущен: композиция обозначается просто посредством записи символов функций рядом.

Тождественные функции 1_S и 1_T в композиции с любой функцией $f\colon S \longrightarrow T$ не меняют ее:

$$f \circ 1_S = 1_T \circ f = f, \quad 1_S \diamond f = f \diamond 1_T = f. \tag{9}$$

Чтобы убедиться в этом, следует заметить, что $(f1_S)s = f(1_S s) = f(s)$ для всех $s \in S$ в силу (8), и применить (6а) и (6б).

Функция $f\colon S \longrightarrow T$ называется *инъективной*, или *инъекцией*, или *вложением*, если из $s \neq s'$ в S следует, что $f(s) \neq f(s')$ в T. Иными словами, инъекция переводит различные элементы своей области в различные элементы кообласти. Инъекция часто называется *взаимно однозначным* отображением S в T. Функция $h\colon S \longrightarrow T$ называется *сюръективной*, или *сюръекцией*, если ее образ совпадает со всей кообластью, т. е. для каждого $t \in T$ существует хотя бы один элемент $s \in S$, такой, что $h(s) = t$. Сюръекции часто обозначаются так: $S \longrightarrow\!\!\!\!\rightarrow T$ и называются отображениями S *на* (все) T. Наконец, *биекцией* называется функция, являющаяся одновременно инъекцией и сюръекцией. Иными словами, биективное отображение *взаимно однозначно* и является отображением *на*. (Чтобы подчеркнуть тот факт, что данная функция $f\colon S \longrightarrow T$ биективна, мы часто будем писать $f\colon S \leftrightarrow T$.)

Например, среди функций из $\mathbf{Z}$ в $\mathbf{Z}$ отображение $n \mapsto -n$ биективно, отображение $n \mapsto 2n$ инъективно, но не сюръективно, а отображение $n \mapsto n^2$ не инъективно и не сюръективно (почему?).

Пример 3. Пусть $\mathbf{P}$ — множество всех положительных целых чисел, $\sigma\colon \mathbf{P} \longrightarrow \mathbf{P}$ — функция следования Пеано, $\sigma(n) = n + 1$ для всех $n \in P$. Это инъекция, но не сюръекция.

Пример 4. Пусть $\mathbf{n} = \{1, 2, \ldots, n\}$ — множество положительных целых чисел $k \leqslant n$. К числу основных функций на нем принадлежит *циклическая перестановка*: $\nu_n\colon \mathbf{n} \longrightarrow \mathbf{n}$. Она сопоставляет

всем числам $k \in \mathbf{n}$, кроме n, следующее $k+1$, а n переводит в 1. Это биекция. При $n=3$ функция ν_3 задается предписанием $1 \mapsto 2$, $2 \mapsto 3$, $3 \mapsto 1$. В общем случае

$$\nu_n\colon\ 1 \mapsto 2,\ \ 2 \mapsto 3,\ 3 \mapsto 4,\ \ldots,\ n-1 \mapsto n,\ \ n \mapsto 1.$$

Характеристические функции. Характеристическая функция подмножества S данного множества U, $e_S\colon U \to \{0, 1\}$, определяется предписанием

$$e_S(x) = \begin{cases} 1, & \text{если } x \in S, \\ 0, & \text{если } x \notin S. \end{cases} \tag{10}$$

Пусть, например, $U = \{1, 2, 3\}$ — множество, которое мы обозначим также символом **3**. Вот списки значений функций e_S, отвечающих всевозможным подмножествам $S \subset U$:

$$\varnothing \mapsto (0, 0, 0),\ \{1\} \mapsto (1, 0, 0),\ \{2\} \mapsto (0, 1, 0),\ \{3\} \mapsto (0, 0, 1),$$
$$\{1, 2\} \mapsto (1, 1, 0),\ \{1, 3\} \mapsto (1, 0, 1),\ \{2, 3\} \mapsto (0, 1, 1),$$
$$\mathbf{3} = \{1, 2, 3\} \mapsto (1, 1, 1).$$

Описанное отображение $b\colon S \to e_S$ также является весьма важной биекцией, к которой мы вернемся в § 1.6. Ее областью является множество $\mathcal{P}(U)$ всех подмножеств $S \subset U$, а ее кообласть (и образ) совпадает с множеством $\mathbf{2}^U$ всех функций $e_S\colon U \to \mathbf{2}$. Например, при $U = \mathbf{3}$ отображение b является биекцией между $\mathcal{P}(\mathbf{3})$ и множеством всех последовательностей из символов 0, 1 длины три. Опуская скобки и запятые, мы можем обозначить любой элемент из $\mathcal{P}(\mathbf{3})$ одной из восьми *двоичных троек*: 000, 100, 010, 001, 110, 101, 011, 111. Это соглашение окажется очень удобным позже, когда мы сможем представить себе список этих троек упорядоченным в словарном, или лексикографическом, порядке:

$$000,\ 001,\ 010,\ 011,\ 100,\ 101,\ 110,\ 111.$$

1.4. ОБРАТНЫЕ ФУНКЦИИ

Если левая композиция функций $f\colon S \to T$ и $g\colon T \to S$ совпадает с тождественной функцией 1_S на S, мы будем говорить, что g — *левая обратная* к f, а f — *правая обратная* к g. Если, кроме того, $f \circ g = 1_T$, мы будем говорить, что g — *двусторонне обратна* к f и, по симметрии, f *двусторонне обратна* к g. Функция, имеющая двусторонне обратную, называется *обратимой*.

Теорема 2. *Функция обратима слева тогда и только тогда, когда она взаимно однозначна (инъективна). Функция обратима справа тогда и только тогда, когда она сюръективна.*

Доказательство. Покажем сначала, что любая обратимая слева функция взаимно однозначна. Пусть $g\colon T \to S$—функция, обратная слева к $f\colon S \to T$. Предположим, что $f(s) = f(s')$. Тогда, по определению,

$$s = 1_S(s) = g(f(s)) = g(f(s')) = 1_S(s') = s'.$$

Таким образом, из предположения, что $f(s) = f(s')$, мы вывели, что $s = s'$. Это доказывает инъективность f.

Остается проверить обратное утверждение, чтобы закончить доказательство первой части теоремы: нам нужно показать, что если f инъективна, то у нее существует левая обратная: $g_1 f = 1_S$. С этой целью выберем элемент $s_1 \in S$ и определим $g_1\colon T \to S$ так:

$$g_1(t) = \begin{cases} s, & \text{если } t = f(s) \text{ для некоторого } s \in S, \\ s_1 & \text{в противном случае, т. е. если } t \notin \operatorname{Im} f. \end{cases}$$

Тогда $(g_1 \circ f)(s) = g_1(f(s)) = s$ для всех $s \in S$, так что, по определению, $g_1 \circ f = 1_S$ и g_1 является левой обратной к f.

Аналогично проверяется вторая часть теоремы. Если $t \in T$ и $fh = 1_T$, то $t = 1_T(t) = f(h(t))$, так что любой элемент $t \in T$ является образом $f(s)$ подходящего элемента $s = h(t) \in S$, и f сюръективна. Обратно, если $f\colon S \to T$ сюръективна, то каждый элемент $t \in T$ является образом $t = f(s)$ хотя бы одного элемента $s \in S$. Для каждого элемента $t \in T$ выберем[1]) из множества всех s, переходящих в t, один представитель и обозначим его, скажем, через $h(t)$. Тогда мы получим функцию $h\colon T \to S$, такую, что $f(h(t)) = t$ для всех $t \in T$, т. е. $fh = 1_T$, как утверждалось.

Следствие. *Функция $f\colon S \to T$ является биекцией тогда и только тогда, когда она обратима слева и справа.*

Теорема 3. *Следующие свойства функции $f\colon S \to T$ эквивалентны:*

(i) *f—биекция;*

(ii) *f обладает левой обратной g и правой обратной h.*

Если эти свойства выполняются, то

(iii) *все обратные к f функции (левые, правые и двусторонние) совпадают. Эта единственная обратная к f функция f^{-1} биективна, и*

$$(f^{-1})^{-1} = f. \tag{11}$$

Доказательство. Эквивалентность свойств (i) и (ii) является просто переформулировкой предыдущего следствия. Из условия (ii) вытекает, что

$$g = g1_T = g(fh) = (gf)h = 1_S h = h. \tag{12}$$

[1]) В этом рассуждении используется аксиома выбора; см. ее критическое обсуждение в гл. 14.

Это показывает, что все левые и правые обратные к f совпадают, и доказывает (iii). Наконец, биективная функция f обратна к своей обратной функции $g=h$, ибо $gf=1_S$ и $fh=1_T$. Следовательно, обратная функция биективна и имеет f в качестве своей единственной обратной, т. е. $(f^{-1})^{-1}=f$.

Частичные функции. Хотя в этой книге мы условились считать, что «функция» ставит в соответствие любому элементу своей области некоторое значение, встречаются важные функции, вроде $1/x$ и $\log x$, определенные для некоторых, но не всех значений x ($x=0$, например). Мы будем называть *частичной функцией* f: $X \to Y$ правило, которое ставит в соответствие каждому элементу x из некоторого *подмножества* $D \subset X$ (*область* f) единственное значение из Y.

1.5. ФУНКЦИИ ИЗ S В S

Рассмотрим теперь более подробно множество S^S всех фуикций с областью и кообластью S. Оно образует алгебраическую систему $[S^S, \circ]$ с бинарной операцией $\circ$ (левая композиция). Поскольку кообласть любой функции из S^S совпадает с областью любой другой функции, композиции $f \circ g$ и $f \diamond g = g \circ f$ [1]) всегда существуют в S^S. Кроме того, обе эти операции удовлетворяют следующему *закону ассоциативности*:

$$f \circ (g \circ h) = (f \circ g) \circ h, \quad f \diamond (g \diamond h) = (f \diamond g) \diamond h \tag{13}$$

для всех f, g, $h \in S^S$. Далее, в этой алгебраической системе тождественная функция 1_S удовлетворяет соотношениям (9):

$$1_S \circ f = f \circ 1_S = f, \text{ т. е. } f1_S = 1_S f = f, \text{ для всех } f \in S^S. \tag{14}$$

Заметим, что уравнения (13) и (14) имеют одинаковый вид для операций $\circ$ и $\diamond$. Как мы уже указывали, эти символы операций обычно опускаются, а результат композиции обозначается просто путем записи символов функций рядом — как это принято для записи умножения в школьной алгебре. Так, (13) превращается в равенство $f(gh)=(fg)h$ для левой и правой композиции, означающее, что обе они являются ассоциативными бинарными операциями. Тождества (14) означают, что функция 1_S *поглощается* при ее композиции с любой другой функцией (как слева, так и справа).

---

[1]) Здесь $f \diamond g$ по существу отличается от $f \circ g$, а не только формой записи, как в § 1.3. — *Прим. перев.*

Квадрат любой функции $f\colon S \longrightarrow S$ определяется как функция $f^2 = f \circ f = f \diamond f$, т. е. как результат двукратного применения f. Когда $f^2 = f$, функция f называется *идемпотентом*, или *проектором*. Например, ортогональный проектор (x, y)-плоскости на x-ось или на y-ось является идемпотентом. Заметим также, что два эти проектора обладают свойством $f \circ g = g \circ f$ (или, что то же самое, $g \diamond f = f \diamond g$), поскольку $f \circ g$ и $g \circ f$ отображают любую точку плоскости в начало координат (0, 0). Вообще пары функций f, g со свойством $fg = gf$ называются *коммутирующими*, или *перестановочными*.

Применительно к любой из двух композиций мы можем назвать функцию $g \in S^S$ *левой обратной* к функции $f \in S^S$, если $fg = 1_S$. Аналогично, *правая обратная* к f — это функция h, такая, что $fh = 1_S$. Согласно теореме 2, f имеет левую обратную функцию относительно левой композиции, и только тогда, когда f является инъекцией, значит, то же относится к существованию правой обратной функции к f отиосительно правой композиции. Аналогично, f имеет обратную справа функцию относительно левой композиции тогда и только тогда, когда f — сюръекция.

Двусторонняя обратная существует в точности для биективных функций независимо от того, какую из композиций (левую или правую) мы рассматриваем. Функция, обратная к f, обозначается через f^{-1}. Она существует в том и только том случае, если f биективна, и удовлетворяет соотношениям

$$f^{-1}f = ff^{-1} = 1_S \text{ (и для левой, и для правой композиции).} \quad (15)$$

Моноиды. Как мы выяснили, множество S^S всех функций из множества S в себя является *алгебраической системой* $[S^S, \circ, 1_S]$ с операцией левой композиции и тождественным элементом (или единицей). Системы, подобные $[S^S, \circ, 1_S]$ и $[S^S, \diamond, 1_S]$, с ассоциативной бинарной операцией и единицей называются *моноидами*. Мы будем систематически изучать их в гл. 7.

Пример 5. Пусть $S = \{a, b\}$; обозначим элементы S^S буквами: $e = 1_S$, α: $a \mapsto a$, $b \mapsto a$; β: $a \mapsto b$, $b \mapsto b$; f: $a \mapsto b$; $b \mapsto a$. Тогда системы $[S^S, \circ, 1_S]$ и $[S^S, \diamond, 1_S]$ имеют следующие таблицы умножения:

$\circ$	e	f	α	β
e	e	f	α	β
f	f	e	β	α
α	α	α	α	α
β	β	β	β	β

$\diamond$	e	f	α	β
e	e	f	α	β
f	f	e	α	β
α	α	β	α	β
β	β	α	α	β

Еще раз напомним, что единица $1 = 1_S$ в моноиде S^S поглощается своими соседями в любом произведении: $1f = f$, $\alpha 1 \beta = \alpha\beta$. Элементы 1, α, β идемпотентны, т. е. $1^2 = 1$, $\alpha^2 = \alpha$ и $\beta^2 = \beta$.

Функция f не идемпотентна, ибо $f^2 = e \neq \beta$. Кроме того, $\alpha \circ \beta \neq \beta \circ \alpha$, так что операция $\circ$ (и противоположная к ней $\diamond$) не коммутативна. Заметим далее, что у функций α и β нет обратных, ни правой, ни левой, а f обратна сама к себе (т. е. $f^2 = 1$).

Конечные множества. Конечные множества играют в прикладной математике особую роль, поскольку только они допускают физическую реализацию. Неформально говоря, множество конечно, если его можно «пересчитать», т. е. если для некоторого положительного целого числа n существует биекция f из множества $\mathbf{n} = \{1, \ldots, n\}$ положительных целых чисел $k \leqslant n$ в множество S. Если обозначить $f(k) \in S$ через s_k, то это означает, очевидно, что все элементы S можно расположить в конечную последовательность $\{s_1, s_2, \ldots, s_n\}$.

С детских лет нам известно, что, в каком бы порядке ни пересчитывать элементы конечного множества, номер последнего по счету элемента остается прежним. Иными словами, между множествами $\mathbf{m} = \{1, \ldots, m\}$ и $\mathbf{n} = \{1, \ldots, n\}$ биекция существует тогда и только тогда, когда $m = n$. Еще одна более яркая формулировка того же принципа: если n птиц сидят в $< n$ гнездах, то хотя бы в одном гнезде сидит не менее двух птиц. В частности, любая инъекция конечного множества S в себя является сюръекцией. Верно также и обратное: любая сюръекция конечного множества S в себя является инъекцией.

В § 1.9 мы вкратце докажем эти основные факты. Пока заметим только, что для инъекций и биекций бесконечных множеств эти утверждения *неверны*. Так, функция следования Пеано $\sigma\colon \mathbf{P} \longrightarrow \mathbf{P}$ из примера 3 имеет бесконечно много левых обратных: положив для любого $m \in \mathbf{P}$

$$\tau_m(n) = \begin{cases} n-1 & \text{при } n > 1, \\ m & \text{при } n = 1, \end{cases} \qquad (15')$$

получаем, очевидно, $\tau_m \circ \sigma = 1_P$.

УПРАЖНЕНИЯ Б

1. Рассмотрим функции $f(n) = 3n$, $g(n) = 3n+1$ и $h(n) = 3n+2$ из $\mathbf{Z}$ в $\mathbf{Z}$. Построить функцию, которая была бы обратной слева к f, g и h одновременно.

2. а) Если fg определена и обе функции f, g имеют левые обратные, показать, что fg имеет левую обратную.

б) Показать на примере, что обратное утверждение верно не всегда: fg может иметь левую обратную функцию даже тогда, когда f ее не имеет.

3. а) Показать, что композиция gf двух любых инъекций $f\colon S \to T$ и $g\colon T \to U$ является инъекцией.

б) Доказать то же утверждение для сюръекций.

4. а) Сколько имеется сюръекций из трехэлементного множества на двухэлементное?

б) Сколько имеется инъекций из трехэлементного множества в четырехэлементное?

5. а) Рассмотрим отображения $x \mapsto x^2$ каждого из следующих множеств в себя:

$$\mathbf{P},\ \mathbf{Z},\ \mathbf{Q},\ \mathbf{R},\ \mathbf{C}.$$

Определить образ каждого из этих отображений и выяснить, являются ли они инъекциями.

б) Тот же вопрос для $x \mapsto x^3$.

6. а) Какие подмножества множества $\mathbf{4} = \{1, 2, 3, 4\}$ представлены следующими двоичными последовательностями: 1001, 0110, 1101, 0010.

б) Показать, что если двоичная последовательность $\mathbf{n} = n_1 n_2 n_3 n_4$ представляет множество S, то последовательность $\mathbf{n}' = (1-n_1)(1-n_2)(1-n_3)(1-n_4)$ представляет его дополнение. Проиллюстрировать это на примерах из п.а).

*7. Распространить определение композиции gf на случай, когда область g содержит кообласть f. Показать, что результат упр. 3.а продолжает быть верным, а результат упр. 3.б становится неверным.

8. Пусть $n(X)$ — число элементов множества X. Тогда следующие тождества верны для любых трех конечных множеств:

а) $n(A)+n(B)=n(A \cap B)+n(A \cup B)$;

б) $n(A \cup B \cup C)+n(A \cap B)+n(B \cap C)+n(C \cap A)=$
$=n(A \cap B \cap C)+n(A \cup B)+n(B \cup C)+n(C \cup A)$.

9. В кровопролитном бою не менее 70% воинов потеряли глаз, не менее 75% — ухо, не менее 80% — руку и не менее 85% — ногу. Оценить снизу число воинов, потерявших одновременно глаз, ухо, руку и ногу (Льюис Кэрролл).

10. а) Доказать, что непрерывная вещественная функция f: $[a, b] \to [c, d]$ — взаимно однозначна тогда и только тогда, когда она монотонна.

б) Доказать, что такая функция является биекцией тогда и только тогда, когда она монотонна и либо (i) $f(a)=c$, $f(b)=d$, либо (ii) $f(a)=d$, $f(b)=c$.

*11. Является ли функция e^x взаимно однозначной на $\mathbf{R}$ и на $\mathbf{C}$? Найдите в каждом случае ее образ.

1.6. СУММЫ, ПРОИЗВЕДЕНИЯ И СТЕПЕНИ

Законы коммутативности и ассоциативности L2 и L3, а также первый закон дистрибутивности L6 имеют хорошо известные аналоги в арифметике. Покажем, что это не случайно: эти и многие другие законы арифметики неотрицательных целых чисел можно вывести из основных свойств множеств и функций.

Будем говорить, что конечное множество S состоит из m элементов, или что его *кардинальное число* равно $\mathbf{m}$, если существует биекция b: $S \leftrightarrow \mathbf{m} = \{1, \ldots, m\}$. Будем считать также, что пустое множество имеет $\mathbf{0}$ элемеитов. Операция сложения в арифметике основана на следующем теоретико-множественном понятии:

Определение. Пусть даны множества S и T. Их *разделенным объединением*, или *суммой*, называется множество $D = S \amalg T$ вместе с фиксированными биекциями i: $S \leftrightarrow S^* \subset D$ и

$j\colon T \leftrightarrow T^* \subset D$, такими, что S^* и T^* — взаимно дополнительные подмножества D.

Ясно, что любые две суммы S и T находятся в канонической биекции. Простейший случай получается тогда, когда $S \subset U$, $T \subset U$ и $S \cap T = \varnothing$. Тогда можно взять просто $S \amalg T = S \cup T$, $i = 1_S$ и $j = 1_T$. Из равенства (4) находим

$$n(S \amalg T) = n(S) + n(T) \tag{16}$$

(в § 2.6 мы покажем, что равенство (16) можно взять в качестве определения сложения).

Законы коммутативности и ассоциативности сложения неотрицательных целых чисел вытекают из (16) и существования очевидных биекций; итак,

$$\alpha\colon S \amalg T \leftrightarrow T \amalg S, \quad \beta\colon S \amalg (T \amalg U) \leftrightarrow (S \amalg T) \amalg U. \tag{16'}$$

Декартовы произведения. Подобно тому как сумма чисел связана с суммой множеств, произведение чисел связано с *декартовым*, или *прямым*, *произведением* множеств. Декартово произведение $S \times T$ множеств S и T определяется как множество всех упорядоченных пар вида (s, t), где s принадлежит S, а t принадлежит T. Так, если $\mathbf{R}$ — множество всех вещественных чисел, то $\mathbf{R} \times \mathbf{R}$ — множество всех упорядоченных пар (x, y) вещественных чисел. Иными словами, $\mathbf{R} \times \mathbf{R}$ есть множество декартовых координат (относительно выбранных осей) всех точек плоскости. Аналогично, $\mathbf{Z} \times \mathbf{Z}$ — решетка точек с целыми координатами (i, j) на этой плоскости.

Как и выше, имеются естественные биекции:

$$S \times T \leftrightarrow T \times S, \quad S \times (T \times U) \leftrightarrow (S \times T) \times U. \tag{17}$$

Для числа элементов декартова произведения двух конечных множеств имеет место такая формула умножения:

$$n(S \times T) = n(S)\, n(T). \tag{17'}$$

Наконец, вместо естественных инъекций $f\colon S \longrightarrow S \cup T$ и $g\colon T \longrightarrow S \cup T$ определены естественные сюръекции:

$$p\colon S \times T \twoheadrightarrow S, \quad p(s, t) = s, \tag{18}$$

$$q\colon S \times T \twoheadrightarrow T, \quad q(s, t) = t. \tag{18'}$$

Эти сюръекции принято называть *проекторами* на сомножители произведения $S \times T$ по аналогии с проекторами плоскости на оси декартовой координатной системы [1]).

[1]) Внимательный читатель заметит, что эта аналогия не вполне точна: оси декартовой системы координат принято представлять себе вложенными в плоскость. В общем случае также имеются инъекции $S \to S \times T$, $s \mapsto (s, t_0)$; $T \to S \times T$, $t \mapsto (s_0, t)$, где $s_0 \in S$, $t_0 \in T$ — какие-то элементы, не определенные однозначно. — *Прим. перев.*

Операция возведения в степень. Множество всех функций из множества S в множество T обозначается через T^S. Если S и T конечны и состоят из $n(S)$, $n(T)$ элементов соответственно, то имеет место формула

$$n(T^S)=n(T)^{n(S)}. \tag{19}$$

Действительно, для каждого $s_i \in S$ значение $f(s_i) \in T$ можно выбрать ровно $n(T)$ способами. Все эти выборы независимы, поэтому f выбирается одним из $n(T)\times \ldots \times n(T)$ ($n(S)$ раз) способов.

Особо интересен для нас случай двухэлементной области $\{0, 1\}=2$. Как в § 1.3, формула

$$e_S(x)=\begin{cases} 1, & \text{если } x\in S, \\ 0, & \text{если } x\in S' \text{ (т. е. } x\notin S), \end{cases} \tag{20}$$

определяет естественную биекцию $b:\mathcal{P}(U)\leftrightarrow 2^U$, которая ставит в соответствие каждому подмножеству $S\subset U$ его характеристическую функцию $e_S\in 2^U$. Эта же биекция отождествляет подмножества множества $\mathbf{n}=\{1, 2, \ldots, n\}$ с двоичными последовательностями длины n. Мы будем постоянно пользоваться ею в последующих главах.

В общем случае имеются естественные биекции:

$$T^{R\sqcup S}\leftrightarrow T^R\times T^S,\quad (T\times U)^S\leftrightarrow T^S\times U^S,\quad (T^S)^R\leftrightarrow T^{S\times R}. \tag{21}$$

Из (19) ясно, что в применении к конечным множествам эти биекции устанавливают обычные правила действия со степенями в области целых положительных чисел. Именно, пользуясь (19), (21), (16) и (17′), получаем

$$a^{r+s}=a^ra^s,\qquad (ab)^s=a^sb^s,\qquad (a^s)^r=a^{sr}. \tag{21′}$$

* Универсальность. Суммы и декартовы произведения обладают интересными «универсальными» средствами относительно отображений. Пусть, как выше, $D=S\sqcup T$, и пусть даны две функции $f: S\longrightarrow X$ и $g: T\longrightarrow X$ с областями S, T и общей кообластью X. Тогда мы можем определить фуикцию $h: D\longrightarrow X$ для элементов $t^*=j(t)\in T^*$ формулой $h(t^*)=g(t)$, а для элементов $s^*=i(s)\in S^*$ формулой $h(s^*)=f(s)$. Тогда $f=h\circ i$, $g=h\circ j$ и $h: D\longrightarrow X$ — единственная функция, обладающая таким свойством. Это доказывает следующий результат:

Теорема 4. *Существует единственный способ дополнить диаграмму из сплошных стрелок на рис.* 1.2,*а пунктирной стрелкой, изображающей функцию* $h:D\longrightarrow X$, *так чтобы полученная диаграмма была коммутативной.*

---

*) Материал, отмеченный звездочкой, можно опускать, не теряя нити изложения.

Пояснение. Диаграмма, состоящая из множеств и отображений, называется *коммутативной*, если любые два отображения, отвечающие композиции указанных в ней отображений и имеющие общую область и кообласть, совпадают. В применении к рис. 1.2,*а* это требование означает, что $h \circ i = f$ и $h \circ j = g$, в применении к рис. 1.2,*б* — что $p \circ h = f$ и $q \circ h = g$.

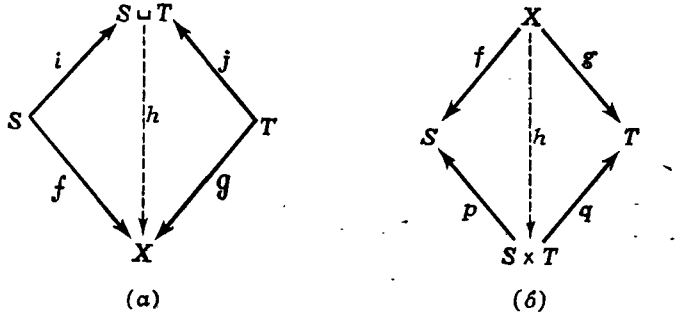


Рис. 1.2. Диаграмма универсальных отображений.

Обратно, можно показать, что если функции $i\colon S \to D$ и $j = T \to D$ обладают тем свойством, что для любого множества X и любых функций $f\colon S \to X$ и $g\colon T \to X$ такая функция $h\colon D \to X$ существует и единственна, то i и j инъективно отображают S и T соответственно в дополнительные подмножества S^* и T^* множества X. Иными словами, свойство, описываемое диаграммой 1.2,*а*, однозначно характеризует разделенные объединения [1]).

Пусть теперь снова $f\colon X \to S$ и $g\colon X \to T$ — две функции с общей областью. Тогда условие

$$h(x) = (f(x),\ g(x)) \text{ для всех } x \in X \qquad (22)$$

определяет единственную функцию $h = f \times g\colon X \to S \times T$. Обозначая через p, q проекторы, определенные формулами (18) и (18′), получаем, очевидно, $p \circ h = f$ и $q \circ h = g$. Обратно, если для функции $h\colon X \to S \times T$ мы имеем $p \circ h = f$ и $q \circ h = g$, то для любого элемента $x \in X$ из равенства $h(x) = (s,\ t)$ следует, что $s = (p \circ h)(x) = f(x)$ и $t = (q \circ h)(x) = g(x)$.

Итак, мы показали, что функция $h = f \times g$, определенная формулой (22), — это единственная функция, которая превращает диаграмму рис. 1.2,*б* в коммутативную в том смысле, что

$$h \diamond p = p \circ h = f, \qquad h \diamond q = q \circ h = q.$$

---

[1]) Доказательство см. на стр. 30 книги Mac Lane S., Birkhoff G., Algebra, Macmillan, 1967. Существенным обстоятельством является здесь единственность универсальных отображений, доказанная на стр. 28 упомянутой книги.

УПРАЖНЕНИЯ В

1. а) Показать, что у множества, состоящего из n элементов, имеется ровно 2^n разных подмножеств.

б) Показать, что если $m < n$, то у множества, состоящего из n элементов, имеется ровно $n!/m!\,(n-m)!$ разных подмножеств из m элементов.

2. а) Доказать со всеми подробностями, что имеется ровно n^m различных функций $f\colon S \longrightarrow T$ из множества, состоящего из m элементов, в множество, состоящее из n элементов.

б) Сколько имеется инъективных функций в п. а?

3. Показать, что (заглавные) буквы латинского алфавита нельзя однозначно представить последовательностями длины четыре, состоящими из символов · (точка) и — (тире). Однако это можно сделать, если разрешить пользоваться последовательностями длины четыре, три и два (азбука Морзе).

4. Совпадают ли множества $A = \varnothing$ и $B = \{\varnothing\}$? Имеется ли биекция из A в B?

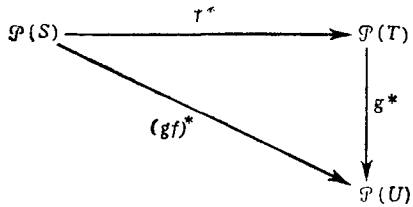
5. Для произвольного множества U указать биекцию $2^U \leftrightarrow \mathcal{P}(U)$, где 2 — множество чисел $\{1, 2\}$.

6. Показать, что функции τ_1, τ_2, определенные предписанием (15′), имеют по две правых обратных.

7. Пусть X, Y и Z — подмножества множества U, такие, что $Y \cap Z = \varnothing$, $Y \cup Z = X$. Построить биекцию b: $\mathcal{P}(X) \leftrightarrow \mathcal{P}(Y) \times \mathcal{P}(Z)$.

8. Пусть X, Y, Z — такие же, как в упр. 7. Построить биекцию $S^X \leftrightarrow S^Y \times S^Z$.

*9. Дана функция $f\colon S \longrightarrow T$. Построить новую функцию $f^*\colon \mathcal{P}(S) \longrightarrow \mathcal{P}(T)$, которая ставит в соответствие каждому подмножеству $R \subset S$ его образ $f^*(R)$, состоящий из всех $t \in T$, для которых существует такой элемент $r \in R$, что $f(r) = t$. Аналогично, построить по функции $g\colon T \longrightarrow U$ функцию g^*. Доказать, что следующая диаграмма коммутативна:



1.7. АКСИОМЫ ПЕАНО

Древнейшая и самая фундаментальная математическая система — множество **P** положительных целых чисел. По словам Кронекера, «бог создал целые положительные числа, все остальное придумал человек». Эту систему [**P**, +, ×] обычно рассматривают как наделенную двумя бинарными операциями: сложением и умножением.

Однако, как показал итальянский математик Дж. Пеано (1858—1932), эту систему можно определить более просто: как алгебру [**P**, σ] с одной унарной операцией σ «счета». Как унарная алгебра [**P**, σ], система целых положительных чисел харак-

теризуется тремя постулатами, налагаемыми на унарную операцию $\sigma \in \mathbf{P}^{\mathbf{P}}$ — *функцию следования* $\sigma\colon \mathbf{P} \to \mathbf{P}$. Это *аксиомы Пеано*:

S1. Если $\sigma(m) = \sigma(n)$, то $m = n$ (σ взаимно однозначна).

S2. Не существует такого $n \in \mathbf{P}$, что $\sigma(n) = 1$.

S3. Пусть подмножество $S \subset \mathbf{P}$ удовлетворяет двум условиям: а) $1 \in S$ и б) из того, что $n \in S$ следует $\sigma(n) \in S$. Тогда $S = \mathbf{P}$.

Аксиома S3 — наиболее сильная из аксиом Пеано. Она называется *аксиомой индукции*, ибо лежит в основе доказательства методом (финитной) индукции (см. ниже).

В терминах унарной операции σ бинарные операции сложения и умножения в множестве $\mathbf{P}$ без труда определяются следующими рекурсивными описаниями.

Пример 6. Для любого $m \in \mathbf{P}$ фуикция $\sigma^m\colon n \mapsto m + n$ определяется простой рекурсией:

$$m + 1 = \sigma^m(1) = \sigma(m), \tag{23}$$

$$m + (n + 1) = \sigma^m(\sigma n) = \sigma(\sigma^m(n)). \tag{24}$$

Для фиксированного $m \in \mathbf{P}$ рассмотрим множество S_m всех $n \in \mathbf{P}$, для которых $\sigma^m(n)$ определена. В силу (23), $1 \in S_m$, а согласно (24), из того, что $n \in S$, следует, что $\sigma(n) \in S_m$. Значит, в силу аксиомы S3, функция $\sigma^m(n) = m + n$ определена для всех $n \in \mathbf{P}$.

Если теперь m будет пробегать все множество $\mathbf{P}$, то мы получим бинарную операцию «сложения» $+\colon \mathbf{P} \times \mathbf{P} \to \mathbf{P}$, свойства которой будут выведены в следующем параграфе из аксиом Пеано.

Пример 7. Для фиксированного $m \in P$ определим функцию $p_m\colon n \mapsto nm$ следующей простой рекурсией:

$$p_m(1) = m, \tag{25}$$

$$p_m(\sigma(n)) = m + p_m(n) = \sigma^m(p_m(n)). \tag{26}$$

Как в примере 6, проверяется, что множество S_m всех $n \in \mathbf{P}$, для которых $p_m(n)$ определена, совпадает со всем $\mathbf{P}$. Поэтому, если m и n пробегают все $\mathbf{P}$, мы получаем бинарную операцию «умножения» на $\mathbf{P}$. Ее свойства также будут выведены из аксиом Пеано в следующем параграфе.

Пример 8. Для любого фиксированного $m \in \mathbf{P}$ определим функцию $e_m\colon m \mapsto m^n$ рекурсией

$$e_1(m) = m, \tag{27}$$

$$e_{\sigma(n)}(m) = me_n(m) = p_m(e_n(m)). \tag{28}$$

Ее обычные свойства также можно доказать по индукции, исходя из этого определения.

1.8. ФИНИТНАЯ ИНДУКЦИЯ

Предположим, что задана последовательность *высказываний* $P(1)$, $P(2)$, $P(3)$, ..., каждое из которых может быть либо истинным, либо ложным. Принцип финитной (конечной) индукции утверждает, что для доказательства истинности высказываний $P(n)$ для всех $n \in \mathbf{P}$ достаточно установить, во-первых, истинность $P(1)$ и, во-вторых, истинность бесконечной последовательности импликаций

$$P(1) \Rightarrow P(2) \Rightarrow P(3) \Rightarrow \ldots \Rightarrow P(n) \Rightarrow P(n+1) \Rightarrow \ldots .$$

Второе условие можно записать как утверждение, что (для всех n) из истинности $P(n)$ следует истинность $P(\sigma(n))$. Здесь, как в § 1.7, $\sigma(n) = \sigma n = n+1$ — функция следования. Вместо $\sigma(m)$ мы иногда будем писать σm, опуская скобки: $m+1 = \sigma^m(1) = \sigma(m) = \sigma m$.

Принцип финитной индукции немедленно следует из аксиомы Пеано S3. Действительно, пусть $T \subset \mathbf{P}$ — множество всех положительных целых чисел n, для которых $P(n)$ истинно. Если $1 \in T$ и если из того, что $n \in T$, следует, что $\sigma(n) \in T$, то, в силу аксиомы S3, $T = \mathbf{P}$.

Индукция является стандартным методом доказательства в арифметике и теории чисел. Чтобы проиллюстрировать это, докажем коммутативность и ассоциативность сложения.

Теорема 5. *Определим сложение в* $[\mathbf{P}, \sigma]$ *формулой* $m+n+\sigma^m(n)$. *Тогда для всех* m, $n \in \mathbf{P}$

$$m+(n+r) = (m+n)+r \quad (\textit{ассоциативность}), \tag{29}$$

$$m+n = n+m \quad (\textit{коммутативность}). \tag{30}$$

Для доказательства ассоциативности обозначим через $P(r)$ высказывание об истинности (29) для этого r и всех m, $n \in \mathbf{P}$. Тогда $P(1)$ утверждает [если $r=1$ в (29)], что $m+\sigma(n) = \sigma(m+n)$; это верно в силу (24).

Примем теперь предположение индукции $P(r)$:

$$m+(n+r) = (m+n)+r,$$

для всех m, n. Снова, замечая, что $m+\sigma(n) = \sigma(m+n)$ согласно (24), находим

$$m+(n+\sigma(r)) \overset{(24)}{=} m+\sigma(n+r) \overset{(24)}{=} \sigma(m+(n+r)) \overset{P(r)}{=} \sigma((m+n)+r),$$

где над каждым знаком равенства отмечено, на основании чего это равенство справедливо (для всех m, $n \in \mathbf{P}$). Снова применяя (24) к $m+n$ и r, получаем

$$\sigma((m+n)+r) \overset{(24)}{=} (m+n)+\sigma(r).$$

Отсюда окончательно

$$m+(n+\sigma(r))=(m+n)+\sigma(r) \quad \text{для всех } m,\ n\in\mathbf{P}.$$

Но это высказывание совпадает с $P(\sigma(r))$. Это завершает доказательство (29).

Чтобы установить (30), начнем с $n=1$. Обозначим через $P(m)$ высказывание $m+1=1+m$ и снова проведем индукцию. Поскольку $1+1=1+1$, утверждение $P(1)$ истинно. Считая $P(m)$ истинным, с помощью (23) и (29) доказываем истинность высказывания

$$(m+1)+1=(1+m)+1=1+(m+1),$$

т. е. $P(m+1)$. Это завершает индукцию и доказывает истинность $P(m)$ для всех m.

Теперь обозначим через $Q(n)$ высказывание $m+n=n+m$ для всех m. Мы доказали $Q(1)$. Считая $Q(n)$ истинным, мы затем можем последовательно доказать, что

$$m+(n+1)=(m+n)+1=(n+m)+1=$$
$$=n+(m+1)=n+(1+m)=(n+1)+m,$$

применяя (29), $Q(n)$, (29), $P(m)$ и (29) (в указанном порядке). Это означает, что $Q(n+1)$ истинно, и завершает индукцию. Но истинность $Q(n)$ для всех n, очевидно, и означает (30).

Аналогичные доказательства можно провести для законов ассоциативности и коммутативности умножения целых положительных чисел:

$$m(nr)=(mn)r \qquad \text{для всех } m,\ n,\ r\in\mathbf{P}, \tag{31}$$

$$mn=nm \qquad \text{для всех } m,\ n\in\mathbf{P}. \tag{32}$$

Три закона возведения в степень (21′) тоже можно доказать (для любых a, r, $s\in\mathbf{P}$), исходя из определений примера 8, методом конечной индукции. Мы, однако, опустим эти доказательства (см. книгу Фермана [6], приведенную в литературе на стр. 38). Вместо этого мы докажем один результат подобного типа, относящийся к функциям. Для любой функции $f\colon S\to S$ положим

$$f^1=f \quad \text{и} \quad f^{\sigma(n)}=f^n\circ f. \tag{33}$$

Это правило определяет $f^n\colon S\to S$ для всех $n\in\mathbf{P}$.

Теорема 6. *Пусть $f\colon S\to S$ — любая функция, отображающая множество S в себя. Тогда*

$$f^m\circ f^n=f^{m+n} \quad \textit{для всех } m,\ n\in\mathbf{P}. \tag{33′}$$

Доказательство. При $n=1$ имеем $f^m\circ f^1=f^m\circ f=f^{\sigma m}=f^{m+1}$; в силу определений f^1, $f^{\sigma m}$ и $m+1$ (как σm). Будем считать

теперь, что (33′) истинно для конкретного $n \in \mathbf{P}$ (предположение индукции). Тогда

$$f^m \circ f^{\sigma(n)} = f^m \circ (f^n \circ f) = (f^m \circ f^n) \circ f = f^{m+n} \circ f = f^{\sigma(m+n)} = f^{(m+n)+1} =$$
$$= f^{m+(n+1)} = f^{m+\sigma(n)}.$$

Здесь мы последовательно пользуемся (33), ассоциативностью композиции (§ 1.3), предположением индукции (33), а затем меняем обозначения, применяем (29) и снова меняем обозначения (в доказательстве теоремы 5 мы проверили, что $\sigma(m+n) = m + \sigma(n)$).

Законы дистрибутивности. Пользуясь индукцией, нетрудно вывести из аксиом Пеано также и дистрибутивные законы. Обозначим через $P(r)$ высказывание, что $m(n+r) = mn + mr$ для всех $m, n \in \mathbf{P}$. Тогда $P(1)$ следует из (26) и (30). Затем, приняв $P(r)$, получаем

$$m(n + \sigma(r)) = m(n + (r+1)) = m((n+r)+1) =$$
$$= m(n+r) + m = mn + mr + m = \text{ (в силу } P(r))$$
$$= mn + (mr + m) = mn + m\sigma r.$$

Символ суммирования. Значение символа $\sum$ определяется рекурсией:

$$\sum_{j=1}^{1} x_j = x_1, \qquad \sum_{j=1}^{m+1} x_j = \left(\sum_{j=1}^{m} x_j \right) + x_{m+1}.$$

После этого индукцией по m без труда доказывается **закон левой дистрибутивности** для m-членных сумм:

$$\left(\sum_{j=1}^{m} x_j \right) y = \sum_{j=1}^{m} (x_j y).$$

Аналогичное рассуждение доказывает правую **дистрибутивность**:

$$x \left(\sum_{k=1}^{n} z_k \right) = \sum_{k=1}^{n} (x z_k).$$

Объединяя эти две формулы, получаем *обобщенный закон дистрибутивности*:

$$\left(\sum_{j=1}^{m} x_j \right) \left(\sum_{k=1}^{n} y_k \right) = \sum_{j=1}^{m} \sum_{k=1}^{n} (x_j y_k).$$

* Обобщенный закон ассоциативности. Из школьной алгебры известен следующий несколько неточно формулируемый принцип: *результат применения любой бинарной ассоциативной операции к последовательности из n членов зависит только от порядка этих членов, но не от порядка, в котором применяется операция.*

Доказательство этого принципа, который называется *обобщенным законом ассоциативности*, на основе простейших трехчленных соотношений ассоциативности (31) требует довольно тонких комбинаторных рассуждений. (На самом деле их требует уже точная формулировка принципа.)

Например, применить бинарную операцию f к четырехчленной последовательности x, y, z, w можно *пятью* существенно разными способами. Полагая для краткости $f(x, y) = (xy)$, мы можем написать их список:

$$(((xy)z)w),\ ((x(yz))w),\ ((xy)(zw)),\ (x((yz)w)),\ (x(y(zw))). \quad (34)$$

Совпадение всех этих выражений, исходя из закона трехчленной ассоциативности, устанавливается с помощью *двух* цепочек равенств:

$$(((xy)z)w) = ((x(yz))w) = (x((yz)w)) = (x(y(zw))), \quad (35)$$

$$(((xy)z)w) = ((xy)(zw)) = (x(y(zw))). \quad (35')$$

В (35) простейший закон ассоциативности применяется для того, чтобы переместить скобки мимо символов или их пар: первоначальная последовательность скобок ((( ))) при этом не меняется, как в формуле (31).

Однако в (35′) мы рассматриваем (xy) как *единый* объект E и применяем (31) к равенству $((Ez)w) = (E(zw))$, в результате чего) перемещается правее w и (правее E, включая все скобки, входящие в E. Обобщение этого приема требует более сложного обращения с индукцией, которое мы обсудим в § 2.10.

Запись. Обобщенный закон ассоциативности в алгебре используется обычно для упрощения записи — в ней опускаются все скобки:

$$\sum_{k=1}^{4} x_k = x_1 + x_2 + x_3 + x_4, \qquad \sum_{k=1}^{n} x_k = x_1 + \ldots + x_n \text{ и т. д.}$$

Чтобы записать сумму двух таких цепочек, мы просто выписываем их одну за другой и ставим между ними знак сложения.

Обобщенный закон коммутативности. Этот закон (для сложения) утверждает, что для любой биекции $\beta\colon \mathbf{n} \to \mathbf{n}$ конечного множества $\mathbf{n} = (1, \ldots, n)$ в себя

$$\sum_{k=1}^{n} a_k = \sum_{k=1}^{n} a_{\beta(k)} = \sum_{k=1}^{n} b_k, \qquad b_k = a_{\beta(k)}. \quad (36)$$

Разобравшись в нем, мы получим гораздо более ясное представление о доказательстве обоих обобщенных законов.

Рассмотрим, например, равенство

$$a_1 + a_2 + a_3 + a_4 + a_5 = a_3 + a_5 + a_2 + a_4 + a_1.$$

Чтобы доказать его, применяя двучленный закон коммутативности

$$x+y=y+x \quad \text{для всех } x, y,$$

мы должны последовательно переставить те пары смежных символов, индексы которых расположены в порядке убывания:

$$a_3+a_5+a_2+a_4+a_1=a_3+a_5+a_2+a_1+a_4=a_3+a_2+a_5+a_1+a_4=$$
$$=a_2+a_3+a_5+a_1+a_4=a_2+a_3+a_1+a_5+a_4=$$
$$=a_2+a_1+a_3+a_4+a_5=a_1+a_2+a_3+a_4+a_5$$

(пятое равенство получается в результате двух перестановок).

Заметим, что и здесь вывод обобщенного закона коммутативности из двучленного тождества $x+y=y+x$ (в предположении, что ассоциативность уже установлена, так что все скобки можно опустить), очевидно, требует более сложной техники доказательства по индукции, чем та, которой мы пользовались для доказательства теоремы 4. Главный момент состоит в доказательстве того, что *любая перестановка разлагается в произведение транспозиций*. Мы установим это в § 7.9.

*1.9. ПРИНЦИП ДИРИХЛЕ; АЛГОРИТМ ДЕЛЕНИЯ

В этом параграфе мы кратко объясним, как вывести из аксиом Пеано еще два важных математических принципа. Мы опускаем подробности, ибо результаты хорошо известны, однако все же рекомендуем внимательно прочитать определения и формулировки основных результатов, которые называются *принципом Дирихле* и *алгоритмом деления*. Ознакомимся с ними.

Неравенство. В § 1.1—1.6 мы часто рассматривали множества $\mathbf{n}=\{1, \ldots, n\}$, $k \leqslant n$. Покажем, как можно определить отношение $m \leqslant n$ в терминах функции следования, используя аксиомы Пеано.

Подмножество $S \subset \mathbf{P}$ называется *σ-замкнутым*, если из $n \in S$ следует, что $\sigma(n) \in S$. Интуитивно ясно, что любое σ-замкнутое подмножество множества $\mathbf{P}$ либо пусто, либо является «финальным сегментом» вида $[n, \infty] \subset \mathbf{P}$, состоящим из некоторого n и всех $k \geqslant n$. Формально мы используем это понятие для введения отношения неравенства $m \leqslant n$: оно означает, что любое σ-замкнутое подмножество множества $\mathbf{P}$, содержащее m, должно содержать также n.

Перечислим некоторые свойства этого отношения:

P1. $n \leqslant n$ для всех $n \in \mathbf{P}$.
P2. Из $m \leqslant n$ и $n \leqslant r$ следует, что $m \leqslant r$.
P3. Из $m \leqslant n$ и $n \leqslant m$ следует, что $m=n$.
P4. Для любых m, $n \in \mathbf{P}$ либо $m \leqslant n$, либо $n \leqslant m$.

Свойство P1 очевидно, потому что любое подмножество множества $\mathbf{P}$, содержащее n, содержит n. Аналогично устанавливается P2: если любое σ-замкнутое подмножество множества $\mathbf{P}$, содержащее m, обязательно содержит n, а любое σ-замкнутое подмножество множества $\mathbf{P}$, содержащее n, обязательно содержит r, то любое σ-замкнутое подмножество множества $\mathbf{P}$, содержащее m, содержит r. Доказательства свойств P3 и P4 труднее: они требуют индукции, и мы их опустим (см. книги Фермана и Глисона, указанные в литературе на стр. 38).

Чтобы понять значение свойства P4, напомним, что для любого $n \in \mathbf{P}$ начальный сегмент $\mathbf{n} = \{1, \ldots, n\}$ есть множество всех $k \leqslant n$ в их естественном порядке. Тогда неравенство $m \leqslant n$ равносильно тому, что $\mathbf{m} \subset \mathbf{n}$. Отсюда следует, что имеется инъекция из $\mathbf{m}$ в $\mathbf{n}$ и сюръекция из $\mathbf{n}$ в $\mathbf{m}$. Докажем теперь обращение этого очевидного результата.

Лемма. *Если существует инъекция* $h: \mathbf{n} \to \mathbf{m}$, *то* $n \leqslant m$.

Будем писать $h < k$, если неверно, что $h \geqslant k$, и $h > k$, если неверно, что $h \leqslant k$. Из P4 следует, что при $h \leqslant k$ либо $h < k$, либо $h = k$, а при $h \geqslant k$ либо $h > k$, либо $h = k$.

Доказательство леммы. Обозначим через $P(n)$ высказывание, что из существования инъекции $f: \mathbf{n} \to \mathbf{m}$ следует $n \leqslant m$. Тогда $P(1)$ очевидно, ибо ни для какого $m \in \mathbf{P}$ неверно, что $1 > m$. Будем считать теперь, что $P(n)$ истинно (предположение индукции). Обозначим через $h: \sigma\mathbf{n} \to \mathbf{m}$ некоторую инъекцию. Положим $h(\sigma n) = s$ и заметим, что в силу инъективности h существует не более одного $r \in \mathbf{n}$ со свойством $h(r) = m$. Рассмотрим теперь два случая $s \neq m$ и $s = m$ и покажем, что в обоих случаях $n \leqslant m - 1$. Если $s \neq m$, то отображение $h'(r) = s$, $h'(k) = f(k)$ при $k \neq r$, определяет инъекцию $\mathbf{n} \to \mathbf{m} - \mathbf{1}$ и $n \leqslant m - 1$ по предположению индукции. Если же $s = m$, то уже отображение $f'(k) = f(k)$ при $k \leqslant n$ определяет инъекцию $\mathbf{n} \to \mathbf{m} - \mathbf{1}$, так что снова $n \leqslant m - 1$. Отсюда вытекает, что $\sigma n \leqslant \sigma(m - 1) = m$, что доказывает лемму.

Выведем теперь из этой ключевой леммы три следствия.

Теорема 7. *Если* $m < n$, *то не существует ни инъекции* $f: \mathbf{n} \to \mathbf{m}$, *ни сюръекции* $g: \mathbf{m} \twoheadrightarrow \mathbf{n}$.

Доказательство. Предположим, что инъекция $f: \mathbf{n} \to \mathbf{m}$ существует. Из леммы следует тогда, что $n \leqslant m$. Это исключает возможность $m < n$ в силу свойств P1—P4. Далее, у любой сюръекции $g: \mathbf{m} \twoheadrightarrow \mathbf{n}$ есть правое обратное отображение $f: \mathbf{n} \to \mathbf{m}$ со свойством $g \circ f = 1_{\mathbf{m}}$. Это отображение f является инъекцией, ибо у него есть левое обратное g. Поэтому снова, согласно лемме, $n \leqslant m$.

Следствие. *Если существует биекция* g: $\mathbf{m} \leftrightarrow \mathbf{n}$, *то* $m = n$.

Теорема 8 (принцип Дирихле). *Пусть* X — *любое конечное множество. Отображение* f: $X \to X$ *взаимно однозначно тогда и только тогда, когда оно сюръективно.*

Следствие. *Пусть* X — *конечное множество. Тогда множества инъективных, сюръективных и биективных отображений* f: $X \to X$ *совпадают.*

На этих результатах основана высокая эффективность техники пересчета конечных множеств в прикладной математике. Для бесконечных множеств они неверны (см. гл. 14).

Третье следствие леммы относится не только к множеству $\mathbf{P}$ всех целых положительных чисел, но также к множеству $\mathbf{N}$ всех целых неотрицательных чисел.

Теорема 9 (алгоритм деления, или алгоритм Евклида). *Для любых* $m \in \mathbf{P}$, $n \in \mathbf{Z}$ *существуют единственные* $q \in \mathbf{Z}$ *и* $r \in \mathbf{N}$, *такие, что* $n = qm + r$, $0 \leqslant r \leqslant m - 1$.

Замечание. Число q называется *неполным частным* от деления n на m, а r — *остатком.*

Набросок доказательства. Определим по индукции функции $q = q(m, n)$ и $r = r(m, n)$ из области $\mathbf{P} \times \mathbf{Z}$ в кообласти $\mathbf{Z}$ и $\mathbf{N}$ соответственно:

$$r(m, 0) = 0, \quad r(m, \sigma n) = \sigma_m(r(m, n)). \tag{37}$$

Здесь σ_m — функция с областью и кообластью $\{0, 1, \ldots, m-1\}$, для которой $\sigma_m(m-1) = 0$, $\sigma_m(k) = \sigma k$, при $k \neq m-1$. Далее, $q(m, 0) = 0$ и

$$q(m, \sigma n) = \begin{cases} q(m, n), & \text{если } r(m, \sigma n) \neq m-1, \\ \sigma q(m, n), & \text{если } r(m, \sigma n) = m-1. \end{cases} \tag{38}$$

УПРАЖНЕНИЯ Д

1. Определив подходящим образом x^n, доказать по индукции следующие тождества в $\mathbf{P}$:

$$1^n = 1, \quad x^m x^n = x^{m+n}, \quad (xy)^n = x^n y^n, \quad (r^s)^n = r^{sn}.$$

2. Определим x^n рекурсией: $x^1 = x$ и $x^{\sigma n} = x^n x$. Доказать по индукции, что если $a^2 = a$, то $a^n = a$ для всех $n \in \mathbf{P}$.

3. а) Доказать по индукции, что $\sum_{k=1}^{n} k = n(n+1)/2$.

б) Доказать по индукции, что $\sum_{k=1}^{n} k^2 = n(n+1)(2n+1)/6$.

4. Доказать, что $\sum_{k=1}^{n} k^3 = [n(n+1)/2]^2$.

5. Полагая по определению $\binom{r}{s} = r!/s!(r-s)!$, доказать по индукции, что

$$\binom{r}{s} + \binom{r}{s-1} = \binom{r+1}{s}$$

для всех $r \in \mathbf{P}$ и $s = 0, \ldots, r$. Затем, используя этот факт, доказать по индукции формулу бинома Ньютона $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k$.

*6. На плоскости n прямых расположены так, что любые две из них пересекаются, но никакие три не пересекаются в одной точке. На сколько частей они разбивают плоскость?

7. Найти ошибку в следующем «доказательстве» предложения $P(n)$: во всяком множестве из n элементов все элементы одинаковы.

Шаг 1. $P(1)$ тривиально верно.

Шаг 2. Применив $P(n-1)$ к подмножеству $a_1, a_2, \ldots, a_{n-1}$ множества $a_1, \ldots, a_n$, заключаем, что $a_1 = a_2 = \ldots = a_{n-1}$.

Шаг 3. Применив $P(n-1)$ к подмножеству $a_2, \ldots, a_{n-1}, a_n$ множества $a_1, \ldots, a_n$, аналогично заключаем, что $a_2 = \ldots = a_n$.

Шаг 4. Соединяя результаты шагов 2 и 3 и пользуясь транзитивностью равенства, находим, что из $P(n-1)$ следует $P(n)$.

8. Доказать, что $\sum_{k=0}^{n} \binom{n}{k}^2 = \binom{2n}{n}$.

9. Доказать по индукции, что из $m+r = m+s$ в $\mathbf{P}$ следует, что $r = s$.

10. а) Доказать по индукции, что композиция инъекций $f_m \circ f_{m-1} \circ \ldots \circ f_1$ является инъекцией.

б) Доказать аналогичное утверждение для сюръекций.

11. Доказать, что в $\mathbf{P}$ справедливы следующие факты:

а) $m < n$ тогда и только тогда, когда $m + r = n$ для некоторого $r \in \mathbf{P}$.

б) $m \leqslant n$ тогда и только тогда, когда $m + r = n$ для некоторого $r \in \mathbf{N}$.

СПИСОК ЛИТЕРАТУРЫ

1. Bartee T. C., Lebow I. L., Reed I. S., Theory and Design of Digital Machines, McGraw-Hill, 1962.
2. Birkhoff G., MacLane S., A Survey of Modern Algebra, 3d ed., Macmillan, 1965.
3. Herstein L. N., Topics in Algebra, Ginn-Blaisdell, 1964.
4. Liu C. L., Introduction to Combinatorial Mathematics, McGraw-Hill, 1968.
5. MacLane S., Birkhoff G., Algebra, Macmillan, 1967.

В этих курсах можно найти дополнительный материал к многим главам нашей книги. Две следующие монографии посвящены тем вопросам, которые рассматривались в первой главе:

6. Feferman S., The number Systems, Addison-Wesley, 1964.
7. Gleason A. M., Fundamentals of Abstract Analysis, Addison-Wesley, 1966.

Аналогичные списки литературы читатель найдет в конце каждой главы.

2

БИНАРНЫЕ ОТНОШЕНИЯ И ГРАФЫ

2.1. ВВЕДЕНИЕ

Бинарным отношением α между множествами X и Y, а также *графиком* этого отношения, называется любое множество упорядоченных пар (x, y), где $x \in X$, $y \in Y$. Если $(x, y) \in \alpha$, мы говорим, что x находится в отношении α к y, и пишем также $x\alpha y$. Если (x, y) не находится в отношении α к y, то мы пишем $x\alpha' y$. Бинарное отношение может задаваться правилом, которое позволяет для каждой пары (x, y) решить, находится ли x в отношении α к y.

Понятие бинарного отношения между X, Y служит обобщением понятия функции $f\colon X \to Y$, изученного в гл. 1. Действительно, каждая функция $f\colon X \to Y$ определяет бинарное отношение α_f между X и Y:

$$x\alpha_f y \quad \text{означает, что} \quad y = f(x). \tag{1}$$

Обратно, пусть дано бинарное отношение α между множествами X и Y. Рассмотрим для каждого $x \in X$ множество всех $y \in Y$ со свойством $x\alpha y$. Это соответствие определяет функцию $f\colon X \to Y$ тогда и только тогда, когда для каждого $x \in X$ существует ровно один элемент $y \in Y$ со свойством $x\alpha y$. Таким образом, понятие бинарного отношения включает понятие функции как (очень важный) частный случай.

Следующий пример из аналитической геометрии показывает, каким образом многозначную функцию (так же, как и однозначную) можно рассматривать как бинарное отношение между ее областью и кообластью.

Пример 1. Пусть $X = Y = \mathbf{R}$ (множество вещественных чисел). Пусть $x\alpha y$ означает, что $x^2 + y^2 = 25$. Итак, $x\alpha y$ тогда и только тогда, когда $y = \pm\sqrt{25 - x^2}$: графиком α является окружность радиуса 5 с центром в начале координат. В этом примере $3\alpha 4$ и $4\alpha(-3)$, но, например, $2\alpha' 3$.

Слово «*график*» в применении к окружности, состоящей из точек, координаты которых (x, y) находятся в отношении $x^2 + y^2 = 25$ (или в функциональном обозначении $y = \pm\sqrt{25 - x^2}$),

как уже говорилось, употребляется и для произвольного бинарного отношения. Хотя бинарное отношение и его график являются эквивалентными понятиями, иногда, особенно при задании отношения с помощью правила, удобно иметь для графика особое обозначение.

Определение. *Графиком* бинарного отношения между множествами X и Y называется множество $S(\alpha)$ всех пар $(x, y) \in X \times Y$, таких, что $x \alpha y$.

Символически

$$S(\alpha) = \{(x, y) \mid x \alpha y\}.$$

Пример 2. Пусть $X = \{a, b\}$, $Y = \{c, d, e\}$. Зададим отношение α списком

$$a \alpha c,\ a \alpha d,\ a \alpha' e,\ b \alpha' c,\ b \alpha' d,\ b \alpha e.$$

Тогда график $S(\alpha)$ имеет вид

$$S(\alpha) = \{(a, c),\ (a, d),\ (b, e)\}.$$

Заметим, что отрицание α' отношения α также является бинарным отношением между X и Y. Переобозначив α' через β, мы получим, что β задается списком

$$a \beta' c,\ a \beta' d,\ a \beta e,\ b \beta c,\ b \beta d,\ b \beta' e.$$

Двойное отрицание α совпадает с α. Это верно и в общем случае; если

$$x \alpha' y \quad \text{означает} \quad \text{«не } x \alpha y\text{»}, \tag{2}$$

то $(\alpha')' = \alpha$.

Любое бинарное отношение ρ между конечными множествами $X = \{x_1, \ldots, x_m\}$ и $Y = \{y_1, \ldots, y_n\}$ можно задать таблицей, строки которой отвечают элементам X, столбцы — элементам Y, а на пересечении x_i-й строки и y_j-столбца записана 1, если $x_i \rho y_j$, и 0, если $x_i \rho' y_j$. Таблицы для отношений α и β примера 2 имеют вид

α	c	d	e
a	1	1	0
b	0	0	1

β	c	d	e
a	0	0	1
b	1	1	0

Табличная запись любого отношения ρ позволяет отождествить ρ (например, α или $\beta = \alpha'$) с характеристической функцией (гл. 1, формула (10)) графика этого отношения: значение этой функции на элементе (x_i, x_j) стоит на пересечении i-й строки и

j-го столбца таблицы:

$$e_{S(\rho)}(x_i, y_j) = \begin{cases} 1, & \text{если } x_i \rho y_j, \\ 0 & \text{в противном случае.} \end{cases} \tag{3}$$

В следующем параграфе мы рассмотрим уравнения (3) с другой точки зрения.

2.2. МАТРИЦЫ ОТНОШЕНИЙ

Пусть заданы пронумерованные конечные множества $X = \{x_1, \ldots, x_m\}$ и $Y = \{y_1, \ldots, y_n\}$. Тогда таблица из нулей и единиц, задающая любое отношение α, представляет собой *$m \times n$-матрицу* $A = \| a_{ij} \|$:

$$a_{ij} = \begin{cases} 1 & \text{если } x_i \alpha y_j, \\ 0, & \text{в противном случае.} \end{cases} \tag{4}$$

В примере 2 матрицы A, B, отвечающие отношениям α, β, имеют вид

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad B = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}.$$

Обратно, любая $m \times n$-матрица $A = \| a_{ij} \|$ из нулей и единиц определяет отношение $\rho(A)$ по формуле (4). Поэтому всякую (прямоугольную) матрицу из нулей и единиц мы будем называть *матрицей отношения*.

Согласно (3) и (4), на пересечении i-й строки и j-го столбца матрицы $M(\alpha) = A$ любого бинарного отношения α между множествами X и Y стоит элемент a_{ij}, значение характеристической функции $e = e_{S(\alpha)}$: $X \times Y \longrightarrow \{0, 1\}$ графика $S(\alpha) \subset X \times Y$ отношения α. Иными словами,

$$e(x_i, y_j) = a_{ij} = \begin{cases} 1, & \text{если } (x_i, y_j) \in S(\rho), \\ 0 & \text{в противном случае.} \end{cases} \tag{5}$$

В примере 2 характеристическая функция $e = e_{S(\alpha)}$ задается следующими предписаниями:

$$e_{S(\alpha)}: \begin{array}{lll} (a, c) \mapsto 1, & (a, d) \mapsto 1, & (a, e) \mapsto 0, \\ (b, c) \mapsto 0, & (b, d) \mapsto 0, & (b, e) \mapsto 1. \end{array}$$

Напомним еще раз, что каждое бинарное отношение α между X и Y определено его графиком $S(\alpha)$ и каждое подмножество $T \subset X \times Y$ является графиком единственного бинарного отношения ρ_T между X и Y, которое определяется условием $x \rho_T y$ тогда и только тогда, когда $(x, y) \in T$.

Поэтому соответствия $\alpha \mapsto S(\alpha)$ и $T \mapsto \rho_T$ определяют взаимно обратные биекции множества *всех бинарных отношений* между X и Y и множества *всех подмножеств* $\mathcal{P}(X \times Y)$ произведения $X \times Y$:

$$\rho_{S(\alpha)} = \alpha, \quad S(\rho_T) = T \text{ для всех } \alpha, T. \tag{6}$$

Кроме того, $S(\alpha') = [S(\alpha)]'$: отрицание отношения в этой биекции отвечает взятию дополнения к графику. По этой причине α' также называется *дополнением* к отношению α.

Булевы операции. Описанные биекции приводят к рассмотрению булевой алгебры, элементы которой являются бинарными отношениями между X и Y. Введем операции $\wedge$ и $\vee$ следующим образом:

$$x(\rho \wedge \sigma)y \text{ означает } x\rho y \text{ и } x\sigma y, \tag{7}$$

$$x(\rho \vee \sigma)y \text{ означает } x\rho y \text{ или } x\sigma y. \tag{7'}$$

Тогда, по определению операций $\cap$ и $\cup$ на множествах,

$$S(\rho \wedge \sigma) = S(\rho) \cap S(\sigma) \quad \text{и} \quad \rho_{T \cap V} = \rho_T \wedge \rho_V, \tag{8}$$

$$S(\rho \vee \sigma) = S(\rho) \cup S(\sigma) \quad \text{и} \quad \rho_{T \cup V} = \rho_T \vee \rho_V \tag{8'}$$

для любых двух отношений между X, Y и любых двух подмножеств T, $V \subset X \times Y$. В частности, для любых двух дополнительных отношений, таких, как α и $\alpha' = \beta$ в примере 2, $S(\alpha \wedge \alpha') = \varnothing$ (пустое множество) и $S(\alpha \vee \alpha') = X \times Y$. Мы будем называть $\alpha \wedge \beta$ пересечением бинарных отношений α, β, а $\alpha \vee \beta$ — их объединением. Наконец, будем писать $\alpha \leqslant \beta$, если $S(\alpha) \subset S(\beta)$. Иными словами,

$$\alpha \leqslant \beta \text{ означает, что из } x\alpha y \text{ следует } x\beta y. \tag{9}$$

Свойства булевых операций, перечисленные в § 1.2, выполняются и здесь.

Теорема 1. *Операции пересечения, объединения и отрицания бинарных отношений удовлетворяют законам идемпотентности, коммутативности, ассоциативности, поглощения (см. теорему 1 гл. 1), а также инволюции и законам де Моргана.*

Следствие. *Бинарные отношения между множествами X и Y образуют булеву алгебру.*

2.3. АЛГЕБРА ОТНОШЕНИЙ

Бинарные отношения между множествами, кроме общих свойств булевых алгебр, обладают многими алгебраическими свойствами. Например, по любому отношению α между множествами X и Y можно определить *обратное* отношение α между Y и X следую-

щим образом:

$$y\breve{\alpha}x \text{ означает, что } x\alpha y. \tag{10}$$

Очевидно, матрица $\| \breve{a}_{ij} \|$ отношения $\breve{\alpha}$ получается транспонированием матрицы $\| a_{ij} \|$ отношения α, т. е. заменой строк столбцами и наоборот (отражением относительно главной диагонали в случае $m=n$). Иными словами, $\breve{a}_{ij}=a_{ji}$.

Только в исключительном случае и отношение, и обратное к нему могут одновременно соответствовать функциям. Тогда, полагая $\alpha=\rho_f$ и $\breve{\alpha}=\rho_g$, имеем: каждый элемент $y\in Y$ (поскольку g—функция) должен отвечать некоторому элементу $x\in X$, и притом единственному (поскольку f—функция). Поэтому функции f и g, отвечающие α и $\breve{\alpha}$, должны быть взаимно обратными *биекциями*. Их матрицы отношений (если они существуют) должны быть квадратными *матрицами перестановок*, у которых в любой строке и в любом столбце имеется ровно одна единица.

Понятие композиции двух функций можно обобщить на отношения. Пусть α, β—отношения между X, Y и Y, Z соответственно. Тогда композицией $\alpha\beta$ называется отношение

$$x(\alpha\beta)z \text{ тогда и только тогда, когда существует такой } y\in Y, \text{ что } x\alpha y \text{ и } y\beta z. \tag{11}$$

Если $\alpha=\rho_f$ и $\beta=\rho_g$ отвечают функциям, то $\alpha\beta=\rho_{g\circ f}$, т. е. композиция отношений отвечает композиции функций.

Читая (11) справа налево, мы убеждаемся, что $z(\breve{\alpha}\breve{\beta})x$ означает существование такого $y\in Y$, что $z\breve{\beta}y$ и $y\breve{\alpha}x$. Отсюда следует тождество

$$\breve{\alpha}\breve{\beta}=\breve{\beta}\breve{\alpha}, \tag{12}$$

которое обобщает соотношение $(fg)^{-1}=g^{-1}f^{-1}$.

Бинарные отношения на S. Бинарное отношение между множеством S и им самим (т. е. $X=Y=S$) называется *отношением на множестве* S. Важным частным случаем этого понятия является отношение равенства e на S: xey означает $x=y$. Ясно, что на множестве $\mathbf{n}$ ему отвечает единичная матрица:

$$I=\| \delta_{ij} \|, \text{ где } \delta_{ij}=\begin{cases} 1, & \text{если } i=j, \\ 0, & \text{если } i\neq j. \end{cases}$$

Это квадратная $n\times n$-матрица с единицами на главной диагонали и нулями на остальных местах.

Определение (11) показывает, что композиция любых двух отношений на множестве S существует. Отношение равенства e удовлетворяет условиям $e\alpha=\alpha e=\alpha$ для всех α. Наконец, справедлив ассоциативный закон:

$$\alpha(\beta\gamma)=(\alpha\beta)\gamma \text{ для любых отношений } \alpha, \beta, \gamma \text{ на } S. \tag{13}$$

Действительно, оба утверждения $x[\alpha(\beta\gamma)]y$ и $x[(\alpha\beta)\gamma]y$ означают, что для подходящих элементов $z_1, z_2 \in S$ имеют место утверждения $x\alpha z_1$, $z_1\beta z_2$ и $z_2\gamma y$.

Резюмируем сказанное:

Теорема 2. *Бинарные отношения на множестве S относительно композиции образуют алгебраическую систему, которая ассоциативна и в качестве единицы имеет отношение равенства.*

В § 1.5, где было показано, что множество функций $f: S \to S$ обладает аналогичными свойствами, такие системы были названы моноидами (моноиды будут изучены в гл. 7). Таким образом, теорему 2 можно переформулировать следующим образом:

Следствие. *Бинарные отношения на любом множестве образуют моноид относительно композиции.*

Существует много важных типов бинарных отношений на S. Если $x\alpha x$ для всех $x \in S$, отношение α называется *рефлексивным*. Если $x\alpha x$ ни для какого x, отношение α называется *иррефлексивным*. Если из $x\alpha y$ следует, что $y\alpha x$, отношение α называется *симметричным*, а в противном случае — *асимметричным*. Если $x = y$, то из $x\alpha y$, очевидно, следует, что $y\alpha x$; обратно, если из $x\alpha y$ и $y\alpha x$ следует, что $y = x$, отношение α называется *антисимметричным*.

Мы видели, что отношение включения между множествами рефлексивно и антисимметрично. Поэтому таково же отношение включения между отношениями.

Бинарное отношение α на множестве S называется *транзитивным*, если из $x\alpha y$ и $y\alpha z$ ($x, y, z \in S$) вместе следует, что $x\alpha z$. В частности, отношение включения $\subset$ между множествами транзитивно. Отношение равенства e (т. е. $=$) рефлексивно, симметрично и транзитивно.

Пусть, например, $T = \{1, 2, 3\}$, и пусть α — отношение на T с графиком

$$S(\alpha) = \{(1,1) \quad (1,2) \quad (2,1) \quad (2,2) \quad (2,3) \quad (3,2) \quad (3,3)\}.$$

Это отношение рефлексивно и симметрично, но не транзитивно, ибо $1\alpha 2$ и $2\alpha 3$, но $1\alpha' 3$. Заметим, что график $S(\alpha^2)$ отношения α^2 совпадает с $T \times T$. Таким образом, α^2 — отношение, выполняющееся на всем T: $x\alpha^2 y$ для всех $x, y \in T$.

Данные выше определения можно переформулировать в терминах операций, введенных в § 2.2. Бинарное отношение α на S рефлексивно тогда и только тогда, когда оно содержит отношение равенства e ($e \leqslant \alpha$) или, что то же самое, когда $e \wedge \alpha = e$ и $e \vee \alpha = \alpha$. Оно симметрично в том и только том случае, когда $\alpha = \breve{\alpha}$. Оно антисимметрично тогда и только тогда, когда $\alpha \wedge \alpha' \leqslant e$.

Оно транзитивно тогда и только тогда, когда $\alpha^2 \leqslant \alpha$ (где α^2, конечно, есть $\alpha\alpha$). Доказательства мы опускаем.

Если $S = \mathbf{n}$, рефлексивность отношения на S означает, очевидно, что все диагональные элементы матрицы этого отношения равны единице. Отношение α симметрично в том и только том случае, когда его матрица $\|a_{ij}\|$ симметрична (т. е. $a_{ij} = a_{ji}$ для всех i, j).

Наконец, понятие декартова произведения функций (гл. 1, (22)) обобщается на отношения следующим образом. Пусть α, β — бинарное отношение между множествами A и Y, а β — между множествами B и Y. Определим *декартово* (или тензорное) *произведение* $\gamma = \alpha \times \beta$ как отношение между множествами $A \times B$ и Y вида

$$(a, b)\,\gamma y \quad \text{означает, что} \quad a\alpha y \text{ и } b\beta y. \tag{14}$$

Аналогично, пусть ξ — бинарное отношение между множествами A и X, η — бинарное отношение между множествами A и Y. Определим отношение $\zeta = \xi \times \eta$ следующим образом:

$$a\zeta\,(x, y) \quad \text{означает, что} \quad a\xi x \text{ и } a\eta y. \tag{14'}$$

УПРАЖНЕНИЯ А

1. Установить, какие из следующих отношений на **P** рефлексивны, иррефлексивны, симметричны, антисимметричны, транзитивны:

   а) $m + n$ четно;
   б) $m + n \leqslant 100$;
   в) $m + n$ нечетно;
   г) m/n является степенью двойки;
   д) m/n четно;
   е) mn нечетно;

2. Какое из следующих отношений на **Z** симметрично:

   а) $m\rho n$ означает, что $m + n$ делится на три;
   б) $m\sigma n$ означает, что $m - n$ делится на три?

Какое из них транзитивно?

3. Пусть ρ, σ — рефлексивные симметричные отношения на множестве S. Показать, что следующие условия равносильны: а) $\rho\sigma$ симметрично; б) $\rho\sigma = \sigma\rho$; в) $\rho\sigma = \rho \vee \sigma$.

4. Построить два симметричных отношения на множестве $\{1, 2, 3\}$, композиция которых несимметрична.

5. Вычислить матрицы отношения $\leqslant$ и функции $m \mapsto -m$ на множествах $\{-1, 0, 1\}$, $\{-2, -1, 0, 1, 2\}$.

6. Доказать со всеми подробностями, что если отношение ρ симметрично, то отношение $\rho \vee \rho^2 \vee \ldots \vee \rho^n$ также симметрично.

7. Доказать, что если ρ, σ, τ — бинарные отношения на X и если $\rho \leqslant \sigma$, то $\rho\tau \leqslant \sigma\tau$ и $\tau\rho \leqslant \tau\sigma$ (монотонность композиции).

8. Показать, что отношение r на множестве U является биекцией тогда и только тогда, когда $r\breve{r} = \breve{r}r = 1_U$ (отношение равенства).

9. а) Доказать, что имеется 256 тернарных операций f на **2**, т. е. функций из области $\mathbf{2}^3$ в кообласть **2**.

*б) Для скольких из таких f существуют функции $g\colon \mathbf{2}^2 \longrightarrow \mathbf{2}$ и $h\colon \mathbf{2}^2 \longrightarrow \mathbf{2}$, такие, что $f(x, y, z) = g(x, h(y, z))$?

2.4. ЧАСТИЧНОЕ УПОРЯДОЧЕНИЕ

Бинарное отношение на множестве S называется *частичным упорядочением* этого множества (или частичным порядком на нем), если оно рефлексивно, антисимметрично и транзитивно. Такие отношения часто обозначаются символом $\leqslant$. Аксиомы частичного порядка могут быть записаны тогда привычным способом:

P1. $x \leqslant x$ для всех $x \in S$.

P2. Если $x \leqslant y$ и $y \leqslant x$, то $x = y$.

P3. Если $x \leqslant y$ и $y \leqslant z$, то $x \leqslant z$.

Пример 3. Обычное отношение $\leqslant$ является частичным упорядочением множества всех положительных целых чисел.

Пример 4. Отношение $m \mid n$ (m делит n) является другим частичным упорядочением на множестве всех положительных целых чисел.

Пример 5. Для любого множества U отношение $S \subset T$ является частичным порядком на множестве $\mathcal{P}(U)$ всех подмножеств множества U.

Определение. Частично упорядоченным множеством называется любая пара $[S, \leqslant]$, где $\leqslant$ — частичный порядок на множестве S.

Имеется много полезных примеров частичных упорядочений. Рассмотрим некоторые из простейших свойств частично упорядоченных множеств (по поводу других см. упражнения Б). Во-первых, отношение, обратное к частичному порядку $\leqslant$, снова является частичным порядком, который называется *двойственным* к первому и обозначается символом $\geqslant$. Таким образом, по определению, $X \geqslant Y$ тогда и только тогда, когда $Y \leqslant X$. Во-вторых, частично упорядоченные множества, состоящие из небольшого числа элементов, удобно описывать *диаграммами*. Маленькие кружки на них означают элементы; линия, ведущая вверх, соединяет элемент с каждым непосредственно следующим за ним бóльшим элементом.

Так, на рис. 2.1,*а* изображено множество $\mathcal{P}(\mathbf{3})$ всех подмножеств множества $\mathbf{3} = \{1, 2, 3\}$, частично упорядоченное относительно отношения включения. Заметим, что перевернув диаграмму, получим диаграмму двойственного порядка.

На рис. 2.1,*б*,*в* изображены диаграммы частично упорядоченного множества $P=[\{2, 3, 5, 7, 14, 15, 21\}, |]$.

Принцип двойственности. Мы уже упоминали, что обращение частичного порядка является частичным порядком. Поэтому в любой общей теореме о частично упорядоченных множествах можно всюду заменить отношение $\leqslant$ отношением $\geqslant$, не

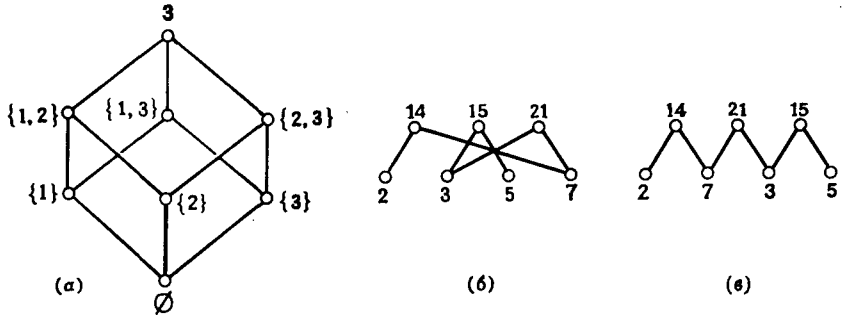


Рис. 2.1. Диаграмма частично упорядоченных множеств.

нарушив ее истинности. Эта математическая «теорема о теоремах» называется *принципом двойственности* в теории частично упорядоченных множеств.

В частично упорядоченном множестве $[\mathcal{P}(\mathbf{3}), \subset]$, изображенном на рис. 2.1, *а*, элементы $\varnothing$ и $\mathbf{3}$ являются *универсальными границами* в том смысле, что $\varnothing \leqslant x \leqslant \mathbf{3}$ для любого элемента $x \in \mathcal{P}(\mathbf{3})$. Это понятие можно определить в общем случае: элементы O и I называются универсальными границами частично упорядоченного множества S (соответственно верхней и нижней), если

$$O \leqslant x \quad \text{и} \quad x \leqslant I \quad \text{для любого } x \in S. \tag{15}$$

Иными словами, O — *наименьший*, а I — *наибольший* элемент множества S.

Лемма. *В любом частично упорядоченном множестве* $[S, \leqslant]$ *может существовать не более одного наименьшего элемента и не более одного наибольшего элемента.*

Доказательство. Пусть O, O^* — два наименьших элемента $[S, \leqslant]$. Тогда $O \leqslant O^*$ (ибо O — наименьший элемент) и $O^* \leqslant O$ (ибо O^* — наименьший элемент). Согласно P2, отсюда следует, что $O = O^*$. Доказательство для I аналогично.

Существуют частично упорядоченные множества без универсальных границ. Таково множество вещественных чисел $[\mathbf{R}, \leqslant]$ с обычным отношением порядка (если оно не расширено формальным присоединением $-\infty$, ∞).

Кроме того, $[\mathbf{R}, \leqslant]$ является *линейно упорядоченным множеством* или *цепью*: кроме свойств P1—P3, оно обладает свойством P4. Для любых x, y либо $x \leqslant y$, либо $y \leqslant x$.

Всякое подмножество упорядоченного множества, очевидно, само упорядочено индуцированным на нем бинарным отношением. Если множество линейно упорядочено, то его подмножества также линейно упорядочены. Все это следует из того, что свойства P1, P2, P3, P4 «наследственны», т. е. сохраняются при ограничении на любое подмножество своей области, если они выполнялись на всей области.

Доминирование. С любым отношением частичного порядка $\leqslant$ связан ряд других бинарных отношений. К ним относятся: отношение $x < y$, означающее, что $x \leqslant y$, но $x \neq y$; отношение $x > y$, означающее, что $y < x$, а также отношение $x \asymp y$ («x и y не сравнимы»), означающее, что ни одно из двух утверждений $x \leqslant y$ и $y \leqslant x$ неверно. Ясно, что если $x, y \in [S, \leqslant]$, то справедлива только одна из следующих альтернатив: $x = y$, $x > y$, $x < y$ или $x \asymp y$. Менее очевидно отношение, связанное с $\leqslant$, а именно доминирование.

Определение. Пусть $P = [S, \leqslant]$—частично упорядоченное множество, a и b—его элементы. Будем говорить, что a *доминирует* над b, если $a > b$, но ни для какого $x \in S$ неверно, что $a > x > b$.

Именно отношение доминирования непосредственно отражено на диаграммах типа рис. 2.1: каждый отрезок соединяет свой нижний конец—элемент a—с верхним концом, элементом b, который доминирует над a. Покажем, что отношение $\leqslant$ однозначно восстанавливается по отношению доминирования в любом конечном частично упорядоченном множестве.

Теорема 3. *Пусть $a < b$ в конечном частично упорядоченном множестве P. Тогда P содержит по крайней мере одну цепь $x_0 = a < x_1 < \ldots < x_l = b$, в которой каждый из элементов x_i $(i = 1, \ldots, l)$ доминирует над x_{i-1}.*

Для доказательства проведем индукцию по количеству n элементов y со свойством $a < y < b$. Если $n = 0$, то b доминирует над a, и утверждение очевидно. Пусть $n > 0$, и пусть $a < c < b$. Тогда количество элементов y, удовлетворяющих условию $a < y < c$, и элементов z, удовлетворяющих условию $c < z < b$, не превосходит $n-1$, ибо c мы исключили. Значит, по предположению индукции существуют конечные цепи, связывающие a с c и c с b, соседние элементы которых находятся в отношении доминирования. Соединяя эти две цепи, получим требуемый результат.

Удобно представлять себе, что если a доминирует над b, то b находится в прямом подчинении к a относительно некоторой

иерархии типа той, которая изображена на рис. 2.1, *б*. Тогда $x \leqslant y$ означает отношение подчинения.

Определение. Элемент m частично упорядоченного множества $[S, \leqslant]$ называется *минимальным*, если не существует такого элемента $x \in S$, что $x < m$. Элемент m называется *максимальным*, если не существует такого элемента $x \in S$, что $x > m$.

Напомним, что $x < m$ означает $x \leqslant m$, но $x \neq m$; $x > m$ означает, что $m \leqslant x$, но $m \neq x$.

Очевидно, что если частично упорядоченное множество обладает универсальной нижней границей, или наименьшим элементом O, то O является единственным минимальным элементом. В неупорядоченном множестве (где $x \leqslant y$ означает $x = y$) любой элемент минимален, а также максимален. В цепи минимальный элемент должен быть наименьшим, и поэтому он единствен.

Следующий принцип, относящийся к конечным частично упорядоченным множествам, играет важную роль. Он утверждает существование согласованной (с порядком) нумерации.

Теорема 4. *Пусть* $[S, \leqslant]$, $S = \{s_1, \ldots, s_n\}$ — *конечное частично упорядоченное множество. Тогда элементы* S *можно занумеровать таким образом:* $S = \{x_1, \ldots, x_n\}$, *что из* $x_i < x_j$ *будет следовать* $i < j$.

Доказательство. Положим $X_m = \{s_1, \ldots, s_m\}$, где первоначальная нумерация $S = \{s_1, \ldots, s_n\}$ выбрана каким угодно способом. Мы построим последовательность биекций β_m, множеств $\mathbf{m} = \{1, \ldots, m\}$ в себя, такую, что каждое подмножество X_m, перенумерованное посредством β_m:

$$X_m = \{x_1^m, \ldots, x_m^m\}, \quad x_i^m = s_{\beta_m(i)},$$

будет удовлетворять сформулированному в теореме утверждению: из $x_i^m < x_j^m$ следует, что $i < j$.

При $m = 1$ биекция β_1 строится однозначно. Предположим, что биекция β_{n-1}: $\mathbf{n-1} \to \mathbf{n-1}$ с требуемым свойством уже построена. Обозначим через k наименьшее из чисел i, обладающих свойством $s_n < x_i^{n-1}$. Построим биекцию β_n: $\mathbf{n} \to \mathbf{n}$ следующим образом:

$$\beta_n(i) = \begin{cases} i, & \text{если } i < k, \\ n, & \text{если } i = k, \\ i+1, & \text{если } i > k. \end{cases}$$

Иными словами, вставим s_n между x_{k-1}^{n-1} и x_k^{n-1}. Проверим, что β_n обладает требуемыми свойствами. Если $x_i^n < x_j^n$ и $\{x_i^n, x_j^n\} \subset X_{n-1}$, то $i < j$ по предположению индукции. Если $s_n = x_k^n < x_j^n$, то $k < j$ по построению. Наконец, если $x_i^n < x_k^n = s_n$, то $x_i^n < x_k^n < x_{k+1}^n$, откуда $x_i^n < x_{k+1}^n$ в силу транзитивности РЗ и, наконец, $i < k+1$ по

предположению индукции, так как $\{x_i^n, x_{k+1}^n\} \subset X_{n-1}$. Следовательно, и здесь $i < k$ (случай $i = k$ невозможен); доказательство завершено.

Следствие. *В любом конечном частично упорядоченном множестве P есть минимальный элемент m: не существует $x \in P$, такого, что $x < m$.*

Верхняя и нижняя грани. Пусть S — некоторое подмножество частично упорядоченного множества P. Назовем $a \in P$ *нижней границей*, или *минорантой*, множества S, если $a \leqslant x$ для всех $x \in S$. Назовем a *верхней границей*, или *мажорантой*, множества S, если $a \geqslant x$ для всех $x \in S$. Назовем элемент $b \in P$ *нижней гранью* S, если (i) он является нижней границей для S и (ii) $b \geqslant \bar{b}$ для любой другой нижней границы $\bar{b}$ множества S. В этом случае мы будем писать $b = \inf S$. Аналогично, назовем $c \in P$ *верхней гранью* множества S, если (i′) c является верхней границей для S и (ii′) $c \leqslant \bar{c}$ для любой другой верхней границы $\bar{c}$. В этом случае мы будем писать $c = \sup S$.

Лемма. *Любое подмножество частично упорядоченного множества имеет не более одной верхней и не больше одной нижней грани.*

Доказательство. Пусть b_1, b_2 — нижние грани множества S. Тогда $b_1 \leqslant b_2$, потому что b_1 — нижняя граница, а b_2 — *наибольшая* нижняя граница. Аналогично, $b_2 \leqslant b_1$. Из свойства P2 следует, что $b_1 = b_2$. Двойственное рассуждение доказывает единственность верхней грани.

Дальнейшее изучение частично упорядоченных множеств будет проведено в § 5.2 и гл. 9.

УПРАЖНЕНИЯ Б

1. Показать, что отношение $i \leqslant j$ на множестве $\mathbf{n} = \{1, 2, \ldots, n\}$ задается треугольной матрицей.

2. а) Показать, что в каждом конечном частично упорядоченном множестве есть цепь наибольшей длины.

б) Показать, что каждое конечное частично упорядоченное множество обладает максимальным элементом.

3. Показать, что в конечном множестве имеется наименьший элемент тогда и только тогда, когда у него есть ровно один минимальный элемент.

4. а) Найти контрпримеры к утверждениям упр. 2 для бесконечных частично упорядоченных множеств.

б) Доказать, что условие «тогда» в упр. 3 перестает быть верным для бесконечных частично упорядоченных множеств.

5. Доказать, что в конечном частично упорядоченном множестве условие $a \leqslant b$ равносильно тому, что либо $a = b$, либо существует конечная цепь $a = x_0 < x_1 < \ldots < x_l = b$, такая, что x_i доминирует над x_{i-1} для всех $i = 1, \ldots, l$.

6. Пусть $<$ — любое иррефлексивное и транзитивное отношение на множестве P. Пусть $x \leqslant y$ означает, что либо $x = y$, либо $x < y$. Доказать, что $\leqslant$ есть частичное упорядочение.

7. Пусть $P = [S, \leqslant]$ и $Q = [T, \leqslant]$ — два частично упорядоченных множества. Обозначим через $[S \times T, \leqslant] = P \times Q$ множество, для которого $(s, t) \leqslant (s', t')$ означает, что $s \leqslant s'$ в P и $t \leqslant t'$ в Q.

а) Показать, что $P \times Q$ — частично упорядоченное множество.

б) Показать, что $P \times Q$ может быть цепью только в том случае, если P или Q состоит из одного элемента.

8. Определим в $\mathbf{N} \times \mathbf{N}$ отношение α условием: $(mn)\,\alpha\,(m_1 n_1)$ означает, что $m \leqslant m_1$ и $n \leqslant n_1$ в $\mathbf{N}$.

а) Показать, что α является частичным упорядочением.

б) Показать, что в частично упорядоченном множестве $[\mathbf{N} \times \mathbf{N}, \alpha]$ любое непустое подмножество обладает минимальным элементом.

* 9. Показать, что $\mathcal{P}(A \cup B) \cong \mathcal{P}(A) \times \mathcal{P}(B)$ для любых двух множеств A, B.

*10. Пусть P и Q — два частично упорядоченных множества, как в упр. 7. Определим лексикографическое произведение $P \otimes Q = [S \times T, \leqslant]$, условившись, что $(s, t) \leqslant (s', t')$, если либо $s < s'$, либо $s = s'$ в P и $t \leqslant t'$ в Q.

а) Доказать, что $P \otimes Q$ — частично упорядоченное множество.

б) Доказать, что если P, Q — конечные цепи, то и $P \otimes Q$ — конечная цепь.

* 11. Пусть $P(m, n)$ — множество высказываний, пронумерованных парами $(m, n) \in \mathbf{N} \times \mathbf{N}$. Предположим, что $P(0, 0)$ истинно и что из истинности $P(m, n)$ следует истинность $P(m+1, n)$ и $P(m, n+1)$. Доказать, что тогда все высказывания $P(m, n)$ истинны.

2.5. РАЗБИЕНИЯ И ОТНОШЕНИЯ ЭКВИВАЛЕНТНОСТИ

Отношение E на множестве S называется *отношением эквивалентности* на S, если оно рефлексивно, симметрично и транзитивно. Очевидно, отношение равенства e является наименьшим отношением эквивалентности на S.

Разбиением Π множества S называется множество подмножеств S_k в S, обладающих следующими свойствами: (i): $S_i \cap S_j = \varnothing$ при $i \neq j$ (S_k попарно не пересекаются); (ii) $\bigcup S_k = S$ (все вместе S_k исчерпывают S). Иными словами, каждый элемент $x \in S$ принадлежит только одному из подмножеств S_k, так что Π разбивает S на различные «части».

По разбиению Π множества S можно построить бинарное отношение $E(\Pi)$: $xE(\Pi)y$ означает, что x, y лежат в одном и том же элементе $S_k \in \Pi$. Ясно, что $E(\Pi)$ есть отношение эквивалентности.

Обратно, по данному отношению эквивалентности E на S построим функцию p_E: $S \longrightarrow \mathcal{P}(S)$, которая ставит в соответствие элементу $x \in S$ множество

$$p_E(x) = \{z \in S \mid zEx\} \subset S. \tag{16}$$

Из рефлексивности E следует, что $x \in p_E(x)$. Поэтому $\bigcup p_E(x) = S$. Предположим теперь, что два подмножества $p_E(x)$ и $p_E(y)$ («классы эквивалентности относительно E») имеют непустое пересечение. Пусть $z \in S$ лежит в обоих классах. Это означает, что xEz и yEz; по симметричности, zEy; из транзитивности E тогда следует, что xEy. Поэтому из yEt по транзитивности находим xEt. Иными словами, $p_E(y) \subset p_E(x)$. Меняя местами x и y, находим подобным же образом, что $p_E(x) \subset p_E(y)$. Следовательно, $p_E(x) = p_E(y)$.

Отсюда вытекает, что разные $p_E(x)$ не пересекаются. Кроме того, так как $x \in p_E(x)$, мы получили следующий результат.

Лемма. *Каждое отношение эквивалентности на S определяет по формуле* (16) *разбиение Π_E множества S на попарно не пересекающиеся подмножества $p_E(x)$.*

Функция $p_E\colon S \to \mathcal{P}(S)$, определенная формулой (16), называется *проектором S на фактормножество S/E.*

Теорема 5. *Существует естественная биекция $b\colon E \to \Pi_E$ между множеством всех отношений эквивалентности на множестве S и множеством разбиений Π; обратная к ней биекция имеет вид $b^{-1}\colon \Pi \to E(\Pi)$.*

Доказательство. Согласно лемме 1, отображение b определено корректно; по замечанию в третьем абзаце этого параграфа отображение b^{-1} также определено корректно. Остается проверить, что эти два отображения взаимно обратны (ср. теорему 3 гл. 1), т. е. что $\Pi_{E(\Pi)} = \Pi$ и $E(\Pi_E) = E$. Но первое очевидно из доказательства леммы, а второе — непосредственно.

Следствие. *Для любого отношения эквивалентности E на S функция $p_E\colon S \to S/E$ отображает S на фактормножество $S/E = \mathcal{P}(E)$, состоящее из классов эквивалентности отношения E.*

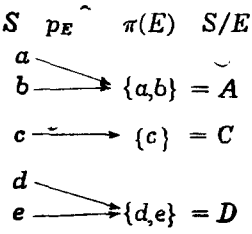
Пример 6. Рассмотрим отношение E_m на $\mathbf{Z}$: xE_my означает $m \mid (x-y)$. Его принято записывать в виде $x \equiv y \bmod m$ и читать «x сравним с y по модулю m». Так как $x - x = m0$, то оно рефлексивно; так как из $x - y = mk$ следует $y - x = m(-k)$, то оно симметрично; наконец, так как из $x - y = mk$ и $y - z = mk'$ следует $x - z = m(k + k')$, то оно транзитивно. Следовательно, оно является отношением эквивалентности. Его классы эквивалентности — это m арифметических прогрессий $\ldots, k-m, k, k+m, k+2m, \ldots$ при $k = 0, 1, \ldots, m-1$.

Пример 7. Пусть $S = \{a, b, c, d, e\}$, и пусть E — отношение эквивалентности с матрицей (*а*). Функция $p_E\colon S \to \mathcal{P}(S)$ описана на диаграмме (*б*), множество S/E обозначено через $\{A, C, D\}$.

Матрица E:

	a	b	c	d	e
a	1	1	0	0	0
b	1	1	0	0	0
c	0	0	1	0	0
d	0	0	0	1	1
e	0	0	0	1	1

(а)



(б)

Факторизация функций. Любую функцию $f\colon S \longrightarrow T$ можно естественным образом разложить в композицию трех отображений. Первое из них является проекцией $p\colon S \longrightarrow S/E_f$ на фактормножество S по отношению эквивалентности E_f:

$$xE_f y \text{ означает, что } f(x) = f(y).$$

Второе отображение является биекцией

$$b\colon S/E_f \leftrightarrow f(S), \quad b(\text{класс } x) = f(x),$$

множества S/E_f с образом $f(S)$, который определяется как множество всех $f(x)$ с $x \in S$. Наконец, третье определяется как естественное вложение $i\colon f(S) \longrightarrow T$. Очевидно, имеет место равенство

$$f = p \diamond b \diamond i, \tag{17}$$

которое называется *каноническим разложением* f. (Заметим, что в соответствии с нашими соглашениями i не обязано совпадать ни с 1_S, ни с 1_T, кроме того случая, когда $S = T$.)

Вообще, если $X \subset S$, *образом* $f(X)$ называется множество всех $f(x)$ с $x \in X$. Если $Y \subset T$ и $f(X) \subset Y$, то функция $g\colon X \longrightarrow Y$, для которой $g(x) = f(x)$ при всех $x \in X$, называется *ограничением* f на X и Y.

УПРАЖНЕНИЯ В

1. Доказать, что если отношение ρ на некотором множестве S рефлексивно и транзитивно, то $\rho \wedge \breve{\rho}$ есть отношение эквивалентности на S.

2. Пусть в упр. А3 ρ и σ — отношения эквивалентности. Доказать, что каждое из условий (б), (в) необходимо и достаточно для того, чтобы $\rho\sigma$ было отношением эквивалентности.

3. Два множества («фигуры») F и G на евклидовой плоскости называются *изометричными* (символически $F \cong G$), если существует биекция $\mu\colon F \leftrightarrow G$, сохраняющая расстояние между любыми двумя точками. Показать, что изометричность есть отношение эквивалентности.

4. Доказать, что у множества, состоящего из n элементов, есть ровно $2^{n-1}-1$ разбиений на два класса эквивалентности.

5. а) Доказать, что отношение α на множестве S является одновременно симметричным и антисимметричным в том и только том случае, когда его матрица $\|a_{ij}\|$ «диагональна», т. е. $a_{ij}=0$ при $i \neq j$.

б) Доказать, что это условие равносильно условию $\alpha \leqslant e$.

*6. Пусть $\pi(n, k)$ — число разбиений множества, состоящего из n элементов, на k непустых подмножеств. Доказать, что

$$\sum_{k=1}^{l} k!\,\pi(n, k) = l^n.$$

2.6. КЛАССЫ ВЫЧЕТОВ И МОРФИЗМЫ

Для каждого целого числа $m > 1$ отношение эквивалентности E_m: $x \equiv y \bmod m$, определенное в примере 6, позволяет построить важную алгебраическую систему классов эквивалентности $\mathbf{Z}/E_m$. Она называется *системой классов вычетов по модулю m* и обозначается $\mathbf{Z}_m$. Опишем ее свойства.

Прежде всего, по определению, класс вычетов числа $n \in \mathbf{Z}$ по модулю m есть множество всех чисел вида $n+km$, где $k \in \mathbf{Z}$. В этом классе имеется единственное наименьшее неотрицательное число $r_m(n)$. При $n \in \mathbf{N}$ оно совпадает с числом $r(m, n)$ из формулы (38) гл. 1. Отсюда видно, что числа $0, 1, \ldots, m-1$ составляют *полную систему представителей классов эквивалентности* $\{n+km\}$, по одному из каждого класса.

Обозначим через $\mathbf{m}=\{0, 1, \ldots, m-1\}$ эту систему представителей и через $\mathbf{Z}_m=[m, \overset{m}{+}, \overset{m}{\times}]$ соответствующую алгебраическую систему с двумя бинарными операциями, сложением и умножением. Они определены следующими формулами:

$$a \overset{m}{+} b = r_m(a+b), \qquad a \overset{m}{\times} b = r_m(a \times b), \tag{18}$$

где $+$, $\times$ — сложение и умножение в $\mathbf{Z}$.

Лемма. Для любых $n, n' \in \mathbf{Z}$

$$r_m(n+n') = r_m(n) \overset{m}{+} r_m(n'), \quad r_m(n \times n') = r_m(n) \overset{m}{\times} r_m(n'). \tag{19}$$

Доказательство. Ясно, что обе части каждого из соотношений (19) лежат в $\mathbf{m}=\{0, 1, \ldots, m-1\}$. С другой стороны, они отличаются от $n+n'$ (соответственно $n \times n'$) на целые кратные m (см. пример 6). Поэтому они должны совпадать.

Символы $\overset{m}{+}$ и $\overset{m}{\times}$ мы использовали для того, чтобы избежать путаницы с соответствующими операциями в $\mathbf{Z}$ (подмножеством которого является $\mathbf{m}$). Однако теперь мы вернемся к стандарт-

ному обозначению $[Z_m, +, \cdot]$ вместо $\left[m, \overset{m}{+}, \overset{m}{\times}\right]$. В этой записи (19) приобретает следующий вид:

$$r_m(n+n') = r_m(n) + r_m(n'), \qquad r_m(nn') = r_m(n)\, r_m(n').$$

Это означает, что r_m является «морфизмом» относительно операций сложения и умножения в следующем общем смысле.

Определение. Пусть $[A, +, \cdot]$, $[B, +, \cdot]$—две алгебраические системы, каждая с бинарными операциями сложения и умножения. Функция θ: $A \to B$ называется *морфизмом* (относительно этих операций), если для любых a, $a' \in A$

$$\theta(a+a') = \theta(a) + \theta(a'), \qquad \theta(aa') = \theta(a)\,\theta(a'). \tag{20}$$

Уравнения (20) равносильны требованию коммутативности диаграммы для сложения, которая выглядит так:

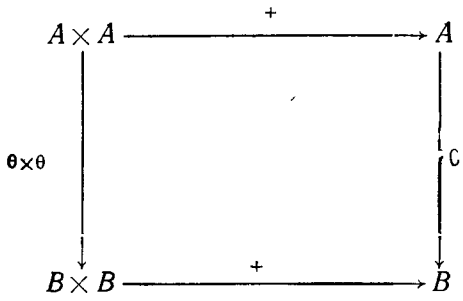


Рис. 2.2.

и аналогичной диаграммы для умножения.

Вообще морфизм есть отображение, сохраняющее одну или несколько операций. Он называется *эпиморфизмом*, если он сюръективен, *мономорфизмом*, если он инъективен, *изоморфизмом*, если он биективен.

В этих терминах мы можем переформулировать нашу лемму так:

Теорема 6. *Отображение, переводящее каждое целое число в его остаток при делении на целое число* $m > 1$, *является эпиморфизмом (относительно операций сложения и умножения)* $\mathbf{Z}$ *на* $\mathbf{Z}_m$.

Таблицы сложения и умножения в $\mathbf{Z}_4$ имеют следующий вид:

$+$	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

$\cdot$	0	1	2	3
0	0	0	0	0
1	0	1	2	3
2	0	2	0	2
3	0	3	2	1

Рис. 2.3.

Свойство подстановки. В § 2.5 мы показали, что каждая функция φ: $S \longrightarrow T$ определяет отношение эквивалентности E_φ на S и естественную биекцию в b: $S/E_\varphi \leftrightarrow \varphi(S)$. Обратно, каждое отношение эквивалентности E на S определяет сюръекцию φ_E: $S \longrightarrow S/E$, причем $E\varphi_E = E$. Зададимся вопросом: какое условие следует наложить на E_φ (соответственно E), чтобы φ (соответственно φ_E) была морфизмом? Покажем, что это условие выражается в виде несложного свойства подстановки.

Рассмотрим сначала пример 6. В этом примере отношение E_m имеет вид

$$xE_m y \quad \text{означает, что} \quad m \mid (y - x). \tag{21}$$

Для него выполнено следующее свойство подстановки:

$$\text{если} \quad xE_m y, \quad \text{то} \quad \varphi(x) E_m \varphi(y). \tag{21'}$$

Кроме того, для него выполнены два свойства подстановки относительно операций $+$ и $\cdot$ из $\mathbf{Z}\times\mathbf{Z}$ в $\mathbf{Z}$:

$$\left.\begin{array}{c}\text{если} \quad xE_m y \quad \text{и} \quad x'E_m y', \quad \text{то} \\ (x+x')\,E_m\,(y+y') \quad \text{и} \quad (xx')\,E_m\,(yy').\end{array}\right\} \tag{22}$$

Эти свойства являются частными случаями следующего общего определения.

Определение. Пусть E — отношение эквивалентности на S, и пусть f — n-арная операция на S. Отношение E обладает *свойством подстановки относительно* f, если для любых x_i, $y_j \in S$

$$\left.\begin{array}{c}\text{из} \quad x_1 E y_1, \; x_2 E y_2, \; \ldots, \; x_n E y_n \quad \text{следует,} \\ \text{что} \quad f(x_1, \ldots, x_n)\, E f(y_1, \ldots, y_n).\end{array}\right\} \tag{23}$$

Если f обладает свойством подстановки относительно E, то на фактормножестве S/E естественно вводится операция $\bar{f}$ фор-

мулой

$$f(p_E(x_1), \ldots, p_E(x_n)) = p_E(f(x_1, \ldots, x_n)),$$

(Читателю следует проверить, что свойство подстановки гарантирует корректность определения.) Поэтому если E обладает свойством подстановки относительно f, то проектор $x \mapsto p_E(x)$ определяет морфизм из $[S, f]$ в $[S/E, \bar{f}]$.

Кардинальные числа. Одно из самых фундаментальных приложений понятий отношения эквивалентности и свойства подстановки на классе всех множеств Γ относится к следующим понятиям.

Определим на Γ отношение E:

SET означает, что существует биекция b: $S \leftrightarrow T$.

В гл. 14 мы покажем, что E есть отношение эквивалентности, обладающее свойством подстановки относительно бинарных операций разделенной суммы, декартова произведения и унарной операции «множество-степень». Отсюда все основные факты арифметики конечных и бесконечных кардинальных чисел (т. е. классов эквивалентности относительно E) выводятся в несколько строк.

Мы вернемся к этим идеям в последующих главах. Они являются основными в современной алгебре.

Морфизмы отношений. Понятие морфизма, определенное выше для систем с операциями, имеет аналог для отношений. Пусть, например, P и Q — два частично упорядоченных множества. Функцию f: $P \longrightarrow Q$ иногда называют *морфизмом, сохраняющим порядок*, если из $x \leqslant y$ в P следует, что $f(x) \leqslant f(y)$ в Q.

Морфизмы бинарных операций — это частичные случаи морфизмов тернарных отношений. Рассмотрим *тернарное отношение* на множестве S как функцию α: $S^3 \longrightarrow \{0, 1\}$; по определению, $(x, y, z)\alpha$, если $\alpha(x, y, z) = 1$, и $(x, y, z)\alpha'$, если $\alpha(x, y, z) = 0$. Тогда, скажем, бинарное «сложение» $+$: $S \times S \longrightarrow S$ на S определяет тернарное отношение α на S, означающее, что $x + y = z$ [1]). Функция θ: $S \longrightarrow T$, очевидно, является морфизмом для операции $+$ (на S и T) в том и только том случае, когда она является морфизмом относительно тернарных отношений $x + y = z$. Последнее по определению означает, что из всякого равенства $a + a' = b$ в S должно следовать равенство $\theta(a) + \theta(a') = \theta(b)$ в T (см. первое равенство (20)).

---

[1]) Иными словами, функция $\alpha(x, y, z)$ есть истинностная функция g высказывания $x + y = z$, равная 1, когда оно верно, и 0, когда оно ложно.

2.7. ЦИКЛИЧЕСКИЕ УНАРНЫЕ АЛГЕБРЫ

Простейший класс алгебраических систем состоит из алгебр с *одной унарной* операцией. В этой главе мы будем называть такие системы просто *унарными* алгебрами[1]).

Прежде всего напомним два простых примера.

Пример 8. *Расширенной алгеброй Пеано* называется система $[\mathbf{N}, \tau]$, где $\mathbf{N}$ — множество неотрицательных целых чисел, и $\tau(n) = n + 1$.

Пример 9. *Циклической алгеброй порядка* m называется система $[\mathbf{m}, \sigma_m]$, где $\mathbf{m} = \{1, \ldots, m\}$ и

$$\sigma_m(k) = \begin{cases} k+1, & \text{если } k \neq m, \\ 1, & \text{если } k = m. \end{cases} \tag{24}$$

На следующих диаграммах представлены две эти алгебры; действие унарной операции изображено стрелками.

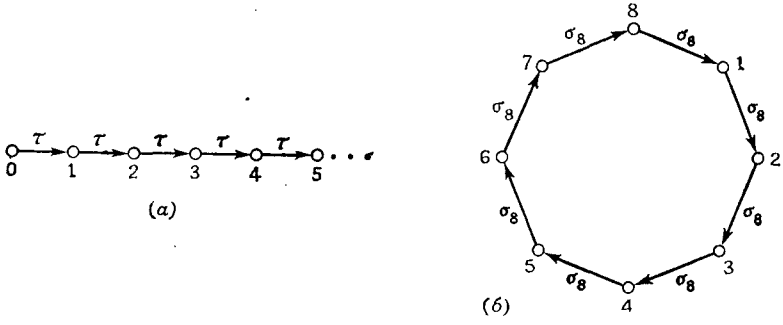


Рис. 2.4.

Пусть $[S, f]$ — некоторая унарная алгебра. Подмножество $T \subset S$ называется *f-замкнутым*, если для всякого $t \in T$ мы имеем $f(t) \in T$. Пустое подмножество и все S, очевидно, f-замкнуты. В $[\mathbf{m}, \sigma_m]$ нет других σ_m-замкнутых подмножеств, кроме пустого подмножества и $\mathbf{m}$. В $[\mathbf{N}, \tau]$ τ-замкнуты подмножества вида $\{k \mid k \geqslant n\}$ для любого $n \in \mathbf{N}$. Элемент $0 \in \mathbf{N}$ *порождает* $\mathbf{N}$ (0 является *образующей* $\mathbf{N}$) в том смысле, что единственное τ-замкнутое подмножество X в $\mathbf{N}$, содержащее 0, совпадает с $\mathbf{N}$.

Цель этого параграфа состоит в классификации всех унарных алгебр, порожденных одним элементом. Начнем с важного предварительного результата. Пусть $[S, f]$ — унарная алгебра, $a \in S$ — любой ее элемент. Рассмотрим функцию

$$\theta\colon\ 0 \mapsto a,\ 1 \mapsto f(a),\ 2 \mapsto f^2(a),\ \ldots,\ n \mapsto f^n(a),\ \ldots \tag{25}$$

[1]) Этот же термин применяется к системам с несколькими унарными операциями, но без операций и отношений большего ранга.

из $\mathbf{N}$ в S. Поскольку $\tau(n)=n+1$, имеем

$$\theta(\tau(n))=\theta(n+1)=f^{n+1}(a)=f(f^n(a))=f(\theta(n)) \qquad (25')$$

для всех $n\in\mathbf{N}$. Следовательно, θ является морфизмом унарных алгебр в следующем смысле.

Определение. Пусть $[S, f]$, $[T, g]$—две унарные алгебры. Отображение θ: $S\longrightarrow T$ называется их *морфизмом*, если

$$\theta(f(s))=g(\theta(s)) \quad \text{для всех } s\in S. \qquad (26)$$

Иными словами, следующая диаграмма должна быть коммутативной

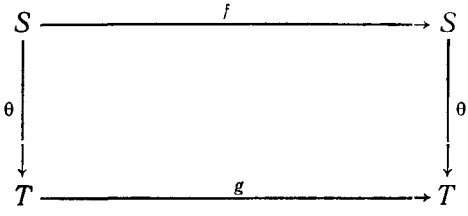


Рис. 2.5.

(Напомним, что диаграмма называется *коммутативной*, если любым путям, соединяющим одну ее вершину с другой, отвечают последовательности отображений, композиция которых не зависит от выбора пути.)

Формулы (18) и (19) означают на этом языке следующее.

Теорема 7. *Пусть a—произвольный элемент любой унарной алгебры $[S, f]$. Тогда существует единственный морфизм θ: $[\mathbf{N}, \tau]\longrightarrow[S, f]$, отображающий $0\in\mathbf{N}$ в a.*

Иными словами, алгебра $[\mathbf{N}, \tau]$ обладает следующим *свойством универсального отображения*: любое отображение ее образующей 0 можно продолжить до морфизма. Это свойство принято выражать, говоря, что $[\mathbf{N}, \tau]$ есть *свободная унарная алгебра* с одной образующей. Позже мы опишем другие свободные алгебры, но здесь ограничимся сказанным.

Следствие. *Всякую унарную алгебру с одной образующей можно представить в качестве эпиморфного образа алгебры* $[\mathbf{N}, \tau]$.

Теперь мы опишем (с точностью до изоморфизма) все возможные унарные алгебры $[S, f]$ с одной образующей $a\in S$. Поскольку S порождена a, имеем $S=\{f^n(a)\}$, множество всех элементов вида $f^n(a)$, замкнутое относительно f.

Случай 1. Все $f^n(a)$ различны. Тогда отображение $n\mapsto f^n(a)$ является изоморфизмом $[\mathbf{N}, \tau]$ в $[S, f]$.

Случай 2. Существует наименьшее p, для которого найдется $k < p$, такое, что $f^p(a) = f^k(a)$. Положим $p = k + m$. Индукция по j приводит к тождествам

$$f^{p+j}(a) = f^j(f^p(a)) = f^j(f^k(a)) = f^{k+j}(a) \text{ для всех } j \in \mathbf{N}. \quad (27)$$

Такая алгебра изображена на рис. 2.6.

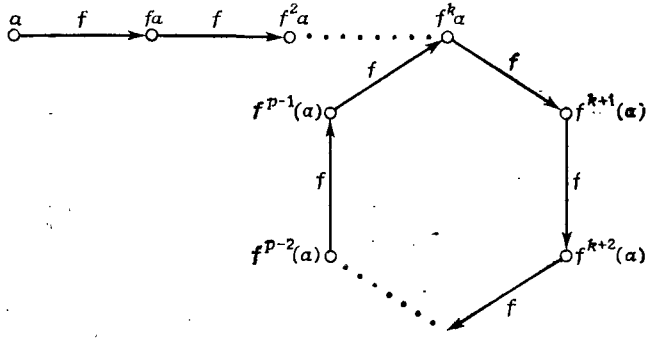


Рис. 2.6.

Ее диаграмма состоит из петли (цикла) длины m, в которую входит «хвост» длины k.

Определим теперь *унарную алгебру* U_m^k как алгебру, элементами которой являются целые числа $0, 1, \ldots, m+k-1$, с унарной операцией $\tau_{m,k}$:

$$\tau_{m,k}(j) = \begin{cases} j+1, & \text{если } j \neq m+k-1, \\ k, & \text{если } j = m+k-1. \end{cases}$$

Мы доказали следующий результат:

Теорема 8 (Дедекинд). *Любая унарная алгебра с одной образующей изоморфна либо* $\mathbf{P}$, *либо одной из алгебр* U_m^k $(m \in \mathbf{P},\ k \in \mathbf{N})$.

УПРАЖНЕНИЯ Г

1. Доказать, что частично упорядоченное множество $[\mathbf{n}, \leqslant]$ не имеет нетривиальных автоморфизмов (изоморфизмов с самим собой), сохраняющих порядок.

2. Доказать, что любая биекция между двумя конечными частично упорядоченными множествами, сохраняющая отношение доминирования, является изоморфизмом относительно отношения порядка. (*Указание*: воспользоваться упр. Б5.)

3. Доказать, что функция f: $0 \mapsto 0$, $1 \mapsto 3$, $2 \mapsto 2$, $3 \mapsto 1$ является изоморфизмом относительно операции сложения в $\mathbf{Z}_4$, но не относительно операции умножения.

4. Функция f: $\mathbf{4} \longrightarrow \mathbf{4}$ задана предписанием $f(1) = 2$, $f(2) = 3$, $f(3) = 2$, $f(4) = 1$.

а) Доказать, что ее степени $f^0 = 1_3$, $f^1 = f$, $f^2 = f \circ f$, $f^3 = f^2 \circ f$ все различны, но $f^4 = f^2$, $f^5 = f^3$.

б) Построить таблицу умножения для моноида, состоящего из всех степеней f.

5. Доказать, что всякая цепь из n элементов изоморфна $[\mathbf{n}, \leqslant]$.

6. Доказать, что отображение $\rho \longrightarrow \tilde{\rho}$ является изоморфизмом из $[\mathcal{P}(S \times S), \Diamond]$ в $[\mathcal{P}(S \times S), \circ]$, а также изоморфизмом относительно операций $\cap$, $\cup$, $'$.

7. Числа $m, n \in \mathbf{N}$ фиксированы. Отношение E на $\mathbf{N}^m$ определяется следующим образом: $(x_1, \ldots, x_m) E (y_1, \ldots, y_m)$ означает, что $n \mid (x_i - y_i)$ для всех $i \in \mathbf{m}$.

а) Показать, что E — отношение эквивалентности.

б) Определим в $\mathbf{N}^m$ сложение и умножение покомпонентно:

$$(x_1, \ldots, x_m) + (y_1, \ldots, y_m) = (x_1 + y_1, \ldots, x_m + y_m),$$
$$(x_1, \ldots, x_m)(y_1, \ldots, y_m) = (x_1 y_1, \ldots, x_m y_m).$$

Показать, что E обладает свойством подстановки относительно этих операций.

8. Пусть E — любое отношение эквивалентности на $\mathbf{Z}_n$.

а) Показать, что если E обладает свойством подстановки относительно сложения в $\mathbf{Z}_n$, то оно обладает этим свойством также относительно умножения.

б) Найти отношение эквивалентности на $\mathbf{Z}_3$, которое обладает свойством подстановки относительно умножения, но не относительно сложения.

9. Пусть $A = [S, f]$ — конечная унарная алгебра, состоящая из k элементов. Определим отношение aRb в A условием: существует такое $n \in \mathbf{N}$, что $f^n(a) = b$.

а) Показать, что это отношение рефлексивно и транзитивно.

б) Показать, что A циклична в том и только том случае, если существует такой $a \in A$, что aRb для всех $b \in A$.

10. Пусть P — любое частично упорядоченное множество. Рассмотрим функцию φ: $P \longrightarrow \mathcal{P}(P)$, которая ставит в соответствие каждому элементу $a \in P$ множество $A = \varphi(a)$ всех элементов $x \leqslant a$ в P. Показать, что это отображение инъективно и переводит $\leqslant$ в $\subset$ (т. е. является мономорфизмом).

2.8. ОРИЕНТИРОВАННЫЕ ГРАФЫ

Ориентированным графом (а также *направленным* графом) называется тройка $\vec{G} = [N, A, \varphi]$, состоящая из (1) множества N *вершин*; (2) множества A *дуг* или (ориентированных) *ребер*; (3) функции φ: $A \longrightarrow N \times N$, которая ставит в соответствие каждому ребру (дуге) $a \in A$ упорядоченную пару (p, q) вершин, называемых *концами* этого ребра (дуги). Дуга с концами (p, p) в одной и той же вершине называется *петлей*. Граф, у которого нет таких дуг, называется *графом без петель*.

На рис. 2.7 изображены четыре ориентированных графа, каждый с четырьмя вершинами. Все эти графы не имеют петель. Кроме того, они просты: последнее, по определению, означает, что любая пара вершин p, q соединена не более чем одним ребром. (Некоторые авторы включают в определение простого графа условие, чтобы он не имел петель, ограничиваясь тем самым в следующей далее теореме 9 антирефлексивными отношениями.)

В простых ориентированных графах мы будем обозначать единственное ребро с концами p, q (если оно существует) через $\overrightarrow{pq}$.

Опишем два важных семейства простых ориентированных графов, каждый из которых зависит от параметра $n \in \mathbf{P}$. *Простой путь* $\vec{\Pi}_n$ длины n состоит из $n+1$ вершин $P_0, \ldots, P_n$ и n ребер, соединяющих соседние вершины $P_k\overrightarrow{P_{k+1}}$. *Простой цикл* Γ_n длины n состоит из n различных вершин $p_1, \ldots, p_n$ и n ребер вида $\overrightarrow{p_kp_{k+1}}$ для $k<n$ и, кроме того, $\overrightarrow{p_np_1}$. (В частности, простой цикл длины 1 является петлей с вершиной p_1.) Все простые пути длины n изоморфны, и все простые циклы длины n изоморфны (см. далее определение).

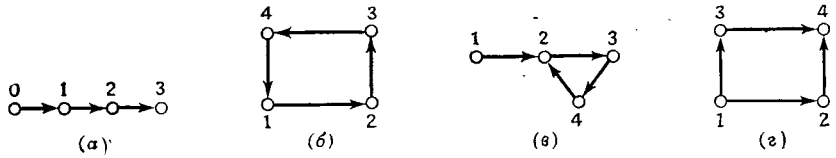


Рис. 2.7. Примеры ориентированных графов.

На рис. 2.7,*а* изображен простой путь длины 3, а на рис. 2.7,*б*—простой цикл длины 4 (граф унарной алгебры, описанной перед формулировкой теоремы 8). На рис. 2.7,*в* изображен граф унарной алгебры U_3^1, а на рис. 2.7,*г*—граф, часто встречающийся в коммутативных диаграммах.

Определение. *Изоморфизмом* ориентированных графов $\vec{G}=[N, A, \varphi]$ и $\vec{G}^*=[N^*, A^*, \varphi^*]$ называется такая пара биекций $\beta\colon N \to N^*$ и $\gamma\colon A \to A^*$, что в $\vec{G}$ ребро $\mathbf{a}$ идет от вершины p к вершине q в том и только том случае, если в $\vec{G}^*$ ребро $\varphi(\mathbf{a})$ идет от вершины $\beta(p)$ к вершине $\beta(q)$. Иными словами, $\varphi(\mathbf{a})=(p, q)$ в $\vec{G}$ равносильно условию $\varphi^*(\gamma(\mathbf{a}))=(\beta(p), \beta(q))$ в $\vec{G}^*$. Еще одна переформулировка—диаграмма на рис. 2.8 коммутативна:

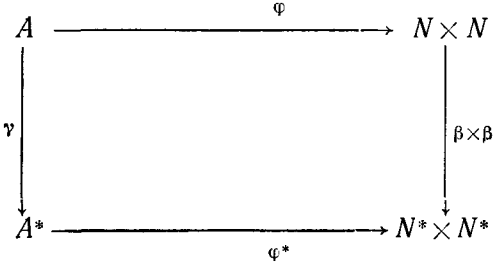


Рис. 2.8.

Если $\vec{G}$, $\vec{G}^*$ — простые графы, как на рис. 2.7, можно указать более простой способ проверки их изоморфности. Биекция $\beta\colon N \longrightarrow N^*$ продолжается до изоморфизма (и притом единственного) тогда и только тогда, когда отображение $\beta\times\beta$: $(p, q)\mapsto(p^*, q^*)$ $(p^*=\beta(p),\ q^*=\beta(q))$ индуцирует биекцию множества всех ребер $\overrightarrow{pq}$ графа G на множество всех ребер $\overrightarrow{p^*q^*}$ графа G^*, так что

$$\varphi^* \circ \gamma = (\beta\times\beta)\circ\varphi.$$

Два ориентированных графа называются *изоморфными*, если между ними существует изоморфизм. Изоморфные ориентированные графы естественно отождествлять. Мы будем рассматривать только свойства, сохраняющиеся при изоморфизме, как, например, свойства быть простым графом, простым путем или простым циклом.

Ориентированные графы и отношения. Всякий ориентированный граф $\vec{G}=[N, A, \varphi]$ определяет на множестве своих вершин бинарное *отношение следования* $\sigma_{\vec{G}}$. По определению,

$$p\sigma_{\vec{G}}q \text{ означает, что } \varphi(\mathbf{a})=(p, q) \text{ для некоторого } \mathbf{a}\in A. \quad (28)$$

Обратно, всякое бинарное отношение ρ на множестве N определяет простой упорядоченный граф $\vec{G}=\vec{G}(\rho)=[N, A(\rho), \psi]$, дуги которого $\overrightarrow{pq}$ отвечают парам $p\rho q$:

$$A(\rho)=\{\overrightarrow{pq} \mid p\in N,\ q\in N,\ p\rho q\},\qquad \psi(\overrightarrow{pq})=(p, q). \quad (29)$$

Нетрудно проверить, что *если* мы исходили из простого ориентированного графа $\vec{G}$, то $\vec{G}\left(\sigma_{\vec{G}}\right)=\vec{G}$. С другой стороны, $\sigma_{\vec{G}(\rho)}=\rho$ для *любого* бинарного отношения ρ. Таким образом, мы установили следующую теорему.

Теорема 9. *Класс множеств с одним бинарным отношением* $[U, \rho]$ *находится в естественном биективном соответствии с классом простых ориентированных графов. Эта биекция задается предписаниями* (28) *и* (29).

Как мы уже отмечали, все графы на рис. 2.7 простые. Вершины их перенумерованы. Поэтому их можно задавать 4×4-матрицами, которые в то же время будут матрицами отношений, отвечающих этим графам в силу теоремы 9:

$$\begin{bmatrix}0&1&0&0\\0&0&1&0\\0&0&0&1\\0&0&0&0\end{bmatrix}\quad \begin{bmatrix}0&1&0&0\\0&0&1&0\\0&0&0&1\\1&0&0&0\end{bmatrix}\quad \begin{bmatrix}0&1&0&0\\0&0&1&0\\0&0&0&1\\0&1&0&0\end{bmatrix}\quad \begin{bmatrix}0&1&1&0\\0&0&0&1\\0&0&0&1\\0&0&0&0\end{bmatrix}$$

(*а*) (*б*) (*в*) (*г*)

Рис. 2.9. Матрицы отношений для графов, изображенных на рис. 2.7.

Если $\vec{G}$ — *любой* ориентированный граф, то $\vec{G}\left(\sigma_{\vec{G}}\right)$ является простейшим ориентированным графом, который получается из $\vec{G}$ отождествлением всех дуг с общим началом и концом.

Матрицы ориентированных графов. Пусть $\vec{G}$ — любой ориентированный граф с множеством вершин N и дуг A. Если же множества упорядочены, с графом $\vec{G}$ можно сопоставить его матрицу инцидентности $B=\|b_{ij}\|$ размера $|N|\times|A|$ [1]):

$$b_{ij}=\begin{cases} -1, & \text{если } a_j\in A \text{ есть начало } n_i\in N,\\ +1, & \text{если } a_j\in A \text{ есть конец } n_i\in N,\\ 0 & \text{в остальных случаях.}\end{cases} \tag{30}$$

Если граф $\vec{G}$ простой, то его можно описать также *матрицей отношения следования* (для любого упорядочения N) размера $|N|\times|N|$. Эта матрица

$$\sigma_{ij}=\begin{cases} 1, & \text{если } n_i\sigma_G n_j,\\ 0 & \text{в противном случае.}\end{cases} \tag{31}$$

Графы отношений доминирования. На рис. 2.9, *а* и *г* изображены отношения доминирования для двух частично упорядоченных множеств: цепи из четырех элементов и булевой алгебры $2^2=\mathcal{P}(2)$ соответственно. Ориентированный граф на рис. 2.10 описывает отношение доминирования в частично упорядоченном множестве $\mathcal{P}(3)$; он напоминает рисунок куба.

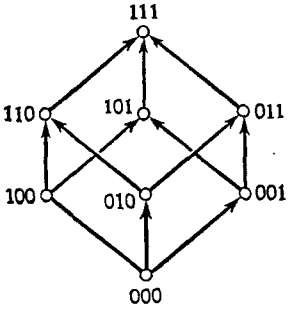


Рис. 2.10.

Вот обобщение этого примера.

Пример 10. Ориентированный граф *n-мерного куба* имеет вершины, перенумерованные всеми двоичными последовательностями $\mathbf{x}=x_1\ldots x_n$ длины n (их 2^n штук). Вершина $\mathbf{x}$ *доминирует* над вершиной $\mathbf{y}$ тогда и только тогда, когда номер $\mathbf{y}$ можно получить из номера $\mathbf{x}$, заменив одну двоичную единицу $x_i=1$ нулем: $y_i=0$. Наш граф является графом этого отношения. Оно совпадает с отношением доминирования в множестве $\mathcal{P}(\mathbf{n})$, частично упорядоченном отношением включения.

---

[1]) В книгах Бусакера и Саати [3, стр. 103] и Бержа [2] (см. литературу в конце главы) -1 отмечает конец дуги, а $+1$ — ее начало.

2.9. ГРАФЫ

Хотя ориентированные графы весьма полезны при описании технических конструкций и процессов, с чисто математической точки зрения следующее понятие изучено подробнее и выглядит более естественным.

Определение. *Графом* $G=[V, E, \theta]$ называется тройка, состоящая из следующих объектов: (1) множество V *вершин*; (2) множество E (неориентированных) *ребер*; (3) функция θ, ставящая в соответствие каждому ребру $a \in E$ *неупорядоченную* пару вершин $(p, q)=(q, p)$, которые называются *концами* этого ребра. Ребро (p, p) называется *петлей*.

Пример 11. *Полный* n-граф — это граф, имеющий n вершин, причем каждая пара вершин соединена ровно одним ребром. На рис. 2.11, *а* и *б* изображены полный 4-граф и полный 5-граф соответственно.

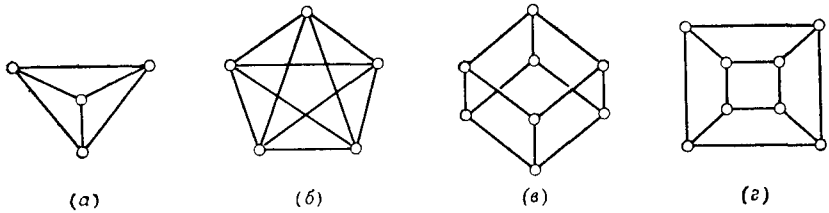


Рис. 2.11. Примеры графов.

Пример 12. Граф n-куба имеет 2^n вершин, перенумерованных всеми двоичными последовательностями длины n. Две вершины соединены ребром тогда и только тогда, когда соответствующие им последовательности отличаются точно в одном месте. На рис. 2.11,*в* изображен граф 3-куба. Графы 2.11,*в* и 2.11,*г* изоморфны (докажите это).

Все графы, изображенные на рис. 2.11, не имеют петель; кроме того, никакие две вершины их не соединены более чем одним ребром. Такие графы называются *простыми*.

Любой граф $G=[V, E, \theta]$, очевидно, определяется отношением *инцидентности* ρ_G между множествами E и V: $l\rho_G p$ означает, что вершина $p \in V$ является концом ребра $l \in E$. Простой граф определяется также отношением смежности α на V: $p\alpha q$ означает, что вершины p и q соединены (единственным) ребром. Отношение иррефлексивно и симметрично.

Если элементы E и V заданы с помощью упорядоченного списка, то отношение инцидентности ρ_G можно описать *матрицей инцидентности* R_G. Отношение смежности вершин простого графа описывается (симметричной и иррефлексивной) *матрицей смежности* $A=\|\alpha_{ij}\|$.

Понятия, введенные для ориентированных графов в § 2.8, можно распространить на неориентированные графы, если считать, что неориентированное ребро $\overline{ab}$ отвечает паре ориентированных ребер $\overrightarrow{ab}$ и $\overleftarrow{ab}$. Тогда мы придем к следующим очевидным результатам:

Теорема 10. *Следующие соответствия определяют взаимно обратные биекции класса всех простых графов с классом всех простых ориентированных графов без петель, на множестве вершин которых отношение связи*[1]) *симметрично:*

(i) $[N, E] \mapsto [N, L]$, *где* $\overrightarrow{pq} \in L$

тогда и только тогда, когда $\overline{pq} \in E$; (32)

(ii) $[N, L] \mapsto [N, E]$, *где* $\overline{pq} \in E$

тогда и только тогда, когда $\overrightarrow{pq} \in L$. (32′)

Простые пути и простые циклы из § 2.8 при этой биекции отвечают следующим понятиям. *Цепь* длины n в графе G есть последовательность вершин $p_i \in V$, $i = 0, \ldots, n$, и ребер $\overline{p_{i-1}p_i} \in E$. Если к тому же $\overline{p_np_0} \in E$, то последовательность этих вершин и ребер $\overline{p_0p_1}, \ldots, \overline{p_np_0}$ называется *циклом* длины $n+1$. Если никакая вершина в этой последовательности не повторяется, соответствующие цепь и цикл называются *простыми*. Ясно, что если вершины p и q вообще можно соединить, то их можно соединить простой цепью. Если две простые цепи имеют общие концы, то их объединение является циклом.

Бихроматические графы. Граф G называется *бихроматическим*, если множество его вершин можно разбить на два непустых дополнительных подмножества R и B («белые» и «черные» вершины, как на шахматной доске), такие, что любая пара соседних вершин имеет разный цвет. Если граф G прост, это означает, что его матрицу смежности можно представить в виде $\begin{pmatrix} 0 & K \\ L & 0 \end{pmatrix}$; два квадрата на главной диагонали состоят из нулей.

УПРАЖНЕНИЯ Д

1. Пусть ρ — отношение смежности на множестве вершин графа с n вершинами. Показать, что $(e \vee \rho)^n = \bigvee_{k=0}^{n} \rho^k$ есть наименьшее транзитивное отношение, содержащее ρ.

---

[1]) В оригинале «linking». — *Прим. перев.*

2. Доказать, что любой путь наименьшей длины от вершины x к вершине y в ориентированном графе $\vec{G}$ является простым путем.

3. Вершина графа называется четной или нечетной в соответствии с тем, четно или нечетно число инцидентных ей ребер. Показать, что у любого конечного графа число нечетных вершин четно.

4. Доказать, что если отношения ρ и σ симметричны, то отношения смежности графов $\vec{G}_\rho + \vec{G}_\sigma$ и $\vec{G}_\rho \times \vec{G}_\sigma$ симметричны.

5. Доказать, что ориентированные графы на рис. 2.7,*б* и *г* неизоморфны.

В упр. 6 и 7 эйлеровым графом называется (неориентированный) граф, обладающий следующим свойством: существует путь, проходящий ровно один раз по каждому из его ребер.

6. Доказать, что конечный связный граф, у которого все вершины четные, эйлеров.

7. Доказать, что конечный связный граф является эйлеровым тогда и только тогда, когда количество его нечетных вершин равно 0 или 2.

8. *Деревом* называется связный граф без простых циклов.

а) Показать, что дерево с n вершинами имеет $n-1$ ребер.

б) Показать, что связный граф с n вершинами и $n-1$ ребрами является деревом.

9. а) Показать, что все деревья с тремя вершинами изоморфны.

б) Найти два неизоморфных дерева с четырьмя вершинами и три—с пятью вершинами.

в) Показать, что каждый граф с четырьмя или пятью вершинами изоморфен одному из графов п. б.

10. Доказать, что если B—матрица инцидентности простого ориентированного графа без петель, то его матрица смежности получается из BB^T путем замены всех элементов на диагонали нулями.

11. Рассмотрим конечный связный граф на плоскости, ребра которого не пересекаются вне вершин и в каждой вершине сходятся не менее двух ребер. Пусть число его вершин равно n_0, а ребер n_1. Доказать, что ограниченная им внутренняя область разбивается на $n_1 - n_0 + 1$ попарно не пересекающихся многоугольников.

12. Напомним, что у матрицы смежности *полного n-графа* все элементы вне главной диагонали равны 1. Доказать, что полный 4-граф нельзя изобразить на плоскости так, чтобы его ребра не пересекались вне вершин.

13. Пусть G—связный ориентированный граф, в каждую вершину которого входит столько же ребер, сколько выходит. Доказать, что в G существует ориентированный цикл, проходящий через каждое ребро ровно один раз.

14. Доказать, что произведение двух бихроматических графов обладает этим же свойством.

* 2.10. ОРИЕНТИРОВАННЫЕ ГРАФЫ, II

Пусть $\vec{G} = [N, A, \varphi]$—ориентированный граф. Пусть $N_1 \subset N$, $A_1 \subset A$ и концы всех $a_1 \in A_1$ лежат в N_1. Ограничивая φ на A_1, N_1, мы получим *подграф* графа $\vec{G}$.

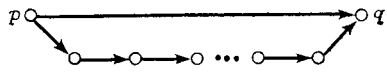
Понятие морфизма графов является частным случаем морфизма отношений. Например, рассмотрение морфизмов простых

путей и циклов в ориентированные графы приводит к следующим определениям.

Определение. *Путем* длины n в ориентированном графе $\vec{G}=[N, A, \varphi]$ называется последовательность $n+1$ вершин $p_0, \ldots, p_n$ (не обязательно различных) и связывающих их ребер $\overrightarrow{p_{i-1}p_i} \in A$ $(i=1, \ldots, n)$. Если $p_0=p_n$, этот путь называется *циклом* длины n. Путь называется простым, если все его вершины разные. Цикл называется *простым*, если вершины $p_0, \ldots, p_{n-1}$ различны.

Таким образом, простые пути и простые циклы длины n в графе $\vec{G}$ суть образы $\vec{\Pi}_n$, $\vec{\Gamma}_n$ при мономорфных вложениях (определения $\vec{\Pi}_n$, $\vec{\Gamma}_n$ даны в § 2.8).

Очевидно, ориентированный граф отношения доминирования любого частично упорядоченного множества не содержит нетривиальных циклов. Он не содержит также подграфов вида



в которых две вершины соединены ребром и простым путем длины $n \geqslant 2$.

Верно и обратное утверждение.

Суммы и произведения. «Синтез» больших графов из малых компонент требует перенесения понятий суммы и декартова произведения с множеств на графы и ориентированные графы.

Пусть $\vec{G}=[S, A, \varphi]$, $\vec{H}=[T, B, \psi]$ — два ориентированных графа. Их *суммой*, или *разделенным объединением*, $\vec{G}+\vec{H}$ называется граф

$$\vec{G}+\vec{H}=[S \sqcup T, A \sqcup B, \varphi \sqcup \psi]. \tag{33}$$

Если ρ — отношение следования в $\vec{G}$, τ — отношение следования в $\vec{H}$, то отношение следования в $\vec{G}+\vec{H}=\vec{G}(\theta)$ определяется условиями:

$$\begin{gathered} s\theta s_1 \text{ означает } s\rho s_1;\ t\theta t_1 \text{ означает } t\tau t_1; \\ \text{ни для какой пары } s \in S,\ t \in T \text{ неверно, что } s\theta t \text{ или } t\theta s. \end{gathered} \tag{33'}$$

Отсюда видно, что если упорядочить вершины $\vec{G}+\vec{H}$ так, чтобы все вершины $\vec{G}$ предшествовали вершинам $\vec{H}$, то матрица отношения следования графа $\vec{G}+\vec{H}$ будет *прямой суммой* соответствующих матриц M, N для $\vec{G}$ и $\vec{H}$ (при индуцированном

упорядочении вершин). Эта матрица имеет *приведенный* вид $\begin{pmatrix} M & 0 \\ 0 & N \end{pmatrix}$. На диагонали стоят квадратные матрицы M и N.

Декартово произведение графов $\vec{G}$, $\vec{H}$ определяется аналогично:

$$\vec{G} \times \vec{H} = [S \times T,\ C,\ \gamma], \tag{34}$$

где C есть сумма $(S \times B) \sqcup (T \times A)$ и

$$\gamma(s,\ b) = ((s,\ \psi_1(b)),\ (s,\ \psi_2(b))) \text{ для } (s,\ b) \in S \times B, \tag{35}$$

$$\gamma(t,\ a) = ((\varphi_1(a),\ t),\ (\varphi_2(a),\ t)) \text{ для } (t,\ a) \in T \times A. \tag{35'}$$

Например, граф 2.7, *г* является декартовым произведением двух простых путей длины 1.

Аналогично можно определить сумму и произведение неориентированных графов. Мы оставляем читателю формулировки определений и свойств этих операций (см. упражнения Д).

Достижимость и связность. Вершина q ориентированного графа $\vec{G}$ называется *достижимой* из вершины p, если $q = p$ или существует путь (а значит, и простой путь) из p в q. Будем писать в этом случае $p \alpha q$.

Алгебра отношений позволяет просто выразить отношение α достижимости на графе $\vec{G}(\rho)$ через отношение следования ρ. Действительно, отношение $p \rho^n q$ означает, что существует путь длины n от p до q, где ρ^n определяется по индукции для $k \in \mathbf{N}$: $\rho^1 = \rho$ и $\rho^{k+1} = (\rho^k)\rho$.

Определим символ $\bigvee_{k=0}^{r}$ при $r \in \mathbf{N}$ аналогично, полагая

$$\bigvee_{k=0}^{1} \rho^k = e \vee \rho, \qquad \bigvee_{k=0}^{r+1} \rho^k = \left(\bigvee_{k=0}^{r} \rho^k \right) \vee \rho^{r+1}, \tag{36}$$

где $\rho^0 = e$ — отношение равенства. Для графа с n вершинами имеем:

$$\alpha = \bigvee_{k=0}^{n} \rho^k; \tag{37}$$

доказательство мы предоставляем читателю. Это означает, что отношение достижимости α на $\vec{G}(\rho)$ есть наименьшее рефлексивное и транзитивное отношение, содержащее ρ, т. е. наименьшее α со свойствами

$$\alpha \geqslant e, \qquad \alpha^2 \leqslant \alpha, \qquad \alpha \geqslant \rho. \tag{38}$$

Если вершины $\vec{G}(\rho)$ перенумерованы: $p_1, \ldots, p_n$, и $R=\|\rho_{ij}\|$ — матрица следования, то матрица достижимости равна

$$A=\bigvee_{k=0}^{n} R^k,$$

где $R^k=R\diamond\ldots\diamond R$ (k раз).

Ориентированный граф $\vec{G}$ называется *сильно связным*, если любая его вершина достижима из любой другой вершины. Если $\vec{G}$ имеет n вершин, сильная связность эквивалентна соотношению $\bigvee_{k=0}^{n} R^k=J$, где J — *универсальное отношение*, матрица которого состоит полностью из 1.

Аналогично можно определить отношение достижимости в неориентированных графах: «q достижима из p» означает «$p=q$ или имеется цепь с концами p, q» (см. теорему 10). Это — отношение эквивалентности на вершинах графа G. Каждый класс эквивалентности вместе с ребрами, соединяющими вершины этого класса, называется *связной компонентой* графа G. Весь граф G разлагается в сумму своих связных компонент $G_1, \ldots, G_r$.

Кратчайшие пути. Назовем *расстоянием* от вершины p до вершины q и обозначим через $d(p, q)$ длину кратчайшего пути от p до q в ориентированном или неориентированном графе. Если упорядоченный граф сильно связен, то расстояние можно вычислить, исходя из отношения следования σ, посредством следующего алгоритма.

Положим сначала $Q_0=\{p\}$. Выпишем затем список Q_1 всех вершин $q\neq p$ с условием $p\sigma q$. После этого составим список Q_2 всех вершин $r\notin Q_0\cup Q_1$ с условием $q\sigma r$ для некоторой $q\in Q_1$. Вообще пусть Q_{n+1} — список всех вершин $t\notin Q_0\cup\ldots\cup Q_n$, для которых $s\sigma t$ с подходящей $s\in Q_n$. Тогда $d(p, t)=m$ для всех $t\in Q_m$.

Деревья. Ориентированный или неориентированный граф без циклов (в том числе без петель) называется *ациклическим*. Связный ациклический граф называется *деревом*; имеется любопытная теория деревьев.

Для нас важнее аналогичный класс ориентированных графов, состоящий из деревьев с *корнем*.

По определению, дерево с корнем — это дерево, у которого имеется единственная вершина — *корень*, из которой любая другая вершина достижима в точности по одному пути. Поскольку всякий путь содержит простой подпуть, этот единственный путь сам должен быть простым. Особый интерес представляют *бинарные* деревья с корнем: в них всякая вершина, не являющаяся

конечной, служит началом ровно двух ребер. Например, разные способы композиции последовательности из n элементов в бинарной алгебре можно описать посредством бинарных деревьев с корнем и n конечными вершинами (упорядоченными раз и навсегда). Такой граф содержит $2n-1$ вершин, не являющихся конечными. Например, на рис. 2.12 изображены бинарные деревья с корнем, отвечающие пяти различным способам перемножения элементов a, b, c, d; ассоциативность умножения не предполагается (ср. гл. 1, (34)).

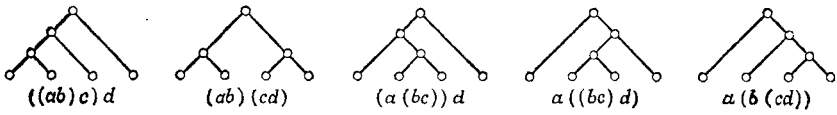


Рис. 2.12. Группировки в попарные произведения.

Помеченные ориентированные графы. Понятие ориентированного графа имеет много важных обобщений. Например, можно отметить специальными метками («окрасить») вершины и/или ребра графа (как в коммутативных диаграммах). Получающийся в результате объект называется помеченным, или окрашенным, (ориентированным) графом. В следующей главе мы будем рассматривать конечные автоматы как такие графы. В гл. 6 мы рассмотрим задачу отыскания кратчайших путей в помеченном направленном графе, ребрам которого приписаны различные *длины*.

Графы игр. Мы закончим эту главу, показав, как можно применить помеченные ориентированные графы к анализу выигрышных стратегий во многих играх двух лиц с полной информацией, как, например, шахматы или шашки.

Свяжем с каждой такой игрой граф, *вершинами* которого являются всевозможные позиции, а *ребрами* — всевозможные допустимые ходы, переводящие одну позицию в другую. Ребра должны быть *помечены* указанием, какой игрок («белые» или «черные») делает ход. Пусть правила игры задают множества конечных позиций, скажем W_0, B_0, в которых соответственно выигрывают «белые» или «черные». Тогда множества W и B *всех* выигрышных (при правильной игре) позиций для двух игроков можно вычислять рекурсивно следующим образом:

(i) $W=\bigcup_{n=0}^{\infty} W_n$, где W_{n+1} состоит из всех позиций, *не входящих* в $\bigcup_{k=1}^{n}(W_k \cup B_k)$ и таких, что

(α) если в этой позиции ход белых, то некоторый ход белых приводит в W_n;

(β) если в этой позиции ход черных, то любой ход черных приводит в W_n;

(ii) $B = \bigcup_{n=0}^{\infty} B_n$, где B_{n+1} определяется по симметрии.

Очевидно, W_m и B_m суть множества позиций, в которых белые (соответственно черные) при безошибочной игре выигрывают в m ходов.

Пример 13. В игре Ним два игрока по очереди берут несколько спичек из разложенных перед ними кучек; каждый раз можно взять любое число спичек, но лишь из одной кучки. Берущий последнюю спичку выигрывает.

Выразим число спичек в каждой кучке в двоичной системе и запишем получившиеся разложения друг под другом. Тогда *выигрышная* позиция для игрока, только что сделавшего ход, характеризуется тем, что сумма двоичных цифр в каждом разряде должна быть четной. Действительно, это так в последней позиции (все кучки пустые). В общем случае второй игрок, встретившись с такой позицией, вынужден будет изменить четность суммы цифр хотя бы в одном столбце, после чего первый игрок сможет взять подходящее количество спичек из самой большой кучки, восстановив условие четности. Так, имея перед собой четыре кучки по 13, 11, 5 и 10 спичек соответственно, он может взять 9 спичек из первой кучки[1]):

$$1101 = 13$$
$$1011 = 11$$
$$0101 = 5$$
$$1010 = 10$$

СПИСОК ЛИТЕРАТУРЫ

1. Beckenbach E. (ed.), Applied Combinatorial Mathematics, Wiley, 1964. (Русский перевод: Прикладная комбинаторная математика, под редакцией Э. Беккенбаха, «Мир», М., 1968.)
2. Berge C., The Theory of Graphs, Wiley-Methuen, 1962. (Русский перевод: Берж К., Теория графов и ее применение, ИЛ, М., 1962.)
3. Busacker R. G., Saaty T. L., Finite Graphs and Networks, McGraw-Hill, 1965.
4. Hall M., Combinatorial Mathematics, Ginn-Blaisdell, 1967. (Русский перевод: Холл М., Комбинаторная математика, «Мир», М., 1970.)
5. Harary M., Graph Theory and Theoretical Physics, Academic Press, 1967.
6. Ore O., Theory of Graphs, Amer. Math. Society, 1962. (Русский перевод: Оре О., Теория графов, «Мир», М., 1965.)

[1]) Дальнейшие подробности см. в книге Hardy, Wright, The Theory of Numbers, sec. 9-8.

3

КОНЕЧНЫЕ АВТОМАТЫ

3.1. ВВЕДЕНИЕ

В предыдущих главах были определены понятия множества, функции, произведения множеств, отношения и графа. Настоящая глава посвящена математическому описанию работы *цифровых вычислительных машин* с помощью этой системы понятий. Мы исключаем из рассмотрения *аналоговые вычислительные машины*, состояния которых могут меняться непрерывно, и *гибридные* устройства, сочетающие цифровые и аналоговые компоненты.

Все многообразие видов цифровых машин можно отнести к одному классу *конечных автоматов*. Это означает, что они обладают следующими общими свойствами.

Во-первых, всякая цифровая вычислительная машина состоит из конечного множества *элементов*, каждый из которых в любой данный момент времени может находиться лишь в одном из конечного числа устойчивых *состояний*. Поэтому и вся машина имеет лишь конечное множество устойчивых состояний.

Во-вторых, каждая цифровая машина работает *последовательно;* ее операции синхронизированы сигналами тщательно настроенных электронных часов (обычно от 10^6 до 10^9 сигналов в секунду)[1]). В соответствии с этим состояния машины меняются в четкой последовательности.

В-третьих, цифровая вычислительная машина является *детерминированным* устройством: при наличии полной информации о внутренних состояниях всех элементов машины и всех ее входов следующее состояние машины определено однозначно.

Цифровые машины делятся на *универсальные* и *специализированные*. Нас будут интересовать прежде всего универсальные машины, поскольку их можно использовать для любых целей.

С функциональной точки зрения современная универсальная машина состоит из пяти типов устройств: (1) *устройства ввода*; (2) *память*; (3) *арифметическое устройство*; (4) *устройство управления*; (5) *устройства вывода*.

---

[1]) Конкретные данные об ЭВМ, приводимые авторами, в значительной степени устарели.— *Прим. перев.*

Устройства ввода служат для ввода данных и команд, записанных на перфолентах, перфокартах, магнитных лентах и др. или представленных в виде электрических сигналов. Устройства ввода могут также измерять те или иные физические величины и переводить результаты в двоичную систему, которую машина способна воспринять. Устройства вывода служат для передачи информации от машины к пользователям. К ним относятся быстродействующие печатающие устройства, пишущие машинки с различными шрифтами, осциллографы и т. д. Мы не будем обсуждать их в этой книге (см., например, [2]).

Чтобы машина выполнила необходимую последовательность команд, в ее память прежде всего должна быть заложена программа, предопределяющая эту последовательность. Вообще говоря, программа должна быть предварительно отперфорирована (или записана на магнитной ленте); затем устройства ввода считают ее и передадут в память.

Машина выполняет эту программу последовательно, команду за командой, начиная с первой. Однако сама программа может предусматривать пропуск или повторение тех или иных шагов (см. гл. 4), и эти пропуски и повторения управляемы; они в принципе зависят от характера информации, хранящейся в памяти.

Одни и те же элементы памяти, согласно предложению покойного Джона фон Неймана, используются и для данных, и для программ. Следует заметить, что часть элементов памяти рассеяна по разным устройствам небольшими группами (ввод-вывод, устройство управления и арифметическое устройство также содержат фрагменты памяти). Память в целом может насчитывать от 10^3 до 10^{10} элементов с двумя устойчивыми состояниями, каждый из которых хранит один бит информации. Число этих элементов определяет величину машины.

Арифметическое устройство машины совершает как арифметические, так и логические операции. Например, арифметическое устройство может сложить, перемножить, вычесть или поделить два числа, а также выполнить различные булевы операции (см. гл. 4).

Наконец, устройство управления интерпретирует команды, считываемые из памяти, вызывая их по порядку или иногда переставляя команды в программе в зависимости от характера данных, запасенных в памяти или вычисленных ранее.

3.2. ДВОИЧНЫЕ ЭЛЕМЕНТЫ И СОСТОЯНИЯ

Почти все цифровые машины конструируются на электронных схемах и других элементах, имеющих *два устойчивых состояния*. Основные причины этого — технологические. Быстродействие электронных схем увеличивается, если в качестве рабочих взяты два

крайних состояния; надежность также возрастает, ибо в этом случае небольшие отклонения характеристик не отражаются на устойчивости схемы. Итак, схемы и сигналы по большей части двузначны. Любой конкретный электрический сигнал, если исключить переходный период, принимает одно из двух значений.

На рис. 3.1 показан наиболее широко используемый сейчас способ представления сигналов. Единицей кодируется положительный сигнал, нулем — нулевой сигнал (земля).

Точнее, устанавливается некоторое пороговое значение сигнала; сигналы выше порога кодируются единицей, ниже порога — нулем. Например, в резисторно-транзисторных схемах порог имеет порядок 0,75 вольт, а в диодно-транзисторных — 1,5 вольт.

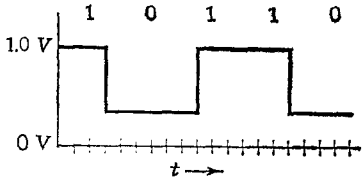


Рис. 3.1. Уровни постоянного тока и соответствующие двоичные значения.

Мы не будем входить в дальнейшие технологические подробности логических схем; для нас важно лишь, что сигналы в машине двузначны, так что переменные в языке для их описания также принимают два значения. То же относится к материальным носителям и преобразователям сигналов: ферритовые сердечники памяти могут находиться в одном из двух магнитных состояний, отверстие в данной позиции перфокарты может быть пробито или не пробито и т. д.

Состояние любой машины (автомата), состоящей из конечного числа r двоичных элементов, математически может быть описано так. Перенумеруем элементы в некотором порядке, $\delta_1, \ldots, \delta_r$. С каждым устойчивым состоянием автомата (т. е. его элементов) свяжем *вектор состояния* $x=(x_1, \ldots, x_r)$, приписав координате x_i значение 1, если δ_i находится в состоянии, помеченном 1, и 0, если δ_i находится в состоянии, помеченном 0.

3.3. КОНЕЧНЫЕ АВТОМАТЫ

Введем теперь математическое понятие, являющееся абстракцией описанных выше представлений.

Определение. *Конечным автоматом* называется набор из пяти объектов $[A, S, Z, \nu, \zeta]$. Здесь

$A=\{a_0, a_1, \ldots, a_n\}$ — конечный список *входных символов* (входной алфавит);

$Z=\{z_0, z_1, \ldots, z_m\}$ — список *выходных символов* (выходной алфавит);

$S=\{s_0, s_1, \ldots, s_r\}$ — множество *внутренних состояний*;

ν: $S\times A \to S$ — *функция перехода* (в следующее состояние);

ζ: $S\times A \to Z$ — *функция выхода*.

Тем самым, конечный автомат математически описывается тремя множествами и двумя функциями. Действие его состоит в том, что он «считывает» последовательность входных символов («программу») и затем «выпечатывает» последовательность выходных символов. Действие происходит последовательно. Конечный автомат, находящийся сначала во внутреннем состоянии s_j, считывает первый входной символ a_k. Функция ζ принимает на паре (s_j, a_k) значение z_l, которое выпечатывается в качестве первого выходного символа. Функция ν принимает на паре (s_j, a_k) значение s_r, которое является следующим внутренним состоянием

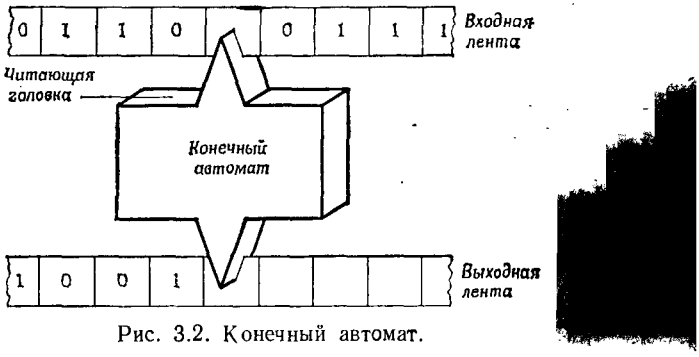


Рис. 3.2. Конечный автомат.

автомата. Затем автомат считывает новый входной символ, выпечатывает выходной, переходит в следующее состояние и т. д., пока не кончится программа.

На рис. 3.2 дан удобный способ представления последовательных тактов работы автомата.

Будем предполагать, что программа записана на *входной ленте*. Автомат считывает с нее входные символы один за другим. По прочтении каждого входного символа выпечатывается выходной символ на *выходной ленте* и автомат переходит в следующее состояние прежде чем считать следующий символ программы. Позже мы введем другие способы представления: графы и таблицы состояний.

(Инженеры и проектировщики вычислительных машин представляют себе входные символы в виде последовательности электрических сигналов или импульсов. Каждый входной символ есть некоторая комбинация импульсов, которые синхронизированы и могут подаваться через регулярные промежутки времени, скажем каждые 10 наносекунд (10^{-8} сек.). Таким образом, для инженера символы из A суть обозначения для комбинаций импульсов.)

В нашем определении подразумевается, что функции ν и ζ в описании автомата M всюду определены: каждый элемент $S \times A$ задает их значения. Такое описание автомата является

полным. Коль скоро задано начальное состояние такого автомата, он способен считывать любую программу и выдавать однозначно определенную цепочку символов. Иными словами, существует функция, которая ставит в соответствие любому начальному состоянию s_i и любой последовательности входных символов вполне определенную последовательность выходных символов. (Дальше мы часто будем говорить «автомат» вместо «конечный автомат».)

Пример 1. Рассмотрим следующий конкретный автомат $M=[A, S, Z, \nu, \zeta]$. Входной алфавит $A=\{0, 1\}$; выходной алфавит $Z=\{0, 1\}$; три внутренних состояния $S=\{s_0, s_1, s_2\}$; функции выхода и перехода задаются предписаниями

$$\begin{aligned}
\nu\colon\ & (s_0, 0) \mapsto s_1 & \zeta\colon\ & (s_0, 0) \mapsto 0\\
& (s_0, 1) \mapsto s_0 & & (s_0, 1) \mapsto 1\\
& (s_1, 0) \mapsto s_2 & & (s_1, 0) \mapsto 1\\
& (s_1, 1) \mapsto s_1 & & (s_1, 1) \mapsto 0\\
& (s_2, 0) \mapsto s_0 & & (s_2, 0) \mapsto 1\\
& (s_2, 1) \mapsto s_2 & & (s_2, 1) \mapsto 0
\end{aligned}$$

Подадим на вход последовательность 0,1,0,1. Если автомат находился сначала в состоянии s_0, то, считав первый символ 0, он перейдет в состояние s_1 и выпечатает 0. Считав затем 1, он останется в состоянии s_1 и выпечатает 0. Считав следующий 0, он перейдет в состояние s_2 и выпечатает 1. Наконец, считав последний символ 1, автомат закончит работу в состоянии s_2, имея на выходной ленте последовательность 0,0,1,0. Таким образом, автомат преобразовал вход 0,1,0,1 (или, короче, 0101) в 0,0,1,0 (или 0010).

Есть два удобных способа описать этот автомат. Прежде всего, можно построить помеченный ориентированный граф, называемый *диаграммой состояний*. Граф для нашего автомата показан на рис 3.3.

Вершины этого графа помечены символами, обозначающими внутренние состояния. Каждое ребро помечено парой симво-

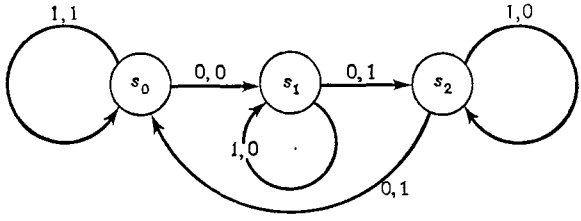


Рис. 3.3. Диаграмма состояний.

лов a, z, где a — входной символ, вызывающий переход в следующее состояние, отвечающее этому ребру, а z — выходной символ, который автомат выпечатывает.

Второй способ описания — *таблица состояний*. Из рис. 3.4, соответствующего нашему автомату, ясно, что это просто табличное представление функций ν и ζ.

Текущее состояние	Следующее состояние		Выход	
	Вход		Вход	
	ν 0	1	ζ 0	1
s_0	s_1	s_0	0	1
s_1	s_2	s_1	1	0
s_2	s_0	s_2	1	0

Рис. 3.4. Таблица состояний.

Оба способа имеют свои преимущества и недостатки. Таблица обычно удобнее при вычислениях, диаграмма нагляднее. Например, по диаграмме легко обнаружить состояния, не достижимые из других состояний. На рис. 3.5 показана диаграмма состояний автомата, у которого состояние s_1 недостижимо, если автомат начинает из состояния s_0 или s_2.

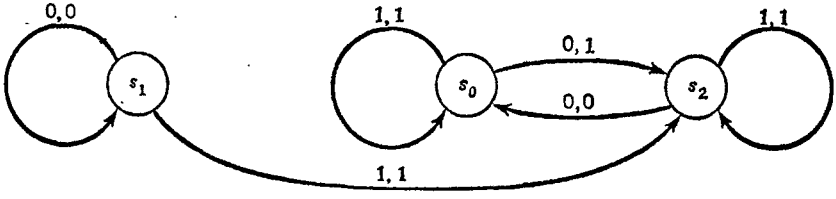
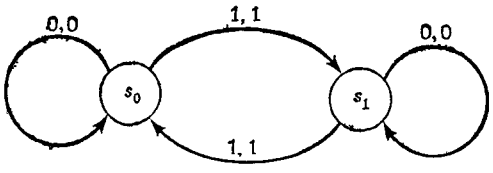


Рис. 3.5. Диаграмма с недостижимыми состояниями.

Пример 2. Автомат с двумя состояниями, изображенный на рис. 3.6, есть автомат *для проверки четности*.

Автомат считывает входную последовательность из нулей и единиц, и его состояние в любой момент времени совпадает с начальным, скажем, s_0, если число считанных к этому моменту единиц четно, и равно s_1, если число считанных единиц нечетно. Выходная последовательность совпадает с входной.

Пример 3. Автомат, изображенный на рис. 3.7, проверяет четность и выпечатывает EVEN (четный) или ODD (нечетный) в ответ на запрос, который соответствует входному символу Q.



Текущее состояние	Следующее состояние		Выход	
	вход 0	вход 1	вход 0	вход 1
s_0	s_0	s_1	0	1
s_1	s_1	s_0	0	1

Рис. 3.6. Автомат для проверки четности.

Считав Q, автомат выпечатывает EVEN, если число ранее считанных единиц было четно, и ODD — если нечетно. Например, входная последовательность $0110Q1110Q$ будет переработана в 0110EVEN1110ODD.

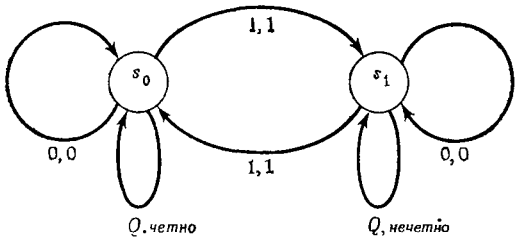


Рис. 3.7. Автомат, выпечатывающий информацию о четности.

УПРАЖНЕНИЯ А

1. Описать автомат M с алфавитами $Z = A = \{0, 1\}$, который, исходя из начального состояния s_0, перерабатывает любую входную последовательность $a^0a^1\ldots$ в последовательность $00a^0a^1a^2a^3\ldots$ (т. е. выход совпадает со входом, задержанным на два такта).

2. Начертить диаграмму состояний автомата с алфавитами $A = \{0, 1\} = Z$, который выпечатывает 1, если непосредственно перед этим он считал четыре последовательных 1; в противном случае выпечатывает 0. Автомат работает так до тех пор, пока не считает три последовательных 0, после чего выпечатывает лишь нули. Пример:

Вход:	0	0	1	0	1	1	1	1	1	0	1	0	0	0	1	0	1	1	1	1	0
Выход:	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0

3. Пусть $M = [A, S, Z, \nu, \zeta]$ — конечный автомат. Подадим на вход бесконечную последовательность $aaaa\ldots$, где $a \in A$.

а) Показать, что последовательность на выходе, начиная с некоторого места, будет периодической.

б) Дать численные оценки длин периода и части, предшествующей установлению периодичности, как функции от числа внутренних состояний автомата.

4. Пусть $[T, \cdot]$—конечная полугруппа, т. е. алгебраическая система из n элементов $t_1, \ldots, t_n$ с ассоциативным умножением $t_i t_j = t_{\varphi(i,\ j)}$.

а) Построить «умножающий автомат» с $n+1$ состояниями и алфавитами из $n+1$ символов, $A = Z = \{0,1,\ldots, n\}$, который по любому входу $0, i(1), \ldots, i(r)$ выпечатывает $i(1)$, $\varphi(i(1), i(2))$, ..., $\varphi(\ldots\varphi(\varphi(i(1), i(2)), i(3))\ldots)$.

б) Показать, что если $[T, \cdot]$ обладает единицей относительно умножения, то можно построить такой автомат с n состояниями и n-символьными алфавитами.

5. Описать автомат с двумя состояниями, который переводит десятичные цифры 0,1, ..., 9, поданные на вход, в двоичные последовательности 0000, 0001, ..., 1001 соответственно, а двоичные последовательности 0000, 0001, ..., 1111,—в десятичные записи 0, 1, ..., 15 соответственно.

6. Построить автомат M с 2^{n+1} внутренними состояниями, который, исходя из начального состояния s_0, переводит два n-битовых числа $\mathbf{i}$, $\mathbf{j}$ в сумму $\mathbf{i}+\mathbf{j}$ и переходит в состояние $s_{\mathbf{i}+\mathbf{j}}$.

7. Пусть $n = rk$. Построить автомат M, который переводит пару двоичных последовательностей $i(1), \ldots, i(n)$; $j(1), \ldots, j(n)$, длины n каждая, в последовательность из r двоичных «слов» длины $2k$, являющуюся двоичным представлением $(\tilde{w}) = (\tilde{w}_{k(1)} \ldots \tilde{w}_{k(r)})$ произведения чисел в двоичной записи $\mathbf{i} = i(1)\ldots i(n)$ и $\mathbf{j} = j(1)\ldots j(n)$.

3.4. ПОКРЫТИЕ И ЭКВИВАЛЕНТНОСТЬ

На вход конечного автомата подается некая последовательность символов входного алфавита A, которую мы обозначим $\mathbf{a}$. Так, если $A = \{0, 1\}$, на вход можно подать, например, $\mathbf{a} = 01101$. Первый символ последовательности $\mathbf{a}$ мы обозначим a^0, следующий a^1 и т. д. Так, при $\mathbf{a} = 01101$ имеем $a^0 = 0$, $a^1 = 1$, $a^2 = 1$, $a^3 = 0$, $a^4 = 1$. Поэтому можно записывать $\mathbf{a} = a^0a^1a^2a^3a^4$ или $\mathbf{a} = a^0, a^1, a^2, a^3, a^4$.

Упорядоченные последовательности символов называют по-разному: *векторы*, (конечные) *последовательности*, *n-ки*, *списки*, *массивы*. В этой главе мы будем обычно называть их *строками*. Словом «вектор» будем пользоваться в основном для обозначения элементов векторных пространств; «последовательностями» мы называем главным образом бесконечные цепи; «массив» будет специальным термином в описании языков программирования. Поэтому входные последовательности будем называть по большей части строками или иногда n-ками.

Итак, на вход подается *строка* $\mathbf{a}$, на выходе выпечатывается строка $\mathbf{z} = z^0, z^1, \ldots, z^{r-1}$ символов алфавита Z; полезно также рассматривать строки $\mathbf{s}$ внутренних состояний $s^i \in S$.

Пусть $M = [A, S, Z, \nu, \zeta]$—некоторый автомат. Тогда по любой входной строке $\mathbf{a} = a^0, a^1, \ldots, a^{r-1}$ длины r и по любому начальному состоянию $s^0 \in S$ однозначно определяется строка длины r внутренних состояний, $\mathbf{s} = s^0, s^1, \ldots, s^{r-1}$, которая

получается последовательным применением отображения ν. Точнее,

$$s^{j+1} = \nu(s^j, a^j), \quad j = 0, \ldots, r-2. \tag{1}$$

Аналогично выходная строка однозначно определяется последовательным применением отображения ζ:

$$z^j = \zeta(s^j, a^j), \quad j = 0, \ldots, r-1. \tag{2}$$

Поэтому, рассматривая автомат как устройство, перерабатывающее пары, состоящие из s^0 и $\mathbf{a} = a^0, \ldots, a^{r-1}$, в строки $\mathbf{s} = s^0, \ldots, s^{r-1}$ и $\mathbf{z} = z^0, \ldots, z^{r-1}$, мы можем определить с помощью (1) и (2) функции

$$\nu_r\colon S \times A^r \to S^r, \quad \zeta_r\colon S \times A^r \to Z^r, \tag{3}$$

которые рекурсивно строятся по ν и ζ, заданным в описании M. Здесь A^r — множество строк длины r символов из A, Z^r и S^r — множества строк длины r символов из Z, S.

Пример 4. Рассмотрим автомат, изображенный на рис. 3.8, с входной строкой $\mathbf{a} = 0110$ и $s^0 = s_0$. Тогда

$$\nu_4(s_0, 0110) = s_0 s_1 s_0 s_0, \quad \zeta_4(s_0, 0110) = 1001. \tag{4}$$

Увеличение числа внутренних состояний автомата не дается даром. В реальном устройстве оно приводит к росту количества

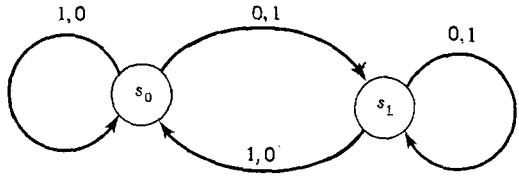


Рис. 3.8. Автомат с двумя состояниями.

требуемых электронных схем, к уменьшению надежности, к усложнению отладки и ремонта и т. д. Поэтому число необходимых состояний автомата, предназначенного для выполнения определенных действий, стремятся уменьшить, не ограничивая его возможностей.

Этим объясняется важность следующей задачи. Предположим, что входной и выходной алфавиты фиксированы. Можно ли заменить данный автомат $M = [A, S, Z, \nu, \zeta]$ автоматом с меньшим числом внутренних состояний $\overline{M} = [A, \overline{S}, Z, \overline{\nu}, \overline{\zeta}]$, но с той же функцией, переводящей входы в выходы? Следующее определение является специализацией этой проблемы.

Определение. Автомат $\overline{M}$ *покрывает* автомат M, если входной и выходной алфавиты у этих автоматов общие и суще-

ствует функция φ: $S \to \bar{S}$, такая, что для любого положительного числа r

$$\zeta_r(s, \mathbf{a}) = \bar{\zeta}_r(\varphi(s), \mathbf{a}) \quad \text{при всех } \mathbf{a} \in A^r. \tag{5}$$

Автомат, который нельзя покрыть меньшим автоматом, называется *минимальным* [1]).

Мы будем писать $\bar{M} \geqslant M$, если $\bar{M}$ покрывает M. Следующий результат очевиден, и мы опустим доказательство.

Лемма 1. *Отношение покрытия рефлексно и транзитивно.*

Определение. Автоматы M и $\bar{M}$ называются *эквивалентными*, если M покрывает $\bar{M}$ и $\bar{M}$ покрывает M. В этом случае мы пишем $M \equiv \bar{M}$.

Это означает, что кроме функции φ: $S \to \bar{S}$ со свойством (5) существует еще функция ψ: $\bar{S} \to S$ со свойством

$$\bar{\zeta}_r(\bar{s}, \mathbf{a}) = \zeta_r(\psi(\bar{s}), \mathbf{a}) \quad \text{при всех } \bar{s} \in \bar{S} \text{ и } \mathbf{a} \in A^r. \tag{6}$$

Следствие. *Отношение эквивалентности автоматов рефлексивно, транзитивно и симметрично.*

Покрытия и морфизмы. Отношения покрытия и эквивалентности автоматов тесно связаны с общим понятием морфизма, введенным в § 2.6. Понятие морфизма можно следующим образом распространить на автоматы M и $\bar{M}$ с общими алфавитами, входным и выходным.

Определение. *Морфизмом* называется такое отображение θ: $S \to \bar{S}$, что

$$\bar{\nu}(\theta(s), a) = \theta(\nu(s, a)) \quad \text{и} \quad \bar{\zeta}(\theta(s), a) = \zeta(s, a) \tag{7}$$

для всех $s \in S$ и $a \in A$. Если θ сюръективно, морфизм называется *эпиморфизмом*. Если θ биективно, морфизм называется *изоморфизмом* (автоматов).

Лемма 1. *Пусть θ — эпиморфизм автомата M на $\bar{M}$. Тогда для любой входной строки $\mathbf{a} = a^0, a^1, \ldots, a^{n-1}$ и начального состояния $s^0 \in S$ выходная строка*

$$\mathbf{z} = z^0, z^1, \ldots, z^{n-1}$$

автомата M совпадает с выходной строкой автомата $\bar{M}$, если начальное состояние $\bar{M}$ равно $\bar{s}^0 = \theta(s^0)$.

---

[1]) В русской литературе φ называется морфизмом по состояниям. — *Прим. перев.*

Доказательство проводится индукцией по n с индуктивным шагом:

$$\bar{s}^{n+1} = \bar{\nu}(\bar{s}^n, a^n) = \bar{\nu}(\theta(s^n), a^n) = \hat{\theta}(\nu(s^n, a^n)) = \theta(s^{n+1}),$$
$$\bar{z}^{n+1} = \bar{\zeta}(\bar{s}^n, a^n) = \bar{\zeta}(\theta(s^n), a^n) = \zeta(s^n, a^n) = z^{n+1}.$$

Следствие. *Любой эпиморфизм* μ: $M \to \bar{M}$ *определяет покрытие автомата* M *автоматом* $\bar{M}$.

Изоморфизм автоматов M и $\bar{M}$ с общими входными и выходными алфавитами есть просто биекция β: $S \leftrightarrow \bar{S}$, такая, что для всякого начального состояния $s^0 \in S$ и всякой входной строки $\mathbf{a} = a^1 a^2 \ldots a^n$ автоматы M и $\bar{M}$ выдают одну и ту же выходную

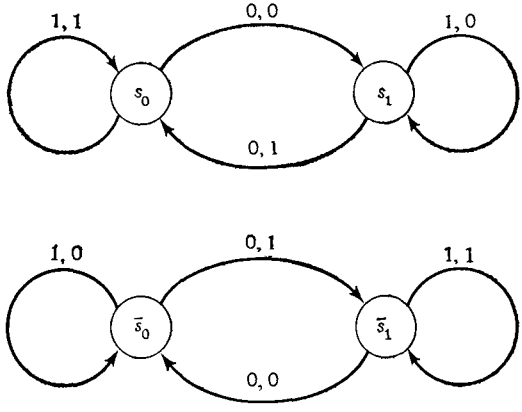


Рис. 3.9.

строку, проходя через соответствующие промежуточные состояния. Согласно этому определению, автоматы, представленные на рис. 3.9 ориентированными графами, изоморфны. Соответствующая биекция имеет вид

$$\beta\colon\ s_0 \mapsto \bar{s}_1 \quad s_1 \mapsto \bar{s}_0. \tag{8}$$

(Более общее определение морфизма автомата $M = [A, S, Z, \nu, \zeta]$ в автомат $\bar{M} = [\bar{A}, \bar{S}, \bar{Z}, \bar{\nu}, \bar{\zeta}]$ было предложено Хартманисом и Стирнсом. Согласно этому определению, морфизм состоит из трех отображений α: $A \to \bar{A}$, β: $S \to \bar{S}$ и ν: $Z \to \bar{Z}$ со свойствами $\bar{\nu}(\beta(s), \alpha(a)) = \beta(\nu(s, a))$, $\bar{\zeta}(\beta(s), \alpha(a)) = \nu(\zeta(s, a))$ для всех $s \in S$, $a \in A$.)

В нашей терминологии автоматы $M = [A, S, Z, \nu, \zeta]$ и $\bar{M} = [A, \bar{S}, Z, \bar{\nu}, \bar{\zeta}]$ с общими алфавитами A и Z изоморфны, если у них одинаковое число внутренних состояний и если сущест-

вует такая биекция β: $S \leftrightarrow \bar{S}$, что любая строка перерабатывается в одну и ту же выходную строку автоматами M, $\bar{M}$ с начальными состояниями s_i, $\bar{s}_j = \beta(s_i)$ соответственно. Это означает, что $\bar{M}$ получается из M просто переменой обозначений для внутренних состояний, как в случае изоморфизма любых алгебраических систем.

3.5. ЭКВИВАЛЕНТНЫЕ СОСТОЯНИЯ

В этом параграфе мы покажем, как решить следующую задачу: по данному описанию автомата M построить новый автомат $\bar{M}$, который (i) покрывает M (возможно, эквивалентен M) и (ii) имеет наименьшее число состояний среди всех автоматов, покрывающих M. Существует систематический и эффективный метод решения этой задачи, если функции ν, ζ всюду определены, как мы предполагали выше (см., однако, § 3.8). Сначала следует определить эквивалентные друг другу состояния автомата M (согласно определению ниже). После этого следует склеить все эквивалентные состояния в одно.

Определение. Состояния s_i и s_j называются *r-эквивалентными*, если для всякой входной строки $\mathrm{a} \in A^r$ длины r имеем

$$\zeta_r(s_i, \mathrm{a}) = \zeta_r(s_j, \mathrm{a}).$$

В этом случае будем писать $s_i E_r s_j$, или $(s_i, s_j) \in G(E_r)$. Если $s_i E_r s_j$ для всех r, будем говорить, что состояния s_i и s_j *эквивалентны* и писать $s_i E s_j$, или $(s_i, s_j) \in G(E)$.

Заметим прежде всего, что E_r и E — действительно отношения эквивалентности. Классы эквивалентности относительно E_1 являются множествами всех пар состояний, перерабатывающих каждый выходной символ в фиксированный выходной символ: $s_i E_1 s_j$ означает, что

$$\zeta(s_i, a) = \zeta(s_j, a)$$

для всех a. В этом можно убедиться непосредственно из таблицы. Для автомата, соответствующего рис. 3.10, имеем $s_0 E_1 s_2$.

Рассмотрим теперь график отношения эквивалентности (§ 2.1). Вместо $(s_i, s_j) \notin G(E_r)$ мы пишем $s_i E_r' s_j$ (как обычно, $G(E_r') \cup G(E_r) = S \times S$). Например, для рис. 3.10

$$G(E_1) = \{(s_0, s_2), (s_2, s_0), (s_0, s_0), (s_1, s_1), (s_2, s_2)\}$$

и

$$G(E_1') = \{(s_0, s_1), (s_1, s_0), (s_1, s_2), (s_2, s_1)\}.$$

M

Текущее состояние	ν		ζ	
	0	1	0	1
s_0	s_2	s_1	0	1
s_1	s_0	s_2	1	0
s_2	s_0	s_1	0	1

$\overline{M}$

Текущее состояние	$\bar{\nu}$		$\bar{\zeta}$	
	0	1	0	1
s_0	$\bar{s}_0$	$\bar{s}_1$	0	1
s_1	$\bar{s}_0$	$\bar{s}_0$	1	0

Рис. 3.10. Таблицы состояний эквивалентных автоматов.

Задача минимизации количества состояний в полностью описанном автомате сводится к определению попарно эквивалентных состояний и последующему их склеиванию. Оказывается, что эффективнее всего начать с выявления *неэквивалентных* состояний.

Чтобы показать это, определим две новые функции ν^* и ζ^*.

Определение. Положим ν^*: $S\times A^r \longrightarrow S$:

$$\nu^*(s^0, \mathbf{a}) = s^{r-1} = \nu(\ldots(\nu(\nu(s^0, a^0), a^1), \ldots), a^{r-1}).$$

Это означает, что $\nu^*(s_i, \mathbf{a})$ есть последнее состояние автомата, начавшего работу в состоянии s_i и считавшего входную строку $\mathbf{a}$ длины r.

Положим далее ζ^*: $S\times A^r \longrightarrow Z$:

$$\zeta^*(s^0, \mathbf{a}) = z^{r-1} = \zeta(\ldots\nu(\nu(s^0, a^0), a^1), \ldots, a^{r-1}).$$

Это означает, что $\zeta^*(s_i, \mathbf{a})$ есть последний символ выходной строки автомата, начавшего работу в состоянии s_i и считавшего ту же входную строку $\mathbf{a}$.

Так, для первого автомата M на рис. 3.9 при $r=3$ имеем $\nu^*(s_0, 101)=s_1$, $\nu^*(s_1, 101)=s_0$. Кроме того, $\zeta^*(s_0, 101)=0$ и $\zeta^*(s_1, 101)=1$.

Теорема 1. *Если* $s_iE's_j$, *то либо* $s_iE_1's_j$, *либо для подходящей строки* $\mathbf{a}=(a^0, \ldots, a^{r-1})$ *имеем* $(\nu^*(s_i, \mathbf{a})\, E_1'\nu^*(s_j, \mathbf{a}))$.

Доказательство. Утверждение $s_iE's_j$ означает, что $\zeta^*(s_i, \mathbf{a})\neq\zeta^*(s_j, \mathbf{a})$ для подходящей строки $\mathbf{a}=(a^0, \ldots, a^{r-1})$. При необходимости мы можем укоротить входную строку $\mathbf{a}$ так, чтобы выходные строки, отвечающие s_i и s_j, отличались только последними символами. Пусть это уже сделано. Если после этого $r=1$, то, очевидно, $s_iE_1's_j$. Если же $r>1$, то $s_iE's_j$, но $s_iE_ks_j$ при $k<r$. Таким образом, последние выходные символы автомата, считавшего a, различны, если он исходил из начальных состояний s_i, s_j соответственно.

Чтобы выходы отличались, $\zeta^*(s_i, \mathbf{a}) \neq \zeta^*(s_j, \mathbf{a})$, должно быть $\nu^*(s_i, \mathbf{a}) E'_1 \nu^*(s_j, \mathbf{a})$. Иначе последний входной символ a^{r-1} даст один и тот же выходной символ.

Это замечание приводит к более сильной теореме.

Теорема 2. *Если* $s_i E'_r s_j$, *но* $s_i E_k s_j$ *для всех* $k < r$, *то* $\nu(s_i, a_l) E'_{r-1} \nu(s_j, a_l)$ *для подходящего* $a_l \in A$.

Переформулировка этого утверждения такова: если $(s_i, s_j) \in G(E'_r) - G(E'_{r-1})$, то для подходящего $a_k \in A$ имеем $(\nu(s_i, a_k), \nu(s_j, a_k)) \in G(E'_{r-1}) - G(E'_{r-2})$.

Эта теорема утверждает, что состояния s_i, s_j, эквивалентные относительно всех входных последовательностей длины $r-1$, могут стать неэквивалентными относительно последовательностей длины r только в том случае, когда имеется символ a_k, переводящий s_i, s_j соответственно в состояния s_l, s_m, не эквивалентные относительно подходящей входной последовательности длины $r-1$. Это означает, что на r-м шаге достаточно исследовать состояния в $G(E_{r-1})$ и установить, найдется ли пара (s_i, s_j), переходящая в пару (s_l, s_m) со свойством $s_l E'_{r-1} s_m$. В этом случае $s_i E'_r s_j$.

Если мы уже определили $G(E'_1)$, то $G(E'_2)$ состоит из $G(E'_1)$ и таких упорядоченных пар (s_i, s_j), что для некоторого a_p имеем $(\nu(s_i, a_p), \nu(s_j, a_p)) \in G(E'_1)$. В общем случае нужно исследовать каждый раз только $G(E'_{r-1}) - G(E'_{r-2})$. Таким способом мы сумеем рекурсивно определить $G(E')$ и, наконец, $G(E)$ — дополнение к $G(E')$ в булевой алгебре подмножеств $S \times S$.

Доказательство. Доказательство теоремы проводится непосредственно. Если пара (s_k, s_l) лежит в $G(E'_{r-1})$, то она не лежит в $G(E'_r) - G(E'_{r-1})$. Значит, нужно рассмотреть лишь такие пары (s_k, s_l), что для некоторой строки $\mathbf{a} \in A^r$ имеем $\zeta^*(s_k, \mathbf{a}) \neq \zeta^*(s_l, \mathbf{a})$, а для всех строк $\mathbf{a} \in A^{r-1}$ имеем $\zeta^*(s_k, \mathbf{a}) = \zeta^*(s_l, \mathbf{a})$. Но это в точности те пары, которые переводятся в $G(E'_1)$ $(r-1)$-м входным символом a^{r-2} и, стало быть, в $G(E'_{r-1}) - G(E'_{r-2})$ некоторым символом $a^0 \in A$.

Лемма. *Если* $G(E'_r) - G(E'_{r-1}) = \varnothing$, *то* $G(E'_r) = G(E'_{r+k})$ *для всех* $k \geqslant 0$.

Действительно, дальнейшие шаги не добавят новых пар состояний, ибо, согласно теореме 2, дополнение $G(E'_{r+1}) - G(E'_r)$ состоит из тех пар, которые переводятся подходящим символом $a_i \in A$ в дополнение $G(E'_r) - G(E'_{r-1})$.

Пример 5. На рис. 3.11 показана таблица состояний некоторого автомата.

Автомат M

Текущее состояние	Следующее состояние		Выход	
	вход		вход	
	0	1	0	1
s_1	s_1	s_2	1	0
s_2	s_1	s_3	1	0
s_3	s_5	s_1	1	0
s_4	s_4	s_2	1	0
s_5	s_4	s_3	1	1

Рис. 3.11. Пример таблицы состояний.

Для автомата $s_1E_1's_5$, $s_2E_1's_5$, $s_3E_1's_5$ и $s_4E_1's_5$. Иными словами,

$$G(E_1') = \{(s_1, s_5),(s_2, s_5),(s_3, s_5),(s_4, s_5), (s_5, s_1),(s_5, s_2), (s_5, s_3), (s_5, s_4)\}.$$

Первое разбиение на классы эквивалентности:

$$C_1^1 = \{s_1, s_2, s_3, s_4\}, \quad C_2^1 = \{s_5\}.$$

Вход 0 переводит s_1 в s_1 и s_3 в s_5: $\nu(s_1, 0) = s_1$ и $\nu(s_3, 0) = s_5$. Поэтому вход 01 дает $\zeta^*(s_1, 01) = 0$ и $\zeta^*(s_3, 01) = 1$, откуда $s_1E_2's_3$ и $(s_1, s_3) \in G(E_2')$. Аналогично $s_2E_2's_3$ и $s_4E_2's_3$, так что

$$G(E_2') = G(E_1') \cup \{(s_1, s_3), (s_3, s_1), (s_2, s_3), (s_3, s_2), (s_3, s_4), (s_4, s_3)\}.$$

Далее, вход 1 переводит s_1 в s_2 и s_2 в s_3, т. е. $\nu(s_1, 1) = s_2$ и $\nu(s_2, 1) = s_3$. Поэтому $s_1E_3's_2$, ибо $\zeta^*(s_1, 101) = 0$ и $\zeta^*(s_2, 101) = 1$. Аналогично $s_2E_3's_4$, откуда следует, что

$$G(E_3') = G(E_2') \cup \{(s_1, s_2)(s_2, s_1)(s_2, s_4)(s_4, s_2)\}.$$

Таким образом, E_3 разбивает S на классы эквивалентности

$$C_1^3 = \{s_1, s_4\}, \quad C_2^3 = \{s_2\}, \quad C_3^3 = \{s_3\}, \quad C_4^3 = \{s_5\}.$$

Дальнейший перебор показывает, что $G(E_3') = G(E_4')$. Таким образом, $G(E_4) = G(E_3)$, $E = E_3$ и s_1Es_4, а остальные пары состояний неэквивалентны.

3.6. ПРОЦЕДУРА МИНИМИЗАЦИИ

Процедура минимизации, которую мы опишем, основана на рассмотрении отношений эквивалентности между упорядоченными парами. Рассмотрим таблицу состояний на рис. 3.12.

Начнем с вычисления $G(E_1)$ и $G(E_1')$. Введем разбиение π_1, которое разобьет состояния в таблице на два класса эквивалентности:

$$\pi_1\colon\ C_1^1=\{1, 3, 5, 7, 8\},\qquad C_2^1=\{2, 4, 6, 9\} \tag{9}$$

(здесь мы пишем 1 вместо s_1, 2 вместо s_2 и т. д.). Это разбиение определяет график $G(E_1)$ соответствующего отношения эквивалентности. Поскольку E_1 рефлексивно и симметрично, его

	Следующее состояние			Выходной символ		
	a_1	a_2	a_3	a_1	a_2	a_3
s_1	s_2	s_2	s_5	1	0	0
s_2	s_1	s_4	s_4	0	1	1
s_3	s_2	s_2	s_5	1	0	0
s_4	s_3	s_2	s_2	0	1	1
s_5	s_6	s_4	s_3	1	0	0
s_6	s_8	s_9	s_6	0	1	1
s_7	s_6	s_2	s_8	1	0	0
s_8	s_4	s_4	s_7	1	0	0
s_9	s_7	s_9	s_7	0	1	1

Рис. 3.12. Таблица состояний автомата, подлежащего минимизации.

график легко восстанавливается по множеству тех пар (s_i, s_j), для которых $\zeta(s_i, a_k)=\zeta(s_j, a_k)$ при всех $a_k\in A$. Обозначим через G_1 подмножество тех $(s_i, s_j)\in G(E_1)$, для которых $i<j$; и вообще через G_l обозначим множество упорядоченных пар $(s_i, s_j)\in G(E_l)$ со свойством $i<j$. Для разбиения (9) имеем

$$G_1=\{(1, 3)(1, 5)(1, 7)(1, 8)(3, 5)(3, 7)(3, 8)(5, 7)(5, 8)(7, 8)\\ (2, 4)(2, 6)(2, 9)(4, 6)(4, 9)(6, 9)\}.$$

$$G_1'=\{(1, 2)(1, 4)(1, 6)(1, 9)(2, 3)(3, 4)(3, 6)(3, 9)(2, 5)(4, 5)\\ (5, 6)(5, 9)(2, 7)(4, 7)(6, 7)(7, 9)(2, 8)(4, 8)(6, 8)(8, 9)\},$$

Так как C_1^1, C_2^1—классы эквивалентности относительно E_1, имеем $s_1E_1s_3$, $s_1E_1s_5$ и $s_1E_1's_2$, $s_1E_1's_4$ и т. д.

Множество G_2' состоит из элементов множества G_1' и еще пар (2, 9), (4, 9) и (6, 9). Например, a_3 переводит (2, 9) в (4, 7), а эта последняя пара принадлежит $G(E_1')$. Добавление этих пар к $G(E_1')$ определяет новое разбиение на классы эквивалентности:

$$\pi_2\colon\ C_1^2=\{1, 3, 5, 7, 8\},\qquad C_2^2=\{2, 4, 6\},\qquad C_3^2=\{9\}. \tag{10}$$

Определим теперь множество G_3'. Оно состоит из элементов множества G_2' и еще пар (2, 6) и (4, 6). Например, a_2 переводит (2, 6) в (4, 9), а эта последняя пара принадлежит G_2'. При разбиении π_3 имеем следующие классы эквивалентности:

$$C_1^3=\{1, 3, 5, 7, 8\}, \quad C_3^3=\{2, 4\}, \quad C_3^3=\{6\}, \quad C_4^3=\{9\}. \quad (11)$$

Множество G_4' состоит из элементов множества G_3' и из пар (1, 5), (1, 7), (3, 5), (3, 7), (8, 5), (8, 7). Поэтому разбиение π_4 состоит из следующих классов эквивалентности:

$$C_1^4=\{1, 3, 8\}, \quad C_2^4=\{2, 4\}, \quad C_3^4=\{5, 7\}, \quad C_4^4=\{6\}, \quad C_5^4=\{9\}.$$

Дальнейший перебор обнаруживает, что $\pi_5=\pi_4$ и $E_4=E$. Конструкция покрывающего автомата теперь несложна. Каждый класс эквивалентности последнего разбиения становится состоянием нового автомата. Например, C_1^4 обозначается через $\bar{s}_1$, C_2^4—через $\bar{s}_2$ и т. д. Получается автомат с пятью состояниями, покрывающий наш первоначальный автомат с девятью состояниями. Поскольку выходы для каждого начального состояния в фиксированном классе эквивалентности не зависят от этого состояния при односимвольных входах, таблица состояний нового автомата, в частности ее выходы, прямо считывается с таблицы состояний первоначального автомата. Чтобы построить функцию перехода в следующее состояние, выберем по состоянию s_i в каждом классе C_i^k, и если элемент $a \in A$ переводит s_i в некоторое состояние из C_m^k, положим $\mu(s_i, a)=\bar{s}_m$. Заметим, что это предписание однозначно: все $s \in C_i^k$ переходят в состояния из C_m^k после считывания $a \in A$.

На рис. 3.13 показан результат этой процедуры, примененной к автомату, представленному на рис. 3.12. Получен *минимальный автомат* для 3.12.

	Следующее состояние			Выходной символ		
	a_1	a_2	a_3	a_1	a_2	a_3
$\bar{s}_1$	$\bar{s}_2$	$\bar{s}_2$	$\bar{s}_3$	1	0	0
$\bar{s}_2$	$\bar{s}_1$	$\bar{s}_2$	$\bar{s}_2$	0	1	1
$\bar{s}_3$	$\bar{s}_4$	$\bar{s}_2$	$\bar{s}_1$	1	0	0
$\bar{s}_4$	$\bar{s}_1$	$\bar{s}_5$	$\bar{s}_4$	0	1	1
$\bar{s}_5$	$\bar{s}_3$	$\bar{s}_5$	$\bar{s}_3$	0	1	1

Рис. 3.13. Минимальный автомат, покрывающий автомат, представленный на рис. 3.12.

Практически не обязательно перечислять все пары из $G(E_t)$ и $G(E_t')$. На каждом шаге достаточно смотреть, переводит ли некоторый вход $a_l \in A$ пару (s_i, s_j) из C_l^k в разные классы эквивалентности C_m^k и C_n^k. Если да, то на следующем шаге s_i и s_j следует развести по разным классам.

Пример 7. Рассмотрим автомат с пятью состояниями, изображенный на рис. 3.14. Имеем:

$$s_0 E_1 s_2,\ s_1 E_1 s_3,\ s_1 E_1 s_4,\ s_3 E_1 s_4.$$

Это приводит к разбиению

$$\pi_1\colon\ C_1^1 = \{s_0, s_2\},\ C_2^1 = \{s_1, s_3, s_4\}.$$

Вход 1 переводит s_3 в s_0, т. е. $\nu(s_3, 1) = s_0$; кроме того, $\nu(s_4, 1) = s_4$. Однако $s_0 E_1' s_4$, так что $s_3 E_2' s_4$ (ибо $\zeta_2(s_3, 10) = 1$ и $\zeta_2(s_4, 10) = 0$).

Текущее состояние	Следующее состояние		Выход	
	0	1	0	1
s_0	s_1	s_2	1	0
s_1	s_4	s_2	0	0
s_2	s_3	s_0	1	0
s_3	s_4	s_0	0	0
s_4	s_4	s_4	0	0

Рис. 3.14. Полностью описанный автомат.

Следующее разбиение π_2 состоит из классов эквивалентности

$$C_1^2 = \{s_0, s_2\},\quad C_2^2 = \{s_1, s_3\},\quad C_3^2 = \{s_4\}. \tag{12}$$

Дальнейшего измельчения не происходит, ибо ν переводит каждый элемент класса эквивалентности в тот же класс. Итак, состояния s_0 и s_2 можно склеить в одно состояние $\bar{s}_0$, а состояния s_1 и s_3 — в состояние $\bar{s}_1$. Состояние s_4 получает первое обозначение $\bar{s}_2$. Новый минимальный автомат, покрывающий автомат на рис. 3.14, представлен рис. 3.15.

Следующее состояние		Выход	
0	1	0	1
$\bar{s}_0$	$\bar{s}_1$ $\bar{s}_0$	1	0
$\bar{s}_1$	$\bar{s}_2$ $\bar{s}_0$	0	0
$\bar{s}_2$	$\bar{s}_2$ $\bar{s}_2$	0	0

Рис. 3.15. Минимальный автомат.

УПРАЖНЕНИЯ Б

1. Минимизировать число состояний следующего автомата:

	Следующее состояние		Выход	
	a_0	a_1	a_0	a_1
1	2	2	1	0
2	3	3	1	0
3	4	4	1	0
4	4	4	0	1
5	5	6	1	1
6	6	5	1	1

2. Минимизировать число состояний следующего автомата:

	Следующее состояние		Выход	
	a_0	a_1	a_0	a_1
1	9	1	1	1
2	2	2	1	0
3	7	5	0	1
4	2	2	1	0
5	2	2	1	0
6	3	9	1	0
7	6	8	1	0
8	9	9	1	0
9	4	6	0	0

3. Пусть M и $\overline{M}$ — автоматы с p и q состояниями соответственно и общим входным алфавитом A. Показать, что если для всех входных строк длины $r \leqslant p+q-1$ выполняются $\zeta_r(s_i, \mathrm{a}) = \overline{\zeta}_r(\overline{s}_j, \mathrm{a})$, то состояние s_i автомата M эквивалентно состоянию $\overline{s}_j$ автомата $\overline{M}$ [1]).

4. Привести таблицы состояний двух автоматов M, $\overline{M}$, которые эквивалентны, но не изоморфны.

---

[1]) Авторы не приводили определения эквивалентности состояний разных автоматов; по-видимому, читатель должен изобрести его. — *Прим. перев.*

5. Нарисовать диаграммы состояний следующих автоматов:

	ν		ζ	
	0	1	0	1
1	1	2	0	0
2	2	3	0	0
3	3	4	0	0
4	4	1	0	1

	ν		ζ	
	0	1	0	1
a	b	c	0	0
b	b	c	0	0
c	b	c	0	1

6. Пусть автоматы M, $\overline{M}$ с одним входиым символом a заданы следующими таблицами:

M	ν	ζ
1	1	0
2	1	1
3	3	1
4	3	1

$\overline{M}$	ν	ζ
$1'$	$2'$	0
$2'$	$2'$	0
$3'$	$1'$	1
$4'$	$4'$	1

Найти все пары (p, q) эквивалентных между собой состояний автоматов M и $\overline{M}$ соответственно [1]). Эквивалентны ли сами автоматы M и $\overline{M}$?

7. Указать морфизм автомата M в автомат $\overline{M}$:

M	ν		ζ	
	0	1	0	1
a	b	c	0	1
b	a	c	0	1
c	c	a	1	0

$\overline{M}$	$\overline{\nu}$		$\overline{\zeta}$	
	0	1	0	1
1	1	2	0	1
2	2	1	1	0

[1]) То же замечание, что и к упр. 3. — *Прим. перев.*

8. Рассмотрим автомат $M=[A, S, \nu, Z, \zeta]$ с $A=\{0, 1\}$, $Z=\{0, 1\}$, $S=\{1, 2, 3, 4, 5\}$:

M	ν		ζ	
	0	1	0	1
1	1	2	0	1
2	1	3	0	1
3	5	1	0	1
4	4	2	0	1
5	4	3	1	1

а) Найти автомат $\overline{M}$ с минимальным числом состояний, который перерабатывает все входы длины 2 в те же выходы, что и автомат M.

б) Эквивалентен ли этот автомат $\overline{M}$ автомату M?

9. а) Доказать, что если автомат M_1 является эпиморфным образом автомата M_2, а M_2 — эпиморфным образом M_3, то M_1 есть эпиморфный образ автомата M_3.

б) Пусть $\overline{M}$ является эпиморфным образом M. Доказать, что $\overline{M}$ покрывает ζ (а) M. (*Указание*: показать индукцией по длине входной строки а, что $\zeta(s, \mathbf{a})=\zeta(f(s), \mathbf{a})\,\overline{\zeta}(f(s), \mathbf{a})$ для всех $s\in S$.)

*в) Построить такие автоматы M, $\overline{M}$, что M покрывает $\overline{M}$, но $\overline{M}$ не является эпиморфным образом автомата M.

10. Построить минимальный автомат $\overline{M}$ для автомата

	ν		ζ	
	a	b	a	b
1	1	6	0	0
2	1	4	0	0
3	2	5	1	0
4	5	8	1	1
5	1	3	1	0
6	8	5	1	1
7	6	3	1	1
8	2	5	1	0

*3.7. МАШИНЫ ТЬЮРИНГА

Исторически понятие конечного автомата развилось из близкого понятия, введенного в 1936 году логиком Тьюрингом. Тьюринг рассматривал гипотетическую «машину», имеющую конечное

множество S внутренних *состояний* и одну бесконечно длинную ленту, разделенную на ячейки, которую машина могла передвигать на одну ячейку вправо или влево за такт. В каждой ячейке машина может записывать символ из конечного алфавита A. Первоначально лента должна быть пустой, кроме конечного числа ячеек, заполненных заранее. (Эти заранее заполненные ячейки можно представлять себе как программу запуска машины.)

Основная разница между машиной Тьюринга и конечным автоматом состоит в том, что: (i) лента машины Тьюринга бесконечна; (ii) машина Тьюринга может двигаться вдоль ленты (или смещать ленту) в любом направлении. Это придает машине бесконечную память, которой можно пользоваться в ходе вычислений. Сверх того, каждую ячейку можно просматривать многократно. Формальное определение машины Тьюринга существует в нескольких вариантах. Вот один из них.

Определение. *Машиной Тьюринга* называется пятерка объектов $[A, S, \nu, \zeta, \delta]$ следующего типа. A есть конечный алфавит символов, которые могут быть записаны в ячейках и являются одновременно входными и выходными: $A=\{a_0, a_1, \ldots, a_n\}$. S есть конечное множество внутренних состояний, $S=\{s_0, s_1, \ldots, s_r\}$; ν—функция из $S\times A$ в S; ζ—функция из $S\times A$ в A; δ—функция из $S\times R$ в множество {П, Л, ОСТАНОВ}, интуитивный смысл которого станет ясен из дальнейшего.

Машина Тьюринга работает следующим образом. Она начинает работу, находясь в начальном состоянии s^0. После считывания первого символа она переходит в новое внутреннее состояние, определяемое функцией ν, записывает в ячейке символ, являющийся значением функции ζ, перемещает ленту направо (П), налево (Л), или остается на месте и прекращает работу (ОСТАНОВ) в зависимости от значения функции δ.

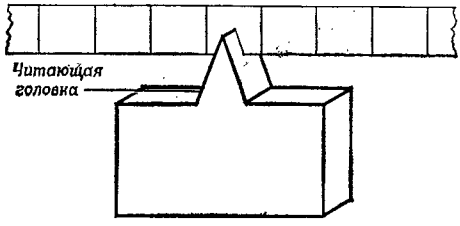


Рис. 3.16.

На рис. 3.16 схематически изображена лента машины Тьюринга и считывающая-записывающая головка.

Еще раз укажем, что работа машины состоит в повторении следующего цикла: считывание символа из ячейки, впечатывание нового символа в эту ячейку, выбор которого определяет функция ζ (может оказаться, что это тот же символ), сдвиг ленты

налево или направо либо остановка. Лента бесконечна в обоих направлениях, однако вначале (и, значит, после любого такта) заполнено лишь конечное число ячеек.

Пример 8. Машина Тьюринга, описанная ниже, считывает входную последовательность нулей и единиц, выпечатывает Ч, если число единиц четное, и Н, если нечетно. Строке из нулей и единиц предшествуют и последуют пустые ячейки, обозначаемые #. Символы Ч или Н печатаются в первой пустой ячейке вслед за входной строкой. Алфавит, таким образом, имеет вид

$$A = \{\#,\ 0,\ 1,\ \text{Ч},\ \text{Н}\}.$$

Внутренние состояния: $S = \{s_0, s_1, s_2\}$; s_0 — начальное состояние. Машина останавливается по сигналу ОСТАНОВ.

ν:	ζ:	δ:
$(s_0, 0) \mapsto s_1$	$(s_0, 0) \mapsto 0$	$(s_0, 0) \mapsto$ Л
$(s_0, 1) \mapsto s_2$	$(s_0, 1) \mapsto 1$	$(s_0, 1) \mapsto$ Л
$(s_1, 0) \mapsto s_1$	$(s_1, 0) \mapsto 0$	$(s_1, 0) \mapsto$ Л
$(s_1, 1) \mapsto s_2$	$(s_1, 1) \mapsto 1$	$(s_1, 1) \mapsto$ Л
$(s_2, 0) \mapsto s_2$	$(s_2, 0) \mapsto 0$	$(s_2, 0) \mapsto$ Л
$(s_2, 1) \mapsto s_1$	$(s_2, 1) \mapsto 1$	$(s_2, 1) \mapsto$ Л
$(s_0, \#) \mapsto s_0$	$(s_0, \#) \mapsto \#$	$(s_0, \#) \mapsto$ Л
$(s_1, \#) \mapsto s_1$	$(s_1, \#) \mapsto$ Ч	$(s_1, \#) \mapsto$ ОСТАНОВ
$(s_2, \#) \mapsto s_2$	$(s_2, \#) \mapsto$ Н	$(s_2, \#) \mapsto$ ОСТАНОВ

Удобнее задавать функции ν, ζ, δ, пользуясь обозначениями Тьюринга. В этом варианте машина Тьюринга задается конечным множеством пятерок $[s_i, a_j, s_r, z_l, t_n]$. В каждой такой пятерке

s_i — текущее состояние машины;
a_j — символ, считываемый из ячейки;
s_r — следующее состояние машины, $s_r = \nu(s_i, a_j)$;
z_l — символ, печатаемый в ячейке, $z_l = \zeta(s_i, a_j)$;
t_n — одна из команд П, Л, ОСТАНОВ.

В этих обозначениях описанная выше машина задается так:

s_0	#	s_0	#	Л
s_0	0	s_1	0	Л
s_0	1	s_2	1	Л
s_1	0	s_1	0	Л
s_1	1	s_2	1	Л
s_2	0	s_2	0	Л
s_2	1	s_1	1	Л
s_1	#	s_1	Ч	ОСТАНОВ
s_2	#	s_2	Н	ОСТАНОВ

Следующий несложный результат показывает, что машины Тьюринга умеют делать все, что умеют конечные автоматы.

Теорема 3. *Пусть* $M=[A, S, Z, \nu, \zeta]$ *—некоторый конечный автомат. Положим*

$$\overline{A}=A\cup Z\cup\{\Lambda\}$$

(Λ *—символ пустой ячейки*) *и*

$$\begin{array}{ll}\bar{\nu}(s_i, a_k)=\nu(s_i, a_k) & \bar{\nu}(s_i, \Lambda)=s_i\\ \bar{\zeta}(s_i, a_k)=\zeta(s_i, a_k) & \bar{\zeta}(s_i, \Lambda)=\Lambda\\ \delta(s_i, a_k)=\Pi & \delta(s_i, \Lambda)=\text{ОСТАНОВ}\end{array}$$

для всех $(s_i, a_k)\in S\times A$. *Тогда машина Тьюринга* $T=[\overline{A}, S, \bar{\nu}, \bar{\zeta}, \delta]$ *ставит в соответствие входным строкам те же выходные строки, что и* M.

Доказательство. Любую входную строку $\mathbf{a}=a^0,a^1\ldots a^r$ автомата M можно записать на ленте T так, чтобы a^j был записан в j-й ячейке. Описанная выше машина T напечатает $z^j=\zeta(s^j, a^j)$ в j-й ячейке, перейдет в состояние $s^{j+1}=\nu(s^j, a^j)$ и передвинется в $(j+1)$-ю ячейку. Добравшись до $(r+1)$-й ячейки, она остановится.

Замечания. Если входной алфавит A автомата M содержит пробел, он должен обозначаться специальным символом в алфавите T, чтобы избежать неоднозначности. Значения функций $\nu(s_i, z_l)$, $\zeta(s_i, z_l)$ мы не определили, ибо для нас они безразличны. Неполностью описанные автоматы с безразличными состояниями будут обсуждены в следующем параграфе.

Пример 9. Простая машина Тьюринга, обладающая бóльшими возможностями, чем любой конечный автомат, определяет по любой входной строке вида ...##111...1100...00##... (где # — пустые ячейки), одинаково ли число нулей и единиц в такой строке. Алфавит ее состоит из символов 0, 1, Ч, Н, # (в частности, машина может «записывать» пустые ячейки, т. е. стирать имеющуюся запись). Внутренние состояния ее:

$$S=\{s_0, s_1, s_2, s_3, s_4, s_5, s_6\}$$

Пятерки, описывающие машину по Тьюрингу:

s_0	#	s_0	#	Л
s_0	0	s_1	#	Л
s_1	0	s_1	0	Л
s_1	1	s_1	1	Л
s_1	#	s_2	#	П
s_2	1	s_3	#	П
s_3	1	s_3	1	П
s_3	0	s_4	0	П
s_4	#	s_0	#	Л
s_3	#	s_5	0	ОСТАНОВ
s_2	0	s_5	0	Л
s_5	#	s_5	Ч	ОСТАНОВ
s_4	1	s_4	1	П
s_4	0	s_4	0	П
s_0	1	s_5	1	П
s_2	#	s_5	Ч	ОСТАНОВ

По окончании работы машина справа от входной строки печатает Н, если число нулей равно числу единиц, Ч—в противном случае, и затем останавливается.

Нетрудно описать машины Тьюринга, вычисляющие разнообразные функции от чисел, подаваемых на вход. Стандартным представлением неотрицательного числа n в машине Тьюринга является последовательность из $n+1$ единиц, стоящих подряд. Два таких числа отделяются нулем. Так, строка ...# # 111011 # # ... представляет упорядоченную пару (2,1). Строка ... # # 11110 1101011 # # ... представляет последовательность (3, 1, 0, 1).

Пример 10. Следующая машина Тьюринга складывает два неотрицательных целых числа, поданных на вход:

s_0	#	s_0	#	Л
s_0	1	s_1	1	Л
s_1	1	s_1	1	Л
s_1	0	s_2	1	Л
s_2	1	s_2	1	Л
s_2	#	s_3	#	П
s_3	1	s_4	#	П
s_4	1	s_5	#	ОСТАНОВ

Она превращает две последовательности единиц, разделенные нулем, в последовательность единиц, представляющую число, равное сумме чисел на входе. Можно описать машину Тьюринга, способную сложить три целых числа, четыре целых числа или даже любое не ограниченное наперед количество целых чисел, но эта задача несколько сложнее.

3.8. НЕПОЛНОСТЬЮ ОПИСАННЫЕ АВТОМАТЫ

В предшествующем изложении мы занимались только полностью описанными автоматами. Практически функции ν и ζ обычно бывают лишь частично определены. Системы обычно проектируются по частям, и некоторые входные строки либо вообще не встречаются, либо встречаются в ситуациях, в которых выход нас не интересует. Это приводит к тому, что некоторые позиции в таблицах состояний (или ребра в диаграммах состояний) отсутствуют. Такие позиции называются *безразличными*, в знак того, что нас не заботит их содержимое. Безразличные позиции в наших таблицах будут обозначаться прочерком.

Текущее состояние	Входы			
	следующее состояние		выход	
	0	1	0	1
s_0	s_1	s_2	0	1
s_1	—	s_1	1	0
s_2	s_0	s_1	—	1

Рис. 3.17. Неполностью описанный автомат.

На рис. 3.17 показана таблица, в которой нам безразлично следующее состояние автомата, исходившего из состояния s_1 и считавшего входной символ 0, а также безразличен выход автомата, находящегося в состоянии s_2 и считавшего символ 0.

Представляет интерес проследить, как разрабатывалась процедура минимизации числа состояний неполностью описанного автомата. Сначала казалось, что следует как-нибудь заполнить безразличные позиции в таблице и затем минимизировать получившийся полностью описанный автомат с помощью ранее предложенной процедуры. Предполагалось, что выбор наименьшего из получившихся минимальных автоматов доставит минимальный автомат в более широком смысле слова. В принципе это свело

бы проблему к задаче сокращения такой прямоугольной процедуры минимизации, во что и было вложено немало усилий.

Однако для многих неполностью описанных автоматов процедура такого вида не приводит к минимизации. Действительно, спецификация всех безразличных позиций обычно уменьшает возможность минимизации. Дело в том, что если безразличная позиция действительно безразлична, то способ ее заполнения может меняться в зависимости от входа, что и позволяет дополнительную минимизацию. Например, если для автомата на рис. 3.17 мы допустим, чтобы выход автомата в состоянии s_2, считавшего 0, был иногда 0, а иногда 1, то автомат можно свести к двум состояниям. Если же настаивать, чтобы выход был всегда одним и тем же, редукция невозможна.

Для описания способа минимизации неполностью описанных автоматов нам потребуются следующие определения.

Определение. Входная последовательность $\mathbf{a} = a^0, \ldots, a^{r-1}$ называется *допустимой* для автомата в начальном состоянии s_j, если функция ν перехода в следующее состояние определена для всех элементов последовательности, кроме, возможно, последнего.

Таким образом, начальное состояние s_j и допустимая входная строка $a^0, \ldots, a^{r-1}$ однозначно определяют строку последовательных внутренних состояний, за исключением того, что последнее состояние может быть неопределенным (безразличным).

Определение. Выходная строка $\mathbf{z}$ *покрывает* выходную строку $\bar{\mathbf{z}}$ (в которой могут быть неопределенные символы), если всякий определенный символ $\bar{z}^j$ в $\bar{\mathbf{z}}$ равен соответствующему символу z^j в $\mathbf{z}$. Например, строка $\mathbf{z} = z^0, z^1, z^2 = 0, 1, 1$ покрывает строку $\bar{\mathbf{z}} = \bar{z}^0, \bar{z}^1, \bar{z}^2 = —, 1, 1$; строка 0, 1, 0 покрывает —, 1, 0; строка 1, 1, 0 покрывает 1, 1, 0 и —, 1, 0. Если $\mathbf{z}$ покрывает $\bar{\mathbf{z}}$, мы пишем $\mathbf{z} \geqslant \bar{\mathbf{z}}$.

Определение. Если $\zeta_r(s_k, \mathbf{a}) \geqslant \bar{\zeta}_r(\bar{s}_j, \mathbf{a})$ для всех $\mathbf{a}$, допустимых для $\bar{s}_j$, мы пишем $s_k \geqslant \bar{s}_j$ и говорим, что s_k *покрывает* $\bar{s}_j$.

Определение. Автомат M *покрывает* автомат $\bar{M}$, если для каждого состояния $\bar{s}_j$ автомата $\bar{M}$ существует такое состояние s_k автомата M, что $\zeta_r(s_k, \mathbf{a}) \geqslant \bar{\zeta}_r(\bar{s}_j, \mathbf{a})$ для всех $\mathbf{a}$, допустимых для $\bar{s}_j$. Иными словами, каждое состояние автомата $\bar{M}$ покрывается некоторым состоянием M. В этом случае мы пишем $M \geqslant \bar{M}$. Очевидно, в случае $M = \bar{M}$ состояние s_k покрывает s_j, если $\zeta_r(s_k, \mathbf{a}) \geqslant \zeta_r(s_j, \mathbf{a})$ для всех $\mathbf{a}$, допустимых для s_j, и мы тогда пишем $s_k \geqslant s_j$.

Рассмотрим теперь автоматы на рис. 3.18.

Текущее состояние	Следующее состояние		Выход	
	0	1	0	1
s_0	s_0	s_1	0	1
s_1	s_1	s_2	—	1
s_2	s_0	s_2	1	1

Автомат M

Текущее состояние	Следующее состояние		Выход	
	0	1	0	1
$\overline{s}_0$	$\overline{s}_0$	$\overline{s}_1$	0	1
$\overline{s}_1$	$\overline{s}_0$	$\overline{s}_1$	1	1

Автомат $\overline{M}$

Рис. 3.18. Неполностью описаиные автоматы.

Автомат $\overline{M}$ покрывает M; состояние $\overline{s}_0$ покрывает s_0 и s_1, а состояние $\overline{s}_1$ покрывает s_2.

Поучительно рассмотреть выходы автоматов M и $\overline{M}$. Вход 0, 1, 0, 1, 0, считанный M в начальном состоянии s_1, перерабатывается в —, 1, 1, 1, —, а тот же выход, считанный $\overline{M}$ в начальном состоянии s_0, перерабатывается в 0, 1, 1, 1, 1. Заметим, что первый безразличный символ на выходе M отвечает 0 в $\overline{M}$, а второй соответствует 1 в $\overline{M}$. В ходе работы безразличная позиция в таблице для M может заполняться как нулем, так и единицей.

Поскольку некоторые из символов выходной строки могут быть безразличными, нам придется рассматривать еще одно отношение между выходными строками, которое мы назовем совместимостью.

Определение. Выходная строка z *совместима* с выходной строкой $\overline{z}$, если в каждой позиции, где символы обеих строк определены, они совпадают. Мы будем писать $z\gamma\overline{z}$, если z совместима с $\overline{z}$.

Примеры. Строка 0, 1, —, 1 совместима с 0, 1, 1, 1; 0, —, 1, — совместима с —, 0, —, 1; но 0, 1, 0, — несовместима с 0, 1, 1, —.

Заметим, что γ не является отношением эквивалентности. Оно рефлексно и симметрично, но не транзитивно.

*3.9. ОТНОШЕНИЯ МЕЖДУ СОСТОЯНИЯМИ И ПРОЦЕДУРА МИНИМИЗАЦИИ

Рассмотрения предыдущего параграфа приводят нас к необходимости более детально изучить несколько отношений между внутренними состояниями неполностью описанного автомата.

Определение. Состояние s_i называется *совместимым по выходу* с состоянием s_j, если $\zeta(s_i, a) \underset{1}{\gamma} \zeta(s_j, a)$ для всех $a \in A$. В этом случае мы пишем $s_i \underset{1}{\sigma} s_j$. Если состояния *несовместимы по выходу*, мы пишем $s_k \underset{1}{\sigma'} s_l$.

Таким образом, два состояния *совместимы по выходу*, если для каждой входной строки длины 1 их выходы совпадают, когда они оба определены. Это отношение можно непосредственно вычислить по таблице состояний.

Рассмотрим, например, рис. 3.19. Для этого автомата $s_0 \underset{1}{\sigma} s_1$ и $s_1 \underset{1}{\sigma} s_2$, но $s_0 \underset{1}{\sigma'} s_2$ (таким образом, отношение σ не транзитивно, хотя оно рефлексивно и симметрично).

	0	1	0	1
s_0	s_0	s_2	0	1
s_1	s_1	s_2	—	1
s_2	s_2	s_0	1	1

Рис. 3.19. Пример неполностью описанного автомата.

Определение. Состояния s_i и s_j называются *совместимыми*, если для всех $\mathbf{a} \in A^r$, допустимых как для s_i, так и для s_j, имеем $\zeta_r(s_i, \mathbf{a}) \gamma \zeta_r(s_j, \mathbf{a})$. В этом случае мы пишем $s_i \sigma s_j$. Если s_k и s_l совместимы не для всех $\mathbf{a}$, мы пишем $s_k \sigma' s_l$. Если состояния s_i и s_j совместимы для всех строк фиксированной длины k, т. е. $\zeta_k(s_i, \mathbf{a}) \gamma \zeta_k(s_j, \mathbf{a})$, где $\mathbf{a}$ — вход длины k, то состояния s_i и s_j называем *k-совместимыми* и пишем $s_i \underset{k}{\sigma} s_j$.

Пусть M, начав работать в состоянии s_i или s_j, для любой входной строки $\mathbf{a}$ дает одинаковые выходы в тех позициях, которые определены. Тогда состояния s_i, s_j совместимы ($s_i \sigma s_j$), и можно *склеить* их в одно состояние. Эта операция не изменит выходы в определенных позициях и даст безразличный выход в тех позициях, которые ранее давали безразличные выходы хотя бы для одного из двух состояний. Таким образом, новый выход будет покрывать оба старых. Обозначим новое состояние через s. Тогда $\zeta_r(s, \mathbf{a}) \geqslant \zeta_r(s_i, \mathbf{a})$ для всех входов $\mathbf{a} \in A^r$, допустимых для s_i, и $\zeta_r(s, \mathbf{a}) \geqslant \zeta_r(s_j, \mathbf{a})$ для всех входов $\mathbf{a} \in A^r$, допустимых для s_j.

Нужен хороший способ проверки совместимости, ибо это отношение указывает возможные способы комбинировать состояния. Однако оно не является отношением эквивалентности и *не* указывает точного способа склеивать состояния.

Рассмотрим рис. 3.20.

Текущее состояние	Следующее состояние				Выход			
	a_1	a_2	a_3	a_4	a_1	a_2	a_3	a_4
s_1	s_2	—	s_3	s_2	0	—	—	—
s_2	s_3	s_5	s_2	—	0	1	0	—
s_3	s_3	s_4	—	s_5	0	1	—	0
s_4	—	s_1	s_2	—	—	1	—	—
s_5	—	—	s_1	—	—	—	1	—

(а) Таблица состояний

$$G\left(\overset{1}{\sigma}\right) = (1, 2)\,(1, 3)\,(1, 4)\,(1, 5)\,(2, 3)\,(2, 4)\,(3, 4)\,(3, 5)\,(4, 5)$$

$$G\left(\overset{1}{\sigma'}\right) = (2, 5)$$

$$G\left(\overset{2}{\sigma'}\right) = (2, 5)\,(1, 3)$$

$$G\left(\overset{3}{\sigma'}\right) = (2, 5)\,(1, 3)\,(1, 5)$$

$$G\left(\overset{4}{\sigma'}\right) = (2, 5)\,(1, 3)\,(1, 5)\,(2, 4)$$

$$G(\sigma) = (1, 2)\,(1, 4)\,(2, 3)\,(3, 4)\,(3, 5)\,(4, 5)$$

(б) Графики отношений

	a_1	a_2	a_3	a_4
(1,2)	(2,3)	—	(2,3)	—
(1,3)	(2,3)	—	—	(2,5)
(1,4)	—	—	(2,3)	—
(1,5)	—	—	(1,3)	—
(2,3)	—	(4,5)	—	—
(2,4)	—	(1,5)	—	—
(3,4)	—	(1,4)	—	—
(3,5)	—	—	—	—
(4,5)	—	—	(1,2)	—

(в) Таблица совместимости

Рис. 3.20. Техника определения совместимых состояний.

Мы хотим определить график отношения совместимости σ для неполностью описанного автомата (рис. 3.20,*а*). Сначала следует определить $G\left(\overset{1}{\sigma'}\right)$, ибо, очевидно, этот график является

подмножеством $G\binom{2}{\sigma'}$. График $G\binom{1}{\sigma'}$ содержит только пару (2, 5), как показано на рис. 3.20,*б*.

В таблице (рис. 3.20,*в*) строки перенумерованы элементами $G\binom{1}{\sigma}$, а столбцы исходными символами. На пересечении столбца a_i и строки (s_l, s_m) указана пара состояний, в которые a_i переводит пару (s_l, s_m). Например, a_1 переводит s_1 в s_2 и s_2 в s_3: поэтому на пересечении (1, 2) и a_1 стоит (2, 3). Если одно или оба из следующих состояний безразличны или если исходная пара переходит в одно и то же состояние, мы ставим в соответствующей позиции прочерк.

Предположим, что некоторый элемент множества $G\binom{1}{\sigma}$ переводится в элемент множества $G\binom{1}{\sigma'}$ некоторой входной строкой. Тогда первоначальный элемент лежит в $G(\sigma')$. Действительно, если после применения этого входа для получившейся пары из $G\binom{1}{\sigma'}$ использовать дающие разные выходы, то получим разные выходы и для первоначальной пары. Например, a_4 переводит (1, 3) в (2, 5), а (2, 5) лежит в $G\binom{1}{\sigma'}$, ибо эти состояния дают разные выходы при входе a_3. Поэтому состояния (1, 3) дадут разные выходы при входе a_4, a_3.

Представим сказанное схематически:

	a_4	a_3
Состояния	1	2
	3	5
Выходы	—	0
	0	1

Мы далее составляем список элементов множества $G\binom{2}{\sigma'}$, начиная с элементов $G\binom{1}{\sigma'}$. (Заметим, что выписаны лишь пары (i, j) с $i < j$. Поскольку наше отношение рефлексивно и симметрично, пары (i, i) и (j, i) с $j > i$ можно опустить.) Добавляются всевозможные пары, которые некоторым входным символом a_i переводятся в элементы $G\binom{1}{\sigma'}$. Затем добавляются пары, которые некоторым символом a_j переводятся в только что полученные пары. Действительно, эти пары различаются подходящим входом длины 3. В нашей таблице на этом шаге добавится пара (1, 5). (Заметим, что строка a_3, a_4, a_3 дает разные выходы для (1, 5).) Этот процесс продолжается, пока продолжают добавляться новые пары. В нашей таблице на следующем шаге добавится (2, 4), ибо a_2 переводит (2, 4) в (1, 5).

Когда новые пары перестанут возникать, процедура кончается. Для автомата с n состояниями это произойдет не позже чем через $n-1$ шагов, ибо к этому времени возможные новые пары исчерпаются. Последнее полученное отношение будет σ'.

Заметим, что выписывать пары, в которых хотя бы одно из состояний безразлично или оба состояния совпадают, нет смысла. Последующие входы не смогут «расщепить» такие пары и их выходы.

Заметим также, что в силу антирефлексивности и симметричности $\overset{i}{\sigma'}$ достаточно выписывать пары лишь в порядке возрастания индексов, а (s_i, s_i) не выписывать вовсе. Сказанное относится и к σ'.

В примере на рис. 3.20 график $G(\sigma')$ содержит (2, 5), (1, 3), (1, 5) и (2, 4), а $G(\sigma)$ — остальные пары.

Расширим теперь определение нашего отношения, учитывая, что склеиваться могут не только пары состояний.

Определение. *Совместимым классом* C_k называется такое множество внутренних состояний $s_i, \ldots, s_m$, что $s_i \sigma s_j$ для всех $s_i, s_j \in C_k$. *Максимальным совместимым классом* C_l называется совместимый класс, не содержащийся в качестве собственного подмножества в другом совместимом классе.

Таким образом, максимальный совместимый класс есть один из наибольших наборов попарно совместимых состояний. Так, если $s_i \sigma s_j$, $s_j \sigma s_k$ и $s_i \sigma s_k$, то (s_i, s_j, s_k) образуют совместимый класс. Однако если $s_i \sigma s_k$, $s_j \sigma s_k$, но $s_j \sigma' s_i$, то в этом подмножестве есть два максимальных совместимых класса: (s_i, s_k) и (s_k, s_j).

Полное множество максимальных совместимых классов есть список самых больших подмножеств состояний, каждое из которых можно склеить в одно состояние. Для автомата, изображенного на рис. 3.20, максимальные совместимые классы таковы: (1, 2), (1, 4), (2, 3) и (3, 4, 5).

Определение. Некоторое множество совместимых классов называется *согласованным*, если для любого класса C_k из этого множества и любых его элементов s_i, s_j внутренние состояния $\nu(s_i, a_k)$, $\nu(s_j, a_k)$ принадлежат подходящему совместимому классу C_l для любого символа a_k.

Определение. Некоторое множество совместимых классов называется *замкнутым*, если всякое внутреннее состояние автомата принадлежит хотя бы одному из этих классов.

Теорема 4. *Пусть задано замкнутое согласованное множество совместимых классов для автомата M. Тогда существует автомат $\overline{M}$, покрывающий автомат M, состояния которого полу-*

чаются склеиванием всех состояний M, содержащихся в одном совместимом классе из данного множества.

Пары s_i, s_j (или множества попарно совместимых состояний) можно склеить в единое состояние. Новая диаграмма состояний получается из старой, если заменить s_j во всех позициях, где оно встречается, на s_i, вписать вместо $\zeta(s_j, a)$ произвольные значения в тех позициях, где они определены для s_j, но не для s_i, и вычеркнуть строку s_j из таблицы.

Более общо, пусть задано замкнутое согласованное множество совместимых классов. Склеим все состояния класса C_i в одно новое состояние $\bar{s}_j$. Определим $\bar{\zeta}(\bar{s}_j, a_k)$ как $\zeta(s_l, a_k)$, если эти значения определены для всех $s_l \in C_i$. Определим $\nu(\bar{s}_j, a_k)$ как $\bar{s}_m$, если $\nu(s_l, a_k) \in C_m$ для всех $s_l \in C_i$. Состояния нового автомата будут находиться в биективном соответствии с классами исходного множества. Каждое новое состояние покрывает все состояния соответствующего класса.

Согласованность необходима для возможности такой склейки. Действительно, если бы для некоторого символа a и состояний s_i, s_j из одного C_k состояния $\nu(s_i, a)$, $\nu(s_j, a)$ принадлежали разным классам C_l, C_m, то нельзя было бы склеить s_i и s_j.

Теорема 5. *Рассмотрим исходное замкнутое согласованное множество, содержащее наименьшее число совместимых классов. Тогда автомат $\overline{M}$, покрывающий M и полученный склеиванием всех состояний каждого класса в одно, будет минимальным.*

Действительно, если автомат $\overline{M}$ покрывает автомат M, то каждое состояние s_j автомата M покрывается подходящим состоянием $\bar{s}_k$ автомата $\overline{M}$. Весь набор состояний, покрываемых $\bar{s}_k$, содержится в полном списке совместимых классов. Все получающиеся таким образом совместимые классы образуют согласованное замкнутое множество. Если $\overline{M}$ минимален, то указанное множество содержит наименьшее возможное число элементов. Таких множеств совместимых классов может быть несколько. Все они приводят к минимальным автоматам, которые не обязаны быть изоморфными.

Рассмотрим автомат рис. 3.20. Одно из возможных предложений состоит в разбиении на классы эквивалентности $C_1 = \{1, 2\}$, $C_2 = \{3, 4, 5\}$, которое привело бы к автомату с двумя состояниями. Однако, $\nu(s_1, a_1) = s_2$ и $\nu(s_2, a_1) = s_3$. Поэтому s_2, s_3 должны были бы лежать в одном классе, а это не так. Значит, это предложение не годится: наше разбиение несогласованно.

Никакая другая пара совместимых классов не покрывает все множество состояний. Поэтому следует рассмотреть разбиения по меньшей мере на три класса. Такое согласованное разбиение

из трех классов существует:

$$C_1 = \{s_1,\ s_2\}, \quad C_2 = \{s_2,\ s_3\}, \quad C_3 = \{s_4,\ s_5\}.$$

На рис. 3.21 показан соответствующий минимальный автомат.

Текущее состояние	Следующее состояние				Выход			
	a_1	a_2	a_3	a_4	a_1	a_2	a_3	a_4
$\bar{s}_1$	s_2	s_3	s_2	s_1	0	1	0	—
$\bar{s}_2$	s_2	s_3	s_1	s_3	0	1	0	0
$\bar{s}_3$	—	s_1	s_1	—	—	1	1	—

Рис. 3.21. Минимальный автомат.

Пока еще не известен никакой систематический метод сокращения перебора для отыскивания минимального замкнутого согласованного множества совместимых классов. Следует просто начать с какого-то самого маленького множества, покрывающего автомат, и увеличивать его, пока не будет достигнута согласованность.

УПРАЖНЕНИЯ В

1. Минимизировать число состояний следующего неполностью описанного автомата:

Текущее состояние	Следующее состояние				Выход			
	a_0	a_1	a_2	a_3	a_0	a_1	a_2	a_3
1	2	1	—	—	0	—	—	1
2	2	1	—	2	—	0	—	—
3	1	4	3	—	1	0	0	1
4	1	4	2	2	0	—	0	—
5	2	—	2	—	—	0	—	1

2. Найти все пары состояний s_i, $\bar{s}_j$ со свойством $s_i \leqslant \bar{s}_j$ для следующих двух неполностью описанных автоматов:

	Следующее состояние		Выход	
	a_0	a_1	a_0	a_1
s_1	s_3	—	1	—
s_2	s_4	s_3	1	0
s_3	s_2	s_4	—	—
s_4	—	—	1	—
s_5	s_1	—	—	1

	Следующее состояние		Выход	
	a_0	a_1	a_0	a_1
$\bar{s}_1$	$\bar{s}_2$	—	1	0
$\bar{s}_2$	—	$\bar{s}_3$	1	—
$\bar{s}_3$	—	—	1	—
$\bar{s}_4$	$\bar{s}_1$	$\bar{s}_2$	—	1

3. Минимизировать число состояний следующего автомата:

	γ		ζ	
	a	b	a	b
1	2	—	0	—
2	1	6	0	0
3	4	—	1	—
4	5	3	0	0
5	—	6	—	1
6	4	—	0	—

$A = \{a, b\}$
$Z = \{0, 1\}$
$S = \{1, 2, 3, 4, 5, 6\}$

4. Минимизировать число состояний следующего автомата:

Состояния	Входы	
	a_0	a_1
s_1	s_2, 0	—
s_2	s_1, 0	s_6, 0
s_3	s_4, 1	—
s_4	s_5, 0	s_3, 0
s_5	—	s_6, 1
s_6	s_4, 0	—

5. Минимизировать число состояний следующего автомата:

	Следующее состояние			Выход		
	a_0	a_1	a_2	a_0	a_1	a_2
1	—	1	2	—	z_0	z_1
2	—	2	1	—	z_0	z_0
3	3	4	—	z_0	z_0	—
4	4	5	—	z_1	z_0	—
5	5	3	—	z_2	z_0	—
6	1	—	1	z_2	—	z_2
7	2	—	2	z_1	—	z_0

6. Минимизировать число состояний следующего автомата:

	Следующее состояние			Выход		
	a_0	a_1	a_2	a_0	a_1	a_2
S_1	S_3	—	S_2	0	—	—
S_2	S_2	S_5	S_3	0	0	0
S_3	S_2	S_4	—	0	—	—
S_4	S_5	S_1	—	0	—	—
S_5	S_1	S_4	—	—	1	1

7. Рассмотрим машины Тьюринга с двумя входными символами X_1, X_2, двумя выходными символами Z_1, Z_2 и двумя внутренними состояниями Q_1, Q_2. Сколько разных таблиц полностью описывают всевозможные такие машины? Сколько разных таблиц неполностью описывают их? Существуют ли эквивалентные машины среди тех, которые описываются полными таблицами? Существуют ли эквивалентные машины среди тех, которые описываются неполными таблицами? Каково отношение между таблицами состояний конкретной машины Тьюринга и общей машины Тьюринга? Сколько существует неэквивалентных машин Тьюринга с двумя входными символами, двумя выходными символами и двумя внутренними состояниями.

8. Рассмотрим машину Тьюринга с четырьмя входными и выходными символами S_1, S_2, S_3, пробел. На вход подаются строки вида

пробел	X_1	X_2	X_3	X_4	X_5	пробел	X_1	X_2	...

где каждая из переменных X_i может принимать любое значение S_1, S_2, S_3. Машина Тьюринга должна переводить любой такой вход в выход

пробел	X_5	X_4	X_3	X_2	X_1	пробел	X_5	X_4	X_3	...

Описать диаграмму состояний и движений ленты для такой машины, реализующей функцию $f(X_1, X_2, \ldots, X_5) = (X_5, X_4, \ldots, X_1)$, печатая один символ X_i за такт.

СПИСОК ЛИТЕРАТУРЫ

1. Arbib M. A. (ed.), Algebraic Theory of Machines, Languages and Semigroups, Academic Press, 1968.
2. Bartee T. C., Digital Computers Fundamentals, 2d ed., McGraw-Hill, 1966.
3. Hartmanis J., Stearns R. E., Algebraic Structure Theory of Sequential Machines, Prentice-Hall, 1966.
4. McCluskey E. J., Introduction to the Theory of Switching Circuits, McGraw-Hill, 1965.
5. McCluskey E. J., Bartee T. C. (eds.), A Servey of Switching Circuit Theory, McGraw-Hill, 1962.
6. McNaughton R., Theory of Automata: A Survey, *Adv. in Computers*, 2: 379—421 (1961).
7. Minsky M., Computation: Finite and Infinite Machines, Prentice-Hall, 1967. [*Русский перевод*: Минский М., Вычисления и автоматы, М., «Мир», 1971.]

4

ЯЗЫКИ ПРОГРАММИРОВАНИЯ

4.1. ВВЕДЕНИЕ

Цифровые вычислительные машины способны совершать арифметические и логические операции с высокой скоростью и точностью. Однако предварительно в машину следует ввести список команд, определяющий ее работу, в специально подготовленном виде. Составление этого списка команд, называемого *программой*, является серьезной проблемой. Пользователь обычно представляет свои соображения о процедуре решения задачи в виде текста на каком-нибудь естественном языке (английский, русский), либо в виде блок-схемы, предписывающей последовательность шагов, либо, наконец, в виде какой-то системы математических соотношений. Необходимость преобразования (трансляции) этой информации в программу, приемлемую для вычислительной машины,— довольно серьезное препятствие для пользователя. Процесс этого преобразования называется *программированием*, а его заключительный этап—фактическое выписывание команд для машины—*кодированием*. Конечно, пользователь желал бы «объясняться» с ЭВМ на языке, естественном для данной задачи. Например, пользователь, желающий обработать числовые данные, переданные с космического корабля, мог бы попросить примерно следующее:

«Вычислить среднее значение данных, отбросить все значения, отличающиеся от среднего больше чем на удвоенное среднеквадратичное отклонение, и снова вычислить среднее оставшихся значений».

Вид этой инструкции весьма далек, однако, от списка двоичных слов, составляющих окончательный вид программы на *машинном языке*.

Для облегчения программирования пользователи довольно давно разработали специальные «искусственные» языки, тексты на которых машина затем сама транслировала на машинный язык. Оказалось, что эти языки более естественны и удобны для пользователей, а трансляция с них со временем стала функцией специальных программ.

Сначала появились совсем простые искусственные языки, называемые *языками ассемблера*, и программы трансляции на машинный язык — *трансляторы*. Рассмотрим простой пример.

В простейшем виде данные и команды для ЭВМ делятся на слова. Каждое слово содержит либо команду, либо какое-то значение — число или имя. Память компьютера составлена из таких слов фиксированной длины, каждому из которых приписывается числовой адрес (номер ячейки). В слове памяти хранится либо команда, либо некая информация. Например, в первых четырех словах памяти могут содержаться четыре целых числа 29, 364, 48, 200 соответственно, или четыре команды. В первом случае мы можем сказать, что в памяти по адресу 1 содержится число 29, по адресу 2 содержится 364 и т. д.

Вот четыре типичные команды на языке ассемблера из реквизита типичной машины.

1. CLA X. Эта команда *очищает* (CLear) специальный регистр, называемый *сумматором* (accumulator), и затем *складывает* (Add) значение, записанное в памяти по адресу X, с (нулевым) содержимым сумматора. Это значение хранится в сумматоре для последующего использования.

2. ADD X. Эта команда *складывает* значение, находящееся в памяти по адресу X, с текущим значением сумматора и записывает сумму в сумматор.

3. MUL X. Эта команда *умножает* значение, находящееся в памяти по адресу X, на текущее значение сумматора и записывает произведение в сумматор.

4. STO X. Эта команда берет число из сумматора и помещает, или *запоминает* (STOrage), его в памяти по адресу X.

В этих примерах X означает число, соответствующее адресу, или ячейке, в памяти. Как уже упоминалось, память цифровых машин состоит из таких ячеек, и в каждую ячейку можно записать некоторую информацию. В нашем случае информация состоит из чисел, обычно в двоичной записи.

Предположим, что по адресу X в памяти хранится число 6, по адресу Y — число 7 и по адресу Z — число 2. Мы можем заставить машину вычислить $(6+7)2$, предъявив ей следующую последовательность команд: «очистить сумматор, сложить его содержимое с числом по адресу X, прибавить число по адресу Y и умножить результат на число по адресу Z». После выполнения этих команд значение $(6+7)2$ окажется в сумматоре. После этого можно командой STO W переслать это число в ячейку W.

На языке ассемблера программист напишет эту последовательность команд так:

```
CLA  X
ADD  Y
MUL  Z
STO  W
```

Предположим, что каждое слово памяти содержит 12 двоичных разрядов (так называемое двенадцатибитовое слово). Тогда приведенная выше последовательность команд должна быть закодирована четырьмя двенадцатибитовыми словами, т. е. 48 разрядами. Перевод этих предложений в сорок восемь нулей и единиц — задача для человека трудная и чреватая ошибками. Ассемблер делает несколько полезных и простых вещей. Он связывает с каждым типом команды, которую может выполнить машина, некоторый мнемонический код, вроде CLA, и дает возможность пользователю писать программу, пользуясь буквами и цифрами (алфавитно-цифровыми метками) вместо адресов. После этого машина сама переводит мнемокод в настоящий двоичный код операции и присваивает конкретные адреса переменным, входящим в программу.

Допустим, что первые четыре разряда слова-команды обозначают код операции, которую нужно выполнить по этой команде. Пусть двоичный код для CLA есть 0100, для ADD—0101, для MUL—1110 и для STO—0110. Программа будет выглядеть, например, так:

```
CLA  X  ↦ 010001000101
ADD  Y  ↦ 010110000110
MUL  Z  ↦ 111010000111
STO  W  ↦ 011011000000
```

Здесь правая часть является результатом работы ассемблера и представляет собой требуемый машинный код. Первые четыре разряда в каждой строке являются кодом очередной операции, в последних восьми разрядах записаны адреса X, Y, Z, W. Если в ячейке X вначале содержится число A, в ячейке Y—число B и в ячейке Z—число C (в предыдущем примере $A=6$, $B=7$ и $C=2$), то после окончания работы программы в ячейке W будет записано число $(A+B)C$.

Мы описали простую программу на одном из ранних языков ассемблера. Пользование такими языками настолько сократило работу программистов и сделало машину доступной такому широкому кругу людей, что языки ассемблера стали развиваться и усложняться, предоставляя пользователям все большие воз-

можности. Однако программы даже для вычисления простых математических выражений на таких языках могут быть очень длинными, в частности, при необходимости условных передач управления.

Существенно, что трансляция на язык ассемблера должна производиться по принципу «один к одному». Иными словами, машинные команды в двоичной записи взаимно однозначно отвечают командам на языке, так что написанная программа по существу мало отличается от машинной. По этой причине языки ассемблера часто называют *машинными*, хотя на самом деле это совершенно другие объекты.

С распространением идеи о том, что использование машины для трансляции может облегчить составление программ, началась работа по созданию *проблемно-ориентированных* языков и трансляторов с них — *компиляторов*. При трансляции с этих языков краткие выражения языка соответствуют большим последовательностям команд, а сами языки обладают значительной гибкостью. Первым из таких языков был ФОРТРАН (акроним для FORmula TRANslation), созданный около 1955 года. Он получил широкое признание, что привело к разработке большого количества новых языков для разного типа задач. Среди них выделяются КОБОЛ для экономических расчетов и АЛГОЛ, ориентированный на нужды научных исследований.

На ФОРТРАНе мы можем просто написать строку $D=(A+B)*C$, и после ее трансляции и выполнения машиной переменная D примет значение $(A+B)\times C$. Удобство ФОРТРАНа не ограничивается возможностью писать алгебраические выражения в виде, близком к обычному; в него заложены также логические средства, которые передаются кратким набором словесных команд вроде go to (перейти к), if (если), read (читать), do (выполнить).

В 1958 году в Цюрихе (Швейцария) собралась группа специалистов для разработки международного языка программирования, предназначенного для научных расчетов. Подготовленный этой группой язык был назван АЛГОЛ (акроним для ALGOrithmic Language), точнее, в первом варианте, АЛГОЛ-58. К 1960 году стало ясно, что первый вариант нуждается в серьезном пересмотре, и новая версия языка была названа АЛГОЛ-60. При внесении последующих изменений и дополнений язык обычно называли просто АЛГОЛом. Те его фрагменты, которые мы обсудим ниже, лишь незначительными деталями отличаются как от АЛГОЛа 60, так и от любого из сейчас употребительных диалектов. Хотя популярность АЛГОЛа как языка программирования не может сравниться с популярностью ФОРТРАНа, он стал международным языком алгоритмов в научной литературе.

Одна из основных целей первых разработчиков АЛГОЛа состояла в создании *машинно-независимого* языка, позволяющего

написать компилятор с него, подходящий для любой машины умеренных размеров. Тогда программисты могли бы писать программы, пригодные для любой машины, имеющей компилятор с АЛГОЛа. КОБОЛ и ФОРТРАН также являются универсальными языками программирования в том смысле, что и АЛГОЛ. АЛГОЛ — стандартный язык международных журналов: он принят в качестве языка публикаций ACM (Association for Computing Machinery). (Постепенно входит в употребление язык PL/1, сходный с АЛГОЛом, но располагающий также рядом возможностей ФОРТРАНа.)

4.2. АРИФМЕТИЧЕСКИЕ ВЫРАЖЕНИЯ

Язык программирования АЛГОЛ строится из следующих основных символов:

1. Буквы: 26 строчных и 26 прописных букв английского алфавита a, b, ..., z и A, B, ..., Z.
   Цифры: 0, 1, ..., 9.
2. Три группы операций:
   арифметические операции: $+$, $-$, $\times$, $/$, $\div$, $\uparrow$
   операции отношения: $=$, $\neq$, $<$, $\leqslant$, $>$, $\geqslant$
   логические операции: $\wedge$, $\vee$, $\neg$, $\equiv$
3. Разделители и скобки: , . ; : : =) ([ ]
4. Выделенные слова (отмечаются всюду жирным шрифтом). В их числе **real, integer, true, go to, for, step, until, begin, end.**

В этом параграфе мы опишем, как в АЛГОЛе формируются *арифметические выражения*. Они могут быть комбинацией констант, как в арифметике, или переменных, как в алгебре, или тех и других. Начнем с простых, но важных замечаний о *константах*.

В программах на АЛГОЛе могут быть константы трех *типов*: **real** (вещественные), **integer** (целые), и **Boolean** (булевы). Булевых констант имеется ровно две **true** (истина) и **false** (ложь). Числовые константы, однако, имеют разные формы записи.

В АЛГОЛе для обозначения целого числа $n \in \mathbf{Z}$ со знаком или без него может использоваться *целая константа*. Она должна быть записана без десятичной точки (иначе она будет интерпретирована как вещественное число). Вот примеры алгольных записей целых констант:

3 0 +16 —1764 12346

Запись 3.0 неприемлема из-за десятичной точки. Она будет интерпретирована как вещественная константа.

Вещественная константа может быть представлена целой константой (если она таковой является) или одним из следующих способов:

(i) как десятичная дробь: 3.0, +3.0, 6.47, —367.4325 или 0.004328,

(ii) как целое число или десятичная дробь, за которыми следует «опущенная десятка» $_{10}$ и затем целый «показатель степени». Такая запись обозначает произведение числа, записанного до $_{10}$, на 10 в указанной степени. Примеры:

$5.34_{10}4$ означает $53400 = 5.34 \times 10^4$,

$-.0687_{10}-5$ означает $-.0687 \times 10^{-5}$,

$2_{10}+15$ означает 2×10^{15},

$365_{10}-12$ означает 365×10^{-12}.

Форма записи (ii) соответствует общепринятой в естественнонаучных текстах. Она особенно полезна для представления очень больших или очень малых чисел, вроде диаметра электрона или галактики в сантиметрах.

При написании программы необходима крайняя тщательность. Поэтому важно иметь в виду следующие дополнительные ограничения, принятые в АЛГОЛе. Десятичная дробь ни при каких обстоятельствах не должна оканчиваться точкой. Так, записи $10._{10}-6$ и $73._{10}-6$ неприемлемы, тогда как 10.0 и $73.00_{10}-6$ вполне допустимы. Далее, в качестве показателей степени допустимы только целые числа, так что запись $5.34_{10}2.5$ не имеет смысла в АЛГОЛе, несмотря на то что $5.34 \times 10^{2.5}$ — вполне определенное вещественное число.

В АЛГОЛе это число все же *можно* записать в виде $5.34 \times 10 \uparrow 2.5$. Вообще всякое арифметическое или алгебраическое выражение нетрудно записать в АЛГОЛе, строго придерживаясь следующих соглашений.

Прежде всего, *арифметическим выражением* в АЛГОЛе называется последовательность алгольных числовых **(integer** или **real)** констант и переменных, соединенных набором бинарных операций сложения (+), вычитания (—), умножения (×), деления (/) и возведения в степень (↑), порядок выполнения которых регулируется скобками.

Удобно записывать арифметические выражения с минимумом скобок. Например, в алгебре мы пишем $a+bc$, а не $a+(bc)$, и, по школьным правилам, мы *не можем* прочесть $a+bc$ как $(a+b)c$. В выражениях, содержащих как сложение, так и умножение, произведения должны вычисляться раньше сумм. Аналогичные правила порядка действий должны быть сформулированы для АЛГОЛа. Они оказываются несколько более подробными, ибо возведение в степень записывается «в строку». В *отсутствие*

скобок предписывается следующий порядок действий: возведение в степень производится прежде всего, умножение и деление предшествуют сложению и вычитанию. Умножение и деление по отношению друг к другу равноправны, то же относится к сложению и вычитанию. Равноправные операции производятся в порядке чтения, слева направо. Итак, действия совершаются в следующем порядке:

1) возведение в степень,
2) умножение и деление,
3) сложение и вычитание, в неопределенных случаях — по порядку слева направо.

Способ применения этих правил станет ясным из следующих примеров. Мы включили в выражения также символы для переменных.

Математическое выражение	Алгольное выражение
$3+\frac{5}{7}$	3+5/7
$\frac{3+5}{7}$	(3+5)/7
$2+3^7$	2+3↑7
$\frac{ab}{c}$	a×b/c
$\frac{a}{bc}$	a/(b×c)
$(a/b)\,c$	a/b×c
$\pi\,(r-p)^2$	3.14159×(r−p)↑2
$\frac{6^{2.437}}{.034^{3.26}}$	6↑2.437/.034↑3.26
$\frac{a^b+c^d}{4q}$	(a↑b+c↑d)/(4×q)

В этих примерах арифметические выражения были записаны с минимумом скобок. Однако экономить на скобках не обязательно: выражение $a+\frac{b}{c}$ в АЛГОЛе можно записать как $a+(b/c)$, а не только $a+b/c$. Следует, однако, соблюдать осторожность: a^{bc} нельзя записать в АЛГОЛе как $a\uparrow b\times c$. Эта запись будет прочитана компилятором как a^bc, ибо возведение в степень в отсутствие скобок предшествует умножению. Правильная запись имеет вид $a\uparrow(b\times c)$. Следует также помнить, что в програм-

мах на АЛГОЛе следует писать $3 \times A$, а не $3A$, чтобы выразить «трижды A».

Символ $\div$. В АЛГОЛе имеется шестая арифметическая операция $\div$. Она оперирует лишь с целыми выражениями, и тогда значением $a \div b$ является **integer** — целая часть отношения a к b. Так, $6 \div 2$ имеет значение 3, $5 \div 2$ — значение 2, и $(3+6) \div 4$ — значение 2.

4.3. ИДЕНТИФИКАТОРЫ И ОПЕРАТОРЫ ПРИСВАИВАНИЯ

Точно так же, как переменные x, y, z, ... в математике, алгольные переменные могут принимать *значения* в некоторых *множествах*, которые должны быть явно указаны; кроме того, алгольные переменные имеют *имена*.

Имя алгольной переменной называется *идентификатором*. Идентификатором может быть последовательность из любого количества букв, цифр и пробелов, однако первый ее символ должен быть буквой. Примеры допустимых идентификаторов:

```
x
maximum
TIME
sum
r37BQ2
Largest Diagonal Element
root1
```

Следующие выражения *не могут быть* идентификаторами:

3ab	(первый символ — не буква),
X, YZ	(запятая не является ни буквой, ни цифрой),
step	(**step** — это выделенное слово, используемое в АЛГОЛе как единый символ, и компилятор не может воспринять его как идентификатор: такие слова резервированы).

В математике переменные могут быть разных типов, обозначая вещественные, комплексные или рациональные числа, элементы групп и т. д. В АЛГОЛе допускаются лишь три основных типа переменных: **integer, real** и **Boolean**. *Целые переменные* принимают целочисленные значения, *вещественные переменные* принимают вещественные числовые значения, точнее, десятичные приближения к ним; *булевы переменные* (которые будут обсуждены в гл. 5) принимают значения **true** и **false.**

Тип каждой переменной должен быть явно указан *описанием типа*. Так, согласно описаниям

real $A1$, SIGMA;
integer SUM, alpha, beta;
Boolean X;

переменные $A1$ и SIGMA должны принимать вещественные значения, SUM, alpha и beta — только целые, а X может принимать лишь значения **true** и **false**.

Перейдем теперь к важнейшей конструкции АЛГОЛа — *операторам присваивания*. Общий вид операторов присваивания таков:

$$\mathcal{V} := \mathcal{E};$$

Здесь $\mathcal{V}$ — имя некоторой алгольной переменной, а $\mathcal{E}$ — некоторое алгольное выражение. Оператор присваивания дает предписание машине вычислить значение выражения справа $\mathcal{E}$ при текущих значениях всех переменных, входящих в $\mathcal{E}$, и присвоить переменной слева $\mathcal{V}$ это значение, которое теперь становится ее текущим значением, предыдущее забывается.

Примеры:

$A := 5;$ присвоить A значение 5,

$B := 5 \times A;$ присвоить B текущее значение A, умноженное на 5,

$C := 5 \times A \times B;$ присвоить C текущее значение $5AB$.

Следует обратить внимание на то, что каждый оператор заканчивается точкой с запятой. Последовательность операторов

$$A := 2;$$
$$B := 3;$$
$$C := A \times B;$$

присваивает переменной C значение 6.

Рассмотрим теперь пример *блока* в АЛГОЛе для иллюстрации уже введенных понятий. Алгольный блок представляет собой последовательность операторов и описаний, которой предшествует выделенное слово **begin** (начало) и которую заключает слово **end** (конец). Позже мы подробнее изучим блочную структуру алгольных программ, а пока повторим простое правило: блок начинается с **begin** и кончается **end**, как в следующем

примере:

```
begin real a, b, c;
  c := 5;
  a := 4.1;
  b := 2×a+7;
  c := 3×a—b;
end
```

Еще раз обратите внимание на точки с запятой. Все операторы в АЛГОЛе, за которыми не следует **end**, должны завершаться точкой с запятой, отмечающей конец оператора. Нужно запомнить также, что все переменные, входящие в блок, должны сопровождаться описаниями типа. В нашем примере a, b, c описаны как вещественные переменные—**real**.

Операторы описания типа не предназначены для выполнения машиной: они лишь доставляют необходимую информацию о типах всех переменных в блоке во время трансляции (преобразовании программы на АЛГОЛе в машинные коды). По этой причине операторы описания типа называются *неисполняемыми*, в противоположность *исполняемым* операторам присваивания. Теперь мы рассмотрим в подробностях действие операторов присваивания в блоке, описанном выше.

$c := 5;$ — переменной c присваивается значение 5,

$a := 4.1;$ — переменной a присваивается значение 4.1,

$b := 2\times a+7;$ — вычисляется значение правой части при текущем значении a, именно 4.1; результат 15.2 присваивается в качестве значения переменной b.

$c := 3\times a-b;$ — переменной c присваивается значение $3\times 4.1 - -15.2 = -2.9$, а ее предыдущее значение 5 забывается. Теперь —2.9 является текущим значением c.

Следующий пример весьма выразительно иллюстрирует смысл операции := оператора присваивания. Запись

$$N := N + 1$$

является совершенно законным оператором присваивания в АЛГОЛе; его выполнение увеличивает текущее значение N на единицу. Так, если текущее значение N до выполнения оператора было 5, то после его выполнения оно станет 6. Совершенно ясно, что эта запись *не* выражает равенства левой и правой частей; символ := означает замену, а не равенство. Запись оператора «присвоить переменной $\mathcal{V}$ значение $\mathcal{E}$» можно также

прочесть как «заменить значение $\mathscr{V}$ значением $\mathscr{E}$», но никогда не «$\mathscr{V}$ равно $\mathscr{E}$».

Смешанные выражения. В АЛГОЛе допускаются разные типы в пределах одного выражения: в него могут входить одновременно вещественные и целые переменные и константы. (В большинстве версий ФОРТРАНа этого нельзя делать.) Тип результата вычислений над операндами разных типов определяется следующими правилами:

1. Результатом операций $+$, $-$, $\times$ является целое число, если оба операнда целые, и вещественное число в остальных случаях.
2. Результатом операции $a \uparrow b$ является целое число, если a целое число, а b — неотрицательное целое число, и вещественное число в остальных случаях.
3. Результатом операции a/b всегда является вещественное число, независимо от типов операндов.
4. Операция $a \div b$ определена только тогда, когда a и b — целые числа; результат — целая часть частного a/b — также типа **integer**.

Рассмотрим в качестве примера выполнение следующих операторов:

$a := 3.6 + 1/4;$ $\quad b := 3 + 2.0;$ $\quad c := 5 \uparrow (-2);$

$i := 5 \uparrow 3;$ $\quad j := (4 \times 3) \div 2;$ $\quad k := 4 \times (3 \div 2);$

В АЛГОЛе значения первых трех — типа **real**: $a = 3.85$, $b = 5.0$ и $c = 0.04$, а последних трех — типа **integer** $i = 125$, $j = 6$ и $k = 4$.

4.4. МАССИВЫ

В АЛГОЛе векторы называются *одномерными массивами*, матрицы — *двумерными массивами*. Индексы их компонент (элементов) заключаются в квадратные скобки. Так, компоненты вектора $x = (x_1, x_2, x_3)$ называются элементами одномерного массива $x[1:3]$ и записываются $x[1]$, $x[2]$, $x[3]$. При первом упоминании массива число его элементов должно быть объявлено в описании типа. Так, описание

real array $X[1:20]$

объявляет, что массив X содержит 20 переменных $X[1]$, ..., $X[20]$, и каждая из них вещественная. Слово **real** здесь можно опустить; транслятор прочитает **real array** $X[1:20]$ так же, как **array** $X[1:20]$. Однако булевы массивы и целые массивы должны быть объявлены полностью. Описание

integer array $X[-5:15]$

объявляет, что $X[-5]$, $X[-4]$, ..., ..., $X[15]$ является набором из 21 целозначных переменных, индексы которых меняются от -5 до 15.

Аналогично описание

array $A[1:10, 1:15]$

относится к вещественной матрице $A[i, j]$ размера 10×15; индекс i пробегает значения от 1 до 10, а индекс j — от 1 до 15. Наконец, описание

integer array $C[-1:8, 0:9, 2:11]$

относится к трехмерному массиву, состоящему из 1000 целозначных переменных $C[i, j, k]$.

Описание массива используется компиляторами или другим процессором в основном для того, чтобы резервировать необходимое количество слов в памяти для записи элементов массива.

Функции. Компиляторы с АЛГОЛа допускают использование некоторых вещественных функций из анализа. К ним обычно относятся

abs(E)	абсолютная величина выражения E,
sqrt(E)	квадратный корень из значения E,
sin(E)	синус значения E,
cos(E)	косинус значения E,
ln(E)	натуральный логарифм значения E,
exp(E)	экспоненциальная функция от значения E.

Аргументы этих функций могут принимать целые или вещественные значения; тип значения функции всегда **real**. Следующие примеры дают образцы алгольных выражений для функций:

Математическое выражение	Выражение на АЛГОЛе
$e^x + y$	exp (x)+y
e^{x+y}	exp (x+y)
$\sin^3\left(\frac{x}{y}\right)$	sin (x/y) ↑ 3
$\sin\frac{x^3}{y}$	sin (x ↑ 3/y)
$\frac{-b+\sqrt{b^2-4ac}}{2a}$	(−b+sqrt (b ↑ 2−4×a×c))/(2×a)
$\sqrt{1+\frac{1}{\sin\lvert x\rvert}}$	sqrt (1+1/sin (abs (x)))

Предостережение. В АЛГОЛе нет средств для изображения комплексных чисел как таковых. Поэтому мы не оберемся хлопот с квадратным корнем из отрицательного числа, если попробуем воспользоваться невинно выглядящей алгольной программой для вычисления корней квадратного уравнения

$$3x^2 + \sqrt{e^{\pi}}x + \ln 8.7 = 0$$

по школьной формуле

$$\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

для корней уравнения $ax^2 + bx + c = 0$:

```
begin real a, b, c, d, root1, root2;
    a := 3;
    b := sqrt (exp (3.14159));
    c := ln (8.7);
    d := sqrt (b ↑ 2 — 4 × a × c);
    root1 := (—b + d)/(2 × a);
    root2 := (—b — d)/(2 × a);
end
```

УПРАЖНЕНИЯ А

1. Записать следующие алгольные выражения в виде обычных математических формул:

   а) $A \times B - C$ г) $A/B - C \uparrow D$
   б) $A \uparrow (B - C)$ д) $A \div B - C \times D$
   в) $A \uparrow B - C$

2. Вычислить значения выражений 1 при $A = 2$, $B = 3$, $C = 4$, $D = 5$.

3. Записать следующие выражения на АЛГОЛе, не пользуясь скобками:

   а) $A\,(B + C) - D$, в) $\dfrac{A^2}{B + C}$,

   б) $\dfrac{A}{B + C} - D$, г) $(A^N - B^N)\,(C - D)$.

4. Записать следующие выражения на АЛГОЛе:

   а) $\sin (A + B^2)$, г) $\sqrt{e^A - \sin A}$,
   б) $e^B - C$, д) $\sin^4 (A + 3^2)$.
   в) Абсолютная величина (натурального логарифма (A^B)),

5. Представить следующие алгольные выражения обычными математическими формулами:

   а) $A - B \times C \times D$ г) $A \times D - E + F \uparrow G$
   б) $A - B + C \uparrow D \times E$ д) $A - B \times C + D$
   в) $A \uparrow D - E \uparrow F$

6. Вычислить значения выражений упр. 5 при $A=2$, $B=3$, $C=-2$, $D=4$, $E=5$, $F=6$, $G=-2$.

7. Записать следующие выражения на АЛГОЛе, не пользуясь скобками:
   а) $A^2+B^C\cdot D$, в) $3\cdot A+BC+(D-E)$,
   б) $A\cdot B+C^{D^E}$, г) $4(A+B)+(6(A-C))^2$.

8. Записать следующие выражения на АЛГОЛе:
   а) $A^2+B^2+(C^3-D)^3$, г) $e^{x/y}$,
   б) $\sin A+\cos^2 B$, д) $\sin^2 e^{x/y}$.
   в) $e^{|x|}$,

9. Записать следующие выражения на АЛГОЛе:
   а) $\sin x+e^{\cos x}$, г) e^x+e^y,
   б) $\sin(x+e^{\cos x})$, д) $e^x\cdot e^y$.
   в) $\sin x+e^{\cos(x+y)}$,

4.5. ОПЕРАТОРЫ ЦИКЛА

В вычислениях часто возникает необходимость повторно выполнять некоторую серию операторов над переменным операндом, значение которого меняется с одним и тем же шагом. Рассмотрим, например, задачу вычисления суммы

$$S=\sum_{i=1}^{n} i^2. \qquad (1)$$

Мы можем начать с оператора присваивания $S:=0$ и затем последовательно выполнять операторы $S:=S+i^2$, $i=1, 2, \ldots, n$. После выполнения k-го оператора переменной S будет присвоено значение $1^2+2^2+\ldots+k^2$ и, наконец, n-е выполнение присвоит S искомое значение $\sum_{i=1}^{n} i^2$.

Оператор цикла в АЛГОЛе позволяет производить такие итерации и является одним из самых сильных средств языка. Один из его видов (не самый общий, но достаточный для наших целей) таков:

for $v:=m$ **step** h **until** n **do** T;

Здесь v может быть вещественной или целой переменной; m, h, n — обычно константы, а T — оператор. Оператор T выполняется для каждого значения переменной v от m до n с шагом h. При $v>n$ управление передается следующему оператору.

Пользуясь оператором цикла, мы можем написать программу на АЛГОЛе для вычисления суммы квадратов совсем просто:

```
S:= 0;
for i:= 1 step 1 until n do S:= S+(i↑2);
```

Вот еще некоторые примеры операторов цикла:

1. **for** $i := 1$ **step** 1 **until** 10 **do** S;
   Оператор предписывает выполнять S для $i = 1, 2, \ldots, 10$.
2. **for** $i := -4$ **step** 2 **until** 7 **do** S;
   Оператор предписывает выполнять S для $i = -4, -2, 0, 2, 4, 6$.
   Заметим, что значение 7 опускается.
3. **for** $x := 0$ **step** .1 **until** 1 **do** S;
   Оператор предписывает выполнять S для $x = 0, 0.1, 0.2, \ldots$
   $\ldots, 0.9, 1.0$.
4. **for** $x := 1$ **step** $-.1$ **until** $-.5$ **do** S.
   Оператор предписывает выполнять S для $x = 1, 0.9, 0.8, \ldots$
   $\ldots, -0.4, -0.5$.
5. **for** $x := 5$ **step** 1 **until** 4 **do** S.
   S вообще не будет выполняться.

Рассмотрим четыре программы на АЛГОЛе.

Пример 1. Следующий алгольный блок порождает массив K с элементами $K[i] = i!$ для $i = 1, 2, \ldots, 10$:

```
begin real array K[1:10];
      integer i;
      K[1]:= 1.0;
      for i:= 2 step 1 until 10 do K[i]:= K[i-1]×i;
end
```

Пример 2. Пусть дана последовательность N результатов наблюдений $x_1, \ldots, x_N (N \leqslant 500)$, и пусть мы хотим вычислить (1) среднее $m = (\sum x_i)/N$ и (2) среднеквадратичное отклонение $s = \left[\sum_{i=1}^{N} (x_i - m)^2\right]^{1/2} / (N-1)$. Это можно выполнить с помощью следующей последовательности операторов АЛГОЛа. Пусть значения переменных m, s, i, N и массив $x[1], \ldots, x[500]$ хранятся в памяти. Переменные $x[i]$, m, s вещественные, а i, N целые. Программа может быть такой:

```
m:= 0;
for i:= 1 step 1 until N do
    m:= m+x[i];
m:= m/N; s:= 0;
for i:= 1 step 1 until N do
    s:= s+(x[i]-m)↑2;
s:= sgrt(s)/(N-1);
```

Пример 3. Пусть мы хотим вычислять значения многочлена, скажем, степени 50, с известными коэффициентами

$$f(y)=\sum_{j=0}^{50} a_j y^j = a_0 + a_1 y + \ldots + a_{50} y^{50}. \qquad (2)$$

Предположим, что значения $a[i]$ хранятся в памяти. Самый примитивный способ вычисления задается программой

```
POLY := 0;
for j := 0 step 1 until 50 do
    POLY := POLY + (a[j] × y ↑ j);
```

Она вычисляет степени y, умножает их на соответствующие коэффициенты и складывает получившиеся значения. Программа очень проста, но требует выполнения неоправданно большого количества операций. Если вычислять $y \uparrow j$ рекурсивно по формулам $y \uparrow 0 = 1$ и $y \uparrow (j+1) = y \times y \uparrow j$, потребуется $N(N-1)/2$ умножений и N сложений, где N — степень многочлена (50 в примере (2)). Следующий способ не столь разорителен.

Пример 4. Значение многочлена

$$f(x) = a_N x^N + a_{N-1} x^{N-1} + \ldots + a_0$$

можно вычислять так:

$$f(x) = x(\cdots x(x(a_N x + a_{N-1}) + a_{N-2}) + \cdots + a_1) + a_0.$$

Это требует N умножений и N сложений: значительная экономия.

Данный способ реализуется следующей программой на АЛГОЛе. Предполагается, что коэффициенты $a_j = A[j]$ хранятся в памяти как вещественные константы и что степень N и аргумент x зафиксированы как значения целой переменной N и вещественной X соответственно. Требуемое вычисление осуществляет следующая последовательность операторов:

```
F := A[N];
for j := N − 1 step − 1 until 0 do
    F := X × F + A[j];
```

Переменная j в операторе цикла принимает значения $N-1$, $N-2$, ..., 1, 0 с шагом -1. Читателю следует убедиться, что после выполнения этой программы переменной F действительно будет присвоено искомое значение. Можно, например, проследить действие этой программы шаг за шагом (т. е. «выполнить» ее вручную) в применении к многочлену небольшой степени, скажем, $4x^3 + 2x^2 + 3x + 1$.

4.6. БЛОЧНЫЕ СТРУКТУРЫ В АЛГОЛе

Программы на АЛГОЛе составлены из блоков. Блоком называется последовательность операторов, которая начинается словом **bigin** с следующим за ним описанием типа и заканчивается словом **end**.

Каждая алгольная программа должна быть оформлена в виде блока.

Пример 5. Имеется два одномерных массива X и Y по 50 элементов каждый. Следующая программа вычисляет значения

$$LX=\sqrt{\sum_{j=1}^{50} X_j^2} \quad \text{(длина вектора } X\text{)}$$

$$LY=\sqrt{\sum_{j=1}^{50} Y_j^2} \quad \text{(длина вектора } Y\text{)}$$

$$\text{INPROD}=\sum_{j=1}^{50} X_jY_j \quad \text{(скалярное произведение } X \text{ и } Y\text{)}.$$

```
begin real array X[1:50], Y[1:50];
      real LX, LY, INPROD;
      integer j;
      LX:= LY:= INPROD:= 0;
      for j:= 1 step 1 until 50 do
         begin LX:= LX+X[j]↑2;
               LY:= LY+Y[j]↑2;
               INPROD:= INPROD+X[j]×Y[j]
         end
      LX:= sqrt(LX); LY:= sqrt(LY)
end
```

Составным оператором называется последовательность операторов, начинающаяся **begin** и оканчивающаяся **end**; отличие от блока лишь в том, что вслед за **begin** нет описаний типа. Операторы, заключенные между **begin** и **end**, синтаксически рассматриваются как единое целое (один оператор). В примере 5 оператор цикла управляет составным оператором, состоящим из трех «простых» операторов. Заметим, что в отсутствие операторных скобок **begin** и **end** управляемым был бы оператор $LX:=LX+X[j]\uparrow 2$; и, разумеется, требуемый результат не был бы получен.

Следует отметить, что блоки и составные операторы допускают рекурсию. Оператор, входящий в составной оператор, сам может

быть составным, как в следующей схеме:

```
begin S1;
      S2;
      begin S3;
            S4;
      end
      S5
end
```

Здесь S_i обозначают операторы. Напомним, что операторы в АЛГОЛе заканчиваются точкой с запятой; ее можно опустить только непосредственно перед **end**.

Программа, сама являющаяся блоком, может содержать несколько подблоков, которые в свою очередь могут содержать подблоки, и т. д. Подробным описанием возможных типов блоков мы не будем здесь заниматься. Отметим лишь, что в больших программах, требующих значительных объемов памяти, блочная структура позволяет эффективно распределить память.

В наших объяснениях не затрагиваются вопросы о структуре операторов ввода-вывода, о считывании данных вычислительной машиной и о выпечатывании результатов или выводе их на дисплей. Это не означает недооценки важности связей машины с внешним миром. Дело в том, что операторы ввода-вывода зависят от характеристик имеющихся устройств (магнитных лент, перфораторов, алфавитно-цифровых печатающих устройств, осциллографов и т. д.). По этой причине процедуры ввода-вывода не были включены в сообщение об АЛГОЛе-60. (Некоторые соглашения о стандартизации этих процедур были достигнуты в 1964 году. Они приведены в книге [10].)

УПРАЖНЕНИЯ Б

1. Написать программы для умножения матрицы

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

а) на вектор-строку $B=[b_{11}, b_{12}, b_{13}]$,
б) на вектор-столбец

$$B=\begin{bmatrix} b_{11} \\ b_{21} \\ b_{31} \end{bmatrix}.$$

2. Написать программу для вычисления суммы элементов матрицы из упр. 1.

3. Написать программу для вычисления значения многочлена $a_0+a_1x+a_2x^2+a_3x^3$ и присваивания этого значения переменной b.

4. Истолковать запись $A:=B+C\uparrow D$ и объяснить, при помощи каких правил АЛГОЛа она строится.

5. Написать программу, умножающую на 6 элементы матрицы A из упр. 1.

6. Написать программу, умножающую на 5 диагональные элементы матрицы A из упр. 1.

7. Написать программу, вычисляющую скалярное произведение $a_1b_1+a_2b_2+a_3b_3$ двух вектор-строк и присваивающую значение результата переменной c.

8. Написать программу для вычисления произведения матриц A и B:
а) в случае когда A, B — вещественные 3×3 матрицы,
*б) в случае когда A, B — матрицы размеров $m\times n$ и $n\times r$ соответственно, $m+n+r\leqslant 100$.

9. Текущее значение вещественной переменной A равно 3.7, а вещественной переменной B равно -6.3.
Объяснить результат выполнения следующих операторов:
а) $A:=B$;
б) $B:=A$;
в) $C:=A$; (где C — целая переменная)
г) $C:=B$; (где C — целая переменная)

10. Перевести следующие записи на АЛГОЛ (← означает «...присваивается значение...»):

а) $z\leftarrow e^{|x\cdot y|\cdot 1/b}$.

б) $F\leftarrow \operatorname{sh} x^2+\operatorname{ch}^2 y$ (предполагается, что готовых подпрограмм для вычисления sh и ch нет).

в) $Y\leftarrow \sum_{k=1}^{N} \cos^k(x^k)/N$ (предполагается, что целой переменной N ранее уже было присвоено положительное значение из P).

г) $h\leftarrow \begin{cases} 1, & \text{если } a,\ b\neq 0 \text{ и } a,\ b \text{ имеют одинаковые знаки,} \\ 0 & \text{в противном случае.} \end{cases}$

*11. Предположим, что имеющийся компилятор не допускает символов элементарных функций sin, cos, abs, ↑, exp и т. д.; в программе разрешается использовать только символы арифметических операций $+$, $-$, $\times$, $/$. Нужно написать программу для вычисления $\cos x$ с точностью до пяти десятичных знаков при $0\leqslant|x|\leqslant 1$.
а) Проанализировать эту задачу математически.
б) Реализовать результат анализа в программе.

4.7. ГРАММАТИКА АЛГОЛа

Создатели АЛГОЛа предложили полное формальное описание правил составления программ на этом языке. Их цель была подобна цели специалистов по математической логике, которые пытались формализовать математические рассуждения таким образом, чтобы подходящая машинная программа могла решать вопрос, является ли каждое данное математическое утверждение осмысленным и, по возможности, истинным или ложным (см. § 14.5).

Обсуждение любого языка L_1 (его «грамматики» и «синтаксиса») по необходимости ведется также на некотором языке, который

называется метаязыком L_0; как правило, метаязык является языком «более высокого уровня». В этом параграфе мы опишем некоторые свойства АЛГОЛа на метаязыке математической лингвистики, а также частично опишем сам этот метаязык.

Определение. *Языком* L называется некое множество строк конечной длины, состоящих из элементов конечного множества V, называемого *алфавитом* языка L.

Множество всех конечных строк, составленных из элементов множества V, обозначается V^*. Таким образом, для математической лингвистики язык с алфавитом V есть некоторое подмножество $L \subset V^*$. Любые две строки $\mathbf{a} = a_1 a_2 \cdots a_m$ и $\mathbf{b} = b_1 b_2 \cdots b_n$ из V^* можно соединить в одну, выписав их элементы подряд:

$$\mathbf{s} = \mathbf{a} * \mathbf{b} = a_1 a_2 \ldots a_m b_1 b_2 \ldots b_n.$$

Операция $*$ называется конкатенацией, или соединением.

Очевидно, она ассоциативна:

$$(\mathbf{a} * \mathbf{b}) * \mathbf{c} = \mathbf{a} * (\mathbf{b} * \mathbf{c}) \text{ для всех } \mathbf{a}, \mathbf{b}, \mathbf{c} \in V^*.$$

Алгебраическая система $[V^*, *]$ замкнута относительно этой операции. Она доставляет образец *полугрупп*, составляющих предмет гл. 7. (В нашем примере $[V^*, *]$ есть *свободная полугруппа, порожденная множеством* V.)

Такое определение языка как произвольного подмножества множества V^* чересчур общо, чтобы быть полезным. Например, в языках программирования множество операторов должно быть определено четкими правилами; анализ выражений на языке также обычно проводится по явным правилам. Математическая лингвистика занимается языками, которые рекурсивно *порождаются* специальными *правилами порождения*, составляющими *синтаксис* языка.

Пример 6. Пусть $V = \{A, B, a, b\}$. Определим L следующими правилами порождения:

(i) $A \in L$ и $B \in L$.

(ii) Если $q \in L$ и $r \in \{a, b\}$, то $q * r \in L$.

Очевидно, L состоит из всех строк, начинающихся с прописной буквы A или B, вслед за которой идут строчные буквы a, b в конечном числе.

Определение. *Грамматикой* языка L называется конечное множество правил, позволяющее рекурсивно породить множество L всех допустимых (осмысленных) выражений (строк) языка.

Язык программирования L состоит из таких строк символов, из которых можно составить программы. Одна из функций компилятора состоит в обнаружении синтаксических ошибок в строках, предположительно подходящих для составления программ, и

в выдаче информации об этом программисту, что облегчает задачу исправления ошибок. (Иногда компиляторы способны сами исправлять некоторые ошибки и сигнализировать о сделанном исправлении.)

Метаязык описания грамматики АЛГОЛа сам в значительной мере символизирован. В нем приняты следующие обозначения:

$\langle X \rangle$	означает «объект типа X»,
$::=$	означает «является»,
$\mid$	означает «или».

Запись в метаязыке

$$\langle \text{цифра} \rangle ::= 0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9$$

означает «цифрой в АЛГОЛе является 0 или 1 или ... или 9». Аналогично буквы определяются метаязыковым выражением

$$\begin{aligned}\langle \text{буква} \rangle ::= {} & a \mid b \mid c \mid d \mid e \mid f \mid g \mid h \mid i \mid j \mid k \mid l \mid m \mid n \\ & \mid o \mid p \mid q \mid r \mid s \mid t \mid u \mid v \mid w \mid x \mid y \mid z \mid A \mid B \mid C \mid D \mid E \\ & \mid F \mid G \mid H \mid I \mid J \mid K \mid L \mid M \mid N \mid O \mid P \mid Q \mid R \mid S \mid T \\ & \mid U \mid V \mid W \mid X \mid Y \mid Z\end{aligned}$$

Этот список состоит из 52 прописных и строчных букв.

Запись в метаязыке

$$\langle \text{арифметическая операция} \rangle ::= + \mid - \mid \times \mid / \mid \div \mid \uparrow$$

перечисляет символы арифметических операций, а запись

$$\langle \text{логическая константа} \rangle ::= \textbf{true} \mid \textbf{false}$$

перечисляет два возможных логических значения.

Все эти записи даны в так называемой *нормальной форме Бэкуса* (или Бэкуса—Наура). В этой форме имя общего объекта определяемого типа пишется слева от символа $::=$ в угловых скобках, а справа от этого символа класс таких объектов либо задается списком (с разделителем $\mid$), либо именами объектов ранее определенных типов в сочетаниях, позволяющих рекурсивно определять новые классы через старые.

Вот пример рекурсивного определения в нормальной форме Бэкуса:

$$\begin{aligned}\langle \text{идентификатор} \rangle ::= {} & \langle \text{буква} \rangle \mid \langle \text{идентификатор} \rangle \\ & \langle \text{буква} \rangle \mid \langle \text{идентификатор} \rangle \langle \text{цифра} \rangle\end{aligned}$$

Смысл этой записи состоит в определении класса объектов, называемых *идентификаторами*. Сначала утверждается, что всякий элемент определенного ранее класса «буква» является идентификатором, затем даны правила порождения, состоящие в том, что приписывание к идентификатору буквы или цифры справа снова дает идентификатор.

Например, A есть идентификатор; приписывая к нему справа B, получаем идентификатор AB; аналогично ABC, $A1$, $AB1$, $A221$ и $A1R1$ допускаются в качестве идентификаторов. Результат применения этих правил можно описать, просто сказав, что идентификатором может быть любая строка, начинающаяся с буквы (слева), вслед за которой идет любая последовательность букв или цифр. Определение идентификатора рекурсивно; последовательное применение правил порождения приведет к бесконечному множеству строк, которые могут быть идентификаторами.

Вот определение основного символа языка:

⟨основной символ⟩::= ⟨буква⟩ | ⟨цифра⟩ |
⟨логическое значение⟩ | ⟨ограничитель⟩.

Первые три класса справа уже были описаны. Ограничители определяются так:

⟨ограничитель⟩::= ⟨операция⟩ | ⟨разделитель⟩ | ⟨описатель⟩
⟨операция⟩::= ⟨арифметическая операция⟩ |
⟨операция следования⟩
⟨операция следования⟩::= **go to** | **if** | **then** | **else** | **for** | **do**
⟨разделитель⟩::= $\hat{}_{10}$ | : | ; | , | · | := | **step** | **until** | **while**
⟨описатель⟩::= **integer** | **real** | **array**

Часть этих понятий имеет отношение к логике; мы вернемся к ним в гл. 5.

Теперь мы можем дать формализованное описание чисел в АЛГОЛе, приведенных в § 4.2. Эта формализация содержится в табл. 4.1. Заметим, что наши правила не объясняют «что такое» числа — они просто являются инструкцией для формирования допустимых строк символов, называемых *числами*.

Рекурсивно применяя правила из табл. 4.1, мы можем породить все целые числа и все десятичные дроби конечной длины.

Таблица 4.1

Числа: синтаксис

⟨целое без знака⟩::=⟨цифра⟩ | ⟨целое без знака⟩ ⟨цифра⟩
⟨целое⟩::=⟨целое без знака⟩ | +⟨целое без знака⟩ | −⟨целое без знака⟩
⟨десятичная дробь⟩::=⟨целое без знака⟩
⟨десятичный порядок⟩::=$\hat{}_{10}$⟨целое⟩
⟨десятичное число⟩::=⟨целое без знака⟩ | ⟨десятичная дробь⟩ | ⟨целое без знака⟩ ⟨десятичная дробь⟩
⟨число без знака⟩::=⟨десятичное число⟩ | ⟨десятичный порядок⟩ | ⟨десятичное число⟩ ⟨десятичный порядок⟩
⟨число⟩::=⟨число без знака⟩ | +⟨число без знака⟩ | −⟨число без знака⟩

Однако любая машина, имея конечную память, не в состоянии представить точно иррациональные числа, вроде $\sqrt{2}$ или π. Такие числа могут быть представлены в ней лишь *приближенно*, своими округленными значениями, которые получаются отбрасыванием всех цифр, кроме конечного числа. В типичном случае двоичная ячейка для хранения числа содержит от 32 до 64 двоичных разрядов, поэтому записанное в ней число с «одинарной точностью» содержит не более 15 десятичных знаков. (Например, ячейки памяти IBM 360-65 состоят из 32 двоичных разрядов, а CDC 6500 — из 60. Однако IBM 85 может производить вычисления над 112-битовыми операндами и 8-битовыми порядками.) Для вычислений большого объема, где нужна большая точность, чтобы предотвратить накопление ошибок, были написаны компиляторы, позволяющие вычислять с двойной точностью. Допустимая длина двоичного слова удвоена. Иногда проводятся вычисления с еще большей точностью и хранением еще большего количества значащих цифр.

В табл. 4.2—4.4 перечислены синтаксические правила порождения арифметических выражений, операторов цикла и операто-

Таблица 4.2

Арифметические выражения: синтаксис

⟨переменная⟩::=⟨идентификатор⟩
⟨операция типа сложения⟩::=+ | −
⟨операция типа умножения⟩::=× | / | ÷
⟨первичное выражение⟩::=⟨число без знака⟩ | ⟨переменная⟩ | ⟨арифметическое выражение⟩
⟨множитель⟩::=⟨первичное выражение⟩ | ⟨множитель⟩ ↑ ⟨первичное выражение⟩
⟨терм⟩::=⟨множитель⟩ | ⟨терм⟩ ⟨операция типа умножения⟩ ⟨множитель⟩
⟨простое арифметическое выражение⟩::=⟨терм⟩ | ⟨операция типа сложения⟩ ⟨терм⟩ | ⟨простое арифметическое выражение⟩ ⟨операция типа сложения⟩⟨терм⟩
⟨условие⟩::=**if** ⟨булево выражение⟩ **then**
⟨арифметическое выражение⟩::=⟨простое арифметическое выражение⟩ | ⟨условие⟩⟨простое арифметическое выражение⟩ **else** ⟨арифметическое выражение⟩

Таблица 4.3

Операторы цикла: синтаксис

⟨элемент списка цикла⟩::=⟨арифметическое выражение⟩ | ⟨арифметическое выражение⟩ **step** ⟨арифметическое выражение⟩ **until** ⟨арифметическое выражение⟩ | ⟨арифметическое выражение⟩ **while** ⟨булево выражение⟩
⟨список цикла⟩::=⟨элемент списка цикла⟩ | ⟨список цикла⟩, ⟨элемент списка цикла⟩
⟨заголовок цикла⟩::=**for** ⟨переменная⟩:=⟨список цикла⟩ **do**
⟨оператор цикла⟩::=⟨заголовок цикла⟩ ⟨оператор⟩

Таблица 4.4

Операторы присваивания: синтаксис

⟨левая часть⟩::=⟨переменная⟩:=
⟨список левой части⟩::=⟨левая часть⟩ | ⟨список левой части⟩ ⟨левая часть⟩
⟨оператор присваивания⟩::=⟨список левой части⟩ ⟨арифметическое выражение⟩

ров присваивания соответственно. Нормальная форма Бэкуса позволяет ввести глубокую рекурсивность. Эти синтаксические правила исчерпывают все типы выражений, которые мы рассматривали.

4.8. ВЫЧИСЛЕНИЕ АРИФМЕТИЧЕСКИХ ВЫРАЖЕНИЙ

В этом параграфе будет объяснено понятие магазинной памяти и его использование в алгоритмах вычисления значений простых арифметических выражений на АЛГОЛе. В следующем параграфе мы покажем, как это же понятие используется в компиляторе при трансляции с АЛГОЛа на язык ассемблера.

Трансляция и вычисление облегчаются тем, что порядок выполнения арифметических операций однозначно определен синтаксисом АЛГОЛа. Напомним, что (при отсутствии скобок) операция ↑ выполняется прежде всех; следующие по порядку операции — это /, ÷, × и затем +, —; внутри этих двух групп операции равноправны и выполняются в порядке написания слева направо. Эти соглашения позволяют указать простой алгоритм для вычисления арифметических выражений.

Для простоты ограничимся выражениями, составленными из однобуквенных переменных, выписанных выше арифметических операций и скобок. Мы будем рассматривать «—» только как бинарную операцию, запретив выражения типа $-b+a$. Нетрудно было бы справиться и с обычным определением «минуса» как унарной и бинарной операции, но эти детали мы опустим.

Процесс вычисления использует *магазинную память*, или *стек*, работающую по принципу «последним пришел — первым обслужен». Подробнее: магазин состоит из упорядоченного множества ячеек памяти. Эти ячейки можно представлять себе в виде вертикальной колоды карт, помещенной на ограничитель. Начальное состояние магазина — когда он пуст, о чем свидетельствует ограничитель. Первая запись помещается непосредственно на ограничитель, а затем ячейки заполняются последовательно одна за другой. При считывании первым читается последний записанный символ, затем предпоследний и. т. д. Считывание уничтожает прочитанный символ. Например, если магазин $S1$ сначала был пустым,

затем мы ввели в него число 16, затем 36, затем произвели считывание, то считано будет число 36. Следующее считывание дает 16. Если мы введем в пустой магазин 6, затем 26 и считаем одно число, это будет 26. Если мы после этого введем 42, 34 и считаем три числа, это будут 34, 42, 6 в данном порядке.

Для алгоритма вычисления арифметических выражений в АЛГОЛе мы воспользуемся двумя магазинами: $S1$ и $S2$. Значения переменных будут записываться в $S1$, а символы операций в $S2$.

АЛГОРИТМ ВЫЧИСЛЕНИЯ АРИФМЕТИЧЕСКИХ ВЫРАЖЕНИЙ

1. Считывать входную строку слева направо.
2. Записать первое считанное числовое значение в $S1$, а первый символ операции в $S2$.
3. Каждое очередное считываемое числовое значение записывать в $S1$.
4. При считывании очередных символов операций поступать так:

а) если очередная операция по правилам порядка действий должна выполняться до операции, записаной последней в $S2$, то записать ее в следующую ячейку $S2$;

б) иначе (если они равноправны или новая менее важная) применить операцию, записанную в $S2$ последней, к соответствующим операндам (два последних элемента из $S1$). Стереть примененную операцию из $S2$, стереть использованные операнды из $S1$. Записать результат в $S1$ и повторить шаг 4.

5. После считывания последнего числового значения входной строки:

а) если магазин $S2$ пуст, алгоритм окончен;

б) иначе считать операцию из $S2$, применить ее к двум операндам из $S1$ и результат записать в $S1$ (напомним, что считывание из магазинной памяти автоматически стирает считанный символ).

в) повторять шаг 5 до опустошения $S2$.

В результате выполнения программы в магазине $S1$ окажется значение данного арифметического выражения. На самом деле тот же алгоритм позволяет вычислять любые выражения с бинарными операциями, для которых установлен порядок действий в любой алгебраической системе, например в булевой алгебре.

Рассмотрим, как будет происходить вычисление выражения $a\times b\times c+d\times e\times f$. Прежде всего считывается a и записывается в $S1$; затем считывается $\times$ и записывается в $S2$. После этого b записывается в $S1$ и считывается следующий символ $\times$. Сравнивая его с хранящимся в $S2$ символом $\times$, получаем, что хранящийся $\times$ следует применить к a, b в $S1$ и записать $a\times b$ в $S1$. Считанный вторым $\times$ помещается в $S2$, затем c помещается в $S1$.

К этому моменту состояние памяти таково:

$S1$	$S2$
c	$\times$
$b\times a$	

Считываемый затем символ $+$ уступает по степени важности хранящемуся $\times$. Операция $\times$ из $S2$ поэтому применяется к $a\times b$ и c и результат $c\times b\times a$ записывается в $S1$. После этого $+$ помещается в $S2$, состояние памяти будет таким:

$S1$	$S2$
$a\times b\times c$	$+$

(мы написали $a\times b\times c$ вместо $c\times b\times a$ по коммутативности).

Теперь d записывается в $S1$, считывается $\times$ и записывается в $S2$, считывается $e\ldots$. После записи $d\times e$ в $S1$ получим:

$S1$	$S2$
$d\times e$	$+$
$a\times b\times c$	

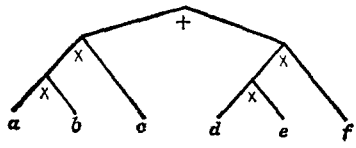
После считывания последнего $\times$ последнее число в $S1$ умножается на f и результат записывается в $S1$. Память к этому моменту такова:

$S1$	$S2$
$d\times e\times f$	$+$
$a\times b\times c$	

Поскольку $S2$ не пуст, хранящаяся в нем операция $+$ применяется к двум числам из $S1$, что дает окончательно

$S1$	$S2$
$a\times b\times c+d\times e\times f$	

Таким образом, вычисление управляется ходом «синтаксического разбора» выражения $a\times b\times c+d\times e\times f$, который можно изобразить в виде дерева



Синтаксические правила АЛГОЛа и правила чтения обеспечивают однозначность разбора.

*4.9. ТРАНСЛЯЦИЯ АРИФМЕТИЧЕСКИХ ВЫРАЖЕНИЙ

Мы описали алгоритм для вычисления значений индивидуальных арифметических выражений в АЛГОЛе. Теперь объясним, как используется этот алгоритм при работе компилятора, предназначенного для перевода любого арифметического выражения на АЛГОЛе в программу на языке ассемблера для вычисления этого выражения.

Будем пользоваться четырьмя командами языка ассемблера, введенными в § 4.1:

CLA X	Значение, хранящееся в ячейке X памяти, записать в сумматор.
ADD X	Значение, хранящееся в ячейке X памяти, сложить с текущим значением сумматора и записать результат в сумматор.
MUL X	Значение, хранящееся в ячейке X памяти, умножить на текущее значение сумматора и записать результат в сумматор.
STO X	Записать текущее значение сумматора в ячейку X памяти.

Ограничимся двумя бинарными операциями $+$ и $\times$ и определим для них две *генерирующие программы*:

Операция	Генерирующая программа	Пояснения
$+$	CLA X	X есть последний адрес, записанный в $S1$;
	ADD Y	Y — предпоследний адрес, записанный в $S1$;
	STO Z	эти адреса стираются и вместо них записывается адрес Z.
$\times$	CLA X MUL Y STO Z	то же самое.

Трансляция производится в соответствии с данным выше описанием алгоритма вычисления со следующими изменениями. Во-первых, компилятор считывает не значения переменных, а их буквенные символы, присваивает каждой из них свой адрес, и магазин хранит эти адреса. В дальнейшем мы будем обозначать через (A) адрес переменной A. Во-вторых, как только операция должна быть использована, компилятор добавляет к списку магазинных команд соответствующую генерирующую программу.

Опишем, как транслируется выражение $A+B\times C$. Считывается A, ей присваивается адрес (A) и запоминается в $S1$. Считывается $+$ и записывается в $S2$. Считывается B, присваивается адрес (B), считывается $\times$ и C. Состояние памяти к этому моменту:

$S1$	$S2$
(C)	$\times$
(B)	$+$
(A)	

При вычислении операция $\times$ выполнялась бы перед $+$. Поэтому транслятор выдает соответствующую генерирующую программу в машинных кодах:

```
CLA (C)
MUL (B)
STO (D)
```

Здесь (D) — некоторый новый адрес, заменивший (C) и (B) в магазине, который теперь находится в состоянии

$S1$	$S2$
(D)	$+$
(A)	

Теперь добавляется генерирующая программа для $+$; окончательно

```
CLA (C)
MUL (B)
STO (D)
CLA (D)
ADD (A)
STO (E)
```

Если значения A, B, C хранились по адресам (A), (B), (C), то после выполнения этой программы значение $A+B\times C$ будет храниться по адресу (E).

Добавление генерирующей программы для остальных арифметических операций дает возможность вычислять значения всевозможных арифметических выражений, записанных без скобок. Правила работы со скобками таковы: (записывается в $S2$, как только она считана;) предписывает исполнять все операции, хранящиеся в $S2$, пока из $S2$ не будет считана (.

Мы описали очень простой пример работы компилятора: синтаксис языка используется для последовательного выяснения того, какие генерирующие программы следует добавлять к списку машинных команд, который в конце концов станет программой

в машинных кодах. В действительности работа компилятора и его структура крайне сложны, и мы лишь проиллюстрировали некоторые заложенные в них принципы.

Правила порядка арифметических действий в отсутствие скобок позволяет без труда транслировать арифметические выражения. Однако с другими классами выражений в АЛГОЛе и вообще в контекстно-свободных языках дело обстоит не так просто. Для перевода очередной строки в машинные коды необходимо определить ее принадлежность к тому или иному синтаксическому классу. Часто не существует прямого способа сделать это. Предположение о том, что некоторая часть данной строки принадлежит к данному синтаксическому классу, должно проверяться изучением соседних частей и выяснением, позволяет ли это предположение провести последовательный синтаксический разбор всей строки. Существуют разные подходы к решению таких задач. Компилятор может осуществлять полный перебор или же перебор по схеме типа дерева, запоминая каждое ответвление и возвращаясь к нему, если очередная попытка синтаксического разбора оказалась неудачной. За дальнейшим обсуждением сложных проблем, возникающих здесь, мы отсылаем читателя к книгам [2] и [5]

УПРАЖНЕНИЯ В

1. Покажите, как порождается выражение $A:=B+C\times D$ правилами грамматики АЛГОЛа, записанными в нормальной форме Бэкуса.

2. Покажите, как запись 10.98 порождается правилами из табл. 4.1.

3. Объясните структуру оператора $A:=B\times 3$, пользуясь табл. 4.4.

СПИСОК ЛИТЕРАТУРЫ

1. Bottenbruch H., Structure and Use of ALGOL 60, *J. ACM*, 9: 161—221 (1962). (Русский перевод: Боттенбрух Х., Структура АЛГОЛ-60 и его использование, М., ИЛ, 1963.)
2. Chomsky N., Syntactic Structures, Mouton and Co., 1957.
3. Dijkstra E. W., A Primer of ALGOL 60 Programming, Academic Press, 1962.
4. Floyd R. W., The Syntax of Programming Languages — a Survey in Programming Systems and Languages, Saul Rosen (ed.), McGraw-Hill, 1967.
5. Ginsburg S., The Mathematical Theory of Context-free Languages, McGraw-Hill, 1966. (Русский перевод: Гинсбург С., Математическая теория контекстно-свободных языков, М., «Мир», 1975.)
6. Knuth D., The Art of Computer Programming, 9 vols., Addison-Wesley. (Русский перевод: Кнут Д., Искусство программирования для ЭВМ, I том, М., «Мир», 1976.)
7. McCracken D. D., A Guide to ALGOL Programming, Wiley, 1962. (Русский перевод: Маккракен Д., Программирование на АЛГОЛе, М., «Мир», 1964.)
8. Naur P., A Course of ALGOL 60 Programming, Regnecentralen 1961.
9. Nicol K., Elementary Programming and ALGOL, McGraw-Hill, 1968.
10. Rutishauser H., Description of ALGOL 60, Springer 1967.

5

БУЛЕВЫ АЛГЕБРЫ

5.1. ВВЕДЕНИЕ

В § 1.2 было неформально введено понятие булевой алгебры и разобраны булевы алгебры подмножеств данного множества. Для удобства мы начнем с формального определения. Бо́льшая часть двух следующих глав посвящена его систематическим приложениям.

Определение. *Булевой алгеброй* $B=[A, \wedge, \vee, ', O, I]$ называется множество A с двумя бинарными операциями $\wedge$, $\vee$, двумя отмеченными элементами («универсальными границами») O, I и одной унарной операцией $'$, причем для любых $x, y, z \in A$

L1. $x \wedge x = x, \quad x \vee x = x$ (идемпотентность),

L2. $x \wedge y = y \wedge x, \quad x \vee y = y \vee x$ (коммутативность),

L3. $x \wedge (y \wedge z) = (x \wedge y) \wedge z$, $x \vee (y \vee z) = (x \vee y) \vee z$ (ассоциативность),

L4. $x \wedge (x \vee y) = x, \quad x \vee (x \wedge y) = x$ (поглощение),

L5а. $x \wedge [y \vee (x \wedge z)] = (x \wedge y) \vee (x \wedge z)$

L5б. $x \vee [y \wedge (x \vee z)] = (x \vee y) \wedge (x \vee z)$ (модулярность),

L6. $x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z)$, $x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z)$ (дистрибутивность),

L7. $x \wedge O = O, \quad x \vee O = x$, $x \wedge I = x, \quad x \vee I = I$ (универсальные границы),

L8. $x \wedge x' = O, \quad x \vee x' = I$ (дополнение),

L9. $(x')' = x$ (инволютивность),

L10. $(x \wedge y)' = x' \vee y', \quad (x \vee y)' = x' \wedge y'$ (закон де Моргана).

Обратите внимание на аксиоматический характер этого определения. Мы считаем заданными пять операций на множестве A (две бинарные, одна унарная, две 0-арных) и перечисляем 21 тождество, сгруппированные в десять аксиом, или постулатов. Алгебраическая система является булевой алгеброй тогда и только тогда, когда она имеет такой набор операций и этот набор удов-

летворяет всем выписанным аксиомам. (Разумеется, способ обозначений этих операций не играет роли, но выбранные нами символы сейчас общеупотребительны.)

Операции $\wedge$ и $\vee$ называются соответственно булевым произведением, или пересечением, и булевой суммой, или объединением. Аналогично называются результаты применения этих операций.

Простейшая нетривиальная булева алгебра описана в примере 1 § 1.2. Напомним его.

Пример 1. *Двухэлементная* булева алгебра $\mathbf{B}=[\{0, 1\}, \wedge, \vee, ', 0, 1]$ имеет следующие операции: $\wedge$ — *«наименьший из...»*, $\vee$ — *«наибольший из...»*, $'$ — *«другой элемент»*. Например, $0\wedge 1=0$, $0\vee 1=1$, $0'=1$.

Тривиальная булева алгебра имеет вид $[\{0\}, \wedge, \vee, ', 0, 0]$ с операциями

$$0\wedge 0=0\vee 0=0'=0$$

и $O=0$, $I=0$. Все тождества L1—L10 выполняются по тривиальной причине: и левые, и правые их части равны 0.

В этой главе мы будем заниматься в основном *конечными* булевыми алгебрами. В гл. 9 мы покажем, что все они (с точностью до изоморфизма) исчерпываются следующим классом примеров (см. теорему 1 гл. 1).

Пример 2. Для любого положительного целого числа n булева алгебра $\mathbf{B}^n=[\mathcal{P}(\mathbf{n}), \wedge, \vee, ', \varnothing, \mathbf{n}]$, содержащая 2^n *элементов*, состоит из всех подмножеств множества $\mathbf{n}=\{1, \ldots, n\}$. Операции $\wedge$, $\vee$, $'$ суть соответственно теоретико-множественное пересечение, объединение и дополнение. Универсальные границы суть пустое множество $\varnothing$ и все множество $I=\mathbf{n}$.

Доказывая теоремы о произвольных булевых алгебрах, следует пользоваться только тождествами L1—L10 и их следствиями. Лишь при этом условии мы можем быть уверены в применимости наших результатов ко всем булевым алгебрам, независимо от того, состоят ли они из множеств, логических высказываний или математических объектов, описывающих электронные схемы, вроде вентильных схем из § 5.4. Такой аксиоматический подход является стандартным в современной алгебре; мы полностью придерживаемся его в этой главе.

Список из 21 тождества L1—L10 очень избыточен. Например, аксиомы L1, L5, L9 и L10 следуют из остальных шести. Покажем это сначала для L1.

Лемма 1 (Дедекинд). *Законы идемпотентности* L1 *следуют из законов поглощения* L4.

Доказательство. Положив $y = x \wedge x$ в первом тождестве $L4$, мы получим

$$x \wedge [x \vee (x \wedge x)] = x.$$

Из второго тождества L4 (с $y = x$) следует, что выражение в квадратных скобках равно x. Значит, $x \wedge [x \vee (x \wedge x)] = x \wedge x$. Поэтому $x = x \wedge x$. Тождество $x = x \vee x$ доказывается аналогично, если в предыдущем рассуждении заменить все $\wedge$ на $\vee$ и наоборот.

Последнее замечание является частным случаем следующего метаматематического принципа двойственности, тесно связанного с принципом двойственности для отношений частичного порядка, определенного в § 2.5. (Метаматематический результат — это «теорема о теоремах».)

Теорема 1. *Любая общезначимая теорема о булевых алгебрах, в формулировке которой участвуют только операции* $\wedge$, $\vee$ *и* $'$, *остается общезначимой, если в ее формулировке всюду заменить* $\wedge$ *на* $\vee$ *и наоборот.*

Доказательство. Множество аксиом L1—L10 в целом не меняется при замене $\wedge$ на $\vee$ и $\vee$ на $\wedge$. Поэтому при такой замене всякое доказательство превращается в доказательство двойственного утверждения.

Покажем теперь, что избыточна аксиома L5. Нам этот результат не понадобится до гл. 9, и здесь мы приводим его лишь для полноты.

Лемма 2. *Аксиома* L5 *следует из* L1—L4 *и* L6.

Доказательство. Достаточно установить первое тождество L5 и воспользоваться принципом двойственности. Пользуясь последовательно L6, L3, L4 и L6, получаем

$$x \wedge [y \vee (x \wedge z)] = x \wedge [(y \vee x) \wedge (y \vee z)] = [x \wedge (y \vee x)] \wedge (y \vee z) = \\ = x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z).$$

В гл. 1 мы показали, что для любого множества X множество 2^X всех функций из X в $2 = \{0, 1\}$ образует булеву алгебру. Мы сейчас подробно исследуем алгебру, которая получается таким способом исходя из множества $X = 2^2$.

Пример 3. Пусть F_2 — множество всех $2^4 = 16$ функций $f : \{00, 01, 10, 11\} \to \{0, 1\}$. В обозначениях гл. 1 это — множество всех функций $f : 2^2 \to 2$; оно само изоморфно булевой алгебре 2^4. В каждой точке $\mathbf{x} \in X = 2^2$ области функции f «дополнение» f' каждой функции f принимает единственное возможное значение, отличное от значения f:

$$f'(\mathbf{x}) = 1 - f(\mathbf{x}) \quad \text{для всех } \mathbf{x} \in 2^2. \tag{1}$$

Знак «минус» понимается в обычном смысле. Булева сумма $e \vee f = j$ функций e, f в каждой точке равна наибольшему из значений e, f в этой точке, булево произведение $e \wedge f = m$ — наименьшему из значений:

$$j(\mathbf{x}) = \max\{e(\mathbf{x}), f(\mathbf{x})\}, \qquad m(\mathbf{x}) = \min\{e(\mathbf{x}), f(\mathbf{x})\}. \tag{2}$$

В частичном списке элементов F_2 в табл. 5.1 даны два примера дополнений и четыре примера булевых сумм и произведений. Кроме перечисленных в этой таблице функций, в F_2 входят постоянные функции 0, 1 и функции α, β, заданные следующей таблицей.

Таблица 5.1

Частичный список элементов F_2

	g	h	g'	h'	$g \wedge h$	$g \wedge h'$	$g' \wedge h$	$g' \wedge h'$	$g \vee h$	$g \vee h'$	$g' \vee h$	$g' \vee h'$
00	0	0	1	1	0	0	0	1	0	1	1	1
01	0	1	1	0	0	0	1	0	1	0	1	1
10	1	0	0	1	0	1	0	0	1	1	0	1
11	1	1	0	0	1	0	0	0	1	1	1	0

	0	1	α	β
00	0	1	0	1
01	0	1	1	0
10	0	1	1	0
11	0	1	0	1

$$\alpha = (g \wedge h') \vee (g' \wedge h) = (g \vee h) \wedge (g' \vee h'),$$
$$\beta = (g \wedge h) \vee (g' \wedge h') = (g \vee h') \wedge (g' \vee h) = \alpha'.$$

Проверка тождеств для α, β доставляет следующую лемму:

Лемма 2. *Булева алгебра F_2 порождена функциями g и h из таблицы* 5.1 *в том смысле, что любой элемент F_2 может быть получен из g, h применением булевых операций к этим образующим.*

В § 5.10 мы убедимся, что F_2 есть «свободная» булева алгебра с двумя образующими, т. е. всякая булева алгебра с двумя образующими является *эпиморфным образом* F_2.

Булево выражение (булев полином)

$$\alpha = (g \wedge h') \vee (g' \wedge h) = g + h \tag{3}$$

называется *симметрической разностью* элементов g и h. Мы обозначаем его символом $\beta(g, h) = g + h$ потому, что значение $g + h$ в любой точке равно *сумме значений g и h по модулю два* в этой точке. Это общий факт: если S, T — подмножества любого множества U с характеристическими функциями e_S, e_T, то

$$e_{S+T} = e_S + e_T \pmod 2, \text{ где } S + T = (S \cap T') \cup (S' \cap T). \tag{3'}$$

Булеву алгебру F_2 из примера 1 можно также изобразить с помощью кругов Эйлера, как на рис. 5.1.

Эти круги делят плоскость на четыре области; 16 элементов естественно сопоставляются с 16 всевозможными объединениями этих четырех областей (см. упр. А4).

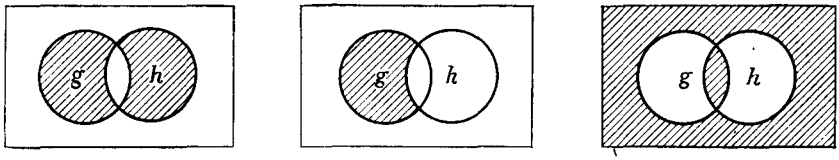


Рис. 5.1. Диаграмма Эйлера для F_2.

5.2. ПОРЯДОК

Аксиомы из § 5.1 для булевых алгебр были написаны в форме *тождеств* относительно трех *операций* $\wedge$, $\vee$, $'$. Существует внешне совершенно другая система аксиом для булевых алгебр, которая формулируется в терминах свойств отношения (частичного) *порядка*. Мы выпишем эту систему аксиом в гл. 9, а здесь лишь укажем важнейшие связи между этими тремя булевыми операциями и отношением порядка.

Интуитивно ясно, что это отношение относится к $\wedge$, $\vee$, $'$ так же, как отношение *включения* подмножеств $S \subset T$ к операциям $S \cap T$, $S \cup T$ и S'.

Например, отношение $S \subset T$ эквивалентно любому из четырех отношений:

$$S \cap T = S, \quad S \cup T = T, \quad S \cap T' = \varnothing, \quad S' \cup T = I.$$

В лемме 2 ниже, исходя только из аксиом L1—L10, мы покажем, что четыре аналогичных отношения эквивалентны в любой булевой алгебре. Начнем доказательство со следующей леммы.

Лемма 1. *В любой булевой алгебре*

$$\text{из } a \wedge x = O \text{ и } a \vee x = I \text{ следует, что } x = a'. \tag{4}$$

Действительно, из посылок (4) последовательно вытекают тождества

$$x = x \wedge I = x \wedge (a \vee a') = (x \wedge a) \vee (x \wedge a') = O \vee (x \wedge a') =$$
$$= (a \wedge a') \vee (x \wedge a') = (a \vee x) \wedge a' = I \wedge a' = a'$$

(воспользоваться L7, L8, L6, посылками, L8, L6, L2, посылками, L7 последовательно). Это доказывает лемму 1.

Полагая $a = I$ и $x = O$ и учитывая, что $I \wedge O = O$ и $I \vee O = I$ в силу L7 и L2, получаем в качестве следствия

$$O' = I, \quad I' = O \text{ в любой булевой алгебре.} \tag{5}$$

Лемма 2. *Во всякой булевой алгебре A следующие четыре соотношения равносильны:*

$$a \wedge b = a, \quad a \vee b = b, \quad a' \vee b = I, \quad a \wedge b' = O. \tag{6}$$

Доказательство. Если $a \wedge b = a$, то $a \vee b = (a \wedge b) \vee b = b$ в силу L4. Далее, если $a \vee b = b$, то

$$a' \vee b = a' \vee (a \vee b) = (a' \vee a) \vee b = I \vee b = I.$$

Если $a' \vee b = I$, то по закону де Моргана L10 находим с помощью (5)

$$a \wedge b' = (a' \vee b)' = I' = O.$$

Наконец, если $a \wedge b' = O$, то

$$a \wedge b = (a \wedge b) \vee O = (a \wedge b) \vee (a \wedge b') = a \wedge (b \vee b') = a \wedge I = a,$$

что завершает цикл импликаций.

Определение. В любой булевой алгебре отношение определяется с помощью любого из соотношений (6).

Это отношение «включения» играет фундаментальную роль, ибо в его терминах можно определить все булевы операции. Это будет полностью доказано в гл. 9. Например, O и I определяются как универсальные границы:

$$O \leqslant a \leqslant I \text{ для всех } a \in A. \tag{7}$$

Действительно, в силу (6) эти соотношения равносильны $O \vee a = a$ и $a \vee I = I$ соответственно, которые выполняются в силу L7 (и L2). (Это доказательство показывает, что два других тождества L7 избыточны.)

Теорема 2. *В любой булевой алгебре $\mathfrak{B} = [A, \wedge, \vee, ']$ отношение $a \leqslant b$ является частичным порядком. Более того, в терминах этого отношения операции $\wedge$ и $\vee$ восстанавливаются так:*

$$\begin{aligned} a \wedge b &= \text{н.н.г. (наибольшая нижняя граница) } \{a, b\}, \\ a \vee b &= \text{н.в.г. (наименьшая верхняя граница) } \{a, b\}. \end{aligned} \tag{8}$$

Доказательство. В силу L1, $a \wedge a = a$, и потому из определения следует, что $a \leqslant a$ для всех $a \in A$. Если $a \leqslant b$ и $b \leqslant a$, то $a \wedge b = a$ и $b \wedge a = b$. Из закона коммутативности L2 и транзитивности равенства получаем $a = b$. Наконец, если $a \leqslant b$ и $b \leqslant c$, то $a \wedge b = a$ и $b \wedge c = b$, откуда, в силу L3, $a = a \wedge b = a \wedge (b \wedge c) = (a \wedge b) \wedge c = a \wedge c$. Значит, $a \leqslant c$, так что $\leqslant$ есть частичный порядок на A.

Остается проверить (8). По теореме 1 (двойственности), из общезначимости любого из тождеств $a \wedge b =$ н.н.г. $\{a, b\}$ и $a \vee b =$ $=$ н.в.г. $\{a, b\}$ следует общезначимость другого. Поэтому достаточно доказать, что $a \wedge b =$ н.н.г. $\{a, b\}$. С этой целью заметим

сначала, что, поскольку в силу L3 и L1, $a\wedge(a\wedge b)=(a\wedge a)\wedge b=a\wedge b$, из (6) следует, что $a\wedge b\leqslant a$. Аналогично, $a\wedge b=(a\wedge b)\wedge b$, откуда $a\wedge b\leqslant b$. Это показывает, что $a\wedge b$ есть некоторая нижняя граница для a и b.

Пусть теперь c—*произвольная* нижняя граница элементов a и b. Используя L3 и предположение относительно c (дважды), находим

$$(a\wedge b)\wedge c=a\wedge(b\wedge c)=a\wedge c=c,$$

откуда $a\wedge b\geqslant c$. Отсюда следует, что $a\wedge b$ является (единственной в силу результатов § 2.4) *наибольшей* нижней границей a и b.

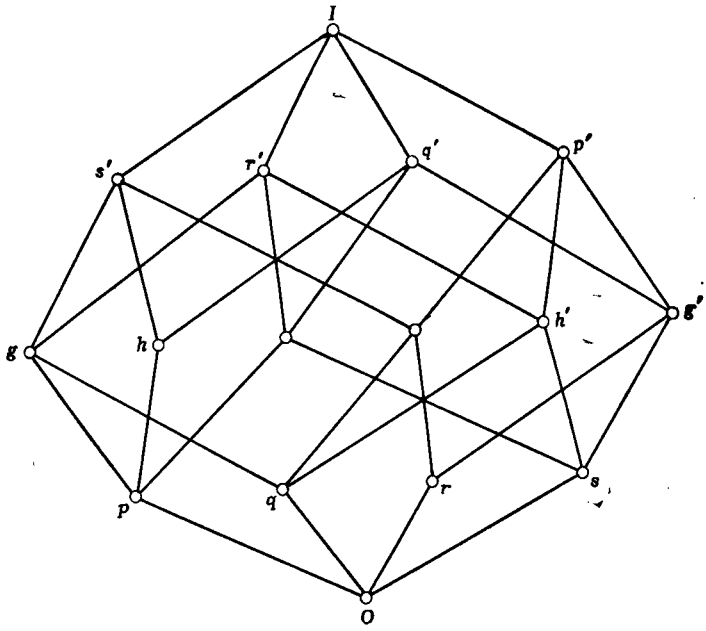


Рис. 5.2. Диаграмма для булевой алгебры 2^4.

Мы можем воспользоваться результатом теоремы 2, чтобы описать булеву алгебру 2^4 примера 3 посредством диаграммы частичного порядка, изображенной на рис. 5.2. В частности, $p=g\wedge h$, $q=g\wedge h'$, $r=g'\wedge h$, $s=g'\wedge h'$.

Лемма 3 (изотонность). *В любой булевой алгебре* A

$$\begin{aligned}&\text{из } b\leqslant c \text{ следует, что } a\wedge b\leqslant a\wedge c \text{ и}\\ &a\vee b\leqslant a\vee c \text{ для всех } a\in A.\end{aligned}\tag{9}$$

Доказательство. Так как $b\leqslant c$, то $b=b\wedge c$, откуда

$$a\wedge b=a\wedge(b\wedge c)=(a\wedge a)\wedge(b\wedge c)=(a\wedge b)\wedge(a\wedge c).$$

Таблица 5.2

Булевы функции от p

$'$		$\wedge$	0	p	p'	1	$\vee$	0	p	p'	1
0	1	0	0	0	0	0	0	0	p	p'	1
p	p'	p	0	p	0	p	p	p	p	1	1
p'	p	p'	0	0	p'	p'	p'	p'	1	p'	1
1	0	1	0	p	p'	1	1	1	1	1	1

Здесь мы пользовались коммутативностью, ассоциативностью и L4. Отсюда, согласно определению, следует, что $a \wedge b \leqslant a \wedge c$. Доказательство $a \vee b \leqslant a \vee c$ получается с помощью двойственного рассуждения.

5.3. БУЛЕВЫ МНОГОЧЛЕНЫ

Булевым многочленом называется всякое выражение, которое можно получить исходя из некоторого множества символов $x_1, \ldots, x_n$ последовательным применением операций $\wedge$, $\vee$ и $'$. В обозначениях Бэкуса (§ 4.9) класс BOOLEPOLY булевых многочленов определяется так:

BOOLEPOLY::=⟨буква⟩ | ⟨BOOLEPOLY⟩∧⟨BOOLEPOLY⟩ |
⟨BOOLEPOLY⟩∨⟨BOOLEPOLY⟩ | ⟨BOOLEPOLY′⟩

Это приводит нас к «правилам порождения» для этого класса алгебраических выражений.

Очевидно, задание булева многочлена $F(x_1, \ldots, x_n)$ определяет для любой булевой алгебры $\mathcal{B}$ функцию $F: \mathcal{B}^n \to \mathcal{B}$ — значения этого многочлена. Это замечание допускает частичное обращение.

Теорема 3. *Каждую функцию $f: 2^n \to 2$ можно представить в виде булева многочлена от координатных функций $\delta_i: \mathbf{x} \to x_i$.*

Доказательство. Поставим в соответствие каждому элементу $\mathbf{v} \in 2^n$ булев многочлен

$$p_{\mathbf{v}}(x_1, \ldots, x_n) = y_1 \wedge \ldots \wedge y_n, \quad \text{где} \quad y_i = \begin{cases} x_i, & \text{если } v_i = 1, \\ x_i', & \text{если } v_i = 0. \end{cases}$$

Он обладает следующим свойством:

$$p_{\mathbf{v}}(\mathbf{v}) = 1 \text{ и } p_{\mathbf{v}}(\mathbf{w}) = 0, \text{ если } \mathbf{w} \neq \mathbf{v} \text{ в } 2^n.$$

Далее, для любой функции $f: 2^n \to 2$ обозначим через $S(f)$ множество всех векторов $\mathbf{v} \in 2^n$, для которых $f(\mathbf{v}) = f(v_1, \ldots, v_n) = 1$.

Рассмотрим функцию

$$F(x_1, \ldots, x_n) = \bigvee_{S(f)} p_{\mathbf{v}}.$$

Согласно приведенному выше вычислению значений $p_{\mathbf{v}}$,

$$F(\mathbf{w}) = \begin{cases} 1, & \text{если } f(\mathbf{w}) = 1, \\ 0, & \text{если } f(\mathbf{w}) = 0, \end{cases}$$

и мы получили требуемое заключение.

Булево *тождество* $F(x_1, \ldots, x_n) = G(x_1, \ldots, x_n)$ есть равенство, справедливое для любых значений $x_1, \ldots, x_n$ в произвольной булевой алгебре. Аксиомы L1—L10 доставляют список из 21 такого тождества, выполняющегося по определению в любой булевой алгебре. Поэтому множество выражений, порождаемых BOOLEPOLY, *избыточно*. Одна из основных целей этой главы состоит в том, чтобы построить систематический алгоритм для приведения каждого булева многочлена F к простому *каноническому виду* $\tilde{F}$. При этом F и G будут представлять одну и ту же функцию из $\mathcal{B}^n$ в $\mathcal{B}$ для всех булевых алгебр $\mathcal{B}$ тогда и только тогда, когда $\tilde{F}$ и $\tilde{G}$ совпадают. Этот алгоритм тесно связан с конструкцией, использованной в доказательстве теоремы 3.

Таким образом, мы опишем систематическую процедуру для определения за конечное число шагов, является ли тождеством любое соотношение вида $F = G$, где F, G—булевы многочлены, т. е. совпадают ли F и G как функции на любой булевой алгебре. Общая задача такого типа для разнообразных алгебраических систем называется *проблемой тождества*, и мы дадим ее решение для булевых алгебр в этой главе.

Таблица 5.2 решает эту задачу для многочленов от одной переменной. Она показывает, что любой такой многочлен от единственной переменной p тождествен либо p, либо p', либо O, либо I. Существует ровно четыре булевых функции одной переменной. Нетрудно проверить эту таблицу, пользуясь аксиомами: таблица дополнений следует из (5) и L9, диагональные элементы таблиц для $\wedge$ и $\vee$ считываются из L1, а остальные элементы—из L7 и L8.

Аналогично, круги Эйлера на рис. 5.1 позволяют предположить, что имеется $2^4 = 16$ различных булевых функций от двух переменных; то же подсказывает таблица из 5.1. Так оно и есть.

В табл. 5.3 эти функции записаны в лексикографическом порядке строк, представляющих их булевы значения. Пользуясь ею, мы получаем простое решение проблемы тождества для $n = 2$. Оказывается, что проверка соотношения $F = G$ сводится к вычислениям F и G как функций на булевой алгебре из примера 1. Многочлены F и G тождественны тогда и только тогда,

Таблица 5.3.

Канонические формы для $F(g, h)$

0000	0	0100	$g' \wedge h$	1000	$g' \wedge h'$	1100	g'
0001	$g \wedge h$	0101	h	1001	$g' + h$	1101	$g' \vee h$
0010	$g \wedge h'$	0110	$g + h$	1010	h'	1110	$g' \vee h'$
0011	g	0111	$g \vee h$	1011	$g \vee h'$	1111	1

когда они совпадают как функции; в табл. 5.3 указаны простые представители для каждого класса тождественных многочленов.

Подобным же образом диаграммы Венна, изображенные на рис. 1.1,*б*, позволяют предположить, что существует ровно $256 = 2^{2^3}$ различных булевых функций от трех переменных и что они находятся в естественном соответствии с подмножествами множества из восьми областей, на которые окружности разбивают прямоугольник U. Это также верно (теорема 9), однако здесь труднее выбрать кратчайшую форму записи для каждой функции. Решение задачи о кратчайшей записи зависит от выбора операций, представляемых индивидуальными символами алфавита. В табл. 5.3 этот алфавит включает символ $+$ (сложение mod 2), а не только стандартные булевы операции $\wedge$, $\vee$, $'$.

В гл. 6 мы займемся задачей о кратчайшей записи для произвольных n в алфавите $\wedge$, $\vee$, $'$. Там будут использованы «суммы произведений» специального вида, удобные для описания электронных схем. В этой главе нашей главной целью будет решение проблемы тождества для булевых многочленов от n переменных. Оставшаяся часть этого параграфа посвящена подготовке к этому решению. Начнем со следующих вспомогательных результатов.

Лемма 1. *Объединение и пересечение конечного семейства элементов в булевой алгебре A зависит только от самих этих элементов, но не от порядка действий над ними.*

Более формально, следующие тождества справедливы для любых конечных семейств S и T элементов булевой алгебры:

$$\bigwedge_S a_i \wedge \bigwedge_T a_j = \bigwedge_{S \cup T} a_k, \qquad \bigvee_S a_i \vee \bigvee_T a_j = \bigvee_{S \cup T} a_k. \tag{10}$$

Здесь и ниже символы $\bigwedge_S a_i$ и $\bigvee_S b_i$ обозначают соответственно наименьшую верхнюю и наибольшую нижнюю границы множества S. Иными словами, для $S = \{a_1, \ldots, a_n\}$

$$\bigwedge_S a_i = a_1 \wedge \ldots \wedge a_n$$

и

$$\bigvee_S a_i = a_1 \vee \ldots \vee a_n.$$

Согласно принципу двойственности (теорема 1), достаточно доказать первое из этих тождеств. На самом деле оба они очевидным образом следуют из L1 и из общих законов коммутативности и ассоциативности § 1.10 (см. также § 2.11). Если элемент a_h входит одновременно в S и T, мы можем поменять местами члены в левой части (10) так, чтобы два вхождения a_h стали соседними и заменить $a_h \wedge a_h$ на a_h, используя L1.

Лемма 2. *Для любых конечных семейств S и T имеем*

$$\left(\bigvee_S a_i\right) \wedge \left(\bigvee_T b_j\right) = \bigvee_{S\times T} (a_i \wedge b_j) \tag{11}$$

и, в силу двойственности,

$$\left(\bigwedge_S a_i\right) \vee \left(\bigwedge_T b_j\right) = \bigwedge_{S\times T} (a_i \vee b_j). \tag{11'}$$

В доказательстве используются только законы коммутативности, ассоциативности и дистрибутивности (L2, L3 и L6); оно проводится так же, как доказательство общего закона дистрибутивности в § 1.8.

Теперь обобщим лемму 1 § 5.2.

Лемма 3. *Если $a \wedge x = a \wedge y$ и $a \vee x = a \vee y$, то $x = y$.*

Доказательство. Используя L4, L6, L2 и свободно заменяя элементы равными им, получаем

$$x = x \wedge (x \vee a) = x \wedge (y \vee a) = (x \wedge y) \vee (x \wedge a) = (\text{в силу L6})$$
$$= (x \wedge y) \vee (y \wedge a) = y \wedge (a \vee x) = y \wedge (a \vee y) = y.$$

Теперь мы сформулируем очень важное обобщение законов де Моргана (L10) и одновременно закона инволюции (L9).

Лемма 4. *Чтобы вычислить дополнение любого булева многочлена, следует заменить все $\vee$ на $\wedge$, все $\wedge$ на $\vee$, над каждой не имеющей штриха буквой поставить штрих, а с каждой имеющей штрих буквы его снять.*

(Предполагается, что в первоначальной записи многочлена все двойные штрихи сняты с помощью L9, а штрихи, относящиеся к выражениям в скобках, внесены внутрь очевидной рекурсивной процедурой.)

Пример. Дополнение к $(x \wedge y') \vee z'$ тождественно $(x' \vee y) \wedge z$.

Доказательство. Проведем индукцию по числу n вхождений переменных в булев многочлен. Для $n = 1$ лемма справедлива, ибо $(x)' = x'$ и $(x')' = x$. Для $n \geqslant 2$ многочлен можно записать в одном из видов $f = p \wedge q$ или $f = p \vee q$. Здесь p и q либо многочлены, над которыми не стоят внешние штрихи, либо переменные, возможно, штрихованные. По законам де Моргана,

$f' = p' \vee q'$ или $f' = p' \wedge q'$. По предположению индукции p' и q' вычисляются, как это описано в лемме. Применение этого рецепта к p и q равносильно применению его к f.

УПРАЖНЕНИЯ А

1. Доказать следующие тождества:
а) для булевой суммы («симметрической разности»):

$$g+h'=g'+h, \quad a+b=a'+b', \quad a+(b+c)=(a+b)+c,$$

б) $(x \wedge y') \vee (x' \wedge y)=(x \vee y) \wedge (x' \vee y')$.

2. а) Написать тождество, двойственное к тождеству упр. 1,б).
б) Вывести второе тождество L5, не используя теорему 1, как в лемме 2 § 5.1.

3. Показать, что в алгебре F_2 (в обозначениях основного текста)

$$\alpha \wedge \beta = 0, \quad \alpha \vee \beta = I, \text{ откуда } \beta = \alpha'.$$

4. Положим $x \wedge y' = x - y$. Показать, что булевы многочлены 0, x, y, $x \wedge y$, $x \vee y$, $x-y$, $y-x$, $x+y$ и их дополнения исчерпывают все булевы многочлены от двух переменных.

5. Доказать тождество

$$(x \wedge y) \vee (y \wedge z) \vee (z \wedge y)=(x \vee y) \wedge (y \vee z) \wedge (z \vee x)$$

пользуясь только аксиомами L1—L6.

6. Доказать, что нетривиальная (конечная) булева алгебра не может состоять из нечетного числа элементов. (*Указание*: объединить в пары x и x'.)

7. Доказать, что тривиальная булева алгебра $\{0\}$ удовлетворяет аксиомам L1—L10.

8. Доказать индукцией по n, что $\left(\bigwedge_{i=1}^{n} x_i\right)' = \bigvee_{i=1}^{n} x_i'$.

9. Доказать соотношения (10) подробно. (*Указание*: провести индукцию по числу элементов в S и T.)

10. Доказать, что тождества (11)—(11′) справедливы в любой булевой алгебре.

11. Используя только аксиомы L1—L4, доказать, что каждое из двух тождеств L5 равносильно следующей самодвойственной импликации:

L5*: если $x \leqslant z$, то $x \vee (y \wedge z)=(x \vee y) \wedge z$.

12. Установить связь принципа двойственности этой главы с принципом двойственности из § 2.4.

5.4. ДИАГРАММЫ ВЕНТИЛЬНЫХ СХЕМ

В гл. 3 мы отмечали, что современные вычислительные машины содержат элементы памяти с двумя устойчивыми состояниями, которые удобно обозначать символами 0,1 («включен — выключен» или «активен — неактивен») либо T, F («истинно — ложно»).

В этом параграфе мы рассмотрим специальный класс логических схем, используемых в цифровых машинах. Они предназначены для того, чтобы по данным входным сигналам, подаваемым на n входов с элементов памяти в состояниях $X_1, \ldots, X_n$, получать один или несколько функционально вполне определенных выходных сигналов $F(X_1, X_2, \ldots, X_n)$, где функции F явно указаны. Такие схемы называются *вентильными схемами* или *вентилями* (в противоположность *последовательностным схемам*, включающим элементы задержки, которые будут обсуждены в § 6.9).

Очевидно, на n двоичных входов можно подать 2^n различных комбинаций сигналов. Поэтому вентили могут реализовать 2^{2^n} различных функций входа-выхода $f: 2^n \longrightarrow 2$. Согласно теореме 3, все эти функции реализуются булевыми многочленами, и формализм булевых алгебр очень удобен для их описания (см. снова § 5.1, где подробно описан случай $n = 1$).

Вентильные схемы полезно представлять в виде (ациклических) *помеченных ориентированных графов* (см. § 2.11), ребра (дуги) которых представляют электрические проводники, а вершины — вентили. Такие графы называются диаграммами. На следующих диаграммах изображены вентили, реализующие три основные

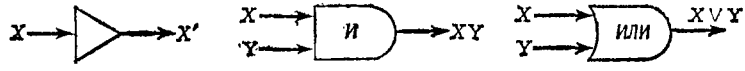


Рис. 5.3. Вентили.

булевы операции $'$, $\wedge$ и $\vee$. Они называются соответственно *инвертором*, *И-вентилем*, *ИЛИ-вентилем*. На рис. 5.3 указаны их стандартные обозначения. Инвертор преобразует 0 на входе в 1 на выходе, и наоборот. Таким образом, вход X преобразуется в выход X'.

И-вентиль и ИЛИ-вентиль производят бинарные операции (см. рис. 5.3). И-вентиль преобразует вход X, Y в выход XY (обычное сокращение для $X \wedge Y$; аналогично, XYZ употребляется вместо $X \wedge Y \wedge Z$ и т. д.) Значение выхода будет равным 1 в том и только том случае, когда оба входа равны 1. ИЛИ-вентиль преобразует вход X, Y в выход $X \vee Y$. Значение выхода будет равно 0 тогда и только тогда, когда оба входа равны 0.

Вентили, изображенные на рис. 5.3, и их комбинации называются вентильными схемами. Операции, производимые такими схемами, описываются многочленами при условии, что ни у каких двух вентилей выходы не соединены и никакой выход вентиля не соединен с его входом даже через цепочку других вентилей (такие обратные связи называются *петлями*; их наличие

часто приводит к тому, что выходы вентильной схемы становятся неопределенными).

На рис. 5.4 показано несколько комбинаций инверторов, И-вентилей и ИЛИ-вентилей вместе с описаниями их выходов как булевых комбинаций входов.

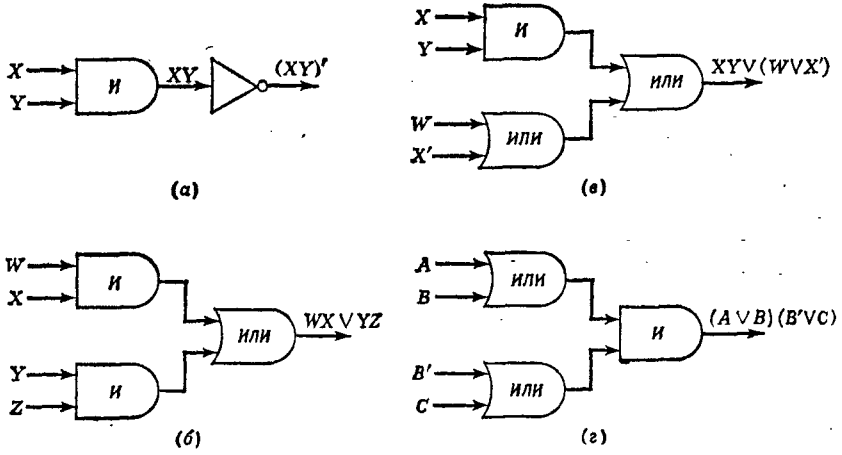


Рис. 5.4. Примеры вентильных схем.

На рис. 5.4,*а* изображен И-вентиль с двумя входами X и Y, выход которого подается на вход инвертора; на выходе последнего получается $(XY)'$. Значение на выходе будет равно 1 тогда и только тогда, когда хотя бы один из входов X, Y равен 0. На рис. 5.4,*б* два И-вентиля соединены с ИЛИ-вентилем. Эта схема называется *вентильной схемой И — ИЛИ*. Функция выхода описывается булевым многочленом $WX \vee YZ$. Поэтому выход будет равен 1, если оба входа W и X равны 1 либо оба входа Y и Z равны 1; в противном случае выходом будет 0.

Правила вычисления функции выхода схемы должны быть ясны из этих примеров. На рис. 5.4,*в* и 5.4,*г* представлено еще два примера.

Анализ вентильных схем такого типа постоянно производится при эксплуатации и наладке вычислительных машин. Действия вентильных схем в машинах изображаются диаграммами описанного вида и булевыми выражениями для выходов, представленными функциями от входов.

Вентили типа И, вентили типа ИЛИ часто имеют более двух входов. Так как операции $\vee$, $\wedge$ коммутативны и ассоциативны, их выходы описываются однозначно без использования скобок.

На рис. 5.5. показаны образцы таких схем.

Имеется хорошо разработанная техника для вычисления булевых функций, реализуемых схемами: это *таблицы комбинаций*. Для каждой из 2^n различных входных комбинаций схемы с n

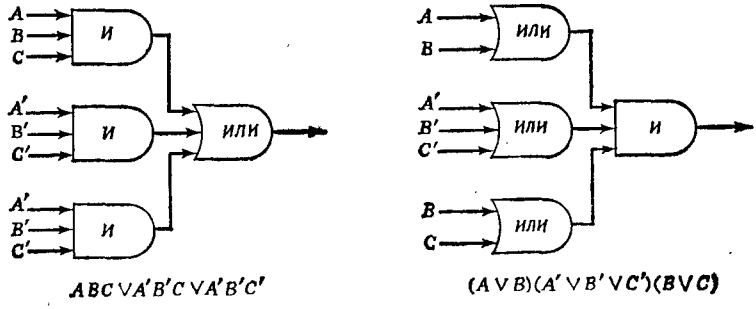
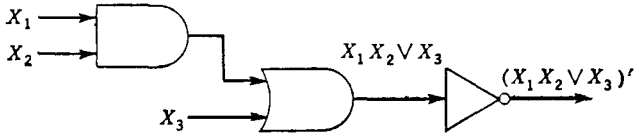


Рис. 5.5. Схемы И—ИЛИ и схемы ИЛИ—И.

входами в таблице указывается значение выхода для схемы или ее части, которое выражается булевой комбинацией. Такая таблица полностью описывает действие схемы. На рис. 5.6 изображена вентильная схема и описывающая ее таблица комбинаций.



Входы

X_1	X_2	X_3	X_1X_2	$X_1X_2 \vee X_3$	$(X_1X_2 \vee X_3)'$
0	0	0	0	0	1
0	0	1	0	1	0
0	1	0	0	0	1
0	1	1	0	1	0
1	0	0	0	0	1
1	0	1	0	1	0
1	1	0	1	1	0
1	1	1	1	1	0

Рис. 5.6.

Способ вычисления значений в каждой строке таблицы очевиден. Столь же очевиден и метод анализа схемы, дающий булево выражение выхода в терминах входов.

5.5. СВЯЗИ С ЛОГИКОЙ

Булевы операции возникают не только в теории множеств, но и в логике. Пусть буквами p, q, r, ... обозначены любые свойства объектов (быть красным, синим, мягким...). Булевы комбинации этих символов имеют следующую интерпретацию:

$p \wedge q$ означает *p и q*;
$p \vee q$ означает *p или q*;
p' означает *не p*.

Эти соглашения тесно связаны с действиями над множествами. Пусть, например, p и q — некоторые свойства, и пусть $S(p)$ означает множество всех объектов (из данного класса U), обладающих свойством p. Тогда

$$\begin{aligned} S(p \vee q) &= S(p) \cup S(q), \\ S(p \wedge q) &= S(p) \cap S(q), \\ S(p') &= [S(p)]'. \end{aligned} \tag{12}$$

Иными словами, отображение $p \mapsto S(p)$ ведет себя как *морфизм* булевых алгебр (мы выражаемся так осторожно, потому что булева алгебра «свойств» не была определена точно — «свойства» не образуют определенного множества).

Заметим, что $\vee$ в булевой алгебре означает «неразделительное или» («и/или»), а «разделительное или» выражается многочленом $p+q=(p \wedge q') \vee (p' \wedge q)$: «$p$ или q, но не оба вместе». В английском и русском языке связка «или» двусмысленна.

Более общая логическая интерпретация булевых операций состоит в рассмотрении общих *высказываний* p, q, r ... Как и в случае электрических сигналов в схемах, предполагается, что высказыванию может отвечать одно из двух истинностных значений, «истина» или «ложь» (T или F). Символическим равенством «$p=q$» можно обозначать утверждение, что p и q логически эквивалентны (более обычна запись $p \Leftrightarrow q$). Аналогично, словесные формулировки «из p следует q», или «если p, то q», записываются $p \Rightarrow q$, что логически эквивалентно $p' \vee q$ («либо q истинно, либо p ложно»).

Из этих примеров видно, что логические средства выражения естественного языка избыточны.

Некоторые высказывания истинны безотносительно к истинности или ложности своих составных частей только в силу своей логической структуры; они называются *тавтологиями*. Простейшая тавтология: $p \vee p'$ («или p или не p»).

Вот еще одна важная тавтология:

$$[(p \Rightarrow q) \wedge (q \Rightarrow r)] \Rightarrow (p \Rightarrow r). \tag{13}$$

Она выражает транзитивность отношения $\Rightarrow$. Для проверки того, что (13)— тавтология, выразим импликации через $\vee$ и $'$ и воспользуемся леммой 4 § 5.3 для установления следующих тождеств:

$$[(p' \vee q) \wedge (q' \vee r)]' \vee (p' \vee r) = (p \wedge q') \vee (q \wedge r') \vee p' \vee r =$$
$$= [(p \wedge q') \vee p'] \vee [(q \wedge r') \vee r] = \quad \text{(лемма 1 § 5.3)}$$
$$= [(p \vee p') \wedge (q' \vee p')] \vee [(q \vee r) \wedge (r' \vee r)] = \quad \text{(L6)}$$
$$= (q' \vee p') \vee (q \vee r) = q' \vee q \vee p' \vee r = I \vee p' \vee r = I.$$

Читателю следует убедиться, что аксиомы L1—L10 булевых алгебр согласуются с законами интуитивной логики не только для *множеств*, но и для *свойств* объектов (предикатов) и для *высказываний*. Это фундаментальное наблюдение и исследование его Булем и другими положило начало современной символической логике.

Мы вернемся к символической логике в гл. 14.

УПРАЖНЕНИЯ Б

1. Показать, что $(p \Rightarrow q) \Rightarrow (q' \Rightarrow p')$ из определений § 5.5.

2. Доказать, что следующие высказывания являются тавтологиями:
   а) $(p \vee p) \Rightarrow p$;
   б) $p \Rightarrow (p \vee q)$;
   в) $(p \Rightarrow q) \Rightarrow (q' \Rightarrow p')$;
   г) $(p \Rightarrow q) \Rightarrow [(r \vee p) \Rightarrow (r \vee q)]$.

3) Высказывания ложные, независимо от истинности или ложности их составных частей, называются *абсурдными*. Доказать, что следующие высказывания абсурдны:
   а) $q' \wedge p \wedge (p \Rightarrow q)$;
   б) $(p' \Rightarrow p) \wedge (p \Rightarrow p')$.

4. Доказать, что высказывание $p \Rightarrow p'$ *не* является абсурдным, и показать на примере, что это соглашение не противоречит интуиции.

5. а) Доказать, что операция

$$\varphi(g, h) = g + h = (g \wedge h') \vee (g' \wedge h)$$

отвечает сложению mod 2 характеристических функций $e_G = g$ и $e_H = h$ любых двух множеств G и H.

б) Доказать, что на первой диаграмме рис. 5.1 заштрихованная область отвечает $g + h$.

5.6. ЛОГИЧЕСКИЕ ВОЗМОЖНОСТИ АЛГОЛа

В § 4.6. мы рассмотрели условный оператор **if ... then... else** в АЛГОЛе и связанный с ним оператор перехода **go to**. Возможность использования условных переходов имеет первостепенную важность для программиста, ибо позволяет ему регу-

лировать порядок совершаемых машиной операций в зависимости от полученных ранее значений переменных.

Теперь мы систематически опишем это и другие логические средства АЛГОЛа, используя введенные выше булевы обозначения. Основное преимущество АЛГОЛа над ФОРТРАНом заключается в богатстве допустимых булевых выражений. Это резко сокращает формулировку сложных логических условий.

В АЛГОЛе текущее значение любого булева выражения или переменной есть либо **true**, либо **false**. Булевы переменные предваряются следующим описанием типа

Boolean $A, B, Y, Z,$

которое означает, что переменные A, B, Y и Z должны принимать булевы значения. Обычно булевы переменные принимают значения в зависимости от истинности или ложности *отношений* между арифметическими выражениями. В АЛГОЛе эти отношения определяются с помощью следующих *операций отношения*:

$$< \quad \leqslant \quad = \quad \geqslant \quad > \quad \neq$$

Например, если X — булева переменная, то оператор присваивания

$$X := 2 \uparrow 4 < 15$$

присваивает X значение **false**. Если Y — булева переменная, то оператор присваивания

$$Y := X \times 2 < A + B$$

присваивает переменной Y значение **true**, если текущее значение $2X$ меньше текущего значения $A+B$, и значение **false**, если $2X \geqslant A+B$.

Булевы переменные можно комбинировать в АЛГОЛе с помощью следующих *логических операций*: $\neg$ (унарная операция отрицания) и $\wedge$, $\vee$, $\supset$, $\equiv$ (бинарные операции). Таблица 5.4 описывает действие этих операций на булевы значения (F означает **false**, T означает **true**).

Таблица 5.4

Действие логических операций

$\neg$		$\wedge$	T	F	$\vee$	T	F	$\supset$	T	F	$\equiv$	T	F
T	F	T	T	F	T	T	T	T	T	T	T	T	F
F	T	F	F	F	F	T	F	F	F	T	F	F	T

Для сокращения количества скобок в АЛГОЛе принят следующий порядок выполнения операций. В первую очередь производятся арифметические действия в порядке, описанном в гл. 4. Все операции отношения равноправны и обрабатываются вслед за арифметическими операциями. Булевы операции обрабатываются в последнюю очередь в следующем порядке; $\neg$, $\wedge$, $\vee$, $\supset$, $\equiv$. Таким образом, $\neg A \vee B$ означает $(\neg A) \vee B$, а $\neg A \vee B \wedge \neg C$ означает $(\neg A) \vee (B \wedge (\neg C))$.

Основные виды *условных операторов* в АЛГОЛе таковы:

if B **then** S;

if B **then** S_1 **else** S_2;

Здесь B означает некоторое булево выражение, а S, S_1, S_2 суть операторы. В первом случае если текущее значение B есть **true**, то выполняется S, если же **false**, то S пропускается и выполняется следующий оператор. Во втором случае если текущее значение B есть **true**, то выполняется S_1 (а S_2 пропускается); если же текущее значение есть **false**, то выполняется S_2, а S_1 пропускается.

Пример условного оператора первого вида:

$$\textbf{if } X \geqslant Y \vee Y \geqslant Z \textbf{ then } W := W \uparrow 2;$$

Он придает W значение W^2, если X не меньше Y или **если** Y не меньше Z.

Пример условного оператора второго вида:

$$\textbf{if } W = Y \vee W = Z \textbf{ then } X := X \uparrow 2 \textbf{ else } X := X \uparrow 3;$$

Он придает X значение X^2, если W совпадает с Y или Z; в противном случае он придает X значение X^3. Еще один пример:

$$A := \textbf{if } J > 6 \textbf{ then } 4 \textbf{ else } 5;$$

Этот оператор эквивалентен словесной формулировке: «если текущее значение переменной J больше 6, присвоить A значение 4, в противном случае присвоить A значение 5». Наконец, если W, X, Y, Z—булевы переменные, то оператор

$$W := \textbf{if } X \wedge Y \vee X \wedge Z \textbf{ then } Y \wedge Z \textbf{ else } X;$$

присваивает X соответствующее булево значение.

Оператор **go to** в АЛГОЛе позволяет предписывать переход к внеочередному оператору и повторять выполнение тех или иных последовательностей операторов. Чтобы воспользоваться этой возможностью, следует поставить перед соответствующим оператором метку, являющуюся строкой символов—букв, цифр или пробелов.

Метка предшествует оператору и отделяется от него двоеточием.

Примеры меток:

L: $X:=A+B;$
EXL: $X:=A\uparrow B;$
QL: $X:=A/B;$
QUO: $X:=A/B;$

Метка означает просто *имя* соответствующего оператора (стоящего после двоеточия). Так, в нашем примере L есть имя оператора «присвоить X значение $A+B$», а EXL является именем оператора «присвоить X значение A^B». Если метка является строкой цифр, начальные нули не имеют значения: метка 056 воспринимается как 56.

Оператор **go to** M предписывает машине выполнить на следующем такте оператор с меткой M:

$Z:=3;$
$X:=5;$
$Y:=6;$
go to $B4;$
$Z:=X+Y;$
$B4$: $Z:=Z+X;$

После выполнения выписанной последовательности операторов переменной Z будет приписано значение 8, ибо программа опускает оператор $Z:=X+Y$, переходя сразу к последнему оператору.

Метки позволяют также использовать операторы вида

if A **then go to** $L;$

где A — некоторое булево выражение, а L — метка. Смысл такого оператора: «если A имеет текущее значение **true**, выполнить оператор с меткой L, если же A имеет текущее значение **false**, перейти к выполнению следующего оператора программы». Пусть, например, W — *вещественная* переменная. Последовательность меток и операторов

$A1$: $W:=2\times W;$
if $W<5_{10}9$ **then go to** $A1;$
end

присвоит W значение наименьшего числа вида 2^nW, превосходящего пяти миллиардов.

Пример 4. Следующая программа на АЛГОЛе вычисляет таблицу значений функции

$$F(x)=\begin{cases} 17.3-(x+1)^x, & x<5, \\ 19.4/(1+x^2), & x\geqslant 5, \end{cases}$$

для x, меняющегося от 0 до 10 с шагом 0.1:

```
begin real x, F;
x := 0;
back: F: = if x < 5 then 17.3—(x+1) ↑ x else 19.4/(1+x ↑ 2);
print (x, F);
if x < 10 then begin x := x+.1; go to back end
end
```

Оператор print — печать (не принадлежащий к числу операторов АЛГОЛа-60) должен выпечатывать значения x и F. Вот несколько более простой вариант этой программы с оператором цикла:

```
begin integer i; real x, F; x: = 0;
for i: = 0 step 1 until 100 do
    begin F := if x < 5 then 17.3—(x+1) ↑ x else 19.4/(1+x ↑ 2);
          print (x, F); x := x+.1;
    end
end
```

5.7. ПРИЛОЖЕНИЯ К БУЛЕВЫМ АЛГЕБРАМ

Логические возможности АЛГОЛа позволяют производить на машине многие вычисления в булевых алгебрах. Мы приведем в этом параграфе несколько простых примеров.

Пример 5. Вычисление булевых операций в множестве $\mathcal{P}(\mathbf{n})$ нетрудно запрограммировать, представив каждое подмножество S множества $\mathbf{n} = \{1, 2, \ldots, n\}$ как одномерный булев массив (**Boolean array**), состоящий из значений характеристической функции $e_S(k)$ этого множества. Она записывается в АЛГОЛе без индексов как $eS(k)$:

$$eS(k) = \begin{cases} \textbf{true}, & \text{если } k \in S, \\ \textbf{false}, & \text{если } k \notin S. \end{cases} \tag{14}$$

Пользуясь известными булевыми формулами

$$e_{S \cap T} = e_S \wedge e_T, \quad e_{S \cup T} = e_S \vee e_T,$$

мы можем выпечатать все элементы множества $S \cap T$, написав программу на АЛГОЛе:

```
for k := step 1 until n do
if eS(k) ∧ eT(k) = true then print k;
```

Аналогично можно написать программу для выпечатывания элементов множеств $S \cup T$ и S' в их естественном порядке.

Подобным же образом можно производить булевы операции над бинарными отношениями, отождествляя каждое бинарное

отношение ρ между множествами **m** и **n** с двумерным булевым массивом, в котором единицы в матрице отношения ρ отвечают булеву значению **true**, а нули — значению **false** (§ 2.1 и 2.2). Почти так же легко вычислить *композицию* бинарных отношений.

Пример 6. Пусть ρ — бинарное отношение на множестве $X=\{x_1, \ldots, x_n\}$, а $R=\|r_{ij}\|$ — его матрица. Вместо нее можно рассмотреть *простой* граф $\vec{G}=\vec{G}(\rho)$ с вершинами $x_1, \ldots, x_n$ и ребрами, соединяющими вершину x_i с вершиной x_j, если $r_{ij}=1$. Двумерный булев массив $B(\rho)=\|b_{ij}\|$, где

$$b_{ij}=\begin{cases}\textbf{true}, \text{ если } x_i \text{ и } x_j \text{ соединены ребром,}\\ \textbf{false} \text{ в противном случае,}\end{cases}$$

описывает граф $\vec{G}$. Матрица R есть *матрица инцидентности* этого графа.

Как говорилось в гл. 2, булева матрица $\|s_{ij}\|$, отвечающая отношению ρ^2, вычисляется по B с помощью формулы

$$s_{ij}=\bigvee_{k=1}^{n}(b_{ik}\wedge b_{kj}).$$

В терминах графа $\vec{G}(\rho)$ $s_{ij}=\textbf{true}$ тогда и только тогда, когда существует путь от вершины x_i к x_j, длина которого точно равна 2.

Пусть булев массив R размера $N\times N$ $(N\leqslant 50)$ содержит матрицу отношения ρ ($1=\textbf{true}$, $0=\textbf{false}$). Следующая программа на АЛГОЛе вычисляет матрицу отношения ρ^2:

```
begin Boolean array R[1:50, 1:50], S[1:50, 1:50];
      integer i, j, k, N;
      for i := 1 step 1 until N do
      for j := 1 step 1 until N do
      begin S[i, j]: = false;
            for k: = 1 step 1 until N do
            S[i, j]: = S[i, j] ∨ (R[i, k] ∧ R[k, j])
      end
end
```

Пример 7. Пусть задана матрица отношения $R=\|r_{ij}\|$ размера $n\times n$. Рассмотрим ее как (симметричный иррефлексивный) двумерный *булев массив*. Поставим задачу выяснить, связен ли отвечающий ей граф.

Можно было бы действовать способом, описанным в гл. 2, и вычислить матрицу

$$e\vee R\vee R^2\vee\ldots\vee R^n, \tag{15}$$

которая состоит из одних единиц в том и только том случае, если граф является связным. Однако это требует n^3 булевых умножений при вычислении очередной матрицы $R^{k+1} = R^k R$, а всего n^4 умножений. Следующий способ гораздо экономнее.

Очевидно, граф связен, тогда и только тогда, когда существует простой путь от вершины 1 к любой другой вершине. Поэтому можно воспользоваться описанным в § 2.10 алгоритмом отыскания кратчайшего пути и вычислять следующие множества: S_k, T_k, U_k.

Прежде всего положим $S_0 = \{1\} = T_0 = U_0$. Затем будем вычислять рекурсивно (например, с помощью программы на АЛГОЛе)

$$T_{k+1} = S_k R, \qquad S_{k+1} = T_{k+1} \cap U_k', \qquad U_{k+1} = S_{k+1} \cup U_k.$$

Здесь

$$S_k R = \{j \in \mathbf{n} \mid iRj \text{ для некоторого } i \in S_k\}.$$

Характеристическая фуикция этого множества вычисляется так: $j \mapsto \bigvee_{i \in S_k} r_{ij}$.

Очевидно, T_{k+1} есть множество всех вершин, соседних с некоторой вершиной S_k; S_{k+1} есть множество всех вершин $h \in \mathbf{n}$, находящихся на *расстоянии*, точно равном $k+1$, от 1. Наконец, U_{k+1} есть множество всех вершин, находящихся на расстоянии $\leqslant k+1$ от 1. Вычисление заканчивается, когда $S_m = \varnothing$. Ясно, что $m \leqslant n$ и связность графа равносильна тому, что U_m включает все вершины.

Поскольку S_k попарно не пересекаются, этот способ требует порядка n^2 булевых умножений и потому гораздо быстрее предыдущего (рассмотрите случай $n = 50$).

УПРАЖНЕНИЯ В

1. Написать программу на АЛГОЛе, выпечатывающую элементы множества $S \cup T$ для любых $S \subset \mathbf{n}$, $T \subset \mathbf{n}$.

2. Написать программу на АЛГОЛе, вычисляющую множество $S' = \neg S$ для любого $S \subset \mathbf{n}$.

3. Пусть $A = \| a_{ij} \|$ и $B = \| b_{ij} \|$ — булевы массивы размера $m \times n$ и $n \times r$. Положим $AB = \| c_{ij} \|$, где $c_{ij} = \bigvee_k a_{ik} b_{kj}$. Написать программу на АЛГОЛе, вычисляющую AB.

4. Объяснить утверждение: «двумерный массив типа **Boolean** является матрицей отношения».

5. Написать программу на АЛГОЛе для вычисления вещественных корней (если они существуют) квадратного уравнения

$$4x^2 + e^{\pi/\sqrt{2}} + \ln 80 = 0.$$

6. Написать программу на АЛГОЛе для вычисления функции

$$F(x)=\begin{cases} 3125-x^x & \text{при } x<5,\\ (x-5)/(1+x^2) & \text{при } x\geqslant 5 \end{cases}$$

для x, меняющегося от 1 до 10 с шагом 0.1.

*7. Пусть R — матрица смежности ориентированного графа $\vec{G}$ с вершинами 1, 2, ..., 50, и пусть $N\geqslant 1$. Написать программу на АЛГОЛе, вычисляющую такую матрицу S, что $S_{ij}=1$ (**true**) тогда и только тогда, когда существует путь от вершины i к вершине j длины $\leqslant N$. Общая организация программы такова:

```
begin integers N, ...;
      Boolean array R [1:50, 1:50], S [1:50, 1:50, 1:50], ...;
Read (N, matrix R);
      .
      .
      .
Print (N, matrix R, matrix S)
end
```

Конкретный вид команд ввода и вывода здесь нас не интересует. Он зависит от особенностей транслятора и имеющихся устройств ввода и вывода.

5.8. БУЛЕВЫ ПОДАЛГЕБРЫ

Пусть $\mathcal{B}=[A, \wedge, \vee, ', O, I]$ — некоторая булева алгебра. Ее *булевой подалгеброй* называется подмножество $S\subset A$, обладающее следующими свойствами:

1. Оно содержит O и I.
2. Оно содержит дополнение x' любого своего элемента x.
3. Вместе с любой парой x, y оно содержит произведение $x\wedge y$ и сумму $x\vee y$.

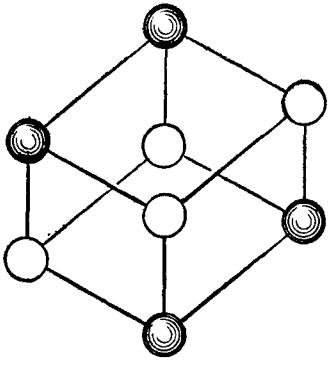


Рис. 5.7.

Из L7 видно, что для любого элемента $x\in A$ с $O<x<I$ четыре элемента O, x, x', I составляют булеву подалгебру A, *порожденную* x (или x'). На рис. 5.7 заштрихованные элементы образуют подалгебру булевой алгебры $\mathcal{B}=[\mathcal{P}(3), \cap, \cup, ', \varnothing, 3]$: Пара элементов O, I также образует подалгебру.

Теорема 4. *Всякая булева подалгебра S булевой алгебры $\mathcal{B}$ является булевой алгеброй $\mathcal{S}=[S, \wedge, \vee, ', O, I]$ относительно операций в $\mathcal{B}$.*

Действительно, в S выполнены аксиомы L1—L10.

Пример 8. Как отмечалось в гл. 1, множество частей $\mathcal{P}(I)$ любого множества I является булевой алгеброй относительно

операций $\cap$ (пересечение), $\cup$ (теоретико-множественное объединение) и дополнения. Поэтому булевы подалгебры $\mathcal{P}(I)$ являются (непустыми) подмножествами множества $\mathcal{P}(I)$, $X, Y, Z, \ldots \subset I$, замкнутыми относительно этих операций. Такие подалгебры называются также *полями* множеств.

Ясно, что подалгебра $\{O, x, x', I\}$, порожденная элементом x, является наименьшей подалгеброй, содержащей x.

В § 5.1 мы убедились, что булева алгебра $F_2 = 2^4$ порождена двумя элементами f и g (см. также рис. 5.2). Обобщим этот результат на любые алгебры $\mathcal{P}(2^r)$.

Теорема 5. *Булева алгебра* $\mathcal{P}(2^r) = B^{2^r}$ *порождена* r *элементами* $X_1, \ldots, X_r$.

Доказательство. Реализуем 2^r как множество двоичных слов длины r, $\mathbf{w} = w_1 \ldots w_r$ ($w_i = 0$ или 1). Обозначим через $X_i \subset 2^r$ множество всех слов, у которых на i-м месте стоит 1. Например, для $r = 3$

$$X_1 = \{100,\ 101,\ 110,\ 111\},$$
$$X_2 = \{010,\ 011,\ 110,\ 111\},$$
$$X_3 = \{001,\ 011,\ 101,\ 111\}.$$

Для каждого конкретного слова $\mathbf{w}$ длины n имеем:

$$\{\mathbf{w}\} = \bigcap_{i=1}^{r} Y_i(\mathbf{w}), \text{ где } Y_i(\mathbf{w}) = \begin{cases} X_i, & \text{если } w_i = 1, \\ X_i', & \text{если } w_i = 0. \end{cases} \tag{16}$$

Поскольку каждое непустое подмножество множества $\mathcal{P}(2^r)$ есть объединение своих одноэлементных подмножеств и $\varnothing = X_i \cap X_i'$, теорема доказана.

Интервалы. Пусть a, b — два элемента частично упорядоченного множества A. *Интервал* $[a, b]$, по определению, есть множество всех $x \in A$ со свойством $a \leqslant x \leqslant b$.

Например, если A — булева алгебра, то $[O, I] = A$.

Лемма 1. *Если* $x \in [a, b]$ *и* $y \in [a, b]$, *то* $x \wedge y$ *и* $x \vee y$ *принадлежат* $[a, b]$.

В самом деле, a — нижняя граница для $\{x, y\}$, а b — верхняя, поэтому a не превосходит *наибольшей* нижней границы $\{x, y\}$, т. е. $x \wedge y$, а b не меньше *наименьшей* верхней границы $\{x, y\}$, т. е. $x \vee y$.

Назовем *решеткой* любую алгебру $[L, \wedge, \vee]$, в которой выполняются аксиомы L1—L4, а *дистрибутивной решеткой* — решетку с дополнительной аксиомой L6. Мы займемся решетками систематически в гл. 9. Пока отметим следствие леммы 1.

Теорема 6. *Любой интервал* $[a, b]$ *в булевой алгебре* A *является дистрибутивной решеткой относительно операций* A.

Доказательство. Согласно лемме 1, интервал $[a, b]$ вместе с любыми двумя элементами x, y содержит их произведение и сумму. Выполнимость тождеств L1—L4 и L6 ясна, поскольку эти тождества выполнены в A.

Заметим, что интервал $[a, b]$ *не является* булевой подалгеброй в A, за исключением тривиального случая $a=O$, $b=I$, когда этот интервал совпадает с A. В противном случае O или I не содержится в $[a, b]$. Однако мы можем определить операцию *относительного* дополнения x в интервале $[a, b]$: $x^*=(a \vee x') \wedge b$, и доказать следующий результат:

Теорема 7. *Для любого интервала* $[a, b]$ *в булевой алгебре* A *алгебраическая система* $[[a, b], \wedge, \vee, *, a, b]$ *сама является булевой алгеброй.*

Мы оставляем доказательство читателю в качестве упражнения (упр. Г7 ниже).

5.9. ДИЗЪЮНКТИВНАЯ НОРМАЛЬНАЯ ФОРМА

В этом параграфе мы завершим решение «проблемы тождества» для булевых многочленов. Для этого очень полезно понятие дизъюнктивности, применимое не только к булевым алгебрам и решеткам, но также к другим частично упорядоченным множествам.

Определение. В частично упорядоченном множестве с универсальной нижней границей O элементы x, y называются *дизъюнктными*, если $x \wedge y = O$ (под $x \wedge y$ понимается наибольшая нижняя граница $\{x, y\}$).

В булевой алгебре A дополнительные элементы a, a' дизъюнктны. Более того, множество всех элементов, дизъюнктных с любым элементом $a \in A$, совпадает с интервалом $[O, a']$:

$$a \wedge x = O \text{ тогда и только тогда, когда } x \leqslant a'. \tag{17}$$

Более общо, в любой дистрибутивной решетке

$$\text{из } a \wedge x = O \text{ и } a \wedge y = O \text{ следует, что } a \wedge (x \vee y) = O, \tag{18}$$

ибо $a \wedge (x \vee y) = (a \wedge x) \vee (a \wedge y) = O \vee O = O$ в силу L6.

Теорема 8. *Пусть* $p_1, \ldots, p_m$ *—попарно дизъюнктные элементы дистрибутивной решетки* L *с нулем. Определим функцию* θ: $\mathcal{P}(\mathbf{m}) \longrightarrow L$ *как отображение, переводящее каждое непустое подмножество* $S \subset \mathbf{m}$ *в* $\theta(S) = \bigvee_{i \in S} p_i$, *а* $\varnothing$ *в* O. *Тогда*

$$\left(\bigvee_{i \in S} p_i\right) \vee \left(\bigvee_{j \in T} p_j\right) = \bigvee_{S \cup T} p_k \tag{19}$$

и

$$(\bigvee_{i\in S} p_i) \wedge (\bigvee_{j\in T} p_j) = \bigvee_{S\cap T} p_k. \tag{19'}$$

Пояснение. Мы полагаем $\bigvee_{\varnothing} p_i = 0$: это согласуется с определением $\vee$ как наименьшей верхней границы.

Доказательство. Равенство (19) следует из коммутативности, ассоциативности и идемпотентности по индукции, как в § 5.3.

Далее, по общему закону дистрибутивности § 1.8 для любых $S \subset \mathbf{m}$, $T \subset \mathbf{m}$ имеем

$$(\bigvee_{i\in S} p_i) \wedge (\bigvee_{j\in T} p_j) = \bigvee_{S\times T} (p_i \wedge p_j).$$

Но в силу дизъюнктности $p_i \wedge p_j = 0$ при $i \neq j$. Поэтому все такие члены мы можем опустить, так что $\bigvee_{S\times T} (p_i \wedge p_j) = \bigvee_{S\cap T} p_k$, если p_i попарно дизъюнктны. Объединяя эти равенства, получаем (19'). Это завершает доказательство теоремы 8.

Теперь мы можем перейти к доказательству нашего основного результата. Чтобы сформулировать его в компактном виде, введем следующие соглашения. Символами $\mathbf{v} = v_1 \ldots v_n$, $\mathbf{w} = w_1, \ldots, w_n$ мы будем обозначать двоичные строки длины n; v_i и w_j суть нули и единицы. Каждой такой строке $\mathbf{w}$ поставим в соответствие булев многочлен $p(\mathbf{w}) = \bigwedge_{\mathbf{n}} z_i$, где $z_i = y_i$, если $w_i = 1$, и $z_i = y_i'$ при $w_i = 0$. Таким способом мы построим 2^n многочленов $p(\mathbf{w})$. Наш основной результат состоит в том, что всякий булев многочлен $f(y_1, \ldots, y_n)$ тождествен булевой сумме некоторого множества многочленов $p(\mathbf{w})$.

Теорема 9. *Всякий булев многочлен от переменных* $Y = \{y_1, \ldots, y_n\}$ *можно представить в так называемой дизъюнктивной нормальной форме:*

$$\bigvee_S p(\mathbf{w}), \text{ где } p(\mathbf{w}) = \bigwedge_{\mathbf{n}} z_i, \quad z_i = \begin{cases} y_i, & \text{если } w_i = 1, \\ y_i', & \text{если } w_i = 0, \end{cases} \tag{20}$$

и притом единственным способом.

Доказательство. Элементы $p(\mathbf{w})$ попарно дизъюнктны: если $\mathbf{v} \neq \mathbf{w}$, то $v_i \neq w_i$ для подходящего i, так что $p(\mathbf{v}) < y_i$ и $p(\mathbf{w}) < y_i'$ или наоборот, откуда $p(\mathbf{v}) \wedge p(\mathbf{w}) = O$. Отсюда и из теоремы 8 следует, что множество всех сумм многочленов $p(w)$ замкнуто относительно операций $\wedge$ и $\vee$. Далее, если S и S' дополнительны в множестве W всех двоичных строк длины n,

то, согласно (19),

$$\bigvee_S p(\mathbf{w}) \vee \bigvee_{S'} p(\mathbf{w}) = \bigvee_W p(\mathbf{w}) = I.$$

Действительно,

$$I = I \wedge \ldots \wedge I = (y_1 \vee y_1') \wedge \ldots \wedge (y_n \vee y_n') = \bigvee_W p(\mathbf{w})$$

в силу общего закона дистрибутивности. Отсюда следует, что $\bigvee_{S'} p(\mathbf{w})$ есть дополнение к $\bigvee_S p(\mathbf{w})$, так что множество всех булевых многочленов, представленных в дизъюнктивной нормальной форме, образует булеву подалгебру всех булевых многочленов от Y.

Остается показать, что все элементы y_i содержатся в множестве (20) с точностью до тождественности. С этой целью заметим, что

$$y_1 = y_1 \wedge I \wedge \ldots \wedge I = y_1 \wedge (y_2 \vee y_2') \wedge \ldots \wedge (y_n \vee y_n').$$

Применяя к последнему выражению общий закон дистрибутивности, находим

$$y_1 = \bigvee_{S(1)} p(\mathbf{w}), \quad S(1) = \{\mathbf{w} \mid w_1 = 1\}.$$

Аналогично, $y_i = \bigvee_{S(i)} p(\mathbf{w})$ в очевидных обозначениях, что завершает доказательство теоремы 9.

Менее очевидно, что 2^{2^n} выражений (20) представляют попарно различные функции. Это, однако, следует из теоремы 5. Формально различные выражения $\bigvee_S p(\mathbf{w})$ представляют различные функции на алгебре $\mathscr{P}(\mathbf{n})$. Это доказывает следующее важное следствие теорем 5 и 9:

Следствие. *Каждый булев многочлен* $p(y_1, \ldots, y_n)$ *можно привести к единственной дизъюнктивной нормальной форме* (20) *последовательным применением тождеств* L1—L10.

Это означает, что $\mathbf{B}^{2^n}$ есть *свободная булева алгебра* с n образующими. В следующем параграфе мы уточним это замечание.

УПРАЖНЕНИЯ Г

1. Доказать, что двоичных строк длины 48 недостаточно, чтобы перенумеровать ими все булевы многочлены от шести переменных. Доказать, что аналогичная нумерация булевых многочленов от 20 переменных требует более 10^6 битов на каждый номер.

2. Доказать, что для хранения всех булевых многочленов от шести переменных требуется больше двух миллионов 64-битовых ячеек памяти.

3. Доказать, что $g + h' = g' + h$ и $g + h = g' + h'$.

4. Описать булев изоморфизм алгебры $F_2 \cong 2^4$ с алгеброй областей, изображенных на рис. 5.1, отображающий g и h на помеченные этими символами круги.

5. Доказать, что для любого булева многочлена $p(x, y)$

$$p(x, y) = [x \wedge p(I, y)] \vee [x' \wedge p(O, y)]$$

и

$$p(x, y) = [p(I, I) \wedge x \wedge y] \vee [p(I, O) \wedge x \wedge y'] \vee \\ \vee [p(O, I) \wedge x' \wedge y] \vee [p(O, O) \wedge x' \wedge y'].$$

6. Установить биекцию между булевыми подалгебрами алгебры 2^n и разбиениями множества атомов (т. е. элементов a, таких, что $[O, a] = \{O, a\}$).

*7. а) Показать, что если $[a, b]$ — любой интервал в булевой алгебре A, то $[[a, b], \wedge, \vee, a, b]$ есть дистрибутивная решетка.

б) Показать, что операция $a \mapsto a^*$ удовлетворяет условиям L7 — L10.

в) Вывести отсюда теорему 7.

*5.10. ПРЯМЫЕ ПРОИЗВЕДЕНИЯ И МОРФИЗМЫ

Пусть A, B — две булевы алгебры. Их *прямым произведением* называется декартово произведение $A \times B$ множеств A, B со следующими операциями:

$$\begin{gathered}(a_1, b_1) \vee (a_2, b_2) = (a_1 \vee a_2, b_1 \vee b_2), \\ (a_1, b_1) \wedge (a_2, b_2) = (a_1 \wedge a_2, b_1 \wedge b_2), \\ (a, b') = (a', b'), \quad O = (O_A, O_B), \quad I = (I_A, I_B).\end{gathered} \tag{21}$$

Справедливость тождеств L1 — L10 очевидна, ибо они выполняются покомпонентно.

Отображение θ: $A \longrightarrow B$ булевых алгебр называется *булевым морфизмом*, если

$$\theta(x \wedge y) = \theta(x) \wedge \theta(y), \qquad \theta(x \vee y) = \theta(x) \vee \theta(y), \qquad \theta(x') = [\theta(x)]'$$

для всех x, $y \in A$. Из этих тождеств следует, что

$$\theta(O_A) = \theta(x \wedge x') = \theta(x) \wedge \theta(x') = \theta(x) \wedge [\theta(x)]' = O_B$$

и, по двойственности, $\theta(I_A) = I_B$.

Как в § 2.6, взаимно однозначный булев морфизм называется (булевым) *моно*морфизмом; морфизм, являющийся отображением на B, называется (булевым) *эпи*морфизмом, а морфизм, являющийся биекцией, называется (булевым) *изо*морфизмом.

Очевидно, отображения p_A: $(a, b) \mapsto a$ и p_B: $(a, b) \mapsto b$ являются булевыми эпиморфизмами p_A: $A \times B \longrightarrow A$ и p_B: $A \times B \longrightarrow B$.

Теорема 10. *Биекция между булевыми алгебрами является булевым изоморфизмом тогда и только тогда, когда она сохраняет частичный порядок (отношение включения).*

Доказательство. Если отношение включения сохраняется, то, согласно теореме 2, сохраняются операции н.в.г., н.н.г. и

дополнения, а также обе универсальные границы. Обратное утверждение тривиально.

Отсюда следует, что любая конечная булева алгебра определяется с точностью до изоморфизма своей диаграммой (ориентированным графом отношения доминирования, или накрытия). Однако эти диаграммы перестают быть наглядными для алгебр порядка 32 или больше.

Заметим еще, что если $\mathcal{P}(U)$ и $\mathcal{P}(V)$—булевы алгебры всех подмножеств $S \subset U$, $T \subset V$ (с операциями пересечения, объединения и дополнения), то отображение $S \sqcup T \longrightarrow (S, T)$ для переменных $S \subset U$, $T \subset V$ определяет изоморфизм $\mathcal{P}(U \sqcup V) \cong \mathcal{P}(U) \times \mathcal{P}(V)$.

Свободные булевы алгебры. Пусть теперь $G = \{g_1, \ldots, g_r\}$—образующие булевой алгебры $\mathcal{P}(\mathbf{2}^r) = \mathbf{B}^{2^r}$, которые обозначались символами $X_1, \ldots, X_r$ в теореме 5. Пусть A — произвольная булева алгебра, и пусть $f\colon G \longrightarrow A$—любое отображение G в A; положим $f(g_i) = y_i$.

Тогда отображение $S \mapsto \bigvee_S p_i$ из теоремы 8 определяет *морфизм* алгебры $\mathcal{P}(\mathbf{2}^r)$ в A относительно операций сложения и умножения. В доказательстве теоремы 9 было показано, что он является также морфизмом относительно дополнения. Поэтому это морфизм булевых алгебр.

Приведенное рассуждение устанавливает следующий результат.

Теорема 11. *Всякое отображение $g_i \mapsto y_i$ множества образующих булевой алгебры $\mathcal{P}(\mathbf{2}^r)$ теоремы 5 в булеву алгебру A продолжается до булева морфизма $\mu\colon \mathcal{P}(\mathbf{2}^r) \longrightarrow A$.*

Алгебраическая система произвольного типа с r образующими, удовлетворяющая заключению теоремы 11, называется *свободной алгеброй* с r образующими. Мы показали, таким образом, что $\mathcal{P}(\mathbf{2}^r) \cong \mathbf{B}^{2^r}$ есть свободная булева алгебра с r образующими.

УПРАЖНЕНИЯ Д

1. Привести следующие булевы многочлены от x, y, z к дизъюнктивной нормальной форме:
   а) y, б) $xy \vee z'$, в) $x + yz'$.

2. Упростить булево выражение $(xy \vee x)(z' \vee zx)$.

3. Доказать, что $xy \vee xz \vee xy' = xy \vee zy'$.

4. Показать, что из каждого из двух дистрибутивных законов L6 следует другой.

5. Пусть $\mathcal{B}$—булева алгебра, $a_1, \ldots, a_n$—ее попарно дизъюнктные элементы и $\bigvee_{i=1}^{n} a_i = I$. Показать, что отображение $\mu\colon S \mapsto \bigvee_S a_i$ является булевым мономорфизмом множества $\mathcal{P}(\mathbf{n})$ всех $S \subset \mathbf{n}$ в $\mathcal{B}$.

6. Доказать со всеми подробностями, что F_2 — свободная булева алгебра с двумя образующими.

7. Вычислить разбиение F_3 на классы эквивалентности:

а) относительно группы перестановок образующих x, y, z;

б) относительно группы, порожденной перестановками x, y, z и отображениями $x \mapsto x'$, $y \mapsto y'$, $z \mapsto z'$.

8. Пусть $\mathcal{A} = [A, \wedge, \vee, ']$, $\mathcal{B} = [B, \wedge, \vee, ']$ — две булевы алгебры и θ: $A \longrightarrow B$ — булев морфизм. Показать, что прообраз $\theta^{-1}(O)$ в A нуля $O \in B$ обладает следующими свойствами:

а) $O \in \theta^{-1}(O)$;

б) если $a \in \theta^{-1}(O)$, то для всех $x \leqslant a$ также $x \in \theta^{-1}(O)$.

в) если $a \in \theta^{-1}(O)$, то $a_1 \vee a_2 \in \theta^{-1}(O)$. Такие подмножества A называются *идеалами*; $\theta^{-1}(O)$ есть *ядро* θ.

9. Показать, что если J — любой идеал в A, то существует булева алгебра $\mathcal{C} = [C, \wedge, \vee, ']$ и булев эпиморфизм $A \longrightarrow C$ с ядром J.

СПИСОК ЛИТЕРАТУРЫ

1. Hohn F. E., Applied Boolean Algebra, Macmillan, 1960.
2. Whitesitt J. E., Boolean Algebra and its Applications, Addison Wesley, 1961.

6

ОПТИМИЗАЦИЯ И ПРОЕКТИРОВАНИЕ ВЫЧИСЛИТЕЛЬНЫХ МАШИН

6.1. ВВЕДЕНИЕ

Теория булевых алгебр получила важные приложения в исследовании и проектировании цифровых вычислительных машин, в особенности вентильных схем, описанных в гл. 5. Проблема оптимизации таких схем по существу является задачей о булевых алгебрах. Мы поясним это утверждение в § 6.3 и 6.4, где будет описана общая методика синтеза вентильных схем, реализующих любую булевую функцию, из нескольких простых элементов.

Предварительно мы обсудим общее понятие оптимизации (§ 6.2) и способ оптимизации для одного класса несложных «многостадийных процессов принятия решений», сводящегося к задачам отыскания кратчайших путей в ориентированных графах, дугам которых приписаны определенные *длины* (или цены).

После этого вводного материала в трех параграфах 6.5—6.7 разработана полезная методика достижения *экономичности* при синтезе вентильных схем. Существует много разных вентильных схем, выход которых как функция от входа реализует данную булеву функцию $f(x_1, \ldots, x_n)$. Задача состоит в отыскании схемы, состоящей из возможно меньшего числа основных элементов, что упрощает и удешевляет конструкцию. Способы решения этой задачи называются *процедурами минимизации*.

В § 6.8 и 6.9 мы займемся более сложной проблемой проектирования *последовательностных* машин. Эти машины содержат как вентили, так и элементы памяти (задержки). Мы опишем способы реализации любого конечного автомата.

6.2. ОПТИМИЗАЦИЯ

Экономичность конструкции является важной инженерной проблемой любого проекта. Обычно стремятся *минимизировать стоимость* достижения определенных технических характеристик. Важная область прикладной математики — *теория оптимизации* — разрабатывает математические средства, помогающие минимизировать

стоимость достижения данных показателей или, наоборот, *максимизировать показатели* при данных ограничениях на ресурсы.

Общую идею оптимизации и математические приемы для решения задач удобно объяснить на примере многостадийных процессов решения следующего типа.

Предположим, что некоторый *процесс* развивается посредством дискретных *переходов* от одного *состояния* данной *стадии* процесса к следующему состоянию следующей стадии. Возможные пути развития такого процесса схематически представляют *ориентированными графами* (или *картами потоков*), вершины которых изображают возможные состояния процесса, а дуги — возможные переходы от одного состояния к другому, *помеченные ценами*. Задача состоит в отыскании такого пути от *начального* до *конечного состояния*, у которого была бы *наименьшая общая стоимость* (сумма цен дуг на этом пути).

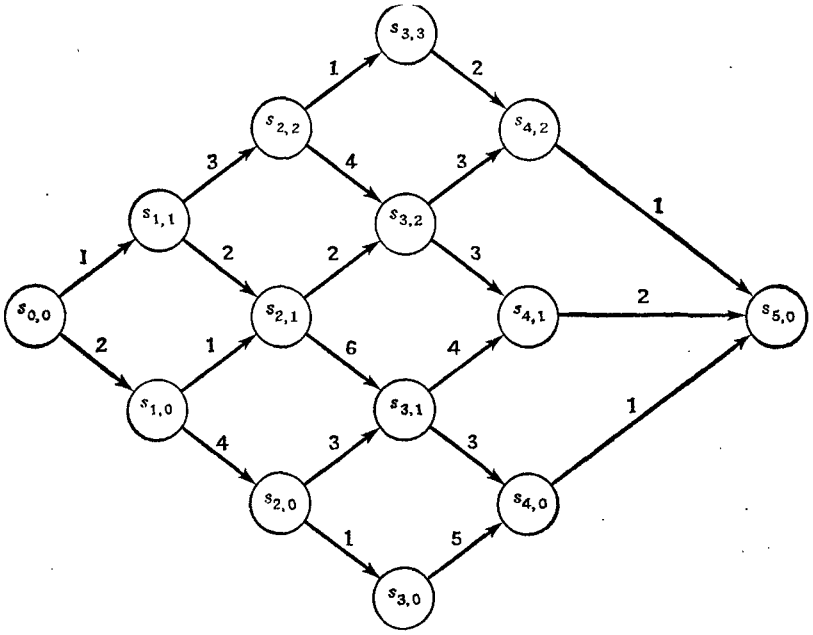


Рис. 6.1. Граф многостадийного процесса.

Если интерпретировать цену как *длину*, то *оптимальное решение* состоит в выборе *кратчайшего* или *оптимального пути*. Эта задача уже встречалась в § 2.10 для случая, когда длины всех дуг одинаковы.

Пример 1. Рассмотрим помеченный ориентированный граф, изображенный на рис. 6.1.

У него есть всего 14 состояний и всего 14 путей от начального состояния $s_{0,0}$ до конечного $s_{5,0}$. Поэтому можно отыскать оптимальный путь, просто вычислив стоимость всех путей и отобрав из них наименьшую. Однако для больших графов такой способ становится непрактичным. Следующий *принцип оптимальности* доставляет более эффективную процедуру.

Теорема 1 (принцип оптимальности). *Любой подпуть оптимального пути оптимален.*

Доказательство очевидно. Пусть $\vec{P}$ — оптимальный путь от s_0 до s_n, состоящий из последовательных дуг $\alpha_i = \overrightarrow{s_{i-1}s_i}$ $(i = 1, \ldots, n)$, и пусть $\vec{Q} = [s_j, s_{j+1}, \ldots, s_k]$ — любой его подпуть (отрезок). Если цену отрезка $\vec{Q}$ можно уменьшить, заменив его отрезком

$$\vec{R} = [\tilde{s}_h = s_j, \tilde{s}_{h+1}, \ldots, \tilde{s}_l = s_k],$$

то новый путь

$$\vec{s} = [s_0, \ldots, s_{j-1}, \tilde{s}_h, \ldots, \tilde{s}_l, s_{k+1}, \ldots, s_n],$$

получающийся из $\vec{P}$ заменой $\vec{Q}$ на $\vec{R}$, будет дешевле (короче) $\vec{P}$, что противоречит предположению об оптимальности $\vec{P}$.

Следствие. *Если цены (длины) всех дуг ориентированного графа положительны, то любой оптимальный путь является простым путем.*

Применим этот принцип оптимальности к графу рис. 6.1, двигаясь справа налево, т. е. находя последовательно оптимальные пути от состояний стадии 4 до $s_{5,0}$, затем от стадии 3 до $s_{5,0}$, от стадии 2 до $s_{5,0}$, от стадии 1 до $s_{5,0}$ и, наконец, от $s_{0,0}$ до $s_{5,0}$.

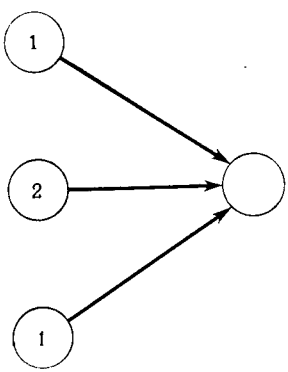


Рис. 6.2. Граф для отыскания функции M от стадии 4.

Определим функцию $M(s_{i,j})$ как стоимость оптимального пути от состояния $s_{i,j}$ к состоянию $s_{5,0}$. Вот некоторые ее значения:

$$M(s_{5,0}) = 0, \quad M(S_{4,2}) = 1,$$
$$M(s_{4,1}) = 2, \quad M(s_{4,0}) = 1.$$

Вычислив эти цены для четвертой стадии процесса, мы перейдем к их вычислению для третьей стадии. С этой целью нужно рассмотреть подграф, изображенный на рис. 6.2, вершины которого $s_{i,k}$ помечены значениями $M(s_{i,k})$. Стоимости $M(s_{i-1,k})$ вычисляются непосредственным перебором для отыскания мини-

мума. Например, от состояния $s_{3,2}$ к стадии 4 имеется две дуги. Цена верхней равна 3, она идет к $s_{4,2}$, где $M(s_{4,2})=1$; общая стоимость равна 4. Нижняя дуга цены 3 идет к состоянию $s_{4,1}$,

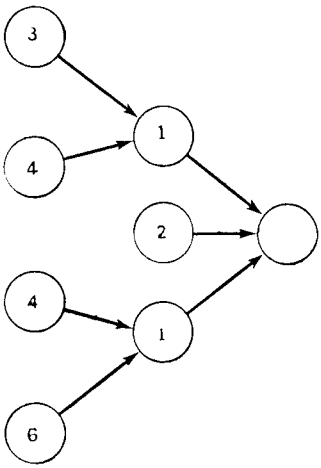


Рис. 6.3. Граф для отыскания функции M от стадии 3.

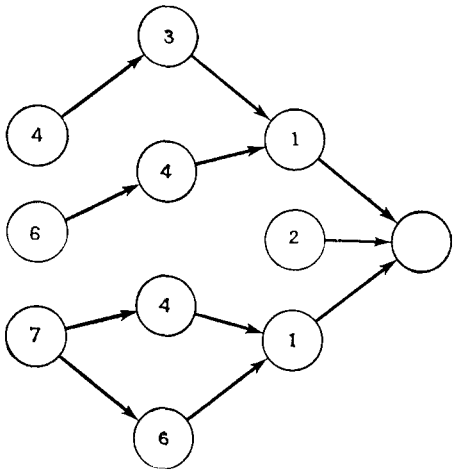


Рис. 6.4. Граф для отыскания функции M от стадии 2.

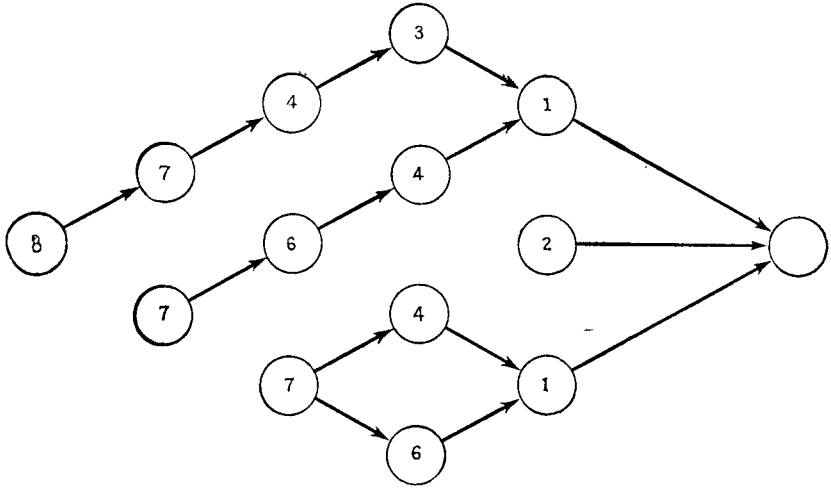


Рис. 6.5. Граф оптимального решения.

где $M(s_{4,1})=2$; общая стоимость равна 5. Поэтому верхний путь дешевле, и $M(s_{3,2})=4$. Так вычисляется подграф для третьей стадии, показанный на рис. 6.3. Следующий граф и значения M

показаны на рис. 6.4. Наконец, последний граф и все значения M показаны на рис. 6.5. Из этого рисунка видно, что оптимальный путь — самый верхний.

6.3. ОПТИМИЗАЦИЯ С ПОМОЩЬЮ ВЫЧИСЛИТЕЛЬНОЙ МАШИНЫ

В этом параграфе рассмотрен еще один пример и описана общая программа для отыскания оптимальных путей в конечных ориентированных графах с известными длинами (ценами) дуг.

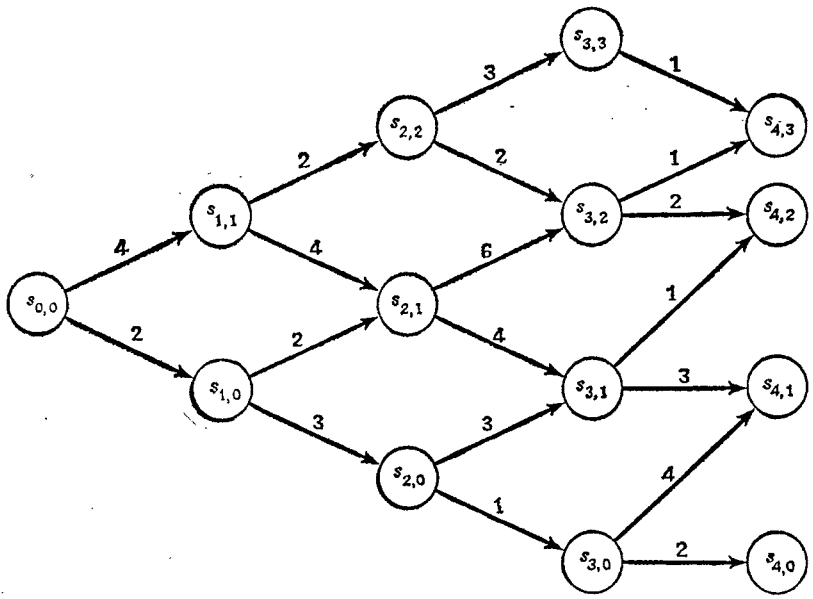


Рис. 6.6. Задача с конечной стадией.

Пример 2. Рассмотрим рис. 6.6. Здесь задача состоит в отыскании оптимального пути от состояния $s_{0,0}$ к любому состоянию $s_{4,k}$ последней стадии. Как и выше, она решается последовательным вычислением оптимального пути от стадии 3 к стадии 4, от 2 к 3, от 1 к 2, от 0 к 1.

Обозначим через $M(s_{i,k})$ стоимость оптимального пути от $s_{i,k}$ до любого из состояний $s_{4,j}$, $i=3, 2, 1, 0$. Очевидно, $M(4, k)=0$ для всех $k=0, 1, 2, 3$. Значения $M(s_{3,k})$ также считываются непосредственно с рисунка: $M(s_{3,0})=2$, $M(s_{3,1})=1$, $M(s_{3,2})=1$, $M(s_{3,3})=1$. Очередной подграф показан на рис. 6.7,*а*.

Подграф для второй стадии показан на рис. 6.7,*б*, а окончательный граф с оптимальным путем — на рис. 6.8.

Описанный способ позволяет резко сократить число шагов, необходимых для отыскания оптимальных путей в больших, но конечных многостадийных процессах принятия решений. Многие такие задачи возникают в транспортных задачах, когда необхо-

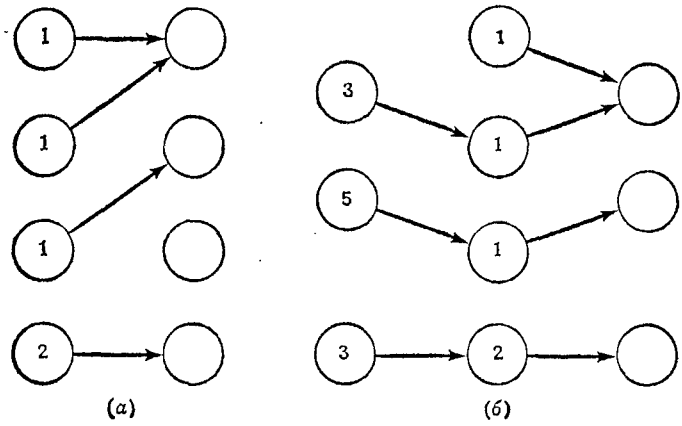


Рис. 6.7. Первые шаги решения.

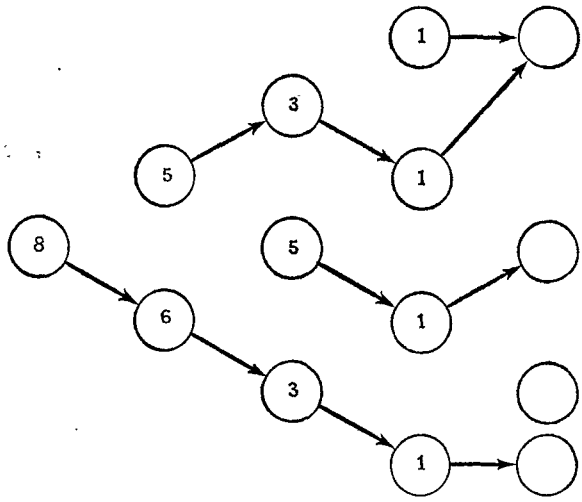


Рис. 6.8. Решение задачи, представленной на рис. 6.6.

димо оптимизировать пути перевозок, потоки и центры распределения тех или иных материальных ценностей с целью минимизации общей стоимости или наибольшего времени доставки. Другие проблемы размещения и организации возникают в многостадийных процессах производства, где целью может быть мак-

симизация объема выпускаемой продукции. (Многие задачи такого рода обсуждены в [2].)

Аналогичные соображения используются в вариационном исчислении, где из принципа оптимальности выводятся уравнения Эйлера—Лагранжа. Задачи оптимального управления предоставляют еще более широкие возможности приложения этих идей (см. [3]). Применение разностных методов к аналитическим задачам такого рода дает промежуточный класс задач, поддающихся решению на машинах.

Общий случай. Заключим этот параграф программой для отыскания кратчайшего пути от вершины a к вершине b в любом ориентированном графе $\vec{G}$, дуги которого α_i имеют заданные положительные длины $\lambda(\alpha_i)$. Обобщение описанного выше способа и использование принципа оптимальности сведут наши действия к следующему (на петли можно не обращать внимания, ибо кратчайшие пути простые).

Стадия 0. Зададим список всех вершин a_i. Положим $\lambda(a)=0$, $\lambda(a_i)=\infty$ для $a_i \neq a$. (Функция $\lambda(a_i)$ будет представлять собой *известную к текущему моменту* оценку сверху для длины оптимального пути от a к a_i.)

Стадия 1. Зададим список всех дуг $\overrightarrow{ax_i}=\xi_i$, выходящих из a. Для каждой вершины x_i по очереди сравним $\lambda(\xi_i)$ с $\lambda(x_i)$. Если $\lambda(\xi_i) \geqslant \lambda(x_i)$, перейдем к x_{i+1}. Если $\lambda(\xi_i) < \lambda(x_i)$, заменим $\lambda(x_i)$ на $\lambda(\xi_i)$ и отметим путь $\overrightarrow{ax_i}$ как кратчайший путь от a до x_i, *найденный к текущему моменту*.

Стадия 2. Для каждой вершины $x_i \neq a$, рассмотренной на стадии 1, зададим список всех дуг $\overrightarrow{x_iy_j}=\eta_{ij}$, выходящих из x_i. Для каждой дуги η_{ij} по очереди вычислим $\mu_{ij}=\lambda(x_i)+\lambda(\eta_{ij})$. Если $\mu_{ij} \geqslant \lambda(y_j)$, перейдем к следующей дуге. Если $\mu_{ij} < \lambda(y_j)$, заменим $\lambda(y_j)$ на μ_{ij} и отметим путь $\overrightarrow{ax_iy_j}$ как кратчайший путь от a до y_j, *найденный к текущему моменту*.

Общая стадия описывается рекурсивно.

Стадия n. Для каждой вершины a_i, у которой $\lambda(a_i)$ была уменьшена на стадии $n-1$, построим список всех дуг $\overrightarrow{a_iz_k}=\zeta_{ik}$, выходящих из a_i. Для всех ζ_{ik} по очереди вычислим $\nu_{ik}=\lambda(a_i)+\lambda(\zeta_{ik})$ и сравним ν_{ik} с $\lambda(z_k)$. Если $\nu_{ik} \geqslant \lambda(z_k)$, перейдем к следующей дуге $(\zeta_{i,\,k+1})$ либо $\zeta_{i+1,\,1}$. Если $\nu_{ik} < \lambda(z_k)$, заменим $\lambda(z_k)$ на ν_{ik} и отметим новый *оптимальный путь* $\overrightarrow{a \ldots a_iz_k}$ *от* a *к* z_k, *найденный к текущему моменту* (т. е. первый найденный путь от a до z_k длины $\leqslant \nu_{ik}$).

На некоторой стадии с номером m, не превосходящим числа вершин графа $\vec{G}$, ни один новый кратчайший путь не появится. Вычисленные к этому моменту $\lambda(a_i)$ будут длинами оптимальных путей.

Приведем теперь программу на АЛГОЛе, реализующую описанный алгоритм. Пусть 1, 2, ... N — вершины ориентированного графа $\vec{G}$. В память машины следует записать число N и $N \times N$-матрицу Z, элемент Z_{ij} которой равен длине дуги от вершины i к вершине j. Если такой дуги вообще нет, следует положить $Z_{ij} = \infty$ (на самом деле — очень большое положительное число, которое еще можно записать, скажем, 10^{20}). Пусть 1 — начальная вершина. Программа будет вычислять длины кратчайших путей от 1 к остальным вершинам $k = 2, 3, \ldots, N$. Используются следующие обозначения:

L_k — текущая оценка длины оптимального пути от 1 к k.

m_k — число вершин на текущем оптимальном пути от 1 к k, исключая 1 и k.

P_k — массив последовательных вершин $P_{k_1}, P_{k_2}, \ldots, P_{k_{m(k)}}$ текущего оптимального пути от 1 к k, исключая 1 и k.

d — массив вершин $d_1, d_2, \ldots, d_r$, для которых оценка длины оптимального пути L_{d_i} была уменьшена при предыдущей итерации.

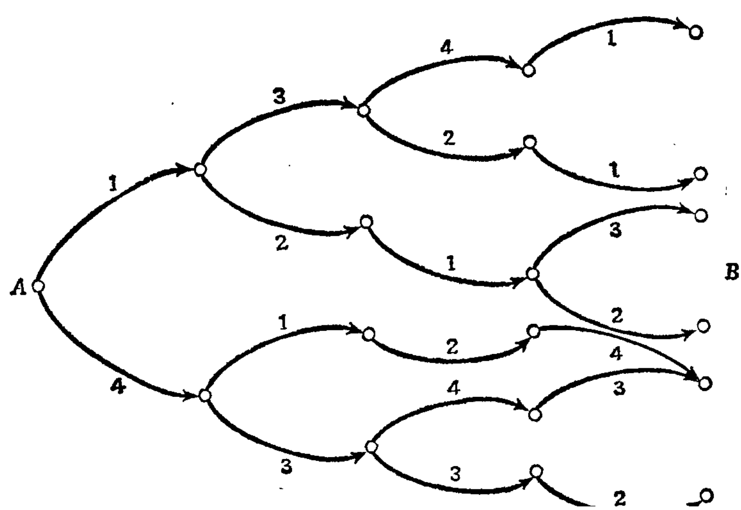
```
begin integer N, i, j, k, q, r, rtemp;
      integer array m[1:50], d[1:50], dtemp[1:50], P[1:50, 1:50];
      real nu; real array Z[1:50, 1:50], L[1:50];
      for i:=2 step 1 until N do
          begin L[i]:=10+20; m[i]:=0; d[i]:=i end
      r:=N-1;
A:    rtemp:=0;
      for q:=1 step 1 until r do
      begin i:=d[q];
            for k:=2 step 1 until N do
            begin nu:=L[i]+Z[i, k]
                  if nu < L[k] then
                   begin L[k]:=nu; m[k]:=m[i]+1;
                         for j:=1 step 1 until m[i] do
                             P[k, j]:=P[i, j];
                             P[k, m[k]]:=i;
                             rtemp:=rtemp+1;
```

```
                    dtemp [rtemp]: =k
                end
            end
    end
    if rtemp > 0 then
            begin r: =rtemp;
                for i: =1 step 1 until r do d[i]: =dtemp [i];
                go to A
            end
end
```

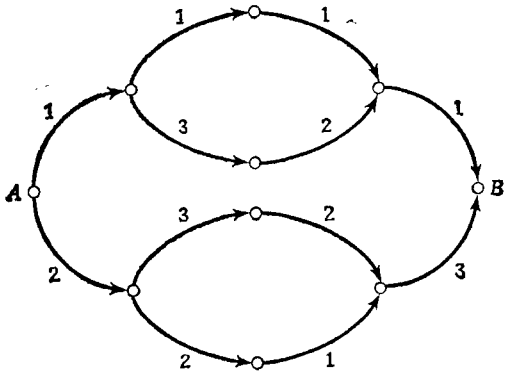
После исполнения этой программы переменной L_k будет присвоено значение длины кратчайшего (оптимального) пути от вершины 1 к вершине k. Если $L_k \geqslant 10^{20}$, такой путь не найден. Промежуточные вершины на этом пути записаны в k-й строке массива P.

УПРАЖНЕНИЯ А

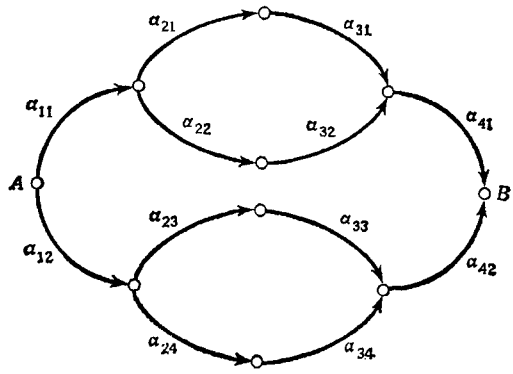
1. Найти кратчайшие пути в следующем помеченном ориентированном графе (числа обозначают длины дуг):



2. Найти кратчайший путь от A до B в следующем помеченном графе:



3. Написать программу на АЛГОЛе для отыскания кратчайшего пути от A до B в следующем помеченном графе:



6.4. ЛОГИЧЕСКОЕ ПРОЕКТИРОВАНИЕ

Связи с теорией булевых алгебр и логикой дали повод называть проектирование вентильных схем вычислительных машин *логическим проектированием*. Существует прямой способ синтеза вентильных схем, реализующих данную булеву функцию $f(x_1, \ldots, x_r) = f(x)$ от r входных булевых переменных, т. е. функцию f: $2^r \longrightarrow 2$. Мы опишем сейчас этот способ с некоторыми подробностями. Схемы синтезируются из элементов И, ИЛИ и инверторов.

Несколько слов относительно обозначений и терминов. В инженерном проектировании употребительна система обозначений, отличающаяся от стандартной математической. Символ бинарной операции $+$ используется вместо символа $\vee$ из гл. 5. Таким образом, $0+0=0$, $0+1=1+0=1+1=1$. Между тем, алгебра-

исты обычно используют $+$ для обозначения операции сложения по модулю 2:

$$0+0=0, \quad 1+1=0,$$
$$0+1=1+0=1,$$

тогда как в инженерном проектировании эта последняя операция обозначается $\oplus$. Далее, операция $\wedge$ обозначается точкой, которая может опускаться: $a \wedge b \wedge c$ можно записать $a \cdot b \cdot c$ или abc. Причины этого связаны с особенностями алфавитов стандартных выходных устройств, которые выпечатывают логические формулы, записанные предварительно на магнитной ленте.

В этой главе мы будем пользоваться знаком $\vee$, как в гл. 5, а знак $\wedge$ будем опускать, как при обычном умножении.

Следующие определения вводят фрагмент употребительной терминологии логического проектирования.

Определение. *Литералом* называется переменная или штрихованная переменная (штрих обозначает дополнение). Например, x, x', y, a', b — 5 различных литералов.

Определение. *Мультипликативным одночленом* называется булево произведение нескольких литералов. *Аддитивным одночленом* называется булева сумма нескольких литералов.

Примеры мультипликативных одночленов: ab, $ab'c'$, xyz и $z'x'$. Примеры аддитивных одночленов: $(a \vee b \vee c)$, $(a' \vee b')$, $(a \vee b' \vee c)$ и $(x \vee y)$.

Определение. *Суммой произведений* называется булева сумма нескольких мультипликативных одночленов. *Произведением сумм* называется булево произведение нескольких аддитивных одночленов.

Примеры сумм произведений: $ab \vee c$, $ab \vee a'b'$, $ac \vee abc \vee a'c'$. Примеры произведений сумм: $(a \vee b)(c \vee d)$, $(a \vee b)(a' \vee b')$, $(a \vee b \vee c)(a' \vee b' \vee c)$ и $(a \vee b' \vee c')$.

Теперь опишем, как осуществляется синтез вентильных схем, реализующих булевы функции от n двоичных переменных. Сохраняется план доказательства теоремы 3 гл. 5, утверждающей существование булевой суммы произведений, эквивалентной любой функции f: $2^n \longrightarrow 2$.

Прежде всего составим таблицу значений функции в зависимости от аргументов. Будем иллюстрировать общие объяснения примером рис. 6.9.

После этого для каждой системы значений переменных, т. е. для каждой строки таблицы, построим мультипликативный одночлен, принимающий значение 1 для данных значений переменных. Так, строке $A=0$, $B=1$, $C=0$ отвечает одночлен $A'BC'$. Строке

$A=1$, $B=1$, $C=0$ отвечает одночлен ABC'. Эти одночлены внесены в столбец «мультипликативные одночлены».

В столбец «аддитивные одночлены» внесены дополнения мультипликативных одночленов той же строки. Так, в строке, где стоит $AB'C'$, добавляется одночлен $(A' \vee B \vee C)$, а в строке

A	B	C	$f(A, B, C)$	Мультипликативные одночлены	Аддитивные одночлены
0	0	0	0	$A'B'C'$	$A \vee B \vee C$
0	0	1	1	$A'B'C$	$A \vee B \vee C'$
0	1	0	0	$A'BC'$	$A \vee B' \vee C$
0	1	1	1	$A'BC$	$A \vee B' \vee C'$
1	0	0	0	$AB'C'$	$A' \vee B \vee C$
1	0	1	1	$AB'C$	$A' \vee B \vee C'$
1	1	0	1	ABC'	$A' \vee B' \vee C$
1	1	1	1	ABC	$A' \vee B' \vee C'$

$$f(A, B, C) = A'B'C \vee A'BC \vee AB'C \vee ABC' \vee ABC = (A \vee B \vee C)(A \vee B' \vee C)(A' \vee B \vee C).$$

Рис. 6.9.

$A'BC'$ добавляется одночлен $(A \vee B' \vee C)$. (Аддитивные одночлены используются для построения соответствующего произведения сумм.)

Представление нашей функции в виде суммы произведений получится, если просто взять сумму мультипликативных одночленов в тех строках, которые отвечают значению функции 1.

В нашей таблице следует выбрать одночлены из второй, четвертой, шестой, седьмой и восьмой строки:

$$A'B'C,\ A'BC,\ AB'C,\ ABC',\ ABC.$$

Таким образом,

$$f(A, B, C) = A'B'C \vee A'BC \vee AB'C \vee ABC' \vee ABC.$$

Значение правой части равно 1 точно при тех значениях аргументов, при которых значение функции f равно 1, потому что один из одночленов справа равен 1 именно при этих значениях. Этот общий алгоритм работает для любой данной функции f: $\mathbf{2}^n \longrightarrow \mathbf{2}$ и приводит к следующему результату.

Теорема 2. *Любую булеву функцию от n переменных можно представить в виде суммы произведений.*

Теперь нетрудно построить вентильную схему, реализующую это булево выражение (см. рис. 6.10). Для каждого мультипликативного одночлена нужен элемент И с таким количеством входов, из скольких литералов состоит одночлен. Кроме того, нужен элемент ИЛИ с таким количеством входов, сколько имеется мультипликативных одночленов. Выход каждого элемента

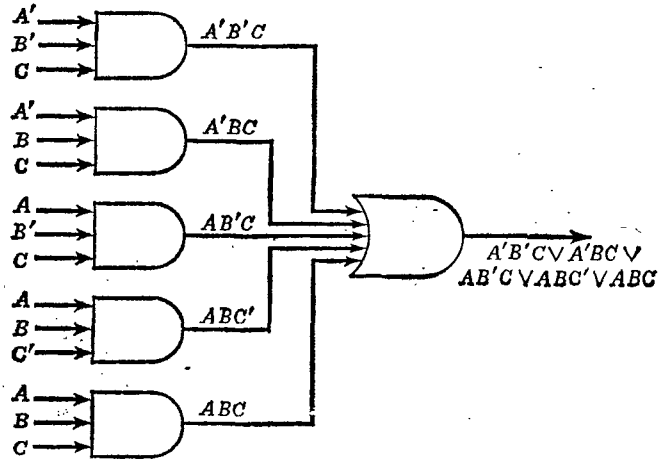


Рис. 6.10. Вентильная схема для функции, представленной на рис. 6.9.

И подается на один из входов элемента ИЛИ. На вход каждого элемента И подаются значения литералов, из которых состоит соответствующий одночлен.

Предполагается, что для значений A, A', B, B', C и C' имеются отдельные каналы. Можно обойтись без каналов для A', B', C', применяя инверторы. В большинстве случаев, однако, такие каналы имеются. Кроме того, с каналов для A, B, C можно снимать значения этих переменных и подавать их на вход любого количества элементов И. (Возникающие здесь технические проблемы, связанные с распределением нагрузки, мы не обсуждаем.)

Аналогично строится произведение сумм, представляющее функцию, и отвечающая ему схема ИЛИ—И. С этой целью описанная выше процедура применяется к функции $f'(A, B, C)$, дополнительной к функции f. Иными словами, отбираются мультипликативные одночлены из строк, где $f'(A, B, C)=1$, т. е. $f(A, B, C)=0$. Их сумма дает представление функции f' в виде суммы произведений: $f'(A, B, C)=(A'B'C' \vee A'BC' \vee AB'C)$. Еще раз применяя дополнение, находим $f''=f$ в виде произве-

дения сумм. Для примера рис. 6.9 это дает

$$f=(A'B'C \vee A'BC' \vee AB'C')' =$$
$$=(A \vee B \vee C)(A \vee B' \vee C)(A' \vee B \vee C)$$

(воспользоваться законом де Моргана). В общем случае это рассуждение приводит к следующему результату.

Теорема 3. *Любая булева функция от n переменных может быть представлена в виде произведения сумм.*

Вместо того чтобы дважды применять операцию дополнения, мы можем просто представить f в виде произведения тех аддитивных одночленов из последнего столбца таблицы, которые стоят в строках, отвечающих нулевым значениям f. Окончательное выражение будет тем же самым. Реализация его требует по одному элементу ИЛИ для каждого аддитивного одночлена с таким количеством входов, сколько литералов составляет этот одночлен, а также одного элемента И с таким количеством входов, сколько имеется элементов ИЛИ.

Функция, представленная рис. 6.9, реализуется схемой типа ИЛИ—И (рис. 6.11).

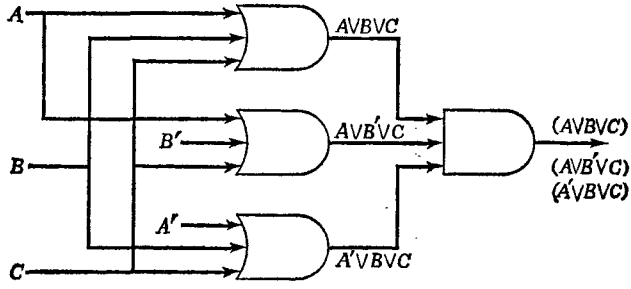


Рис. 6.11. Схема типа ИЛИ—И.

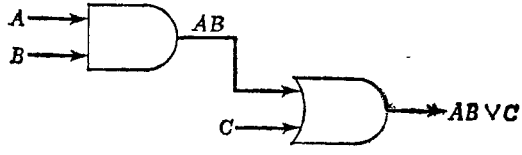


Рис. 6.12. Упрощенная схема.

Недостаток описанного метода—его неэкономичность. С помощью булевых преобразований можно значительно уменьшить число необходимых элементов и упростить схему. Например, на рис. 6.12 показана схема, реализующая ту же функцию f, что и две предыдущих схемы.

6.5. ЭЛЕМЕНТЫ НЕ-И И НЕ-ИЛИ

В предыдущем параграфе мы представляли булевы функции с помощью двух бинарных операций (сумма и произведение) и одной унарной (дополнение). Было показано, что их достаточно для синтеза любых булевых функций от n переменных. Естественно возникает вопрос, нельзя ли обойтись меньшим числом операций. Законы де Моргана позволяют немедленно ответить утвердительно. Можно обойтись произведением и дополнением, ибо $(X'Y')' = X \vee Y$, так что суммы выражаются через них. Аналогично $(X' \vee Y')' = XY$, так что можно обойтись суммой и дополнением. Таким образом, в любом выражении, содержащем произведения, суммы и дополнения, систематической процедурой можно исключить либо суммы, либо произведения.

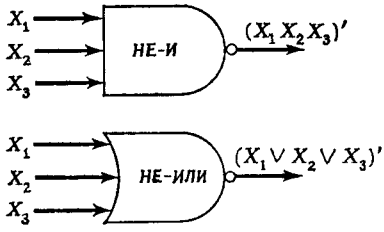


Рис. 6.13. Универсальные элементы.

Встает естественный вопрос, нельзя ли обойтись всего одной операцией для построения всех булевых функций? Ответ снова утвердителен: достаточна операция Пирса $\downarrow$: $x \downarrow y = x'y'$, а также «штрих Шеффера» $|$: $z \mid y = z' \vee y'$. Для доказательства универсальности $\downarrow$ заметим, что $x' = x \downarrow x$, после чего, построив дополнение, имеем $xy = x' \downarrow y'$. Из соображений двойственности видно, что $\mid$ универсальна.

Широко употребительны элементы, реализующие эти операции не обязательно в бинарном варианте. Они называются элементами НЕ-И и НЕ-ИЛИ. На рис. 6.13 приведено их схематическое представление.

Следует помнить, что ни одна из этих операций не ассоциативна:

$$(x \downarrow y) \downarrow z \neq x \downarrow (y \downarrow z).$$

Поэтому элементы НЕ-И, НЕ-ИЛИ с несколькими входами нельзя реализовать параллельным соединением таких элементов с двумя входами.

Элемент НЕ-ИЛИ с n входами реализует функцию $P(x_1, x_2, \ldots x_n) = x_1' \cdot x_2' \ldots \cdot x_n'$, а элемент НЕ-И — функцию $N(x_1, x_2, \ldots, x_n) = x_1' \vee x_2' \vee \ldots \vee x_n'$.

На рис. 6.14 показана двухуровневая схема типа НЕ-И — НЕ-И и двухуровневая схема типа НЕ-ИЛИ — НЕ-ИЛИ. Заметим, что схема НЕ-И — НЕ-И реализует сумму произведений входов. Поэтому, заменив левые элементы на элементы И, а правый — на элемент ИЛИ, мы реализуем ту же функцию. Анало-

гичное рассуждение показывает, что схему НЕ-ИЛИ—НЕ-ИЛИ можно заменить схемой ИЛИ—И.

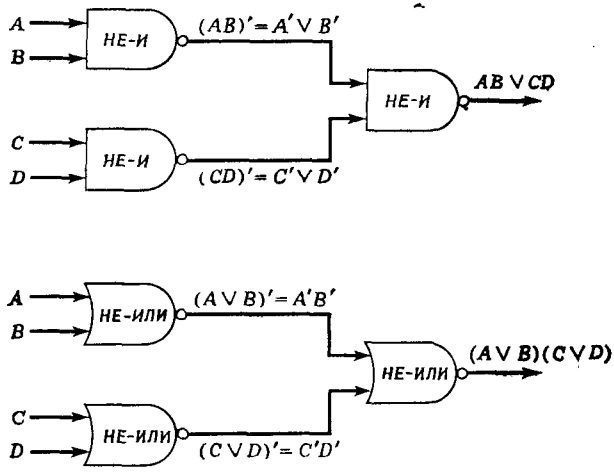


Рис. 6.14. Универсальная схема.

Резюмируем сказанное. Пусть функция f: $2^n \longrightarrow 2$ от n двоичных переменных $x_1, x_2, \ldots, x_n$ представлена в виде

$$f = \bigvee_l Z_l, \text{ где } Z_l = \bigwedge_{i=1}^{n} Y_{i\,(l)} \text{ и } Y_{i\,(l)} = x_i \text{ или } x_i'. \qquad (1)$$

В наших теперешних обозначениях это же представление имеет вид

$$f = \sum_l \left(\prod_{i=1}^{n} Y_{i\,(l)} \right), \text{ где } Y_{i\,(l)} \text{—литералы } (x_i \text{ или } x_i'). \qquad (2)$$

Так как каждый мультипликативный одночлен реализуется одним элементом И, а в сумме имеется не более 2^n одночленов, любую булеву функцию f: $2^n \longrightarrow 2$ можно реализовать, использовав не более 2^n элементов И с не более n входами (на самом деле всегда можно обойтись $n-1$ элементами и часто—гораздо меньшим их количеством).

По двойственности, всякую функцию f: $2^n \longrightarrow 2$ можно представить в виде

$$f = \prod_l \left(\sum_{i=1}^{n} Z_{i\,(l)} \right), \text{ где } Z_i \text{—литералы } (x_i \text{ или } x_i'). \qquad (3)$$

Напомним, наконец, что любой булев многочлен от $x_1, \ldots, x_n$ можно привести к виду (2) и (3), пользуясь только тождествами L1—L10 (аксиомами булевых алгебр).

Заметим, что значительная часть рассмотренных в этом параграфе булевых функций принадлежала к классу симметрических функций в смысле следующего определения.

Определение. Функция $f\colon 2^n \to 2$ называется *симметрической*, если

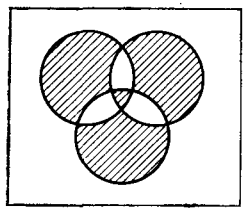
$$f(X_1, \ldots, X_n) = f(X_{\sigma 1}, \ldots, X_{\sigma n})$$

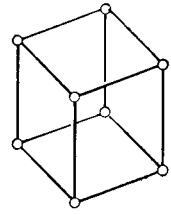
для любой перестановки σ аргументов X_i (общее обсуждение перестановок $\sigma\colon \mathbf{n} \to \mathbf{n}$ см. в § 7.9). Функция называется $\pm$*симметрической*, если она инвариантна относительно перестановок литералов $Y_1, Y_2, \ldots, Y_{2n}$, где Y_i есть X_i или X_i'. Она называется *частично симметрической*, если она симметрична по подмножеству, содержащему не менее двух своих аргументов.

Важный пример симметрической функции

$$\alpha(a, b, c) = a + b + c = abc \vee ab'c' \vee a'bc' \vee a'b'c$$

представлен следующей диаграммой Венна:





Ее носитель, подмножество 2^3, на котором она принимает значения 1, состоит из всех «красных» вершин бихроматического графа для 2^3.

Можно положить

$$a + b + c + d = (a + b + c) + d = (a + b) + (c + d)$$

и аналогично определить булеву сумму mod 2 любого числа n переменных. Схемы, реализующие такие функции, требуют удивительно большого числа элементов.

6.6. ПРОБЛЕМА МИНИМИЗАЦИИ

Основная проблема минимизации при проектировании вентильных схем состоит в следующем: как по заданному булеву многочлену найти эквивалентное выражение, которое в некотором смысле было бы «самым простым»?

Определение. *Канонической суммой произведений* от n переменных называется булева сумма (без повторений) мультипликативных одночленов, в каждый из которых каждая переменная входит ровно по одному разу.

В гл. 5 такие выражения назывались «дизъюнктивными нормальными формами».

Способ, описанный в предыдущем параграфе, приводит к выражениям именно такого вида. Мы приведем два определения минимальности сумм произведений; двойственные определения относятся к произведениям сумм. Оба определения не учитывают операций дополнения (триггеры обычно имеют два противоположных выхода, так что значения переменной и ее дополнения выдаются одновременно).

Определение минимальности. 1. Выражение Ф типа «сумма произведений» называется *минимальным по литералам*, если никакое эквивалентное ему выражение Ψ не содержит меньшего числа литералов.

Определение минимальности. 2. Выражение Ф типа «сумма произведений» называется *минимальным по одночленам*, если никакое эквивалентное ему выражение Ψ не содержит меньшего числа мультипликативных одночленов либо такого же числа мультипликативных одночленов с меньшим числом литералов.

Эти два вида минимальности согласуются с требованиями простоты реализации: при первом минимизируется число входов, а при втором — число элементов И и входов.

Часто используются еще два вида минимальности: «минимум суммы числа литералов и одночленов» и «минимум числа необходимых электронных схем (диодов, транзисторов...)». Оба они легко сводятся к предыдущим, и к ним применим тот же способ минимизации. Мы опишем такой способ лишь для сумм произведений. Для произведений сумм все изменения согласуются с двойственностью.

Определение. Выражение Ψ *влечет* выражение Ф, если не существует такого набора значений переменных, при котором Ψ принимает значение 1, а Ф — значение 0. Мультипликативный одночлен α, который влечет f, называется *импликантом* функции f.

Теорема 4. *Пусть* Ф — *сумма произведений, представляющая функцию* f. *Тогда любой мультипликативный одночлен, входящий в* Ф, *влечет* f.

Пояснение. Одночлен $\alpha = ab'c'$ влечет, например, $ab'c' \vee a'b'c$, а $\alpha = ab$ влечет $ab \vee ab'c' \vee a'b'$.

Доказательство. Если при данных значениях переменных одночлен β принимает значение 1, то любая сумма одночленов, в которую входит β, также принимает значение 1.

Минимальное выражение Ф типа сумма произведений для функции f должно состоять только из суммы импликантов функции f, ибо Ф обязательно влечет f.

Определение. *Простым импликантом* функции f называется такой импликант α, который перестает быть импликантом после удаления любого литерала.

Простые импликанты f — это самые короткие из импликантов, состоящих из одних и тех же литералов.

Существует простой способ проверять, является ли данный мультипликативный одночлен импликантом выражения Ф. Припишем значения 1 всем нештрихованным литералам одночлена и значения нуль всем литералам, помеченным штрихом. Одночлен примет значение 1. После этого проверим, принимает ли Ф значение 1 при любых значениях остальных литералов.

Например, $\alpha = xy$ влечет $\Phi = xyz \vee xyz'$, потому что при $x = 1$, $y = 1$ имеем $\Phi = 1 \cdot 1 \cdot z \vee 1 \cdot 1 \cdot z' = z \vee z' = 1$. Одночлен $\beta = x'y$ влечет $\Psi = wx'yz \vee wx'yz' \vee w'x'yz \vee w'x'yz'$, потому что при $x = 0$, $y = 1$ имеем

$$\Psi = w \cdot 1 \cdot 1 \cdot z \vee w \cdot 1 \cdot 1 \cdot z' \vee w' \cdot 1 \cdot 1 \cdot z \vee w' \cdot 1 \cdot 1 \cdot z' =$$
$$= wz \vee wz' \vee w'z \vee w'z' = w(z \vee z') \vee w'(z \vee z') = w \vee w' = 1.$$

В этих двух примерах импликанты α, β являются простыми, ибо сокращение α или β на один из литералов нарушает их свойство быть импликантами.

Теорема 5. *Функция f эквивалентна сумме своих простых импликантов.*

Доказательство. Очевидно, сумма простых импликантов влечет f. Далее, любой набор значений переменных, при которых функция f равна 1, обращает в 1 некоторый ее импликант. Поэтому, сокращая этот импликант до простого, мы можем аналогичным образом набрать достаточное количество простых импликантов, так что их сумма будет принимать значение 1 тогда и только тогда, когда значение f равно 1.

Определение. Выражение Ф типа «сумма произведений», эквивалентное функции f, называется *неизбыточным*, если:

(i) любой мультипликативный одночлен, входящий в Ф, является простым импликантом функции f,

(ii) устранение любого одночлена из Ф нарушает эквивалентность $\Phi = f$.

Теорема 6. *Всякое выражение* Ф *типа «сумма произведений» минимальное по литералам или по одночленам, неизбыточно.*

Доказательство. Если Ф не является неизбыточной, то либо один из входящих в нее одночленов не является простым импликантом, либо один из одночленов можно вычеркнуть, не изменив представляемой функции. И в том, и в другом случае Ф не минимальна по литералам и одночленам.

Чтобы найти для f неизбыточное представление, следует вычислить все простые импликанты и затем отыскать их неизбыточное подмножество.

Распознавание избыточности не столь просто, как это может показаться на поверхностный взгляд. Например, представление $\Psi = ab' \vee ac' \vee b'c$ является избыточным, хотя каждый одночлен в нем является простым импликантом. Действительно, одночлен ab' можно вычеркнуть, не изменив функции. (Если одночлен α является лишним в выражении Φ, то α влечет Φ, и обратно; так можно проверять выражения на избыточность.)

Итак, мы разделили задачу на следующие этапы:

(i) порождение множества простых импликантов,

(ii) порождение неизбыточных сумм таких импликантов,

(iii) выбор из этих сумм минимальных.

Начнем с методики порождения импликантов.

Определение. Мультипликативный одночлен α *поглощает* мультипликативный одночлен β, если каждый литерал, входящий в β, входит также в α. Например, abc поглощает ab, $ac'd'$ поглощает ad', ab поглощает ab и т. д.

Определение. Мультипликативный одночлен α является *дополнением мультипликативного одночлена* β *по отношению к выражению* Φ, если α поглощает β и если каждая переменная, входящая в Φ, входит в α.

Например, все одночлены $abc'd'$, $abcd'$, $abc'd$, $abcd$ являются дополнениями ab по отношению к выражению $\Phi = abc' \vee d \vee cd \vee ab$.

Теорема 7. *Пусть* Φ *— представление функции* f *в виде канонической суммы произведений. Если одночлен* α *влечет* f, *то все его дополнения входят в* Φ.

В самом деле, пусть какое-то дополнение β не входит в Φ. Рассмотрим такую систему значений переменных из Φ, для которой $\beta = 1$. Тогда $\alpha = 1$, но $\Phi = 0$, так что α не влечет Φ.

Например, если ac' влечет функцию f от переменных a, b, c, d, то все одночлены $ab'c'd'$, $ab'c'd$, $abc'd'$, $abc'd$ должны входить в дизъюнктивную нормальную форму Φ для f.

6.7. ПЕРЕЧИСЛЕНИЕ ПРОСТЫХ ИМПЛИКАНТОВ

Опишем теперь, как систематически перечислять простые импликанты данной булевой функции f. Начнем с канонического представления Φ для f в виде суммы произведений. После этого, объединяя в нем попарно одночлены всевозможными способами, для каждой пары вида $\gamma x \vee \gamma x'$ выпишем γ. К получившемуся множеству более коротких одночленов снова применим это правило, и т. д. Закончив список, вычеркнем из него все одночлены, ко-

торые поглощают какой-нибудь одночлен из списка: они не могут быть простыми импликантами.

Пример 6. Применим этот рецепт к канонической сумме произведений $\Phi = abcd \vee ab'cd' \vee ab'cd \vee abcd' \vee a'b'cd'$. Объединяя первый и третий одночлены, получим acd. Первый и четвертый дают abc, второй и третий $ab'c$ и т. д. Окончательный список импликантов:

$abcd$	acd	ac
$ab'cd'$	abc	
$ab'cd$	$ab'c$	
$abcd'$	$b'cd'$	
$a'b'cd'$	acd'	

В нем имеется два одночлена, которые не поглощают других одночленов из списка: это ac и $b'cd'$. Они являются простыми импликантами, и представление $ac \vee b'cd'$ для Φ оказывается здесь уже минимальным, хотя в общем случае процедура на этом не кончается.

Этот метод был предложен Куайном; Мак-Клоски ввел в него дальнейшие усовершенствования.

Шаг 1. Упорядочим переменные и будем писать их в каждом одночлене в выбранном порядке. После этого представим каждый одночлен последовательностью из 1,0 и —, ставя на i-м месте 1, если i-я переменная входит в одночлен без штриха, 0, если входит со штрихом, и знак —, если не входит. Например, выражение $\Phi = abc \vee ac' \vee ad'$ перепишем в виде $111— \vee 1—0— \vee 1— —0$, а выражение $\Psi = ad' \vee a'c \vee abcd$ — в виде $1— —0 \vee 0—1— \vee 1111$.

Шаг 2. Разобъем двоичные выражения, отвечающие одночленам, на классы по числу единиц, и расположим списки этих классов в порядке возрастания этого числа. Для выражения

$$\Phi = abc'd \vee ab'c'd \vee abc'd' \vee ab'c'd' \vee a'b'cd' \vee a'b'c'd \vee a'b'c'd'$$

получится список

0 0 0 0
0 0 0 1
0 0 1 0
1 0 0 0
1 1 0 0
1 0 0 1
1 1 0 1

Шаг 3. Тождество $yx \vee yx' = y$ может быть применимо только к парам, элементы которых лежат в соседних классах. Если мы поместим все y, происшедшие из данной пары соседних классов, в один класс, то на следующем этапе придется снова сравнивать только пары из соседних классов. Члены, входящие в пары такого вида, можно пометить знаком $\surd$; в дальнейшем они заведомо не останутся в списке простых импликантов. Предыдущий список после обработки будет выглядеть так:

$\surd$ 0 0 0 0	$\surd$ 0 0 0 —	— 0 0 —
$\surd$ 0 0 0 1	0 0 — 0	1 — 0 —
$\surd$ 0 0 1 0	$\surd$ — 0 0 0	
$\surd$ 1 0 0 0	$\surd$ — 0 0 1	
$\surd$ 1 1 0 0	$\surd$ 1 — 0 0	
$\surd$ 1 0 0 1	$\surd$ 1 0 0 —	
$\surd$ 1 1 0 1	$\surd$ 1 1 0 —	
	$\surd$ 1 — 0 1	

Простые импликанты: $a'b'd'$, $b'c'$ и ac'.

Таков *алгоритм Куайна — Мак-Клоски* перечисления простых импликантов. Заметим, что подходящие пары отыскиваются простым наблюдением: элементы пары должны отличаться ровно в одной позиции, и в этой позиции не должно быть прочерков.

Шаг 4. Перечислив все простые импликанты для Ф, нам остается выбрать из них такое подмножество, что Ф влечет сумму его элементов. Сверх того, выбранное выражение должно быть минимальным (в смысле одного из определений).

Эта задача решается с помощью таблиц простых импликантов, примеры которых показаны на рис. 6.15.

	$a'b'c'd'$	$a'b'c'd$	$a'b'cd'$	$a'b'cd$	$abc'd'$	$abc'd$	$ab'c'd$	$abc'd$
$a'b'd'$	×		×					
$b'c'$	×	×		×			×	
ac'				×	×	×	×	

	0000	0001	0010	1000	1100	1001	1101
00—0	×		×				
—00—	×	×		×		×	
1—0—				×	×	×	×

Рис. 6.15. Таблицы простых импликантов для примера 6.

Столбцы такой таблицы перенумерованы мультипликативными одночленами, входящими в каноническое представление Ф. Строки перенумерованы простыми импликантами. Крестики стоят в тех позициях, где одночлен поглощает простой импликант.

Определение. Будем говорить, что одночлен α *покрывает* одночлен β, если β поглощает α.

Выражение Φ влечет Ψ—сумму простых импликантов, если каждый одночлен из Φ покрывается подходящим одночленом из Ψ. Действительно, иначе существуют такие значения переменных, которые придают непокрытому одночлену γ (и потому Φ) значение 1, тогда как Ψ принимает значение 0.

В частности, в таблице импликантов выбранное подмножество строк должно быть таким, чтобы в каждом столбце хоть один крестик находился в выбранной строке.

Таким образом, приходим к следующему правилу: если в столбце имеется лишь один крестик, то соответствующая строка должна быть выбрана обязательно. Множество таких строк (точнее, импликантов), назовем *ядром* задачи.

На рис. 6.15 это замечание заставляет выбрать все импликанты, ибо $a'b'c'd$ покрывается только $b'c'$, $a'b'cd'$ покрывается только $a'b'd'$, а $abc'd'$ и $abc'd$ покрываются только ac'. Значит, все простые импликанты лежат в ядре, и минимальное в любом смысле выражение имеет вид $a'b'd' \vee b'c' \vee ac'$.

Во многих случаях, однако, после выделения ядра еще останутся члены выражения Φ, подлежащие покрытию. Может оказаться, что останутся простые импликанты, которые не покрывают ни один из одночленов в Φ, не покрываемых элементами ядра. Эти импликанты влекут сумму элементов ядра и являются *совершенно лишними*: они не войдут ни в одно минимальное выражение. Пример этого показан на рис. 6.16, *а*. Устранив импли-

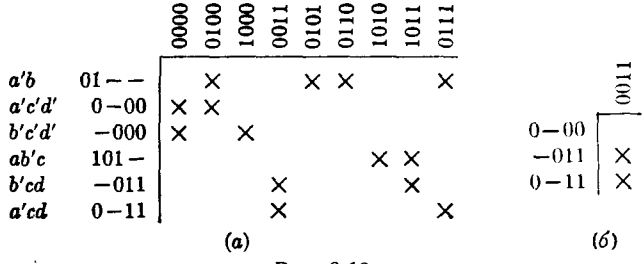


Рис. 6.16.

канты $a'b$, $b'c'd'$ и $ab'c$, составляющие ядро, и все члены канонического разложения, кроме $a'b'cd$, мы придем к таблице, представленной рис. 6.16,*б*. Импликант $a'c'd'$ не покрывает одночлена и потому является совершенно лишним.

Из рис. 6.16, *б* видно, что к ядру следует добавить либо $b'cd$, либо $a'cd$. Это приводит к двум неизбыточным выражениям:

$$a'b \vee b'c'd' \vee ab'c \vee b'cd, \quad a'b \vee b'c'd' \vee ab'c \vee a'cd.$$

Оба они минимальны по любому из двух признаков.

Нетрудно также извлечь из таблицы список одночленов, покрывающих данную функцию. Рассмотрим, например. рис. 6.17.

	a	b	c	d	e
A	×		×	×	
B		×	×		
C	×		×		×
D		×		×	

Рис. 6.17. Схема таблицы простых импликантов.

В этой таблице простые импликанты обозначены буквами A, B, C, D, а члены канонического выражения — буквами a, b, c, d, e. На языке булевой алгебры содержание таблицы состоит в том, что a покрывается $A \vee C$, b покрывается $B \vee D$, c покрывается $A \vee B \vee C$ и т. д. Рассмотрим произведение

$$(A \vee C)(B \vee D)(A \vee B \vee C)(A \vee D)(C).$$

Разложим его в сумму произведений, учтя правила поглощения $\alpha\beta \vee \alpha = \alpha$. Каждый одночлен этой суммы покрывает функцию, представленную таблицей. В нашем примере получится

$$ABC \vee CD.$$

УПРАЖНЕНИЯ Б

1. Найти минимальные суммы произведений для следующих выражений:
   а) $f(a, b, c) = \sum(0, 3, 5, 6)$,
   б) $f(a, b, c) = \sum(0, 3, 4, 6)$,
   в) $f(a, b, c, d) = \sum(0, 1, 2, 3, 4, 6, 7, 8, 9, 11, 15)$,
   г) $f(a, b, c, d, e) = \sum(0, 1, 3, 7, 8, 9, 10, 11, 14, 15, 16, 18, 22, 23, 24, 25, 26, 30)$.

Смысл записей справа: превратить каждое десятичное число в двоичное, расшифровать его по Мак-Клоски. Например, $0 = a'b'c'$, $3 = a'bc$, $5 = ab'c$. Знак $\sum$ означает булево суммирование всех выражений в скобках.

2. Построить вентильную схему с выходами Z_1, Z_2, пользуясь только элементами И, ИЛИ, для следующих булевых выражений:

$$Z_1 = ABC + A'BC + AB'C + A'B'C + ABC',$$
$$Z_2 = AB'C + AB'C' + A'B'C + A'B'C' + ABC'.$$

Считать, что A, B, C подаются с триггеров, так что для A, A', B, B', C, C' имеются отдельные каналы. По возможности использовать элементы для нескольких целей. Минимизировать число элементов, не обращая внимания на число входов.

3. а) Выбрать множество импликантов α_j, покрывающее все β_i, из следующей таблицы простых импликантов:

	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
α_1	×	×	×	×	×			
α_2	×					×		
α_3		×	×		×		×	×
α_4			×	×				×
α_5				×		×	×	

б) Объяснить и доказать следующее правило обработки таблиц простых импликантов: если столбец G_i содержит крестик в каждой строке, в которой другой столбец G_j содержит крестик, то столбец G_i можно вычеркнуть.

4. Построить двухуровневую схему И—ИЛИ с двумя выходами, реализующими следующие булевы выражения:

$$f_1(w, x, y, z) = wx'y'z' + wx'y'z + wx'yz' + wx'yz + wxy'z' + w'xy'z,$$
$$f_2(w, x, y, z) = wx'yz' + wx'yz + wxyz' + wxyz + wxy'z' + w'xy'z.$$

Минимизировать общее число элементов, используя их для параллельного обслуживания нескольких выражений.

5. а) Найти минимальную сумму произведений для выражения

$$f = b'd \vee abd \vee a'bc'd \vee a'bcd'.$$

б) Построить минимальную двухуровневую схему для него, пользуясь лишь элементами НЕ-ИЛИ.

6. а) Пользуясь методикой простых импликантов, найти минимальные суммы произведений и произведения сумм для функции

$$f(a, b, c, d) = \sum(1, 12, 13, 14, 15).$$

б) Построить двухуровневые схемы, реализующие эту функцию, следующих типов:

(1) И—ИЛИ,
(2) ИЛИ—И,
(3) НЕ-И—НЕ-ИЛИ,
(4) НЕ-ИЛИ—НЕ-ИЛИ.

7. Построить следующую *селекторную схему*. Пусть a_1, a_2, a_3 — *селекторные* переменные, они изменяются в множестве $a_i a_j = 0$ при $i \neq j$ и $a_1 + a_2 + a_3 = 1$. Переменные x_1, x_2, x_3 могут меняться как угодно. Требуется реализовать функцию

$$f(a_1, a_2, a_3, x_1, x_2, x_3) = x_i, \qquad \text{если } a_i = 1,$$

т. е. включение a_i на селекторе должно приводить к выдаче x_i. Спроектировать двухуровневые схемы типов И—ИЛИ и ИЛИ—И.

*8. Пусть $\mathcal{A}$ — алгоритм, применение которого к любой булевой функции F доставляет ее минимальное представление в виде суммы произведений: символически $\mathcal{A}(F) = f$. Пусть $\mathcal{A}'$ — следующий составной алгоритм: (1) вычислить F'; (2) вычислить $\mathcal{A}(F')$; (3) вычислить $[\mathcal{A}(F')]' = \mathcal{A}'(F) = g$. Доказать, что g есть минимальное представление F в виде произведения сумм.

*9. Оценить сверху число необходимых И-элементов, ИЛИ-элементов и инверторов, достаточное для представления любой булевой функции от 2, 3, r переменных.

*6.8. СОЮЗ

В предыдущем параграфе были обсуждены проблема минимизации булевых выражений и способы такой минимизации. Метод перечисления и отбора простых импликантов поддается дальнейшим усовершенствованиям, часть которых обсуждается в настоящем параграфе, а другая часть — в упражнениях.

Определение. Пусть α, β — два мультипликативных одночлена. Предположим, что имеется ровно одна переменная, такая, что она входит в α, а ее дополнение входит в β. Тогда *союз* α и β, $\sigma(\alpha, \beta)$, определяется как произведение двух одночленов, один из которых получается вычеркиванием этой переменной из α, а другой — вычеркиванием ее дополнения из β. Повторяющиеся литералы в произведении оставляются по одному разу.

Например, союзы пар abc и $ab'c'$, xyz и $x'y'$, xyz и xz неопределенны; союз abc и $bc'd$ равен abd, союз xyz и $xy'z$ есть xz; $\sigma(x'y'z, x'yw) = wx'z$ и т. д.

Теорема 8. *Если* α *и* β *входят в сумму произведений* Φ, *то* $\Phi \vee \sigma(\alpha, \beta) = \Phi$.

Доказательство. Заметим, что $\sigma(\alpha, \beta)$ влечет $\alpha \vee \beta$. Действительно, пусть x — такая переменная, что $\alpha = \pi x$, $\beta = \pi x'$. Тогда $\sigma(\alpha, \beta) = \pi\gamma$. Если $\pi\gamma = 1$ при данных значениях переменных, то

$$\alpha \vee \beta = (1 \cdot x) \vee (1 \cdot x') = 1,$$

что доказывает требуемое.

Теорема 9. *Пусть* Φ *— сумма произведений. Полную систему простых импликантов для* Φ *можно найти, последовательно совершая следующие действия:*

(i) *если одночлен* α *поглощает* β, *вычеркнуть* α,

(ii) *если для двух одночленов* λ, π *их союз определен и не поглощает ни один из имеющихся одночленов, добавить этот союз.*

Пример 7. Рассмотрим выражение $yz \vee y'z' \vee x'u \vee wyz \vee xwz' \vee uwxy = \Phi$. Союз wyz и xwz' равен wxy. Этот одночлен поглощается $uwxy$. Новое выражение имеет вид

$$yz \vee y'z' \vee x'u \vee wyz \vee xwz' \vee wxy.$$

Для него

$$\sigma(x'u, xwz') = uwz', \quad \sigma(x'u, wxy) = uwy.$$

Это приводит к выражению

$$yz \vee y'z' \vee x'u \vee wyz \vee uwz' \vee uwy \vee xwz' \vee wxy,$$

члены которого составляют полный список простых импликантов для Ф.

Доказательство. Нужно доказать три утверждения.

1) Если некоторый простой импликант не входит в выражение Ф, то в нем существуют два одночлена, союз которых определен и может быть добавлен к Ф.

2) Если некоторый одночлен α, входящий в Ф, не является простым импликантом, то

а) либо α поглощает некоторый простой импликант, уже содержащийся в Ф,

б) либо можно образовать и добавить некоторый союз в соответствии с условием теоремы.

3) Последовательность шагов типа (i) и (ii) обрывается.

Труднее всего доказать первое утверждение. Пусть Ф — сумма произведений, представляющая функцию f. Пусть γ — простой импликант, не входящий в Ф. Мы хотим убедиться, что из членов Ф можно образовать союз и добавить его к Ф. Так как γ, будучи простым импликантом, не поглощает членов выражения Ф, существует мультипликативный одночлен (возможно, сам γ), поглощающий γ, но не поглощающий ни один член из Ф. Добавляя литералы из Ф к γ, мы можем построить одночлен δ максимальной длины, поглощающий γ, но не поглощающий ни один член из Ф. Заметим, что δ не может поглощаться γ, потому что все пополнения простых импликантов f влекут f и потому поглощают некоторый член любой нормальной формы f.

Выберем переменную, не входящую в δ, скажем x. Одночлены δx и $\delta x'$ поглощают γ, а также некоторые члены выражения Ф, ибо длина δ была выбрана максимальной. Пусть δx поглощает α, $\delta x'$ поглощает β и α, β входят в Ф. Тогда α содержит x, β содержит x', а остальные литералы не могут входить одновременно в α и β, в один из членов со штрихом, в другой — без него. Кроме того, невозможно, чтобы $\alpha = x$, $\beta = x'$: иначе $\alpha \vee \beta = 1$, так что $f = 1$. Поэтому можно построить союз α и β, который поглощается δ, но ни одним из членов выражения Ф.

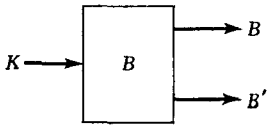
Доказательство утверждения (2) проводится непосредственно. Любой член α выражения Ф, не являющийся простым импликантом, должен поглощать некоторый простой импликант ξ. Если ξ входит в Ф, то α можно вычеркнуть из Ф. Если ξ не входит в Ф, то можно применить (1) и образовать некоторый союз.

Утверждение (3) устанавливается следующими рассуждениями. Из $2n$ литералов можно образовать лишь конечное число муль-

типликативных одночленов без повторяющихся литералов. Пусть член α входит в Ф. Его можно вычеркнуть из Ф, только если α поглощает некоторый член β; β можно вычеркнуть, только если β поглощает некоторый член γ и т. д. Все члены в любой такой последовательности поглощаются α, так что α не может появиться вновь. Кроме того, α не может добавиться вновь, ибо он поглощается самим α. Наконец, образование союзов может привести лишь к членам из конечного множества. Поэтому через конечное число шагов процесс закончится.

6.9. ТРИГГЕРЫ

В этом параграфе приступим к обсуждению технической реализации любого конечного автомата (см. гл. 3) в виде комбинации вентильных схем и подходящих элементов памяти с двумя устойчивыми состояниями. В следующем параграфе приводятся дальнейшие подробности.



(а)

$K(i)$	$B(i+1)$	$B'(i+1)$
1	1	0
0	0	1

(б)

Рис. 6.18. Триггер.

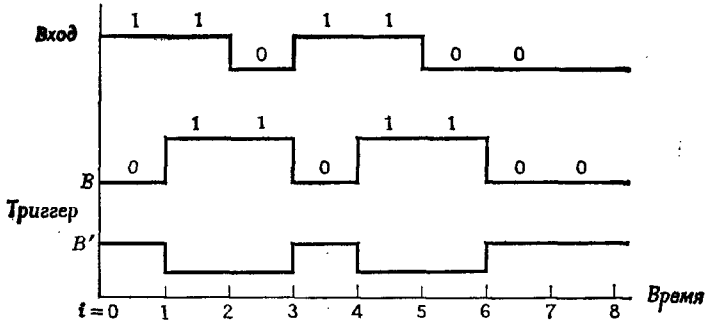


Рис. 6.19. Работа триггера.

На рис. 6.18, *а* дано стандартное представление электронной схемы, которая называется *триггером*. У простейшего триггера имеется один вход K и два выхода B, B'. Значение $K(i)$, которое подается на вход в момент i, запоминается в триггере K до момента $i+1$. В момент $i+1$ на вход подается $K(i+1)$, что и будет состоянием триггера до момента $i+2$, и т. д. «Моменты» задаются импульсами электронных часов.

Триггер имеет два выхода, B и B'. Выход, обозначенный B, выдает в момент $i+1$ сигнал $K(i)$, а выход, обозначенный B', выдает в момент $i+1$ сигнал $K(i)'$, дополнительный к $K(i)$.

На рис. 6.19 показаны возможные последовательности импульсов на входе-выходе.

Триггеры являются самыми употребительными элементами активной памяти в современных цифровых машинах. Описанный здесь тип входа является простейшим. Читатель, интересующийся другими вариантами, найдет их в литературе (см., например, книгу [1]).

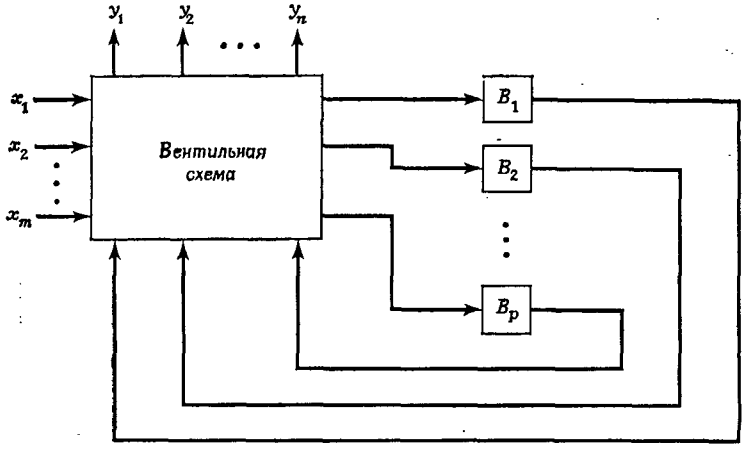


Рис. 6.20. Логическая схема последовательностной машины.

На рис. 6.20 показано, как триггеры используются в реализации последовательностных схем. Триггеры $B_1, \ldots, B_p$ соединены с вентильной схемой как входными, так и выходными каналами; сверх того, она имеет m дополнительных двоичных входов $x_1, \ldots, x_m$ и n выходов $y_1, \ldots, y_n$. В общей сложности вентильная схема имеет $m+p$ входов и $n+p$ выходов.

6. 10. ПРОЕКТИРОВАНИЕ ПОСЛЕДОВАТЕЛЬНОСТНЫХ МАШИН

Электронные схемы типа изображенной на рис. 6.20 могут реализовать функцию состояний любого конечного автомата. Точнее говоря, имеет место

Теорема 10. *Функция состояний* $\nu\colon A\times S\longrightarrow S$ *любого конечного автомата может быть реализована последовательностной схемой, синтезированной из инверторов, элементов И, ИЛИ и триггеров.*

Доказательство. Рассмотрим таблицу (или граф) для q состояний, описывающую функцию ν: $A \times S \to S$. Чтобы последовательностная машина имела q внутренних состояний, должно быть обеспечено неравенство $2^p \geqslant q$, где p — количество триггеров. Будем изображать состояние, отвечающее данным значениям B_i, двоичным числом $\sum_{i=0}^{p-1} B_i 2^i$. Удобно считать, что это двоичное число изображает индекс состояния. Например, автомат с пятью внутренними состояниями может отвечать такой таблице:

	B_0	B_1	B_2
s_0	0	0	0
s_1	1	0	0
s_2	0	1	0
s_3	1	1	0
s_4	0	0	1

Значения остальных входов и выходов X, Y можно закодировать аналогично, если сначала они были введены абстрактно. В большинстве случаев, однако, они естественно, кодируются двоичными последовательностями. Типичный пример представлен на рис. 6.21.

Текущее состояние	Следующее состояние		Выход	
	0	1	0	1
s_0	s_0	s_1	1	0
s_1	s_2	s_1	0	1
s_2	s_0	s_1	1	1

(*а*)

	B_0	B_1
s_0	0	0
s_1	1	0
s_2	0	1

(*б*)

Рис. 6.21. Таблица состояний, подлежащая реализации.

Пример 8. Рассмотрим таблицу состояний (рис. 6.21, *а*). Пусть состояния закодированы, как на рис. 6.21, *б*. Эта таблица состояний определяет следующую частичную функцию ν:

X	B_0	B_1		Y	B_0^*	B_1^*
0	0	0	$\mapsto$	1	0	0
0	1	0	$\mapsto$	0	0	1
0	0	1	$\mapsto$	1	0	0
1	0	0	$\mapsto$	0	1	0
1	1	0	$\mapsto$	1	1	0
1	0	1	$\mapsto$	1	1	0

Ее можно физически реализовать посредством трех вентильных схем, как это описано в предыдущих параграфах. Соответствующие булевы функции имеют вид:

$$Y = X'B_0'B_1' \vee X'B_0'B_1 \vee XB_0'B_1 \vee XB_0B_1',$$
$$B_0^* = XB_0'B_1' \vee XB_0B_1' \vee XB_0'B_1,$$
$$B_1^* = X'B_0B_1'.$$

Здесь B_0^* и B_1^* — это состояния, в которых триггеры должны находиться в следующий момент времени. Объединив наши три вентильные схемы воедино (как на рис. 6.20) и соединив очевидным образом их входы и выходы с входами и выходами двух триггеров, мы получим схему, реализующую таблицу состояний на рис. 6.21.

Пример 9. Спроектируем последовательностную машину для сложения двух неотрицательных чисел и двоичной записи C и D. Мы будем считать, что двоичная запись читается в направлении, противоположном обычному: например, $C(0)=0$, $C(1)=1$, $C(2)=1$, $C(3)=0$, $C(i)=0$ для $i>3$ изображает число 6. Выходная переменная Y должна принимать значение $C+D$. Машина должна иметь два внутренних состояния (чтобы в случае необходимости держать «в уме» единицу следующего разряда). Таблица состояний показана на рис. 6.22.

$C_iD_i=$	Следующие состояния CD				Выход CD			
	00	01	10	11	00	01	10	11
s_0	s_0	s_0	s_0	s_1	0	1	1	0
s_1	s_0	s_1	s_1	s_1	1	0	0	1

Рис. 6.22. Таблица состояний двоичного сумматора.

Функции ν: $C\times D\times S\to S$ и ζ: $C\times D\times S\to Y$ без труда считываются из таблицы, представленной на рис. 6.22. Они сведены в следующую таблицу.

C	D	S		Y	S
0	0	s_0	$\mapsto$	0	s_0
0	0	s_1	$\mapsto$	1	s_0
0	1	s_0	$\mapsto$	1	s_0
0	1	s_1	$\mapsto$	0	s_1
1	0	s_0	$\mapsto$	1	s_0
1	0	s_1	$\mapsto$	0	s_1
1	1	s_0	$\mapsto$	0	s_1
1	1	s_1	$\mapsto$	1	s_1

Чтобы представить эти функции в виде сумм произведений, мы закодируем $s_0=0$, $s_1=1$ состояниями B', B триггера B. Тогда

$$Y=C'D'B\vee C'DB'\vee CD'B'\vee CDB,$$
$$B=CDB'\vee CD'B\vee C'DB\vee CDB.$$

На рис. 6.23 показана блок-схема последовательностной машины, складывающей двоичные числа. Начальное состояние триггера B должно быть 0.

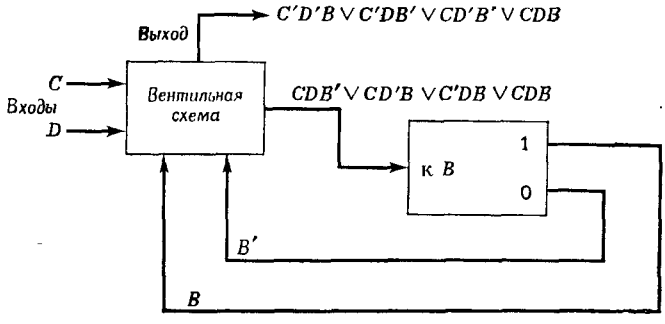


Рис. 6.23. Двоичный сумматор.

В предыдущем описании способ кодирования внутренних состояний наборами состояний триггеров был произвольным. Проблема минимизации, состоящая в таком выборе кодирования, чтобы требуемая вентильная схема была минимальной, не решена.

Описанный алгоритм позволяет для любой таблицы или графа состояний спроектировать электронную схему, реализующую эту таблицу. Таким образом, конечные автоматы, описанные в гл. 3,

физически реализуемы. К описанным нами схемам следует еще подключить считывающие и печатающие устройства.

Добавив выходные каналы, управляющие движением ленты, мы можем также реализовать любую машину Тьюринга, при условии, что лента может наращиваться с обоих концов и потому может считаться потенциально бесконечной.

УПРАЖНЕНИЯ В

1. а) Начертить диаграмму состояний автомата, который будет подавать на выход 0 до тех пор, пока на его единственный вход не будет подано подряд четыре единицы. Иными словами, пусть t — переменная, принимающая целые значения 1, 2, ..., n, ...; $X(t) = 0, 1$ — значение на входе, $Z(t)$ — значение на выходе. Тогда $Z(t)$ должно быть нулем до тех пор, пока не окажется $X(k) = X(k+1) = X(k+2) = X(k+3) = 1$ для некоторого k, и тогда $Z(k+3+j) = 0$ для всех $j \geqslant 0$.

Однако, если на вход до этого будет подано два нуля подряд, то выход должен все время оставаться нулевым.

б) Спроектировать схему, реализующую этот автомат, из элементов И, ИЛИ и триггеров.

2. а) Спроектировать последовательностную машину с двумя входами X_1, X_2 и одним выходом Z, такую, что 1 должна появиться на выходе только после подачи на входы (X_1, X_2) последовательности (0,1), (1,0), (1,1). В начальном состоянии $Z = 0$ и $(X_1, X_2) = (0, 0)$. Значение 1, однажды появившись на выходе, должно оставаться таким все время. Начертить диаграмму состояний машины, минимизировать число состояний, закодировать их наборами состояний триггеров, после чего спроектировать реализующую схему из элементов И, ИЛИ и триггеров.

3. Спроектировать последовательностную схему из элементов И, ИЛИ и триггеров, которая подает на выход 1, получив на входе последовательность, состоящую из нечетного числа единиц, зажатых между двумя нулями, т. е. схему, *распознающую* последовательности

$$S = \{01^{2n+1}0 \mid n \geqslant 0\}.$$

4. Найти минимальные выражения для следующих булевых функций:

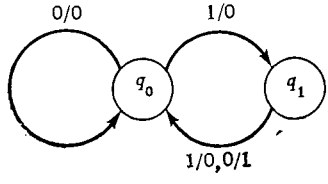
а) $f(a, b, c) = \sum (0, 3, 5)$,

б) $f(a, b, c, d) = \sum (0, 1, 3, 5, 6, 9, 10)$,

в) $f(a, b, c, d, e) = \sum (0, 2, 4, 6, 7, 8, 10, 12, 14, 30)$.

5. Рассмотрим бинарное отношение совместимости между мультипликативными одночленами от конечного числа переменных. Является ли оно рефлексивным? антирефлексивным? симметричным? антисимметричным? транзитивным?

6. Спроектировать последовательностную схему из элементов И, ИЛИ и триггеров, реализующую диаграмму состояний:



7. а) Спроектировать последовательностную схему из элементов НЕ-И и триггеров, дающую на выходе 1 только после того, как на вход будет подана последовательность из четного числа единиц, за которой следует единственный нуль, т. е. схему, *распознающую* последовательности

$$S = \{1^{2n}0 \mid n \in \mathbf{N}\}.$$

б) Объяснить, почему не может существовать конечного автомата, распознающего последовательности $S = \{1^n0^n \mid n \in \mathbf{N}\}$.

СПИСОК ЛИТЕРАТУРЫ

1. Barti T. C., Digital Computer Fundamentals, 2ed. McGraw-Hill, 1966.
2. Bellmann R. E., Dynamic Programming, Princeton University Press. [Русский перевод: Беллман Р., Динамическое программирование, М., ИЛ, 1960.]
3. Lee E. B., Markus L. A., Foundations of Optimal Control Theory, Wiley, 1967.

7

МОНОИДЫ И ГРУППЫ

7.1. БИНАРНЫЕ АЛГЕБРЫ

Эта глава в основном посвящена группам. Напомним, что булевы алгебры описывают поведение *множеств* относительно операций пересечения, объединения и дополнения. С теми же основаниями можно сказать, что группы описывают поведение *биекций* относительно операции композиции. Эти биекции, в частности, могут быть *симметриями* геометрических или алгебраических конфигураций.

Прежде чем вводить группы, рассмотрим свойства общих бинарных операций.

Определение. *Бинарной операцией* на множестве S называется произвольная функция β: $S^2 \longrightarrow S$.

Иными словами, бинарная операция есть правило, которое любым двум элементам x, $y \in S$ ставит в соответствие элемент $\beta(x, y) \in S$, результат применения β к x, y.

Часто удобно представлять себе операцию как умножение и вместо $\beta(x, y)$ писать $x \circ y$ или $x\beta y$ или просто xy. Следует, однако, помнить, что бинарная операция не обязана быть ни коммутативной, ни ассоциативной. Для конкретной операции может случиться, что $xy \neq yx$ и $x(yz) \neq (xy)z$, т. е.

$$\beta(x, y) \neq \beta(y, x), \quad \beta(x, \beta(y, z)) \neq \beta(\beta(x, y), z).$$

Определение. *Бинарной алгеброй* $[S, \beta]$ называется множество S с бинарной операцией β: $S^2 \longrightarrow S$.

Если β ассоциативна, бинарная алгебра $[S, \beta]$ называется *полугруппой*. Элемент $1_l \in S$, для которого $\beta(1_l, x) = x$ при всех $x \in S$, называется *левой единицей*; элемент 1_r, для которого $\beta(x, 1_r) = x$ при всех $x \in S$, называется *правой единицей*. Полугруппа, содержащая элемент 1 со свойством $x1 = 1x = x$ для всех $x \in S$, называется *моноидом* Повторим это определение полностью.

Определение. *Моноидом* $[S, \beta]$ называется множество с бинарной операцией $\beta(x, y) = xy$, которая удовлетворяет следующим условиям:

M1. $x(yz)=(xy)z$ для всех x, y, $z\in S$ (ассоциативность).

M2. Для некоторого $1\in S$ и всех $x\in S$ справедливы равенства $1x=x1=x$.

Заметим, что из M2 следует единственность единицы: если $1,1'$ — две единицы, то $1=1\cdot 1'=1'$, где первое равенство следует из того, что $1'$ — правая единица, а второе из того, что 1 — левая единица. Если единицы в полугруппе нет, мы можем ее «присоединить», добавив к S элемент 1 и распространив на него операцию β условиями $1x=x1$ для всех $x\in S$ и $1\cdot 1=1$. В результате мы получим моноид $[S\cup\{1\},\beta]$ вместо $[S,\beta]$.

Пример 1. По любой булевой алгебре $[A,\wedge,\vee,',O,I]$ можно построить два (двойственных) моноида $[A,\wedge]$ и $[A,\vee]$ с единицами I и O соответственно. Двойственность здесь понимается в смысле § 2.5 и 5.2.

Пример 2. Для любого множества X множество X^X всех функций $f\colon X\to X$ (унарных операций на X) является моноидом $[X^X,\circ]$ относительно левой композиции. Оно является также моноидом относительно правой композиции. Действительно, для всех f, g, $h\in X^X$

$$f\diamond(g\diamond h)=(h\circ g)\circ f=h\circ(g\circ f)=(f\diamond g)\diamond h, \tag{1}$$

$$1_X\diamond f=f\circ 1_X=f=1_X\circ f=f\diamond 1_X. \tag{2}$$

Пусть, например, $X=2=\{0,1\}$. Обозначим четыре функции $f\in 2^2$ так:

$$1\colon\ 0\mapsto 0,\quad 1\mapsto 1;\qquad r\colon\ 0\mapsto 1,\quad 1\mapsto 0;\qquad z_0\colon\ 0\mapsto 0,\quad 1\mapsto 0;\qquad z_1\colon\ 0\mapsto 1,\quad 1\mapsto 1:$$

Тогда моноиды $[2^2,\circ]$ и $[2^2,\diamond]$ задаются следующими таблицами умножения:

$\circ$	1	r	z_0	z_1
1	1	r	z_0	z_1
r	r	1	z_1	z_0
z_0	z_0	z_0	z_0	z_0
z_1	z_1	z_1	z_1	z_1

$\diamond$	1	r	z_0	z_1
1	1	r	z_0	z_1
r	r	1	z_0	z_1
z_0	z_0	z_1	z_0	z_1
z_1	z_1	z_0	z_0	z_1

Заметим, что моноиды из примера 1 *коммутативны* (ибо $x\wedge y=y\wedge x$ и $x\vee y=y\vee x$ для всех x, $y\in A$), тогда как моноиды из примера 2 не коммутативны, за исключением тривиального случая, когда X состоит из одного элемента. Теорема 2 гл. 2 доставляет еще один класс примеров.

Пример 3. Для любого множества X множество $\mathrm{Rel}(X)$ всех бинарных отношений на X является (некоммутативным)

моноидом относительно правой композиции. Единицей в нем является отношение равенства.

Естественно поставить вопрос, может ли моноид (или любая бинарная алгебра) иметь более одной единицы. Ответ таков:

Лемма 1. *Если у бинарной алгебры есть левая единица и правая единица, то они определены однозначно и совпадают с (единственной) двусторонней единицей.*

Доказательство. Пусть 1_l и 1_r—левая и правая единицы соответственно. Тогда $1_l 1_r = 1_r$, ибо 1_l—левая единица, и $1_l 1_r = 1_l$, ибо 1_r—правая единица. Отсюда следует требуемое.

Определение. В бинарной алгебре $[S, \beta]$ *левым нулем* называется элемент 0_l со свойством $0_l x = 0_l$ для всех $x \in S$, а *правым нулем*—элемент 0_r со свойством $x 0_r = 0_r$ для всех $x \in S$. *Нулем* называется элемент, являющийся одновременно левым и правым нулем.

В примере 1 элемент O является нулем для $[A, \wedge]$, а элемент I—нулем для $[A, \vee]$. В примере 2 функции z_0 и z_1 являются левыми нулями для левой композиции и правыми нулями для правой композиции. В примере 3 нулевое отношение с пустым графиком является (двусторонним) нулем.

Лемма 2. *Если у бинарной алгебры есть левый нуль 0_l и правый нуль 0_r, то оба они определены однозначно и совпадают с (единственным) двусторонним нулем.*

Доказательство. Справедливы равенства $0_l 0_r = 0_l$, ибо 0_l—левый нуль, и $0_l 0_r = 0_r$, ибо 0_r—правый нуль. Отсюда следует требуемый результат.

7.2. ЦИКЛИЧЕСКИЕ МОНОИДЫ; ПОДМОНОИДЫ

В любом моноиде M степени элемента $a \in M$ определяются для всех $n \in \mathbf{N}$ рекурсией:

$$a^0 = 1,\ a^1 = a,\ a^2 = aa,\ \ldots,\ a^{n+1} = a^n a. \tag{3}$$

Моноид называется *циклическим*, если он состоит из степеней c^n некоторого своего элемента c (в этом случае говорят, что M порожден элементом c).

Теорема 1. *В любом моноиде $a^m a^n = a^{m+n}$ для всех $m, n \in \mathbf{N}$.*

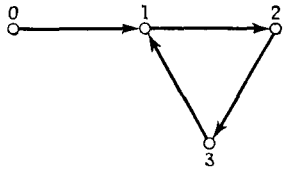
Доказательство проводится индукцией по n при любом фиксированном $m \in \mathbf{N}$. Обозначим через $P_m(n)$ утверждение $a^m a^n = a^{m+n}$. Утверждение $P_m(0)$ верно, ибо $a^m a^0 = a^m 1 = a^m$. Если $P_m(n)$ верно, то

$$a^m a^{n+1} = a^m (a^n a) = (a^m a^n) a = a^{m+n} a = a^{m+n+1},$$

где четыре выписанные равенства получаются из (3), ассоциативности, $P_m(n)$ и (3) соответственно.

Следствие. *Всякий циклический моноид коммутативен.*

Пример 4. Пусть f: $4 \to 4$—функция, заданная таблицей на рис. 7.1, *а*. Все степени $f^0 = 1$, $f^1 = f$, $f^2 = f \circ f$, $f^3 = f^2 \circ f$ различны, но $f^4 = f$. Поэтому циклический моноид, состоящий из всех степеней f, можно задать таблицей умножения 7.1, *б* (см. также § 2.6).



(*а*)

$\circ$	1	f	f^2	f^3
1	1	f	f^2	f^3
f	f	f^2	f^3	f
f^2	f^2	f^3	f	f^2
f^3	f^3	f	f^2	f^3

(*б*)

Рис. 7.1. Пример циклического моноида.

Пример 5. Неотрицательные целые числа образуют относительно сложения циклический моноид $\mathbf{N}$ с единицей 0, порожденный элементом 1.

Читатель, пораженный неожиданным переходом от мультипликативной к (традиционной в данном случае) аддитивной записи, легко сообразит, что $+(x, y) = x + y$ столь же правомерная бинарная операция на $\mathbf{N}$, как и $\cdot(x, y) = xy$. Относительно последней $[\mathbf{N}, \cdot]$ является нециклическим моноидом с единицей 1 и нулем 0. Очевидно, единицей в $[\mathbf{N}, +]$ служит 0, а нуля нет совсем (его роль играл бы присоединенный символ ∞ с обычными правилами действия).

Пусть теперь C—любой циклический моноид, порожденный элементом c. По теореме 1 умножение на c (с любой стороны) переводит c^r в c^{r+1}. Рассмотрим функцию f_c: $C \to C$ и действие f_c и ее степеней f_c^h: $c^r \mapsto c^{r+h}$ на

$$C = \{f_c^h(1)\} = \{1_C,\ f_c(1),\ f_c^2(1),\ \ldots,\ f_c^h(1),\ \ldots\}.$$

Если все c^r различны, то $C \cong [\mathbf{N}, +]$ по теореме 1. В противном случае существует наименьшее число $s \in \mathbf{P}$, такое, что $c^s = c^m$ для некоторого $m < s$ (т. е. $C = \{1_C, c, \ldots, c^{s-1}\}$). Число s является *порядком* (числом элементов) моноида C. Индукцией по j можно доказать формулу $c^i c^j = c^{\varphi(i, j)}$, где

$$\varphi(i, j) = i + j - kn \tag{4}$$

и $n = s - m$, а $k \in \mathbf{N}$—наименьшее целое число, такое, что $k > (i + j - s)/n$. (В частности, если $i + j < s$, то $k = 0$.) Таким образом, мы доказали следующий результат.

Лемма. *Всякий бесконечный циклический моноид изоморфен* $[\mathbf{N}, +]$. *Всякий конечный циклический моноид порядка* s *изоморфен для подходящих неотрицательных целых чисел* $m < s$ *и* $n = s - m$ *моноиду с законом умножения* (4). *Этот моноид мы обозначим символом* $C_{m,n}$, *заметив, что числа* m, n *определяют его с точностью до изоморфизма.*

Доказательство опирается на результаты § 2.7, где было показано, что моноид $C_{m,n}$ можно представить ориентированным графом $G_{m,n}$, который состоит из начального сегмента длины m, за которым следует цикл («петля») длины n, как показано на рис. 7.2.

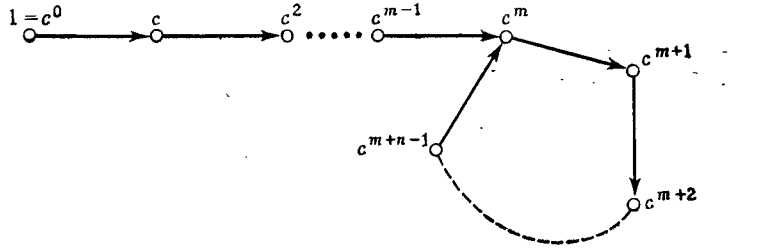


Рис. 7.2.

Умножение любого элемента $x \in C_{m,n}$ на c переводит вершину x графа на один шаг по ребру, исходящему из x, а умножение на c^l отвечает переходу на l таких шагов. Очевидно, переход на l шагов от вершины петли возвращает нас в исходную вершину тогда и только тогда, когда петля проходится конечное число раз. В терминах $C_{m,n}$ это означает, что при $l \neq 0$

$c^i c^l = c^i$ тогда и только тогда, когда $n \mid l$ и $m \leqslant i < m+n$.

В частности, выбрав в качестве l то единственное число среди $\{m, m+1, \ldots, m+n-1\}$, которое делится на n, и полагая в этом равенстве $i = l$, находим $c^l c^l = c^l$. Мы получили такой результат.

Теорема 2. *В каждом циклическом моноиде* $C_{m,n}$ *есть ровно один идемпотент, кроме единицы, при* $m \neq 0$ *и только единичный идемпотент при* $m = 0$.

Следствие 1. *В каждой конечной циклической полугруппе есть хотя бы один идемпотент (полугруппа* C *называется циклической, если* $C = \{c^r\}$, $r \in \mathbf{P}$).

Следующий результат более интересен.

Следствие 2. *Пусть* M *— конечный моноид. Тогда для всякого* $x \in M$ *и подходящего* $n \in \mathbf{P}$ *элемент* x^n *является идемпотентом.*

Доказательство. Очевидно, каждый элемент $x \in M$ порождает циклическую полугруппу $\{x^n\}$, $n = 1, 2, 3, \ldots$. Поэтому требуемый результат сразу же вытекает из следствия 1.

Универсальность. Теорема 1 показывает, что моноид $[\mathbf{N}, +]$ обладает следующим свойством *универсальности*. Для любого отображения его образующей $1 \mapsto a$ в любой другой моноид $[M, \cdot]$ функция θ: $n \mapsto a^n$ согласована с бинарными операциями в M, $\mathbf{N}$, ибо

$$(m+n)\theta = a^{m+n} = a^m \cdot a^n$$

в силу теоремы 1. Такие отображения называются *морфизмами* (см. § 2.6). Другое описание свойства θ «быть морфизмом» состоит в том, что диаграмма

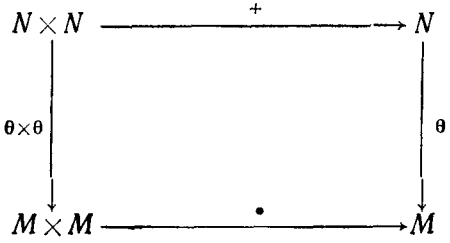


Рис. 7.3.

коммутативна в обычном смысле слова.

Читатель легко проверит, что $[\mathbf{P}, +]$ обладает тем же универсальным свойством в классе полугрупп.

Это свойство позволяет назвать $[\mathbf{N}, +]$ свободным моноидом с одной образующей, а $[\mathbf{P}, +]$—свободной полугруппой с одной образующей.

Легко доказать, что свободные моноид и полугруппа определяются свойством универсальности однозначно с точностью до изоморфизма. Действительно, пусть A, B—два таких универсальных объекта с образующими a, b соответственно. Тогда отображения $a \mapsto b$ и $b \mapsto a$ продолжаются до морфизмов α: $A \to B$ и β: $B \to A$, композиции которых суть $\alpha \diamond \beta = 1_A$ и $\beta \diamond \alpha = 1_B$ соответственно. Поэтому α, β—взаимно обратные изоморфизмы, как и утверждалось.

Это рассуждение применимо к алгебраическим системам с любым числом образующих.

7.3. ГРУППЫ

Элемент a моноида M называется *обратимым слева*, если $xa = 1$ для подходящего $x \in M$, и *обратимым справа*, если $ay = 1$ для подходящего $y \in M$. Элемент, обратимый слева и справа,

называется просто *обратимым*. Покажем, что свойства обратимости, доказанные для моноида функций X^X в § 1.5, верны в общем случае.

Лемма 1. *Любой обратимый элемент a в моноиде M обладает двусторонним обратным элементом a^{-1}, таким, что $aa^{-1} = a^{-1}a = 1$. Любой левый или правый обратный элемент к a совпадает с a^{-1}.*

Доказательство. Пусть b_l и b_r—левый и правый обратные элементы к a соответственно. Тогда

$$b_l = b_l 1 = b_l(ab_r) = (b_l a) b_r = 1 b_r = b_r.$$

Поэтому, полагая $a^{-1} = b_l = b_r$, получаем $aa^{-1} = a^{-1}a = 1$. Для любого левого обратного x к a имеем

$$x = x(ab_r) = (xa) b_r = 1 b_r = b_r = a^{-1}.$$

Случай правого обратного рассматривается аналогично.

Моноид, в котором всякий элемент обратим, называется *группой*. Конечные группы составляют наиболее изученный класс бинарных алгебр. Снова приведем полное определение группы.

Определение. Группой, т. е. моноидом, в котором все элементы обратимы, называется система $[G, \cdot, 1, {}^{-1}]$ со следующими свойствами:

G 1.	$x(yz) = (xy)z$	для всех $x, y, z \in G$	(ассоциативность);
G 2.	$1x = x1 = x$	для всех $x \in G$	(единица);
G 3.	$xx^{-1} = x^{-1}x = 1$	для всех $x \in G$	(обратные).

В классе конечных моноидов группы выделяются следующим простым свойством.

Теорема 3. *Конечный моноид M является группой тогда и только тогда, когда 1 является его единственным идемпотентом.*

Доказательство. Если 1—единственный идемпотент моноида M, то, согласно следствию 2 теоремы 2, любой элемент $x \in M$ имеет обратный, являющийся степенью x: $x^{n-1}x = xx^{n-1} = 1$. Значит, M—группа. Обратно, если M—группа с единицей 1 и e—ее идемпотент, то для правого обратного элемента e^{-1} к e имеем

$$e = e1 = e(ee^{-1}) = (ee)e^{-1} = ee^{-1} = 1.$$

Значит, 1—единственный идемпотент в M.

Назовем *подмоноидом* моноида $M = [M, \beta]$ любое подмножество S, для которого (i) $1 \in S$ и (ii) из $x \in S$ и $y \in S$ следует, что $xy = \beta(x, y) \in S$. Очевидно, всякий элемент a любого моноида M порождает в M циклический подмоноид, состоящий из всех степеней a.

Лемма 2. *В любом моноиде M обратимые слева и обратимые справа элементы образуют подмоноиды L и R моноида M.*

Доказательство. Очевидно, $1 \in L$ и $1 \in R$, ибо $1 \cdot 1 = 1$. Далее, если $xa = 1$ и $yb = 1$, то

$$(yx)(ab) = y(xa)b = y\,1\,b = yb = 1.$$

Значит, L — подмоноид. Аналогичное рассуждение применимо к R.

Обратимые элементы M составляют подмоноид $L \cap R$. Из доказательства леммы 2 видно, что

$$(ab)^{-1} = b^{-1}a^{-1} \text{ для любых обратимых элементов } a,\ b \in M. \quad (5)$$

Это тождество относится также к левым обратным и правым обратным элементам, однако, учитывая их возможную неоднозначность, его следует понимать в том смысле, что $b^{-1}a^{-1}$ есть один из обратимых к ab.

Пример 6. Пусть $[\mathbf{Z}_n, \cdot]$ — мультипликативный моноид классов вычетов по модулю n. Его подмоноид (группа) обратимых элементов состоит из (классов) всех положительных целых чисел $k < n$, взаимно простых с n (обычная запись $(k, n) = 1$).

Идемпотенты e в полугруппе S, т. е. элементы e, удовлетворяющие условию $ee = e$, играют важную роль. Если e — идемпотент, то $e^n = e$ для всех n (индукция). Иными словами, циклический подмоноид S, порожденный e, состоит только из 1 и e. Следующий результат также устанавливается без труда.

Лемма 3. *В любом коммутативном моноиде M идемпотенты образуют подмоноид.*

Доказательство. Очевидно, $1^2 = 1$ в любом моноиде. Далее, в коммутативном моноиде из $e^2 = e$ и $f^2 = f$ следует, что

$$(ef)^2 = (ef)(ef) = e(fe)f = e(ef)f = (ee)(ff) = ef.$$

Более общо, это рассуждение показывает, что если два идемпотента e, f в моноиде или полугруппе перестановочны (т. е. $ef = fe$), то их произведение также является идемпотентом.

Определение. В бинарной алгебре $[S, \cdot]$ элемент $a \in S$ называется

(1) сократимым справа, если из $xa = ya$ следует $x = y$;

(1′) сократимым слева, если из $ax = ay$ следует $x = y$.

Лемма 4. *В любом моноиде M всякий обратимый справа элемент сократим справа, а обратимый слева элемент сократим слева.*

Лемма 5. *В любом конечном моноиде M*

(2) *всякий обратимый справа элемент сократим слева;*

(2′) *всякий обратимый слева элемент сократим справа.*

Следствие. *В любом конечном моноиде все правые обратные любого элемента совпадают со всеми его левыми обратными.*

Доказательства мы предоставляем читателю.

Опишем для иллюстрации программу на АЛГОЛе, использующую результат теоремы 3. Массив T размера 50×50 содержит таблицу умножения моноида $[M, \mu]$, где $M=\{1, 2, \ldots, N\}$ $(N\leqslant 50)$, μ: $M\times M\longrightarrow M$: $(k, j)\longrightarrow T[k, j]$. Единица $1\in M$ такова, что $T[1, j]=T[j, 1]=j$ для всех $j\in M$. Программа присваивает булевой переменной GROUP значение *true*, если моноид $[M, \mu]$, описанный массивом T, является группой, и значение *false* в противном случае.

```
begin integer array T[1:50, 1:50];
      integer i, N; Boolean GROUP;
for i = 2 step 1 until N do
      if T[i, i] = i then begin GROUP: = false; go to F end;
GROUP: = true;
F: end
```

УПРАЖНЕНИЯ А

1. Показать, что присоединение единицы превращает любую полугруппу в моноид.

2. Показать, что если z — левый нуль полугруппы S, то все его левые кратные xz также являются левыми нулями.

3. Показать, что если множество T состоит более чем из одного элемента, то моноид всех отображений T в себя содержит более одного левого нуля, но ни одного правого нуля.

4. Показать, что в моноиде из примера 3 умножение некоммутативно: $fg\neq gf$ для надлежащим образом выбранных f, g.

5. Показать, что в моноиде, содержащем более одного элемента, левый нуль не может иметь левого обратного элемента.

6. Показать, что в любом моноиде отображение g_a: $x\mapsto xa$ (a фиксирован, x — переменный элемент) является биективным тогда и только тогда, когда a имеет правый обратный элемент).

7. Пусть S — любой конечный моноид.

а) Показать, что элемент $a\in S$ имеет левый обратный в том и только том случае, если $x\mapsto ax$ биективно.

б) Показать, что элемент $a\in S$ имеет левый обратный в том и только том случае, если он имеет правый обратный.

8. Привести пример полугруппы с левой единицей и правым нулем, не являющейся моноидом.

*9. Дать определение свободного коммутативного моноида с двумя образующими (см. § 2.7) и показать, что он изоморфен $[\mathbf{N}, +]\times[\mathbf{N}, +]$.

*10. а) Дать определение свободной бинарной алгебры с одной образующей F_1^2. (*Указание*: см. рис. 7.4.)

б) Доказать, что F_1^2 изоморфна множеству всех строк вида α, $(\alpha\alpha)$, $((\alpha\alpha)\alpha)$, $(\alpha(\alpha\alpha))$, ..., «правильно составленных» из символа α и скобок),(с операцией «заключение в скобки и соединение»: $\beta_1\circ\beta_2=(\beta_1)(\beta_2)$.

7.4. МОРФИЗМЫ; ПРЯМЫЕ ПРОИЗВЕДЕНИЯ

В этом параграфе мы изучим морфизмы полугрупп и моноидов и обсудим его связь с прямыми произведениями.

Определение. *Морфизмом полугрупп* называется отображение θ: $S \to T$ одной полугруппы в другую, которое переводит операцию в S в операцию в T. В мультипликативной записи это означает

$$\theta(ss') = \theta(s)\theta(s') \text{ для всех } s,\ s' \in S. \tag{6}$$

Если S, T — моноиды с единицами 1_S, 1_T соответственно, то отображение θ: $S \to T$ называется *морфизмом моноидов*, если выполняется условие (6) и, кроме того,

$$\theta(1_S) = 1_T. \tag{7}$$

Как и в § 2.6 морфизмы, являющиеся отображениями на, называются *эпиморфизмами*, инъективные морфизмы — *мономорфизмами*, а биективные морфизмы — *изоморфизмами*. Покажем, что для любых моноидов S, T понятия *изоморфизма полугрупп* и *изоморфизма моноидов* равносильны.

Лемма 1. *Для любых двух моноидов S, T биекция β: $S \leftrightarrow T$ является морфизмом моноидов тогда и только тогда, когда она является морфизмом полугрупп.*

Доказательство. Пусть $\beta^{-1}(1_T) = s \in S$. Тогда, по определению,

$$\beta(ss') = \beta(s)\beta(s') = 1_T\beta(s') = \beta(s') \in T \text{ для всех } s' \in S.$$

Поскольку β — биекция, отсюда следует, что

$$ss' = \beta^{-1}(\beta(ss')) = \beta^{-1}(\beta(s')) = s' \text{ для всех } s' \in S.$$

Аналогично, $s's = s'$ для всех $s' \in S$. Поэтому $s = 1_S$ по лемме 1 § 7.1. Итак, $\beta^{-1}(1_T) = 1_S$; следовательно, $\beta(1_S) = 1_T$, так что β — морфизм моноидов.

Лемма 2. *Множество всех морфизмов моноида $M = [S, \cdot]$ в себя является подмоноидом в S^S.*

Доказательство. Из (6) следует, что

$$\theta'(\theta(ss')) = \theta'(\theta(s)\theta(s')) = \theta'(\theta(s))\,\theta'(\theta(s')).$$

Лемма 3. *Отображение θ^{-1}, обратное к любому изоморфизму θ: $S \to T$ полугрупп (моноидов), является изоморфизмом полугрупп (моноидов).*

Доказательство. Для любых t, $t' \in T$ существуют такие s, $s' \in S$, что $t = \theta(s)$ и $t' = \theta(s')$. Из (6) следует, что

$$\theta^{-1}(tt') = \theta^{-1}(\theta(s)\theta(s')) = \theta^{-1}(\theta(ss')) = ss' = \theta^{-1}(t)\,\theta^{-1}(t').$$

Введем еще два важных класса морфизмов. Морфизм полугруппы (или любой алгебраической системы) в себя называется *эндоморфизмом*. Любой биективный эндоморфизм, т. е. изоморфизм алгебраической системы с собой, называется *автоморфизмом*.

Лемма 4. *Автоморфизмы любой полугруппы S образуют группу, называемую группой автоморфизмов полугруппы S.*

Доказательство. Тождественное отображение S, очевидно, является единицей для автоморфизмов. По лемме 1 композиция двух автоморфизмов является автоморфизмом. По лемме 2 обратное отображение к любому автоморфизму является автоморфизмом.

Лемма 5. *Пусть $\theta\colon M \to N$ — любой морфизм моноидов. Тогда образ относительно θ любого подмоноида $S \subset M$ является подмоноидом $\theta(S) \subset N$ и обратный образ $\theta^*(T)$ любого подмоноида $T \subset N$ является подмоноидом в M.*

Доказательство непосредственно следует из определений, и мы предоставляем его читателю. Аналогичные результаты справедливы для морфизмов полугрупп. Заметим, что пересечение двух непустых подполугрупп данной полугруппы может быть пустым, тогда как все моноиды содержат единицу.

Следующий результат несколько сильнее.

Теорема 4. *Пусть $M = [S, \cdot, 1]$ — любой моноид и $\theta\colon S \to T$ — сюръективное отображение M на бинарную систему $[T, \beta]$, удовлетворяющее соотношениям*

$$\theta(s, s') = \beta(\theta(s), \theta(s')) \text{ в } T. \tag{8}$$

Тогда $[T, \beta]$ — полугруппа, а $[T, \beta, \theta(1)]$ — моноид. *Кроме того, если M — группа, то $[T, \beta]$ — также группа.*

Доказательство. Поскольку θ сюръективно, для любых элементов $t_i \in T$ $(i = 1, 2, 3)$ существуют элементы $s_i \in S$ со свойством $\theta(s_i) = t_i$.

Применяя (8) дважды, находим

$$\beta(t_1, \beta(t_2, t_3)) = \theta(s_1(s_2 s_3)) = \theta((s_1 s_2) s_3) = \beta(\beta(t_1, t_2) t_3).$$

Следовательно, β ассоциативна, т. е. $[T, \beta]$ — полугруппа. Обозначая $\theta(1)$ через $e \in T$, получаем

$$\beta(e, t_i) = \theta(1 s_i) = \theta(s_i) = t_i \text{ для всех } t_i.$$

Следовательно, e является левой единицей. Аналогично устанавливается, что e — правая единица. Наконец, если M — группа, то существование обратных элементов в T доказывается переносом их из M.

Установим теперь «теорему о представлении», показывающую, что все общие свойства композиции функций переносятся на моноиды.

Теорема 5. *Любой моноид M изоморфен подмоноиду M^M всех функций $f\colon M \to M$ относительно левой или правой композиции (теорема Кэли).*

Доказательство. Рассмотрим отображение $\mu\colon s \mapsto f_s$, где f_s — функция $x \mapsto xs$ из M в M. По ассоциативности $x(st) = (xs)t$ для всех $x \in M$. Следовательно, $f_{st} = f_s \diamond f_t$ и μ есть морфизм (переводящий умножение в правую композицию). Так как из $f_s = f_t$ следует, что $s = 1s = 1t = t$, где 1 — единица в M, это мономорфизм. Аналогично, отображение $\nu\colon s \mapsto g_s$, где $g_s(x) = sx$ для всех $x \in M$, является мономорфизмом моноидов, переводящим умножение в левую композицию.

Пример 7. Пусть M — моноид, состоящий из единицы 1 и r левых нулей $z_1, \ldots, z_r$, так что $z_i z_j = z_i$ для всех $i, j \in \{1, 2, \ldots, r\}$. Тогда мономорфизм μ теоремы 5 отображает 1 в тождественное отображение, а z_k в функцию $\varphi_k \in M^M$, принимающую значение z_k на 1 и z_j на z_j.

Еще один интересный пример получится, если применить эту конструкцию к случаю $X = 2$ в примере 2. Четыре функции 1, r, z_0, z_1 задаются следующей таблицей:

	1	r	z_0	z_1
0	0	1	0	1
1	1	0	0	1

Пусть умножение в M означает правую композицию. Тогда теорема 5 позволяет представить элементы M в виде функций на множестве $1, r, z_0, z_1$, а не на множестве $2 = \{0, 1\}$.

Определение. *Прямым произведением* $S \times T$ двух полугрупп S, T называется множество всех пар (s, t) с покомпонентным умножением:

$$(s, t)(s', t') = (ss', tt') \text{ для всех } s, s' \in S,\ t, t' \in T. \tag{9}$$

Если S, T — моноиды, то $S \times T$ также моноид с единицей $(1_S, 1_T)$. В этом случае $S \times T$ содержит *подмоноиды* $S_1 \cong S$ и $T_1 \cong T$, состоящие из пар $(s, 1_T)$, $(1_S, t)$ соответственно. В любом случае отображения $(s, t) \mapsto s$ и $(s, t) \mapsto t$ являются эпиморфизмами $S \times T \to S$ и $S \times T \to T$.

Для моноидов все эти морфизмы показаны на диаграмме рис. 7.4, *а*. Для любых полугрупп имеется очевидный изоморфизм.

$$S \times (T \times U) \cong (S \times T) \times U\colon\ (s, (t, u)) \mapsto ((s, t), u).$$

На диаграмме рис. 7.4, *б* отражено еще одно важное свойство прямых произведений. Пусть $\sigma\colon A \to S$, $\tau\colon A \to T$ — морфизмы полугрупп. Тогда отображение $a \to (\sigma(a), \tau(a))$ является морфизмом $\alpha\colon A \to S \times T$, и притом единственным, делающим диаграмму 7.4, *б* коммутативной.

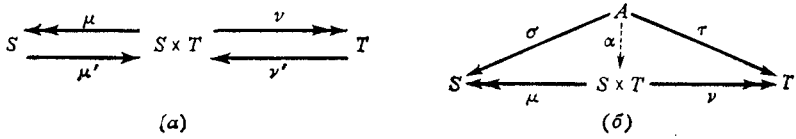


Рис. 7.4.

7.5. ПРИМЕРЫ ГРУПП; АКСИОМЫ

Мы приведем несколько примеров групп и начнем их систематическое изучение, напомнив их *аксиоматическое определение* и выведя из него ряд общих элементарных свойств.

Многие группы знакомы нам из элементарной алгебры; в основном они *коммутативны*, или *абелевы*: $xy = yx$ для всех x, y.

Пример 8. Ненулевые комплексные числа $z = x + y\sqrt{-1}$ образуют коммутативную группу $[\mathbf{C}^*, \cdot]$ относительно операции умножения.

Аналогично, ненулевые вещественные числа и ненулевые рациональные числа образуют группу относительно умножения, однако для ненулевых целых чисел это неверно.

Многие другие коммутативные группы в алгебре принято записывать аддитивно и использовать 0 для обозначения единичного элемента.

Пример 9. Целые числа образуют коммутативную группу $[\mathbf{Z}, +]$ относительно сложения.

Пример 10. Комплексные числа образуют коммутативную группу $[\mathbf{C}, +]$ относительно сложения.

Еще один важный класс групп возникает из симметрий алгебраических и геометрических объектов. Среди них одной из основных является группа перестановок, т. е. симметрий множества без дополнительной структуры.

Пример 11. Для любого множества X множество всех биекций $\beta\colon X \to X$ является группой. Если X состоит из n элементов, эта группа называется *симметрической группой* степени n и часто обозначается символом S_n. Она состоит из $n!$ элементов.

Пример 12. *Диэдральная группа* Δ_n состоит из симметрий правильного n-угольника Π_n.

Чтобы изучить ее строение, представим себе чертеж n-угольника Π_n с горизонтальной нижней стороной и поместим в его центр начало координат. На рис. 7.5 изображены случаи $n=3$ и $n=4$.

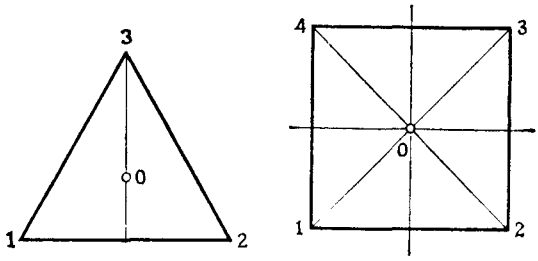


Рис. 7.5.

Будем отсчитывать углы вращения относительно начала координат против часовой стрелки. Очевидно, вращение R на угол $(360/n)^\circ$ является симметрией Π_n вместе со всеми своими степенями R^k. Кроме того, отражения V относительно вертикальной оси симметрии и $n-1$ других осей симметрии, проходящих через центр, являются симметриями. Эти оси образуют с вертикалью углы $(180/n)^\circ$. Кроме этих $2n$ симметрий, других нет.

Действия R^k и V на любой вектор, образующий угол α с вертикалью, очевидно, таковы:

$$R^k(\alpha)=\alpha+(k/n)\,360^\circ, \qquad V(\alpha)=-\alpha.$$

Из этих формул ясно, что

$$R^n=V^2=1,$$

$$R^kV=VR^{-k} \text{ переводит } \alpha \text{ в } -\alpha+\frac{k}{n}\,360^\circ.$$

Полную таблицу умножения для диэдральной **группы Δ_n** легко построить, исходя из следующих соотношений:

$$R^n=1, \quad V^2=1, \quad RV=VR^{-1}, \tag{10}$$

где 1 — тождественное преобразование. Каждый элемент группы можно записать либо в виде R^k, либо в виде VR^k для подходящих $k=0, 1, \ldots, n-1$. Правила умножения таковы:

$$R^hR^k=R^{h+k} \text{ (или } R^{h+k-n} \text{ для } h+k\geqslant n), \quad (VR^h)\,R^k=VR^{h+k}, \tag{10'}$$

$$R^h(VR^k)=(R^hV)\,R^k=(VR^{-h})\,R^k=\begin{cases} VR^{k-h} & \text{при } k\geqslant h, \\ VR^{n+k-h} & \text{при } k<h. \end{cases} \tag{10''}$$

Для удобства дальнейших ссылок еще раз напомним общее определение группы из § 7.3.

Определение. *Группой* (в мультипликативной записи) $G=[G, \cdot]$ называется множество G с бинарной ассоциативной операцией умножения, содержащее *единицу* 1, т. е. элемент со свойством

$$x1=1x=x \text{ для всех } x\in G, \tag{11}$$

и для каждого x содержащее *обратный* элемент x^{-1} со свойством

$$xx^{-1}=x^{-1}x=1. \tag{12}$$

Перечислим теперь некоторые общие свойства групп. Напомним, что, как показано в § 7.1, всякий моноид и тем более всякая группа содержит единственную единицу (11), так что единственность постулировать не нужно. Далее, согласно изложенному в § 7.3, обратный элемент x^{-1} (12) к любому $x\in G$ также определен однозначно, так как для любого другого обратного мы имеем

$$y=y1=y(xx^{-1})=(yx)x^{-1}=1x^{-1}=x^{-1}.$$

Есть и другие свойства, общие для всех групп, но не обязательно для моноидов. Среди них основными являются следующие правила сокращения.

Правило 1. *Все элементы любой группы G удовлетворяют следующим правилам левого и правого сокращения:*

$$\left.\begin{array}{l} \text{из } ax=ay \text{ следует } x=y, \\ \text{из } xa=ya \text{ следует } x=y. \end{array}\right\} \tag{13}$$

Докажем правило левого сокращения. Если $ax=ay$, то

$$x=1x=(a^{-1}a)x=a^{-1}(ax)=a^{-1}(ay)=(a^{-1}a)y=1y=y.$$

Читателю рекомендуется уточнить для себя, на чем основан каждый шаг в предыдущей цепочке равенств.

Заметим, что единственность левого и правого обратного следует из правил сокращения, применимых к равенствам $ax=ay=1$ и $xa=ya=1$.

Правило 2. *Для любых данных элементов a, b группы G уравнения $xa=b$ и $ay=b$ имеют единственное решение $x=ba^{-1}$ и $y=a^{-1}b$ соответственно.*

Доказательство. Очевидно, $x=ba^{-1}$ есть решение уравнения $xa=b$, ибо

$$(ba^{-1})a=b(a^{-1}a)=b1=b.$$

Обратно, если $xa=b$, то

$$x=xe=x(aa^{-1})=(xa)a^{-1}=ba^{-1}.$$

Таким образом, это решение единственно. Аналогично разбирается уравнение $ay=b$.

Из правил 1 и 2 следует, что каждый *левый сдвиг* φ_a: $c \mapsto ax$ группы G и каждый ее *правый сдвиг* ψ_a: $x \mapsto xa$ является биекцией. Слово «сдвиг» используется здесь по аналогии со следующим частным случаем. Если G — аддитивная группа всех упорядоченных пар (x, y) вещественных чисел по сложению, то отображение

$$\tau_{a,b}\colon (x, y) \longrightarrow (x+a, y+b) = (x, y) + (a, b)$$

является сдвигом (или переносом) (x, y)-плоскости (комплексной плоскости) из примера 4 в обычном геометрическом смысле.

Правило 3. *Пусть* 1 *— единица группы,* a, b *— ее любые элементы. Тогда*

$$1^{-1} = 1, \quad (ab)^{-1} = b^{-1}a^{-1}. \tag{14}$$

Доказательство. Так как 1 — единица, то $1 \cdot 1 = 1$. Это означает, что 1 является (единственным) левым и правым обратным элементом к 1. Далее,

$$(ab)(b^{-1}a^{-1}) = a(bb^{-1})a^{-1} = a1a^{-1} = aa^{-1} = 1.$$

Тем самым доказано, что элемент $b^{-1}a^{-1}$ является правым обратным к ab. Аналогично устанавливается, что это левый обратный.

Другие системы аксиом. Несколько уточняя приведенные выше рассуждения, можно показать, что группы характеризуются менее избыточными системами аксиом. Вот два примера.

1. Всякая полугруппа $[S, \cdot]$, у которой есть *левая* единица e ($ex = x$ для всех x) и *левый* обратный элемент x^{-1} ($x^{-1}x = e$) для каждого элемента, является группой.

2. Пусть S — полугруппа, в которой для любых $a, b \in S$ существуют такие x, $y \in S$, что $xa = b$ и $ay = b$. Тогда S является группой.

7.6. ПОДГРУППЫ

Подгруппой группы G называется любое подмножество $S \subset G$, являющееся группой относительно умножения в G. В частности, если x, $y \in S$, то $xy \in S$. Далее, единица 1 должна лежать в S, ибо она является единственным идемпотентом в G. Наконец, так как x^{-1} есть единственное решение в G уравнения $xy = 1$, а 1 — единственная единица в G, S должно содержать вместе с каждым элементом x элемент x^{-1}.

Обратно, пусть $S \subset G$ удовлетворяет этим трем условиям. Тогда умножение в S автоматически ассоциативно, 1 является единицей и все $x \in S$ обратимы в S. Таким образом, мы доказали следующую теорему:

Теорема 6. *Подмножество* S *группы* G *является подгруппой* G *тогда и только тогда, когда оно удовлетворяет следующим*

условиям: (i) $1 \in S$; (ii) *из* $x \in S$ *следует, что* $x^{-1} \in S$; (iii) *из* $x \in S$ *и* $y \in S$ *следует, что* $xy \in S$.

Определение степени элемента (§ 7.2) в группах можно распространить на все целые показатели. Положим

$$a^m = aaa\ldots a \quad (m \text{ множителей}), \qquad a^0 = 1, \qquad a^{-m} = (a^{-1})^m. \tag{15}$$

Теорема 1 допускает следующее усиление:

Теорема 7. *Пусть* G*—группа; тогда*

$$a^r a^s = a^{r+s}, \quad (a^r)^s = a^{rs} \quad \textit{для всех } a \in G. \tag{16}$$

Доказательство. Если r, $s \in \mathbf{N}$, то первое равенство (16) следует из теоремы 1. Если $r = -m$ и $s = -n$ оба отрицательные, то в силу (14)

$$a^r a^s = (a^{-1})^m (a^{-1})^n = (a^{-1})^{m+n} = a^{-(m+n)} = a^{r+s}.$$

Остается рассмотреть случай, когда показатели степени имеют разные знаки, скажем $r = m$ и $s = -n$, где $m > 0$ и $n > 0$. В этом случае

$$a^m a^{-n} = a^m (a^{-1})^n = \overbrace{(a\ldots a)}^{m}\,\overbrace{(a^{-1}\ldots a^{-1})}^{n}.$$

Пользуясь ассоциативностью, мы можем последовательно сокращать стоящие рядом a и a^{-1}. При $m \geqslant n$ в конце концов остается a^{m-n}, а при $m < n$ остается $n-m$ множителей a^{-1}, т. е. $(a^{-1})^{n-m} = a^{-(n-m)}$. В обоих случаях получаем $a^m a^{-n} = a^{m+(-n)}$.

Второе равенство (16) еще проще. При $s > 0$ из первой части (16) находим

$$a^r a^r \ldots a^r \ (s \text{ множителей}) = a^{r+\cdots+r} = a^{rs}.$$

При $s < 0$ можно написать аналогичное произведение из $-s$ множителей a^{-r}. При $s = 0$ результат очевиден.

Следствие. *В любой группе* G *степени каждого элемента образуют подгруппу.*

Определение. Группа C, состоящая из всех степеней a^r элемента a, называется *циклической* группой, *порожденной* элементом a. Порядок циклической подгруппы $\{a^r \mid r \in \mathbf{Z}\}$, порожденной элементом a, называется также *порядком* этого элемента.

Заметим, что аддитивная циклическая группа $[\mathbf{Z}, +]$ не является циклическим моноидом. Эта циклическая группа бесконечна. Остальные циклические группы, изоморфные группам $\mathbf{Z}_m$ классов вычетов $\bmod m$ по сложению, являются циклическими, если их рассматривать как моноиды (см. гл. 2).

Определение. *Морфизмом групп* называется любое отображение $\mu\colon G \to H$ одной группы в другую, которое удовлетворяет следующим условиям: (i) $\mu(1_G) = 1_H$; (ii) для всех $g \in G$ $\mu(g^{-1}) = [\mu(g)]^{-1}$ в H; (iii) для всех $g, g' \in G$

$$\mu(gg') = \mu(g)\,\mu(g'). \tag{17}$$

Покажем, что условия (i) и (ii) на самом деле следуют из (iii).

Лемма. *Если отображение f групп обладает свойством* (iii), *то оно обладает также свойствами* (i) *и* (ii).

Доказательство. Так как $\mu(1_G)\,\mu(1_G) = \mu(1_G 1_G) = \mu(1_G)$, то образ 1_G является единственным идемпотентом в H, т. е. 1_H. Аналогично для всякого $g \in G$ из равенства

$$1_H = \mu(1_G) = \mu(gg^{-1}) = \mu(g)\,\mu(g^{-1})$$

следует, что элемент $\mu(g^{-1})$ является обратным к $\mu(g)$ в H.

Согласно этой лемме, для любых групп G, H морфизм полугрупп $\mu\colon G \to H$ автоматически является морфизмом групп. Заметим, что для моноидов это уже неверно. Если M — моноид с двумя идемпотентами 1_M и $e \neq 1_M$, то отображение $x \mapsto e$ для всех $x \in M$ является морфизмом полугрупп, но не моноидов.

Теорема 8. *Если $\mu\colon G \to H$ — любой морфизм групп, то* (i) *образ $\mu(S)$ всякой подгруппы $S \subset G$ является подгруппой в H*; (ii) *обратный образ $\mu^{-1}(T)$ всякой подгруппы $T \subset H$ является подгруппой в G.*

Доказательство. Предыдущая лемма показывает, что $\mu(S)$ удовлетворяет условиям теоремы 6, откуда следует (i). Утверждение (ii) также получается из теоремы 6. В самом деле, $1_G \in \mu^{-1}(T)$, ибо $\mu(1_G) = 1_H$ и $1_H \in T$. Далее, из $\mu(x) \in T$ и $\mu(y) \in T$ следует, что

$$\mu(x\ y) = \mu(x)\,\mu(y) \in T,$$

откуда $xy \in \mu^{-1}(T)$ в силу (14). Наконец, из $\mu(x) \in T$ вытекает, что $\mu(x^{-1}) = [\mu(x)]^{-1} \in T$. Поэтому $\mu^{-1}(T)$ удовлетворяет условиям теоремы 6.

Следствие. *Если $\theta\colon G \to H$ — эпиморфизм полугрупп и G — группа, то H также является группой.*

Доказательство. Так как $\theta(1_G g) = \theta(1_G)\,\theta(g)$ и $\theta(g 1_G) = \theta(g)\,\theta(1_G)$ для всех $\theta(g) \in H$, то $\theta(1_G)$ есть единица в H. Аналогично, для всякого $\theta(g) \in H$ элемент $\theta(g^{-1})$ является двусторонним обратным к $\theta(g)$. Даже ассоциативность умножения в H следует из ассоциативности умножения в G, и ее можно не постулировать.

Докажем, наконец, результат, утверждающий, что $[\mathbf{Z}, +]$ есть *свободная группа* с одной образующей.

Теорема 9. *Пусть G — группа и $a \in G$. Тогда существует единственный морфизм* $\mu\colon [\mathbf{Z}, +] \longrightarrow G$, *такой, что* $\mu(1) = a$.

Доказательство. Индукцией по положительным и отрицательным показателям степени сразу же убеждаемся, что любой морфизм μ, обладающий таким свойством, должен удовлетворять условиям $\mu(m) = a^m$ для всех m; поэтому такой морфизм, если он существует, единствен. Но в силу (15) отображение $m \mapsto a^m$ является морфизмом, что и требовалось доказать.

УПРАЖНЕНИЯ Б

1. Задать списком восемь симметрий квадрата и описать геометрически действие каждой.

2. Показать, что группа симметрий равностороннего треугольника (относительно композиции) изоморфна группе всех биекций $f\colon \mathbf{3} \longrightarrow \mathbf{3}$.

3. а) Показать, что в циклических группах порядков 5, 6 и 14 образующую можно выбрать 4, 2 и 6 способами соответственно.

б) Показать, что группы автоморфизмов этих групп циклические.

4. а) Показать, что группа $\mathrm{Aut}[\mathbf{Z}, +]$ всех автоморфизмов аддитивной группы $[\mathbf{Z}, +]$ целых чисел имеет порядок 2.

б) Показать, что автоморфизмы группы $[\mathbf{Z}_8, +]$ образуют нециклическую группу порядка 4.

5. Описать все изоморфизмы между группами $[\mathbf{Z}_4, +]$ и $[\mathbf{Z}_5^*, \cdot]$. (*Объяснение*: $\mathbf{Z}_5^* = \{1, 2, 3, 4\} \bmod 5$ с умножением в качестве композиции.)

6. Показать, что группа симметрий правильного тетраэдра изоморфна группе всех перестановок множества $\{1, 2, 3, 4\}$.

7. Построить некоторый эпиморфизм «аффинной» группы всех линейных преобразований вида $x \mapsto ax + b$ ($a \neq 0$, b вещественные) на группу $[\mathbf{Z}_2, +]$.

8. Показать, что эндоморфизмы любой группы образуют моноид относительно композиции.

В упр. 9 через $\tilde{M} = \{0, 1\}$ *обозначен моноид из двух элементов с таблицей умножения*

$\circ$	0	1
0	0	0
1	0	1

9. а) Пусть $M = [\mathbf{2}^{\mathbf{2}}, \circ]$ — моноид всех функций $f\colon \mathbf{2} \longrightarrow \mathbf{2}$ относительно левой композиции. Показать, что отображение $\theta\colon M \longrightarrow \tilde{M}$, переводящее биекции множества 2 в $1 \in M$, а все остальные f в 0, является эпиморфизмом M на $\tilde{M}$.

б) Построить аналогичный эпиморфизм из $M_n = [\mathbf{n}^{\mathbf{n}}, \diamondsuit]$ на M.

10. а) Пусть M — любой *конечный* моноид, не являющийся группой, и пусть M^* — множество его обратимых элементов. Показать, что отображение $\theta\colon M \longrightarrow \tilde{M}$, $\theta(M^*) = 1$, $\theta(M \setminus M^*) = 0$, является эпиморфизмом.

б) Показать, что для бесконечных моноидов это утверждение может нарушаться. (*Указание*: рассмотреть моноид всех отображений $\mathbf{N}$ в $\mathbf{N}$.)

7.7. АБЕЛЕВЫ ГРУППЫ

Многие важные группы коммутативны; их называют также *абелевыми* группами. В частности, алгебраические системы с естественной операцией сложения часто оказываются абелевыми группами относительно этой операции (например, множество вещественных матриц одного и того же порядка). В таких аддитивных абелевых группах удобно обозначать бинарную операцию знаком $+$, единицу группы знаком 0, а элемент, обратный к x, символом $-x$.

В этом параграфе мы перечислим некоторые специальные свойства абелевых групп, используя аддитивные обозначения. Аксиомы абелевых групп в них приобретают следующий вид:

$$x+y=y+x, \tag{18а}$$

$$x+(y+z)=(x+y)+z, \tag{18б}$$

$$x+0=0+x=x, \tag{18в}$$

$$x+(-x)=(-x)+x=0. \tag{18г}$$

Далее, целые степени a^r ($r \in \mathbf{Z}$) элемента a в аддитивных обозначениях превращаются в ra, а формулы (15) имеют такой вид:

$$(r+s)a=ra+sa, \quad s(ra)=(sr)a \text{ для всех } r,\ s \in \mathbf{Z}. \tag{19}$$

Формально эти тождества выглядят как законы дистрибутивности и ассоциативности. Кроме того, в абелевых группах справедлива формула $(ab)^r=a^rb^r$, которая может быть и неверной в *не*абелевой группе. Эта формула в аддитивной записи выглядит так:

$$r(a+b)=ra+rb \text{ для всех } r \in \mathbf{Z};\ a, b \in A. \tag{20}$$

Равенство (20), которое легко доказать с помощью индукции, можно переформулировать в терминах морфизмов следующим образом. Напомним, что эндоморфизм группы A есть морфизм $A \to A$.

Теорема 11. *Отображение $A \to A$: $a \mapsto ra$ для любого целого числа r является эндоморфизмом абелевой группы A.*

Следствие. *В любой абелевой группе A для любого целого числа r множество rA всех элементов вида ra ($r \in \mathbf{Z}$, $a \in A$) и множество $0:r$ всех элементов $a \in A$, обладающих свойством $ra=0$, являются подгруппами этой группы.*

Кроме того, из (20) следует, что в любой абелевой группе A множество всех элементов конечного порядка составляет подгруппу. Она называется *подгруппой кручения* в A.

Укажем еще несколько примеров абелевых групп. Ясно, что всякая циклическая группа абелева (см. (15)). Путем построения прямых произведений циклических групп можно получить много других примеров (§ 7.4). Следующая теорема почти очевидна.

Теорема 12. *Прямое произведение $G\times H$ любых двух групп G и H является группой. Прямое произведение абелевых групп абелево.*

Доказательство. Как и для моноидов, $(1_G, 1_H)$ есть единица в $G\times H$. Обратным к элементу $(g, h)\in G\times H$ является (g^{-1}, h^{-1}) (мы перешли к мультипликативной записи). Наконец, если $gg'=g'g$ для всех $g, g'\in G$ и $hh'=h'h$ для всех $h, h'\in H$, то

$$(g, h)(g', h')=(gg', hh')=(g'g, h'h)=(g', h')(g, h)$$

для всех (g, h) и (g', h') в $G\times H$.

Пример 13. Для любого простого числа p прямое произведение $Z_p\times\ldots\times Z_p$ n экземпляров циклической группы $Z_p=[\mathbf{Z}_p, +]$ называется *элементарной абелевой группой* порядка p^n. В частности, элементарные абелевы группы порядка 2^n получаются из булевых алгебр порядка 2^n с помощью следующей конструкции.

Пример 14. Пусть A — любая булева алгебра. Тогда система $G=[A, +]$ со следующей самодвойственной операцией сложения является абелевой группой:

$$a+b=(a\wedge b')\vee(a'\wedge b)$$

(см. § 5.1). Элемент 0 является единицей группы в G, и каждый элемент x обратен самому себе: $x+x=0$ для всех $x\in A$.

Легко проверить, что $C_2\times C_3\cong C_6$: единица $(1, 1)\in C_2\times C_3$ имеет порядок 6 и порождает всю группу C_6. Более общо, пусть m, n — взаимно простые целые положительные числа. Тогда группа $C_m\times C_n\cong C_{mn}$ циклическая с образующей $(1, 1)$ порядка mn, ибо $(r, r)=(0, 0)$ тогда и только тогда, когда m и n являются делителями r.

Отсюда следует, что любое прямое произведение циклических групп можно представить в виде прямого произведения циклических групп, порядки которых суть степени простых чисел. Основная теорема об абелевых группах утверждает, что и, обратно, любая *конечная* абелева группа изоморфна прямому произведению циклических групп, порядки которых суть степени простых чисел. Это обстоятельство позволяет без труда перечислить *все* абелевы группы данного конечного порядка.

УПРАЖНЕНИЯ В

1. а) Показать, что если $a^2=1$ для всех элементов a группы G, то G коммутативна.

б) То же для моноидов.

2. Доказать тщательно и строго, со всеми подробностями, что каждая подгруппа циклической группы циклическая.

3. Показать, что группа G абелева в том и только том случае, если отображение $a\mapsto a^2$ является эндоморфизмом G.

4. Показать, что группа G абелева тогда и только тогда, когда отображение $a \mapsto a^{-1}$ является эндоморфизмом G.

5. Сколько имеется (с точностью до изоморфизма) абелевых групп порядка 36? порядка 300? порядка p^6 (если p — простое число)?

6. а) Показать, что в любой аддитивной абелевой группе G для любого $n \in \mathbf{P}$ множество $0{:}\, n = \{x \in G \mid nx = 0\}$ является подгруппой.

б) Показать, что в группе симметрий квадрата множество всех элементов порядка 2 не является подгруппой.

7. В булевой алгебре A положим $a + b = (a \wedge b') \vee (a' \wedge b)$. Показать, что $[A, +]$ является абелевой группой, в которой каждый неединичный элемент имеет порядок 2.

8. Показать, что если $F = [\mathbf{Z}_2, +] \times [\mathbf{Z}_2, +]$ (4-группа), то $\mathrm{Aut}\, F$, группа автоморфизмов F, изоморфна симметрической группе порядка 3.

9. Назовем *антиавтоморфизмом* полугруппы S всякую биекцию $\theta{:}\, x \mapsto x'$, такую, что $(xy)' = y'x'$.

а) Показать, что моноид $\mathbf{n}^{\mathbf{n}}$ всех функций $f{:}\, \mathbf{n} \to \mathbf{n}$ не имеет нетривиальных антиавтоморфизмов. (*Указание*: рассмотреть его левые и правые нули.)

*б) Показать, что у любой конечной группы, имеющей более двух элементов, имеется нетривиальный антиавтоморфизм. (Разрешается пользоваться основной теоремой об абелевых группах.)

7.8. ДЕЙСТВИЯ ГРУПП НА МНОЖЕСТВАХ

Мы введем сейчас фундаментальное понятие действия группы на множестве. Пусть $f{:}\, X \to X$ — любая биекция множества на себя. Тогда ее положительные и отрицательные степени f^r, f^{-r} вместе с тождественным отображением $1_X = f^0$ образуют циклическую группу, действующую на X. Точнее, как и в § 7.2, мы имеем

$$f^s(f^r(x)) = f^{r+s}(x) \text{ для всех } x \in S;\ r, s \in \mathbf{Z}, \tag{21}$$

т. е.

$$f^r \diamond f^s = f^s \diamond f^r = f^{r+s} \text{ для всех } r, s \in \mathbf{Z}.$$

Это позволяет определить два морфизма, μ и $\nu = \mu^{\mathrm{opp}}$ *аддитивной* группы $[\mathbf{Z}, +]$ в *мультипликативную* симметрическую группу, состоящую из всех биекций X, относительно правой и левой композиции соответственно.

Иными словами, отображение $r \mapsto f^r$ определяет *действие* (или представление перестановки: см. § 7.9) группы $[\mathbf{Z}, +]$ на множестве X. Это понятие обобщается на случай любого моноида $[M, \cdot]$.

Определение. Пусть M — мультипликативный моноид, X — множество. Отображение $\mu{:}\, M \to X^X$ называется *действием* M на X, если $\mu(1) = 1_X$ и либо

$$\mu(rs) = \mu(r) \diamond \mu(s) \text{ для всех } r, s \in M, \tag{22}$$

либо

$$\mu(rs) = \mu(r) \circ \mu(s) \text{ для всех } r, s \in M. \tag{22'}$$

Точнее, в случае (22) μ определяет *правое* действие M на X, а в случае (22′) — *левое* действие M на X.

Таким образом, правое (левое) действие M на X есть морфизм μ моноида M в моноид X^X относительно правой (левой) композиции соответственно. Если $M=G$ есть группа, то из соотношений

$$\mu(g)\diamond\mu(g^{-1})=\mu(gg^{-1})=\mu(1)=1_X \text{ и } \mu(g^{-1})\diamond\mu(g)=1_X$$

видно, что все отображения $\mu(g)$ при действии справа или слева должны быть биекциями. Для любой функции $f\colon X\to X$ отображение $r\mapsto f^r$ определяет действие аддитивного моноида $[\mathbf{N}, +]$ на X^X, где, как обычно, $f^0=1_X$.

Любое действие $\mu\colon G\to X^X$ группы G на множестве X определяет следующее важное *отношение эквивалентности* γ на X:

$$x\gamma y \text{ означает, что } \mu_g(x)=y \text{ для некоторого } g\in G, \qquad (23)$$

где μ_g означает то же, что и раньше $\mu(g)$.

Иными словами, $x\gamma y$, если некоторое преобразование из G переводит x в y. Отношение (23), очевидно, рефлексивно, ибо $1_X(x)=x$ и $\mu_1=1_X$. Оно симметрично, ибо из $\mu_g(x)=y$ следует, что $\mu_{g^{-1}}(y)=x$ (так как $\mu_{g^{-1}}=(\mu_g)^{-1}$). Оно транзитивно, ибо из $\mu_g(x)=y$ и $\mu_h(y)=z$ следует, что $\mu_{gh}(x)=(\mu_g\diamond\mu_h)(x)=z$ для любого правого действия, и

$$\mu_{hg}(x)=(\mu_h\circ\mu_g)(x)=\mu_h(\mu_g(x))=\mu_h(y)=z, \qquad hg\in G,$$

для любого левого действия G и X. Итак, мы доказали такой результат:

Теорема 13. *Для любого действия группы G на множестве X отношение* (23) *является отношением эквивалентности на X.*

Классы эквивалентности множества X относительно отношения эквивалентности (23) называются *орбитами* данного действия. Действие называется *транзитивным*, если все X составляет одну орбиту, т. е. $x\gamma y$ для всех $x, y\in X$ (для всех $x, y\in X$ существует элемент $g\in G$, такой, что $\mu_g(x)=y$).

Например, представление Кэли группы G правыми сдвигами на множестве G определяет транзитивное правое действие группы G на множестве своих элементов: $\mu_g(x)=xg$ для всех $g\in G$. Аналогично, левые сдвиги $\nu_g(x)=gx$ для всех $x, g\in G$ определяют транзитивное левое действие G на себе. Эти действия *просто* транзитивны в том смысле, что для всех $x, y\in G$ существует ровно один элемент $g\in G$, обладающий свойством $xg=y$, и ровно один элемент $h\in G$, обладающий свойством $hx=y$.

Евклидова группа всех движений плоскости как твердого тела по себе транзитивна на множестве точек плоскости. Орбиты этой группы на множестве треугольников (или других плоских фигур) определяют отношение, которое в геометрии называется *конгру-*

энтностью. Подгруппа всех параллельных переносов плоскости просто транзитивна на точках плоскости; это представление $[\mathbf{R}, +]^2$ является частным случаем представления Кэли. Из (23) ясно, что две прямые лежат в одной орбите тогда и только тогда, когда они параллельны. Наконец, многоугольник является правильным в том и только том случае, если его группа изометрий транзитивна на вершинах и сторонах.

7.9. ПЕРЕСТАНОВКИ

Биекция β: $X \leftrightarrow X$ конечного множества в себя обычно называется *перестановкой* (элементов) множества X. Степени такой биекции β образуют циклическую группу перестановок. Рассмотрим действие циклической группы $\{\beta^r\}$ как морфизм моноида $[\mathbf{Z}, +]$ в X^X. Орбиты такого действия называются *циклами* перестановки β. Из теоремы 13 следует, что X является разделенным объединением своих циклов.

Стандартный способ указания перестановки β множества $\mathbf{n} = \{1, \ldots, n\}$ состоит в последовательной записи в скобках сначала цикла

$$\gamma_1 = (1, 1\beta, 1\beta^2, \ldots, 1\beta^{m_1 - 1}),$$

где 1β означает $\beta(1)$, $1\beta^2$ означает $\beta(\beta(1))$ и т. д., затем (не пересекающегося с первым) цикла γ_2, который начинается с первого числа $j_2 \in \mathbf{n}$, не входящего в γ_1, затем цикла, который начинается с первого числа $j_3 \in \mathbf{n}$, не входящего ни в γ_1, ни в γ_2, и т. д., пока не будет исчерпано все $\mathbf{n}$. Циклы, состоящие из одного элемента, можно опустить: в них входят те элементы, которые β оставляет на месте. Разделяющие символы и запятые тоже можно опустить, если это не приведет к двусмысленности.

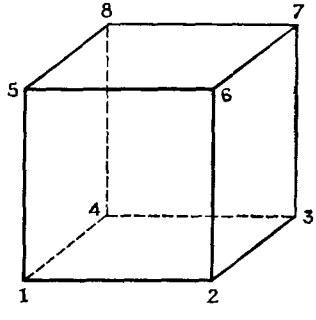


Рис. 7.6.

Пример 16. Симметрии куба на рис. 7.6 индуцируют 48 различных перестановок его вершин. Так, вращение на 90° против часовой стрелки вокруг вертикальной оси индуцирует перестановку

$$(1234)(5678) = (5678)(1234).$$

Вращение на 120° вокруг диагонали $\overrightarrow{17}$ индуцирует перестановку $(245)(386)$ и т. д.

Заметим, что *непересекающиеся циклы*, рассматриваемые как перестановки, *попарно коммутируют*: $\gamma_k \gamma_l = \gamma_l \gamma_k$ для всех k, l.

Поэтому порядок записи циклов в разложении данной перестановки несуществен. Цикл длины 2 называется *транспозицией*. Имеет место следующий результат.

Лемма. *Любую перестановку* β *множества* $\mathbf{n}$ *можно записать в виде произведения* m *транспозиций, где* $m \leqslant n(n-1)/2$.

Доказательство. В списке $1\beta^{-1}$, $2\beta^{-1}$, $3\beta^{-1}$, ..., $n\beta^{-1}$ будем переносить $n=(n\beta)\beta^{-1}$ на одно место направо (меняя его местами с правым соседом) столько раз, сколько это возможно. Это число окажется самым последним в списке через не более чем $n-n\beta \leqslant n-1$ транспозиций. Затем повторим эту процедуру для $n-1$, для $n-2$ и т. д. После самое большее

$$m=(n-1)+(n-2)+\ldots+1=\frac{n(n-1)}{2}$$

транспозиций мы переместим каждое число $k\beta^{-1}$ на k-е место, получив в результате перестановку β: $k\beta^{-1}\mapsto k$.

Четные и нечетные перестановки. Перестановка φ множества $\mathbf{n}$ называется *четной* или *нечетной* в зависимости от того, является ли число пар $(i\varphi, j\varphi)$ с $i, j \in \mathbf{n}$, $i<j$ и $i\varphi > j\varphi$, четным или нечетным. Положим

$$K_n=\prod_{\substack{i<j\\ i,\,j\in\mathbf{n}}}(j-i)=1^{n-1}2^{n-2}3^{n-3}\ldots(n-1)=\prod_{k=1}^{n-1}k!. \qquad (24)$$

Так как множество всех *неупорядоченных* пар (i, j) с $i \neq j$ переходит в себя под действием φ, ясно, что $\prod_{i<j}(j\varphi-i\varphi)=K_n$ или $-K_n$ в зависимости от того, четна или нечетна φ. В первом случае мы пишем $\operatorname{sgn}\varphi=+1$, во втором $\operatorname{sgn}\varphi=-1$.

Каждая *транспозиция* $\tau=(i, j)$ нечетна: она обращает знак $j-i$, а также всех $j-k$ и $k-i$ с $i<k<j$, так что в произведении будет $2(j-i)-1$ отрицательных чисел. Таким образом, $\tau(K_n)=-K_n$ и, более того, $\tau(\varphi(K_n))=-\varphi(K_n)$, для любой перестановки φ, ибо φ просто переписывает те же числа в другом порядке. Пользуясь леммой и этим результатом, получаем, что φ четна или нечетна в зависимости от того, разлагается она в произведение четного или нечетного числа транспозиций. В качестве следствия мы получаем такой результат:

Теорема 14. *Функция* $\varphi\mapsto\operatorname{sgn}\varphi$ *является эпиморфизмом группы перестановок на группу по умножению* $\{\pm 1\}$. *Четные перестановки образуют подгруппу группы* S_n.

Эта подгруппа называется *знакопеременной* группой степени n и обозначается через A_n.

УПРАЖНЕНИЯ Г

1. Пусть $\alpha = (12)(354)$. Перечислить все степени α в стандартных обозначениях (произведение попарно не пересекающихся циклов).

2. Показать, что любой 3-цикл $\gamma = (pqr)$ можно представить в виде коммутатора $\alpha = \sigma^{-1}\tau^{-1}\sigma\tau$ двух 2-циклов (транспозиций).

3. а) Представить (1234567) в виде произведения 3-циклов (возможно, пересекающихся).

б) Представить (1234)(56) и (1234)(5678) в виде произведения 3-циклов.

*4. Показать, что любую четную перестановку можно представить в виде произведения 3-циклов (возможно, пересекающихся).

5. Показать, что соотношения $s^2 = t^2 = (st)^3 = 1$ определяют группу, изоморфную симметрической группе степени 3.

6. Показать, что симметрическая группа S_n порождена циклическими перестановками (12) и $(23\ldots n)$.

*7. Построить действие группы $x \mapsto \dfrac{ax+b}{cx+d}$, где a, b, c, d вещественны и $ad-bc=1$, на сумме $R \sqcup \{\infty\}$.

7.10. ТЕОРЕМА ЛАГРАНЖА

Пусть G — любая группа и S — ее подгруппа. Рассмотрим действие справа S на G правыми сдвигами $x \mapsto xs$ $(s \in S)$. Из тождества $x(ss') = (xs)s'$ видно, что это определение корректно (оно индуцирует морфизм $S \to G^G$ относительно правой композиции). Это действие является ограничением на S представления Кэли. Единица $1 \in S$ индуцирует тождественное отображение группы G.

Орбита любого элемента g состоит из всех gs с $s \in S$. Это множество обозначается через gS и называется *левым* смежным классом подгруппы S, содержащим g. Теорема 13 в применении к этому случаю доставляет следующий результат:

Лемма 1. *Множество левых смежных классов gS относительно любой подгруппы S группы G образует разбиение множества G.*

Иными словами, объединение левых смежных классов совпадает с G и два класса либо не пересекаются, либо совпадают. Это нетрудно проверить и непосредственно. Если элемент $u = gs = hs'$ принадлежит одновременно gS и hS, то $gs'' = hs's^{-1}s''$ для всех $s'' \in S$; так как $s's^{-1}s'' \in S$, отсюда видно, что $gS \subset hS$. Аналогично, $hS \subset gS$, так что левые смежные классы с непустым пересечением совпадают. Поскольку $g = g1 \in gS$, их объединение равно G.

Назовем *индексом* $[G : S]$ подгруппы S в G число различных левых смежных классов относительно S. Очевидно, $[G : 1] = o(G)$ — порядок G. Так как левый сдвиг G на любой элемент $g \in G$ является биекцией G, каждый левый смежный класс gS в S состоит

в точности из $o(S)=[S:1]$ элементов. Объединяя это наблюдение с леммой 1, мы получаем теорему Лагранжа в следующей форме:

Теорема 15. *Для любой подгруппы S группы G*

$$[G:1]=[G:S][S:1]; \tag{25}$$

иными словами, порядок G равен произведению порядка S на индекс S в G.

Следствие 1. *Порядок любого элемента группы G делит порядок G.*

Действительно, порядок элемента, по определению, равен порядку порожденной им циклической подгруппы, который делит $o(G)$ в силу теоремы 15.

Следствие 2. *Всякая группа простого порядка p является циклической.*

Действительно, порядок любого неединичного элемента c такой группы должен быть больше единицы и делить p; поэтому он совпадает с p, так что этот элемент порождает всю группу.

Следствие 3. *Пусть G — нетривиальная группа. У нее нет собственных подгрупп тогда и только тогда, когда $o(G)$ — простое число.*

Мы опускаем доказательство.

Пример 17. Обозначим через S подгруппу диэдральной группы Δ_n, состоящую из 1 и отражения относительно оси симметрии, проходящей через фиксированную вершину 1. Тогда для всякого элемента $g\in\Delta_n$ класс Sg состоит из всех симметрий, переводящих 1 в g (1); класс gS — из всех симметрий, переводящих в 1 ту же вершину, что и g. Наконец, множество $g^{-1}Sg$ всех $g^{-1}sg$ с $s\in S$ состоит из 1 и ее отображения относительно оси симметрии, проходящей через вершину g (1).

Теперь мы докажем результат, который будет полезен в теории конечных полей, излагаемой в гл. 12.

Теорема 16. *Пусть G — нециклическая абелева группа конечного порядка g. Тогда существует такой собственный делитель h числа g, что $x^h=1$ для всех $x\in G$.*

Доказательство. Представим $g=p_1^{e_1}p_2^{e_2}\ldots p_r^{e_r}$ в виде произведения степеней различных простых чисел, $p_1<p_2<\ldots<p_r$. Обозначим через e_i' наибольшее целое число, такое, что G содержит элемент g_i порядка $p^{e_i'}$. Из теоремы 15 следует, что $e_i'\leqslant e_i$. Разберем два возможных случая.

Случай 1. $e_i' = e_i$ для всех i. Пусть $y_1, \ldots, y_r$ — элементы порядков $p_1^{e_1}, \ldots, p_r^{e_r}$. Так как они коммутируют, элемент $y = y_1 \ldots y_r$ должен иметь порядок $p_1^{e_1} \ldots p_r^{e_r}$; поэтому G циклическая.

Случай 2. $e_i' < e_i$ для некоторого i, так что число $h = \prod p_i^{e_i'}$ является собственным делителем g. Пусть порядок элемента $x \in G$ равен $p_1^{c_1} \ldots p_r^{c_r}$, где c_i зависит от x. Полагая $n_i = \prod_{j \neq i} p_j^{c_j}$, находим, что элемент $y_i = x^{n_i}$ имеет порядок $p_i^{c_i}$, поэтому $e_i \leqslant e_i'$. Следовательно, порядок x делит h и h — собственный делитель g, не зависящий от x.

УПРАЖНЕНИЯ Д

1. а) Найти в S_5 такой 3-цикл γ, что $\gamma^{-1}(123)\gamma = (124)$.
б) Вывести отсюда, что при $n \geqslant 5$ все 3-циклы сопряжены в знакопеременной группе A_n степени n.

Коммутантом $G' = [G, G]$ *группы* G *называется подгруппа* G, *порожденная всеми коммутаторами* $a^{-1}b^{-1}ab$.

2. а) Показать, что если G абелева, то $G' = \{1\}$.
б) Показать, что $\alpha(G') = G'$ для любого автоморфизма α группы G.
В следующих упражнениях используются определения из § 7.11.
в) Показать, что G' является нормальной подгруппой G и что факторгруппа G/G' абелева.
г) Показать, что если N — нормальная подгруппа G, то факторгруппа G/N абелева в том и только том случае, если $N \supset G'$.

3. Показать, что для всех n подгруппа A_n в S_n является коммутантом S_n. (*Указание*: воспользоваться упр. Г4 и Д1.)

*4. Пусть H — нормальная подгруппа A_n, $n \geqslant 5$.
а) Показать, что если H содержит 3-цикл, то $H = A_n$. (*Указание*: воспользоваться упр. Г4 и Д1.)
б) Показать, что если $H \neq \{1\}$, то H содержит 3-цикл.
в) Вывести отсюда, что $H = A_n$, т. е. группа A_n простая при всех $n \geqslant 5$.

5. Сколько имеется неизоморфных помеченных квадратов, вершины которых помечены красным, желтым или зеленым цветом? Задать их списком.

6. Тот же вопрос для помеченных шестиугольников.

*7. Пусть S — подгруппа группы G. Введем отношение xE_Sy: $xy^{-1} \in S$.
(i) Показать, что E_S — отношение эквивалентности на G.
(ii) Описать классы эквивалентности.
(iii) В каких случаях отношение E_S обладает свойством подстановки а) для унарной операции $x \mapsto x^{-1}$? б) для операции xy?

7.11. НОРМАЛЬНЫЕ ПОДГРУППЫ

Пусть a — элемент группы G. Отображение C_a: $x \mapsto a^{-1}xa$ группы в себя называется *сопряжением справа* посредством a. Из равенства $a^{-1}(xy)a = (a^{-1}xa)(a^{-1}ya)$ видно, что C_a является эндо-

морфизмом группы G (т. е. ее морфизмом в себя). Это верно и в моноидах для любого обратимого элемента a.

Далее, $C_b(C_a(x)) = b^{-1}(a^{-1}xa)b = C_{ab}(x)$ для всех $a, b, x \in G$. Это означает, что отображение γ: $a \mapsto C_a$ определяет морфизм γ: $G \longrightarrow G^G$. В частности, C_a есть биекция G, т. е. автоморфизм; обратная биекция совпадает с C_a^{-1}. Итак, мы доказали следующий результат:

Теорема 17. *В любой группе G правое сопряжение посредством любого элемента $a \in G$ определяет автоморфизм C_a: $x \mapsto a^{-1}xa$ этой группы. Отображение $a \mapsto C_a$ определяет правое действие группы G на самой себе.*

Определение. Автоморфизмы вида $x \mapsto a^{-1}xa$ называются *внутренними автоморфизмами* группы (или моноида) G.

Очевидно, единственным внутренним автоморфизмом абелевой группы является тождественное отображение. Однако у каждой неабелевой группы есть нетривиальные внутренние автоморфизмы.

Орбиты группы G относительно внутренних автоморфизмов называются *классами сопряженных элементов*. Заметим, что понятия внутреннего автоморфизма и класса сопряженных элементов не изменятся, если в предыдущих определениях заменить сопряжение справа сопряжением слева, потому что $bxb^{-1} = (b^{-1})^{-1}x(b^{-1})$.

Определение. Подгруппа S группы G называется *нормальной* в G, если $g^{-1}Sg = S$ для всех $g \in G$. Это отношение записывается так: $S \lhd G$.

Лемма 1. *Любое из следующих двух условий на подгруппу S группы G равносильно ее нормальности*:

(i) $gS = Sg$ *для всех* $g \in G$;

(ii) $g^{-1}Sg \subset S$ *для всех* $g \in G$.

Доказательство. Если $S \lhd G$, то $S = g^{-1}Sg$ для всех $g \in G$, откуда

$$gS = gg^{-1}Sg = Sg,$$

для всех $g \in G$; значит, из $S \lhd G$ следует (i). Из (i) сразу же следует (ii). Наконец, из (ii) следует $S \lhd G$, поскольку если $g^{-1}Sg \subset S$ для всех $g \in G$, то, в частности, $(g^{-1})^{-1}Sg^{-1} \subset S$, откуда

$$S = g^{-1}(g^{-1})^{-1}Sg^{-1}g \subset g^{-1}Sg,$$

так что $g^{-1}Sg = S$. Мы установили импликации $S \lhd G \Rightarrow$ (i) $\Rightarrow$ $\Rightarrow$ (ii) $\Rightarrow S \lhd G$ и завершили доказательство леммы.

Заметим, что, согласно лемме 1(i), правые и левые смежные классы нормальной подгруппы совпадают; поэтому можно говорить просто о смежных классах.

Теорема 18. *Пусть* $\theta\colon G \to H$ *— любой морфизм групп. Тогда обратный образ* $\theta^{-1}(1_H) = K$ *единицы* 1_H *относительно* θ *является нормальной подгруппой в* G.

Доказательство. Ясно, что если $x \in K$ и $y \in K$, то

$$\theta(xy) = \theta(x)\theta(y) = 1_H 1_H = 1_H.$$

Далее, $\theta(x^{-1}) = [\theta(x)]^{-1} = 1_H^{-1} = 1_H$. Поэтому K — подгруппа. Аналогично, если $x \in K$ и $g \in G$, то

$$\theta(g^{-1}xg) = [\theta(g)]^{-1}\theta(x)\theta(g) = [\theta(g)]^{-1}1_H\theta(g) = [\theta(g)]^{-1}\theta(g) = 1_H.$$

Поэтому K — нормальная подгруппа.

Определение. Обратный образ 1_H относительно морфизма $\theta\colon G \to H$ называется *ядром* θ.

Факторгруппы. Обращение теоремы 18 также верно и доставляет важный способ построения групп. Для доказательства нам понадобится вспомогательная конструкция.

Пусть S, T — два непустых подмножества моноида M. Определим их произведение: $ST = \{st \mid s \in S, t \in T\}$. Оно также является непустым подмножеством M. Часто оказывается полезным следующее замечание.

Лемма. *Непустые подмножества любого моноида* M *сами образуют моноид* $\mathcal{P}(M) \setminus \{\varnothing\}$ *относительно умножения.*

Действительно, $(ST)U = S(TU)$ есть множество всех произведений stu, где $s \in S$, $t \in T$, $u \in U$. Единицей служит множество $\{1\}$. Однако заметим, что если через S^{-1} обозначить множество всех обратных s^{-1} $(s \in S)$, то неверно, что $SS^{-1} = 1$ (кроме случая $S = \{s\}$). В группе G включение $SS^{-1} \subset S$ имеет место тогда и только тогда, когда S — подгруппа G.

Теорема 19. *Пусть* N *— любая нормальная подгруппа группы* G. *Тогда существует эпиморфизм* $\theta\colon G \to H$ *с ядром* N, *где* H *— некоторая группа, которая определяется по* N *однозначно с точностью до изоморфизма и обозначается через* G/H.

Доказательство. Рассмотрим отображение $\theta\colon x \mapsto Nx$ группы G на множество ее смежных классов относительно N. Определим умножение смежных классов, как выше: $(Nx)(Nx') = \{yy' \mid y \in Nx, y' \in Nx'\}$. Нетрудно проверить, что произведением будет снова смежный класс:

$$(Nx)(Nx') = N(xN)x' = N(Nx)x' = NNxx' = Nxx'.$$

Таким образом, смежные классы образуют полугруппу, и θ является эпиморфизмом полугрупп. Но так как G — группа, то, согласно следствию из теоремы 6, множество смежных классов Nx тоже является группой.

Пусть теперь $\theta\colon G \to H$ — любой эпиморфизм с ядром N. Тогда $\theta(x) = \theta(y)$ в H в том и только том случае, если

$$\theta(xy^{-1}) = \theta(x)\theta(y^{-1}) = \theta(x)[\theta(y)]^{-1} = \theta(y)[\theta(y)]^{-1} = 1_H.$$

Это равенство означает, что $xy^{-1} \in N$, $x \in Ny$, т. е. $Nx = Ny$. Значит, разбиение G на классы эквивалентности относительно отображения θ (см. § 2.3 и 2.4) совпадает с разбиением на смежные классы относительно N. Тождественная биекция этих двух разбиений согласуется с законом умножения смежных классов в G, с одной стороны, и элементов H, с другой стороны. Поэтому она определяет изоморфизм H и группы смежных классов.

Пример 18. В аддитивной группе $[\mathbf{Z}, +]$ обозначим через (n) подгруппу всех кратных числа $n \in \mathbf{P}$. Тогда $\mathbf{Z}/(n)$ — аддитивная группа классов вычетов $\bmod n$.

Пример 19. Каждая перестановка $\pi \in S_n$ чисел $1, \ldots, n$ переводит многочлен $P = \prod_{i<j}(x_i - x_j)$ либо в P, либо в $-P$. Это определяет эпиморфизм $S_n \to (\pm 1)$ группы S_n на мультипликативную циклическую группу второго порядка. Ядром этого эпиморфизма, по определению, является знакопеременная группа $A_n \lhd S_n$, состоящая из всех четных перестановок этих чисел.

УПРАЖНЕНИЯ Е

1. Перечислить все элементы группы S_5, сопряженные с (12) (34).

2. Показать, что в симметрической группе S_n две перестановки, представленные в виде произведения непересекающихся циклов, сопряжены тогда и только тогда, когда наборы длин циклов у этих перестановок совпадают.

3. Описать все классы сопряженных элементов в группе S_4 и указать число элементов в каждом.

4. Тот же вопрос для S_5.

5. Показать, что если N — максимальная нормальная подгруппа в G, то G/N — простая группа, т. е. у нее нет нетривиальных нормальных подгрупп.

6. Показать, что подгруппа S в группе G нормальна, если она содержит все сопряженные элементы для некоторой системы образующих группы G.

7. а) Описать шесть симметрий куба, оставляющих на месте одну его вершину.

б) Показать, что у куба есть ровно 48 симметрий.

8. Показать, что у правильного октаэдра также есть ровно 48 симметрий, и связать их с симметриями куба.

9. Показать, что сдвиги $(x, y) \mapsto (x+a, y+b)$ (x, y)-плоскости образуют нормальную подгруппу в группе всех аффинных преобразований вида

$$(x, y) \mapsto (\alpha x + \beta y + a, \gamma x + \delta y + b), \quad \alpha\delta \neq \beta\gamma.$$

10. а) Показать, что если n не делится на простое число p, то $n^{p-1} \equiv 1 \bmod p$. (*Указание:* рассмотреть мультипликативную группу $[\mathbf{Z}_p - \{0\}, \cdot]$.)

б) Показать, что если p — простое число, то $n^p \equiv n \bmod p$ для всех $n \in \mathbf{Z}$. (*Указание:* воспользоваться п. а и разобрать отдельно оставшийся случай.)

11 Предположим, что $2^r + 1 = p$ есть простое число.

а) Показать, что в группе $[\mathbf{Z}_p - \{0\}, \cdot]$ порядок 2 равен $2r$.

б) Используя упр. 10, вывести отсюда, что $2r$ является делителем числа $p - 1 = 2^r$.

в) Заключить отсюда, что $r = 2^s$ есть степень двойки. (Числа $p = 2^{2^s} + 1$ называются числами Ферма.)

СПИСОК ЛИТЕРАТУРЫ

1. Chevalley C., Fundamental concepts of Algebra, Academic Press, 1956.
2. Hall M., The Theory of Groups, Macmillan, 1960. (Русский перевод: Холл М., Теория групп, ИЛ, М., 1962.)
3. Higman B., Applied Group-Theoretic and Matrix Methods, Dover 1955.
4. Ляпин С., Теория полугрупп, «Наука», М., 1960.
5. Weyl H., Symmetry, Princeton University Press, 1952. (Русский перевод: Вейль Г., Симметрия, «Наука», М., 1968.)
6. Wielandt H., Finite Permutation Groups, Academic Press, 1964.

8

ДВОИЧНЫЕ ГРУППОВЫЕ КОДЫ

8.1. ВВЕДЕНИЕ

Эта глава посвящена одному из методов решения следующей важной проблемы в теории передачи информации: двоичное кодирование и декодирование, обеспечивающие надежную передачу по *каналам* с «шумом». Типичная ситуация такова: мы хотим передать *сообщение*, которое может быть *строкой* символов некоторого конечного алфавита: {0, 1}, или {строчные и/или прописные латинские буквы}, или {арабские цифры} и т. п. Например, сообщение может быть текстом на английском или русском языке (тогда в алфавит следует включить пробелы и знаки препинания). Сообщения могут также состоять из двоичных или десятичных строк ограниченной длины, как в тех случаях, когда информация передается с одной вычислительной машины на другую или результаты телеметрических измерений передаются с космической станции.

Так или иначе, *передача данных* сводится к передаче по некоторому каналу связи знаков некоторого конечного алфавита. Практически всегда канал связи не идеален в том смысле, что с ненулевой вероятностью q переданный символ будет принят неправильно. Если передаваемые сигналы длинны, то даже по видимости малая вероятность ошибки, скажем $q = 10^{-6}$ на символ, может оказаться нетерпимой. Так, например, обстоит дело в системах связи ЭВМ—ЭВМ.

Например, одна вычислительная машина может быть связана с другой через спутник. В этом случае обычно используется двоичный алфавит {0, 1}. Канал связи физически реализуется электромагнитным полем между поверхностью Земли и спутником. Электромагнитные сигналы, соответствующие 0 и 1, накладываясь на внешнее поле, могут исказиться и ослабиться до неузнаваемости. Такие системы особенно чувствительны к солнечным пятнам, атмосферным условиям и т.д.

Двоичные симметричные каналы. При математическом анализе систем связи обычно пользуются упрощенными моделями. Для двоичного алфавита {0, 1} простейшая достаточно

реалистическая модель называется *двоичным симметричным* каналом. Она используется наиболее часто, и только ею мы будем заниматься в этой главе.

Пусть двоичные сигналы 0, 1 последовательно передаются по каналу связи на приемник. На рис. 8.1 представлена ситуация, когда каждый символ принимается правильно с вероятностью p и ошибочно с вероятностью $q = 1 - p$. Сверх того предполагается, что ошибки в передаче последовательных символов происходят *независимо*, в смысле следующего определения.

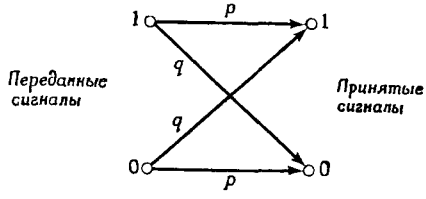


Рис. 8.1. Вероятности перехода в двоичном симметричном канале.

Определение. Пусть $\mathscr{E}_1$, $\mathscr{E}_2$, $\mathscr{E}_3$, ... — некоторая последовательность испытаний, и пусть T — некоторое событие, которое может произойти или не произойти при испытании $\mathscr{E}_i$. Обозначим через p_i вероятность события T, а через $q_i = 1 - p_i$ вероятность того, что оно не произойдет. Тогда испытания называются *независимыми* по отношению к событию T, если для любых непересекающихся подмножеств I, J испытаний вероятность того, что при $\mathscr{E}_i \in I$ событие T произойдет, а при $\mathscr{E}_j \in J$ не произойдет, равна $\prod_I p_i \prod_J q_j$.

Это условие равносильно следующему требованию: на множестве только таких испытаний, что событие T происходит при $\mathscr{E}_i \in I$ и не происходит при $\mathscr{E}_j \in J$, *условная вероятность* того, что при $\mathscr{E}_k (k \notin I \sqcup J)$ произойдет событие T, равна p_k.

Для определенности предположим, что вероятность ошибки при передаче единственного двоичного символа 0 или 1 (одного *бита* информации) равна $q = 1\% = 0.01$ и что мы хотим быть уверены в абсолютной точности передачи последовательностей из 10 000 символов. Тогда при прямой передаче последовательности символа за символом она будет правильно принята с ничтожной вероятностью:

$$P_0 = (1 - 0.01)^{10\,000} \simeq 10^{-4.4} < 0.004\%.$$

Этот числовой результат является частным случаем классической формулы Бернулли для независимых испытаний (см. любой учебник по теории вероятностей). Пусть

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!} = \frac{n\,(n-1)\ldots(n-k+1)}{1 \cdot 2 \cdot 3 \ldots k}.$$

Имеет место следующий важный факт.

Теорема 1. *Пусть по двоичному симметричному каналу передается n-битовая последовательность. Вероятность того, что она будет принята с ровно k ошибками, равна*

$$P_k = \binom{n}{k} p^{n-k} q^k = \binom{n}{k} p^{n-k} (1-p)^k = \binom{n}{k} (1-q)^{n-k} q^k. \quad (1)$$

Для больших k и n величина биномиального коэффициента $\binom{n}{k}$ довольно точно вычисляется по формуле Стирлинга. Соответствующие оценки приводят к закону больших чисел, утверждающему, что для больших n доля ошибочно воспринятых символов будет почти наверняка близка к q. Так, при $p = 0.99$ в последовательности из 10 000 двоичных символов около 100 символов будут приняты неверно; почти наверняка число ошибок будет лежать между 50 и 150 (здесь существенно предположение о независимости ошибок).

Общая схема передачи числовых данных представлена на рис. 8.2. Имеется информация, подлежащая передаче по каналу,

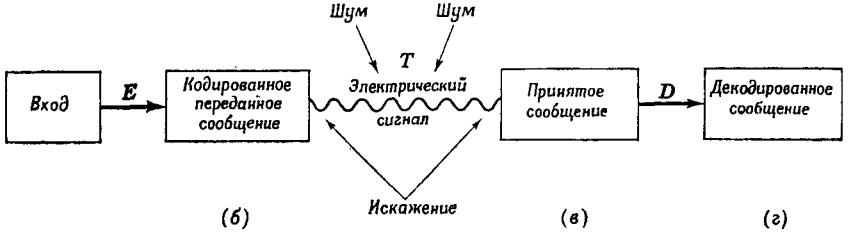


Рис. 8.2. Схема канала связи.

шум в канале, прием сигналов на фоне шума, перекодирование принятых сигналов с шумом в двоичную последовательность, некоторые знаки которой могут оказаться ошибочными из-за шума.

Общая задача, встающая перед проектировщиком такой системы связи, состоит в следующем. Из-за физических ограничений, наложенных на систему (максимальная мощность передатчика, затухание сигнала в канале связи и неустранимый шум), проектировщик не может надеяться на точное воспроизведение приемником переданного сигнала (хотя бы и с уменьшенной амплитудой). Если искажения велики, приемник будет делать ошибки. Проектировщик должен либо сознательно принять определенную долю ошибок, либо придумать способ уменьшить ее.

8.2. КОДИРОВАНИЕ И ДЕКОДИРОВАНИЕ

Во многих системах, обрабатывающих информацию, наподобие систем связи со спутниками и больших вычислительных систем, за ошибки приходится расплачиваться очень дорогой ценой.

Поэтому важно добиться, чтобы они были крайне маловероятны.

Эта глава в основном посвящена описанию эффективных методов увеличения надежности передачи информации с использованием *систематических кодов* разного типа. Большая часть из них принадлежит к классу *групповых кодов*, основанных на использовании теоремы Лагранжа (§ 7.10).

Идея, положенная в основу использования любого систематического кода, такова. Последовательности символов, подлежащие передаче, кодируются более длинными последовательностями тех же символов (обычно 0 и 1) по определенной схеме *кодирования*. Приемник способен распознавать и/или исправлять ошибки, вызванные шумом, анализируя дополнительную информацию, содержащуюся в добавочных символах. Принятая длинная последовательность декодируется по схеме *декодирования* в первоначально переданную, т. е. в последовательность до стадии кодирования.

Определение. Двоичным (m, n)-*кодом* называется пара, состоящая из схемы кодирования $E\colon 2^m \to 2^n$ и схемы декодирования $D\colon 2^n \to 2^m$, где 2^n — множество всех двоичных последовательностей длины n. Функции E и D выбираются так, чтобы функция $H = E \diamond T \diamond D$, где T — «функция ошибок», с вероятностью, близкой к единице, была тождественной. Поскольку кодирование и декодирование производятся в контролируемых условиях, можно считать, что они осуществляются безошибочно. Таким образом, математическую модель системы связи можно представить блок-схемой (рис. 8.3), где E, D — детерминированные функции.

Рис. 8.3. Математическая модель системы связи.

Коды делятся на два больших класса. *Коды с исправлением ошибок* имеют целью восстановить с вероятностью, близкой к единице, посланное сообщение. *Коды с обнаружением ошибок* имеют целью выявить с вероятностью, близкой к единице, наличие ошибок. Для иллюстрации приведем два примера.

Пример 1. Простой код с обнаружением ошибок основан на схеме *проверки четности*, применимой к сообщениям $\mathbf{a} = (a_1, \ldots, a_m) = a_1 \ldots a_m$ любой фиксированной длины m. Схема кодирования определяется так:

$$E\colon (a_1, \ldots, a_m) = a_1 \ldots a_m = \mathbf{a} \mapsto \mathbf{b} = b_1 \ldots b_{m+1},$$

где

$$b_i = a_i \text{ при } i = 1, \ldots, m, \tag{2}$$

$$b_{m+1} = \begin{cases} 0, \text{ если } \sum_{i=1}^{m} a_i \text{ четная,} \\ 1, \text{ если } \sum_{i=1}^{m} a_i \text{ нечетная.} \end{cases} \tag{2'}$$

Например, E при $m = 2$ определяется предписаниями: $00 \mapsto 000$, $01 \mapsto 011$, $10 \mapsto 101$, $11 \mapsto 110$. Из определений (2) и (2′) видно, что поразрядная сумма любой кодированной последовательности $\mathbf{b} = E\mathbf{a}$ должна быть четной.

Соответствующая схема декодирования такова: $D\colon \mathbf{b} \mapsto \mathbf{c}$, где

$$b_i = c_i \quad \text{при} \quad i = 1, \ldots, m. \tag{2''}$$

Если сумма $\sum_{i=1}^{m+1} b_i$ нечетна, приемник укажет *наличие ошибки передачи*. Разумеется, если сумма $\sum b_i$ четна, мы не можем быть уверены в том, что ошибка не произошла. Но, скажем, при $m = 2$ и при вероятности q ошибочного приема одного символа доля неверно принятых сообщений будет $q^3 + 3q^2p + 3qp^2$ (три, две или одна ошибка). Из них незамеченными при нашей схеме окажутся только ошибки точно в двух знаках, не изменяющие четности. Поэтому доля ошибочных сообщений, оставшихся незамеченными, относительно всех ошибочных сообщений будет

$$\frac{3q^2p}{q^3 + 3q^2p + 3qp^2} < \frac{q}{q+p} = q.$$

Значит, вероятность *пропуска* ошибки будет $< q$.

Теперь подробнее разберем пример кода с исправлением ошибок. Простейший пример такого кодирования состоит в повторении сигнала, хотя этот способ очень неэффективен (далек от оптимального).

Пример 2. Рассмотрим двоичный симметричный канал, типа описанного в § 8.1, для передачи строк двоичной информации. Иногда полезен следующий $(m, 3m)$-код с *тройным повторением*. Любое сообщение разбивается на блоки по m последовательных символов в каждом, и каждый блок передается трижды: это определяет функцию E. Функция D такова. Принятая строка разбивается на блоки длины $3m$. Если очередной блок состоит из трех одинаковых строк длины m, эта строка является результатом его декодирования. В общем случае по тройке символов c_i, c_{i+m}, c_{i+2m} в этом блоке восстанавливается символ, чаще всего (два или три раза) встречающийся в этой тройке, и ставится на i-е место в декодированном блоке.

Вероятность того, что символ в данной позиции будет принят трижды правильно, равна p^3. Вероятность ошибки только в первый раз равна p^2q, так что вероятность ровно одной ошибки есть $3p^2q$. Поэтому вероятность правильного приема символа в данной позиции равна p^3+3p^2q, а вероятность ошибочного приема равна $3pq^2+q^3$. Предположим, что $q=0.1$. Тогда в каждой позиции символ будет принят трижды правильно с вероятностью 0.729 и дважды правильно с вероятностью 0.243. Он будет принят дважды неправильно с вероятностью 0.027 и трижды неправильно с вероятностью 0.001. Таким образом, наш код снижает вероятность ошибки на один символ с 10% до $\approx 2.8\%$.

Аналогично пятикратная передача и декодирование по принципу «большинство голосов» даст вероятность ошибки $q^5+5q^4p+10q^3p^2=0.00856$, т. е. меньше 1%. В результате вероятность правильной передачи *строки длины* 10 возрастет с $(0.9)^{10}\simeq 35\%$ до $(0.972)^{10}\simeq 74\%$ при тройных повторениях и до $(0.99144)^{10}\simeq 91.5\%$ при пятикратных повторениях.

В заключение заметим, что тройное повторение обеспечивает исправление одной ошибки в каждой позиции за счет трехкратного удлинения времени передачи.

В примере 4 мы покажем, как исправление одной ошибки с той же надежностью достигается применением подходящего (3,6)-кода, который лишь удваивает время передачи.

УПРАЖНЕНИЯ А

1. Пусть двоичный симметричный канал (рис. 8.1) используется для передачи строк из двух символов. Показать, что таблица вероятностей приема имеет следующий вид:

Сигналы	00	01	10	11
00	p^2	pq	pq	q^2
01	pq	p^2	q^2	pq
10	pq	q^2	p^2	pq
11	q^2	pq	pq	p^2

2. Следующий массив называется *треугольником Паскаля:*

$$
\begin{array}{ccccccccccc}
 & & & & & 1 & & & & & \\
 & & & & 1 & & 1 & & & & \\
 & & & 1 & & 2 & & 1 & & & \\
 & & 1 & & 3 & & 3 & & 1 & & \\
 & 1 & & 4 & & 6 & & 4 & & 1 & \\
1 & & 5 & & 10 & & 10 & & 5 & & 1 \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot
\end{array}
$$

а) Добавить к нему еще три строки.
б) Объяснить его связь с формулой Бернулли.

3. Предположим, что по двоичному симметричному каналу передаются строки длины 14.

а) Какова вероятность того, что ровно пять символов будут приняты неправильно?

б) Какова вероятность того, что не больше пяти символов будут приняты неправильно?

в) Сколько имеется строк, отличающихся от данной не больше, чем в четырех позициях?

4. Пусть строка длины три передается по двоичному симметричному каналу, но вероятность ошибки меняется со временем, так что i-й символ принимается правильно с вероятностью p_i, $i = 1, 2, 3$.

а) Показать, что матрица $T = \| p_{hl} \|$ симметрична.

б) Показать, что T — стохастическая матрица, т. е. $\sum_l p_{hl} = 1$ для всех h, причем $0 \leqslant p_{hl} \leqslant 1$.

8.3. БЛОЧНЫЕ КОДЫ

Описанные выше примеры принадлежат к классу *блочных кодов*. По определению, блочный код заменяет каждый блок из m символов некоторым более длинным блоком из n символов, который после передачи подлежит декодированию. Мы будем ниже рассматривать только такие коды. (Существуют также «последовательные» коды, в которых символы сообщения чередуются с контрольными, а значение очередного символа зависит от всего предшествующего фрагмента сообщения).

Из соображений простоты и надежности большинство систем связи конструируется для передачи двоичных последовательностей. Блочный (m, n)-код, как мы уже говорили, определяется двумя функциями:

$$E\colon 2^m \longrightarrow 2^n, \quad D\colon 2^n \longrightarrow 2^m, \quad \mathbf{m} \leqslant \mathbf{n} \tag{3}$$

(случай $\mathbf{m} = \mathbf{n}$ используется в шифровании, где цель кодирования состоит в обеспечении секретности сигнала). Должно быть выполнено соотношение $E \diamond D = D \circ E = 1$, так чтобы сообщение было принято правильно при отсутствии помех.

В этой постановке одна из проблем оптимизации для кодов с исправлением ошибок в двоичных симметричных каналах может быть сформулирована так. При данных m и n найти такие D и E, чтобы вероятность приема сигнала с ошибкой была минимальной.

Расстояние между словами. К числу ключевых понятий теории кодирования принадлежит понятие расстояния между двоичными словами. Всякое двоичное слово длины n

$$\mathbf{a} = (a_1, \ldots, a_n)$$

можно рассматривать как вершину графа-*гиперкуба* всех двоичных слов длины n (см. § 2.8). *Расстояние* между двумя словами $\mathbf{a}$ и $\mathbf{b}$ тогда будет просто равно числу позиций, в которых $a_i \neq b_i$. Расстояние минимально и равно единице, если слова отличаются в точно одной позиции.

Можно сформулировать это определение иначе, отождествив двоичные слова длины n с элементарной абелевой группой периода 2 относительно покоординатного сложения по модулю 2.

Определение. *Весом* $w(\mathbf{a})$ слова $\mathbf{a}=(a_1, \ldots, a_n)$ называется число единиц среди его координат. *Расстоянием* $d(\mathbf{a}, \mathbf{b})$ между двоичными словами одинаковой длины называется вес их суммы, т. е. $d(\mathbf{a}, \mathbf{b})=w(\mathbf{a}+\mathbf{b})$.

Очевидно, это определение совпадает с предыдущим. Например, $w(0101)=2$, $w(1101)=3$, $d(1011, 1111)=1$, $d(0000, 0011)=2$, $d(1101, 1101)=0$. Заметим, что сдвиг $\mathbf{x}\mapsto\mathbf{x}+\mathbf{c}$ не меняет расстояний:

$$d(\mathbf{a}+\mathbf{c}, \mathbf{b}+\mathbf{c})=w(\mathbf{a}+\mathbf{b}+\mathbf{c}+\mathbf{c})=w(\mathbf{a}+\mathbf{b})=d(\mathbf{a}, \mathbf{b}).$$

Напомним, что в обозначениях § 8.2 вероятность k ошибок при приеме равна $\binom{n}{k}p^{n-k}q^k$, а вероятность $\leqslant l$ ошибок равна

$$p^n+\binom{n}{1}p^{n-1}q+\binom{n}{2}p^{n-2}q^2+\ldots+\binom{n}{l}p^{n-l}q^l. \qquad (4)$$

Воспользуемся функцией расстояния $d(\mathbf{b}, \mathbf{b}^*)$ для представления вероятностей ошибок. Вероятность того, что переданное слово $\mathbf{b}$ будет принято как $\mathbf{b}^*$, равна $p^{n-d(\mathbf{b}, \mathbf{b}^*)}q^{d(\mathbf{b}, \mathbf{b}^*)}$; например, 0011 будет принято как 1011 с вероятностью p^3q.

Заметим теперь, что для возможности *обнаружения* ошибки в одной позиции минимальное расстояние между кодовыми словами должно быть 2. Иначе ошибка в одной позиции может превратить одно кодовое слово в другое, и она не будет обнаружена.

Теорема 2. *Для того чтобы код давал возможность обнаруживать все ошибки в* $\leqslant k$ *позициях, необходимо и достаточно, чтобы наименьшее расстояние между двумя кодовыми словами было* $k+1$.

Для такого кода вероятность того, что ошибки в сообщении останутся необнаруженными, равна

$$q^n+\binom{n}{1}pq^{n-1}+\ldots+\binom{n}{k+1}p^{n-k-1}q^{k+1}. \qquad (5)$$

При малых q и умеренных k определяющим членом здесь является $\binom{n}{k+1}p^{n-k-1}q^{k+1}$.

Теорема 3. *Для того чтобы код давал возможность исправлять все ошибки в* $\leqslant k$ *позициях, необходимо и достаточно, чтобы наименьшее расстояние между двумя кодовыми словами было* $2k+1$.

Если это условие на E выполнено, то в качестве D следует взять функцию $a\mapsto$(ближайшее слово из образа E). Например,

для простого (1, 3)-кода E: $0\mapsto 000$, $1\mapsto 111$, а в качестве D служит функция

$$\begin{array}{ll} 000\mapsto 0 & 111\mapsto 1 \\ 001\mapsto 0 & 011\mapsto 1 \\ 010\mapsto 0 & 101\mapsto 1 \\ 100\mapsto 0 & 110\mapsto 1 \end{array}$$

Она исправляет ошибки в одной позиции.

Из теоремы 1 вытекает

Теорема 4. *Если код исправляет $\leqslant k$ ошибок, то для двоичного симметричного канала вероятность приема слова длины n, декодирование которого не совпадает с посланным сигналом, не превосходит*

$$\binom{n}{k+1} p^{n-k-1}q^{k+1} + \ldots + \binom{n}{1} pq^{n-1} + q^n. \tag{6}$$

Соответственно вероятность *правильного* приема не меньше, чем

$$p^n + \binom{n}{1} p^{n-1}q + \ldots + \binom{n}{k} p^{n-k}q^k. \tag{7}$$

Оценку можно улучшить, если $2k+1$ — это только минимальное расстояние между кодовыми словами. Однако не многие коды, основанные на чисто алгебраических процедурах (не табличные), работают заметно лучше.

С точки зрения групповой структуры удобно рассматривать строки ошибок. Данное сообщение $\mathbf{a} = a_1a_2\ldots a_m$ перекодируется в кодовое слово $\mathbf{b} = b_1b_2\ldots b_n$. Канал связи при передаче добавляет к нему *строку ошибок* $\mathbf{e} = e_1e_2\ldots e_n$, так что приемник принимает сигнал $\mathbf{r} = r_1r_2\ldots r_n$, где $r_i = b_i + e_i$. Система, *исправляющая* ошибки, переводит слово $r_1r_2\ldots r_n$ в ближайшее кодовое слово $b_1, b_2, \ldots, b_n$. Система, только *обнаруживающая* ошибки, смотрит лишь, является ли принятое слово кодовым, и сигнализирует о наличии ошибки, если это не так.

Пусть, например, передаваемое слово $\mathbf{a} = 01$ кодируется словом $\mathbf{b} = 0110$, а строка ошибок есть $\mathbf{e} = 0010$. Тогда будет принято слово $\mathbf{r} = 0100$. Система, *исправляющая* ошибки, переведет его в 0110 и затем восстановит переданное слово 01.

Если система только обнаруживает ошибки, то любая строка ошибок $\mathbf{e}$ с единственной единицей приведет к слову $\mathbf{b}^* = \mathbf{b} + \mathbf{e}$, которое не является кодовым. Например, рассмотрим (2,3)-код с проверкой четности:

$$E\colon 00\mapsto 000,\ 10\mapsto 101,\ 01\mapsto 011,\ 11\mapsto 110.$$

Множество кодовых слов есть 000, 011, 101, 110. Ни одна из строк ошибок 001, 010, 100, 111 не переводит ни одно кодовое слово

в кодовое слово. Поэтому однократная (а также тройная) ошибка будет обнаружена.

Код, обнаруживающий две ошибки, — это такой код, что ни одна строка ошибок с одной или двумя единицами не переводит одно кодовое слово в другое.

Пример 3. Следующий (2,5)-код обнаруживает две ошибки:

$$E: \quad \begin{array}{ll} 00 \mapsto 00000 = \mathbf{b}^1 & 01 \mapsto 01011 = \mathbf{b}^2 \\ 10 \mapsto 10101 = \mathbf{b}^3 & 11 \mapsto 11110 = \mathbf{b}^4 \end{array}$$

Эта же схема кодирования способна исправлять однократную ошибку, потому что любые два кодовых слова отличаются по меньшей мере в трех позициях. Из того, что $d(\mathbf{b}^i, \mathbf{b}^j) \geqslant 3$ при $\mathbf{b}^i \neq \mathbf{b}^j$, следует, что однократная ошибка приведет к приему слова, которое находится на расстоянии 1 от единственного кодового слова, которое и было передано.

Поэтому схема декодирования, состоящая в том, что принятое слово переводится в ближайшее к нему кодовое слово, будет исправлять однократную ошибку. В двоичном симметричном канале вероятность правильной передачи одного блока будет не меньше чем p^5+5p^4q.

Кодовое слово длины n будет передано правильно с вероятностью p^n. Ошибки в данных k позициях будут сделаны с вероятностью $p^{n-k}q^k$. При $k>0$ легко видеть, что вероятность ошибки в данных позициях уменьшается с ростом числа позиций. Это объясняет с вероятностной точки зрения, почему принятое слово лучше всего переводить в ближайшее к нему кодовое слово: вероятность того, что было передано более далекое слово, меньше.

УПРАЖНЕНИЯ Б

1. Доказать, что если расстояние между кодовыми словами равно 7, то код способен обнаруживать до шести ошибок и исправлять до трех ошибок.

2. Рассмотрим (8,9)-код с проверкой на четность. Какова вероятность того, что не будет обнаружена ошибка при передаче блока длины 9?

3. Рассмотрим (4,5)-код с проверкой на четность и (4,12)-код с троекратной передачей. При $p=0.9$, $q=0.1$ вычислить вероятности того, что ошибочно переданное слово не будет обнаружено.

4. Рассмотрим такой (4,8)-код, что минимальное расстояние между кодовыми словами равно 4. Он способен исправлять однократную ошибку и обнаруживать тройную ошибку. Вычислить вероятности пропуска ошибок.

8.4. МЕТОДИКА МАТРИЧНОГО КОДИРОВАНИЯ

Ранее мы описывали каждую схему кодирования таблицами, задающими кодовое слово для каждого блока: $\mathbf{b}=E[\mathbf{a}]$. Для блоков большой длины этот способ требует большого объема памяти и неэкономичен.

Гораздо меньшего объема памяти требует разработанная методика *матричного кодирования*. Пусть E — некоторая $m \times n$-матрица, где e_{ij} — элемент на пересечении i-й строки и j-го столбца — равен 0 или 1. Обозначая знаком $+$ сложение по модулю 2, мы можем определить схему кодирования уравнениями

$$b_j = a_1 e_{1j} + a_2 e_{2j} + \ldots + a_m e_{mj} = \sum_{i=1}^{m} a_i e_{ij}, \quad j = 1, \ldots, n. \qquad (8)$$

Она отвечает умножению строки на кодирующую матрицу E справа и определяет (m, n)-код.

Пример 4. Рассмотрим следующую 3×6-матрицу:

$$E = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

Если $\mathbf{a} = 100$, то (8) превращается в $b_j = e_{1j}$. Таким образом, слова $\mathbf{a}^1 = 100$, $\mathbf{a}^2 = 010$ и $\mathbf{a}^4 = 001$ кодируются соответственно первой, второй и третьей строками матрицы E. Полный список кодирования таков:

$$\mathbf{a}^0 = 000 \mapsto 000000$$
$$\mathbf{a}^1 = 100 \mapsto 100110$$
$$\mathbf{a}^2 = 010 \mapsto 010011$$
$$\mathbf{a}^3 = 110 \mapsto 110101$$
$$\mathbf{a}^4 = 001 \mapsto 001111$$
$$\mathbf{a}^5 = 101 \mapsto 101001$$
$$\mathbf{a}^6 = 011 \mapsto 011100$$
$$\mathbf{a}^7 = 111 \mapsto 111010$$

Этот список показывает преимущество матричного кодирования: достаточно запомнить m кодовых слов вместо 2^m слов. Ясно, что это общий факт.

Код не должен приписывать одно и то же кодовое слово разным словам. Простой способ добиться этого состоит в том, чтобы первые m столбцов матрицы E образовывали единичную матрицу. Мы опускаем несложное доказательство.

Групповая операция. Напомним (пример 13, § 7.7), что множество всех двоичных слов $\mathbf{a} = a_1 a_2 \ldots a_m$ образует *абелеву группу* $\mathbf{Z}_2^m$ относительно покоординатного сложения по модулю 2. Это относится и к кодовым словам.

Предположим, что $\mathbf{a}=\mathbf{a}'+\mathbf{a}''$. Тогда для $\mathbf{b}=\mathbf{a}E$, $\mathbf{b}'=\mathbf{a}'E$, $\mathbf{b}''=\mathbf{a}''E$, очевидно, имеем

$$b_j \equiv \sum_{i=1}^{m} a_i e_{ij} \equiv \sum_{i=1}^{m} (a_i' + a_i'') e_{ij} \equiv$$
$$\equiv \sum_{i=1}^{m} a_i' e_{ij} + \sum_{i=1}^{m} a_i'' e_{ij} \equiv b_j' + b_j''.$$

Таким образом,

$$E[\mathbf{a}] = \mathbf{b} = \mathbf{b}' + \mathbf{b}'' = E[\mathbf{a}'] + E[\mathbf{a}'']. \qquad (9)$$

В терминологии гл. 7 имеет место

Теорема 5. *Пусть E — кодирующая $m \times n$-матрица, у которой есть единичная $m \times m$-подматрица. Тогда отображение $\mathbf{a} \mapsto \mathbf{a}E = E[\mathbf{a}]$ является мономорфизмом группы сообщений в группу кодирующих слов.*

8.5. ГРУППОВЫЕ КОДЫ

Теорема 5 показывает, что матричные коды являются групповыми кодами в смысле следующего определения.

Определение. Блочный код называется *групповым*, если его кодовые слова образуют группу.

Следующий результат позволяет определить способность таких кодов к обнаружению и исправлению ошибок.

Теорема 6. *Если код является групповым, то наименьшее расстояние между двумя кодовыми словами равно наименьшему весу ненулевого кодового слова.*

Доказательство. Это очевидно из соотношения $d(\mathbf{b}^i, \mathbf{b}^j) = w(\mathbf{b}^i + \mathbf{b}^j)$.

Наименьший вес ненулевого кодового слова в примере 4 равен 3. Следовательно, минимальное расстояние тоже равно 3, и код способен исправлять однократную ошибку и обнаруживать двойную.

Легко определить, какие ошибки останутся незамеченными; для групповых кодов они отвечают в точности тем строкам ошибок, которые сами являются кодовыми словами. Так, в примере 4 строка ошибок $\mathbf{e}=100110$ переводит любое кодовое слово в кодовое слово.

Вероятность того, что ошибка останется необнаруженной, равна сумме вероятностей всех строк ошибок, то есть кодовых слов. В примере 4 это $4p^3q^3 + 3p^2q^4$.

Оптимальное декодирование. Рассмотрим теперь задачу оптимизации декодирования группового кода E: $\mathbf{a} \mapsto \mathbf{a}E$

с двоичной матрицей кодирования E. Мы хотим минимизировать вероятность того, что $D[\mathbf{a}E] \neq \mathbf{a}$. Мы будем предполагать, что вероятности передачи любого сигнала одинаковы и что канал — двоичный симметричный.

Схема декодирования исходит из таблицы $\mathbf{C}$ всех слов, которые могут быть приняты. Так как кодовые слова образуют подгруппу $\mathbf{B} \subset \mathbf{C}$, мы можем расположить $\mathbf{C}$, записав в одну строку элементы смежного класса по модулю $\mathbf{B}$. Первая строка отвечает нулевому классу:

$$\mathbf{0},\ \mathbf{b}^1,\ \mathbf{b}^2,\ \ldots,\ \mathbf{b}^{2^m-1}.$$

Если $\mathbf{c}^i \in \mathbf{C}$, $\mathbf{c}^i \notin \mathbf{B}$, то строка, содержащая слово $\mathbf{c}^i$, имеет вид

$$\mathbf{0}+\mathbf{c}^i,\ \mathbf{b}^1+\mathbf{c}^i,\ \mathbf{b}^2+\mathbf{c}^i,\ \ldots,\ \mathbf{b}^{2^m-1}+\mathbf{c}^i.$$

Лидером этого класса называется слово наименьшего веса. Можно считать, что он записан первым. По теореме Лагранжа, если строки такой таблицы попарно не совпадают, то они попарно не пересекаются. Таким образом, каждый элемент из $\mathbf{C}$ однозначно представляется в виде суммы $\mathbf{c}^i+\mathbf{b}^j$, где $\mathbf{c}^i$ — лидер соответствующего смежного класса и $\mathbf{b}^j \in \mathbf{B}$. Окончательная схема декодирования имеет вид

$\mathbf{0}$	$\mathbf{b}^1$	$\mathbf{b}^2$	$\ldots$	$\mathbf{b}^{2^m-1}$
$\mathbf{c}^1$	$\mathbf{b}^1+\mathbf{c}^1$	$\mathbf{b}^2+\mathbf{c}^1$	$\ldots$	$\mathbf{b}^{2^m-1}+\mathbf{c}^1$
$\mathbf{c}^2$	$\mathbf{b}^1+\mathbf{c}^2$	$\mathbf{b}^2+\mathbf{c}^2$	$\ldots$	$\mathbf{b}^{2^m-1}+\mathbf{c}^2$
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
$\mathbf{c}^{2^{n-m}-1}$	$\mathbf{b}^1+\mathbf{c}^{2^{n-m}-1}$	$\mathbf{b}^2+\mathbf{c}^{2^{n-m}-1}$	$\ldots$	$\mathbf{b}^{2^m-1}+\mathbf{c}^{2^{n-m}-1}$

Декодирование слова $\mathbf{c}=\mathbf{b}^i+\mathbf{c}^j$ состоит в выборе кодового слова $\mathbf{b}^i$ в качестве переданного и последующем применении операции E^{-1}.

Заметим, что в ходе доказательства теоремы Лагранжа был установлен более общий факт.

Лемма 1. *В любом групповом коде, не обязательно двоичном, любой элемент* $\mathbf{c} \in \mathbf{C}$ *однозначно представляется суммой* $\mathbf{c}=\mathbf{e}^i+\mathbf{b}$ *кодового слова* $\mathbf{b} \in \mathbf{B}$ *и лидера* $\mathbf{e}^i$. (Предполагается, что лидеры в каждом классе выбраны каким-нибудь способом заранее.)

8.6. ТАБЛИЦЫ ДЕКОДИРОВАНИЯ

Вышеизложенный метод удобно реализовать с помощью специальной таблицы декодирования. Для (3,6)-кода из примера 4 она выглядит так:

Строка кодовых слов ⟶								
	000000	100110	010011	011100	001111	101001	110101	111010
	100000	000110	110011	111100	101111	001001	010101	011010
	010000	110110	000011	001100	011111	111001	100101	101010
	001000	101110	011011	010100	000111	100001	111101	110010
	000100	100010	010111	011000	001011	101101	110001	111110
	000010	100100	010001	011110	001101	111011	110110	111000
	000001	100111	010000	011101	001110	101000	110100	111011
	000101	100011	010110	011001	001010	101100	110000	111111
	↑ *Столбец лидеров*							

Чтобы *декодировать* принятое слово $\mathbf{b}^j+\mathbf{e}$, следует отыскать его в таблице и выбрать в качестве переданного кодовое слово в том же столбце и в первой строке.

Например, если принято слово 110011, считается, что было передано слово 010011; если принято 100101, переданным считается 110101; если принято 110101, считается, что оно и было передано, и т. п.

Какие строки ошибок такой код может исправлять? Очевидно, в точности лидеры смежных классов. Строка ошибок 000001 обязательно будет исправлена, тогда как строка ошибок 010001 после декодирования приведет не к тому сигналу, который был передан. Таким образом, наш код исправляет все одинарные ошибки, а также двойную ошибку 000101 — единственный лидер веса $w>1$. Имеет место общий результат.

Теорема 7. *Групповое кодирование со схемой декодирования посредством лидеров исправляет в точности те строки ошибок, которые являются лидерами.*

Вероятность правильного декодирования переданного по двоичному симметричному каналу слова равна сумме вероятностей всех лидеров (включая нулевой). В примере 4 это будет (для слов длины 6) $p^6+6p^5q+p^4q^2$.

Качество этой схемы декодирования оценивает

Теорема 8. *Предположим, что лидеры являются словами наименьшего веса в своем классе. Тогда кодовое слово, стоящее в данном столбце, является ближайшим кодовым словом ко всем словам этого столбца.*

Иными словами, описанный метод обеспечивает выбор самого близкого кодового слова к принятому. Напомним, что такой выбор оптимален, ибо минимизирует вероятность ошибки при передаче по двоичному симметричному каналу.

Доказательство. Пусть мы передали слово $\mathbf{b}^i$ и приняли $\mathbf{b}^i+\mathbf{e}$. Расстояние от $\mathbf{b}^i+\mathbf{e}$ до $\mathbf{b}^i$ равно $w(\mathbf{e})$. Расстояние от $\mathbf{b}^i$ до любого другого кодового слова равно весу разности

этих слов, которая лежит в том же смежном классе, что и $\mathbf{b}^l$. Поэтому этот вес не меньше веса лидера.

Геометрическая интерпретация. Опишем наши объекты в геометрических терминах. *Кодирующая* схема двоичного (m,n)-кода определяет инъекцию множества сигналов 2^m длины m в *гиперкуб* $\mathbf{C}=2^n$. В графе, вершины которого отвечают вершинам куба, а ребра — его ребрам, расстояние между вершинами совпадает с расстоянием между соответствующими словами.

Схема *декодирования* D отображает 2^n назад на 2^m, так что прообраз каждого слова длины m состоит в точности из тех слов, которые будут декодированы данным словом. Чтобы кодирование исправляло все ошибки веса $\leqslant w$, каждый такой класс эквивалентности должен содержать «шар» радиуса w с центром E[a], состоящий из всех слов (вершин), расстояние которых от кодового слова не превосходит w. Таких слов имеется $1+n+\binom{n}{2}+\ldots+\binom{n}{w}$. Код называется *совершенным*, если все классы эквивалентности являются такими шарами. Совершенные коды строить очень трудно.

Чтобы блочный (m,n)-код был совершенным, необходимо выполнение условия $1+n+\ldots+\binom{n}{w}=2^{n-m}$. Целых чисел w $\left(1<w<\frac{n-1}{2}\right)$, для которых $1+n+\ldots+\binom{n}{w}$ является степенью двойки, немного. Общий метод построения *оптимальных* двоичных блочных (m,n)-кодов, минимизирующих вероятность ошибочного декодирования, неизвестен. Задача сводится к проблеме «плотнейшей» упаковки «сфер» радиуса w в n-мерном кубе.

Пример 5. На рис. 8.4 показан двоичный групповой код, его кодирующая матрица и стандартный массив (таблица). Если α_i — число лидеров веса i, то вероятность правильного декодирования равна $P=\sum\alpha_i p^{n-i}q^i$.

Для нашего кода $\alpha_0=1$, $\alpha_1=6$, $\alpha_2=1$, остальные α_i нулевые. Код является почти совершенным: он исправляет все одинарные ошибки и одну двойную, и лучшего (3,6)-кода не существует.

УПРАЖНЕНИЯ В

1. Рассмотрим следующую кодирующую матрицу:

$$G=\begin{bmatrix}1&0&0&0&0&1&1&1\\0&1&0&0&1&0&1&1\\0&0&1&0&1&1&0&1\\0&0&0&1&1&1&1&0\end{bmatrix}.$$

СТАНДАРТНЫЙ МАССИВ

		Лидеры ↓						
Кодовые слова	$l_0 = 000000$	100110	010011	001101	110101	101011	011110	111000
	$l_1 = 000001$	100111	010010	001100	110100	101010	011111	111001
	$l_2 = 000010$	100100	010001	001111	110111	101001	011100	111010
	$l_3 = 000100$	100010	010111	001001	110001	101111	011010	111100
	$l_4 = 001000$	101110	011011	000101	111101	100011	010110	110000
	$l_5 = 010000$	110110	000011	011101	100101	111011	001110	101000
	$l_6 = 100000$	000110	110011	101101	010101	001011	111110	011000
	$l_7 = 100001$	000111	110010	101100	010100	001010	111111	011001

Кодирующая матрица $E = \begin{bmatrix} 100110 \\ 010011 \\ 001101 \end{bmatrix}$ $\qquad \mathbf{a}^i E = \mathbf{b}^i$

Матрица проверки четности $F = \begin{bmatrix} 101100 \\ 110010 \\ 011001 \end{bmatrix}$, $\qquad \mathbf{b}^i F^T = 0 \qquad EF^T = 0$

Рис. 8.4. (3,6)-код и таблица декодирования к примеру 5.

Каждое кодовое слово состоит из четырех символов сообщения и четырех контрольных символов. Пусть это используется как код, обнаруживающий ошибки двоичного симметричного канала. Какова вероятность ошибочного приема?

2. Рассмотрим кодирующую матрицу, которая получается из матрицы G упр. 1 вычеркиванием нижней строки и четвертого столбца. Кодовое слово состоит из трех символов сообщения и четырех контрольных символов.

а) Записать этот код в стандартной табличной форме. Предполагается, что используются все возможности для исправления ошибок (можно исправлять не только ошибки в одной позиции).

б) Какова вероятность ошибочного декодирования?

в) Если код используется лишь для обнаружения ошибок, какова вероятность того, что ошибка пройдет незамеченной?

3. а) Доказать, что любой сдвиг $\mathbf{x} \mapsto \mathbf{x}+\mathbf{c}$ в $[\mathbf{Z}_2^l, \cdot]$ сохраняет расстояние между словами.

б) Доказать, что минимальное расстояние от данного кодового слова до остальных кодовых слов не зависит от выбора этого слова в двоичном групповом коде.

Через Δ_l обозначено некоторое множество вершин в n-кубе, любые две из которых отстоят друг от друга не меньше чем на l.

4. Доказать, что при $n=3$ множество Δ_3 состоит не больше чем из двух элементов.

5. Доказать, что Δ_5 может состоять не больше чем из $2^{n+1}/(n^2+n+2)$ элементов. $\left[\textit{Указание}: \binom{n}{2}+\binom{n}{1}+\binom{n}{0}=(n^2+n+2)/2.\right]$

Код C называется эквивалентным коду C^*, если существует биекция между кодовыми словами C и C^*, сохраняющая попарные расстояния.

6. Пусть G — некоторая кодирующая матрица. Показать, что следующие операции над ней приводят к эквивалентному коду:

а) перестановка строк G,

б) перестановка столбцов G,

в) замена одной строки ее суммой с другой строкой.

7. Пусть I_m — квадратная единичная матрица, G — некоторая $m\times l$-матрица. Кодирующая матрица вида $[I_m G]$ называется *стандартной*. Такова, например, матрица G в упр. 1. Построить стандартную матрицу G^*, порождающую код, эквивалентный коду с матрицей

$$G=\begin{bmatrix}0&0&0&1&1&1\\0&1&1&0&1&1\\1&1&0&0&0&1\end{bmatrix}.$$

8. а) Пусть G — некоторая кодирующая матрица. Показать, что существует *матрица проверки четности* H со следующими свойствами:

(i) если размер G равен $m\times n$, то размер H равен $n\times(n-m)$,

(ii) $GH=0$,

(iii) пусть h_i, $1\leqslant i\leqslant n-m$, — столбцы матрицы H.

Тогда они линейно независимы, т. е. $\sum_{i=1}^{n-m} a_i h_i \neq 0$ для любых констант a_i, не все из которых нулевые. Матрица H называется *матрицей проверки четности*, ибо для всех кодовых слов $\mathbf{c}$ матрицы G имеем $\mathbf{c}H=0$.

б) Пусть дана матрица проверки четности. Показать, что минимальный вес ненулевого кодового слова равен наименьшему числу строк H, которые в сумме дают нуль.

9. Рассмотреть код, обнаруживающий ошибки, с матрицей

$$G=\begin{bmatrix}1&0&0&0&1&1\\0&1&0&1&0&1\\0&0&1&1&1&1\end{bmatrix}.$$

Какова вероятность того, что слово длины шесть будет принято как правильное, тогда как на самом деле остались необнаруженными ошибки?

10. Пусть $A_{m,\,m+k}=I_mA_{m,\,k}$ — матрица кода с проверкой на четность, исправляющего r ошибок. Показать, что матрица проверки четности

$$\left[\begin{array}{c|c|c}I_{m,\,m}&A_{m,\,k}&0_{m,\,1}\\\hline 1_{1,\,m+k}&\multicolumn{2}{c}{1}\end{array}\right]$$

доставляет код, позволяющий исправить r ошибок и обнаружить $r+1$ ошибок. Здесь I_m — единичная матрица размера m.

11. Рассмотрим двоичный групповой код с матрицей

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

а) Найти стандартную матрицу G, дающую эквивалентный код (см. упр. 7).

б) Найти матрицу проверки четности для кода п. а.

в) Найти кодовое слово с контрольными символами 110. Показать, что оно лежит в пространстве строк G и в нуль-пространстве H.

*12. Пусть в системе связи k-битовые слова кодируются n-битовыми. Пусть C_0, C_1, ..., C_q ($q=2^k-1$) — кодовые слова. Декодирующая таблица имеет вид

$$\begin{array}{cccc}C_0 & C_1 & \dots & C_q\\ C_0+S_1 & C_1+S_1 & \dots & C_q+S_1\\ \dots & \dots & \dots & \dots\\ C_0+S_r & C_1+S_r & \dots & C_q+S_r\end{array}$$

Здесь $r=2^{n-k}-1$, $C_0=00\dots0$, C_0+S_i имеет наименьший вес в i-й строке. Построить такую таблицу для кода из упр. 11.

13. Пусть H — матрица проверки четности для некоторого кода. Слово rH называется *синдромом* слова r. Показать, что имеется естественная биекция между синдромами и лидерами смежных классов.

8.7. КОДЫ ХЭММИНГА

Изучив способ декодирования групповых кодов, опишем класс совершенных кодов, остроумная конструкция которого принадлежит Хэммингу.

Минимальное расстояние между кодовыми словами в коде Хэмминга равно 3, так что они исправляют одинарную ошибку. Коды Хэмминга *совершенны*: существует $(m=2^r-1-r,\ n=2^r-1)$-код для любого r, способный исправлять ошибку ровно в одной позиции и больше никаких ошибок; сверх того, никакой другой $(2^r-1-r,\ 2^r-1)$-код не может исправлять все одинарные ошибки и еще какие-нибудь ошибки. Это означает, что у деко-

дирующей таблицы для такого кода лидерами будут **0** и все 2^r-1 строк ошибок ровно с одной единицей. Вообще совершенный код, исправляющий все ошибки в не более чем k позициях, должен иметь в качестве лидеров все строки с $\leqslant k$ единицами; никакой другой код такого типа не может быть лучше совершенного. (Код называется «квазисовершенным», если, кроме того, он способен исправлять некоторые ошибки в $k+1$ позициях, но не больше.)

Декодирующая схема, обнаруживающая ошибку в единственной позиции, для кодов Хэмминга также проста. В классе кодов Хэмминга имеются коды любой длины, и способ кодирования и декодирования для них в основном тот же. Но мы ограничимся описанием совершенных кодов; остальные строятся аналогично, и сведения о них включены в упражнения.

Процедура построения кода Хэмминга такова.

1. Выберем целое положительное число r. Сообщения будут словами длины 2^r-1-r, а кодовые слова — длины 2^r-1.

2. В каждом кодовом слове $\mathbf{b}=b_1, b_2, \ldots, b_{2^r-1}$ символы $b_{2^0}, b_{2^1}, b_{2^2}, \ldots, b_{2^{r-1}}$ являются контрольными, остальные — в естественном порядке — символами сообщения. Например, при $r=4$ символы b_1, b_2, b_4, b_8 — контрольные, а $b_3, b_5, b_6, b_7, b_9, b_{10}, b_{11}, b_{12}, b_{13}, b_{14}, b_{15}$ — символы сообщения.

3. Построим матрицу M из 2^r-1 строк и r столбцов. В i-м столбце стоят символы двоичного разложения числа i. Матрицы для $r=2$, 3 и 4 таковы:

$$M_{3,2}=\begin{bmatrix}01\\10\\11\end{bmatrix} \qquad M_{7,3}=\begin{bmatrix}001\\010\\011\\100\\101\\110\\111\end{bmatrix} \qquad M_{15,4}=\begin{bmatrix}0001\\0010\\0011\\0100\\0101\\0110\\0111\\1000\\1001\\1010\\1011\\1100\\1101\\1110\\1111\end{bmatrix}.$$

4. Запишем систему уравнений в $\mathbf{b}M=0$, где M — матрица из п. 3. Она состоит из r уравнений: например, для $r=3$:

$$b_4+b_5+b_6+b_7=0,$$
$$b_2+b_3+b_6+b_7=0,$$
$$b_1+b_3+b_5+b_7=0.$$

5. Чтобы закодировать сигнал, возьмем в качестве b_j для $j\neq 2^i$ соответствующие символы и отыщем b_{2^i} из написанной системы уравнений. Так как в каждое уравнение входит ровно одно b_{2^i}, то это сделать нетрудно. В выписанной системе b_4 входит в первое уравнение, b_2 — во второе, b_1 — в третье.

Наименьший ненулевой вес кодового слова в нашем примере равен 3.

Декодирование несложно. Пусть принято слово $\mathbf{b}+\mathbf{e}$, где $\mathbf{b}$ — переданное слово. Так как $\mathbf{b}M=\mathbf{0}$, то $(\mathbf{b}+\mathbf{e})M=\mathbf{b}M+\mathbf{e}M=\mathbf{e}M$. Если результат нулевой, как происходит при правильном приеме, считается, что ошибок не было. Если вектор ошибок имеет единицу в i-й позиции, $\mathbf{e}=00\ldots1\ldots00$, то при его умножении на M получится i-я строка матрицы M, т. е. двоичное разложение числа i. Тогда следует изменить символ в i-й позиции слова $\mathbf{b}+\mathbf{e}$.

Пример 6. (4,7)-код Хэмминга имеет в качестве одного из кодовых слов $\mathbf{b}=0001111$. Матрица M имеет вид

$$\begin{bmatrix}001\\010\\011\\100\\101\\110\\111\end{bmatrix}$$

Имеем $\mathbf{b}M=000$. Добавим к $\mathbf{b}$ строку ошибок $\mathbf{e}=0010000$. Тогда $\mathbf{b}+\mathbf{e}=0011111$ и $(\mathbf{b}+\mathbf{e})M=011$. Это — двоичное представление числа 3, так что ошибка находится в третьей позиции. Если $\mathbf{e}=0000001$, то $(\mathbf{b}+\mathbf{e})M=111$, т. е. 7 в двоичной записи: ошибка находится в седьмой позиции.

Если ошибка допущена больше чем в одной позиции, декодирование даст неверный результат. Например, если строка ошибок будет кодовой, то $(\mathbf{b}+\mathbf{e})M=\mathbf{0}$, и декодирование не изменит принятого слова. Если ошибки допущены в двух позициях, код все равно укажет одну позицию, и притом неправильно.

К этому коду можно добавить проверку четности. Получится код наименьшего веса 4, способный исправлять ошибку в одной позиции и обнаруживать ошибки в двух позициях.

Опишем этот (4,7)-код геометрически. Для этого рассмотрим 7-мерный *гиперкуб* $\mathbf{C}=2^7$, вершины которого отвечают двоичным словам длины 7. У него имеется 128 вершин и $7\times 64=448$ ребер. Пусть L — множество лидеров, состоящее из $\mathbf{0}$ и семи его ближайших соседей. Векторы ошибок $\mathbf{e}\in\mathbf{C}$ имеют по единице в одной позиции. Групповые сдвиги $L+\mathbf{b}$ на кодовые слова $\mathbf{b}\in\mathbf{B}$ образуют 16 непересекающихся «уголков», каждый из которых состоит из некоторого кодового слова и ближайших его соседей.

УПРАЖНЕНИЯ Г

1. **Матрица проверки** четности для (4,7)-кода Хэмминга имеет вид

$$M^T=\begin{bmatrix}0&0&0&1&1&1&1\\0&1&1&0&0&1&1\\1&0&1&0&1&0&1\end{bmatrix}.$$

Она приводит к системе уравнений:

$$\begin{aligned}b_1&=a_1+a_2+a_4, & b_4&=a_1, & b_7&=a_4.\\ b_2&=a_1+a_3+a_4, & b_5&=a_2,\\ b_3&=a_2+a_3+a_4, & b_6&=a_3,\end{aligned}$$

Слово $\mathbf{c}=\mathbf{b}+\mathbf{e}$, принятое с одной ошибкой, декодируется так. Вычислим $\mathbf{c}M=\mathbf{s}$. Слово $\mathbf{s}$ есть *синдром* $\mathbf{c}$. Число $k=\sum_{i=1}^{3}s_i2^{3-i}$ — номер позиции, где была допущена ошибка. Распространить эту процедуру на расширенный код, исправляющий одинарную ошибку и обнаруживающий двойную. Построить затем таблицу смежных классов и описать их биекцию с синдромами.

2. Пусть H — матрица проверки четности для двоичного группового кода. Показать, что смежный класс с синдромом $\mathbf{v}$ содержит вектор веса w тогда и только тогда, когда некоторая линейная комбинация w столбцов H равна $\mathbf{v}$.

3. Показать, что кодовые слова двоичного группового кода либо все имеют четный вес, либо половина — четный, а половина — нечетный. (*Указание*: установить, что кодовые слова четного веса образуют подгруппу.)

4. Построить (10,14)-код Хэмминга с исправлением одинарных ошибок, дав таблицу кодовых слов. Предположим, что возможности исправления ошибок в системе используются полностью. Какова вероятность правильного приема (при двоичном симметричном канале)?

5. Рассмотрим (3,6)-код с обнаружением ошибок, имеющий матрицу

$$G=\begin{bmatrix}1&0&0&0&1&1\\0&1&0&1&0&1\\0&0&1&1&1&1\end{bmatrix}.$$

Какова вероятность того, что ошибка в передаче слова длины 6 не будет обнаружена?

6. Код Хэмминга описывается следующими матрицами:

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} \qquad \begin{bmatrix} 001 \\ 010 \\ 011 \\ 100 \\ 101 \\ 110 \\ 111 \end{bmatrix}$$

а) Кодирующая матрица б) Матрица проверки четности

Пусть при приеме следующих слов была допущена ошибка в одной позиции:

а) 0 1 1 1 1 1 0,
б) 0 0 0 1 1 1 1.

Какие слова были переданы?

7. Показать, что если $n > 1$, то $(n, 3n)$-код с декодированием «по большинству голосов» (пример 2) не может быть совершенным.

8. Показать, что двоичный групповой код с минимальным расстоянием 7 способен обнаруживать $\leqslant 6$ ошибок и исправлять $\leqslant 3$ ошибок.

Упражнения 9—11 *относятся к двоичному групповому коду с матрицей*

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

9. Задать таблицей все сообщения и их коды.

10. Построить вентильную схему, дающую на выходе символы проверки четности, когда на вход подается слово длины 5.

11. Вычислить вероятности обнаружения и исправления ошибок для этого кода.

СПИСОК ЛИТЕРАТУРЫ

1. Abramson N., Information Theory and Coding, McGraw-Hill, 1965.
2. Berlekamp E. R., Algebraic Coding Theory, McGraw-Hill, 1968. (Русский перевод: Берлекэмп Э. Алгебраическая теория кодирования, М., «Мир», 1971.)
3. Fano R., The Transmission of Information, MIT Press, 1963. (Русский перевод: Фано Р., Передача информации. Статистическая теория связи, М., «Мир», 1965.)
4. Peterson W. W., Error-correcting Codes, MIT Press 1961. (Русский перевод: Питерсон У., Коды, исправляющие ошибки, М., «Мир», 1964.)

РЕШЕТКИ

9.1. РЕШЕТКИ И ЧАСТИЧНО УПОРЯДОЧЕННЫЕ МНОЖЕСТВА

Понятия частичного порядка и частично упорядоченного множества были определены в § 2.4 и там же были описаны некоторые их основные свойства. В § 5.2 мы заметили, что любая булева алгебра естественно частично упорядочена отношением $a \leqslant b$, которое означает, что $a \wedge b = a$, или эквивалентно, $a \vee b = b$. Относительно этого частичного порядка

$$a \wedge b = \text{н. н. г. } \{a, b\}, \quad a \vee b = \text{н. в. г. } \{a, b\}.$$

Эти результаты следуют только из аксиом L1—L4 § 5.1; аксиомы L5—L10 не используются. Мы назовем *решеткой* любую алгебраическую систему $L = [L, \wedge, \vee]$, бинарные операции которой удовлетворяют аксиомам L1—L4. Таким образом, *булевы алгебры являются решетками* относительно бинарных операций $\wedge$ и $\vee$. Прежде чем продолжать чтение, читателю рекомендуется заново просмотреть указанные параграфы.

В этой главе мы изучим как общие свойства решеток, так и свойства их некоторых специальных классов (в том числе булевых алгебр). Решетки естественно появляются во многих областях математики. Следующие два примера являются типичными.

Пример 1. Пусть $V = \mathbf{R}^X$ — множество всех вещественнозначных функций f, g, h, ... с областью X. Определим отношение $f \leqslant g$ следующим образом: $f(x) \leqslant g(x)$ для всех $x \in X$. Эквивалентное определение в терминах н. н. г. и н. в. г.: $f \wedge g = h$ означает, что $h(x) = \min \{f(x), g(x)\}$ для всех $x \in X$, а $f \vee g = j$ означает, что $j(x) = \max \{f(x), g(x)\}$ для всех $x \in X$.

Пример 2. Пусть G — любая группа. Обозначим через $L(G)$ множество всех подгрупп S, T, ... группы G. По определению $S \leqslant T$ означает, что $S \subset T$. Очевидно, $L(G)$ является частично упорядоченным множеством. Ясно, кроме того, что пересечение $S \cap T$ двух подгрупп S и T является подгруппой, которая содержит любую подгруппу, содержащуюся одновременно в S и T.

Поэтому $S \cap T = \text{н.н.г.}\{S, T\} = S \wedge T$ в частично упорядоченном множестве $L(G)$. С другой стороны, множество $\{s_1 t_1 s_2 t_2 \ldots\}$ всевозможных произведений элементов $s_i \in S$ и $t_i \in T$ является подгруппой группы G, содержащейся в любой подгруппе G, которая содержит S и T. Поэтому оно совпадает с $S \vee T = \text{н. в. г.}\{S, T\}$. Таким образом, частично упорядоченное множество $L(G)$ является решеткой.

Другие образцы решеток даны в примерах 3, 4 и 5 § 2.4. Позже мы пополним их список, но сначала напомним следующий принцип из гл. 2.

Принцип двойственности. Отношение, обратное к частичному порядку, снова является частичным порядком.

Этот принцип позволяет вдвое сократить доказательства многих теорем; он обобщает принцип двойственности для булевых алгебр из § 5.1. Далее мы покажем, как он применяется к решеткам.

Сначала, однако, рассмотрим класс отношений, более общих, чем отношения порядка. (В свою очередь частично упорядоченные множества значительно шире класса всех решеток). Типичный пример доставляет отношение *делимости* в любом коммутативном моноиде, например $M = [\mathbf{Z}, \cdot]$. В таком моноиде отношение $a \mid b$ означает, что $ax = b$ для некоторого $x \in M$.

Лемма 1. *В любом коммутативном моноиде M отношение $a \mid b$ рефлексивно и транзитивно.*

Доказательство. Поскольку $a1 = a$, мы имеем $a \mid a$ для всех $a \in M$. Далее, из $ax = b$ и $by = c$ следует, что $a(xy) = (ax)y = by = c$. Поэтому из $a \mid b$ и $b \mid c$ следует, что $a \mid c$. Следовательно, отношение $\mid$ является отношением квазипорядка на M, и множество $[M, \mid]$ квазиупорядочено в смысле следующего определения.

Определение. *Квазипорядком* множества S называется любое рефлексивное и транзитивное бинарное отношение $<$. *Квазиупорядоченным множеством* $Q = [S, <]$ называется пара, состоящая из множества и квазипорядка на нем.

Очевидно, всякое частично упорядоченное множество является квазиупорядоченным множеством. Обратное неверно, ибо квазипорядок может не быть антисимметричным (как, например, в множестве $[\mathbf{Z}, \mid]$).

Лемма 2. *В квазиупорядоченном множестве $Q = [S, <]$ введем отношение $x \sim y$; оно означает, что $x < y$ и $y < x$. Тогда:*

а) *$\sim$ является отношением эквивалентности на S.*

б) *Если E и F — два класса эквивалентности относительно $\sim$, то либо ни для каких $x \in E$, $y \in F$ неверно, что $x < y$, либо для всех $x \in E$, $y \in F$ верно, что $x < y$.*

в) *Пусть* $E \leqslant F$ *означает, что* $x < y$ *для некоторых (и потому всех)* $x \in E$, $y \in F$. *Фактормножество* $S/\sim$ *частично упорядочено относительно этого отношения.*

Доказательство. Сначала докажем утверждение а). Рефлексивность отношения $\sim$ следует из того, что $x < x$ для всех $x \in S$.

Если $x \sim y$ и $y \sim z$, то $x < y$ и $y < z$ (по определению), откуда $x < z$. Аналогично $z < x$, так что $x \sim z$; поэтому отношение $\sim$ транзитивно. Симметричность следует из определения.

Перейдем к доказательству утверждения б). Если $x < y$ для некоторых $x \in E$, $y \in F$, то

$$x_1 < x < y < y_1 \text{ для всех } x_1 \in E,\ y_1 \in F,$$

откуда $x_1 < y_1$ по транзитивности.

Наконец, докажем в). Так как $x \sim x$, имеем $E \leqslant E$ для всех E. Из $E \leqslant F$ и $F \leqslant G$ следует, что $x < y < z$ для всех $x \in E$, $y \in F$, $z \in G$, откуда $x < z$. Мы установили транзитивность. Наконец, если $E \leqslant F$ и $F \leqslant E$, то для всех $x \in E$, $y \in F$ имеем $x < y$ и $y < x$, так что $x \sim y$ и $E = F$.

У леммы 2 имеется ряд приложений. Вот одно из них.

Пример 3. Отношение $a \mid b$ в коммутативном моноиде приводит к отношению $\sim$, которое в алгебраической теории чисел называется «ассоциированностью».

9.2. РЕШЕТКИ КАК ЧАСТИЧНО УПОРЯДОЧЕННЫЕ МНОЖЕСТВА

Мы уже отмечали, что при выполнении аксиом L1—L4 операции $\wedge$ и $\vee$ восстанавливаются по отношению порядка $\leqslant$, если используются понятия н.н.г. и н.в.г. В этом параграфе будет показано, что *можно обойтись и без аксиом* L1—L4: они выполнены для всякого частично упорядоченного множества L, в котором у любых двух элементов имеется н.н.г. и н.в.г. Идею доказательства этого факта можно проиллюстрировать на следующем примере.

Пример 4. Рассмотрим частично упорядоченное множество $[\mathbf{P}, \mid]$ всех положительных целых чисел относительно делимости. Напомним, что наибольший общий делитель $d = $ н.о.д. (m, n) двух положительных целых чисел m и n обладает следующими свойствами:

(i) $d \mid m$, $d \mid n$;

(i′) если $c \mid m$ и $c \mid n$, то $c \mid d$.

Аналогично наименьшее общее кратное $r = $ н. о. к. (m, n) чисел m и n обладает следующими свойствами:

(ii) $m \mid r$, $n \mid r$;
(ii′) если $m \mid s$ и $n \mid s$, то $r \mid s$.

Обобщение этих понятий на любые частично упорядоченные множества таково.

Определение. Пусть $[P, \leqslant]$ — частично упорядоченное множество $a, b \in P$. Элемент $d \in P$ называется *наибольшей нижней границей*, или *пересечением*, или *нижней гранью* a и b (обозначение: $d = a \wedge b$), если

$$d \leqslant a, \quad d \leqslant b \tag{1}$$

и

$$\text{из } x \leqslant a \text{ и } x \leqslant b \text{ следует, что } x \leqslant d. \tag{1′}$$

По двойственности элемент $s \in P$ называется *наименьшей верхней границей*, или *объединением*, или *верхней гранью* a и b (обозначение: $s = a \vee b$), если

$$a \leqslant s, \quad b \leqslant s \tag{2}$$

и

$$\text{из } a \leqslant x \text{ и } b \leqslant x \text{ следует, что } s \leqslant x. \tag{2′}$$

Не следует путать операции объединения и пересечения соответственно с операциями взятия наибольшего или наименьшего элемента: они совпадают там, где определены и те, и другие, но первые определены чаще. В цепях, т. е. частично упорядоченных множествах с условием

P4. для всех x, y либо $x \leqslant y$, либо $y \leqslant x$,

эти операции действительно не различаются. Объединение двух элементов в цепях равно наибольшему из них, а пересечение — наименьшему. Вообще говоря, в решетках это не так: см. по этому поводу упражнения.

Сформулируем теперь наш первый основной результат.

Теорема 1. *Если следующие выражения определены в частично упорядоченном множестве* P, *то*

L1. $x \wedge x = x$, $x \vee x = x$;
L2. $x \wedge y = y \wedge x$, $x \vee y = y \vee x$;
L3. $(x \wedge y) \wedge z = x \wedge (y \wedge z)$, $(x \vee y) \vee z = x \vee (y \vee z)$;
L4. $x \wedge (x \vee y) = x$, $x \vee (x \wedge y) = x$.

Доказательство. В силу принципа двойственности достаточно доказать лишь одно из тождеств под каждым номером; мы будем доказывать первое. Согласно P1, $x \leqslant x$. Поэтому из $d \leqslant x$ и $d \leqslant x$ следует, что $d \leqslant x$, откуда $x = x \wedge x$. Далее, $x \wedge y = y \wedge x$; это следует из того, что определение н.н.г. $\{x, y\}$ не зависит от порядка x и y. Аналогично $x \wedge (y \wedge z)$ и $(x \wedge y) \wedge z$ совпадают

с н.н.г. $\{x, y, z\}$. Общее определение н. н. г. для любого подмножества S элементов частично упорядоченного множества таково:

Определение. a = н. н. г. S означает, что:
(i) $a \leqslant x$ для всех $x \in S$;
(i′) из $v \leqslant x$ для всех $x \in S$ следует, что $v \leqslant a$.

Таким образом, мы установили L3. Наконец, $x \leqslant x \wedge (x \vee y)$. Действительно, $x \leqslant x$ и $x \leqslant x \vee y$. Обратно, если $b \leqslant x$ и $b \leqslant x \vee y$, то $b \leqslant x$. Это означает, что x = н. н. г. $\{x, x \vee y\}$; первое равенство из L4 доказано.

Дадим теперь основное определение этой главы, которое мотивируется доказанной теоремой.

Определение. *Решеткой* называется частично упорядоченное множество, в котором любые два элемента a, b обладают н. н. г. (обозначаемой через $a \wedge b$) и н. в. г. (обозначаемой через $a \vee b$).

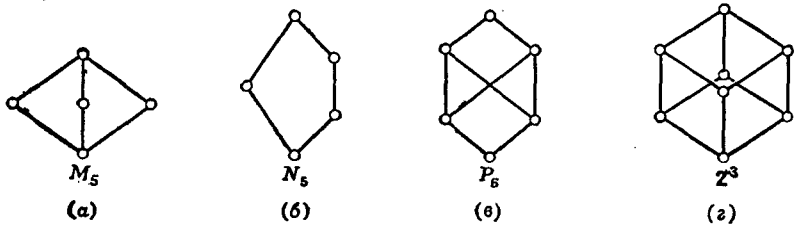


Рис. 9.1. Примеры частично упорядоченных множеств.

На рис. 9.1 изображены диаграммы четырех частично упорядоченных множеств. Первые два из них доставляют важные примеры решеток; четвертое также решетка (и даже булева алгебра 2^3). Третье множество решеткой не является.

Из теоремы 1 вытекает

Следствие. *Формулы* L1—L4 *теоремы* 1 *тождественно истинны в любой решетке.*

Мы займемся выяснением того, какие свойства булевых алгебр (установленные в гл. 5) выполняются в более общих решетках. Первый шаг состоит в проверке того, что данное определение решеток равносильно определению с помощью аксиом L1—L4.

Сразу же видно, что все диаграммы на рис. 9.1 являются частично упорядоченными множествами, если отношение $a \leqslant b$ определено, как в § 2.4: существует «лестница», ведущая от a до b, все «ступеньки» которой направлены вверх. Существование нижней грани для любых пар элементов первых двух диаграмм очевидно из соображений симметрии. Существование верхней грани получается по двойственности (переворачивая диаграмму вверх ногами, мы получаем изоморфную диаграмму).

Гораздо больше усилий потребовалось бы, чтобы написать таблицы операций $\wedge$, $\vee$ и проверить все тождества L1—L4. Таблица, относящаяся к рис. 9.1, *г*, имеет размер 8×8 и имеется 8^3 частных случаев тождества L3. Проверка частичной упорядоченности значительно экономнее.

УПРАЖНЕНИЯ А

1. Определение полурешеток дано в начале следующего параграфа. Проверить, что полурешетки можно охарактеризовать следующими двумя тождествами: (а) $xx=x$ для всех x; (б) $(xy)z=z(xy)$ для всех x, y, z.

2. Доказать, что если операции $\wedge$, $\vee$ в алгебре $[A, \wedge, \vee]$ удовлетворяют тождествам L2—L4, то они удовлетворяют тождествам L1.

3. Показать, что на рис. 9.1 диаграммы M_5, N_5 являются решетками, а P_6 нет.

4. Дать подробное доказательство того, что диаграмма рис. 9.1, *в* не является решеткой.

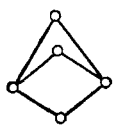
В упражнениях 5—7 *для каждой группы* G *построить диаграмму* $L(G)$ *решетки всех ее подгрупп.*

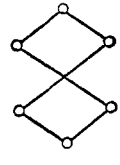
5. а) $G=D_6$, диэдральная группа порядка 6.
б) $G=D_8$, диэдральная группа порядка 8 (группа симметрий квадрата).

6. а) $G=Z_2\times Z_3$; б) $G=Z_2\times Z_4$.

7. а) $G=Z_{30}$; б) $G=Z_{36}$.

8. а) Какое из частично упорядоченных множеств, отвечающих изображенным ниже диаграммам, является решеткой?
б) Какие из элементов этой решетки имеют дополнения?





9. Показать, что частично упорядоченное множество, имеющее планарный граф и обладающее универсальными границами, является решеткой.

10. Пусть Q — любое рефлексивное и транзитивное отношение на множестве S.
а) Показать, что отношение $E=Q\cap Q^*$, означающее, что xQy и yQx, является отношением эквивалентности.
б) Показать, что для этого отношения выполнено свойство подстановки, т. е. из xEy и xQa следует yQa.
в) Показать, что система $[S/E, Q]$ является частично упорядоченным множеством.

9.3. РЕШЕТКИ И ПОЛУРЕШЕТКИ

Полурешеткой называется множество S с идемпотентной, коммутативной и ассоциативной бинарной операцией. Обозначая результат этой операции над a, b через ab (а в более сложных случаях используя скобки), мы получим, что полурешетка определяется следующими аксиомами:

$$a^2 = a,\ ab = ba,\ a(bc) = (ab)c \text{ для всех } a, b, c \in S. \tag{3}$$

Иными словами, полурешеткой называется коммутативная полугруппа, все элементы которой идемпотентны.

Лемма. *Пусть* $x_1, \ldots, x_n$*—элементы некоторой полурешетки, а* f, g*—любые два выражения, в которые входят все* $x_1, \ldots, x_n$. *Тогда* $f = g$.

Доказательство очевидно.

Важнейшие свойства полурешетки связаны с понятием делимости. Оно вводится так же, как в общих коммутативных моноидах (см. § 9.1): в полурешетке S отношение $a \mid b$ («a делит b») означает, что $ax = b$ для подходящего $x \in S$.

Заметим теперь, что отношение делимости в любой полурешетке удовлетворяет следующим соотношениям:

$$a \mid a; \tag{4}$$

$$\text{если } a \mid b \text{ и } b \mid a, \text{ то } a = b; \tag{5}$$

$$\text{если } a \mid b \text{ и } b \mid c, \text{ то } a \mid c. \tag{6}$$

Действительно, (4) следует из того, что $aa = a$. Кроме того, если $ax = b$ и $by = a$, то, используя ассоциативность, получаем

$$a = by = bby = ba = axa = aax = ax = b.$$

Соотношение (5) доказано. Далее, если $ax = b$ и $by = c$, то

$$a(xy) = (ax)y = by = c.$$

Это доказывает (6). Наконец заметим, что

$$a \mid ab;\ b \mid ab;\ \text{если } a \mid c \text{ и } b \mid c, \text{ то } ab \mid c. \tag{7}$$

Последнее вытекает из того, что если $ax = c$ и $by = c$, то $(ab)(xy) = ax(by) = cc = c$. Таким образом, ab есть наименьшее общее кратное a и b. Мы доказали, таким образом, первое утверждение следующей теоремы.

Теорема 2. *Любая полурешетка является частично упорядоченным множеством относительно делимости, и* $ab = \text{н.о.к.}(a, b)$.

Обратно, пусть P*—частично упорядоченное множество, в котором любые два элемента* a, b *имеют нижнюю грань* $a \wedge b$.

Тогда $[P, \wedge]$ *является полурешеткой, в которой* $a \mid b$ *тогда и только тогда, когда* $a \geqslant b$ *относительно частичного порядка в* P.

Для доказательства второго утверждения заметим, что тождество $a = a \wedge a$ выполняется очевидным образом, ибо $x \leqslant a$ равносильно $x \leqslant a$ и $x \leqslant a$. Аналогично $a \wedge b = b \wedge a$, ибо определение нижней грани элементов a и b симметрично по a и b. Далее, элементы $a \wedge (b \wedge c)$ и $(a \wedge b) \wedge c$ совпадают с нижней гранью тройки $\{a, b, c\}$. Таким образом, $[P, \wedge]$ есть полурешетка. Наконец, если $a \geqslant b$, то $a \wedge b = b$ (по транзитивности), так что $a \mid b$. Обратно, если $a \wedge x = b$, то $a \geqslant b$, так как b должен быть нижней границей $\{a, x\}$.

Напомним, что в § 9.2 решетка была определена как частично упорядоченное множество, в котором любые два элемента a, b обладают нижней гранью $a \wedge b$ и верхней гранью $a \vee b$. Из теоремы 2 вытекает следующее утверждение.

Следствие. *Любая решетка является полурешеткой относительно операции* $\wedge$, *а также относительно операции* $\vee$. *Эти операции связаны с порядком* $\leqslant$ *следующим образом: отношения*

$$a \geqslant b, \quad a \wedge b = b, \quad a \vee b = a \text{ эквивалентны.} \tag{8}$$

Нетрудно доказать более сильный результат:

Теорема 3. *Решеткой является любое множество* L *элементов с операциями* $\wedge$, $\vee$, *которые удовлетворяют аксиомам* L1—L4. *Иными словами, решетка есть полурешетка относительно двух операций* $\wedge$, $\vee$, *для которых выполняется следующий закон поглощения:*

$$a \wedge (a \vee b) = a \vee (a \wedge b) = a \text{ для всех } a, b \in L. \tag{9}$$

Доказательство. Очевидно, если эти условия выполнены, то L является решеткой. Обратно, из теоремы 2 и ее следствия сразу вытекают все утверждения, кроме закона поглощения (9). Чтобы вывести и его, заметим, что $a \wedge (a \wedge b) = (a \wedge a) \wedge b = a \wedge b$ в любой полурешетке. Согласно (8), отсюда следует, что $a \vee (a \wedge b) = a$. Второе тождество (9) получается из соображений двойственности.

Пример 5. Пусть $\mathcal{P}(I)$ — множество всех подмножеств любого множества I, и пусть $S \leqslant T$ означает, что $S \subset T$. Тогда $\mathcal{P}(I)$ является решеткой, в которой $S \cap T$ есть нижняя грань S и T, а $S \cup T$ — верхняя грань.

9.4. ПОДРЕШЕТКИ И ПРЯМЫЕ ПРОИЗВЕДЕНИЯ

В общем случае *подалгеброй* алгебраической системы A называется подмножество A, замкнутое относительно операций в A. Таким образом, *подрешетка* решетки L есть подмножество $S \subset L$,

такое, что

$$\text{если } a, b \in S, \text{ то } a \wedge b \in S \text{ и } a \vee b \in S. \tag{10}$$

Подрешетка множества $\mathcal{P}(U)$ частей данного множества U есть такое множество подмножеств U, которое вместе с S и T содержит также $S \cap T$ и $S \cup T$. Такие подрешетки называются *кольцами подмножеств* (обычно требуют также, чтобы в кольце содержались $\varnothing$ и U).

Прямые произведения алгебраических систем A, B одного типа также определяются естественным образом. Если рассматривать $A \times B$ как множество, то оно совпадает с декартовым произведением множеств A, B, т. е. состоит из упорядоченных пар (a, b), где $a \in A$, $b \in B$. Операции в $A \times B$ производятся покомпонентно. В частности, если A, B—решетки, то структура решетки на $A \times B$ определяется следующими правилами:

$$(a_1, b_1) \wedge (a_2, b_2) = (a_1 \wedge a_2, b_1 \wedge b_2), \tag{11}$$

$$(a_1, b_1) \vee (a_2, b_2) = (a_1 \vee a_2, b_1 \vee b_2). \tag{11'}$$

Аксиомы решетки в $A \times B$ проверяются тривиально, ибо они выполняются покомпонентно.

Например, решетка на рис. 9.1, *г* есть просто прямое произведение $2 \times 2 \times 2 = 2^3$ трех экземпляров решетки из двух элементов. По этой причине ее диаграмма напоминает рисунок куба, который является произведением трех отрезков.

Если U, V—два множества, то прямое произведение решеток $\mathcal{P}(U)$ и $\mathcal{P}(V)$ с операциями $\cap$, $\cup$ изоморфно решетке $\mathcal{P}(U \sqcup V)$.

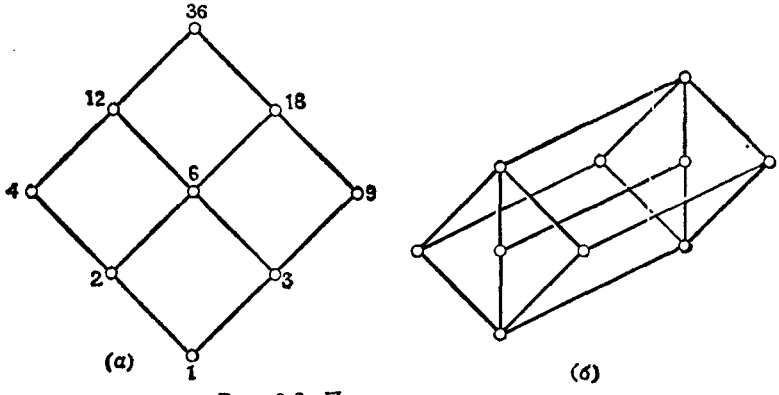


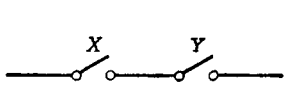
Рис. 9.2. Прямые произведения.
а) делители 36; б) $2 \times M_5$.

Решетка делителей числа 36, отвечающая частично упорядоченному множеству с отношением делимости, изоморфна прямому произведению $3^3 = 3 \times 3$ двух цепей из трех элементов (рис. 9.2, *а*).

На рис. 9.2, *б* изображено прямое произведение 2 с пятиэлементной решеткой M_5, представленной на рис. 9.1, *б*.

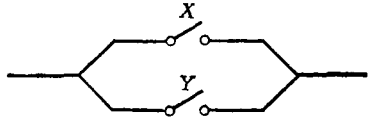
Элементы И, ИЛИ. В § 6.4 было установлено, что любую булеву функцию $f\colon 2^n \longrightarrow 2$ можно реализовать посредством подходящей вентильной схемы из элементов И, ИЛИ и инверторов. Предположим теперь, что мы не располагаем ни дополнениями входных переменных, ни инверторами. Поставим вопрос: какие булевы функции можно реализовать, пользуясь только элементами И, ИЛИ? Иными словами, какие функции $f\colon 2^n \longrightarrow 2$ можно построить из функций $f_i\colon \mathbf{x} = (x_1, \ldots, x_n) \mapsto x_i$, используя только операции $\wedge$ и $\vee$? Так как функции f_i и операции $\wedge$, $\vee$, $'$ порождают алгебру всех таких функций, т. е. свободную булеву алгебру 2^{2^n} с n образующими (гл. 5), мы хотим отыскать ее (дистрибутивную) *подрешетку*, порожденную элементами $f_1, \ldots, f_n$.

Реализация функции $X \wedge Y$ достигается последовательным соединением вентильных схем для X и Y, а реализация $X \vee Y$ — их параллельным соединением:



(*а*)

последовательное соединение



(*б*)

параллельное соединение

Поэтому введенный выше класс функций реализуется *последовательно-параллельными* схемами.

При $n = 2$ интересующие нас функции исчерпываются списком x, y, $x \wedge y$, $x \vee y$. При $n = 3$ таких функций уже 18: четыре самодвойственных элемента

$$x,\ y,\ z$$

$$(x \wedge y) \vee (y \wedge z) \vee (z \wedge x) = (x \vee y) \wedge (y \vee z) \wedge (z \vee x),$$

семь элементов

$$x \wedge y,\ y \wedge z,\ z \wedge x,\ x \wedge y \wedge z,$$
$$x \wedge (y \vee z),\ y \wedge (z \vee x),\ z \wedge (x \vee y)$$

и семь двойственных к ним элементов.

В гл. 5 мы вывели из аксиом L1—L4, что функции $x \wedge y$ и $x \vee y$ изотонны (сохраняют порядок) по обеим переменным. Индукция показывает, что то же верно для любого булева многочлена от любого числа нештрихованных переменных (здесь не используется закон дитрибутивности L6). Замечательно, что верно и обратное утверждение.

Теорема 4. *В свободной булевой алгебре всех функций* $f\colon 2^n \to 2$ *функции* $f_i\colon (x_1, \ldots, x_n) \mapsto x_i$ *и операции* $\wedge$, $\vee$ *порождают подрешетку, состоящую из всех изотонных функций.*

Доказательство мы опускаем. При $n = 2$ это утверждение можно проверить прямым рассмотрением 16 таблиц в примере 3 § 5.1. При $n = 3$ аналогичная проверка для всех типов симметрии булевых многочленов также возможна, но уже довольно кропотлива.

УПРАЖНЕНИЯ Б

1. Доказать со всеми подробностями, что любая цепь является дистрибутивной решеткой.

2. Доказать, что решетка является цепью тогда и только тогда, когда любое ее подмножество является подрешеткой.

3. *Интервал* $[a, b]$ в частично упорядоченном множестве состоит из всех x, удовлетворяющих условию $a \leqslant x \leqslant b$.

а) Доказать, что пересечение двух интервалов в решетке является интервалом (возможно, пустым).

б) Доказать, что любой интервал в решетке является подрешеткой.

4. Доказать со всеми подробностями, что отношения эквивалентности в любом конечном множестве S образуют решетку (относительно такого отношения $E_1 \leqslant E_2$ означает, что классы E_1 полностью содержатся в классах E_2). Она не является подрешеткой решетки всех бинарных отношений на S. (*Указание*: см. § 2.10.)

5. Пусть X, Y — множества, $f\colon X \to Y$ — функция. Показать, что множество образов $f(S)$ всех подмножеств $S \subseteq X$ образует подрешетку в $\mathcal{P}(Y)$. Когда она является булевой подалгеброй в $[\mathcal{P}(Y), \cap, \cup, ']$?

6. Доказать, что в дистрибутивной решетке L выполняется тождество

$$(x \wedge y) \vee (y \wedge z) \vee (z \wedge x) = (x \vee y) \wedge (y \vee z) \wedge (z \vee x),$$

но оно не выполняется в решетках M_5 и N_5, изображенных на рис. 9.1.

7. Доказать, что все нормальные подгруппы группы G образуют подрешетку в $L(G)$, решетке всех подгрупп группы G.

8. Доказать, что в *конечной* решетке все интервалы (см. упр. 3), включая пустые, образуют решетку относительно включения.

9. Пусть $M = [S, \cdot]$ — мультипликативный моноид. Показать, что отношения эквивалентности на S, обладающие свойством подстановки относительно умножения, образуют подрешетку в решетке, описанной в упр. 4.

9.5. ДИСТРИБУТИВНЫЕ РЕШЕТКИ

В пятиэлементных решетках M_5 и N_5 (рис. 9.1) нетрудно указать тройки элементов, для которых *не выполнены* дистрибутивные законы L6:

$$x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z), \tag{12}$$

$$x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z). \tag{12'}$$

С другой стороны, мы знаем, что они выполняются во всех решетках $\mathcal{P}(U)$, а потому и в решетке $\mathcal{P}(\mathbf{3}) = \mathbf{2}\times\mathbf{2}\times\mathbf{2}$ на рис. 9.1, *г*.

Вообще решетка называется *дистрибутивной*, если в ней выполнены дистрибутивные законы (12) и (12′); в противном случае она не дистрибутивна. Мы убедились, что существуют как дистрибутивные, так и недистрибутивные решетки.

Подрешетка дистрибутивной решетки, очевидно, дистрибутивна. Поэтому *всякое кольцо множеств является дистрибутивной решеткой.*

Теорема 5. *В любой решетке справедливы законы полудистрибутивности:*

$$x \wedge (y \vee z) \geqslant (x \wedge y) \vee (x \wedge z), \tag{13}$$

$$x \vee (y \wedge z) \leqslant (x \vee y) \wedge (x \vee z). \tag{13'}$$

Кроме того, из (12) *следует* (12′), *и обратно.*

Доказательство. Так как неравенства (13), (13′) двойственны друг другу, достаточно проверить первое из них. Очевидно, x является верхней границей элементов $x \wedge y$ и $x \wedge z$, а потому также их наименьшей верхней границей $(x \wedge y) \vee (x \vee z)$. Аналогично, из неравенств $y \geqslant x \wedge y$ и $z \geqslant x \wedge z$ следует, что любая верхняя граница элементов y и z должна мажорировать $(x \wedge y) \vee (x \wedge z)$. Поэтому $y \vee z$ есть верхняя граница для $(x \wedge y) \vee (x \wedge z)$. Это показывает, что $(x \wedge y) \vee (x \wedge z)$ является *нижней* границей для x и $y \vee z$. Отсюда неравенство (13) следует по определению нижней грани.

Покажем теперь, что из (12) следует (12′). Действительно, используя (12), находим

$$\begin{aligned}(a \vee b) \wedge (a \vee c) &= [(a \vee b) \wedge a] \vee [(a \vee b) \wedge c] = \\ &= a \vee [(a \wedge c) \vee (b \wedge c)] = \quad [\text{в силу (9) и (12)}] \\ &= [a \vee (a \wedge c)] \vee (b \wedge c) = a \vee (b \wedge c)\ [\text{в силу (9)}].\end{aligned}$$

По двойственности (12) выводится из (12′), а это завершает доказательство.

Лемма 1. *В любой дистрибутивной решетке*

$$\text{из } a \wedge x = a \wedge y \quad \text{и} \quad a \vee x = a \vee y \text{ следует, что } x = y. \tag{14}$$

Доказательство. Используя последовательно L4, коммутативность и первое равенство (14), (12′), коммутативность и второе равенство (14), (12′), коммутативность, первое равенство (14)

и, наконец, L4, получаем

$$x = x \vee (x \wedge a) = x \vee (y \wedge a) =$$
$$= (x \vee y) \wedge (x \vee a) = (y \vee x) \wedge (y \vee a) =$$
$$= y \vee (x \wedge a) = y \vee (y \wedge a) = y.$$

Особый интерес представляет случай $a \wedge x = a \wedge y = O$, $a \vee x = a \vee y = I$, где O, I — универсальные границы. *Дополнением* элемента a в решетке L с универсальными границами O, I называется такой элемент $x \in L$, что $a \wedge x = O$ и $a \vee x = I$. Очевидно, O и I дополнительны друг другу.

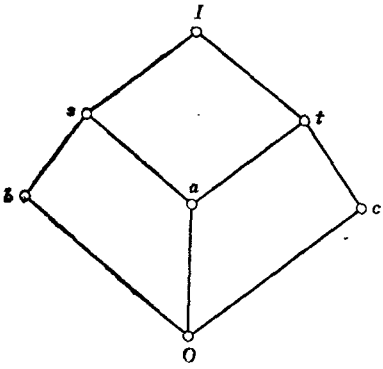


Рис. 9.3.

Из леммы 1 вытекает

Следствие 1. *В любой дистрибутивной решетке L элемент a может иметь не более одного дополнения.*

Если дополнение к a существует, оно обычно обозначается через a'. Таким образом, в дистрибутивной решетке

$$a \wedge a' = O, \quad a \vee a' = I. \quad (15)$$

Следствие 2. *Любой изоморфизм частично упорядоченных множеств, являющихся булевыми алгебрами, определяет изоморфизм булевых алгебр.*

В большинстве решеток, например в цепях длины $n > 2$, существуют элементы без дополнений. В решетке на рис. 9.3 единственным таким элементом является a.

Лемма 2. *В любой дистрибутивной решетке множество всех элементов с дополнениями образует подрешетку.*

Доказательство. Пусть a, a' и b, b' — пары дополнительных элементов. Тогда

$$(a \wedge b) \wedge (a' \vee b') = (a \wedge b \wedge a') \vee (a \wedge b \wedge b') = O \vee O = O,$$
$$(a \wedge b) \vee (a' \vee b') = (a \vee a' \vee b') \wedge (b \vee a' \vee b') = I \vee I = I.$$

Следовательно, элементы $a \wedge b$ и $a' \vee b'$ дополнительны. Аналогично устанавливается, что $a \vee b$ и $a' \wedge b'$ дополнительны друг к другу. Доказательство завершено.

В недистрибутивной решетке на рис. 9.3 элементы с дополнениями не образуют подрешетку.

9.6. МОДУЛЯРНЫЕ И ГЕОМЕТРИЧЕСКИЕ РЕШЕТКИ

В этом параграфе мы опишем два важных класса решеток, которые не являются дистрибутивными, но обладают другими полезными свойствами.

Первый из них — это класс модулярных решеток. Решетка называется *модулярной*, если она удовлетворяет двум тождествам аксиомы L5 § 5.1. Представим эту аксиому в более простом виде.

Лемма. *В любой решетке каждое из двух тождеств аксиомы* L5 § 5.1 *равносильно условию*

L5*. *Если* $a \leqslant c$, *то* $a \vee (b \wedge c) = (a \vee b) \wedge c$.

Доказательство. Поскольку $x \leqslant x \vee z$, из L5* следует, что

$$x \vee [y \wedge (x \vee z)] = (x \vee y) \wedge (x \vee z).$$

Это и есть второе тождество L5.

Обратно, если $a \leqslant c$, то $a \vee c = c$. Второе тождество L5 при $a \leqslant c$ поэтому сводится к $a \vee (b \wedge c) = (a \vee b) \wedge c$, т. е. к L5*. Таким образом, L5* равносильно второму тождеству L5.

Наконец, так как условие L5* самодвойственно, из принципа двойственности следует, что и первое тождество L5 также эквивалентно L5*.

Важнейший класс модулярных решеток описывается следующей теоремой.

Теорема 6. *Нормальные подгруппы любой группы* G *образуют модулярную решетку* $M(G)$.

Доказательство. Очевидно, $M(G)$ является подрешеткой решетки $L(G)$ всех подгрупп G (см. пример 2 § 9.1). Мы должны установить, что если $H \lhd G$, $K \lhd G$ и $L \lhd G$, то

$$\text{из } H \leqslant L \text{ следует, что } H \vee (K \wedge L) = (H \vee K) \wedge L. \quad (16)$$

Но, очевидно, $H \leqslant H \vee K$, $H \leqslant L$, $K \wedge L \leqslant K \leqslant H \vee K$ и $K \wedge L \leqslant L$. Поэтому H и $K \wedge L$ являются нижними границами для $(H \vee K) \wedge L$, так что их объединение $H \vee (K \wedge L)$ также является нижней границей. Это означает, что $H \vee (K \wedge L) \leqslant (H \vee K) \wedge L$ (это рассуждение проходит для любой решетки).

Чтобы доказать равенство (16), в силу P2 достаточно установить включение

$$(H \vee K) \wedge L \subset H \vee (K \wedge L).$$

Но если $x \in (H \vee K) \wedge L$, то $x = hk = l$ $(h \in H,\ k \in K,\ l \in L)$, поскольку из $H \lhd G$, $K \lhd G$ следует, что $H \vee K = HK$. Но из $H \subset L$ вытекает, что $h \in L$, так что $k = h^{-1}l \in L$. Поэтому $k \in K \wedge L$, откуда

$$x = hk \in H \vee (K \wedge L),$$

что и завершает доказательство.

Пятиэлементная решетка на рис. 9.1, *а* изоморфна решетке всех (а также нормальных) подгрупп 4-группы, абелевой группы всех симметрий прямоугольника.

Пример 6. Рассмотрим все подгруппы аддитивной группы двоичных троек. Имеется семь таких подгрупп порядка 2 и семь порядка 4. Группы порядка 2 состоят из 000 и одного из элементов множества {001, 011, 010, 111, 101, 110, 100}. Подгруппы порядка 4 состоят из всевозможных троек $x = (x_1, x_2, x_3)$, координаты которых удовлетворяют условию вида

$$a_1x_1 + a_2x_2 + a_3x_3 \equiv 0 \pmod 2$$

для некоторого $\mathbf{a} = a_1a_2a_3 \neq 000$. Тройку **а** можно взять из множества {100, 101, 110, 111, 011, 001}. На рис. 9.4, *а* эти тройки

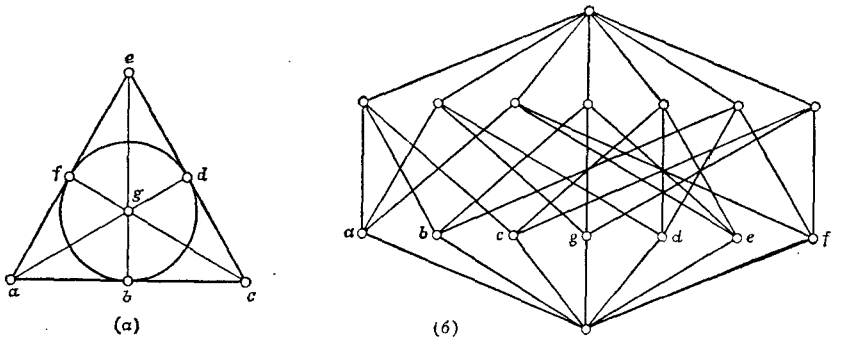


Рис. 9.4. Решетка подгрупп $\mathbf{Z}_2^3$.

представлены шестью отрезками и одной окружностью. На рис. 9.4, *б* изображена диаграмма для отношения порядка соответствующей решетки, которая называется плоскостью Фано.

Займемся теперь решетками разбиений. Понятие *разбиения*, введенное в гл. 2, принадлежит к числу основных понятий теории множеств. Начнем со следующего результата.

Теорема 7. *Упорядочим множество всех разбиений данного множества следующим отношением частичного порядка:* $\pi \leqslant \pi^*$ (π *есть подразбиение* π^*) *означает, что*

$$\textit{из } x\pi y \textit{ следует, что } x\pi^* y. \tag{17}$$

Тогда разбиения образуют решетку.

Заметим, что отношение $\leqslant$ совпадает с отношением включения для бинарных отношений, которое уже было определено в гл. 2.

Здесь мы отождествляем π, π^* с отношениями эквивалентности $E(\pi)$, $E(\pi^*)$, построенными по π, π^*-соответственно.

Доказательство. Определим $\pi \wedge \pi^*$ условием:

$$x(\pi \wedge \pi^*)y \text{ означает, что } x\pi y \text{ и } x\pi^* y. \tag{18}$$

Нетрудно проверить, что $\pi \wedge \pi^*$ является отношением эквивалентности, т. е. оно рефлексивно, симметрично и транзитивно. Далее, пусть

$$x(\pi \vee \pi^*)y \text{ означает, что для подходящих } z_1, \ldots, z_{2n} \tag{18'}$$
$$x \pi z_1,\ z_1\pi^* z_2,\ z_2\pi z_3,\ \ldots,\ z_{2n-1}\pi^* z_{2n},\ z_{2n}\pi y.$$

Решетка на рис. 9.5, *а* изоморфна решетке всех разбиений π_4 множества, состоящего из четырех элементов. Для трехэлементного множества соответствующая решетка π_3 изображена на рис. 9.1, *а*.

Сформулируем следующее обобщение результата из примера 2.

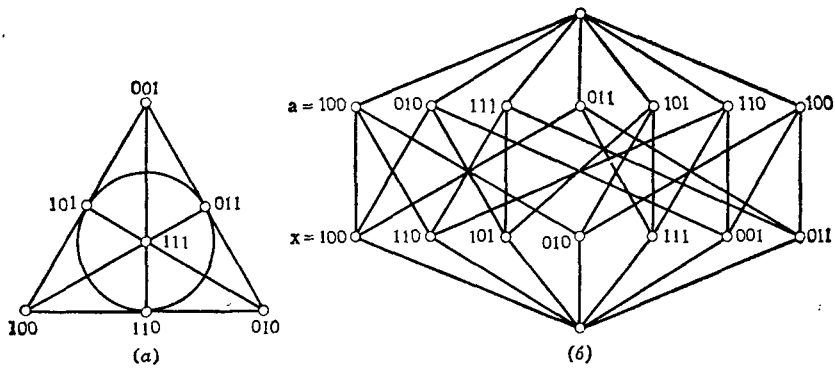


Рис. 9.5. Решетка подгрупп Δ_4.

Теорема 8. *Решетка всех подгрупп конечной группы G порядка n является подрешеткой решетки π_n всех разбиений G как множества.*

На рис. 9.4, *б* изображена решетка всех подгрупп группы симметрий квадрата (диэдральной группы порядка 8).

УПРАЖНЕНИЯ В

1. а) Пусть A, B — цепи. Доказать, что любой их морфизм, сохраняющий порядок, является морфизмом решеток.

б) Сколько имеется морфизмов $\mathbf{4} \to \mathbf{3}$, сохраняющих порядок?

в) Сколько среди них эпиморфизмов?

2. Доказать, что 1_n является единственным автоморфизмом множества n, сохраняющим порядок.

3. Пусть P — частично упорядоченное множество с универсальными границами, и в нем нет цепей, длины которых больше 2.

а) Показать, что P — решетка.

б) Показать, что P либо является цепью 3, либо обладает нетождественным автоморфизмом.

4. Пусть L — дистрибутивная решетка, и пусть $a < b$ в L. Показать, что отображение $x \mapsto (x \vee a) \wedge b$ является морфизмом решеток, который проектирует всю L на отрезок $[a, b]$.

5. Показать со всеми подробностями, что если f — изоморфизм решеток, то f^{-1} также изоморфизм решеток.

6. Пусть $h\colon L \to M$ — морфизм решеток, и пусть L обладает универсальными границами.

а) Показать, что если $h(x') = [h(x)]'$ хотя бы для одной пары дополнительных элементов x, x' в L, то M обладает универсальными границами $h(O_L) = O_M$ и $h(I_L) = I_M$.

б) Обратно, показать, что если $h(O_L) = O_M$ и $h(I_L) = I_M$, то $h(x') = [h(x)]'$ для любой пары дополнительных элементов x, $x' \in L$.

7. Пусть $h\colon L \twoheadrightarrow M$ — эпиморфизм решеток и L имеет универсальные границы. Показать, что M обладает универсальными границами и что $h(x') = [h(x)]'$ в M для любого элемента x, имеющего дополнение в L.

8. Доказать со всеми подробностями, что морфизмы любой решетки L в себя образуют моноид.

9. а) Пусть L, M — дистрибутивные решетки, а g и $h\colon L \to M$ — морфизмы решеток. Будем писать $g \leqslant h$, если $g(x) \leqslant h(x)$ для всех $x \in L$. Показать, что множество всех морфизмов решеток из L в M само образует решетку относительно этого отношения порядка.

б) Необходимо ли условие дистрибутивности в п. а?

*10. Показать, что решетка всех булевых подалгебр булевой алгебры $\mathcal{P}(n)$ двойственна решетке всех отношений эквивалентности на множестве n. (*Указание*: попробуйте рассмотреть случай $n = 4$).

11. а) Построить сохраняющий порядок эпиморфизм из  на  .

б) Показать, что любой эпиморфный образ решетки  изоморфен либо ей самой, либо одноэлементной решетке o.

12. Показать, что у свободной булевой алгебры с двумя образующими (см. § 5.1) имеется 24 автоморфизма и 12 пар образующих.

13. Построить нетривиальные эпиморфизмы решетки  .

14. Пусть $P = \{x_1, x_2, \ldots, x_N\}$ — некоторое частично упорядоченное множество. Положим $\varphi(1) = 1$ и затем по рекурсии

$$\varphi(i+1) = \begin{cases} i+1, & \text{если } x_{i+} < x_{\varphi(i)}, \\ \varphi(i) & \text{в противном случае.} \end{cases}$$

Доказать по индукции, что $x_{\varphi(N)}$ — некоторый минимальный элемент множества P.

15. Массив X содержит 100 различных целых чисел $X[1], X[2], \ldots, X[100]$. Написать программу на АЛГОЛе, вычисляющую индекс m наименьшего целого числа $X[m]$ в массиве X.

*9.7. БУЛЕВЫ РЕШЕТКИ

Рассмотрим булевы алгебры как частично упорядоченные множества, т. е. системы с одним отношением порядка. Так как операции $\wedge$ и $\vee$ совпадают с операциями взятия нижней и верхней грани соответственно, булевы алгебры являются *решетками*. Поставим вопрос, какие решетки отвечают булевым алгебрам. Ясно, что такие решетки должны быть *дистрибутивными*. Далее, они должны быть решетками *с дополнением* в следующем смысле.

Определение. *Решеткой с дополнением* называется решетка, которая обладает универсальными границами O, I и в которой каждый элемент a имеет хотя бы одно дополнение x, т. е.

$$a \wedge x = O, \qquad a \vee x = I. \tag{19}$$

Булевой решеткой называется дистрибутивная решетка с дополнением.

Теорема 9. *В любой булевой решетке справедливы тождества*

$$(a')' = a, \tag{20}$$

$$(a \wedge b)' = a' \vee b', \qquad (a \vee b)' = a' \wedge b'. \tag{21}$$

Доказательство. Равенство (20) следует из единственности дополнений (лемма 1, следствие 1, § 9.5) и симметричности отношения дополняемости. Таким образом, функция c: $A \to A$, $a \mapsto a'$, является биекцией и обратная к ней совпадает с нею. Кроме того, она обращает порядок. Действительно, из $a \leqslant b$ следует, что $a \wedge b' \leqslant b \wedge b' = O$. Поэтому $a \wedge b' = O$ и

$$b' = b' \wedge I = b' \wedge (a \vee a') = (b' \wedge a) \vee (b' \wedge a') =$$
$$= O \vee (b' \wedge a') = b' \wedge a',$$

так что $b' \leqslant a'$.

Поэтому c является изоморфизмом решетки и двойственной к ней решетки, что доказывает (21).

Следствие. *Класс решеток, которые получаются из булевых алгебр «забвением» операции дополнения, совпадает с классом булевых решеток.*

*9.8. МОРФИЗМЫ И ИДЕАЛЫ

Функция $\theta\colon L \to M$ из решетки L в решетку M называется *морфизмом решеток*, если для всех $x, y \in L$

$$\theta(x \wedge y) = \theta(x) \wedge \theta(y), \quad \theta(x \vee y) = \theta(x) \vee \theta(y). \tag{22}$$

Это свойство равносильно коммутативности диаграмм

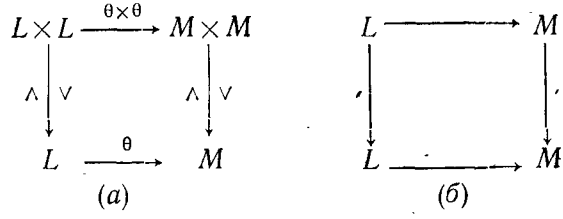


Рис. 9.6. Диаграммы морфизмов решеток.

Всякий морфизм решеток сохраняет порядок, т. е. изотонен.

Однако изотонное изображение может не быть морфизмом решеток. Пусть, например, $P = [S, \leqslant]$ — частично упорядоченное множество. Рассмотрим отображение θ_P: $a \mapsto \theta_P(a)$, ставящее в соответствие элементу $a \in S$ множество $\theta_P(a) = A \subset S$ всех $x \leqslant a$ в $[S, \leqslant]$. Ясно, что по транзитивности из $a \leqslant b$ в S следует $\theta_P(a) \subset \theta_P(b)$ в $\mathcal{P}(S)$. Значит, θ_P — изотонное отображение S в $\mathcal{P}(S)$. Если S — решетка, то $\theta_P(a \wedge b) = \theta_P(a) \cap \theta_P(b)$ (докажите это). Однако, вообще говоря, $\theta_P(a \vee b)$ строго больше, чем $\theta_P(a) \cup \theta_P(b)$ (например, в решетках на рис. 9.1, *а* и *б*). Поэтому θ_P *не является* морфизмом решеток.

Можно показать, что любая конечная дистрибутивная решетка изоморфна некоторому кольцу множеств. Используя аксиому выбора (гл. 14), можно доказать, что *любая* дистрибутивная решетка изоморфна кольцу множеств. Поэтому тождества, определяющие дистрибутивные решетки, полностью характеризуют алгебраические свойства операций $\cap$ и $\cup$ над множествами. Мы не будем приводить здесь доказательства этих фактов. Однако в § 9.9 будет показано, что всякая конечная булева решетка (алгебра) L изоморфна множеству частей некоторого множества (а именно множества атомов этой алгебры).

Назовем *идеалом* в решетке L непустое подмножество $J \subset L$, обладающее следующими свойствами:

(i) если $a \in J$ и $b \in J$, то $a \vee b \in J$;

(ii) если $a \in J$ и $x \leqslant a$, то $x \in J$.

Теорема 10. *Пусть* $\theta\colon L \to M$ *— морфизм решеток, образ которого содержит* $O \in M$. *Тогда* $\theta^{-1}(O)$ *является идеалом в* L.

Доказательство. Пусть $K=\theta^{-1}(O)$ (это множество называется *ядром* θ). Оно непусто по предположению. Условие (i) выполнено, ибо из $\theta(a)=\theta(b)=O$ следует, что $\theta(a\vee b)=\theta(a)\vee\vee\theta(b)=O\vee O=O$. Условие (ii) также выполнено, так как если $\theta(a)=0$, то из $x\leqslant a$ следует, что $\theta(x)\leqslant\theta(a)=0$.

Булевы морфизмы. *Булев морфизм* θ: $L\to M$ есть морфизм решеток, удовлетворяющий дополнительному условию

$$\theta(a')=[\theta(a)]' \tag{23}$$

для всех $a\in L$.

Ядром θ является $\theta^{-1}(O)$. Оно непусто, ибо заведомо содержит $O\in L$:

$$\theta(O)=\theta(a\wedge a')=\theta(a)\wedge\theta(a')=\theta(a)\wedge[\theta(a)]'=0.$$

Следствие. *Ядро K любого булева морфизма является идеалом.*

Обратно, любой идеал $J\subset A$ булевой алгебры A является ядром некоторого булева эпиморфизма θ: $A\to B$, образ которого $B\cong A/J$ однозначно с точностью до изоморфизма восстанавливается по ядру. Легче всего установить это, рассмотрев A как *коммутативное кольцо* (см. гл. 10) относительно умножения $xy=x\wedge y$ и сложения $x+y=(x\wedge y')\vee(x'\wedge y)$.

* 9.9. КОНЕЧНЫЕ БУЛЕВЫ АЛГЕБРЫ

В этом параграфе мы воспользуемся введенными выше понятиями для доказательства того, что любая конечная булева алгебра изоморфна множеству частей некоторого конечного множества и потому состоит из 2^n элементов для некоторого целого n.

В доказательстве существенно используются свойства морфизмов и прямых произведений.

Лемма 1. *В любой дистрибутивной решетке L для любого $a\in L$ отображение $x\mapsto(x\wedge a,\ x\vee a)$ определяет мономорфизм решеток* θ: $L\to[O,a]\times[a,I]$.

Доказательство. В силу законов дистрибутивности L6,

$$x\wedge y\mapsto((x\wedge y)\wedge a,\ (x\wedge y)\vee a)=$$
$$=((x\wedge a)\wedge(y\wedge a),\ (x\vee a)\wedge(y\vee a)).$$

По определению прямого произведения решеток последнее выражение совпадает с $\theta(x)\wedge\theta(y)$. Таким образом, $\theta(x\wedge y)=\theta(x)\wedge\theta(y)$. Аналогично устанавливается, что $\theta(x\vee y)=\theta(x)\vee\theta(y)$. Поэтому θ есть морфизм решеток. Его мономорфность следует из того, что если $x\wedge a=y\wedge a$ и $x\vee a=y\vee a$, то

$$x=x\wedge(x\vee a)=x\wedge(y\vee a)=(x\wedge y)\vee(x\wedge a)=$$
$$=(y\wedge x)\vee(y\wedge a)=y\wedge(x\vee a)=y\wedge(y\vee a)=y.$$

Лемма 2. *Пусть L—дистрибутивная решетка, a—элемент с дополнением a'. Тогда для любого $x \in L$ существует единственный морфизм решеток* θ: $3^2 \to L$, *такой, что*

$$(2, 0) \mapsto a, \quad (1, 1) \mapsto x, \quad (0, 2) \mapsto a',$$

как показано на рис. 9.7.

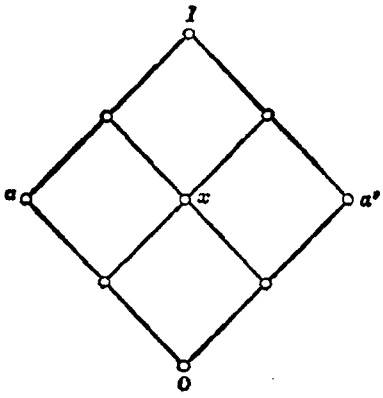


Рис. 9.7.

Лемма 3. *В любой булевой решетке L и для любого элемента $a \in L$ отображение* β: $x \to (x \wedge a, x \vee a)$ *является изоморфизмом.*

Доказательство. Из леммы 1 мы знаем, что это мономорфизм. Поэтому достаточно проверить, что это эпиморфизм, т. е. что любая пара (y, z) с $y \leqslant a$ и $z \geqslant a$, является образом относительно β некоторого элемента $x \in L$. Мы покажем, что в качестве x можно взять $y \vee (z \wedge a')$, т. е. что отображение γ: $(y, z) \mapsto y \vee (z \wedge a')$ является левым обратным к β относительно правой композиции. (Ниже мы покажем, что β—биекция, так что γ двусторонне обратно к β.) Разобьем проверку на два шага.

Прежде всего имеют место тождества

$$\begin{aligned}[y \vee (z \wedge a')] \wedge a &= (y \vee z) \wedge (y \vee a') \wedge a = (y \vee a') \wedge a = \\ &= (y \vee a') \wedge (y \vee z) \wedge a = (y \vee a') \wedge a = \\ &= (y \wedge a) \vee (a' \wedge a) = y \vee 0 = y. \end{aligned} \tag{i}$$

Доказательство (i). Первое равенство следует из дистрибутивности, второе из L2—L3, третье из того, что $y \vee z \geqslant z \geqslant a$, откуда $(y \vee z) \wedge a = a$. Пользуясь теперь дистрибутивностью и неравенством $y \leqslant a$, мы сразу же получаем последние два равенства.

Остается доказать тождества

$$\begin{aligned}[y \vee (z \wedge a')] \vee a &= (a \vee y) \vee (z \wedge a') = a \vee (z \wedge a') = \\ &= (a \vee z) \wedge (a \vee a') = z \wedge I = z. \end{aligned} \tag{ii}$$

Доказательство (ii). Первое равенство следует из L2—L3, если опустить скобки и поменять местами подходящие члены. Второе вытекает из $y \leqslant a$. Остальные равенства двойственны последним равенствам в доказательстве тождеств (i).

Теорема 11. *В любой булевой алгебре L для любого элемента $a \in L$ с дополнением a' отображение* α: $x \mapsto [x \wedge a, x \wedge a']$

является изоморфизмом; обратный к которому имеет вид $(y, z) \mapsto y \vee z$:

$$\alpha\colon L \cong [0, a] \times [0, a'] = A \times A'. \tag{24}$$

Доказательство (24). Из леммы 3 следует, что отображение $\beta\colon x \mapsto (x \wedge a, x \vee a)$ определяет изоморфизм булевых алгебр (см. § 9.5, лемма 1, следствие 2). Иными словами, $\beta\colon L \cong [0, a] \times [a, I]$. С другой стороны, отображение $\alpha'\colon x \mapsto x \wedge a'$ определяет изоморфизм $\alpha'\colon [a, I] \cong [0, a']$, обратный к которому имеет вид $\alpha\colon x \mapsto x \vee a$, поскольку α изотонен и

$$(x \wedge a') \vee a = (x \vee a) \wedge (a' \vee a) = x \vee a = x \quad \text{для } x \in [a, I],$$

$$(y \vee a) \wedge a' = (y \wedge a') \vee (a \wedge a') = y \wedge a' = y \quad \text{для } y \in [0, a'].$$

Теорема 12. *Любая конечная булева алгебра* L *изоморфна* 2^n *для подходящего* $n \in \mathbf{N}$.

Доказательство проводится индукцией по числу элементов L. Если $L \cong \mathbf{1}$ или $\mathbf{2}$, то результат тривиален для $n = 0, 1$ соответственно.

В противном случае L содержит некоторый элемент a, такой, что $0 < a < I$. По теореме 11 $L \cong A \times A'$, где A и A' имеют порядок, меньший, чем порядок L. В силу предположения индукции $A \cong 2^r$ и $A' \cong 2^s$ для подходящих r, s. Поэтому $L \cong A \times A' = 2^{r+s}$.

В качестве последнего результата несколько обобщим построение, использованное в теореме 11.

Теорема 13. *Пусть* a *— элемент булевой алгебры* L *и* a' *— его дополнение. Рассмотрим отображения*

$$\begin{aligned} &\alpha\colon x \mapsto x \wedge a, \quad \alpha'\colon x \mapsto x \wedge a', \\ &\alpha^\dagger\colon x \mapsto x \vee a, \quad \alpha^*\colon x \mapsto x \vee a'. \end{aligned} \tag{25}$$

Тогда $\operatorname{Im}\alpha = [0, a]$, $\operatorname{Im}\alpha' = [0, a']$, $\operatorname{Im}\alpha^\dagger = [a, I]$ *и* $\operatorname{Im}\alpha^* = [a', I]$. *Относительно правой композиции имеем*

$$\alpha\alpha^\dagger = \alpha^\dagger\alpha = p_a, \quad \alpha'\alpha = \alpha^*\alpha' = p_{a'}, \tag{26}$$

$$\alpha\alpha' = \alpha'\alpha = p_0, \quad \alpha^\dagger\alpha = \alpha^*\alpha^\dagger = p_1, \tag{26'}$$

$$\alpha\alpha^* = \alpha^*, \quad \alpha^*\alpha = \alpha, \quad \alpha'\alpha^\dagger = \alpha^\dagger, \quad \alpha^\dagger\alpha' = \alpha'. \tag{26''}$$

Набросок доказательства. Тождества (26)—(26″) проверяются непосредственно. Для проверки тождества $\alpha\alpha^* = \alpha^*$, например, вычислим

$$(x \wedge a) \vee a' = (x \vee a') \wedge (a \vee a') = (x \vee a') \wedge I = x \vee a'.$$

Проверка первого утверждения аналогична доказательству леммы 3.

Следствие. *Множество* $A = \{I_L, \alpha, \alpha', \alpha^\dagger, \alpha^*, p_a, p_{a'}, p_0, p_1\}$ *образует моноид относительно правой композиции.*

СПИСОК ЛИТЕРАТУРЫ

1. Abbott J. C., Sets, Lattices and Boolean Algebra, Allyn and Bacon, 1969.
2. Birkhoff G., Lattice Theory, 3d ed., Amer. Math. Soc., 1967.
3. Lieber L. R., Lattice Theory (illus. by H. G. Lieber), Galois Institute, 1959.
4. Rutherford D. E., Introduction to Lattice Theory, Hafner, 1965.
5. Szasz G., Introduction to Lattice Theory, 2d ed., Academic Press, 1963.

10

КОЛЬЦА И ИДЕАЛЫ

10.1. ВВЕДЕНИЕ

В следующих четырех главах мы изучим класс алгебраических систем, называемых *кольцами*. В кольцах определены две основные операции—*сложение* и *умножение*, которые удовлетворяют большинству известных законов элементарной алгебры. Однако умножение часто не предполагается коммутативным, а возможность деления постулируется редко. Довольно типичным представителем коммутативных колец является $\mathbf{Z}$.

Определение. *Кольцом* называется алгебраическая система $R=[R, +, \cdot]$ с двумя бинарными операциями $+$ и $\cdot$, которые удовлетворяют следующим условиям.

(i) $+$ определяет на R структуру абелевой группы;
(ii) $\cdot$ определяет на R структуру моноида;
(iii) справедливы законы дистрибутивности

$$a(b+c)=ab+ac, \quad (a+b)c=ac+bc \tag{1}$$

для любых $a, b, c \in R$.

Целые числа $\mathbf{Z}=[\mathbf{Z}, +, \cdot]$ с обычным сложением и умножением образуют кольцо. Как мы уже отмечали, возможность деления не постулируется.

Кольцо называется *коммутативным*, если

$$ab=ba \text{ для всех } a, b \in R. \tag{2}$$

Кольца $\mathbf{Z}$, $\mathbf{Z}_n$ (целые числа по модулю n, определенные в § 2.6) и другие хорошо известные числовые системы коммутативны. Однако многие другие кольца некоммутативны. Если дано любое нетривиальное кольцо R, коммутативное или нет, то из R можно следующим образом построить некоммутативное кольцо.

Пример 1. Обозначим через $M_2(R)$ множество всех 2×2-матриц $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ с элементами $a, b, c, d \in R$. Определим сложение и

умножение матриц формулами

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} + \begin{bmatrix} a' & b' \\ c' & d' \end{bmatrix} = \begin{bmatrix} a+a' & b+b' \\ c+c' & d+d' \end{bmatrix}, \tag{3}$$

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} a' & b' \\ c' & d' \end{bmatrix} = \begin{bmatrix} aa'+bc' & ab'+bd' \\ ca'+dc' & cb'+dd' \end{bmatrix}. \tag{3'}$$

Тогда $M_2(R)$ является некоммутативным кольцом.

Пример 2. Пусть R — любое кольцо и n — положительное целое число. Обозначим через $M_n(R)$ множество всех $n \times n$-матриц $A = \|a_{ij}\|$, $B = \|b_{ij}\|$ и т. д. Определим сложение и умножение матриц формулами

$$\|a_{ij}\| + \|b_{ij}\| = \|a_{ij} + b_{ij}\|, \tag{4}$$

$$\|a_{ij}\| \|b_{ij}\| = \left\| \sum_{k=1}^{n} a_{ik} b_{kj} \right\|. \tag{4'}$$

Тогда $M_n(R)$ является кольцом с единицей, которая задается единичной матрицей

$$I = \|\delta_{ij}\|, \text{ где } \delta_{ij} = \begin{cases} 1 \text{ при } i = j, \\ 0 \text{ при } i \neq j. \end{cases} \tag{5}$$

Одна из наших основных целей будет состоять в описании построения колец с различными желаемыми свойствами, в частности *полей*. Поля — это коммутативные кольца, в которых ненулевые элементы образуют группу относительно умножения. К этому классу принадлежат не только вещественные числа $\mathbf{R}$, комплексные числа $\mathbf{C}$, рациональные числа $\mathbf{Q}$, но также поля $\mathbf{Z}_p$ целых чисел по модулю простого числа p и многие другие конечные поля. Мы покажем, что такие поля, особенно поля «характеристики 2», порядок которых является степенью двойки, очень полезны для построения кодов.

Мы изучим также различные промежуточные классы колец, в частности области целостности и евклидовы области. В евклидовых областях деление не всегда выполнимо, но возможно «деление с остатком» и верна теорема о разложении в произведение степеней простых элементов. К важнейшим евклидовым областям относятся обыкновенные целые числа и многочлены от одной переменной с коэффициентами в любом поле.

Удобно изучать кольца аксиоматически: это позволяет выявить естественную степень общности каждого конкретного результата. Утверждая, что некоторая теорема верна, скажем, в любой евклидовой области, мы просто подразумеваем, что ее можно вывести из аксиом для евклидовых областей.

Начнем с доказательства нескольких простых результатов, справедливых для всех колец.

Некоторые элементарные свойства колец являются просто частными случаями свойств групп и моноидов. Например, для любых $a, b \in R$ уравнение $a+x=b$ имеет в R единственное решение:

$$x=(-a)+b. \tag{6}$$

Отсюда, в частности, вытекает, что:

(i) если $z+a=a$, то $z=0$;

(ii) если $a+x=0$, то $x=-a$;

(iii) если $a+b=a+c$, то $b=c$ (правило сокращения для сложения).

Как и в общих моноидах, единица 1 в кольце определена однозначно. Точнее:

$$\text{если } au=a \text{ для всех } a\in R, \text{ то } u=1. \tag{7}$$

Действительно, полагая $a=1$, находим $u=1u=1$.

Следующий результат использует закон дистрибутивности (1) — единственную аксиому, связывающую сложение с умножением.

Лемма 1. *«Аддитивная единица» 0 любого кольца R является одновременно «нулем» для умножения, т. е.*

$$0b=0=b0 \text{ для всех } b\in R. \tag{8}$$

Доказательство. Поскольку $a=a+0$ для всех $a\in R$, находим

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

Но по определению 0 мы имеем $ab=ab+0$. Следовательно, $ab+0b=ab+0$. Пользуясь правилом сокращения для сложения, выводим отсюда, что $0b=0$. Аналогично устанавливается, что $0=b0$.

Лемма 2. *Для всех $a, b\in R$*

$$(-a)(-b)=ab. \tag{9}$$

Доказательство. Так как $a+(-a)=0$, имеем

$$0=0\cdot(-b)=[a+(-a)](-b)=a(-b)+(-a)(-b).$$

С другой стороны,

$$a(-b)+ab=a[(-b)+b]=a0=0.$$

Применяя правило сокращения, получаем требуемый результат.

10.2. ОБЛАСТИ ЦЕЛОСТНОСТИ И ПОЛЯ

Понятие кольца является весьма общим, и потому кольца имеют не так уж много интересных свойств. По этой причине мы сосредоточим внимание на двух специальных классах колец,

о которых можно сказать значительно больше. К ним относятся, во-первых, поля (упомянутые во введении к этой главе) и, во-вторых, области целостности, важнейшее свойство которых состоит в том, что их можно расширять до полей (см. § 10.3). Начнем с формального определения.

Определение. *Областью целостности* называется *коммутативное кольцо*, в котором выполняется следующее правило сокращения:

$$\text{если } ab = ac \text{ и } a \neq 0, \text{ то } b = c. \tag{10}$$

Кольцо целых чисел $\mathbf{Z}$ и кольцо многочленов $\mathbf{Q}[x]$ от одной переменной с рациональными коэффициентами являются областями целостности. Кольцо $\mathbf{Z}_4$ классов вычетов по модулю 4 таковым не является, ибо $2 \cdot 2 = 2 \cdot 0$ в $\mathbf{Z}_4$, тогда как $2 \neq 0$.

Теорема 1. *Коммутативное кольцо R является областью целостности тогда и только тогда, когда произведение ненулевых сомножителей в нем всегда отлично от нуля, т. е.*

$$\text{если } a \neq 0, \ b \neq 0, \text{ то } ab \neq 0. \tag{11}$$

Доказательство. Если (11) неверно, то $ab = 0 = a0$ для подходящих a, $b \neq 0$, что противоречит (10) при $c = 0$. Обратно, если (10) неверно, то $a(b-c) = ab - ac = 0$, тогда как a и $b-c$ отличны от нуля.

Условие (11) можно перефразировать так: ненулевые элементы любой области целостности образуют подмоноид мультипликативного моноида умножения всех ее элементов.

Условие (11) можно записать в эквивалентной форме:

$$\text{если } ab = 0, \text{ то либо } a = 0, \text{ либо } b = 0. \tag{11'}$$

Отсюда вытекает

Следствие. *В области целостности нет идемпотентов, кроме 0 и 1.*

Доказательство. По определению, идемпотент u удовлетворяет уравнению $u^2 = u$, т. е.

$$0 = u^2 - u = u(u-1).$$

Из (11') следует, что либо $u = 0$, либо $u - 1 = 0$, и доказательство завершено.

Кольца $\mathbf{Z}_m$. Условие (11') в применении к случаю кольца $\mathbf{Z}_m = [\mathbf{Z}_m, +, \cdot]$ (см. § 2.6) означает, что $\mathbf{Z}_m$ является областью целостности тогда и только тогда, когда из $m \mid ab$ следует, что либо $m \mid a$, либо $m \mid b$ в $\mathbf{Z}$. Известно (см. также лемму 1 § 10.10), что числа $m \in \mathbf{P}$, обладающие указанным свойством,— это в точности *простые* числа. Отсюда вытекает

Лемма. *Кольцо* $\mathbf{Z}_m$ *является областью целостности в том и только том случае, если* m *— простое число.*

Определение. *Полем* называется коммутативное кольцо, у которого ненулевые элементы образуют группу относительно умножения.

Так как в любой группе справедливо правило сокращения, всякое поле является областью целостности.

Самые известные примеры полей — поле $\mathbf{Q}$ всех рациональных чисел, поле $\mathbf{R}$ всех вещественных чисел и поле $\mathbf{C}$ всех комплексных чисел. Менее известны поля $\mathbf{Z}_p$ классов вычетов по модулю простого числа p. Мы докажем, что $\mathbf{Z}_p$ является полем в теореме 2. Ниже даны таблицы сложения и умножения в $\mathbf{Z}_5$:

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

·	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

Рис. 10.1.

Из определения поля и свойств групп (гл. 7) вытекает

Лемма 3. *В любом поле деление на любой ненулевой элемент возможно, и притом единственным способом.*

Это свойство можно положить в основу определения не обязательно коммутативных полей с делением.

Теорема 2. *Любая конечная область целостности* D *является полем.*

Доказательство. По теореме 1 ненулевые элементы области целостности D образуют конечный моноид $[D^*, \cdot]$, в котором все правые сдвиги $x \mapsto xb$ и левые сдвиги $x \mapsto ax$ взаимно однозначны. Из конечности следует тогда, что они являются даже биекциями. В частности, $ax = xa = 1$ для некоторого $x \in D^*$. Значит, каждый ненулевой элемент $a \in D^*$ имеет обратный, так что D является полем.

Единственное решение x уравнения $bx = a$ в поле обозначается через a/b («частное от деления a на b»).

Теорема 3. *В любом поле частные подчиняются следующим правилам* ($b \neq 0$, $d \neq 0$):

(i) $a/b = c/d$ *тогда и только тогда, когда* $ad = bc$;

(ii) $a/b \pm c/d = (ad \pm bc)/bd$;

(iii) $(a/b)(c/d) = ac/bd$;
(iv) $a/b + (-a/b) = 0$;
(v) $(a/b)(b/a) = 1$, *если* $a, b \neq 0$.

Доказательство. Докажем все эти формулы по очереди.
(i) Если $a/b = c/d$, то

$$ad = a(b^{-1}b)d = (ab^{-1})(bd) = (cd^{-1})(db) = c(d^{-1}d)b = cb.$$

Здесь третье равенство следует из посылки, а остальные являются общими тождествами в коммутативных кольцах. Обратно, если $ad = bc$, то

$$ab^{-1} = a(dd^{-1})b^{-1} = (ad)(d^{-1}b^{-1}) = (cb)(b^{-1}d^{-1}) = c(bb^{-1})d^{-1} = cd^{-1}.$$

(ii) Пусть $bx = a$, $dy = c$. Тогда левая часть (ii) равна $x \pm y$. Имеем

$$bd(x \pm y) = bdx \pm bdy = bxd \pm bc = ad \pm bc.$$

Это означает, что правая часть (ii) также совпадает с $x \pm y$.

(iii) По определению, $(a/b)(c/d) = xy$, где $bx = a$, $dy = c$ и $b, d \neq 0$. Следовательно, $ac = (bx)(dy) = (bd)(xy)$ и $bd \neq 0$. Это означает, что $xy = ac/bd$.

(iv) Аналогично, $(a/b) + (-a/b) = x + y$, где $bx = a$ и $by = -a$. Значит, $0 = a + (-a) = bx + by = b(x + y)$. Так как $b \neq 0$, то, согласно теореме 1, отсюда следует, что $x + y = 0$.

(v) Наконец, $(a/b)(b/a) = xy$, где $bx = a$ и $ay = b$. Поэтому $ab = (bx)(ay) = b(xa)y = (ab)(xy)$. Но $ab = (ab)1$ и $ab \neq 0$ в силу (11) и наших предположений. Следовательно, $xy = 1$ по правилу сокращения (10), примененному к уравнению $(ab)(xy) = (ab)1$, в котором $ab \neq 0$. Это завершает доказательство.

УПРАЖНЕНИЯ А

1. Доказать, что в любом коммутативном кольце

$$(a+b)^n = a^n + \binom{n}{1} a^{n-1}b + \binom{n}{2} a^{n-2}b^2 + \ldots + b^n.$$

2. Выписать разложение $(a+b)^3$ в некоммутативном кольце.

3. а) Показать, что если A некоторая 2×2-матрица, коммутирующая с $\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$, то A диагональна.

б) Верен ли этот результат над любым полем? Над любым кольцом? Почему?

4. Доказать, что в любом коммутативном кольце для всех $m, n \in \mathbf{N}$

$$a^m a^n = a^{m+n}, \quad (ab)^n = a^n b^n, \quad (a^m)^n = a^{mn}.$$

5. Доказать, что для $a \neq 0$, $b \neq 0$ в любом поле и для всех $r, s \in \mathbf{Z}$

$$a^r a^s = a^{r+s}, \quad (ab)^r = a^r b^r, \quad (a^r)^s = a^{rs}.$$

Кольцо, в котором $x^2=x$ для всех x, называется булевым кольцом.

6. Доказать, что любое булево кольцо коммутативно.

7. Доказать, что $x+x=0$ для всех x в булевом кольце.

8. Показать, что булево кольцо, имеющее больше двух элементов, не может быть областью целостности. (*Указание*: рассмотреть уравнение $x(x-1)=0$.)

9. Рассмотрим одноэлементную систему $\{0\}$, где $0+0=0=0\cdot 0$. Является ли она кольцом? Областью целостности? Полем?

*10. Показать, что в определении кольца достаточно предполагать, что $[R, +]$ является (не обязательно коммутативной) группой. (*Указание*: разложить $(1+1)(a+b)$ двумя разными способами.)

*10.3. ПОЛЯ ЧАСТНЫХ

Формулы (i)—(v) теоремы 3 указывают способ, с помощью которого можно по любой данной области целостности D построить поле $Q=Q(D)$, называемое *полем частных* для D. Кольцо D канонически вкладывается в свое поле частных, и при вложении D в любое поле поле частных отождествляется с замыканием D относительно операции деления. В этом смысле $Q(D)$ является *минимальным* полем, содержащим D.

Мы начнем с формулировки основной теоремы. Само доказательство, хотя оно и приведено для полноты, может служить главным образом иллюстрацией использования техники отношений эквивалентности со свойством подстановки для построения новых алгебраических систем из уже известных.

Теорема 4. *Пусть D — область целостности, и пусть Q — множество формальных частных a/b, c/d, ... ее элементов с ненулевыми знаменателями. Определим отношение эквивалентности E на Q условием* (i) *теоремы 3. Тогда формулы* (ii)—(iii) *задают на фактормножестве Q/E операции сложения, вычитания и умножения, определенные этим отношением эквивалентности. Относительно указанных операций множество Q/E является полем. Оно называется полем частных кольца D и обозначается через $Q(D)$.*

Мотивировка. В случае $D=\mathbf{Z}$ эта конструкция является стандартным построением рациональных чисел из целых. Общий случай наиболее наглядно прослеживается на этом хорошо известном примере. В применении к кольцу $\mathbf{R}[x]$ всех многочленов с вещественными коэффициентами это построение доставляет поле $\mathbf{R}(x)=Q(\mathbf{R}[x])$ всех рациональных функций от одной переменной с вещественными коэффициентами (они определены всюду, где их знаменатели не обращаются в нуль). Конструкция имеет и другие важные приложения.

Доказательство. Отображение $a/b \mapsto (a, b)$ отождествляет множество формальных частных с множеством $D\times D^*$, где $D^*=D-\{0\}$. Формулы (ii) и (iii) теоремы 3 определяют систему

$[D\times D^*, +, \cdot]$ с двумя бинарными операциями «сложения» и «умножения». Заметим, что свойство (11) областей целостности используется, например, при доказательстве того, что дробь $ac/bd = (a/b)(c/d)$ имеет ненулевой знаменатель bd, т. е. что $(ac, bd) = (a, b)\cdot(c, d) \in D\times D^*$.

Система $[D\times D^*, +, \cdot]$ не является кольцом. Утверждается лишь, что если мы определим отношение эквивалентности E условием «$(a/b)\,E\,(c/d)$, если $ad = bc$», то классы эквивалентности уже будут образовывать коммутативное кольцо и даже поле. Мы проверим по очереди следующие утверждения:

(vi) E является отношением эквивалентности

(vii) Отношение E обладает свойством подстановки относительно $+$ и $\cdot$

(viii) Факторалгебра $(D\times D^*)/E$ является коммутативным кольцом с нулем $0/1$ и единицей $1/1$

(ix) $Q(D) = (D\times D^*)/E$ является полем, а отображение μ: $a \mapsto a/1$ определяет мономорфизм колец из D в $Q(D)$.

(vi) Рефлексивность E следует из того, что $ab = ba$. Симметричность E следует из того, что $ad = bc$ равносильно $cb = da$. Остается проверить транзитивность. Пусть $(a/b)\,E\,(c/d)$ и $(c/d)\,E\,(e/f)$. Тогда $ad = bc$ и $cf = de$, откуда $adf = bcf = bde$. Сокращая это соотношение на $d \neq 0$, получаем $af = be$, откуда $(a/b)\,E\,(e/f)$. Это доказывает, что E — отношение эквивалентности на $D\times D^*$. Оставшаяся часть доказательства относится к фактормножеству $(D\times D^*)/E$; в ней используется понятие свойства подстановки из § 2.5 и 2.6.

(vii). Как в § 2.6, свойство подстановки используется для доказательства того, что сложение и умножение классов эквивалентности, определенное на их представителях формулами (ii) и (iii) теоремы 3, однозначно. Иными словами, следует установить, что $(a/b) + (c/d)\,E\,(a/b) + (e/f)$ и $(a/b)(c/d)\,E\,(a/b)(e/f)$, если $(c/d)\,E\,(e/f)$. Мы оставляем доказательства читателю; они проводятся с помощью рассуждений, аналогичных тем, которые уже были использованы.

(vii) Из формул (ii), (iii) и теоремы 4 § 7.4 следует, что $(D\times D^*)/E$ является группой относительно $+$ и моноидом относительно $\cdot$. Нетрудно проверить, что

$$(0/1) + (a/b) = (0b + a)/b = a/b$$

и

$$(1/1)(a/b) = (a/b).$$

Остается доказать дистрибутивность. Стандартные выкладки показывают, что

$$(a/b)\,[(c/d) + (e/f)]\,E\,(acf + ade)/(bdf)$$

и

$$(a/b)(c/d) + (a/b)(e/f)\,E\,(acdf + bdae)/(b^2df).$$

Но для любого $b \neq 0$ в D и для всех a, $j \in D^*$ нетрудно убедиться, что $(a/j)\, E\, [(ab)/(bj)]$. Тем самым проверка дистрибутивности завершена.

(ix) Поскольку $(a/b)(b/a) = (ab/ba)$ и $(ab/ba)\, E\, (1/1)$, наше фактормножество является полем. Отображение μ, определенное в (ix), есть гомоморфизм колец, потому что

$$(a/1) + (b/1) = (a1 + b1)/(1 \cdot 1)\, E\, (a + b)/1$$

и

$$(a/1)(b/1) = (ab)/(1 \cdot 1)\, E\, (ab)/1.$$

Наконец, μ является мономорфизмом, ибо если $\mu(a) = 0/1$, то $a \cdot 1 = 0$.

10.4. ПОДКОЛЬЦА

Понятие подкольца аналогично понятиям подгруппы, булевой подалгебры или подмоноида. Подкольцом кольца R называется подмножество R, содержащее 0 или 1 и замкнутое относительно операций сложения, умножения и $x \mapsto -x$.

Можно потребовать несколько меньше:

Определение. *Подкольцом* S кольца R называется подмножество R, содержащее 1 и такое, что вместе с любыми двумя элементами x, y в него входят также разность $x - y$ и произведение xy.

Очевидно, отсюда следует, что подкольцо содержит $1 - 1 = 0$ и вместе с любыми x, y также $x - (0 - y) = x + y$.

Лемма 3. *Всякое подкольцо кольца R само является кольцом относительно операций в R.*

Доказательство. Подкольцо замкнуто относительно всех операций в R; аксиомы выполняются на всем R, а потому и на подкольце.

Ясно, что подкольцо коммутативного кольца коммутативно. Аналогично, подкольцо области целостности само является областью целостности. Поэтому такие подкольца называются также *подобластями*.

Подкольцо S поля F не обязано само быть полем. Если оно все же им является, оно называется *подполем*.

Для этого необходимо и достаточно, чтобы из $x \in S$ следовало, что $x^{-1} \in S$ для всех $x \neq 0$.

Подкольцо, порожденное единицей. Если кольцо R нетривиально (т. е. $R \neq \{0\}$), то любое его подкольцо содержит единицу 1 кольца R. Рассмотрим аддитивную подгруппу U груп-

пы $[R, +]$, порожденную 1. Она состоит из всех «кратных» единицы:

$$n1 = \underbrace{1 + \ldots + 1}_{n \text{ раз}} \ (n \in \mathbf{P}); \ (-n)\,1 = -(n1) = n(-1), \ 01 = 0. \quad (12)$$

Так как $1 \neq 0$, ее порядок не меньше двух.

Определение. *Характеристикой* кольца R называется порядок подгруппы, порожденной единицей в $[R, +]$.

Кольца $\mathbf{Z}$, $\mathbf{Q}$, $\mathbf{R}$ и $\mathbf{C}$ имеют характеристику ∞ [1]), кольцо $\mathbf{Z}_m$ — характеристику m. Если кольцо R имеет характеристику m, то то же верно и для любого его подкольца или надкольца. Проверку мы предоставляем читателю.

Лемма 4. *Характеристика кольца R совпадает с наименьшим положительным целым числом n (если оно существует), для которого*

$$na = a + \ldots + a = 0 \quad \textit{при всех} \quad a \in R. \quad (12')$$

Доказательство. Пусть m — характеристика R. Если она конечна, то $m1 = 0$, так что для всех $a \in R$

$$ma = a + \ldots + a = 1a + \ldots + 1a =$$
$$= (1 + \ldots + 1)\,a = (m1)\,a = 0a = 0,$$

где суммы состоят из m членов. Поэтому (12) выполняется для $n = m$. С другой стороны, если $0 < n < m$, то $n1 \neq 0$. Это доказывает лемму 4.

Из определения ясно, что отображение $n1 \leftrightarrow n \pmod m$ является изоморфизмом групп $[\mathbf{Z}_m, +]$ и $[U, +]$. Покажем, что оно является также изоморфизмом *колец*, т. е. что $(n1)(r1) = (nr)\,1 = (nr)_m 1$ и $1 \cdot 1 = 1$. Второе равенство тривиально. Доказательство первого сводится к разбору девяти случаев, в зависимости от того, положительным, отрицательным или нулевым является значение n и r соответственно. Если $n > 0$ и $r > 0$, то

$$(n1)(r1) = \overbrace{(1 + \ldots + 1)}^{n \text{ раз}}\,\overbrace{(1 + \ldots + 1)}^{r \text{ раз}} = \overbrace{(1 + \ldots + 1)}^{nr \text{ раз}} = (nr)\,1. \quad (13)$$

Аналогично если $r = -s$ $(s > 0)$, то

$$(n1)(r1) = (n1)(s(-1)) = ns[1(-1)] = ns(-1) = (-ns)\,1 = (nr)\,1. \quad (13')$$

Остальные случаи разбираются аналогично.

Резюмируем доказанное.

---

[1]) В случае когда группа, порожденная единицей, бесконечна, принято говорить, что кольцо имеет характеристику нуль, а не бесконечность.— *Прим. перев.*

Теорема 5. *Пусть R — кольцо характеристики m. Тогда его подкольцо, порожденное единицей, изоморфно $\mathbf{Z}_m$ и содержится в любом подкольце кольца R.*

Если R — область целостности, то это подкольцо не может содержать делителей нуля. Поэтому имеет место

Теорема 6. *Характеристика любой области целостности является либо простым числом, либо ∞.*

Следствие. *Подкольцо, порожденное единицей в любом конечном поле, изоморфно $\mathbf{Z}_p$ для подходящего простого числа p и потому является подполем.*

Определение. *Центром* кольца R называется множество всех элементов $a \in R$, для которых

$$ax = xa \text{ при всех } x \in R. \tag{14}$$

Центр коммутативного кольца R совпадает с R. В § 10.8 мы покажем, что если F — поле, то центр некоммутативного кольца $R = M_n(F)$ состоит из скалярных матриц $a_{ij} = \lambda\delta_{ij}$, $\lambda \in R$.

Теорема 7. *Центр любого кольца R является его подкольцом.*

Действительно, если $ax = xa$ и $bx = xb$ для всех x, то

$$(a \pm b)x = ax \pm bx = xa \pm xb = x(a \pm b),$$
$$(ab)x = a(bx) = a(xb) = (ax)b = (xa)b = x(ab).$$

Кроме того, $1x = x1$ для всех x.

УПРАЖНЕНИЯ Б

1. Доказать, что следующие подмножества являются подкольцами в кольце матриц $M_n(R)$:
   а) множество диагональных матриц $\|a_{ij}\|$ с $a_{ij} = 0$ при $i \neq j$;
   б) множество треугольных матриц $\|a_{ij}\|$ с $a_{ij} = 0$ при $i < j$.

2. Доказать, что характеристики колец R и $M_n(R)$ совпадают.

3. Является ли множество всех $n \times n$-матриц отношений кольцом относительно сложения mod 2 и композиции?

4. Какое подкольцо $\mathbf{Q}$ порождено 1/3? 1/3 и 1/4?

5. а) Какое подкольцо $\mathbf{R}$ порождено $\sqrt{2}$?
   б) Показать, что $\mathbf{Q}[\sqrt{2}] = \{a + b\sqrt{2} \mid a, b \in \mathbf{Q}\}$ является подполем в $\mathbf{R}$.

6. Какое подкольцо $\mathbf{C}$ порождено i?

7. Какое подкольцо $\mathbf{C}$ порождено $\mathbf{Q}$ и i? Является ли оно полем?

8. Доказать со всеми подробностями утверждение (v) теоремы 3 (можно считать, что утверждения (i)—(iv) уже доказаны).

9. Доказать, что неотрицательные рациональные числа представляются в виде суммы нескольких квадратов рациональных чисел.

*10. Пусть S—коммутативная полугруппа, в которой из $a+x=a+y$ следует, что $x=y$ (правило сокращения). Доказать, что S можно вложить в группу G. (*Указание:* построить G в виде $\{a-b\}/E$, $a, b \in S$, где E—надлежащим образом определенное отношение эквивалентности.)

10.5. МОРФИЗМЫ КОЛЕЦ

Мы уже занимались морфизмами групп, моноидов, булевых алгебр и решеток. В этом параграфе мы определим морфизмы колец, приведем несколько примеров и выведем элементарные свойства морфизмов.

В следующем параграфе будет показано, что морфизмы колец связаны с прямыми суммами так же, как морфизмы групп. В § 10.7 мы обнаружим, что эпиморфные образы любого кольца R можно восстановить (с точностью до изоморфизма) по идеалам в этом кольце, аналогично тому, как эпиморфные образы группы восстанавливаются по ее нормальным подгруппам.

Начнем с определения морфизмов.

Определение. Пусть R и S—кольца. *Морфизмом* θ: $R \to S$ называется отображение, для которого

$$\theta(x+y)=\theta(x)+\theta(y),\ \theta(1)=1,\ \theta(xy)=\theta(x)\theta(y)$$

при всех $x, y \in R$. (15)

Лемма 1. *Пусть* θ—*морфизм колец. Тогда* $\theta(0)=0$, $\theta(-x)=-\theta(x)$ *и, если* x *обратим,* $\theta(1/x)=1/\theta(x)$.

Доказательство. В силу (15), $\theta(0)=\theta(0+0)=\theta(0)+\theta(0)$. По правилу сокращения отсюда следует, что $\theta(0)$ есть нуль в S. Далее, если x обратим в R, то

$$1=\theta(1)=\theta(x\cdot(1/x))=\theta(x)\theta(1/x),$$

так что $\theta(x)$ также обратим и $[\theta(x)]^{-1}=\theta(x^{-1})$. Аналогично проверяется, что $\theta(-x)=-\theta(x)$.

Понятия мономорфизма, эпиморфизма, изоморфизма, эндоморфизма и автоморфизма определяются точно так же, как для других алгебраических систем.

Следующее рассуждение обобщает доказательство (из § 2.6) того, что $\mathbf{Z}_n$ является кольцом.

Теорема 8. *Пусть* R—*кольцо,* $[S, +, \cdot]$—*некоторая система со сложением и умножением и* θ: $R \to S$ *эпиморфизм. Тогда* S—*кольцо.*

Доказательство. Очевидное обобщение теоремы 4 гл. 7 показывает, что S—абелева группа относительно сложения. По аналогичной причине S является моноидом с единственной единицей 1 (см. гл. 7) относительно умножения. Наконец, дистрибутивность в S устанавливается следующим способом. Пусть a,

$b, c \in S$. Существуют такие $x, y, z \in R$, что $\theta(x) = a$, $\theta(y) = b$ и $\theta(z) = c$. Тогда

$$a(b+c) = \theta(x)[\theta(y) + \theta(z)] = \theta[x(y+z)] = \theta(xy + xz),$$

ибо θ — морфизм, а R — кольцо.

Далее, по тем же причинам

$$\theta[xy + xz] = \theta(x)\theta(y) + \theta(x)\theta(z) = ab + ac.$$

Теорема 9. *В любой области целостности D простой характеристики p отображение $x \to x^p$ является мономорфизмом* μ: $D \to D$.

Доказательство. По формуле бинома для любого $n \in \mathbf{P}$

$$(a \pm b)^n = a^n \pm \binom{n}{1} a^{n-1}b + \binom{n}{2} a^{n-2}b^2 \pm \ldots \pm \binom{n}{n} (\pm b)^n. \quad (16)$$

Равенство (16) без труда устанавливается индукцией по n. Умножение элемента на биномиальный коэффициент, как и вообще на натуральное число, означает в любом кольце суммирование соответствующего количества экземпляров этого элемента. Следовательно, в нашей сумме останутся лишь крайние члены, если мы проверим, что $\binom{n}{k}$ делится на p при всех $1 < k < p$. Но $p!$ делится на p, тогда как ни $k!$, ни $(p-k)!$ не делятся на p при этих k. Таким образом,

$$(a \pm b)^p = a^p \pm b^p, \text{ если } \operatorname{char} R = p. \quad (17)$$

С другой стороны,

$$(ab)^p = a^p b^p \text{ в любом коммутативном кольце} \quad (17')$$

(и даже в любом коммутативном моноиде).

Из формул (17) и (17′) следует, что μ является морфизмом колец. Мы завершим доказательство теоремы 9, показав, что он взаимно однозначен и потому является мономорфизмом.

Действительно, если $x^p = y^p$, то $(x-y)^p = x^p - y^p = 0$. Так как D — область целостности, отсюда вытекает, что $x - y = 0$, т. е. $x = y$, как и утверждалось.

10.6. ПРЯМЫЕ СУММЫ

В предыдущих главах мы определили «прямые произведения» групп, решеток, булевых алгебр и т. д. В этом параграфе мы опишем их аналог для колец. Хотя речь по-прежнему будет идти о декартовом произведении множеств элементов, получившееся кольцо принято называть прямой суммой, а не произведением.

Определение. Пусть R и S—кольца. Их *прямой суммой* $R \oplus S$ называется множество упорядоченных пар (x, y), $x \in R$, $y \in S$ с операциями, определенными следующим образом:

$$(r, s) \pm (r', s') = (r \pm r', s \pm s'), \tag{18}$$

$$(r, s)(r', s') = (rr', ss'). \tag{18'}$$

Единицей в $R \oplus S$ служит пара $(1_R, 1_S)$, где 1_R—единица в R, а 1_S—единица в S.

Иными словами, как аддитивная группа $[R \oplus S, +]$ является прямым произведением аддитивных групп $[R, +]$ и $[S, +]$. Аналогично мультипликативный моноид $[R \oplus S, \cdot]$ есть прямое произведение моноидов $[R, \cdot]$ и $[S, \cdot]$. Чтобы убедиться в том факте, что наша конструкция приводит к кольцу, нужно еще проверить два закона дистрибутивности. Вот одна из проверок:

$$\begin{aligned}(r, s)[(r', s') + (r'', s'')] &= (r(r' + r''), s(s' + s'')) = \\ &= (rr' + rr'', ss' + ss'') = \\ &= (r, s)(r', s') + (r, s)(r'', s'').\end{aligned}$$

Прямые суммы колец связаны некоторыми каноническими морфизмами со своими слагаемыми так же, как это имеет место для других алгебраических систем. А именно существуют очевидные эпиморфизмы ρ: $(r, s) \mapsto r$ и σ: $(r, s) \mapsto s$ из $R \oplus S$ на R и на S. Обратно, если дано любое кольцо T и два морфизма α: $T \to R$, β: $T \to S$, то существует такой морфизм τ: $T \to R \oplus S$, что $\rho \circ \tau = \alpha$ и $\sigma \circ \tau = \beta$, а именно $\tau(t) = (\alpha(t), \beta(t))$. Это утверждение называется *свойством универсальности* для прямых сумм. Оно иллюстрируется диаграммой на рис. 10.2.

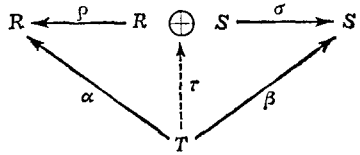


Рис. 10.2. Морфизмы и прямые суммы.

Заметим, что множество элементов вида $(r, 0)$ в $R \oplus S$ *не является* подкольцом, хотя оно изоморфно R как система со сложением и умножением. Это объясняется тем, что указанное множество не содержит единицы $(1_R, 1_S)$ кольца $R \oplus S$. Элементы $e = (1_R, 0)$ и $e' = (0, 1_S)$ кольца $R \oplus S$ обладают интересными свойствами:

(i) e и e' *идемпотентны*: $ee = e$ и $e'e' = e'$;
(ii) e и e' лежат в *центре* $R \oplus S$;

(iii) $e+e'=1$: элементы $e'=1-e$ и $e=1-e'$ являются дополнительными;

(iv) $ee'=0$.

Обратно, пусть e—любой идемпотент, лежащий в центре некоторого кольца A, и пусть $e'=1-e$. Тогда

$$e'^2=(1-e)^2=(1-e)(1-e)=1-e-e+e^2=1-2e+e=1-e=e',$$

так что e' также является идемпотентом. Далее, для всякого $x\in A$

$$xe'=x(1-e)=x-xe=x-ex=(1-e)x=e'x,$$

так что e' лежит в центре A. Наконец, отображение $x\mapsto(xe, xe')$ определяет изоморфизм кольца A с кольцом $R\oplus S$, где $R=[(e), +, \cdot, e]$, $S=[(e'), +, \cdot, e']$ суть множества всех кратных e и e' (главные идеалы), рассматриваемые как кольца с единицами e и $e'=1-e$ соответственно. Итак, мы доказали следующую теорему.

Теорема 10. *Разложения $A=R\oplus S$ данного кольца A в прямую сумму главных идеалов биективно соответствуют разложениям 1 в сумму двух идемпотентов e и $e'=1-e$, лежащих в центре A.*

Аналогичные результаты справедливы для прямых сумм любого числа колец. Разложение $A=R_1\oplus\ldots\oplus R_h$ отвечает выбору h ненулевых идемпотентов $e_1, \ldots, e_h$, лежащих в центре A и таких, что

$$e_i^2=e_i,\ e_ie_j=0 \quad \text{при } i\neq j,\ e_1+\ldots+e_h=1. \tag{19}$$

Кроме того, имеются канонические изоморфизмы:

$$R\oplus S\cong S\oplus R,\ R\oplus(S\oplus T)\cong(R\oplus S)\oplus T \text{ и т. д.} \tag{20}$$

Если все R_i являются копиями *одного* кольца R, то вместо $R\oplus\ldots\oplus R$ (h раз) мы пишем R^h. В силу (20)

$$R^h\oplus R^k=R^{h+k} \text{ для любых } h,\ k\in\mathbf{P}. \tag{21}$$

Булевы кольца. Рассмотренные выше понятия имеют полезные применения к специальному классу коммутативных колец характеристики 2, называемых *булевыми кольцами*, которые тесно связаны с булевыми алгебрами. Эта связь подсказывается рассмотрением с двух точек зрения множества всех двоичных векторов $\mathbf{x}=(x_1, \ldots, x_h)$ длины h. Это множество можно рассматривать либо как булеву алгебру $[2^h, \wedge, \vee, ', O, I]$, либо как (булево) кольцо $\mathbf{Z}_2^h$ с операциями

$$\mathbf{x}\cdot\mathbf{y}=\mathbf{x}\wedge\mathbf{y}=(x_1y_1, \ldots, x_hy_h)=(x_1\wedge y_1, \ldots, x_h\wedge y_h) \ldots$$

и

$$\mathbf{x}+\mathbf{y}=(\mathbf{x}\wedge\mathbf{y}')\vee(\mathbf{x}'\wedge\mathbf{y}).$$

Заметим, что умножение в кольце идемпотентно: так как $x_i^2=x_i$ для всех $i=1, \ldots, h$, то $\mathbf{x}^2=\mathbf{x}$ для всех $\mathbf{x}$. Аксиоматизируя эту ситуацию, мы приходим к следующему определению.

Определение. *Булевым кольцом* A называется кольцо, в котором все элементы идемпотентны:

$$a^2=a \text{ для всех } a\in A. \tag{22}$$

Лемма. *Всякое булево кольцо коммутативно и имеет характеристику два.*

Доказательство. Имеем $1+1=(1+1)^2=(1+1)+(1+1)$, откуда по правилу сокращения $0=1+1$. Значит, характеристика A равна двум. Аналогично, для любых a и b

$$a+b=(a+b)(a+b)=aa+ab+ba+bb=a+b+ab+ba.$$

Сокращая, находим $ab+ba=0$; но, кроме того, $ab+ab=0$. Отсюда $ba=ab$.

Следствие. *Всякое булево кольцо является полурешеткой относительно умножения.*

Связь булевых колец с булевыми алгебрами описывается следующей теоремой.

Теорема 11. *Пусть $\mathcal{A}$—класс булевых колец, $\mathcal{B}$—класс булевых алгебр с точностью до изоморфизма. Тогда замена законов композиции*

$$a\wedge b=ab, \quad a\vee b=a+b+ab, \tag{23}$$

$$ab=a\wedge b, \quad a+b=(a\wedge b')\vee(a'\wedge b) \tag{23'}$$

определяет взаимно обратные биекции α: $\mathcal{A}\to\mathcal{B}$ и β: $\mathcal{B}\to\mathcal{A}$.

Доказательство. Пусть сначала $A=[S, +, \cdot]$—булево кольцо с единицей 1. По предыдущему следствию система $B=[S, \wedge]$, построенная из A по формулам (23), будет полурешеткой по булеву умножению $\wedge$. Чтобы проверить, что $[S, \wedge, \vee]$ является решеткой, достаточно установить эквивалентность равенств $a+b+ab=a$ и $ab=b$ в A, а это очевидно.

Аналогично проводится и обратное рассуждение.

10.7. ИДЕАЛЫ И ФАКТОРКОЛЬЦА

В гл. 7 мы показали, что каждый эпиморфный образ любой группы G изоморфен *факторгруппе* G/N по подходящей *нормальной подгруппе* N и что все нормальные подгруппы $N\lhd G$

отвечают некоторым эпиморфизмам. В этом параграфе будут доказаны аналогичные результаты для колец. Роль нормальных подгрупп в кольцах играют идеалы.

Определение. *Идеалом* в кольце R называется непустое подмножество $H \subset R$, такое, что:
(i) если $x \in H$ и $y \in H$, то $x \pm y \in H$;
(ii) если $x \in H$ и $y \in R$, то $xy \in H$ и $yx \in H$.

В любом кольце R все кольцо R и подмножество $\{0\}$ являются идеалами по очевидным причинам. Отличные от $\{0\}$ и R идеалы называются *собственными идеалами*.

Пусть θ: $R \to S$ — любой морфизм колец. Прообраз нуля $\theta^{-1}(0)$ в R называется *ядром* θ.

Теорема 12. *Ядро любого морфизма колец* θ: $R \to S$ *является идеалом в* R.

Доказательство. Рассматривая θ как морфизм аддитивных групп $[R, +]$ в $[S, +]$, из теоремы 18 гл. 7 выводим, что ядро θ является аддитивной подгруппой в R. Это доказывает условие (i) в определении идеала. Далее, если $x \in H$ и $y \in R$, то, по определению морфизма,

$$\theta(xy) = \theta(x)\theta(y) = 0 \cdot \theta(y) = 0.$$

Поэтому $xy \in H$. Доказательство включения $yx \in H$ аналогично.

Важный класс идеалов в коммутативных кольцах строится на основе следующего легко доказываемого факта.

Теорема 13. *В коммутативном кольце* R *множество* $(a) = \{xa\}$ *всех кратных* $xa = ax$ *любого фиксированного элемента* $a \in R$ *является идеалом в* R.

Доказательство. Если $xa \in (a)$ и $ya \in (a)$, то $xa \pm ya = (x \pm y)a \in (a)$; кроме того, для любого элемента $r \in R$ имеем $r(xa) = (rx)a \in R$ и $(xa)r = a(xr) \in R$.

Идеалы (a) называются *главными идеалами*. Если коммутативное кольцо R не является полем, то оно содержит некоторый необратимый элемент $a \neq 0$, среди кратных которого нет единицы. Поэтому (a) является собственным идеалом в R.

Напротив, *поля не имеют собственных идеалов*. В самом деле, предположим, что H — собственный идеал поля F.

Пусть $h \in H$ — ненулевой элемент. Так как F — поле, h^{-1} имеется в F; значит, H должен содержать $(xh^{-1})h$ для любого $x \in F$; поэтому $H = F$ — противоречие. Таким образом, поле F не имеет идеалов, кроме $F = (1)$ и $\{0\}$.

Следовательно, класс коммутативных колец, не имеющих собственных идеалов, совпадает с классом полей. Покажем на примере, что в некоммутативном случае положение дел сложнее.

Пример 3. Рассмотрим полное матричное кольцо $M_n(F)$ (пример 2 § 10.1), состоящее из всех $n \times n$-матриц $A = \|a_{ij}\|$ с элементами из поля F. Пусть в этом кольце E^{hk} означает матрицу, у которой на месте (ij) стоит 1, если $h = i$, $k = j$, и нуль в остальных случаях; $e_{ij}^{hk} = \delta_{hi}\delta_{jk}$. Пусть $H \subset M_n(F)$ является некоторым ненулевым идеалом и $A \neq 0$ — матрица из этого идеала с ненулевым элементом a_{ij}. Тогда идеал H должен содержать все матрицы $E^{hi}AE^{jk} = a_{ij}E^{hk}$. Выберем любую матрицу $B = \|b_{kh}\|$. Так как $a_{ij} \neq 0$, мы можем положить $c_{hk} = a_{ij}^{-1}b_{hk}$ и затем, обозначив через $c_{hk}I$ матрицу с элементами c_{hk} на диагонали и нулями вне нее, мы можем представить B в виде

$$B = \sum_{h,\,k} (c_{hk}IE^{hi}AE^{jk}).$$

Таким образом, $H = M_n(F)$. Отсюда следует, что полное матричное кольцо над полем F не имеет собственных идеалов.

Перейдем теперь к двум основным результатам, обращающим теорему 12. Они гарантируют существование и единственность эпиморфизмов.

Теорема 14. *Пусть* θ: $R \to S$ *и* θ': $R \to S'$ *— два эпиморфизма колец с общей областью* R *и ядром* H. *Тогда их образы* S *и* S' *канонически изоморфны.*

Доказательство. Прежде всего, по теореме 18 гл. 7 имеется канонический изоморфизм b: $S \leftrightarrow S'$ аддитивных групп: $x \leftrightarrow x'$, если $\theta^{-1}(x)$ и $\theta'^{-1}(x')$ совпадают. Остается проверить, что этот же изоморфизм совместим с умножением и потому является изоморфизмом колец. По определению морфизма колец, если $\theta^{-1}(x) = H + a$ и $\theta^{-1}(y) = H + b$, то $\theta^{-1}(xy)$ содержит ab, так что $\theta^{-1}(xy) = H + ab$. Аналогично если $x' = b(x)$ и $y' = b(y)$, то из $\theta'^{-1}(x') = H + a$ и $\theta'^{-1}(y') = H + b$ следует, что $\theta'^{-1}(x'y') = H + ab$. Следовательно, $b(xy) = b(x'y')$, так как прообразы обоих элементов одинаковы.

Прежде чем переходить ко второму результату, докажем нетрудную лемму.

Лемма. *Пусть* H *— аддитивная подгруппа кольца* R. *Тогда разбиение* R *на смежные классы по модулю* H *удовлетворяет условиям подстановки*:

$$\text{если } a \equiv b \pmod{H}, \text{ то } ar \equiv br \pmod{H} \text{ для всех } r \in R \quad (24)$$

в том и только в том случае, если H *— правый идеал.*

Доказательство. Условие (24) равносильно тому, что из $a - b \in H$ следует $ar - br = (a - b)r \in H$ для всех $r \in R$. Но это и означает, что H — правый идеал.

Теорема 15. *Пусть H — идеал в кольце R. Тогда множество аддитивных смежных классов $H+x$ образует факторкольцо R/H с операциями*

$$(H+x)+(H+y)=H+(x+y), \tag{25}$$

$$(H+x)\cdot(H+y)=H+(xy). \tag{25'}$$

Кроме того, отображение $x\mapsto x+H$ является эпиморфизмом кольца R на R/H.

Доказательство. В абелевой группе $[R, +]$ любая подгруппа H нормальна. Поэтому (25) определяет абелеву группу R/H, а отображение $x\mapsto x+H$ является эпиморфизмом аддитивных абелевых групп. Остается проверить, что формулы (25') однозначно определяют умножение на аддитивных смежных классах по H. Иными словами, мы должны показать, что разбиение R на смежные классы по H обладает свойством подстановки для умножения: если $H+x=H+x'$ и $H+y=H+y'$, то $H+xy=H+x'y'$. Действительно, пусть $x'=h+x$ и $y'=h'+y$ $(h, h'\in H)$. Тогда

$$H+x'y'=H+(h+x)(h'+y),$$

$$H+hh'+hy+xh'+xy=H+xy,$$

потому что $(hh'+hy+xh')\in H$ согласно определению идеала.

УПРАЖНЕНИЯ В

1. Подробно разобрать свойства морфизмов, указанных на рис. 10.2.

2. Показать, что в любом булевом кольце R для любого элемента $a\neq 0, 1$ отображение $x\mapsto(xa, x-xa)$ определяет разложение R в прямую сумму $R=A\oplus B$ двух (собственных) эпиморфных образов R.

3. Показать, что любое двухэлементное булево кольцо изоморфно $[\mathbf{Z}_2, +, \cdot]$.

4. Показать, что любое конечное булево кольцо изоморфно прямой сумме $R\cong \mathbf{Z}_2\overbrace{\oplus\ldots\oplus}^{r}\mathbf{Z}_2$ для некоторого $r\in\mathbf{N}$.

5. Построить изоморфизм колец $\mathbf{Z}_2\oplus\mathbf{Z}_3\cong\mathbf{Z}_6$.

6. а) Показать, что смежные классы по идеалам обладают следующим свойством: если $x\in H+a$ и $y\in H+b$, то $H+xy=H+ab$.

б) Интерпретировать это свойство как свойство подстановки для умножения.

7. Пусть A — абелева группа. Показать, что ее эндоморфизмы $\alpha, \beta, \gamma, \ldots$ образуют кольцо со сложением $(\alpha+\beta)(x)=\alpha(x)+\beta(x)$ и умножением $(\alpha\beta)(x)=\alpha(\beta(x))$.

8. Показать, что морфизм μ теоремы 4 обладает следующим универсальным свойством: если $\varphi\colon D\to F$ — любой мономорфизм области D в поле F, то существует такой мономорфизм колец $\theta\colon Q(D)\to F$, что $\varphi=\theta\circ\mu$.

9. Показать, что в определении *эпиморфизма* колец условие $\theta(1)=1$ можно отбросить.

10.8. ДЕЛИМОСТЬ

Оставшаяся часть главы посвящена *коммутативным* кольцам. Наша основная цель состоит в доказательстве теоремы об однозначном разложении на простые элементы для класса евклидовых областей. В гл. 11 будет показано, что этот класс содержит кольца многочленов от одной переменной с коэффициентом в любом поле.

Начнем с некоторых общих сведений о разложениях, относящихся к любой области целостности D.

Напомним, что запись $a \mid b$ (a делит b) означает существование элемента $x \in D$, удовлетворяющего условию $ax = b$. Отношение $a \mid b$ рефлексивно (ибо $a1 = a$) и транзитивно, ибо из $ax = b$ и $by = c$ следует, что

$$a(xy) = (ax)y = by = c.$$

Далее, если $a \mid b$ и $a \mid c$, то $a \mid (b \pm c)$, ибо из $b = ax$ и $c = ay$ следует $b \pm c = ax \pm ay = a(x \pm y)$. [Это просто переформулировка утверждения о том, что все кратные a образуют идеал, «главный идеал (a)».]

Назовем элементы a и b *ассоциированными* и будем писать $a \sim b$, если $a \mid b$, и $b \mid a$ одновременно. Следующий результат очевиден.

Лемма 1. *В коммутативном кольце R имеем $a \mid b$ тогда и только тогда, когда $(a) \supset (b)$; $a \sim 1$ тогда и только тогда, когда a обратим.*

Лемма 2. *В области целостности условие $a \sim b$ равносильно тому, что $au = b$ для некоторого обратимого элемента u.*

Доказательство. Если $au = b$ и u обратим, то $bu^{-1} = auu^{-1}$, так что $a \sim b$. Обратно, если $ax = b$ и $by = a$, то $a(xy) = (ax)y = by = a = a1$, откуда $xy = 1$ по правилу сокращения; следовательно, x обратим.

Все эти результаты по существу тривиальны, как и многие другие очень общие результаты. Перейдем к обобщениям хорошо известной, но глубокой теоремы элементарной арифметики о существовании и единственности разложения целых чисел на простые множители.

Доказательство ее, восходящее к Евклиду, основано на алгоритме деления с остатком, уже обсуждавшемся в первой главе.

Алгоритм деления. Для данных целых чисел a, b, $b > 0$, существуют такие целые q и r, что

$$a = bq + r, \quad 0 \leqslant r < b. \tag{26}$$

Пользуясь им, легко доказать следующий важный результат:

Теорема 16. *Всякое непустое множество H целых чисел, замкнутое относительно вычитания, либо состоит из одного нуля, либо совпадает с множеством всех кратных своего наименьшего положительного элемента.*

Доказательство. Если $H \neq \{0\}$, то для подходящего $a \neq 0$ множество H должно содержать a и $-a$; в частности H должно содержать положительные целые числа. Пусть b — *наименьшее* такое число. Тогда, очевидно, $(b) \subset H$. Пусть $a \in H$ — любое число. Деля a на b с остатком, как в (26), находим $a - qb = r \in H$. Так как $0 \leqslant r < b$ и b — наименьшее положительное число в H, то $r = 0$. Следовательно, $(b) = H$, как и утверждалось.

Следствие 1. *Все идеалы в кольце* $\mathbf{Z}$ *главные.*

Объединяя следствие 1 и теорему 14, получаем

Следствие 2. *Любой эпиморфный образ кольца* $\mathbf{Z}$ *изоморфен одному из колец* $\mathbf{Z}_m$.

10.9. ЕВКЛИДОВЫ ОБЛАСТИ

Доказанные для $\mathbf{Z}$ утверждения могут быть значительно обобщены. Чтобы получить эти обобщения, нужно аксиоматизировать те свойства $\mathbf{Z}$, которые существенно использовались в доказательствах. Мы начнем со следующего определения.

Определение. *Евклидовой областью* называется область целостности D вместе с *нормой* v: $D^* \to \mathbf{N}$ (D^* — множество ненулевых элементов D, $\mathbf{N}$ — множество неотрицательных целых чисел), которая удовлетворяет следующим условиям:

$$v(xy) \geqslant v(x) \quad \text{для всех } x, y \in D^*; \tag{27}$$

$$\left.\begin{array}{l}\text{для всех } a \in D \text{ и } b \in D^* \text{ существует элемент } q \in D, \\ \text{такой, что } a = bq + r, \text{ где } r = 0 \text{ или } v(r) < v(b).\end{array}\right\} \tag{28}$$

Пример 4. Кольцо $\mathbf{Z}$ является евклидовой областью с нормой $v(a) = |a|$, абсолютное значение a. Действительно, $|xy| = |x| \cdot |y|$ и $|y| \geqslant 1$ при $y \neq 0$, откуда следует (27). Условие (28) следует из существования алгоритма деления.

Пример 5. Кольцо $\mathbf{Z}[\sqrt{-1}]$ *гауссовых целых чисел* является подобластью в $\mathbf{C}$, состоящей из всех комплексных чисел вида $m + n\sqrt{-1}$ $(m, n \in \mathbf{Z})$. Она является евклидовой областью с нормой $v(m + n\sqrt{-1}) = m^2 + n^2$.

Действительно, заметим, что $v(z) = |z|^2$ и $|zz'|^2 = |z|^2 \cdot |z'|^2$. Так как $m^2 + n^2 \geqslant 1$, если m, n целые и не равны одновременно нулю, получаем условие (27).

Для доказательства (28) рассмотрим множество всех кратных bq в комплексной плоскости, где b—фиксированное гауссово целое число, а q переменное. Оно представляет собой вершины квадратной решетки со стороной $|b|$. Поэтому любое число $a \in \mathbf{Z}[\sqrt{-1}]$ находится на расстоянии не более чем $|b|/\sqrt{2}$ от ближайшей точки решетки bq. Мы получаем условие (28) для $r = a - bq$, ибо $v(r) = |r|^2 \leqslant |b|^2/2 = v(b)/2$.

Пример 6. Многочлены от одной переменной $F[x]$ с коэффициентами в поле F образуют евклидову область относительно нормы «$v(p)$ = степень многочлена p» (степенью многочлена $a_0 + \ldots + a_n x^n$ называется наибольшее i с $a_i \neq 0$).

Эту область мы изучим в гл. 11. Заметим, что для нее $v(pq) = v(p) + v(q)$.

Установим теперь простой общий результат об евклидовых областях.

Лемма 1. *В любой евклидовой области D для $x, y \in D^*$*

$$v(xy) = v(x), \tag{29}$$

если y обратим, и $v(xy) > v(x)$ в противном случае.

Доказательство. Если y обратим, то $v(x) = v(xyy^{-1}) \geqslant v(x, y)$, а это вместе с (27) доказывает (29). Обратимость y, очевидно, равносильна тому, что $xy \mid x$. Поэтому осталось рассмотреть случай $xy \nmid x$. Но в этом случае в силу (28)

$$x = q(xy) + r, \text{ где } v(r) < v(xy).$$

Кроме того, $r = x - qxy = x(1 - qy)$, так что $v(r) \geqslant v(x)$ согласно (27). Наконец, $v(xy) > v(r) \geqslant v(x)$, и доказательство завершено.

Следствие. *В евклидовой области элемент x обратим тогда и только тогда, когда $v(x) = v(1)$.*

Например, обратимые элементы в $\mathbf{Z}$—это ± 1, в $\mathbf{Z}[\sqrt{-1}]$—это ± 1 и $\pm i$; наконец, обратимые многочлены—это ненулевые константы (многочлены нулевой степени).

Теорема 17. *В евклидовой области D все идеалы главные.*

Доказательство аналогично доказательству теоремы 16. Пусть H—любой идеал в D. Если $H = \{0\}$, результат тривиален. В противном случае H содержит элемент $b \neq 0$ с наименьшим возможным значением $v(b)$. Очевидно, $H \supset (b)$. Чтобы доказать обратное включение, возьмем произвольный элемент $a \in H$ и разделим его на b с остатком r. Случай $v(r) < v(b)$ в (28) невозможен, ибо $r \in H$, а b был выбран с наименьшей нормой. Поэтому обязательно $r = 0$, т. е. $a = bq \in (b)$, что доказывает требуемый результат.

Алгоритм Евклида. Из теоремы 17 следует, что сумма $(a)+(b)$ любых двух главных идеалов в D снова является главным идеалом. Обозначим его через (c). Ясно, что $c=sa+tb$ и c является делителем как a, так и b. Если d — любой другой общий делитель a и b, т. е. $a=qd$, $b=q'd$, то

$$c=sa+tb=sqd+tq'd=(sq+tq')\,d.$$

Поэтому c делится на любой общий делитель элементов a и b. Это дает основание называть c наибольшим общим делителем a и b в смысле следующего определения.

Определение. В области целостности D элемент c называется *наибольшим общим делителем* элементов a и b (символически $c=$ н.о.д. (a, b)), если он делит a и b и делится на любой другой общий делитель a и b.

Заметим, что любой элемент $cu=c'$, ассоциированный с любым наибольшим делителем a и b (здесь u обратим), сам также является наибольшим общим делителем. Например, согласно этому определению, в $\mathbf{Z}$ вместе с c наибольшим общим делителем будет также $-c$. Конечно, в $\mathbf{Z}$ символом (a, b) принято обозначать положительный н.о.д. (см. ниже программу его вычисления на АЛГОЛе).

Из теоремы существования н.о.д. в евклидовой области можно извлечь алгоритм для их вычисления, если известно, как вычислять q и r в (28). Этот алгоритм для $\mathbf{Z}$ называется *алгоритмом Евклида*. Опишем его, напомнив, что (a) означает главный идеал, состоящий из всех кратных числа a.

Если $a=0$, то $(a)+(b)=(b)$, так что можно положить $c=b$; аналогично, если $b=0$, можно положить $c=a$. В остальных случаях $a, b\in D^*$ и определены $v(a)$, $v(b)$. Не ограничивая общности, можно считать, что $v(b)\leqslant v(a)$.

Разделим a на b с остатком. Если $a=qb$, то $(a)\subset(b)$ и $(a)+(b)=(b)$, так что можно положить $c=b$. В противном случае $a=qb+r_1$, где $v(r_1)<v(b)$. Кроме того, $(a)+(b)=(r_1)+(b)$, где

$$\min\{v(r_1), v(b)\}=v(r_1)<\min\{v(a), v(b)\}.$$

Эту процедуру можно повторять, пока мы не придем к нулевому остатку r_k, причем

$$(a)+(b)=(b)+(r_1)=(r_1)+(r_2)=\ldots=(r_{k-1}).$$

Очевидно, $r_{k-1}=$ н.о.д. (a, b).

Ниже приведена программа на АЛГОЛе для вычисления положительного н.о.д. двух целых чисел a и b с помощью алгоритма Евклида. (Здесь для н.о.д. используется обозначение gcd. — *Перев.*)

```
begin integer a, b, gcd, aa, bb, temp;
```

```
    aa: = abs (a); bb: = abs (b);
    if aa = 0 ∨ bb = 0 then
    begin gcd: = aa + bb; go to F end;
    if aa < bb then begin temp: = aa; aa: = bb; bb: = temp
                    end;
D:  r: = aa − (aa ÷ bb) × bb;
    if r ≠ 0 then begin aa: = bb; bb: = r; go to D end;
    gcd: = b;
F:  end
```

УПРАЖНЕНИЯ Г

1. Показать, что в области целостности $\mathbf{Z}[x]$
   а) н. о. д. $(7x^3+1, 2x)=1$,
   б) не существует таких многочленов $a(x)$, $b(x)$, что
   $$a(x)(7x^3+1)+2xb(x)=1.$$

2. Показать, что отношение ассоциированности $a \sim b$ в любой области целостности обладает следующими свойствами подстановки:
   если $a \sim b$, то $ac \sim bc$;
   если $a \sim b$ и $c \sim d$, то $ac \sim bd$.

3. Показать, что если c, c' — два наибольших общих делителя элементов a, b некоторой области, то $c \sim c'$.

4. а) Пусть p — простое число в евклидовой области D. Доказать, что уравнение $x^2=p$ не разрешимо в поле частных $Q(D)$.
   б) Доказать, что $\sqrt[3]{3}$ — иррациональное число.

5. а) Показать, что $\sqrt{3} \notin \mathbf{Q}(\sqrt{2})$ (см. упр. Б5, б)).
   б) Показать, что биекция $a+b\sqrt{2} \leftrightarrow a+b\sqrt{3}$ не является изоморфизмом полей $\mathbf{Q}(\sqrt{2})$ и $\mathbf{Q}(\sqrt{3})$.
   *в) Показать, что изоморфизма полей $\mathbf{Q}(\sqrt{2}) \cong \mathbf{Q}(\sqrt{3})$ не существует.

6. Вычислить н. о. д. $(108996, 76219)=d$, представив его в виде $d = 108996s+76219t$ (s, $t \in \mathbf{Z}$).

7. а) Вычислить таблицу обратных элементов в $\mathbf{Z}_{37}$. (*Указание*: вычислить 2^{-1}, 3^{-1} и воспользоваться формулой $(xy)^{-1}=x^{-1}y^{-1}$.)

*10.10. ОБЛАСТИ С ОДНОЗНАЧНЫМ РАЗЛОЖЕНИЕМ

В любой области целостности D всякий ненулевой элемент x можно очевидным образом разложить в произведение $x=uy$ любого обратимого элемента u и $y=u^{-1}x$. Мы назовем *простым* всякий ненулевой элемент $p \in D^*$, который сам не обратим и не может быть разложен в произведение двух необратимых множителей. Если p, q — два простых элемента в D и $px=q$, т. е. $p \mid q$, то,

очевидно, x обратим, так что $p \sim q$. Поэтому простые элементы D разбиваются на классы по отношению ассоциированности так, что простые из разных классов не могут делить друг друга.

Вот несколько первых простых чисел в $\mathbf{Z}$: ± 2, ± 3, ± 5, ± 7, ± 11 и т. д. В $\mathbf{Z}$ имеет место известная теорема об однозначном разложении: каждое целое число однозначно разлагается в произведение степеней простых чисел (с точностью до замены p ассоциированным числом $-p$; в $\mathbf{Z}$ кроме 1 имеется еще только обратимый элемент -1). Мы покажем, что аналогичный результат справедлив в любой евклидовой области, в основном благодаря существованию алгоритма Евклида. Доказательство проводится по индукции и использует следующую лемму.

Лемма 1. *Пусть p — простой элемент евклидовой области D, и пусть $p \mid ab$. Тогда либо $p \mid a$, либо $p \mid b$.*

Доказательство. Допустим, что $p \nmid a$. Рассмотрим $c = $ н. о. д. (p, a). Так как $p \nmid a$, он должен быть делителем p, *не* ассоциированным с p и потому обратимым. С другой стороны, в силу алгоритма Евклида, $c = sa + tp$ для подходящих $s, t \in D$. Поэтому

$$b = c^{-1}cb = c^{-1}(sa + tp)b = c^{-1}sab + c^{-1}tpd.$$

Мы предположили, что $p \mid ab$; кроме того, $p \mid c^{-1}tpd$ (это очевидно); значит, $p \mid b$. Лемма 1 доказана.

Назовем элементы a, d в евклидовой области D *взаимно простыми* (и будем писать $(a, d) = 1$), если $\{sa + td\} = D$. Предполагая, что $(a, d) = 1$ и $d \mid ab$, точно так же, как в лемме, находим

$$b = (sa + td)b = sab + tdb,$$

откуда $d \mid b$ (ибо d делит оба члена суммы sab и tdb). Таким образом, если $(a, a) = 1$, то $d \mid b$. Из этого обобщения леммы 1 в качестве следствия получаем следующий результат.

Лемма 2. *Пусть $(a, a') = 1$ и $a \mid c$, $a' \mid c$. Тогда $aa' \mid c$.*

Доказательство. Поскольку $a \mid c$, имеем $ab = c$ для подходящего b. Тогда $a' \mid ab$ и $(a, a') = 1$. Значит, $a' \mid b$ по доказанному выше. Следовательно, $b = a'x$, так что $c = aa'x$ для подходящего $x \in D$. Это означает, что $aa' \mid c$.

Следствие. *Пусть $a_1, \ldots, a_n$ — попарно взаимно простые делители элемента c в евклидовой области D. Тогда $\left(\prod_{i=1}^{n} a_i\right) \mid c$.*

Лемма 3. *В любой евклидовой области D всякий ненулевой необратимый элемент a можно разложить в произведение простых сомножителей.*

Доказательство. Если a простой, утверждение очевидно. В противном случае $a = bc$, где элементы $b, c \in D^*$ необратимы.

Тогда, в силу леммы 1 § 10.9, $v(b) < v(a)$ и $v(c) < v(a)$. Проводя индукцию по значению нормы, мы можем считать, что b и c разлагаются в произведение простых элементов. То же верно тогда и для $bc = a$.

Лемма 3 составляет часть следующего основного результата о существовании и единственности разложений.

Теорема 18. *В любой евклидовой области всякий ненулевой необратимый элемент a можно представить в виде произведения обратимого элемента и степеней попарно неассоциированных простых элементов. В этом разложении классы простых элементов и их степени определены однозначно.*

Доказательство. Рассмотрим два разложения:

$$a = \prod_{i=1}^{m} p_i = \prod_{j=1}^{n} q_j, \tag{30}$$

где p_i, q_j простые (хотя бы одно разложение существует по лемме 3). Так как $q_n \mid a$ (в силу леммы 1 и индукции по n), q_n должен делить некоторый $p_{i(n)}$. Согласно первому абзацу этого параграфа, $q_n = up_{i(n)}$, где u обратим. Поэтому

$$a = p_{i(n)} \prod_{k=1}^{m-1} p_{i(k)} = p_{i(n)}\, u \prod_{j=1}^{n-1} q_j. \tag{31}$$

В силу закона сокращения находим

$$b = \prod_{k=1}^{m-1} p_{i(k)} = \prod_{j=1}^{n-1} q'_j, \qquad q'_1 = uq_1, \qquad q'_j = q_j, \qquad j > 1. \tag{32}$$

Обозначим через $P(m)$ утверждение, что для данного m в формуле (30) $n = m$ и простые p_i, q_j попарно ассоциированы. Случай $P(1)$ тривиален. Предполагая, что $P(m-1)$ верно, и применяя его к (32), получаем, что $m-1 = n-1$ и что простые $p_{i(k)}$, q'_j попарно ассоциированы. Возвращаясь к (31), получаем утверждение $P(m)$, которое тем самым доказано. Это теорема единственности.

Выбирая из каждого класса ассоциированных простых по одному элементу и собирая их вместе с учетом обратимых множителей u, мы можем переписать произведение (30) в виде

$$a = \prod \bar{u}_i p_i^{e_i} = u \prod p_i^{e_i},$$

где различные p_i уже не ассоциированы. Это — разложение в произведение степеней простых. Его единственность следует из предыдущего.

*10.11. ПРОСТЫЕ И МАКСИМАЛЬНЫЕ ИДЕАЛЫ

В § 10.7 мы показали, что эпиморфные образы кольца R изоморфны факторкольцам R/H по различным идеалам H. Ограничиваясь *коммутативными* кольцами R, мы рассмотрим два вопроса об этих факторкольцах: (i) факторы по каким идеалам являются областями целостности? (ii) полями?

Ответ дан в теореме 19. Начнем со следующих определений.

Определение. Идеал H называется *простым*, если из $ab \in H$ следует, что либо $a \in H$, либо $b \in H$. Идеал H в кольце R называется *максимальным*, если любой идеал H', промежуточный между H и R, $H \subset H' \subset R$, совпадает либо с H, либо с R.

Теорема 19. *Пусть R — коммутативное кольцо, H — его идеал.*

1) *R/H есть область целостности тогда и только тогда, когда идеал H простой.*

2) *R/H есть поле тогда и только тогда, когда идеал H максимальный.*

Доказательство. Факторкольцо R/H является областью целостности в том и только том случае, если оно не имеет делителей нуля: из $(H+a)(H+b) = H$ следует, что либо $H+a = H$, либо $H+b = H$. Но это условие равносильно утверждению, что из $ab \in H$ следует либо $a \in H$, либо $b \in H$. Оно справедливо, согласно определению, в точности для простых идеалов, что доказывает первое утверждение.

Рассмотрим теперь максимальный идеал H в кольце R и некоторый ненулевой элемент $H+a \in R/H$. Легко видеть, что множество $S = \{h+ax \mid h \in H,\ x \in R\}$ является идеалом. Этот идеал содержит H и строго больше H, потому что $a \notin H$. Так как H максимален, отсюда следует, что $S = R$. Значит, $1 = h+ax$ для некоторых $h \in H$, $x \in R$. Следовательно,

$$H+1 = H+h+ax = H+ax = (H+a)(H+x).$$

Поэтому всякий ненулевой элемент $a+H$ факторкольца R/H обратим, так что оно является полем.

Обратно, пусть R/H — поле. Обозначим через M любой идеал, строго больший, чем H. Пусть $a \in M$ и $a \notin H$. Так как R/H — поле, уравнение

$$(H+a)(H+x) = H+b$$

разрешимо для любого $b \in R$ (заметим, что $H+a \neq H$, ибо $a \notin H$). Следовательно, $H+ax = H+b$. Но $H+ax \subset M$, так как $H \subset M$

и $a \in M$. Значит, $M = M + b$ для любого $b \in R$, а это возможно лишь при $M = R$. Поэтому H является максимальным идеалом.

Следствие. *Всякий максимальный идеал прост.*

Это следует из теоремы 19, ибо поля являются областями целостности.

* 10.12. ГАУССОВ МЕТОД ИСКЛЮЧЕНИЯ

С точки зрения приложений одно из важнейших свойств полей состоит в том, что система n линейных уравнений с n неизвестными вида

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= b_1, \\ a_{21}x_2 + a_{22}x_2 + \ldots + a_{2n}x_n &= b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots & \\ a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n &= b_n, \end{aligned} \tag{33}$$

как правило, имеет одно явно вычислимое решение над полем, которому принадлежат коэффициенты системы.

Сначала рассмотрим подробнее случай $n = 2$.

Пусть a, b, c, d, e, f — константы, принадлежащие полю F. Мы ищем решения x, y системы

$$ax + by = e, \qquad cx + dy = f. \tag{34}$$

Умножив первое уравнение на d, второе на b и производя вычитание, мы получим $(ad - bc)x = de - bf$. Умножив первое уравнение на c, второе на a и производя вычитание, получим $(ad - bc)y = af - ce$. Значит, если определитель $ad - bc = \Delta$ матрицы коэффициентов $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ системы (34) отличен от нуля, то система имеет единственное (векторное) решение (x, y):

$$x = \frac{de - bf}{\Delta}, \qquad y = \frac{af - ce}{\Delta}, \qquad \Delta = ad - bc. \tag{35}$$

(Хотя в принципе этот единственный кандидат на решение исходной системы мог бы не быть решением, нетрудно проверить непосредственной подстановкой, что решением он является.)

Если определитель Δ системы равен нулю, то система (34) либо вообще не имеет решений, либо их число больше одного; матрица системы в этом случае называется вырожденной.

Аналогичные результаты имеют место в общем случае. Однако мы не будем развивать общую теорию определителей, а вместо этого опишем алгоритм исключения по Гауссу, который позволяет вычислять решения. Следует помнить, что в случае вещественных коэффициентов реальные вычисления являются приближенными.

Запишем общую систему m линейных уравнений с n неизвестным в виде

$$\sum_{j=1}^{n} a_{ij} \cdot x_j = b_i, \qquad i = 1, \ldots, m. \tag{36}$$

Будем рассуждать по индукции.

Если все коэффициенты a_{1j} первого уравнения обращаются в нуль, то либо $b_1 \neq 0$, и тогда (36) не имеет решений, либо же $b_1 = 0$, и тогда любое решение оставшихся $m-1$ уравнений будет решением и первого уравнения.

Если же, скажем, $a_{1j} \neq 0$ (при $F = \mathbf{R}$ удобно выбрать j с максимальным $|a_{1j}|$), то первое уравнение (36) равносильно уравнению

$$x_j = a_{1j}^{-1} \left(b_1 - \sum_{k \neq j} a_{1k} x_k \right). \tag{37}$$

В условиях (37) система оставшихся уравнений равносильна системе

$$\sum_{k \neq j} (a_{ik} - a_{ij} a_{1j}^{-1} a_{1k}) x_k = b_i - a_{ij} a_{1j}^{-1} b_1, \qquad i = 2, \ldots, m. \tag{38}$$

Это система $m-1$ линейных уравнений с $n-1$ неизвестными того же типа, что и (36).

Поэтому мы можем применить к ней ту же процедуру, исключив еще одну неизвестную, и т. д.

При $m = n$ после n-кратного повторения этого процесса мы, вообще говоря, получим уравнение, дающее значение некоторой неизвестной. Подставляя его в предыдущее уравнение, получим значение следующей неизвестной, и т. д. Мы опускаем описание того, что происходит в вырожденном случае.

Приведем программу на АЛГОЛе, реализующую этот алгоритм:

```
procedure gauss(u, a, y);
real array a, y; integer u;
comment Эта процедура предназначена для решения системы линейных уравнений методом последовательного исключения неизвестных. Матрица коэффициентов и свободных членов есть a. Число неизвестных есть u. Вектор решений есть y. Если система не имеет решений или имеет более одного решения, производится передача управления «go to error», где error — метка вне программы;
    begin
        integer i, j, k, m, n;
        n:= 0;
ck 0:   n:= n + 1;
        for k:= n step 1 until u do if a[k, n] ≠ 0 then go to ck1;
```

```
        go to error;
ck 1:   if k=n then go to ck 2;
        for m:=n step 1 until u+1 do
    begin
        temp:=a[n, m]; a[n, m]:=a(k, m); a[k, m]:=temp
    end;
ck 2:   for j:=u+1 step -1 until n do a[n, j]:=a[n, j]/a[n, m];
        for i:=k+1 step 1 until u do
        a[i, j]:=a[i, j]-a[i, n]×a[n, j];
        if n≠u then go to ck 0;
                for i:=u step -1 until 1 do
    begin y[i]:=a[i, u+1]/a[i, i];
                for k:=i-1 step -1 until 1 do
                a[k, u+1]:=a[k, u+1]-a[k, i]×y[i]
    end end;
```

УПРАЖНЕНИЯ Д

1. Доказать, что в евклидовой области $(m, n)=1$ тогда и только тогда, когда $(m, p_i)=1$ для всех простых делителей p_i элементов n.

2. Решить систему уравнений $x+2y=0$, $2x+y=1$
а) над $\mathbf{Z}_5$; б) над $\mathbf{Z}_3$; в) над $\mathbf{Z}_7$.

3. а) Доказать, что следующие таблицы описывают поле:

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

·	0	1	2	3
0	0	0	0	0
1	0	1	2	3
2	0	2	3	1
3	0	3	1	2

б) Обозначив это поле через $GF(4)$, найти мономорфизм из $\mathbf{Z}_2$ в $GF(4)$.
в) Решить систему из упр. 2 в $GF(4)$.

4. Выписать все невырожденные 3×3-матрицы над $\mathbf{Z}_2$.

СПИСОК ЛИТЕРАТУРЫ

1. Jacobson N., Lectures on Abstract Algebra, 3 vols., Van Nostrand 1953-61.
2. Jans J. P., Rings and Homology, Holt, 1968.
3. McCoy N. H., The Theory of Rings, Macmillan, 1964.
4. Zariski O., Samuel P., Commutative Algebra, 2 vols., Van Nostrand, 1959-60. (Русский перевод: Зарисский О., Самюэль П., Коммутативная алгебра, т. 1 и 2, ИЛ, М., 1963.)

11

ПОЛИНОМИАЛЬНЫЕ КОЛЬЦА И ПОЛИНОМИАЛЬНЫЕ КОДЫ

11.1. КОЛЬЦО $R[x]$

В этой главе рассматриваются только коммутативные кольца. Пусть R такое кольцо.

Определение. Стандартным *многочленом* (или *полиномом*) степени $\leqslant m$ от *неизвестной* x над *коммутативным* кольцом R называется выражение вида

$$a_0 + a_1 x + \ldots + a_m x^m = \sum_{k=0}^{m} a_k x^k, \quad a_k \in R. \tag{1}$$

Элементы a_k называются *коэффициентами* многочлена (1); все они или часть их могут быть нулевыми.

Многочлен (1) часто обозначается символически $a(x)$. При этом не имеется в виду, что многочлен рассматривается как функция из R в R (см. § 11.7). Стандартный многочлен (1) однозначно определяется массивом (вектором) своих коэффициентов:

$$a = (a_0, a_1, \ldots, a_m).$$

Такое представление многочленов используется в АЛГОЛе (см. гл. 4).

Каноническая форма многочлена (1) определяется следующим образом. Отыщем наибольшее k, такое, что $a_k \neq 0$, скажем, $k = n$, и напишем

$$a(x) = a_0 + a_1 x + \ldots + a_n x^n, \quad a_n \neq 0. \tag{2}$$

Если же все a_k обращаются в нуль, канонической формой является 0. Степенью многочлена $a(x)$ называется число n, определенное в (2), если оно существует; если каноническая форма $a(x)$ есть 0, степень по определению равна $-\infty$. Степень $a(x)$ обозначается $\deg a$.

Два стандартных многочлена называются *равными*, если у них одинаковая каноническая форма. Следующий результат очевиден.

Лемма 1. *Многочлены $a(x)$ и $b(x)$ равны тогда и только тогда, когда $a_k = b_k$ для всех k, при которых a_k и b_k определены, а все остальные a_k, b_k равны нулю.*

Степени равных многочленов одинаковы, и их коэффициенты с номерами, не превосходящими степени, попарно одинаковы.

Символом $R[x]$ обозначается множество всех многочленов от x с коэффициентами из кольца R.

Чтобы сложить два многочлена, достаточно сложить коэффициенты при одинаковых степенях x. Точнее, если степень $a(x)$ не превосходит m, а степень $b(x)$ не превосходит r, то их сумма $c(x)=a(x)+b(x)$ есть

$$c(x)=c_0+c_1x+\ldots+c_sx^s, \tag{3}$$

где $s=\max(m, r)$ и

$$c_k=\begin{cases} a_k+b_k & \text{при } k\leqslant \max(m, r) \text{ всегда,} \\ a_k & \text{при } m<k\leqslant r, \text{ если } m<r, \\ b_k & \text{при } r<k\leqslant m, \text{ если } r<m. \end{cases} \tag{3'}$$

Аналогично произведение $p(x)=a(x)b(x)$ определяется формулой

$$p(x)=p_0+p_1x+\ldots+p_{m+r}x^{m+r}, \tag{4}$$

где

$$p_k=\sum_{i+j=k}a_ib_j. \tag{4'}$$

Очевидно, при $k\leqslant\min(m, r)$ имеем $p_k=a_0b_k+\ldots+a_kb_0$. Кроме того, $p_{m+r}=a_mb_r$.

Пример 1. Пусть $R=\{0, 1\}=Z_2$, $f(x)=1+0\cdot x+1\cdot x^2$ и $g(x)=1+1\cdot x+1\cdot x^2$. Тогда

$$f(x)+g(x)=(1+x^2)+(1+x+x^2)=x$$

и

$$f(x)g(x)=(1+x^2)(1+x+x^2)=1+1\cdot x+1\cdot x^3+1\cdot x^4.$$

Имеют место тривиальные оценки:

$$\deg(a+b)\leqslant\max(\deg a, \deg b), \tag{5}$$

$$\deg(ab)\leqslant\deg a+\deg b. \tag{5'}$$

Поскольку в области целостности из $a_m\neq 0$ и $b_r\neq 0$ следует, что $a_mb_r\neq 0$, из (4) и (4') вытекает

Лемма 2. *Если D — область целостности, то в $D[x]$ имеем*

$$\deg(ab)=\deg a+\deg b.$$

Теорема 1. *Операции* (3), (3'), (4) *и* (4') *определяют на множестве $R[x]$ многочленов над коммутативным кольцом R структуру коммутативного кольца. Многочлены нулевой степени в $R[x]$ с нулем образуют подкольцо констант, изоморфное R.*

Доказательство состоит из длинной серии легких проверок. Например, следует доказать, что равенство обладает свойством подстановки относительно сложения и умножения (если отождествлять $R[x]$ с фактормножеством стандартных многочленов по отношению равенства). Это без труда следует из леммы 1.

Мы опустим детали проверки аксиом, ибо в гл. 13 будет установлен более общий результат (о кольцах формальных рядов). Утверждение о константах очевидно.

Таблица 11.1. Суммы и произведения в $\mathbf{Z}[x]$ (программа для вычислений с многочленами на АЛГОЛе).

```
begin integer array a[—1:100], b[—1:100], s[—1:100], p[1:200],
                                    f[—1:100], g[—1:100];
      integer i, k, initial, final;
 if a[—1] < b[—1]
   then for i:=1 step 1 until b[—1] do
          begin f[i]:=a[i]; g[i]:=b[i] end;
   else for i:=1 step 1 until a[—1] do
          begin f[i]:=b[i]; g[i]:=a[i] end;
 comment если deg a < deg b, то f присваивается значение a,
   а g—значение b, иначе f присваивается значение b, а g—зна-
   чение a, теперь вычисляется s=f+g=a+b;
 for k:=0 step 1 until f[—1] do s[k]:=a[k]+b[k];
 for k:=f[—1]+1 step 1 until g[—1] do s[k]:=g[k];
 s[—1]:=g[—1];
bk: if s[s[—1]]=0
     then begin s[—1]:=s[—1]—1; go to bk end;
 comment вычисляется p=fg=ab;
 if (f[—1]=0 ∧ f[0]=0) ∨ (g[—1]=0 ∧ g[0]=0)
   then begin p[—1]:=0; p[0]:=0; go to F end;
 p[—1]:=f[—1]+g[—1];
 for k:=0 step 1 until p[—1] do
 begin s[k]:=0;
 initial:=if k≤g[—1] then 0 else k—g[—1];
 final:=if k≤f[—1] then k else f[—1];
 for i:=initial step 1 until final do
   s[k]:=s[k]+f[i]×g[k—i]; end
end
```

В этой программе многочлен

$$a(x)=a_0+a_1x+\ldots+a_mx^m\in\mathbf{Z}[x]$$

представлен массивом целых чисел $(m=a_{-1},\ a_0,\ a_1,\ \ldots,\ a_m)$. Допускаются лишь массивы с $m\neq 0$ и $a_m\neq 0$; $\deg a,\ \deg b\leqslant 100$.

11.2. ПОЛИНОМИАЛЬНЫЕ КОЛЬЦА НАД ПОЛЯМИ

В § 10.9 мы определили понятие *нормы* и назвали *евклидовой областью* всякую область целостности, допускающую норму. Мы показали, что в евклидовых областях имеется алгоритм деления с остатком и верна теорема об однозначном разложении, как в кольце $\mathbf{Z}$ с нормой «модуль».

Теорема 2. *Над любым полем F кольцо многочленов $F[x]$ является евклидовой областью с нормой $v(a)=\deg a$.*

Доказательство. Для ненулевых многочленов $a,\ b\in F[x]$ имеем

$$\deg ab=\deg a+\deg b, \tag{6}$$

причем $\deg ab=\deg a$ тогда и только тогда, когда степень многочлена b равна нулю, т. е. b — константа. Поэтому deg удовлетворяет требованиям (27) гл. 10 (определение нормы).

Следствие. *В кольце $F[x]$ обратимы ненулевые константы, и только они.*

Для завершения доказательства теоремы 2 нужно еще проверить справедливость условия (28) гл. 10. Это утверждает следующая лемма.

Лемма 3 (алгоритм деления). *Пусть $a(x),\ b(x)$ — многочлены в $F[x]$, $b(x)\neq 0$. Тогда существуют неполное частное $q(x)\in F[x]$ и остаток $r(x)\in F[x]$ со свойствами: либо $r(x)=0$, либо $\deg r(x)<\deg b(x)$ и*

$$a(x)=b(x)\,q(x)+r(x),\qquad \deg r(x)<\deg b(x). \tag{7}$$

Доказательство. Частное $q(x)$ и остаток $r(x)$ можно вычислить с помощью следующего школьного алгоритма.

Положим $a(x)=\sum\limits_{k=0}^{m}a_kx^k$ и $b(x)=\sum\limits_{k=0}^{n}b_kx^k$, где $b_n\neq 0$. Рассмотрим два случая.

Случай 1. Если $m<n$, полагаем $q(x)=0$ и $r(x)=a(x)$.

Случай 2. Если $m\geqslant n$, положим

$$a_1(x)=a(x)-b_n^{-1}a_mx^{m-n}b(x)=a(x)-q_1(x)\,b(x).$$

Степень этого многочлена не больше $m-1$, ибо старшие коэффициенты у $a(x)$ и $q_1(x)\,b(x)$ совпадают и равны $b_n^{-1}a_mb_n=a_m$.

После не более чем $m-n+1$ таких шагов степень остатка окажется меньше m.

Любой многочлен степени n можно однозначно представить в виде $cm(x)$, где c — старший коэффициент, а $m(x)$ — многочлен со старшим коэффициентом единица:

$$m(x)=m_0+m_1x+\ldots+m_{n-1}x^{n-1}+x^n. \tag{8}$$

Неприводимые многочлены. Простые элементы в кольце $F[x]$ имеют специальное название.

Определение. Многочлен $p(x)$ в кольце $F[x]$ называется *приводимым* (над F), если $p(x)=a(x)b(x)$ для подходящих непостоянных многочленов $a(x)$, $b(x)\in F[x]$. В противном случае многочлен $p(x)$ называется *неприводимым* (также над F).

Обратим внимание читателя, что приводимость или неприводимость данного многочлена $p(x)$ существенно зависит от поля F. Например, над полем вещественных чисел $\mathbf{R}$ многочлен x^2+1 неприводим, но над $\mathbf{C}$ он становится приводимым: $x^2+1=(x+i)(x-i)$. Многочлен x^2-2 неприводим над $\mathbf{Q}$ и над $\mathbf{Z}_3$, тогда как $x^2-2=(x+3)(x+8)$ над $\mathbf{Z}_{11}$ и $x^2-2=(x+\sqrt{2})(x-\sqrt{2})$ над $\mathbf{R}$. Аналогично x^2+x+1 неприводим над $\mathbf{Z}_2$, тогда как $x^2+x+1=(x+2)^2$ — над $\mathbf{Z}_3$.

Из равенства $\deg ab=\deg a+\deg b$ ясно, что линейные многочлены неприводимы над любым полем. Согласно «основной теореме алгебры», над полем $\mathbf{C}$ неприводимы только линейные многочлены. Над $\mathbf{R}$ неприводимы только линейные многочлены и квадратные многочлены x^2+px+q с отрицательным дискриминантом p^2-4q. Над полем рациональных чисел $\mathbf{Q}$ есть бесконечно много неприводимых многочленов любой степени.

В этой книге нас будут больше всего интересовать многочлены над полем $\mathbf{Z}_2$.

УПРАЖНЕНИЯ А

1. Доказать неравенства (5) и (5′).

2. Доказать, что факторкольцо $R[x]/(x^4+x^3+x+1)$ не может быть полем ни для какого коммутативного кольца R.

3. Вычислить образ $(2x+1)^{-1}$ в факторкольце $F[x]/(x^3-2)$, где
а) $F=\mathbf{Q}$, б) $F=\mathbf{Z}_5$, в) $F=\mathbf{Z}_7$.

4. Показать, что для любого поля F поле частных кольца $F[x]$ есть поле рациональных функций от x (см. теорему 3 гл. 10).

5. Пусть $F^*=F-\{0\}$, $(F[x])^*=F[x]-\{0\}$, и пусть M — моноид многочленов со старшим коэффициентом единица по умножению. Установить изоморфизм моноидов $(F[x])^*=F^*M$.

6. а) Доказать, что $(x^m-1)\mid(x^n-1)$ тогда и только тогда, когда $m\mid n$ (над любым полем коэффициентов).

б) Вывести отсюда, что $(k^m-1) \mid (k^n-1)$ для любого целого числа $k>1$ тогда и только тогда, когда $m \mid n$.

7. Обозначим через $C[0, 1]$ кольцо всех непрерывных вещественнозначных функций на замкнутом отрезке $0 \leqslant x \leqslant 1$ с законами композиции

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

а) Является ли $C[0, 1]$ областью целостности? Обоснуйте ответ.

б) Покажите, что отображение $f \mapsto f(a)$: $C[0, 1] \longrightarrow\!\!\!\!\rightarrow \mathbf{R}$ является эпиморфизмом, ядро которого является максимальным идеалом.

11.3. ПОЛИНОМИАЛЬНЫЕ КОДЫ

В этом параграфе опишем специальный класс групповых кодов — так называемые *полиномиальные коды*. При кодировании сообщения отождествляются с многочленами, а само кодирование состоит в умножении на фиксированный многочлен. Мы опишем метод обнаружения ошибок ниже. Как обычно, для блочных (m, n)-кодов в каждом блоке имеется $k=n-m$ контрольных символов.

Как в гл. 8, пусть $a_0, \ldots, a_{m-1}$ — символы сообщения $\mathbf{a}$. Пусть сначала алфавит двоичный, отождествим его с $\mathbf{Z}_2$. Тогда слова сообщения можно отождествить с многочленами степени $\leqslant m-1$ над $\mathbf{Z}_2$:

$$\mathbf{a} \leftrightarrow a_0+a_1x+\ldots+a_{m-1}x^{m-1} \tag{9}$$

(нижние индексы $0, 1, \ldots, m-1$ здесь удобнее, в отличие от индексов $1, \ldots, n$ гл. 8).

Так, последовательность 01101 при $m=5$ отвечает многочлену $x+x^2+x^4$.

В общем случае алфавит полиномиального кода отождествляется с некоторым конечным полем (скажем, $\mathbf{Z}_3$ или $\mathbf{Z}_5$).

Определение. Пусть F — некоторое конечное поле. Фиксируем многочлен степени k:

$$g(x)=g_0+g_1x+\ldots+g_kx^k \in F[x], \quad g_0 \neq 0, \quad g_k \neq 0. \tag{10}$$

Полиномиальный код с *кодирующим многочленом* $g(x)$ кодирует слово сообщения $\mathbf{a}$ вида (9) многочленом

$$b(x)=b_0+b_1x+\ldots+b_{n-1}x^{n-1}=a(x)\,g(x) \tag{11}$$

или словом $\mathbf{b}$, составленным из коэффициентов этого многочлена [1]).

Подчеркнем, что в (10) требуется выполнение неравенств $g_0 \neq 0$ и $g_k \neq 0$. Если $g_0=0$, то все кодовые слова будут начинаться с нуля, и этот первый символ не будет нести никакой информации.

---

[1]) См. также конец § 11.5, где дано другое, более близкое к общепринятому описание схемы кодирования полиномиальных кодов.— *Прим. перев.*

Аналогично при $g_k = 0$ последний символ не будет нести информации.

Пример 2. Пусть $\mathbf{Z}_2$ — двоичный алфавит $[0, 1]$. Рассмотрим кодирующий многочлен $g(x) = 1 + x^2 + x^3$. Сообщение $a_0a_1 \ldots a_4 = 01011$, отвечающее многочлену $a(x) = x + x^3 + x^4$, будет закодировано коэффициентами многочлена $b(x) = a(x)\,g(x) = x + x^5 + x^7$, т. е. словом $b_0b_1 \ldots b_7 = 01000101$.

Теорема 3. *Полиномиальный код с кодирующим многочленом*

$$g(x) = g_0 + g_1x + \ldots + g_kx^k$$

является матричным кодом с кодирующей матрицей размера $m \times (m + k)$:

$$G = \begin{bmatrix} g_0 & g_1 & g_2 \cdots & g_k & 0 & 0 & 0 \\ 0 & g_0 & g_1 \cdots & g_{k-1} & g_k & 0 & 0 \\ \vdots & \vdots & & & & & \vdots \\ 0 & 0 & 0\ g_0 & g_1 \cdots & g_{k-1} & & g_k \end{bmatrix}$$

Ненулевые элементы в j*-й строке составляют блок* $g_0g_1 \ldots g_k$*, расположенный от* j*-го до* $(j + k)$*-го места.*

Действительно, в j-й строке стоят коэффициенты произведения x^j на $g(x)$ в соответствии с описанием матриц кодирования, которое было дано в гл. 8.

Заметим, что ненулевые элементы очередной строки получаются *сдвигом* вправо на единицу из ненулевых элементов предыдущей строки. Как мы покажем в § 11.5, это облегчает физическую реализацию полиномиальных кодов электронными схемами, позволяя пользоваться регистрами сдвига.

Пример 3. (3,6)-код с кодирующим многочленом $1 + x + x^3$ отвечает матрице

$$G = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \end{bmatrix}$$

Этот код задается таблицей:

$a \mapsto b = aG$	$a \mapsto b = aG$
$000 \mapsto 000000$	$100 \mapsto 110100$
$001 \mapsto 001101$	$101 \mapsto 111001$
$010 \mapsto 011010$	$110 \mapsto 101110$
$011 \mapsto 010111$	$111 \mapsto 100011$

Рассмотрим некоторые свойства полиномиального кодирования. Прежде всего из определения ясно, что кодовые слова образуют аддитивную подгруппу элементарной абелевой группы $[\mathbf{Z}_2^n, +]$.

Теорема 4. *Минимальное расстояние между двумя кодовыми словами полиномиального кода с кодирующим многочленом* $g(x)$ *совпадает с минимумом весов многочленов* $a(x)g(x)$.

Отметим следующий более специальный результат.

Теорема 5. *Если многочлен с коэффициентами в* $\mathbf{Z}_2$ *делится на* $1+x$, *то он имеет четное число ненулевых коэффициентов.*

Доказательство. Пусть $f(x)=f_0+f_1x+\ldots+f_nx^n$. Предположим, что $f(x)=(x+1)h(x)$ для некоторого $h(x)$. Придадим x значение 1. Тогда, с одной стороны,

$$f(1)=(1+1)h(1)=0,$$

а с другой стороны, $f(1)=f_0+f_1+\ldots+f_n$. Это доказывает требуемое.

Если каждое кодовое слово имеет четное число единиц, то это есть код с проверкой на четность, обнаруживающий любое нечетное число ошибок передачи.

Пример 4. (3,5)-код с кодирующим многочленом $g(x)=1+x$ для сообщений длины 3 задается таблицей

$a\mapsto b=aG'$	$a\mapsto b=aG'$
$000\mapsto 0000$	$100\mapsto 1100$
$001\mapsto 0011$	$101\mapsto 1111$
$010\mapsto 0110$	$110\mapsto 1010$
$011\mapsto 0101$	$111\mapsto 1001$

$$G=\begin{bmatrix}1100\\0110\\0011\end{bmatrix}$$

11.4. ПРЕИМУЩЕСТВА ПОЛИНОМИАЛЬНЫХ КОДОВ

Этот параграф посвящен описанию некоторых технических преимуществ полиномиальных кодов.

Теорема 6. *Если* $g(x)$ *не является делителем ни одного многочлена вида* x^k-1 *при* $k<n$, *то для* (m, n)*-кода, порожденного* $g(x)$, *минимальное расстояние между словами не меньше трех.*

Доказательство. Множество кодовых слов имеет вид $a(x)g(x)$, $\deg a(x)<m$. Поскольку код является групповым, минимальное расстояние между кодовыми словами совпадает с минимальным весом кодового слова.

Если бы существовало кодовое слово веса 2, то $g(x)$ делил бы многочлен вида $e(x)=x^i+x^j=x^i(1+x^{j-i})$. Так как $g_0=1$, $g(x)$ должен делить $1+x^{j-i}$ в противоречие с предположением.

Кодовое слово веса 1 также не может существовать, ибо иначе $g(x)$ должен был бы делить x^i.

Примитивные многочлены. Многочлен $g(x)$ степени k над $\mathbf{Z}_2$ называется *примитивным*, если $g(x)\,|\,(x^m-1)$ для $m=2^k-1$, но ни для какого меньшего значения m.

Пример 5. Многочлен $1+x^2+x^3\in\mathbf{Z}_2[x]$ примитивен: он делит x^7-1, но не делит x^j-1 для $j<7$.

Мы вернемся к примитивным многочленам в § 11.8 и затем в гл. 12, где доказывается существование примитивных многочленов любой степени и описываются их замечательные свойства. Пока мы покажем только, что примитивный многочлен $1+x^2+x^3$ порождает (4,7)-код Хэмминга, описанный в § 8.7.

Этот код задается следующей таблицей:

$0000\mapsto 0000000$	$1000\mapsto 1011000$
$0001\mapsto 0001011$	$1001\mapsto 1010011$
$0010\mapsto 0010110$	$1010\mapsto 1001110$
$0011\mapsto 0011101$	$1011\mapsto 1000011$
$0100\mapsto 0101100$	$1100\mapsto 1110100$
$0101\mapsto 0100111$	$1101\mapsto 1111111$
$0110\mapsto 0111010$	$1110\mapsto 1100010$
$0111\mapsto 0110001$	$1111\mapsto 1101001$

Непосредственно видно, что наименьшее расстояние между кодовыми словами равно 3.

Матрица этого кода

$$G=\begin{bmatrix}1&0&1&1&0&0&0\\0&1&0&1&1&0&0\\0&0&1&0&1&1&0\\0&0&0&1&0&1&1\end{bmatrix}$$

после подходящей перестановки строк и столбцов переходит в матрицу (4,7)-кода Хэмминга из § 8.6.

Теорема 7. *Рассмотрим полиномиальный* (m, n)*-код, порожденный многочленом* $g(x)$. «*Строка ошибок*» $\mathbf{e}=e_0e_1\ldots e_{n-1}$ *останется необнаруженной в том и только том случае, если соответствующий многочлен ошибок делится на* $g(x)$.

Доказательство. Строка ошибок остается необнаруженной (в групповом коде) только в том случае, если она является кодовым словом, что доказывает требуемое.

Обнаружение ошибок производится посредством алгоритма деления с остатком: мы делим многочлен, отвечающий принятому слову, на $g(x)$; если остаток степени $< \deg g$ оказывается ненулевым, то при передаче произошло искажение.

Существуют очень эффективные полиномиальные коды с *обнаружением* ошибок. Назовем *экспонентой* многочлена g наименьшее положительное целое число e, для которого $g(x) \mid (x^e - 1)$. В частности, $g \in \mathbf{Z}_2[x]$ примитивен, если его степень равна k, а экспонента $e = 2^k - 1$.

Назовем две смежных ошибки *двойной* ошибкой. (В реальных системах ошибки имеют тенденцию группироваться; идеализация двоичного симметричного канала зачастую бывает далека от действительности.)

Теорема 8. *Пусть кодирующий многочлен полиномиального (m, n)-кода имеет вид $g(x) = (1+x)h(x)$, где экспонента e многочлена $h(x)$ больше n. Тогда можно обнаружить любую комбинацию из двух простых или двойных ошибок.*

Доказательство. Многочлен ошибок $e(x) = x^i + x^j$ обнаруживается, ибо $h(x)$ не может делить $x^i(1 + x^{j-i})$. Многочлены ошибок $x^i + x^{i+1} + x^j$, $x^i + x^j + x^{j+1}$ или x^i обнаруживаются, потому что у них нечетное число ненулевых членов, так что они не могут делиться на $1 + x$. Многочлен ошибок

$$e(x)\,x^i + x^{i+1} + x^j + x^{j+1}$$

представляется в виде $(1+x)(x^i + x^j)$. Он не может делиться на $g(x)$, ибо $x^i + x^j$ не может делиться на $h(x)$, поскольку $j < n - 1$.

11.5. РЕГИСТРЫ СДВИГА

Полиномиальное кодирование состоит в умножении переданного сообщения на фиксированный многочлен. В этом параграфе описаны электронные схемы — *регистры сдвига*, позволяющие реализовать это умножение физически.

На рис. 11.1 показана последовательностная схема, умножающая $g(x) = 1 + x^2 + x^3$ на входной многочлен $a(x) = a_0 + a_1x + \ldots + a_mx^m$.

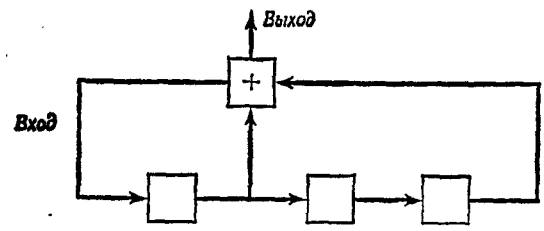


Рис. 11.1. Логическая схема, реализующая умножение на $1 + x^2 + x^3$.

На вход по очереди подаются коэффициенты многочлена $a(x)$ в порядке убывания степени; когда коэффициенты исчерпаются, подаются еще четыре нуля. Элементы памяти суть триггеры, задерживающие вход на один такт. На выходе появляются коэффициенты произведения $b(x)=a(x)g(x)$ в порядке убывания степени. Коэффициент b_{m+3} выдается в тот момент, когда на вход подается a_m. Коэффициент b_0 появляется последним, после $m+4$ тактов.

На рис. 11.2 показан фрагмент более общей схемы для умножения переменного входного многочлена $a(x)$ на фиксированный

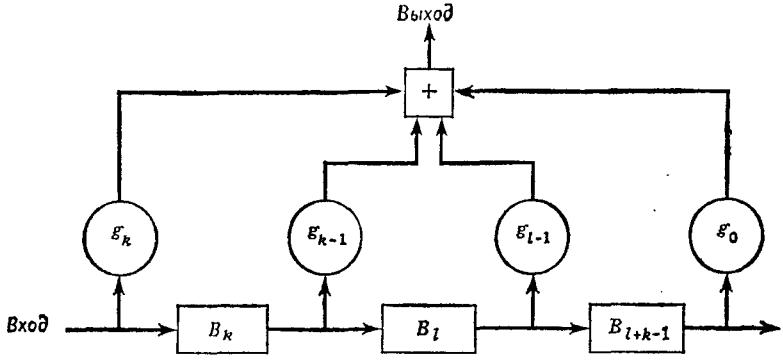


Рис. 11.2. Логическая схема, реализующая умножение на фиксированный многочлен $g(x)$.

многочлен $g(x)$. Поле коэффициентов может быть любым. Элемент, помеченный g_i (g_i — коэффициент многочлена $g(x)$), умножает на g_i выход элемента памяти B_{i+1}. Элементы памяти уже не обязаны быть триггерами: это электронные схемы, способные запоминать символы алфавита на один такт. Для поля $\mathbf{Z}_2$ схема значительно упрощается: если $g_i=1$, то элемент, умножающий на g_i, есть просто проводник, а если $g_i=0$, то соответствующий элемент отсутствует.

Отметим одну важную тонкость. При простом умножении на $g(x)$ входные символы перемешиваются и заменяются их линейными комбинациями. Поэтому на самом деле используется другая процедура кодирования. Именно, многочлен $a(x)$ кодируется многочленом $b(x)=x^{n-m}a(x)-r(x)=q(x)g(x)$, где $r(x)$ — остаток от деления $x^{n-m}a(x)$ на $g(x)$, а $q(x)$ — неполное частное.

При таком методе кодирования кодовые слова состоят из всех многочленов, делящихся на $g(x)$, так что к ним применимы установленные ранее результаты. В то же время вектор коэффициентов $b_{n-k}, \ldots, b_n$ совпадает с входным вектором $a_0, \ldots, a_{m-1}$, так что ими можно пользоваться немедленно: процедура декоди-

рования очень проста. На рис. 11.3 показан фрагмент схемы для деления входного многочлена $a(x)\,x^{n-m}$ на $g(x)$. Коэффициенты делимого подаются по очереди в порядке убывания степени, затем подается $k=n-m$ нулей. Коэффициенты остатка запоминаются в элементах памяти. Их можно считывать параллельно или использовать позже.

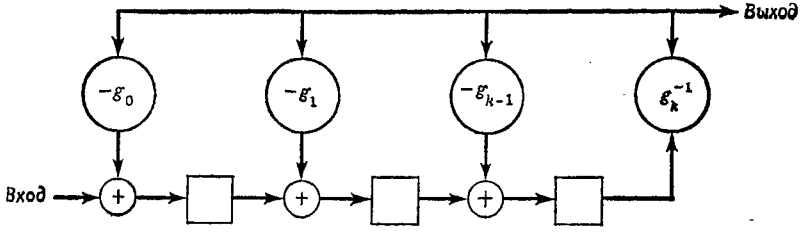


Рис. 11.3. Логическая схема, реализующая деление с остатком на $g(x)$.

УПРАЖНЕНИЯ Б

В упражнениях рассматриваются многочлены над $\mathbf{Z}_2$.

1. Показать, что неприводимый многочлен $1+x^3+x^4+x^5$ делит $x^{13}-1$, но не делит x^r-1 для $r<13$.

2. Пусть $g(x)=1+x+x^2+x^4=(1+x)(1+x^2+x^3)$ определяет (3,7)-код.
   а) Показать, что наименьший вес ненулевого кодового слова равен 4.
   б) Построить 3×7-матрицу G этого кода.

3. а) Представить матрицу G из упр. 2 в эквивалентной приведенной форме $[I_3P]$, где I_3 единичная 3×3-матрица, а P — контрольная 3×4-матрица.
   б) Построить 7×3-матрицу H проверки на четность с линейно независимыми столбцами, для которой $GH=0$.

4. Многочлен $1+x+x^4$ примитивен над $\mathbf{Z}_2$.
   а) Построить регистр сдвига и последовательную схему для соответствующего (11,15)-кода.
   б) Выписать матрицу проверки четности для этого кода.
   в) Привести кодирующую матрицу к стандартному виду.

5. Показать, что код, отвечающий многочлеиу $p(x)=x^i+x^j$, $i>j>0$, позволяет *обнаружить нечетное* число ошибок в любых позициях.
   б) Проверить детально, что множество кодовых слов полиномиального кода образует группу.
   в) Сколько кодовых слов имеется у полиномиальиого (k, n)-кода над полем $\mathbf{Z}_p$?

11.6. ТЕОРЕМА ОБ ОДНОЗНАЧНОМ РАЗЛОЖЕНИИ ДЛЯ МНОГОЧЛЕНОВ

Согласно теореме 2 §11.2, кольцо многочленов над любым полем F является *евклидовой областью*. Отсюда вытекает ряд следствий.

Следствие 1. *В кольце* $F[x]$ *все идеалы главные.*
Это вытекает из теоремы 10 гл. 10.

Следствие 2. *Наибольший общий делитель любых двух многочленов* $a(x)$, $b(x) \in F[x]$ *существует и представляется в виде* $r(x)a(x)+s(x)b(x)$, *где* $r(x)$, $s(x) \in F[x]$.

Это представление можно найти посредством алгоритма Евклида, если имеется алгоритм для вычислений в поле коэффициентов.

Следствие 3. *Если* $p(x)$ *неприводим в* $F[x]$, *то*

$$p(x) \mid a(x)b(x) \text{ влечет либо } p(x) \mid a(x), \text{ либо } p(x) \mid b(x). \quad (12)$$

Доказательство. Это частный случай леммы 2 §10.6.

Следствие 4. *Пусть* $p_k(x) \mid f(x)$ *для* $k=1, \ldots, r$, *где* $p_k(x)$ *попарно не ассоциированные неприводимые многочлены. Тогда* $\left[\prod_{k=1}^{r} p_k(x)\right] \mid f(x)$.

Теорема 9 (теорема об однозначном разложении). *Любой непостоянный многочлен* $p(x)$ *в* $F[x]$ *можно представить в виде произведения константы и неприводимых многочленов со старшими коэффициентами единица. Это разложение единственно с точностью до порядка множителей.*

Это следует из теоремы 19 гл. 10.

Теорема об однозначном разложении справедлива для многочленов от любого числа переменных над любым полем. Более того, если эта теорема верна в области D, то она верна и в $D[x]$.

11.7. КОМПЛЕКСНЫЕ КОРНИ ИЗ ЕДИНИЦЫ

Корнем из единицы степени m в поле F называется такой элемент $x \in F$, что $x^m = 1$. Известно, что в $\mathbf{C}$ такие корни имеют вид $z_k = e^{2\pi ik/m}$. Они представляются вершинами правильного m-угольника радиуса 1 в комплексной плоскости. С алгебраической точки зрения они образуют *циклическую* группу порядка m с образующей $z_1 = \zeta = e^{2\pi i/m}$.

Как было отмечено в гл. 7, циклическая группа $\mathbf{Z}_m$ порождается любым из своих $\varphi(m)$ образующих элементов, где φ — функция Эйлера. Именно если k взаимно прост с m, то z_k порождает $\mathbf{Z}_m$, т. е. всякий корень из единицы степени m является степенью z_k.

Эти образующие называются *примитивными* корнями из единицы степени m. Для $m=8$ — это ζ, ζ^3, ζ^5 и ζ^7. Для $m=15$ — это ζ^k с $k=1, 2, 4, 7, 8, 11, 13$ и 14. Тем самым, $\varphi(8)=4$, $\varphi(15)=8$.

Над любым полем F корни x_k многочлена $p(x)=0$ отвечают линейным делителям $x-x_k$ этого многочлена. В частности, для

z^m-1 над комплексными числами имеем

$$\prod_{k=0}^{m-1}(z-\zeta^k)=\prod_{k=0}^{m-1}(z-z_k)=z^m-1. \tag{13}$$

Это вытекает из следующей теоремы.

Теорема 10. *Пусть R—коммутативное кольцо, $a\in R$, $p(x)\in R[x]$. Тогда $(x-a)\mid p(x)$ в том и только том случае, когда $p(a)=0$ в R.*

Доказательство. Если $(x-a)\mid p(x)$ в $R[x]$, то $p(x)=(x-a)q(x)$ и $p(a)=(a-a)q(a)=0$.

Обратно, для любого $p(x)$ имеем

$$\begin{aligned}p(x)-p(a)&=\sum_{k=0}^{n}a_kx^k-\sum_{k=0}^{n}p_ka^k=\sum_{k=1}^{n}p_k(x^k-a^k)=\\&=\sum_{k=1}^{n}p_k[(x-a)(x^{k-1}+x^{k-2}a+\ldots+a^{k-1})]=\\&=(x-a)\sum_{k=1}^{n}p_k\left(\sum_{i=0}^{k-1}x^{k-i}a_i\right).\end{aligned}$$

Значит, $(x-a)\mid p(x)-p(a)$ в $R[x]$ для любого $a\in R$. Поэтому при $p(a)=0$ имеем $(x-a)\mid p(x)$.

В частности, над $\mathbf{C}$ для $z_k=e^{2\pi ik/n}$ имеем

$$(z-z_k)\mid(z^m-1),\ k=0,\ 1,\ \ldots,\ m-1. \tag{14}$$

Из следствия 4 §11.7 вытекает $\prod_{k=0}^{m-1}(z-z_k)\mid(z^m-1)$. Так как оба многочлена имеют степень m и старшие коэффициенты, равные единице, то они должны совпадать. Это доказывает равенство (13).

Круговые многочлены. Неприводимые над $\mathbf{Q}$ делители z^m-1 называются круговыми многочленами. Разберем для примера случай $m=15$.

Пример 6. Над любым полем имеется разложение

$$x^{15}-1=(x-1)(x^2+x+1)(x^4+x^3+x^2+x+1)c_8(x), \tag{15}$$

где $c_8(x)$—*круговой многочлен* степени 8:

$$c_8(x)=x^8-x^7+x^5-2x^4+x^3-x+1. \tag{15'}$$

Над $\mathbf{C}$ имеем $c_8(x)=\prod_P(z-z_k)$, где P—множество примитивных корней из единицы степени 15. Можно доказать, что корнями любого кругового многочлена являются в точности все примитивные корни из единицы подходящей степени.

Над $\mathbf{Z}_2$, однако, многочлен $c_8(x)$ *приводим:*

$$c_8(x) = (x^4 + x^3 + 1)(x^4 + x + 1) = x^8 + x^7 + x^5 + x^3 + x + 1. \quad (16)$$

Корни сомножителей (16) являются примитивными корнями степени 15 из единицы в конечном поле из 16 элементов, которое будет описано в гл. 12.

Вот список примитивных многочленов степеней 5, 6, 7 над $\mathbf{Z}_2$, делящих $z^{31}-1$, $z^{63}-1$, $z^{127}-1$ соответственно:

$$x^5 + x^4 + 1,\ x^6 + x + 1,\ x^7 + x^6 + 1.$$

УПРАЖНЕНИЯ В

1. Найти в $\mathbf{Q}[x]$ с помощью алгоритма Евклида наибольший общий делитель многочленов

$$x^5 + 3x^4 + 2x^3 - 2x^2 - 3x - 1,\ 3x^7 + 6x^6 + 5x^5 + 7x^4 + 8x^3 + 5x^2 + 4x + 2.$$

2. а) Показать, что в кольце $F[x]$ включение $(p(x)) \subset (q(x))$ равносильно делимости $q(x) \mid p(x)$.

б) Показать, что если $p(x)$ приводим, то идеал $(p(x))$ не максимален.

3. Показать, что квадратный или кубический многочлен $p(x)$ в $F[x]$ неприводим в том и только том случае, если $p(c) \neq 0$ для всех $c \in F$.

4. Вывести отсюда, что $x^3 + x + 1$ неприводим над $\mathbf{Z}_2$.

5. Написать таблицы сложения и умножения для кольца $\mathbf{Z}_2[x]/(x^3 + x + 1)$.

6. а) Показать, что $x^2 + x + 1$ — единственный неприводимый квадратичный многочлен над $\mathbf{Z}_2$.

б) Вычислить классы x^{-1} и $(1 + x)^{-1}$ в $GF(4) = \mathbf{Z}_2[x]/(x^2 + x + 1)$.

7. а) Показать, что $x^3 + x + 1$ и $x^3 + x^2 + 1$ неприводимы над $\mathbf{Z}_2$.

б) Показать, что других непрнводимых кубических многочленов над $\mathbf{Z}_2$ нет.

8. Показать, что $x^4 + x + 1$ неприводим над $\mathbf{Z}_2$. (*Указание:* рассмотреть возможные линейные и неприводимые квадратные делители.)

*11.8. ПОЛИНОМИАЛЬНЫЕ ФУНКЦИИ

В этом параграфе мы рассмотрим многочлены над кольцом R как *функции*. Множество R^S *всех* функций на произвольном множестве S со значениями в коммутативном кольце R само является коммутативным кольцом с операциями

$$(f + g)(s) = f(s) + g(s), \quad (17)$$

$$(fg)(s) = f(s)\, g(s). \quad (18)$$

В частности, это так при $S = R$. В кольце R^R тождественная функция 1_R: $x \mapsto x$ для всех $x \in R$ играет особую роль. *Полиномиальной функцией* называется любой элемент f: $R \to R$ из *подкольца* $R\langle x\rangle$, порожденного константами (функциями вида $\tilde{a}(x) = a$

для всех $x \in R$) и x (или 1_R). Многочлен

$$p(x) = p_0 + p_1 x + \ldots + p_n x^n$$

определяет полиномиальную функцию $\tilde{p}$: $a \mapsto \tilde{p}(a)$, $a \in R$, $\tilde{p} \in R^R$.

Пример 7. Пусть $R = \mathbf{Z}_3$. Многочлен $1 + x$ определяет функцию $\tilde{f}$: $R \to R$:

$$\tilde{f}(0) = 1, \quad \tilde{f}(1) = 2, \quad \tilde{f}(2) = 0.$$

Другая точка зрения на полиномиальные функции такова. Любой элемент a коммутативного кольца R определяет отображение «значение в точке a» f_a: $R[x] \to R$, что записывается как $f_a(p(x)) = \tilde{p}(a)$. Следующий результат очевиден.

Лемма. *Для всякого $a \in R$ отображение f_a: $p \mapsto \tilde{p}(a)$ является морфизмом колец. Отображение* μ: $p \mapsto \tilde{p}$, $R[x] \to R^R$, *также является морфизмом колец, образ которого совпадает с кольцом полиномиальных функций.*

Иными словами, операции над многочленами согласованы с операциями над представленными ими функциями.

В школьной алгебре многочлены обычно не отличают от функций. Это оправдано над $\mathbf{Q}$ и $\mathbf{R}$, где отображение μ является мономорфизмом.

Сейчас мы выведем простое достаточное условие на кольцо R для того, чтобы μ было мономорфизмом. Очевидно, необходимо и достаточно, чтобы

$$\text{из } p(a) = q(a) \text{ для всех } a \in R \text{ следовало } p(x) = q(x) \text{ в } R[x]. \quad (19)$$

Эквивалентное условие:

$$\text{из } p(a) = 0 \text{ для всех } a \in R \text{ следует } p(x) = 0 \text{ в } R[x]. \quad (20)$$

Это условие, очевидно, выполнено даже не для всех полей. Например, многочлен $x^3 - x$ представляет нулевую функцию на $\mathbf{Z}_3$. Вообще если R — любое *конечное* коммутативное кольцо из n элементов $a_1, \ldots, a_n$, то многочлен $(x - a_1) \ldots (x - a_n)$ представляет нулевую функцию на R. Покажем на основании теоремы 10, что для бесконечных областей целостности это уже не может произойти.

Будем говорить, что элемент $a \in R$ является *нулем* многочлена $p(x)$, если $\tilde{p}(a) = 0$.

Теорема 11. *Многочлен $p(x)$ степени n над областью целостности D имеет не более n нулей.*

Доказательство. Пусть a — нуль многочлена $p(x)$. Тогда по теореме 10 имеем $p(x) = (x - a)\, q(x)$, где степень q равна $n - 1$. Индукция по степени дает требуемое.

Следствие. *Если два многочлена над бесконечной областью целостности D представляют одну и ту же функцию, то они совпадают.*

Доказательство. Если их разность ненулевая и имеет степень n, то она имеет не более n нулей и потому не может представлять нулевую функцию.

Интерполяционная формула Лагранжа. Из доказанного следует, что существует не более одной полиномиальной функции на области целостности D, которая принимает в точках $x_0, x_1, \ldots, x_n$ значения $y_0, \ldots, y_n$ и имеет степень $\leqslant n$. Если D — *поле*, то такую функцию можно указать явно методом Лагранжа.

Теорема 12. *Пусть F — поле, $x_0, \ldots, x_n \in F$ и $y_0, \ldots, y_n \in F$. Тогда существует многочлен $p(x)$ степени $\leqslant n$ с условиями $p(x_0) = y_0, \ldots, p(x_n) = y_n$.*

Доказательство. Рассмотрим сначала многочлены

$$q_i(x) = (x-a_0) \ldots (x-a_{i-1})(x-a_{i+1}) \ldots (x-a_n). \qquad (21)$$

Если $i \neq j$, то $q_i(a_j) = 0$; если $i = j$, то $q_i(a_j) \neq 0$. Пусть $L_i = q_i(a_i)$. Ясно, что многочлен

$$p(x) = \sum_{i=0}^{n} \frac{y_i}{L_i} q_i(x) = \sum_{i=0}^{n} y_i \prod_{j \neq i} \frac{x-a_j}{a_i-a_j} \qquad (22)$$

принимает требуемые значения. Это — *интерполяционная формула Лагранжа.*

*11.9. ФОРМАЛЬНЫЕ ПРОИЗВОДНЫЕ

Как в анализе, мы можем определить *формальную производную* многочлена $f(x) = \sum_k f_k x^k$ с коэффициентами в поле F формулой

$$f'(x) = \sum_k k f_k x^{k-1} = \sum_l (l+1) f_{l+1} x^l. \qquad (23)$$

Здесь kf_k означает сумму k экземпляров f_k.

Легко проверить, что формальные производные удовлетворяют обычным тождествам:

$$(f+g)' = f' + g', \quad (cf)' = cf' \text{ (линейность)} \qquad (24)$$

и

$$(fg)' = fg' + gf'. \qquad (25)$$

Теорема 13. *Над полем F характеристики ∞ непостоянный многочлен $f(x)$ не имеет кратных множителей в том и только том случае, если н.о.д. $(f, f^1) = 1$.*

Доказательство. Разложим f на неприводимые множители:

$$f = cp_1(x)^{e_1} \ldots p_r(x)^{e_r}, \quad c \neq 0.$$

Случай 1. Если $e_i > 0$ для некоторого i, то $p_i(x)$ делит не только f, но и f', ибо

$$f' = c \sum_{k=1}^{r} \left[e_k p_k^{e_k - 1} p_k' \left(\prod_{j \neq k} p_j^{e_j} \right) \right].$$

Значит, $p_i(x) \mid (f, f')$.

Случай 2. Если все $e_i = 1$, то из этого же равенства находим

$$f'(x) \equiv cp_i'(x) \prod_{j \neq i} p_j(x) \not\equiv 0 \mod p_i(x).$$

Мы пользуемся тем, что $p_i(x)$ неприводимы и попарно не ассоциированы. Кроме того, f' не может быть тождественным нулем, если характеристика F не конечна. Таким образом, неприводимые делители многочлена f не могут делить f', так что н.о.д. $(f, f') = 1$.

УПРАЖНЕНИЯ Г

1. Показать, что для вычисления значений многочлена четвертой степени достаточно трех умножений и пяти сложений.
(*Указание:* $q(x) = a_0\{[g(x) + \lambda_2][g(x) + x + \lambda_3] + \lambda_4\}$, где $g(x) = x(x + \lambda_1)$ с подходящим λ_1.)

2. Разложить $x^6 - 1$ на неприводимые множители:
а) над $\mathbf{R}$; б) над $\mathbf{Q}$; в) над $\mathbf{C}$.

3. Показать, что в любом поле корни уравнения $x^n = 1$ образуют циклическую группу порядка, делящего n.

4. Пусть D — область целостности, $R \subset D$ — подкольцо, $c \in D - R$. Продолжить отображения 1_R: $R \to R$ и $x \mapsto c$ до эпиморфизма колец f: $R[x] \twoheadrightarrow R[c]$.

12

КОНЕЧНЫЕ ПОЛЯ

12.1. РАСШИРЕНИЯ ПОЛЕЙ

Пусть F, G — два поля и $F \subset G$. Тогда G называется *расширением* поля F. Так, $\mathbf{C}$ является расширением поля $\mathbf{R}$, а $\mathbf{R}$ — расширением поля $\mathbf{Q}$. Всякое поле F является расширением своего простого подполя, порожденного единицей. Простое подполе характеристики p канонически изоморфно $\mathbf{Z}_p$, а поле характеристики ∞ канонически изоморфно $\mathbf{Q}$.

Ряд аспектов теории расширений полей можно проследить на примере комплексных чисел.

Пример 1. Поле $\mathbf{C}$ состоит из выражений вида $z = x + iy$ $(i = \sqrt{-1};\ x, y \in \mathbf{R})$. Сложение и умножение в $\mathbf{C}$ определяются правилами:

$$(x_1 + iy_1) + (x_2 + iy_2) = (x_1 + x_2) + i(y_1 + y_2), \tag{1}$$

$$(x_1 + iy_1)(x_2 + iy_2) = (x_1x_2 - y_1y_2) + i(x_1y_2 + y_1x_2). \tag{2}$$

Отображение $x \mapsto x + i0$ является вложением $\mathbf{R}$ в $\mathbf{C}$ в качестве подполя. Поле $\mathbf{C}$ можно рассматривать как двумерное векторное пространство над $\mathbf{R}$ с базисом $1 = 1 + i0$ и $i = 0 + i1$. Поскольку $i^2 = -1 + i0$, мы можем писать $i = \sqrt{-1}$. Как поле (и даже как коммутативное кольцо) $\mathbf{C}$ порождается $\mathbf{R}$ и i. Мы записываем это так: $\mathbf{C} = \mathbf{R}[i]$.

На самом деле *любое* расширение G поля F можно рассматривать как *векторное пространство* над F. Это означает, что $[G, +]$ — абелева группа и что определено умножение ее элементов на элементы F (скаляры), подчиняющиеся тождествам

$$a(b\xi) = (ab)\xi, \qquad (a + b)\xi = a\xi + b\xi,$$

$$a(\xi + \eta) = a\xi + a\eta, \qquad 1\xi = \xi$$

для всех $a, b \in F$ и $\xi, \eta \in G$.

Пусть V — векторное пространство над полем F. Его (конечным) *базисом* называется такое множество n векторов $\beta_1, \ldots, \beta_n$, что любой $\xi \in V$ однозначно представим в виде линейной комбинации $\xi = x_1\beta_1 + \ldots x_n\beta_n$, где $x_i \in F$. Одна из первых теорем о ли-

нейных пространствах утверждает, что если существует *один* конечный базис V над F, то *любой другой* базис имеет то же число элементов. Это число называется тогда *размерностью* V над F. Показывается, что в n-мерном пространстве V любые $n+1$ векторов линейно зависимы, т. е. удовлетворяют соотношению вида

$$f_0\xi_0 + \ldots + f_n\xi_n = 0,$$

где не все f_i равны нулю.

Пространство V заведомо имеет *конечный* базис, если существует такое конечное множество векторов $\xi_1, \ldots, \xi_r$, что любой его элемент $\xi \in V$ представляется в виде

$$\xi = c_1\xi_1 + \ldots + c_r\xi_r \quad (c_i \in F).$$

Если пространство не имеет конечного базиса, оно называется *бесконечномерным*.

Пример 2. Многочлены степени $\leqslant n$ над любым полем F образуют $(n+1)$-мерное векторное пространство над F с базисом $1, x, x^2, \ldots, x^n$ (базис, конечно, не определяется однозначно: например, $1, 1+x, 1+x^2, \ldots, 1+x^n$—тоже базис). Коммутативное кольцо *всех* многочленов $F[x]$ является бесконечномерным векторным пространством над полем F.

В этой книге мы не будем заниматься бесконечномерными пространствами. Любое конечное поле F, очевидно, конечномерно над своим простым подполем $\mathbf{Z}_p$.

Если поле G является расширением поля F и имеет размерность n как линейное пространство над F, то n называется *степенью этого расширения*: $[G:F]=n$.

Так, комплексное поле $\mathbf{C}$ является расширением степени 2 (квадратичным расширением) поля $\mathbf{R}$. Оно порождено одним элементом, скажем $\sqrt{-1}$. Вообще, поле G называется *простым* расширением поля F, если существует такой элемент $c \in G$, что $G = F(c)$—наименьшее подполе G, содержащее F и c. Это определение равносильно возможности представить любой элемент $x \in G$ в виде частного:

$$x = \frac{p(c)}{q(c)} = \frac{\sum_{k=0}^{n} p_k c^k}{\sum_{k=0}^{n} q_k c^k}, \tag{3}$$

где p, q—многочлены от c с коэффициентами $p_k, q_k \in F$ и $q(c) \neq 0$.

В этой главе мы будем в основном заниматься *конечными полями*. Пусть G—конечное поле простой характеристики p. Пусть n—его степень над простым подполем $\mathbf{Z}_p$. Записывая элементы G в виде векторов $\mathbf{x} = (x_1, \ldots, x_n)$ из коэффициентов раз-

ложения по какому-нибудь базису над $\mathbf{Z}_p$, находим, что *любое конечное поле характеристики p имеет порядок p^n для некоторого n*. Опишем явно конечные поля $GF(4)$ и $GF(8)$, состоящие из четырех и восьми элементов соответственно.

Пример 3. В поле $\mathbf{Z}_2$ имеем $x(x+1)=0$ для $x=0$ и $x=1$. Поэтому уравнение $x(x+1)=1$ не имеет корней в $\mathbf{Z}_2$, а многочлен x^2+x+1 неприводим. Мы можем присоединить его корень j к полю $\mathbf{Z}_2$. Иными словами, построим факторкольцо $GF(4)=\mathbf{Z}_2[x]/(x^2+x+1)$ (см. ниже § 12.3). Таблицы сложения и умножения для $GF(4)$ нетрудно вычислить непосредственно; в ответе всюду j^2 заменяется на $j+1=-j-1$ в силу соотношения $j^2+j+1=0$ над $\mathbf{Z}_2$. Результат таков:

Таблица 12.1

Сложение и умножение в $GF(4)$

+	0	1	j	$j+1$
0	0	1	j	$j+1$
1	1	0	$j+1$	j
j	j	$j+1$	0	1
$j+1$	$j+1$	j	1	0

×	0	1	j	$j+1$
0	0	0	0	0
1	0	1	j	$j+1$
j	0	j	$j+1$	1
$j+1$	0	$j+1$	1	j

Аддитивная группа $GF(4)$ изоморфна $\mathbf{Z}_2\times\mathbf{Z}_2$. Ненулевые элементы образуют мультипликативную группу порядка 3 с образующей j (или $j+1=j^2$). Поэтому $GF(4)$ является *полем*.

Пример 4. Кубический многочлен x^3+x^2+1 неприводим над $\mathbf{Z}_2$, ибо не имеет корней в $\mathbf{Z}_2$. Определим $GF(8)$ («поле Галуа порядка 8») как факторкольцо:

$$GF(8)=\mathbf{Z}_2[x]/(x^3+x^2+1).$$

Общий элемент его можно записать в виде $a_2\bar{x}^2+a_1\bar{x}+a_0$, $\bar{x}=x \bmod (x^3+x^2+1)$ или просто в виде строки $a_2a_1a_0=a$, как и сделано в табл. 12.2. При умножении двух элементов они перемножаются как многочлены в $\mathbf{Z}_2[x]$, произведение делится с остатком на x^3+x^2+1, и остаток, в котором $\bar{x}$ подставлен вместо x, считается результатом умножения. В табл. 12.2 представлено умножение (таблица сложения тривиальна — как в $[\mathbf{Z}_2, +]^3$).

Ненулевые элементы $GF(8)$ образуют *циклическую группу* порядка 7 с единицей 001. (Эта единица имеется в каждом столбце и строке таблицы, кроме первых, так что в моноиде ненулевых элементов все элементы обратимы и он действительно является группой.) Цикличность следует из того, что 7 — простое число (гл. 7, теорема 15, следствие 2).

Таблица 12.2

Таблица умножения в GF (8)

×	000	001	010	011	100	101	110	111
000	000	000	000	000	000	000	000	000
001	000	001	010	011	100	101	110	111
010	000	010	100	110	101	111	010	011
011	000	011	110	101	001	010	111	100
100	000	100	101	001	111	011	010	110
101	000	101	111	010	011	110	100	001
110	000	110	001	111	010	100	011	101
111	000	111	011	100	110	001	101	010

12.2. ПРОСТЫЕ РАСШИРЕНИЯ

Пусть G — расширение степени n поля F, и пусть $c \in G$. Элементы $1, c, c^2, \ldots, c^n$ в G должны быть линейно зависимы над F, т. е. удовлетворять соотношению вида

$$f(c) = \sum_{k=0}^{n} f_k c^k = 0.$$

Таким образом, справедлива

Теорема 1. *Если поле G — конечное расширение F степени n, то любой элемент $c \in G$ является корнем многочлена над F степени, не большей n.*

Пусть теперь G — *простое* расширение F степени n, т. е. $G = F[c]$ для подходящего $c \in G$. Деля уравнение для c на его старший коэффициент, мы можем считать, что этот коэффициент равен единице:

$$m(c) = c^s + \alpha_1 c^{s-1} + \ldots + \alpha_s = 0, \qquad \alpha_i \in F. \tag{4}$$

Здесь через $s \leqslant n$ обозначена *наименьшая* возможная степень. Она, однако, должна совпадать с n: действительно, значение в c любого многочлена от x с коэффициентами из F равно в G значению его остатка от деления на $m(x)$:

$$p(c) = q(c)\, m(c) + r(c) = r(c), \text{ ибо } m(c) = 0. \tag{5}$$

Значит, G линейно порождается элементами $1, c, \ldots, c^{s-1}$, а потому $[G : F] \leqslant s$ и, окончательно, $n \leqslant s$. Отсюда мы получаем следующую теорему.

Теорема 2. *Пусть G — простое расширение поля F степени n, $G = F[c]$. Тогда c является корнем подходящего многочлена*

$m(x)$ степени n, старший коэффициент которого равен единице:

$$m(c) = c^n + \alpha_1 c^{n-1} + \ldots + \alpha_n = 0, \quad \alpha_i \in F. \tag{6}$$

Покажем теперь, как восстановить G по $m(x)$.

Теорема 3. *В предположениях теоремы 2 многочлен $m(x)$ неприводим и имеет место канонический изоморфизм $G \cong F[x]/(m(x))$ поля F с факторкольцом кольца $F[x]$ по главному идеалу, порожденному $m(x)$.*

Доказательство. Предположим, что $m(x)$ приводим. Тогда c будет корнем одного из собственных делителей $m(x)$. Так как степень его меньше n, то это противоречит тому, что $[G:F] = n$.

Далее, рассмотрим гомоморфизм μ: $F[x] \to G$, где $\mu(x) = c$. Он является *эпиморфизмом*, поскольку $G = F[c]$. Его ядро состоит из всех многочленов $p(x)$ с $p(c) = 0$. Этот идеал главный, поскольку все идеалы в $F[x]$ главные. Следовательно, он состоит из всех кратных многочлена наименьшей степени с корнем c, т.е. многочлена $m(x)$. Согласно теореме 14 гл. 10, получаем

$$G \cong F[x]/(m(x)),$$

что завершает доказательство.

Простые трансцендентные расширения. Пусть F — любое поле. Поле частных $Q(F[x])$ кольца многочленов над F называется *простым трансцендентным расширением* поля F посредством переменной x и обозначается через $F(x)$. Если F бесконечно, $F(x)$ можно отождествить с полем рациональных функций на F со значениями в F: функций, представленных отношениями многочленов (всюду, кроме тех точек, где знаменатель обращается в нуль).

*12.3. ВЫЧИСЛЕНИЯ В $R[x]/(m(x))$

Опишем подробнее, как представлять элементы простых расширений и как вычислять их суммы, разности, произведения и частные. Наши формулы применимы к любому факторкольцу $R[x]/(m(x))$ по главному идеалу, порожденному многочленом, старший коэффициент которого равен единице, над кольцом R:

$$m(x) = x^n + \sum_{k=1}^{n} c_k x^{n-k}, \quad c_k \in R. \tag{7}$$

Они основаны на следствии из формулы (7):

$$x^n \equiv \sum_{k=1}^{n} (-c_k) x^{n-k} \bmod (m(x)); \tag{8}$$

теперь можно заменить класс любой степени x^{n+h} $(h \geqslant 0)$ в $R[x]/(m(x))$ линейной комбинацией степеней x^{n-k}, вычитая подходящие кратные $m(x)$. Отсюда вытекает, что всякий аддитивный смежный класс по модулю $(m(x))$ в кольце $R[x]$ представлен ровно одним многочленом (лидером класса) степени $\leqslant n-1$:

$$a_0 + a_1 x + \ldots + a_{n-1} x^{n-1}, \quad a_i \in R, \quad i = 0, \ldots, n-1. \tag{9}$$

Относительно сложения смежные классы образуют n-мерный «свободный модуль» над R, на лидерах сложение осуществляется по простой формуле:

$$\sum_{k=0}^{n-1} a_k x^k + \sum_{k=0}^{n-1} b_k x^k = \sum_{k=0}^{n-1} (a_k + b_k) x^k, \tag{10}$$

а умножение на элементы $r \in R$ — по формуле

$$r \sum_{k=0}^{n-1} a_k x^k = \sum_{k=0}^{n-1} (r a_k) x^k. \tag{10'}$$

Базис этого модуля состоит из многочленов

$$1, \ x, \ x^2, \ \ldots, \ x^{n-1}.$$

Умножение в терминах лидеров немногим сложнее. Они перемножаются как многочлены, и результат заменяется остатком от деления на $m(x)$. Приведем, например, явные формулы для $m(x) = x^3 + x + 1$. Используя соотношение $x^3 = -x - 1$, находим последовательно:

$$\begin{aligned}
1 \cdot (b_0 + b_1 x + b_2 x^2) &= b_0 + b_1 x + b_2 x^2, \\
x \cdot (b_0 + b_1 x + b_2 x^2) &= b_0 x + b_1 x^2 + b_2 x^3 = \\
&= -b_2 + (b_0 - b_2) x + b_1 x^2, \\
x^2 \cdot (b_0 + b_1 x + b_2 x^2) &= x[-b_2 + (b_0 - b_2) x + b_1 x^2] = \\
&= -b_2 x + (b_0 - b_2) x^2 + b_1 (-x - 1).
\end{aligned}$$

Отсюда, в силу билинейности,

$$\begin{aligned}
\left(\sum_{k=0}^{2} a_k x^k\right)\left(\sum_{k=0}^{2} b_k x^k\right) &= a_0 [b_0 + b_1 x + b_2 x^2] + \\
&+ a_1 [(-b_2) + (b_0 - b_2) x + b_1 x^2] + \\
&+ a_2 [(-b_1) + (-b_2 - b_1) x + (b_0 - b_2) x^2],
\end{aligned} \tag{11}$$

или, объединяя подобные члены,

$$\begin{aligned}
&(a_0 b_0 - a_1 b_2) + (a_0 b_1 + a_1 b_0 - a_1 b_2 - a_2 b_1 - a_2 b_2) x + \\
&+ (a_0 b_2 + a_1 b_1 + a_2 b_0 - a_2 b_2) x^2.
\end{aligned} \tag{12}$$

Формулы (10) и (12) дают правила сложения и умножения в кольце $R[x]/(x^3+x+1)$.

В общем случае они выглядят так.

Теорема 4. *Пусть* $m(x)=x^n+\sum_{k=1}^{n} c_k x^{n-k}$ *— многочлен со старшим коэффициентом единица над коммутативным кольцом* R. *Тогда факторкольцо* $R[x]/(m(x))$ *канонически отождествляется с пространством многочленов* $\sum_{k=0}^{n-1} a_k x^k$ *степени* $\leqslant n-1$ *с коэффициентами* $a_k \in R$. *Сложение в этом пространстве определяется формулами* (10), *а умножение — формулами*

$$\left(\sum_{k=0}^{n-1} a_k x^k\right)\left(\sum_{k=0}^{n-1} b_k x^k\right)=\sum_{k=0}^{n-1}\left(\sum_{i=0}^{k} a_i b_{k-i}\right)x^k+\sum_{k=n}^{2n-2}\left(\sum_{i=k-n+1}^{n-1} a_i b_{k-i}\right)x^k, \quad (13)$$

в которых старшие степени x *последовательно заменяются по формулам*

$$x^n=-\sum_{k=1}^{n} c_k x^{n-k}=p_1(x),$$

$$x^{n+1}=\sum_{k=2}^{n}(-c_k)\,x^{n-k+1}+c_1\sum_{k=1}^{n}(c_k x^{n-k})=p_2(x), \quad (13')$$

а далее по рекурсии при $j=1, \ldots, n-3$:

$$p_{n+j+1}(x)=x p_{n+j}(x)$$

и

$$x(a_0+a_1x+\ldots+a_{n-1}x^{n-1})=\sum_{i=1}^{n-1} a_{i-1}x^i-a_{n-1}\left(\sum_{k=1}^{n} c_k x^{n-k}\right). \quad (14)$$

Доказательство мы предоставляем читателю; оно проводится индукцией по j при данном $m(x)$.

УПРАЖНЕНИЯ А

1. Вычислить $(1+\sqrt[3]{3})^{-1}$ в $\mathbf{Q}[\sqrt[3]{3}]$.

2. а) Пусть α — примитивный элемент в поле $GF(q)$ $(q=p^n)$ над простым подполем. Показать, что α^j примитивен тогда и только тогда, когда $(j, q-1)=1$.

б) Вывести отсюда, что в $GF(q)$ имеется ровно $\varphi(q-1)$ примитивных элементов, где φ — функция Эйлера.

3. а) Найти примитивный элемент α в $\mathbf{Z}_3[x]/(x^2-2)$.

б) Представить все степени $\alpha^2, \ldots, \alpha^8$ в виде $a+b\sqrt{2}$, где $a, b \in \mathbf{Z}_3$.

в) Однозначно ли такое представление?

4. а) Можно ли вложить $GF(4)$ в качестве подполя в $GF(8)$?

б) Для каких $m, n \in \mathbf{P}$ поле $GF(p^m)$ можно вложить в $GF(p^n)$? Обоснуйте ответ.

5. Показать, что отображение $x \mapsto x^3$ не является автоморфизмом аддитивной группы поля Галуа $GF(2^n)$ при $n > 1$.

6. Для каких r отображение $x \mapsto x^r$ является автоморфизмом мультипликативного моноида поля Галуа $GF(2^n)$?

12.4. ТЕОРЕМА СУЩЕСТВОВАНИЯ

В этом параграфе мы опишем в общем виде процедуру расширения полей с помощью неприводимых многочленов, частными случаями которой являются процедуры построения $GF(4)$ и $GF(8)$ из § 12.2. Напомним сначала один результат из § 10.1.

Лемма 1. *Любой максимальный идеал M коммутативного кольца R прост.*

Из нее сразу же следует

Теорема 5. *В области целостности $F[x]$, где F — любое поле, следующие условия на непостоянный многочлен $p(x)$ эквивалентны:*

(i) *$p(x)$ неприводим;*
(ii) *главный идеал $(p(x))$ максимален;*
(iii) *главный идеал $(p(x))$ прост.*

Доказательство. Пусть $p(x)$ неприводим, и пусть идеал $P = (p(x))$ строго содержит Q. Тогда существует элемент $q(x) \in Q$, не делящийся на $p(x)$. Значит, $p(x)$ и $q(x)$ взаимно просты. В силу результата из § 10.9 некоторая линейная комбинация $p(x)$ и $q(x)$ равна 1; но она принадлежит Q вместе с $p(x)$ и $q(x)$. Поэтому $Q = F[x]$. Значит, P максимален. Мы показали, что из (i) следует (ii).

Из (ii) следует (iii) согласно лемме 1.

Наконец, из (iii) следует (i). Действительно, предположим, что $p(x) = q(x)r(x)$. Если идеал P прост, то либо $q \in P$, либо r должен лежать в P, но это невозможно для собственных делителей p. Доказательство завершено.

Теперь мы можем доказать нашу основную теорему существования.

Теорема 6. *Пусть F — любое поле, $m(x)$ — многочлен, неприводимый над F, старший коэффициент которого равен единице. Тогда факторкольцо $F[x]/(m(x)) = G$ является простым расширением поля F, которое мономорфно погружено в G.*

Доказательство. Согласно теореме 1 гл. 11, $G = F[x]/(m(x))$ есть коммутативное кольцо, содержащее F. Поскольку $m(x)$ неприводим, из теоремы 5 следует, что идеал $(m(x))$ максимален. Теорема 19 гл. 10 показывает тогда, что $G = F[x]/(m(x))$ является полем. Простота расширения G следует

из того, что G порождено над F образом x. Доказательство завершено.

Вычисление обратных элементов. В § 12.3 мы объяснили, как вычислять суммы, разности и произведения в кольце $F[x]/(m(x))$. Однако вычисление обратных элементов менее тривиально. В примерах 1 и 2, где поля настолько малы, что их можно задать таблицами, обратные элементы можно находить из этих таблиц.

В общем случае если ненулевой элемент поля G представлен многочленом $a(x)$, то $a(x)$ должен быть взаимно прост с $m(x)$. С помощью алгоритма Евклида отыщем многочлены $s(x)$ и $t(x)$, такие, что $s(x)a(x)+t(x)m(x)=1$. Так как $m(x)=0$ в G, элемент, обратный к классу $a(x)$, представлен классом $s(x)$.

12.5. КОНЕЧНЫЕ ПОЛЯ

Конечное поле, по определению, состоит из конечного числа элементов. Оно имеет простую характеристику p и его порядок равен $q=p^n$. Конечные поля называются также *полями Галуа*. Зафиксируем некоторое конечное поле порядка p^n и обозначим его через $GF(p^n)$.

Теорема 7. *Любой элемент x поля $G=GF(p^n)$ удовлетворяет уравнению $x^q=x$, где $q=p^n$.*

Доказательство. Любой ненулевой элемент $x\in G$ лежит в мультипликативной группе поля порядка $q-1$ и потому удовлетворяет соотношению $x^{q-1}=1$, из которого следует, что $x^q=x$. Нуль также удовлетворяет этому уравнению.

Следствие. *Для любого $x_i\in G$ в кольце $G[x]$ мы имеем $(x-x_i)\mid(x^q-x)$.*

Доказательство. Так как $x_i^q-x_i=0$, это следует из теоремы 10 гл. 11.

Теорема 8. *Имеем $x^q-x=\prod_{x_i\in G}(x-x_i)$, где $q=p^n$.*

Доказательство. Согласно следствию из теоремы 7 и следствию 4 §11.6, многочлен x^q-x делится на произведение $\prod_{x_i\in G}(x-x_i)$, так как сомножители попарно взаимно просты. С другой стороны, степени этих многочленов одинаковы, а старшие коэффициенты равны единице.

Теорема 9. *Мультипликативная группа G^* любого конечного поля $GF(p^n)$ циклична.*

Доказательство. Эта группа абелева порядка $q-1$. Допустим, что она не циклична. Тогда по теореме 16 гл. 7 должен существовать собственный делитель r числа $q-1$, такой, что $x_i^r=1$ для всех $x_i \neq 0$ в G^*. Такое же рассуждение, как и выше, показывает тогда, что

$$\prod_{x_i \in G} (x-x_i) \mid (x^r-1), \qquad r<q-1.$$

Но это невозможно, ибо произведение имеет большую степень, чем r.

Следствие 1. *Всякое конечное поле характеристики p является простым алгебраическим расширением поля $\mathbf{Z}_p$.*

Действительно, если c порождает мультипликативную циклическую группу G^*, то подкольцо, порожденное c, содержит $G^* \sqcup \{0\} = GF(p^n)$.

Объединяя следствие 1 с теоремой 4, получаем

Следствие 2. *Любое конечное поле $GF(p^n)$ изоморфно $\mathbf{Z}_p[x]/(m(x))$, где $m(x)$ — подходящий многочлен степени n, неприводимый над $\mathbf{Z}_p$.*

Заметим, что непременно $m(x) \mid (x^{q-1}-1)$.

Следствие 3. *Для всякого $m \mid (p^n-1)$ поле $GF(p^n)$ содержит $\varphi(m)$ элементов мультипликативного порядка m, где φ — функция Эйлера.*

Доказательство. Пусть $[G^*, \cdot]$ — мультипликативная группа $GF(p^n)$ и c — ее образующая порядка $q-1=p^n-1$. Для каждого делителя m числа $q-1$ порядок m будут иметь в точности элементы $c^{k\frac{q-1}{m}}$, где $k=1, 2, \ldots, m$ и k взаимно просты с m. Таких чисел k имеется $\varphi(m)$.

Пример 5. Рассмотрим поле $GF(16)=GF(2^4)$. Как в примере 6 гл. 11, имеем

$$x^{15}-1=(x-1)(x^2+x+1)(x^4+x^3+x^2+x+1)c_8(x). \tag{15}$$

Многочлен $c_8(x)$ неприводим над $\mathbf{Z}$, однако над $\mathbf{Z}_2$

$$c_8(x)=(x^4+x^3+1)(x^4+x+1). \tag{16}$$

Из теоремы 4 теперь получаем:

$$\begin{aligned} GF(16) &\cong \mathbf{Z}_2[x]/(x^4+x^3+1) \cong \\ &\cong \mathbf{Z}_2[x]/(x^4+x+1) \cong \\ &\cong \mathbf{Z}_2[x]/(x^4+x^3+x^2+x+1). \end{aligned} \tag{17}$$

Имеем $GF(4) \subset GF(16)$; это подполе состоит из 0,1 и двух корней многочлена x^2+x+1. Их мультипликативный порядок в G^*

равен 3. Но $GF(8) \not\subseteq GF(16)$, ибо порядок любого собственного расширения $GF(8)$ равен 8^r, $r > 1$.

При изоморфизме $\mathbf{Z}_2[x]/(x^4+x^3+1) \cong GF(16)$ класс x имеет мультипликативный порядок 15 в G^*, совпадающий с порядком всей группы. Такие элементы называются *примитивными*. Однако при естественном изоморфизме

$$\mathbf{Z}_2[x]/(x^4+x^3+x^2+x+1) \cong GF(16)$$

класс x переходит в элемент мультипликативного порядка 5, ибо

$$x^5-1=(x-1)(x^4+x^3+x^2+x+1)=0.$$

Этот элемент не примитивен в $GF(16)$.

Можно показать, что элемент x примитивен в $\mathbf{Z}_2[x]/(m(x)) \cong GF(2^n)$ тогда и только тогда, когда $m(x)$ — примитивный многочлен.

УПРАЖНЕНИЯ Б

1. Для каких из значений $k=1, 2, 3, 4, 5, 6$ факторкольцо $\mathbf{Z}_7[x]/(x^2+k)$ является полем?

2. Описать поле Галуа $\mathbf{Z}_5[x]/(x^2+3)$, дать таблицу обратных элементов $(a+b\sqrt{-3})^{-1}$ и степеней α^r (α — примитивный элемент).

3. Пусть x_1, x_2, x_3 — корни кубического многочлена $x^3+x+1=0$. Найти многочлен с корнями x_1^2, x_2^2, x_3^2. (*Указание*: отыскать уравнение вида $x^6+c_1x^4+c_2x^2+c_3=0$, которое следует из $x^3=-x-1$.)

4. а) Доказать, что если многочлен $p(x)=a_0x^n+\ldots+a_n$ неприводим над полем F, то $a_n \neq 0$.

б) Доказать, что корни уравнения $a_nx^n+a_{n-1}x^{n-1}+\ldots+a_0$ обратны корням $p(x)$.

5. а) Доказать, что если α — корень уравнения $x^n-1=0$, то любая степень α является корнем этого уравнения.

б) Доказать, что любой корень этого уравнения является степенью примитивного.

Кольцом кватернионов над коммутативным кольцом R называется совокупность выражений вида $q=a_0+a_1i+a_2j+a_3k$, $a_i \in R$, с дистрибутивным умножением, определенным следующим образом: $i^2=j^2=k^2=-1$, $ij=-ji=k$, $jk=-kj=i$, $ki=-ik=j$, причем все элементы R перестановочны с i, j, k.

* 6. Доказать, что в кольце кватернионов над $\mathbf{R}$ все ненулевые элементы обратимы.

* 7. Показать, что это неверно для колец кватернионов над $\mathbf{Z}_p$.

12.6. ВЫЧИСЛЕНИЯ В $GF(2^n)$

Рассмотрим поле Галуа $G=GF(2^n)$ характеристики 2 как векторное пространство над $\mathbf{Z}_2=\{0, 1\}$ с фиксированным базисом $1, c, \ldots, c^{n-1}$, где c — корень неприводимого многочлена степени n

на $\mathbf{Z}_2$. Так, в примере 5 можно использовать $m(x)=x^4+x^3+1$ (см. (17)).

Отождествим каждый многочлен $\sum_{k=1}^{n} b_k c^{n-k}$ (как элемент поля $GF(2^n)$) с двоичным словом $b=b_1b_2\ldots b_n$. Например, в $GF(16)\cong \mathbf{Z}_2[x]/(x^4+x^3+1)$ нуль отождествляется с 0000, единица — с 0001, c — с 0010. Нетрудно вычислить далее:

$$\begin{array}{lll} c=0010, & c^6=1111, & c^{11}=1101,\\ c^2=0100, & c^7=0111, & c^{12}=0011,\\ c^3=1000, & c^8=1110, & c^{13}=0110,\\ c^4=1001, & c^9=0101, & c^{14}=1100,\\ c^5=1011, & c^{10}=1010, & c^{15}=0001. \end{array}$$

Сложение совпадает со сложением векторов над $\mathbf{Z}_2$. Умножение производится по формуле $c^rc^s=c^{r+s}$, где показатели складываются по модулю 15. Чтобы легче отыскивать векторы в таблице, можно переписать ее в другом порядке:

$$\begin{array}{lll} 0001=c^{15}, & 0110=c^{13}, & 1011=c^5,\\ 0010=c, & 0111=c^7, & 1100=c^{14},\\ 0011=c^{12}, & 1000=c^3, & 1101=c^{11},\\ 0100=c^2, & 1001=c^4, & 1110=c^8,\\ 0101=c^9, & 1010=c^{10}, & 1111=c^6. \end{array}$$

12.7. КОДЫ БОУЗА — ЧОУДХУРИ — ХОККЕНГЕМА

Один класс эффективных полиномиальных кодов с исправлением многократных ошибок был открыт около 1960 г. независимо Боузом, Чоудхури и Хоккенгемом. Существует систематический способ построения таких БЧХ-кодов любой длины. Число контрольных символов зависит от числа ошибок, которое мы хотим обнаруживать или исправлять. В случае кода с исправлением ошибок метод определения ошибочных символов достаточно прост.

Алфавит БЧХ-кода отождествляется с некоторым конечным полем $GF(q)$. На практике из-за использования двоичных устройств q обычно является степенью двойки. Кодирующий многочлен $g(x)$ имеет коэффициенты в этом поле, а кодовые слова состоят из всех кратных многочлена $g(x)$, как в гл. 11.

Расстоянием между двумя словами длины n называется число позиций, в которых они отличаются. Например, над $GF(3^6)$ слова 012021 и 021012 отстоят друг от друга на расстояние 4. Покажем, как построить код, у которого минимальное расстояние между словами не меньше d.

Кодирующий многочлен этого кода строится так. Выберем r так, чтобы $q^r = p^{rs} \geqslant d+1$. Затем возьмем примитивный элемент $\alpha \in GF(q^r)$, мультипликативный порядок которого равен $q^r - 1$. Обозначим через $m_1(x)$ минимальный многочлен для α, через $m_2(x)$ — для α^2 и т. д. Мы покажем, что кодирующий многочлен

$$g(x) = \text{н.о.к.}\,[m_1(x),\ m_2(x),\ \ldots,\ m_{d-1}(x)] \qquad (18)$$

доставляет код с кодовыми словами длины $q^r - 1$ и расстоянием между словами, не меньшим d.

Пример 6. Построим двоичный ($q = 2$) БЧХ-код с длиной кодовых слов $n = 15$ и наименьшим расстоянием $d = 5$. Выберем в $GF(2^4)$ примитивный элемент α, корень многочлена $x^4 + x^3 + 1$. Построим далее минимальные многочлены $m_i(x)$ для α^i.

Так как в поле характеристики p имеем $f(\alpha)^{p^i} = f(\alpha^{p^i})$ для любого многочлена f, степени α^2 и α^4 удовлетворяют тому же уравнению, что и α, откуда

$$m_1(x) = m_2(x) = m_4(x) = x^4 + x^3 + 1.$$

Аналогично, α^3, α^6, α^{12} и $\alpha^{24} = \alpha^9$ имеют общий минимальный многочлен, который равен

$$x^4 + x^3 + x^2 + x + 1.$$

Отсюда вытекает, что наименьшее общее кратное для $m_1(x)$, $m_2(x)$, $m_3(x)$, $m_4(x)$ совпадает с $m_1(x)\,m_3(x)$:

$$g(x) = m_1(x)\,m_3(x) = (x^4 + x^3 + 1)(x^4 + x^3 + x^2 + x + 1) = \\ = x^8 + x^4 + x^2 + x + 1.$$

Степень этого многочлена равна 8. Поэтому каждое кодовое слово длины 15 имеет $15 - 8 = 7$ сигнальных символов. Например, при кодировании умножением на $g(x)$ слово 10001000 с $a(x) = 1 + x^4$ будет закодировано словом 111001100000100, ибо

$$a(x)\,g(x) = x^{12} + x^6 + x^5 + x^2 + x + 1 = b(x).$$

Наименьшее расстояние для этого кода будет не меньше 5. Докажем это в общем виде.

Теорема 10. *Пусть α — примитивный элемент в $GF(q^r)$, и пусть длина кодовых слов не превосходит $q^r - 1$. Тогда БЧХ-код с алфавитом $GF(q)$ и кодирующим многочленом*

$$g(x) = \text{н.о.к.}\,[m_1(x),\ m_2(x),\ \ldots,\ m_{d-1}(x)]$$

имеет минимальное расстояние, равное по крайней мере d.

Доказательство. Прежде всего заметим, что $g(x)$ — многочлен над $GF(q)$ наименьшей степени с корнями α, α^2, α^3, ..., α^{d-1}.

Поэтому каждый кодовый многочлен обращается в нуль в α, α^2, α^3, ..., α^{d-1}.

Мы хотим доказать, что любой многочлен с корнями α, α^2, ..., α^{d-1} имеет не менее d ненулевых коэффициентов. Допустим, что это не так, и придем к противоречию.

Если многочлен $c(x)$ имеет менее d ненулевых членов, его можно записать в виде

$$c(x) = b_1 x^{n_1} + b_2 x^{n_2} + \ldots + b_{d-1} x^{n_{d-1}}.$$

Пусть α, ..., α^{d-1}—его корни. Тогда коэффициенты должны удовлетворять системе линейных соотношений:

$$\sum_{i=1}^{d-1} b_i \alpha^{kn_i} = 0, \qquad k = 1, \ldots, d-1.$$

Матрица коэффициентов A невырожденна, ибо ее определитель Вандермонда равен

$$|A| = \prod_{i>j} (\alpha^{n_i} - \alpha^{n_j}) \neq 0, \qquad n(j) < n(i) < q^r.$$

Следовательно, все b_i равны нулю. Мы пришли к противоречию и завершили доказательство.

12.8. СВОЙСТВА НАИМЕНЬШЕГО РАССТОЯНИЯ

Ограничимся теперь БЧХ-кодами двоичного типа.

Теорема 11. *Можно построить двоичный БЧХ-код с кодовыми словами длины $2^m - 1$ и нечетным минимальным расстоянием d, у которого число контрольных символов не больше $[(d-1)/2]\,m$.*

Пояснение. Причина предположения о нечетности состоит в том, что минимальное расстояние для двоичного кода на самом деле автоматически нечетно.

Доказательство. Согласно теореме 10, для кода со словами длины $2^m - 1$ и расстоянием d достаточно иметь $m(d-1)$ контрольных символов. Учтем теперь, что в случае, когда характеристика равна 2, если $p(\alpha) = 0$, $\alpha \in GF(2^m)$, то $p(\alpha^2) = p(\alpha^4) = \ldots = p(\alpha^{2^i}) = 0$. Значит, минимальный многочлен для α будет одновременно минимальным многочленом для α^2; многочлен для α^3—многочленом для α^6; многочлен для α^5—многочленом для α^{10} и т. д.

Степень $m_i(x)$ не превосходит m, ибо $\alpha^i \in GF(2^m)$. Наименьшее общее кратное $m_1(x)$, ..., $m_{d-1}(x)$ делит произведение $(d-1)/2$ из этих многочленов и потому имеет степень $\leqslant [(d-1)/2]\,m$.

В примере 6 имеем $2^m-1=15$, $m=4$ и $k=8$. Наименьшее расстояние равно 5, число контрольных символов — в точности $[(d-1)/2]\,m=[(5-1)/2]\cdot 4=8$.

Коды Хэмминга являются БЧХ-кодами. Это коды над $\mathbf{Z}_2$ с наименьшим расстоянием 3; они способны исправлять одинарную ошибку.

Эффективность БЧХ-кодов с обнаружением ошибок можно продемонстрировать на следующем примере. В европейских системах передачи данных широко используется двоичный (231,255)-код. Он построен над $\mathbf{Z}_2$ с помощью примитивного элемента α из $GF(2^8)$ мультипликативного порядка $2^8-1=255$. Степень кодирующего многочлена равна 24. Его корнями являются также α, α^2, α^3, α^4, α^5 и α^6, что обеспечивает минимальное расстояние 7 и обнаружение до шести ошибок. Разумеется, будут обнаружены многие другие комбинации ошибок. Поскольку в общем числе слов длины 255 доля кодовых слов составляет $2^{-24}\approx 1/16\times 10^6$, то при вводе случайных слов лишь примерно одно из шестнадцати миллионов оказалось бы кодовым. В течение многих лет эксплуатации европейских систем связи не было случая, чтобы ошибка передачи прошла незамеченной при декодировании.

Используются также другие коды, обнаруживающие меньшее число ошибок. Кодирующие и декодирующие устройства производятся разными фирмами.

БЧХ-коды умеренной длины не слишком далеки от «совершенных» или «квазисовершенных» кодов гл. 8. Первый фактически реализованный БЧХ-код был (92,127)-кодом с исправлением пяти ошибок, отвечающий многочлену

$$g(x)=m_1(x)\,m_3(x)\,m_5(x)\,m_7(x)\,m_9(x)=$$
$$=(1+x+x^7)(1+x+x^2+x^3+x^7)(1+x^2+x^3+x^4+x^5+x^7)\times$$
$$\times(1+x+x^2+x^4+x^5+x^6+x^7)(1+x+x^2+x^3+x^4+x^5+x^7).$$

При передаче сообщения длины 1000 по двоичному симметричному каналу с вероятностью ошибки $q=0.01$ в одной позиции вероятность ошибочного приема (с использованием этого кода) примерно равна 1.2×10^{-6}. Если вероятность ошибки в одной позиции равна 0.006, вероятность ошибочного приема снижается до 7.7×10^{-9}. Наконец, при $q=0.002$ получаем 5.8×10^{-14}.

С ростом длины кодовых слов качество БЧХ-кодов ухудшается. Поэтому не прекращается поиск кодов, эффективных для слов большой длины.

УПРАЖНЕНИЯ В

1. Рассмотрим БЧХ-код, порожденный над $\mathbf{Z}_2$ многочленом

$$p(x)=1+x+x^2+x^4+x^5+x^8+x^{10}.$$

Показать, что он отвечает примитивному элементу $\alpha \in GF(2^4)$, корню многочлена $m_1(x) = 1 + x + x^2 + x^3 + x^4$. Показать, что $m_3(x) = 1 + x + x^4$ и $m_5(x) = 1 + x + x^2$ — многочлены с корнями α^3, α^5.

2. Построить БЧХ-код с исправлением двойных ошибок и кодовыми словами длины 15. Как находится его порождающий многочлен?

*12.9. ПОЛЯ РАЗЛОЖЕНИЯ

Оставшаяся часть этой главы посвящена введению в один из самых красивых разделов чистой математики — теорию Галуа. Галуа создал эту теорию, применил ее к доказательству неразрешимости в радикалах общего уравнения пятой степени, а также ввел конечные поля и изучил их структуру. По этой причине их часто называют *полями Галуа*.

Мы покажем в конце концов, что для каждой степени простого числа p^n существует единственное поле Галуа $GF(p^n)$ порядка p^n. Его можно охарактеризовать как поле разложения уравнения $x^{p^n} = x$. В конкретных вычислениях удобнее задавать его как поле разложения примитивного неприводимого многочлена с коэффициентами в $\mathbf{Z}_p$. Вот таблица таких многочленов для $p = 2$:

Степень		Степень	
2:	$1 + x + x^2$	9:	$1 + x^4 + x^9$
3:	$1 + x + x^3$	10:	$1 + x^4 + x^{10}$
4:	$1 + x + x^4$	11:	$1 + x^2 + x^{11}$
5:	$1 + x^2 + x^5$	12:	$1 + x + x^4 + x^6 + x^{12}$
6:	$1 + x + x^6$	13:	$1 + x + x^3 + x^4 + x^{13}$
7:	$1 + x^3 + x^7$	14:	$1 + x + x^6 + x^{10} + x^{14}$
8:	$1 + x^2 + x^3 + x^4 + x^8$	15:	$1 + x + x^{15}$

Более обширные таблицы можно найти в книге Marsh R. W., Table of Irreducible Polynomials over $GF(2)$ through degree 19, NASA 1957; Peterson W. W., Error-Correcting Codes, M.I.T Press, 1961.

Определение. Расширение N поля F называется *полем разложения* многочлена $f(x)$ с коэффициентами в F, если:

(i) $f(x)$ разлагается на линейные множители $(x - x_1) \ldots (x - x_n)$ в $N[x]$;

(ii) N порождено F и x_i (как поле).

Поле разложения — это наименьшее поле, в котором $f(x)$ разлагается на линейные множители.

Пример 8. Пусть $f(x) = x^3 - 2$, $F = \mathbf{Q}$. Поле разложения имеет степень 6 над $\mathbf{Q}$ и линейно порождено элементами 1, ω, $\sqrt[3]{2}$,

$\sqrt[3]{2}\,\omega$, $\sqrt[3]{4}$, $\sqrt[3]{4}\,\omega$, где $\omega=(1+\sqrt{3}\,i)/2$ — примитивный кубический корень из единицы. Имеем $\mathbf{Q}(\omega)=\mathbf{Q}[x]/(x^2+x+1)$, ибо $\omega^2+\omega+1=0$ и многочлен x^2+x+1 неприводим над $\mathbf{Q}$. Поле $\mathbf{Q}(\omega)$ является полем разложения для этого многочлена, а также для x^3-1.

Теорема 12. *Пусть многочлен* $m(x)=x^r+\sum_{k=0}^{r-1} f_k x^k$, $f_k\in F$, *неприводим над полем* F. *Тогда над расширением поля* $G=F[x]/(m(x))$ *многочлен* $m(t)\in G[t]$ *приводим: он делится на* $t-\bar{x}$, $\bar{x}=x \bmod m(x)$.

Доказательство. Ясно, что $(t-\bar{x})\mid m(t)$ в $G[t]$. Поэтому остается применить теорему 10 гл. 11.

Пример 9. Пусть $F=\mathbf{Z}_p(u)$ — простое трансцендентное расширение поля $\mathbf{Z}_p$. Многочлен x^p-u над F неприводим. Пусть $G=F[t]/(t^p-u)$. Над G имеем $x^p-u=x^p-\bar{t}^p=(x-\bar{t})^p$. Поэтому G является полем разложения x^p-u над F.

Пример 10. Пусть $GF(p^n)=G$ — конечное поле порядка $p^n=q$. По теореме 6

$$x^q-x=\prod_{x_i\in G}(x-x_i). \tag{19}$$

Поэтому *всякое конечное поле порядка* $p^n=q$ *является полем разложения для* x^q-x.

В следующем параграфе мы покажем, что поле разложения определяется однозначно с точностью до изоморфизма, а сейчас завершим доказательство его существования.

Теорема 13. *Пусть* F — *любое поле и* $f(x)$ — *любой многочлен степени* n *над этим полем. Тогда его поле разложения существует и имеет степень не больше* $n!$ *над* F.

Доказательство. Не теряя общности, можно считать, что $f(x)$ имеет старший коэффициент, равный единице. Разложим $f(x)=p_1(x)\ldots p_r(x)$ на неприводимые множители и построим поле $G_1=F[x]/(p_1(x))$, скажем, для первого множителя. Мы получим поле, в котором у $p_1(x)$ имеется корень y_1 и которое порождено F и этим корнем.

Пусть $f(x)=(x-y_1)f_1(x)$, $f_1(x)=p_1'(x)\ldots p_k'(x)$. Применим этот же прием к $f_1(x)$, построив поле G_2, и т. д. Поле G_n будет полем разложения $f(x)$ степени, не большей $n!$; доказательство завершено.

Кратные корни. Пусть $G=F[x_1, \ldots, x_n]$ — поле разложения многочлена f над F: $f(x)=(x-x_1)\ldots(x-x_n)$. Допустим, что среди сомножителей $x-x_i$ есть одинаковые, т. е. $x_i=x_j$,

$i \neq j$. Тогда

$$f'(x) = (x - x_i)^2 g(x) + 2(x - x_i).$$

Таким образом, мы доказали следующую лемму (см. теорему 13 гл. 11).

Лемма 1. *Если $f(x)$ имеет кратный корень x_i, то f и f' делятся на $x - x_i$.*

Лемма 2. *Если $q = p^n$, то многочлен $x^q - x$ не имеет кратных корней в поле характеристики p.*

Действительно, $(x^q - x)' = qx^{q-1} - 1 = -1$.

Следствие. *В любом поле разложения многочлен $x^q - x$ над $\mathbf{Z}_p$ разлагается на q различных линейных множителей.*

Теорема 14. *Пусть $q = p^n$; тогда любое поле разложения G многочлена $x^q - x$ состоит из p^n элементов.*

Доказательство. Это поле содержит все корни, а их имеется ровно q в силу следствия леммы 2. С другой стороны, множество этих корней образует поле, ибо оно замкнуто относительно сложения, умножения и взятия обратного элемента (для ненулевых корней).

Следствие. *Существуют поля Галуа любого порядка p^n.*

Примитивные многочлены. Уравнение $x^{p^n - 1} = 1$ имеет $\varphi(p^n - 1)$ *примитивных корней* в $GF(p^n)$. Все они различны по следствию из леммы 2. Каждый из них порождает поле $GF(p^n)$ над $\mathbf{Z}_p$ и является одним из n корней неприводимого многочлена степени n. Отсюда получаем такой результат.

Теорема 15. *Над $\mathbf{Z}_p$ имеется $\varphi(p^n - 1)/n$ многочленов степени n со старшим коэффициентом, равным единице, корни которых имеют период $p^n - 1$.*

* 12.10. ИЗОМОРФИЗМ ПОЛЕЙ РАЗЛОЖЕНИЙ

Пусть G, H — два поля разложения многочлена $f(x)$. Мы хотим доказать, что они изоморфны. Идея состоит в «продолжении» изоморфизмов. Объясним ее на примере простого расширения.

Пусть $u_1, \ldots, u_n$ и $v_1, \ldots, v_n$ — корни $f(x)$ в полях G и H соответственно. Предположим, что u_1 и v_1 являются корнями одного и того же неприводимого многочлена F. Тогда отображение

$$a_0 + a_1 u_1 + \ldots + a_{n-1} u_1^{n-1} \mapsto a_0 + a_1 v_1 + \ldots + a_{n-1} v_1^{n-1} \qquad (20)$$

определяет изоморфизм $F(u_1) \to F(v_1)$:

$$G_1 = F(u_1) \cong F[x]/(p_1(x)) \cong F(v_1) = H_1.$$

Он тождествен на F, т. е. является продолжением тождественного изоморфизма 1_F: $F \longrightarrow F$.

Чтобы иметь возможность итерировать эту конструкцию, нужно несколько усилить ее.

Лемма. *Пусть дан изоморфизм полей* $\alpha: F \longrightarrow F'$, *и пусть* u, v — *корни неприводимых многочленов* $p(x)$, $q(x)$ *соответственно:*

$$p(x) = \sum_{k=0}^{n} b_k x^k, \qquad q(x) = \sum_{k=0}^{n} b'_k x^k, \qquad b'_k = \alpha(b_k). \tag{21}$$

Тогда отображение

$$a_0 + a_1 u + \ldots + a_{n-1} u^{n-1} \mapsto a'_0 + a'_1 v + \ldots + a'_{n-1} v^{n-1},$$

где $a'_k = \alpha(a_k)$, *является продолжением* α *до изоморфизма* $\overline{\alpha}$: $F(u) \cong F(v)$.

Применяя этот результат несколько раз, получаем следующее утверждение.

Теорема 16. *Пусть* α: $F \longrightarrow F'$ — *изоморфизм полей, переводящий многочлен* $p(x)$ *над* F *в многочлен* $q(x)$ *над* F'. *Тогда его можно продолжить до изоморфизма поля разложения* p *с полем разложения* q *над* F.

Единственность поля разложения (с точностью до изоморфизма) следует из частного случая этой теоремы: $\alpha = 1_F$.

Группы Галуа. Пусть теперь $p(x)$ неприводим над F, и пусть $N = F(u_1, \ldots, u_n)$ — поле разложения p над F. *Группой Галуа* $G(N/F)$ называется группа всех автоморфизмов N, тождественных на F.

Теорема 17. *Группа Галуа поля разложения любого неприводимого многочлена над* F *транзитивна на корнях этого многочлена.*

Действительно, из леммы следует существование автоморфизма φ, тождественного на F и переводящего любой корень u_1 в любой другой.

В частности, порядок группы Галуа делится на n и делит $n!$, где r — степень многочлена.

Нетрудно вычислить группу Галуа любого конечного поля $GF(p^n)$ над его простым подполем $\mathbf{Z}_p$. Пусть $c \in GF(p^n)$ порождает мультипликативную группу ненулевых элементов поля. Любой автоморфизм α поля должен переводить c в некоторую степень c^r. Поэтому для любого $x = c^k$ имеем

$$\alpha(x) = \alpha(c^k) = [\alpha(c)]^k = c^{rk} = x^r,$$

а также, очевидно, $\alpha(0) = 0^r$, $\alpha(1) = 1^r$. Итак, любой автоморфизм имеет вид $x \mapsto x^r$ для подходящего r.

Так как $x^{r+q-1}=x^r$ и композиция α: $x\mapsto x^r$ и β: $x\mapsto x^s$ имеет вид $x\mapsto x^{rs}$, группа Галуа изоморфна некоторой факторгруппе подгруппы $(\mathbf{Z}/(q-1)\mathbf{Z})^*$. Отсюда следует, что она циклична. Поскольку она транзитивна на n корнях минимального многочлена для c, она должна быть циклической порядка n.

Пример 11. Пусть $F(x_1, \ldots, x_n)$ — поле рациональных функций от n переменных над F характеристики ∞. Пусть

$$S=F(s_1, \ldots, s_n)$$

— подполе, порожденное F и элементарными симметрическими многочленами $s_j=x_1^j+\ldots+x_n^j$ $(j=1, \ldots, n)$. Основная теорема о симметрических функциях утверждает, что S содержит все рациональные функции, инвариантные относительно перестановок переменных. Отсюда можно вывести, что группа Галуа поля $F(x_1, \ldots, x_n)$ над S является симметрической группой степени n.

УПРАЖНЕНИЯ Г

1. Пусть A — любое множество автоморфизмов поля F. Показать, что элементы $x\in F$, инвариантные относительно всех $\alpha\in A$, образуют подполе $S(A)\subset F$.

2. Пусть T — любое подполе поля F. Показать, что автоморфизмы α поля F, для которых $\alpha(x)=x$ при всех $x\in T$, образуют группу $G(T)$, подгруппу $\operatorname{Aut} F$.

3. В обозначениях упр. 1 и 2 показать, что

$$G(S(A))\supset A, \qquad S(G(T))\supset T,$$
$$S(G(S(A)))=S(A), \qquad G(S(G(T)))=G(T).$$

(*Указание:* из $T\supset T_1$ следует, что $G(T)\subset G(T_1)$; из $A\supset A_1$ следует, что $S(A)\subset S(A_1)$.)

4. Пусть α алгебраично над $\mathbf{Z}_p$. Показать, что группа Галуа $\mathbf{Z}_p[\alpha]$ над $\mathbf{Z}_p$ изоморфна $\operatorname{Aut}\mathbf{Z}_p[\alpha]$ (можно считать известным, что $\mathbf{Z}_p[\alpha]$ — поле разложения).

5. Вычислнть группу Галуа $\mathbf{Z}_5[x]/(x^2+3)$ над $\mathbf{Z}_5$.

6. Вычислить группу Галуа $GF(16)$ над $GF(4)$. Явно описать нетривиальный автоморфизм.

7. Каково поле разложения многочлена x^6-1 над $\mathbf{Q}$? Над $\mathbf{Z}_7$?

*8. Сделать упр. 6 для $GF(81)$ над $GF(9)$.

9. Описать группу Галуа $\mathbf{Q}[\omega]/\mathbf{Q}$, если

а) $\omega=(1+\sqrt{-3})/2$; б) ω — примитивный корень из 1 степени 5.

*10. Описать группу Галуа поля разложения x^3-2.

*11. То же для x^5-3.

СПИСОК ЛИТЕРАТУРЫ

1. Artin E., Galois Theory, Notre Dame Math. Lectures, № 2, 1944, 1959.
2. Berlekamp E. R., Algebraic Coding Theory, McGraw-Hill, 1968. (Русский перевод: Берлекэмп Э., Алгебраическая теория кодирования, «Мир», М., 1971.)
3. Peterson W. W., Error-Correcting Codes, M.I.T. Press, 1961. (Русский перевод: Питерсон У., Коды, исправляющие ошибки, «Мир», М., 1964.)

13

РЕКУРРЕНТНЫЕ ПОСЛЕДОВАТЕЛЬНОСТИ

13.1. РАДАР и СИСТЕМЫ СВЯЗИ

Важным изобретением начала 1940 годов был радар, или радиолокатор — электронная система для обнаружения и локации объектов в атмосфере и пространстве. Его действие основано на испускании радиоволн и приеме части отраженной объектом энергии (см. рис. 13.1). Радиолокация используется ныне при слежении за спутниками, в аэропортах и многих других областях.

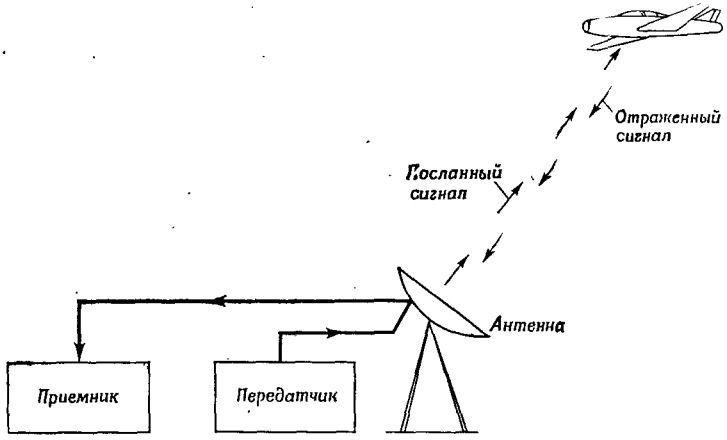


Рис. 13.1. Схема радиолокационной системы.

Антенна радиолокационной системы излучает электромагнитную энергию узким пучком. Направление этого пучка меняется при движении антенны. Частота f излучения находится в пределах $2 \cdot 10^8 - 10^{10}$ герц; длина волны $\lambda = c/f$ — соответственно в пределах 3 см — 1,5 м (ибо скорость света $c \sim 300\,000$ км/сек).

Когда пучок облучает некоторый объект (цель), часть его энергии отражается назад к антенне. Поэтому сама по себе регистрация отраженной энергии служит указанием на наличие объекта в направлении излучающей антенны. Приемная антенна (которая может совпадать с излучающей), как правило, также

обладает резко выраженной направленностью действия, что увеличивает точность локации направления, т. е. угловых координат объекта относительно антенны.

Третья необходимая координата — радиальная, т. е. расстояние от антенны до объекта. Принцип ее определения состоит в измерении промежутка времени между моментом излучения импульса энергии и моментом прихода его отражения. Разделив результат на 2 и умножив на скорость света, получим расстояние.

Здесь приходится преодолевать ряд трудностей. Энергия отраженного сигнала весьма мала; ее надо зарегистрировать на фоне естественного шума; с удалением объекта эта энергия все уменьшается. Поэтому для уверенного обнаружения нужно либо суммировать полученный сигнал по длинным интервалам времени, либо посылать короткие, но очень мощные импульсы.

Проектировщик радиолокационной системы вынужден сделать выбор. Чем короче импульсы, тем лучшее разрешение по расстоянию достижимо; чем они длиннее (при заданной мощности), тем больше предельное расстояние, на котором может быть обнаружен объект. Предпочтительнее всего было бы посылать импульсы одновременно очень короткие и очень мощные, но здесь нас ограничивают возможности передатчиков.

Были предложены различные способы обойти эту трудность. Один из них состоит в том, чтобы модулировать частоту импульса и затем сравнивать характер модуляции отраженного сигнала со стандартным. Это позволяет сохранить точность измерения расстояния, увеличив длину импульса.

Другой способ широко использовался для локации на дальних расстояниях, в частности Луны и Венеры. Он состоит в посылке длинной серии импульсов, закодированной таким образом, чтобы сравнение с отраженным сигналом можно было произвести однозначно.

В этой главе изучается способ кодирования для порождения *периодичных* последовательностей с очень *длинным периодом*. В § 13.6 мы объясним, как они используются в радиолокации. В § 13.9 объясняется, как сравнивать посланный сигнал с отраженным, который ослаблен и искажен шумом.

13.2. РАЗНОСТНЫЕ КОДЫ

Разностный код отображает сигнал длины m, $\mathbf{a}=a_0, a_1, \ldots, a_{m-1}$, в бесконечную *рекуррентную* последовательность $\mathbf{s}=s_0, s_1, s_2, \ldots$, которая определяется по рекурсии системой соотношений *(разностных уравнений)* вида

$$c_0 s_i + c_1 s_{i-1} + \ldots + c_m s_{i-m} = 0, \quad i = m, m+1, \ldots, c_m, \ c_0 \neq 0. \quad (1)$$

Алфавит A, которому принадлежат s_i и c_i, отождествляется с некоторым *конечным полем* $CF(q)$ (см. гл. 12). Начинается кодовая последовательность с блока

$$s_0 = a_0,\ s_1 = a_1,\ \ldots,\ s_{m-1} = a_{m-1}. \tag{1'}$$

Пример 1. Рассмотрим разностное уравнение

$$s_i + 2s_{i-1} + s_{i-2} = 0 \tag{2}$$

над $GF(3) = \mathbf{Z}_3$. Если $s_0 = 1$, $s_1 = 1$ (т. е. $a_0 = 1$, $a_1 = 1$), кодовая последовательность имеет вид $\mathbf{s} = 1, 1, 0, 2, 2, 0, 1, 1, \ldots$. При $s_0 = 1$, $s_1 = 2$ получится $\mathbf{s} = 1, 2, 1, 2, 1, 2, \ldots$. При $s_0 = 0$, $s_1 = 2$ получится $\mathbf{s} = 0, 2, 2, 0, 1, 1, 0, 2, \ldots$ и т. д. Так как имеется $3^2 = 9$ начальных блоков длины 2, разностное уравнение (2) определяет одну из девяти возможных последовательностей.

Кодирующие функции описанного типа физически реализуются с помощью регистров сдвига на триггерах. В случае алфавита $\mathbf{Z}_2$ схема выглядит совсем просто.

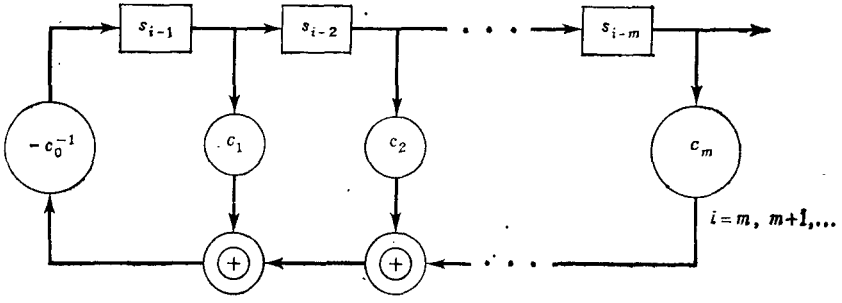


Рис. 13.2.

На рис. 13.2. показана схема, реализующая функцию (1). Имеется m элементов памяти, которые обозначены $s_{i-1}, \ldots, s_{i-m}$. Кроме того, имеется m элементов, умножающих свой вход на коэффициент c_j. Все полученные выходы суммируются и умножаются на $-c_0^{-1}$, что дает очередное значение s_i. В момент времени $t = i$ (или на i-м такте) значение, которое было запомнено ранее в s_j, сдвигается в s_{j-1}, а в s_{i-1} подается

$$-c_0^{-1}(c_{-1}s_{i-1} + c_2 s_{i-2} + \ldots + c_m s_{i-m}).$$

На рис. 13.3 показана схема для примера 1.

Пример 2. Если $GF(2) = \mathbf{Z}_2$, то $c_i = 0$ или 1, так что схема приобретает простой вид. На рис. 13.4 показан регистр сдвига с обратной связью для уравнения $s_i + s_{i-1} + s_{i-3} = 0$. Начав со значений $s_0 = 1$, $s_1 = 0$, $s_2 = 0$, мы получим последовательность

$$\mathbf{s} = \{s_i\}_{i=0}^{\infty} = 100111101 \ldots .$$

Коэффициент $c_i = 0$ или 1 реализуется наличием или отсутствием отвода от выхода триггера.

Кодовые последовательности разностного типа периодичны, как правило, с длинным периодом, что относится к числу их важнейших преимуществ. Метод нахождения периода обсуждается в § 13.6; здесь мы ограничимся доказательством периодичности.

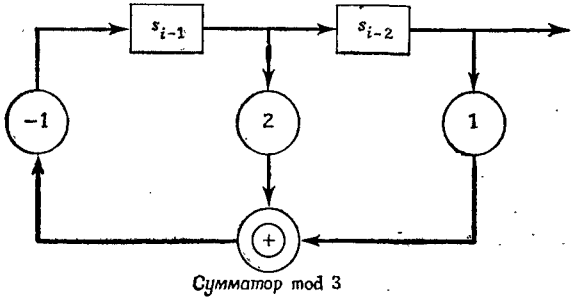


Рис. 13.3.

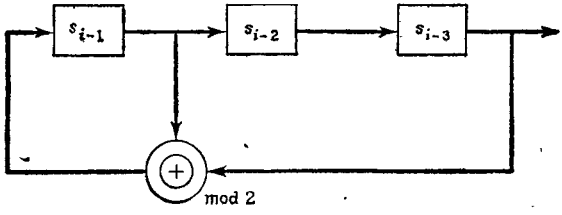


Рис. 13.4.

Определения. Последовательность $\mathbf{s} = s_0 s_1 \ldots$ *периодична*, если существует такое r, что $s_i = s_{i+r}$ для всех $i \geqslant 0$. Наименьшее такое $r = p(\mathbf{s})$ называется *периодом* последовательности.

Последовательность называется *периодичной с некоторого места*, если существуют такие t и r, что

$$s_i = s_{i+r} \tag{3}$$

для всех $i \geqslant t$.

Теорема 1. *Если* $c_m \neq 0$, *то последовательность, определенная разностным кодом* (1), *периодична.*

Пояснение. Последовательность $\mathbf{s}$, очевидно, периодична с некоторого места, ибо множество блоков длины m конечно; поэтому все дело в доказательстве чистой периодичности.

Доказательство. Пусть α—функция «сдвига на единицу» отвечающая уравнению (1):

$$\alpha\colon (s_0, \ldots, s_{m-1}) \mapsto (s_1, \ldots, s_{m-1}, -c_0^{-1}(c_1 s_{m-1} + \ldots + c_m s_0)).$$

Она определяет эндоморфизм аддитивной группы m-векторов над полем $GF(q)$. Ядро этого эндоморфизма нулевое: если $\alpha(s_0,\ldots,s_{m-1})=\mathbf{0}$, то $s_1+0,\ \ldots,\ s_{m-1}=0$ и далее из $-c_m s_0=0$ следует, что $s_0=0$. Следовательно, α есть автоморфизм. Но $\alpha^{r+1}=\alpha$ для подходящего r, откуда следует, что α^r — тождественное отображение. Так как $\alpha^k(s_0,\ldots,s_{m-1})=(s_k,\ldots,s_{m+k-1})$, мы получаем, что r есть период любой кодовой последовательности!

13.3. РАЗНОСТНЫЕ УРАВНЕНИЯ

Уравнения (1) определяют над любым кольцом D, в котором содержится c_i и элемент c_0 обратим, бесконечную последовательность элементов кольца, если только задан ее начальный сегмент длины m:

$$\mathbf{a}=(a_0,a_1,\ldots,a_{m-1})=(s_0,s_1,\ldots,s_{m-1}).$$

Им нно

$$s_n=-\sum_{j=1}^{m-1} c_j s_{n-j} \text{ для всех } n\geqslant m. \tag{4}$$

Если основным кольцом является поле F, то множество всех таких последовательностей $S(\mathbf{c})$, где $\mathbf{c}=(c_0,\ldots,c_m)$, образует m-мерное векторное пространство над F. Если это поле конечно и состоит из q элементов, то число кодовых последовательностей равно q^m.

Следующий классический пример относится к $\mathbf{R}$, $\mathbf{Q}$, $\mathbf{Z}$ или $\mathbf{N}$.

Пример 3. *Уравнение Фибоначчи* имеет вид

$$s_{n+1}=s_n+s_{n-1},\qquad s_i\in\mathbf{R},\ n\in\mathbf{P}. \tag{5}$$

Начальные значения $s_0=0$, $s_1=1$ приводят к последовательности Фибоначчи $0,1,1,2,3,5,8,13,21,34,55,89,\ldots$.

Мы можем рассмотреть ее в $GF(p)$, т. е. в $\mathbf{Z}_p$. При $p=11$ получим

$$0,1,1,2,3,5,8,2,10,1,0,1,1,2,3,5,8\ldots .$$

Эта последовательность *периодична* с периодом $T=10$.

Пусть F^ω — пространство бесконечных последовательностей $s_0s_1s_2\ldots$ элементов из F. Определим *операторы правого и левого сдвига*:

$$X\colon\ (s_0,s_1,s_2,\ldots)\mapsto(0,s_0,s_1,\ldots), \tag{6}$$

$$X^{-1}\colon\ (s_0,s_1,s_2,\ldots)\mapsto(s_1,s_2,s_3,\ \ldots). \tag{6'}$$

Очевидно, $X^{-1}\circ X=X\diamond X^{-1}=1$. Уравнение (1) можно записать в виде

$$[c_0I+c_1X^{-1}+\ldots+c_mX^{-m}]\,\mathbf{s}=\mathbf{0}=(000\ldots). \tag{7}$$

Многочлен $c(x)=c_0+c_1x+\ldots+c_mx^m$ называется *левым* характеристическим многочленом уравнения (1), а многочлен с обратными корнями:

$$\bar{c}(x)=c_0x^m+c_1x^{m-1}+\ldots+c_m,$$

называется *правым* характеристическим многочленом. Полезно рассматривать и тот, и другой.

Характеристические решения. Над комплексным полем правый характеристический многочлен позволяет вычислять элементарные решения уравнения (1).

Теорема 2. *Последовательность* $(1,\alpha,\alpha^2,\ldots)$ *удовлетворяет уравнению* (1) *в том и только том случае, если* $c(\alpha)=0$.

Доказательство. Подставляя в (1), находим

$$\sum_{j=0}^{m} c_j s_{i-j} = \sum_{j=0}^{m} c\alpha^{i-j} = \alpha^{i-m}\, c(\alpha) \text{ для всех } i,$$

что делает утверждение очевидным.

Например, для уравнения Фибоначчи характеристический многочлен равен x^2-x-1, а его корни суть $\alpha_{1,\,2}=(1\pm\sqrt{5})/2$. Поэтому общее комплексное решение уравнения (5) таково:

$$s_k = a\left(\frac{1+\sqrt{5}}{2}\right)^k + b\left(\frac{1-\sqrt{5}}{2}\right)^k, \tag{8}$$

где a, b — любые комплексные константы. Так как второй член стремится к нулю, любое решение с $a\neq 0$ будет расти как $s_{k+1}/s_k \longrightarrow (1+\sqrt{5})/2 \simeq 1.618034\ldots$. Например, $89/55 = 1.6181818\ldots$.

Для уравнения $u_{n+1}=2u_n+u_{n-1}$ корнями характеристического уравнения будут $1\pm\sqrt{2}$. При начальных значениях 0, 1 продолжением будет 2, 5, 12, 29, 70, 169, ... Отношение

$$169:70 = 2.4142857142857\ \ldots$$

является хорошим приближением к $1+\sqrt{2}=2.414214\ \ldots$.

УПРАЖНЕНИЯ А

1. Вычислить первые 10 членов каждой из 8 последовательностей над $\mathbf{Z}_2$, удовлетворяющих $s_i=s_{i-2}+s_{i-3}$.

2. Для того же уравнения над $\mathbf{Z}_3$ выписать первые 10 членов каждой из трех последовательностей, начинающихся с 100, 010, 001. Пользуясь ими, вычислить последовательность 111 ... .

3. Определить период каждой из последовательностей в упр. 1.

4. То же для упр. 2.

5. а) Каковы вещественные корни характеристического уравнения для $s_i=2s_{i-1}+2s_{i-2}$?

б) Найти рациональное приближение к $\sqrt{3}$ с точностью 0.001 с помощью последовательности $1,3,\ldots$.

6. а) Каковы вещественные корни характеристического уравнения для $s_i = s_{i-1} + 3s_{i-2}$?

б) Найти рациональное приближение к $\sqrt{13}$ с точностью 0.001 с помощью последовательности $1,2,\ldots$.

7. а) Показать, что характеристический многочлен для $s_n = s_{n-3}$ приводим над любым полем F.

б) Использовать это для отыскания базиса решений.

в) Обсудить вопрос о периодах этих решений над $\mathbf{Z}_p$ (p — любое простое число).

8. Пусть $s_n = a_1 s_{n-1} + a_2 s_{n-2}$. Найти рекуррентное соотношение для $r_i = s_i/s_{i-1}$ в виде $r_i = F(r_{i-1})$.

*9. Метод Ньютона для вычисления большего корня уравнения $x^2 - a_1 x - a_2 = \varphi(x) = 0$ состоит в применении итерации $X_i = F(X_{i-1})$, $F(x) = x - [\varphi(x)/\varphi'(x)]$. Показать, что он сходится быстрее, чем метод упр. 8.

13.4. ФОРМАЛЬНЫЕ СТЕПЕННЫЕ РЯДЫ

Над коммутативным кольцом R можно рассматривать не только многочлены, но также формальные степенные ряды, т. е. выражения вида

$$a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n + \ldots = \sum_{n=0}^{\infty} a_n x^n, \qquad a_n \in R. \tag{9}$$

Слово «формальный» употребляется здесь, чтобы подчеркнуть следующее обстоятельство: сходимость или расходимость ряда не принимается во внимание и чаще всего (скажем, над конечными полями) даже неопределена.

Формальные ряды очень полезны при анализе рекуррентных последовательностей.

Определение. Характеристической функцией последовательности $\mathbf{s} = s_0, s_1, s_2, \ldots$ называется формальный ряд

$$s(x) = s_0 + s_1 x + s_2 x^2 + \ldots = \sum_{n=0}^{\infty} s_n x^n. \tag{10}$$

Заметим, что оператор сдвига (6) действует на $s(x)$ как умножение на x.

В следующем параграфе мы покажем, что характеристические функции возвратных последовательностей представляются в виде отношений многочленов степени $\leqslant m-1$: $a(x)/c(x)$; здесь $c(x)$ — характеристический многочлен (7).

Начнем с проверки того, что множество формальных рядов $R[[x]]$ над коммутативным кольцом R само является коммутативным кольцом.

Сложение и умножение в $R[[x]]$ определены формулами:

$$\sum_{k=0}^{\infty} a_k x^k + \sum_{k=0}^{\infty} b_k x^k = \sum_{k=0}^{\infty} (a_k + b_k) x^k, \tag{11}$$

$$\sum_{k=0}^{\infty} a_k x^k \cdot \sum_{k=0}^{\infty} b_k x^k = \sum_{k=0}^{\infty} \left(\sum_{i=0}^{k} a_i b_{k-i} \right) x^k. \tag{12}$$

Единица в $R[[x]]$ имеет вид

$$1 = 1 + 0x + 0x^2 + \dots . \tag{13}$$

Пример 4. Пусть $R = \{0, 1\} = \mathbf{Z}_2$, $f(x) = 1 + x^3$ и

$$g(x) = 1 + x + x^2 + \dots + x^n + \dots .$$

Тогда $f(x) + g(x) = x + x^3 + x^4 + \dots + x^n + \dots$, и

$$f(x)\, g(x) = 1 + x + 0 \cdot x^2 + 0 \cdot x^3 + \dots + 0 \cdot x^n + \dots ,$$

т. е. $g(x) = \frac{1+x}{1+x^2}$. Заметим, что $1 + x^2 = (1+x)^2$, так что даже $g(x) = \frac{1}{1+x}$.

Теорема 3. *Формулы* (11) *и* (12) *определяют на* $R[[x]]$ *структуру коммутативного кольца.*

Доказательство. Прежде всего, $R[[x]]$ есть абелева группа со сложением (11). Действительно, нулем является ряд $0 = \sum_{k=0}^{\infty} 0 \cdot x^k$; обратен к $\sum_{k=0}^{\infty} a_k x^k$ ряд $\sum_{k=0}^{\infty} (-a_k) x^k$; сумма, очевидно, ассоциативна.

Далее, $R[[x]]$ — коммутативный моноид с единицей $1 + \sum_{k=1}^{\infty} 0x^k$ по умножению (12). Ассоциативность и дистрибутивность проверяются непосредственно.

Обратимые элементы. Ряд $p(x) = \sum_{k=0}^{\infty} p_k x^k \in R[[x]]$ обратим тогда и только тогда, когда существует такой ряд $q(x) = \sum_{k=0}^{\infty} q_k x^k$, что $p(x)\, q(x) = 1$, т. е.

$$\begin{aligned} p_0 q_0 &= 1 \\ p_0 q_1 + p_1 q_0 &= 0 \\ p_0 q_2 + p_1 q_1 + p_2 q_0 &= 0 \\ &\dots \\ p_0 q_n + p_1 q_{n-1} + \dots + p_n q_0 &= 0 \\ &\dots \end{aligned} \tag{14}$$

Рассматривая (14) как систему уравнений для q_i при данных p_i, мы убеждаемся, что необходимое и достаточное условие ее разрешимости состоит в том, чтобы p_0 был обратим в R. После этого остальные p_i определяются по рекурсии: $q_0 = p_0^{-1}$, $q_n = -p_0^{-1}\sum_{i=1}^{n} p_i q_{n-i}$, $n \geqslant 1$. Итак, мы установили следующий результат.

Теорема 4. *Ряд* $\sum_{k=0}^{\infty} p_k x^k$ *обратим в* $R[[x]]$ *тогда и только тогда, когда* p_0 *обратим в* R.

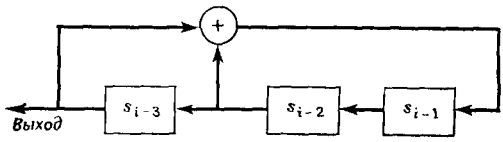
Следствие 1. *Если* F *—поле, то ряд* $\sum_{k=0}^{\infty} p_k x^k \in F[[x]]$ *обратим тогда и только тогда, когда* $p_0 \neq 0$.

Следствие 2. *Любой ненулевой идеал в кольце* $F[[x]]$ *имеет вид* (x^n). *Здесь* n *—наименьшее целое число, такое, что в идеале есть ряд вида* $\sum_{k=n}^{\infty} p_k x^k$ с $p_n \neq 0$.

УПРАЖНЕНИЯ Б

1. Написать 9 последовательностей над $GF(3)$, определенных уравнением $s_i + s_{i-1} + 2s_{i-2} = 0$.

2. Над $GF(2)$ многочлен $1 + x^2 + x^3$ связан со следующим регистром сдвига:



Он также отвечает разностному уравнению

$$c_0 s_i + c_1 s_{i-1} + c_2 s_{i-2} + c_3 s_{i-3} = 0.$$

а) Каковы значения c_0, c_1, c_2 и c_3?

б) Элемент s_i фигурирует в разностном уравнении. Почему в регистре нет соответствующего элемента?

3. Пусть начальные значения $s_{i-3} = 1$, $s_{i-2} = 0$, $s_{i-1} = 0$ (регистр из упр. 2).

а) Какой будет выходная последовательность?

б) Каков ее период?

в) Какую последовательность состояний сменит регистр, пока он не вернется к начальному состоянию?

4. Сравнить выходную последовательность упр. А3 с последовательностью коэффициентов ряда для $(1 + x^2)/(1 + x^2 + x^3)$.

5. Нарисовать регистр сдвига, отвечающий $1 + x^3 + x^4$.

а) Вычислить последовательные состояния регистра с начальным состоянием 0001 вплоть до возвращения к начальному состоянию.

б) Вычислить выходную последовательность.

6. Вычислить $s_0s_1s_2\ldots$ для последовательностей из упр. 5.

7. Построить регистр сдвига из элементов ИСКЛЮЧАЮЩЕЕ ИЛИ, порождающий последовательность

$$11110011110011110\ldots$$

Использовать как можно меньше триггеров. Выписать соответствующий ему многочлен.

8. Рассматриваются регистры сдвига с начальным состоянием $0\ldots01$.
а) Рассмотреть последовательности, отвечающие x^2+1 и x^3+1. Какова их сумма?
б) Какой многочлен порождает сумму? (возможно, со сдвигом).

9. Каков ряд, обратный к $1+x^2$ в $\mathbf{Z}_2[[x]]$?

10. То же в $\mathbf{Z}_3[[x]]$?

11. Какие из следующих рядов обратимы в $\mathbf{Z}[[x]]$? Почему?
а) $1+x^2$, б) $-1+x^3$, в) $2+x^4$,
г) $3+2x^3$; д) $4+6x^4$.

13.5. ПРИЛОЖЕНИЕ К РАЗНОСТНЫМ КОДАМ

Рассмотрим уравнение (1) над полем F с $c_0=1$ и $c_m\neq 0$. Правый и левый характеристический многочлены:

$$c(x)=c_0+c_1x+c_2x^2+\ldots+c_mx^m,$$
$$\overline{c}(x)=c_m+c_{m-1}x+\ldots+c_0x^m,$$

обратимы в $F[[x]]$. Начиная с этого места, мы ограничимся рассмотрением правого многочлена и слово «правый» будем опускать.

Теорема 5. *Последовательность с характеристической функцией*

$$s(x)=\frac{a(x)}{c(x)}=s_0+s_1x+s_2x^2+\ldots,$$

где $a(x)$ — многочлен степени $l<m$, удовлетворяет уравнению (1), *если $c(x)$ — характеристический многочлен для* (1).

Доказательство. Имеем

$$c(x)s(x)=a(x)=a_0+a_1x+\ldots+a_{m-1}x^{m-1}. \tag{15}$$

Поэтому коэффициент при x^n в $c(x)s(x)$ равен

$$c_0s_n+c_1s_{n-1}+\ldots+c_ns_0=0 \text{ для } n\geqslant m. \tag{16}$$

Это и есть уравнение (1) с $c_0=1$.

Следствие. *Любое решение* $\mathbf{s}=s_0,s_1,s_2,\ldots$ *уравнения* (1) *имеет характеристическую функцию*

$$s(x)=s_0+s_1x+\ldots+s_kx^k+\ldots=\frac{a(x)}{c(x)}, \tag{16'}$$

где $c(x)$ — характеристический многочлен для (1).

Пример 5. Пусть $F=\mathbf{Z}_2$, и пусть характеристический многочлен равен $1+x+x^2$. Последовательность с начальным блоком $\mathbf{a}=10$, т. е. $a(x)=1$, отвечает функции

$$s(x)=\frac{1}{1+x+x^2}=(1+x+x^2)^{-1}. \tag{17}$$

Нетрудно проверить, что в $\mathbf{Z}_2[[x]]$

$$s(x)=1+x+x^3+x^4+x^6+x^7+\ldots .$$

Последовательность коэффициентов $\mathbf{s}=110110110\ldots$ *периодична*. Вычисление обратного элемента можно провести, например, деля «уголком».

Пример 6. Рассмотрим уравнение над $\mathbf{Z}_2$:

$$s_i+s_{i-1}+s_{i-3}=0. \tag{18}$$

Его характеристический многочлен есть $1+x+x^3$. Любой многочлен степени $\leqslant 2$ приводит к решению, например, $a(x)=1+x$. Деля «уголком», находим

$$1+x^3+x^4+x^5+x^7+x^{10}+\ldots ,$$

что дает последовательность 10011101001 ... Вот список начальных членов всех восьми кодовых последовательностей:

00000000000 ...
10011101001 ...
01001110100 ...
11010011101 ...
00111010011 ...
10100111010 ...
01110100111 ...
11101001110 ...

Заметим, что каждая из семи ненулевых последовательностей получается сдвигом любой из них. Кроме того, все они периодичны с периодом 7. Мы объясним причины этого в следующем параграфе.

13.6. РЕКУРРЕНТНЫЕ ПОСЛЕДОВАТЕЛЬНОСТИ

Пусть $\mathbf{s}=s_0s_1s_2\ldots$ — некоторая последовательность элементов поля или кольца. Заметим, что множество ее периодов, если оно непусто, состоит из всех кратных наименьшего периода. Действительно, наибольший общий делитель всех периодов есть

период. Это так для н. о. д. двух периодов n, n', ибо если $d=(n, n')$, то $d=kn-ln'$ для подходящих $k>0$, $l>0$, откуда

$$s_i=s_{i+kn}=s_{i+kn-ln'}=s_{i+d} \quad \text{для всех } i\in \mathbf{N}.$$

Отсюда следует, что это справедливо для н. о. д. всех периодов.

Назовем *рекуррентной последовательностью* над полем F всякую последовательность $\mathbf{s}=(s_0, s_1, s_2, \ldots)$, которая удовлетворяет одному или нескольким уравнениям типа (1):

$$s_i+\sum_{j=1}^{m} c_j s_{i-j}=0, \qquad i=m,\ m+1,\ m+2,\ \ldots. \tag{19}$$

Пусть $s(x)$—ее характеристическая функция, $c(x)$—левый характеристический многочлен уравнения (1), рассматриваемый как элемент $F[[x]]$.

Рассмотрим множество *всех* многочленов $\bar{p}(x)=\sum_{k=0}^{r} p_k x^k$, таких, что

$$(p_0 I+p_1 X^{-1}+p_2 X^{-2}+\ldots+p_r X^{-r})\,\mathbf{s}=\mathbf{0}, \tag{20}$$

где X^{-1}—*левый* оператор сдвига.

Лемма. *Множество $J(\mathbf{s})$ этих многочленов является идеалом в кольце $F[x]$.*

Доказательство. Если $\bar{p}(X^{-1})\mathbf{s}=\mathbf{0}$ и $\bar{q}(X^{-1})\mathbf{s}=\mathbf{0}$, то $[\bar{p}(X^{-1})\pm\bar{q}(X^{-1})]\,\mathbf{s}=\mathbf{0}$. Далее, для любого $\bar{r}(x)$

$$[\bar{r}(X^{-1})\bar{p}(X^{-1})]\,\mathbf{s}=\bar{r}(X^{-1})[\bar{p}(X^{-1})\,\mathbf{s}]=\bar{r}(X^{-1})\,\mathbf{0}=\mathbf{0}.$$

(Из рекуррентности последовательности следует, что это множество содержит ненулевой элемент.)

Так как $F[x]$—евклидово кольцо и потому кольцо главных идеалов, отсюда вытекает, что для любой данной рекуррентной последовательности существует единственный многочлен со старшим коэффициентом единица $\bar{c}_s(x)$, такой, что

$$\bar{p}(X^{-1})\mathbf{s}=\mathbf{0} \text{ тогда и только тогда, когда } \bar{c}_s(x)\mid\bar{p}(x). \tag{21}$$

Определение. Многочлен $\bar{c}_s(x)$ со свойством (21) называется *минимальным* многочленом рекуррентной последовательности $\mathbf{s}$.

Отсюда вытекает

Теорема 6. *Если характеристический многочлен уравнения* (1) *неприводим, то любое нетривиальное решение этого уравнения удовлетворяет в точности тем линейным уравнениям типа* (1), *характеристические многочлены которых делятся на $c(x)$.*

Сдвиги. Напомним, что $s_i^* = s_{i+t}$, если $\mathbf{s}^* = X^{-t}\mathbf{s}$. Заметим, что не обязательно $X^t\mathbf{s}^* = \mathbf{s}$, ибо начальные блоки этих последовательностей могут не совпадать.

Пример 7. Над $\mathbf{Z}_2$ уравнение $s_{i+4} = s_i$ с характеристическим многочленом $x^4 + 1 = x^4 - 1$ имеет два решения 0000 ... и 1111 ... периода 1, два решения 0101 ... и 1010 ... периода 2 и двенадцать решений периода 4:

$$00010001 \ldots, \quad 00110011 \ldots, \quad 01110111 \ldots, \qquad (22)$$

и их сдвиги.

Поскольку $x^4 + 1 = (x+1)^4$ над полем $\mathbf{Z}_2[x]$, минимальные многочлены этих последовательностей суть $x+1$, $x^2+1 = (x+1)^2$, $x^3 + x^2 + x + 1 = (x+1)^3$ и $x^4 + 1$.

Легко показать, что сдвиг периодичной последовательности не меняет ее минимального многочлена. Это неверно, однако, для *периодичных с некоторого места* последовательностей. Так, минимальный многочлен для 0111111 ... равен $x^2 + x$ (уравнение $s_{i+2} + s_{i+1} = 0$).

13.7. ПЕРИОДЫ ПОСЛЕДОВАТЕЛЬНОСТЕЙ, СВЯЗАННЫХ С ВЗАИМНО ПРОСТЫМИ МНОГОЧЛЕНАМИ

Вопрос о периодах рекуррентных последовательностей представляет большой интерес. Многое можно сказать о них, если характеристический многочлен последовательности примитивен или по крайней мере известны его неприводимые множители (см. [7]).

Теорема 7. *Пусть* $f(x)$, $g(x)$ *— взаимно простые ненулевые многочлены над полем* F, *и пусть* $h(x) = f(x)\,g(x)$. *Тогда* $S(h(x)) = S(f(x)) + S(g(x))$, *где* $S(f(x)) + S(g(x))$ *— множество всех последовательностей вида* $\mathbf{s} + \mathbf{s}^*$, $\mathbf{s} \in S(f(x))$, $\mathbf{s}^* \in S(g(x))$.

Замечание. Поскольку $f(x)$, $g(x)$ взаимно просты, их произведение совпадает с наименьшим общим кратным.

Пример 8. Многочлены $1 + x + x^2$ и $1 + x + x^3$ взаимно просты. Их произведение равно $1 + x^4 + x^5$. Множество $S(1 + x + x^2)$ состоит из нулевой последовательности и трех последовательностей периода 3, и $S(1 + x + x^3)$ состоит из нулевой последовательности и семи последовательностей периода 7. Их пересечение содержит лишь нулевую последовательность. Множество $S(1 + x^4 + x^5)$ содержит оба предыдущих множества и еще 21 последовательность периода 21, попарные суммы последовательностей из $S(1 + x + x^2)$ и $S(1 + x + x^3)$. Например,

$$\begin{array}{l} 011011011011011011011011\ldots \in S(1+x+x^2) \\ +\,111010011101001110100111010\ldots \in S(1+x+x^3) \\ \hline 100001000110010101111100001\ldots \in S(1+x^4+x^5) \end{array}$$

Последовательности периода 21 можно получить путем 21 сдвига выписанной последовательности.

Доказательство теоремы 7. Прежде всего заметим, что $S(f(x)) \cap S(g(x)) = 0$. Действительно, если $\mathbf{s}$ лежит в $S(f(x))$ и $S(g(x))$, то минимальный многочлен $c_s(x)$ должен быть общим делителем $f(x)$ и $g(x)$, т. е. 1, а тогда $\mathbf{s} = \mathbf{0}$. С другой стороны, $S(h(x))$ содержит сумму $S(f(x)) + S(g(x))$, а размерность ее равна сумме размерностей слагаемых, именно $\deg f + \deg g = \deg h$. Значит, эта сумма исчерпывает $S(h(x))$, что завершает доказательство.

Следствие. *Если* $f_1(x), \ldots, f_r(x)$ *попарно взаимно просты в* $F[x]$, *то* $S(f_1 \ldots f_r) = S(f_1) + \ldots + S(f_r)$.

Порядок. Назовем *порядком* многочлена $f(x) = \sum_{k=0}^{m} f_k x^k$ с $f_0 f_m \neq 0$ наименьшее $e \in \mathbf{P}$, такое, что $f(x) \mid (x^e - 1)$. Будем писать $\operatorname{ord} f(x) = e$.

Теорема 8. *Пусть* $\mathbf{s} \in S(f(x))$, $\mathbf{s}^* \in S(g(x))$, *где* $f(x)$ *и* $g(x)$ *взаимно просты. Тогда период* $p(\mathbf{s} + \mathbf{s}^*)$ *равен наименьшему общему кратному наименьших периодов* $p(\mathbf{s})$ *и* $p(\mathbf{s}^*)$.

Доказательство. Период $p(\mathbf{s} + \mathbf{s}^*) = t$, очевидно, делит н.о.к. r наименьших периодов $p(\mathbf{s})$ и $p(\mathbf{s}^*)$. Покажем, что он также делится на периоды $p(\mathbf{s})$ и $p(\mathbf{s}^*)$.

Для этого достаточно проверить, что t является одним из периодов $\mathbf{s}$ и $\mathbf{s}^*$.

Имеем $s_i + s_i^* = s_{i+t} + s_{i+t}^*$. Поэтому $s_i - s_{i+t} = s_{i+t}^* - s_i^*$. Значит, последовательность $s_i - s_{i+t}$ лежит одновременно в $S(f(x))$ и $S(g(x))$ и потому обращается в нуль. Таким образом, t является периодом как $\mathbf{s}$, так и $\mathbf{s}^*$.

Лемма 1. *Если* $\mathbf{s} \in S(f(x))$, *то* $p(\mathbf{s})$ *делит* $\operatorname{ord} f(x)$.

Доказательство. Пусть $e = \operatorname{ord} f(x)$. Так как $f(x)$ делит $x^e - 1$, имеем $S(f(x)) \subset S(x^e - 1)$. Но e является наименьшим общим периодом всех последовательностей из $S(x^e - 1)$.

Лемма 2. *Пусть* $f(x)$ *— минимальный многочлен для последовательности* $\mathbf{s}$. *Тогда*

$$\operatorname{ord} f(x) = p(\mathbf{s}).$$

Доказательство. Так как $\mathbf{s} \in S(x^{p(\mathbf{s})} - 1)$ и $f(x)$ делит $x^{p(\mathbf{s})} - 1$, имеем $\operatorname{ord} f(x) \leqslant p(\mathbf{s})$. С другой стороны, по лемме 1 $p(\mathbf{s})$ делит $\operatorname{ord} f(x)$.

Следствие. *Если многочлен* $f(x)$ *неприводим, то любая ненулевая последовательность* $\mathbf{s} \in S(f(x))$ *имеет период* $p(\mathbf{s}) = \operatorname{ord} f(x)$.

Действительно, тогда $f(x)$ минимален для этой последовательности, и результат следует из леммы 2.

13.8. ПОСЛЕДОВАТЕЛЬНОСТИ С МАКСИМАЛЬНЫМ ПЕРИОДОМ

Обычно желательно получить от регистра сдвига данной сложности кодовые последовательности с возможно большей длиной периода. Поэтому встает вопрос, каков максимальный период последовательности, порождаемой многочленом степени n.

Теорема 9. *Пусть многочлен $f(x)$ над $GF(q)$ делит $x^{q^n-1}-1$, $\deg(f(x))=n$, и $f(x)$ не делит x^t-1 с $t<q^n-1$. Тогда $p(\mathbf{s})=q^n-1$ для всех ненулевых $\mathbf{s}\in S(f(x))$.*

Доказательство. Если $f(x)$ делит $x^{q^n-1}-1$, то $p(\mathbf{s})$ делит q^n-1 для всех $\mathbf{s}\in S(f(x))$. Допустим, что $p(\mathbf{s})=t<q^n-1$. Тогда $\mathbf{s}\in S(x^t-1)$, так что $f(x)\,|\,(x^t-1)$ вопреки предположению. Значит, $p(\mathbf{s})=q^n-1$ для всех ненулевых $\mathbf{s}$. Любая такая последовательность называется *последовательностью с максимальным периодом*.

Пример 9. Многочлен x^4+x^3+1 над $\mathbf{Z}_2$ делит $x^{15}+1$, но ни один многочлен x^k+1 с $k<15$. С ним связано уравнение $s_i+s_{i-3}+s_{i-4}=0$. Одна из последовательностей имеет вид 100010011010111100 01... . Период ее равен 15 в соответствии с теорией. Все множество $S(x^4+x^3+1)$ состоит из 15 сдвигов этой последовательности и нуля.

Рассмотрим еще раз функцию α: $F^n\to F^n$, связанную с $f(x)$:

$$\alpha(s_0,\ \ldots,\ s_{n-1})=[s_1,\ \ldots,\ s_{n-1},\ -f_0^{-1}(f_1s_{n-1}+\ldots+f_ns_0)].$$

Мы показали, что при $f_0f_n\neq 0$ она является перестановкой множества из q^n-1 ненулевых векторов. Поэтому наибольшее возможное значение для числа t с $\alpha^t=1$, но $\alpha^w\neq 1$ при $w<t$ есть $t=q^n-1$. Это показывает, что период, больший q^n-1, не может реализоваться для многочленов степени n над полем из q элементов.

Это рассуждение показывает также, что последовательные блоки длины n в одном периоде последовательности $\mathbf{s}\in S(f(x))$ с $p(f(x))=q^n-1$ составляют множество всех n-блоков по одному разу. По этим причинам линейные последовательности максимального периода использовались для порождения псевдослучайных чисел.

Заметим еще, что множество последовательностей $S(f(x))$ замкнуто относительно сдвига, сложения и умножения на скаляры.

УПРАЖНЕНИЯ В

1. а) Построить регистр сдвига, отвечающий многочлену степени 3, который, исходя из разных ненулевых начальных состояний, порождает последовательности с разными периодами.

б) Выписать последовательные состояния этого регистра.

в) Существует ли примитивный многочлен с такими свойствами? Почему?

2. а) Многочлен $1+x+x^4$ примитивен над $\mathbf{Z}_2$. Построить отвечающий ему регистр сдвига, который порождает последовательность с периодом 15.

б) Представить элементы $GF(2^4)$ внутренними состояниями регистра сдвига 0000, 1000, 0100 и т. д.

3. а) Построить матрицу M_1 над $\mathbf{Z}_2$ со свойством $\gamma^i M_1 = \gamma^{i+1}$, $i \bmod 15$, где γ^i из упр. 2,б рассматриваются как векторы.

б) Построить схему из элементов ИСКЛЮЧАЮЩЕЕ ИЛИ, реализующую эту матрицу.

4. а) Построить матрицу M_2, для которой $\gamma^i M_2 = \gamma^{2i}$ для $i \bmod 15$.

б) Построить матрицу M_3, для которой $(\gamma^i M_3)^2 = \gamma^i$ для $i \bmod 15$, т. е, матрицу, которая извлекает квадратный корень в поле $GF(16)$.

13.9. АВТОКОРРЕЛЯЦИОННАЯ ФУНКЦИЯ

В системах связи и радиолокационных системах применяется специальная методика анализа сигналов, в которой важную роль играет понятие *автокорреляционной функции*.

Определение. Пусть $f(t)$ — вещественная периодическая функция (сигнал) с периодом T. Ее автокорреляционной функцией называется

$$A(\tau) = \frac{1}{T}\int_0^T f(t) f(t+\tau)\, dt. \qquad (23)$$

На рис. 13.5,*а* показан график серии импульсов с периодом T. На рис. 13.5,*б* показан график автокорреляционной функции.

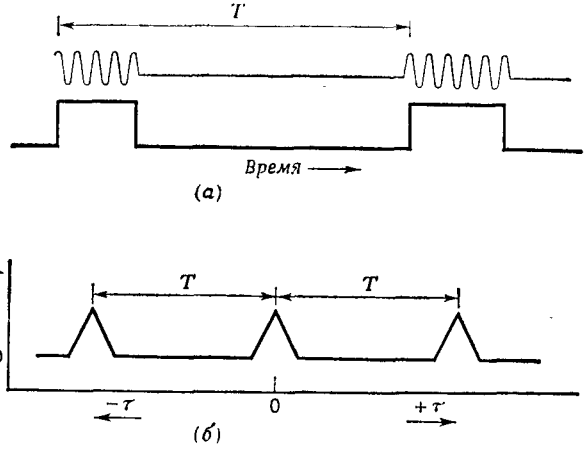


Рис. 13.5. Автокорреляционная функция серии импульсов.

В этой главе обсуждались только функции, принимающие конечное множество значений. Время считалось дискретным. Заменив $f(t)$ на $s(i)$, $i = 0, 1, \ldots, m \ldots$, мы приходим к следующему

дискретному варианту автокорреляционной функции:

$$A(j)=\frac{1}{r}\sum_{i=0}^{r-1} s(i)\cdot s(i+j).$$

Здесь $s(i)$ — элементы рекуррентной последовательности, а r — ее период.

Простой пример рекуррентной периодической последовательности с периодом r:

$$s(j)=0 \quad \text{при} \quad j\neq mr,\ s(mr)=1,\ m=0,\ 1,\ 2,\ \ldots,\ k.$$

Она изображает периодическую последовательность импульсов, как на рис. 13.5. Автокорреляционная функция равна $1/r$ при $j=0,\ r,\ 2r,\ \ldots$ и нулю при остальных j.

Радиолокация посредством периодической последовательности импульсов позволяет однозначно определять расстояния, только если отраженный импульс возвращается раньше, чем испускается следующий импульс. Это означает, что период последовательности должен быть достаточно велик, чтобы позволить измерение больших расстояний.

Рассмотрим радиолокационную систему, которая посылает периодическую последовательность импульсов и получает (после усиления) отраженный сигнал, являющийся точной копией (по форме) испущенного. Пусть $\mathbf{s}$ — посланный сигнал, $\mathbf{s}^*$ — отраженный. Тогда $s(j)=s^*(j-d)$, где d — величина задержки.

Автокорреляционная функция для $\mathbf{s}$ или $\mathbf{s}^*$ принимает значения $A(j)=1/r$ для $j=d,\ 2d,\ 3d,\ \ldots$, что и определяет расстояние до объекта.

В действительности принимаемый сигнал не только ослаблен, но и искажен помехами. Поэтому приемное устройство должно позволять «угадывать» значения задержки. Принятая последовательность $\mathbf{s}^*$ будет отличаться от посланной $\mathbf{s}$. Назовем *последовательностью ошибок*

$$e(j)=s^*(j)-s(j).$$

Если $s(j)$ есть последовательность периодических импульсов, а $e(j)$ содержит много единиц, т. е. ошибок, то придется рассмотреть много периодов для правильного определения задержки, чтобы исключить влияние случайного шума.

Естественно попытаться отыскать класс последовательностей, которые позволяют более эффективно исключать помехи. Они должны иметь большой период и хорошую автокорреляционную функцию. В следующем параграфе мы покажем, как использовать рекуррентные последовательности максимального периода для этой цели.

*13.10. ТЕОРЕМА ОБ АВТОКОРРЕЛЯЦИИ

Покажем, что автокорреляционная функция периодической последовательности максимального периода имеет тот же вид, что и функция последовательности «один импульс за период». Энергию отраженного сигнала можно суммировать по периоду, что увеличивает способность системы обнаруживать объект и следить за ним.

Напомним, что последовательности максимального периода порождаются разностными кодами, характеристические многочлены которых *примитивны* степени n и периода p^n-1 над $\mathbf{Z}_p$. (Дальнейшие подробности см. в статье [5].)

Ограничимся двоичными последовательностями максимального периода, т. е. полем $\mathbf{Z}_2$. Пусть $f(x)$ — многочлен степени n, делящий $x^{q-1}-1$, $q=2^n$, но не x^t-1 для $t<2^n-1$. Имеется 2^n-1 возможных ненулевых начальных блоков длины n и 2^n-1 соответствующих им многочленов степени $\leqslant n-1$. Так как $\operatorname{ord} f = p(s) = 2^n-1$, любая ненулевая последовательность $\mathbf{s}^* \in S(f(x))$ является левым сдвигом любой другой такой последовательности $\mathbf{s}$. Сложив две последовательности $\mathbf{s}$ и $\mathbf{s}^*$, мы снова получим последовательность такого типа либо $\mathbf{0}$ при $\mathbf{s}=\mathbf{s}^*$.

Каждая последовательность $\mathbf{s} \in S(f(x))$ содержит в одном периоде ровно $2^{n-1}-1$ нулей и 2^{n-1} единиц.

Отсюда вытекает следующая *теорема об автокорреляции*.

Теорема 10. *Пусть $S(f(x))$ — множество последовательностей максимального периода, $\deg(f(x))=n$. Определим дискретную автокорреляционную функцию формулой*

$$A(r)=\sum_{i=0}^{2^n-1} B(s_i+s_{i+r}), \text{ где } B(0)=+1,\ B(1)=-1. \qquad (24)$$

Здесь $+1$ и -1 рассматриваются как вещественные числа. Тогда

$$A(r)=\begin{cases} 2^{n-1}, & \text{если } (2^n-1)\mid r,\\ -1 & \text{в противном случае.}\end{cases}$$

Заметим, что B определяет морфизм из аддитивной группы $[\mathbf{Z}_2, +]$, в мультипликативную группу $\{-1, +1\}$.

Прибавив к $\{s_i\}_{i=0}^{2^n-1}$ сдвиг $\{s_{i+r}\}_{i=0}^{2^n-1}$, мы получим последовательность, которая снова является одной из последовательностей такого типа (либо нулевой). Если она ненулевая, то содержит $2^{n-1}-1$ нулей и 2^{n-1} единиц, и сумма значений B будет -1. Если она нулевая, сумма значений B будет 2^n-1.

На рис. 13.6 изображена автокорреляционная функция для последовательности максимального периода.

Многочлен x^3+x+1 из примера 6 порождает последовательность 1010011101001110... и ее семь сдвигов. Любая последовательная семерка символов в ней содержит четыре единицы и три нуля. Сумма значений B по ним равна -1. Функция $A(r)$ будет равна -1, кроме значений $r=0, 7, 14\ldots$, где она равна 7.

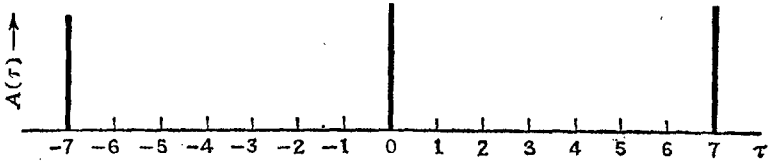


Рис. 13.6. Автокорреляционная функция последовательности максимального периода.

Примитивный многочлен степени 20 над $\mathbf{Z}_2$, например $x^{20}+x^3+1$, порождает последовательность периода $2^{20}-1\simeq 1000000$. Автокорреляционная функция равна -1 для всех сдвигов, кроме сдвигов на кратность периода, где ее значения равны $2^{20}-1$. Этого примерно достаточно для локации Луны.

Многочлен степени 30 дает период $2^{30}-1\simeq 10^9$. Период такого порядка достаточен для локации Венеры. Системы связи со спутниками используют многочлены степени 50. При передаче 10^6 импульсов в секунду период будет порядка года. Заметим, что прямой анализ этих последовательностей для определения их характеристик совершенно немыслим; ими можно пользоваться лишь благодаря хорошей теории используемых многочленов.

*13.11. ФОРМАЛЬНЫЕ РЯДЫ ЛОРАНА

В кольце $F[[x]]$ формальных рядов обратимы только ряды с $p_0\neq 0$; $F[[x]]$ не является полем. Расширим $F[[x]]$ до *поля* $F((x))$ *формальных рядов Лорана* над F. Его элементами являются выражения вида

$$p((x))=\sum_{k=-\infty}^{\infty} p_k x^k,\quad p_k\in F, \tag{28}$$

в которых все p_k *с* $k<0$, *кроме конечного числа, равны нулю.* Если все $p_k=0$, ряд равен нулю. Иначе *наименьшее* k с $p_k\neq 0$ называется *порядком* ряда $p((x))$.

Сложение и умножение в $F((x))$ определяется формулами

$$\sum_{k=-\infty}^{\infty} f_k x^k+\sum_{l=-\infty}^{\infty} g_l x^l=\sum_{j=-\infty}^{\infty}(f_j+g_j)x^j, \tag{29}$$

$$\sum_{k=-\infty}^{\infty} f_k x^k\cdot\sum_{l=-\infty}^{\infty} g_l x^l=\sum_{j=-\infty}^{\infty} h_j x^j, \tag{30}$$

где $h_j=\sum f_i g_{j-i}$, сумма по всем i с $f_i g_{j-i}\neq 0$.

Теорема 11. *Множество всех формальных рядов Лорана над полем F является полем.*

Доказательство. Все свойства проверяются немедленно, кроме, возможно, существования обратных. Но если ряд $\sum p_k x^k$ имеет порядок s, то ряд $x^{-s}\sum p_k x^k$ обратим, и $x^{-s}(x^s(\sum p_k x^k)^{-1})$ будет обратным к $\sum p_k x^k$.

УПРАЖНЕНИЯ Г

1. Начертить график автокорреляционной функции сигнала, который равен 0 в течение 9 секунд, 10 — 1 секунду и т. д., с периодом десять.

2. Начертить график автокорреляционной функции, отвечающий многочлену $1+x+x^4$.

3. То же для многочлена $1+x+x^2+x^3+x^4$. Зависит ли график от начального блока последовательности? Почему?

4. Нули следующих многочленов имеют мультипликативные порядки 3, 4 и 5 соответственно:

$$x^2+x+1,\quad x^3+x^2+x+1,\quad x^4+x^3+x^2+x+1.$$

Пользуясь ими, построить регистр сдвига, дающий хотя бы одну последовательность периода 30. Указать многочлен и разностное уравнение.

5. Доказать, что извлечение квадратного корня в конечном поле характеристики 2 является линейной операцией над $\mathbf{Z}_2$.

6. Многочлен $1+x^3+x^5$ примитивен над $\mathbf{Z}_2$. Представить элементы поля $GF(2^5)$ внутренними состояниями соответствующего регистра сдвига. Пользуясь этим представлением, построить схему из элементов ИСКЛЮЧАЮЩЕЕ ИЛИ, извлекающую квадратные корни.

7. а) Многочлен $1+x^2+x^5$ неприводим над $\mathbf{Z}_2$. Порождает ли он последовательность максимального периода?

б) Построить регистр сдвига для вычисления $(1+x^2+x^5)^{-1}$ в $\mathbf{Z}_2[[x]]$. Каково его начальное состояние?

8. Показать, что многочлен $q(x)=x^4+x^3+x^2+x+1$ неприводим над $GF(2)=\mathbf{Z}_2$, но распадается на линейные множители над $GF(2^4)$ и корни его не примитивны.

9. а) Построить регистр сдвига для $q(x)$.

б) Каково начальное состояние этого регистра, если он порождает последовательность коэффициентов $(1+x)/q(x)$?

в) Каковы периоды последовательностей, порождаемых этим регистром?

г) Каково наименьшее r с $q(x)\mid(x^r+1)$? Докажите. [*Указание*: воспользоваться п.в.]

СПИСОК ЛИТЕРАТУРЫ

1. Gill A., Linear Sequential Circuits, McGraw-Hill, 1966.
2. Golomb S. W., Shift Register Sequences, Holden-Day, 1967.
3. Golomb S. W. (ed.), Digital Communications, Prentice-Hall, 1964.
4. Hall M., Jr., Combinatorial Theory, Blaisdell, 1967. (Русский перевод: Холл М., Комбинаторика, М., «Мир», 1970.)
5. Huffman D. A., The Synthesis of Linear Sequential Coding Networks, Proc. Third Symposium on Information Theory, pp. 77—95, Academic Press, New York, 1956.
6. Mann H. (ed.), Combinatorial Structures in Planetary Reconnaissance, Wiley, 1968.
7. Zierler, N. Linear Recurring Sequences, *J. Soc. Industrial Applied Math.*, 7, № 1, pp. 31—48, March 1959.

14

ВЫЧИСЛИМОСТЬ

14.1. АРИФМЕТИКА КАРДИНАЛЬНЫХ ЧИСЕЛ

В § 2.6 мы упоминали о возможности построить арифметику конечных и бесконечных кардинальных чисел, исходя из основных свойств множеств и функций, установленных в гл. 1 и 2. Напомним, что в классе всех множеств Γ имеется отношение E:

SET означает, что между S и T существует биекция b: $S \leftrightarrow T$. (1)

Это отношение, очевидно, рефлексивно (1_S есть биекция $S \leftrightarrow S$), симметрично (b — обратная к b: $S \leftrightarrow T$ биекция) и транзитивно (если b — биекция $S \leftrightarrow T$, c — биекция $T \leftrightarrow U$, то $c \circ b$ — биекция S и U). Поэтому E — отношение эквивалентности на Γ.

Следовательно, мы можем образовать класс Γ/E, элементами которого являются классы эквивалентности Γ относительно E[1]). Элементы Γ/E, по определению, называются *кардинальными числами*. Кроме того, E удовлетворяет свойствам подстановки (гл. 2) относительно различных бинарных операций над множествами.

Теорема 1. *Теоретико-множественные операции разъединенной суммы, декартова произведения и множества-степени обладают свойством подстановки относительно отношения эквивалентности* (1).

Доказательство. Пусть SES^* и TET^*, т. е. существуют биекции b: $S \leftrightarrow S^*$ и c: $T \leftrightarrow T^*$. Тогда отображения, определенные в 1.6:

$$b \sqcup c\colon\ S \sqcup T \leftrightarrow S^* \sqcup T^*, \tag{2}$$

$$b \times c\colon\ S \times T \leftrightarrow S^* \times T^*, \tag{3}$$

$$c^b\colon\quad T^S \leftrightarrow T^{*S^*}, \tag{4}$$

являются биекциями в силу их конструкции.

---

[1]) Предлагаемый авторами способ конструкции кардинальных чисел наивен и не поддается прямой формализации ни в одном из двух стандартных языков теории множеств — Цермело — Френкеля и Гёделя — Бернайса. Вообще все обсуждение бесконечных множеств в этой главе страдает от невыявленности основных принципов. По этому поводу см. Френкель А., Бар-Хилел И., Основания теории множеств, М., «Мир», 1966. — *Прим. перев.*

Эти свойства подстановки позволяют определить бинарные операции сложения, умножения и возведения в степень кардинальных чисел и вывести из общих свойств множеств и функций стандартные арифметические тождества:

$$x+y=y+x, \qquad xy=yx \qquad \text{(коммутативность)}, \tag{5}$$

$$x+(y+z)=(x+y)+z, \quad x(yz)=(xy)z \qquad \text{(ассоциативность)}, \tag{6}$$

$$x(y+z)=xy+xz, \quad (x+y)z=xz+yz \qquad \text{(дистрибутивность)}, \tag{7}$$

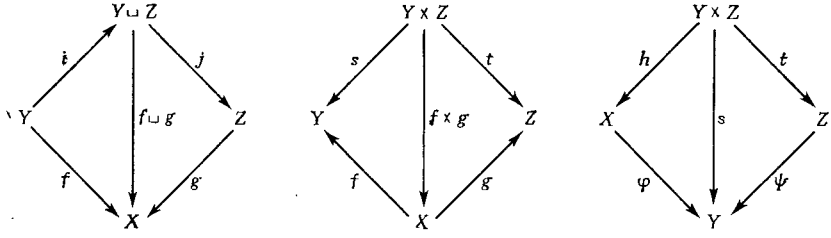
и законы экспоненцирования:

$$x^{y+z}=x^y x^z, \qquad (x^y)^z=x^{yz}, \qquad (xy)^z=x^z y^z, \tag{8}$$

не пользуясь кропотливыми индуктивными проверками, наброски которых были даны в §1.8 и 1.9.

Теорема 2. *Теоретико-множественные операции* $X \sqcup Y$, $X \times Y$ *и* Y^X *определяют однозначные бинарные операции* $x+y$, xy *и* y^x *на классе* Γ/E *всех кардинальных чисел. Эти операции удовлетворяют всем тождествам* (5)—(8).

Набросок доказательства. Первое утверждение следует из теоремы 7 гл. 2 и обсуждения, предшествующего формуле (23) этой главы. Второе утверждение следует из существования ряда очевидных биекций, большая часть которых была указана в гл. 1 (см. упр. A1—A3 ниже).



(а) $x^{y+z}=x^y x^z$ (б) $(yz)^x=y^x z^x$ (в) $x^{(yz)}=(x^y)^z$

Рис. 14.1.

Нетрудно также установить существование *коммутативных диаграмм* типа рис. 14.1, связывающих между собой некоторые из упомянутых биекций.

Наиболее интересно, что эти результаты применимы не только к конечным, но и к *бесконечным множествам*. Это позволяет развить арифметику бесконечных кардинальных чисел. Некоторые результаты о них приведены в § 14.2 и 14.3.

14.2. СЧЕТНАЯ БЕСКОНЕЧНОСТЬ

Назовем *мощностью* множества S кардинальное число S, т. е. класс эквивалентности S относительно E.

Два важнейших бесконечных кардинальных числа — это $\mathbf{d}$ (мощность $\mathbf{P}$) и $\mathfrak{c}$ (мощность $\mathbf{R}$). Первое называется *счетной бесконечностью*, второе — *континуумом*. Большая часть стандартных бесконечных множеств биективна либо $\mathbf{P}$, либо $\mathbf{R}$. Так, $\mathbf{N}$, $\mathbf{Z}$ и $\mathbf{Q}$ биективны $\mathbf{P}$, тогда как $\mathbf{C}$ и $\mathbf{R}^n$ биективны $\mathbf{R}$.

Например, отображение β: $n \mapsto n+1$ и обратное к нему β^{-1}: $m \mapsto m-1$ определяют биекцию $\mathbf{N}$ и $\mathbf{P}$. Расположив элементы $\mathbf{Z}$ в последовательность $0, 1, -1, 2, -2, \ldots$ мы получаем очевидную биекцию ν: $n \mapsto 2n + (1+\operatorname{sgn} n)/2$ из $\mathbf{Z}$ в $\mathbf{P}$ (считаем $\operatorname{sgn} 0 = 1$).

Несколько труднее установить биекцию $\mathbf{Q}$ и $\mathbf{P}$.

Имея в виду представление рациональных чисел в виде отношений целых чисел, докажем здесь несколько более простой результат.

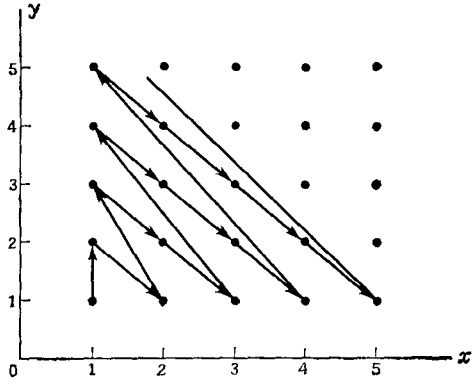


Рис. 14.2.

Лемма 1. *Существует биекция* α: $\mathbf{N} \leftrightarrow \mathbf{N}^2$.

Напомним, что $\mathbf{N}^2 = \mathbf{N} \times \mathbf{N}$ есть множество всех упорядоченных пар (m, n) с $m, n \in \mathbf{N}$.

Доказательство. Расположим пары из $\mathbf{N}^2$ в последовательность, как показано на рис. 14.2. Это расположение отвечает отображению α множества пар (i, j) на множество натуральных чисел:

$$\alpha(i, j) = \left[\frac{(i+j)(i+j+1)}{2}\right] + j.$$

Покажем теперь, что бесконечное множество не может иметь меньшую мощность, чем $\mathbf{P}$, т. е. что $\mathbf{d}$ — самое маленькое бесконечное кардинальное число. Чтобы указать на это свойство, Кантор, создавший теорию бесконечных кардинальных чисел, обозначил его $\mathbf{d} = \aleph_0$ («алеф нуль»).

Теорема 3. *Для любого бесконечного множества U существует инъекция* f: $\mathbf{P} \to U$.

Хотя этот результат представляется очевидным, формальное доказательство его требует применения принципа, вызвавшего много споров — так называемой аксиомы выбора.

Аксиома выбора. *Для любого множества U существует функция φ: $\mathcal{P}(U) \to U$, ставящая в соответствие каждому непустому подмножеству $S \subset U$ некоторый элемент из этого подмножества*: $s = \varphi(S) \in S$.

Применяя аксиому выбора, мы можем доказать теорему 3 так: положим $f(1) = \varphi(U)$, $S_1 = U - \{f(1)\}$ и затем по рекурсии

$$f(n+1) = \varphi(S_n'), \quad S_{n+1} = S_n U \{f(n+1)\}.$$

Рекурсия показывает, что $f(n) \neq f(k)$ при $k < n$, так что f является инъекцией $\mathbf{P}$ в U.

Следствие 1. *Если множество U бесконечно, то существует инъекция g: $U \to U$, не являющаяся сюръекцией, и сюръекция j: $U \to\!\!\!\to U$, не являющаяся инъекцией.*

Следствие 2. *Чтобы множество S было конечным, необходимо и достаточно выполнение любого из двух условий*: (1) *любое инъективное отображение f: $S \to S$ сюръективно*; (2) *любое сюръективное отображение g: $S \to S$ инъективно.*

Из существования биекций $\mathbf{P} \leftrightarrow \mathbf{N}$, $\mathbf{P} \leftrightarrow \mathbf{Z}$ и $\mathbf{P} \leftrightarrow \mathbf{Q}$ вытекают следующие формулы трансфинитной арифметики.

Теорема 4. *Пусть $n \in \mathbf{P}$ — положительное целое число. Тогда*

$$n + \mathbf{d} = \mathbf{d} + \mathbf{d} = \mathbf{d}, \quad n\mathbf{d} = \mathbf{d}\mathbf{d} = \mathbf{d}^n = \mathbf{d}. \tag{9}$$

Доказательство. Биекция множества $\{1, \ldots, n\} \sqcup \mathbf{P}$ с $\mathbf{P}$ определяется формулами: $k \mapsto k$ $(k \leqslant n)$ и $l \mapsto l + n$ при $l > n$.

Биекцию множества $\mathbf{P} \sqcup \mathbf{P}$ с $\mathbf{P}$ можно построить, отобразив первый экземпляр $\mathbf{P}$ на нечетные положительные числа $2k - 1$, а второй экземпляр $\mathbf{P}$ на четные числа $2k$. Отсюда следует, что $\mathbf{d} + \mathbf{d} = \mathbf{d}$.

Биекция $\mathbf{n} \times \mathbf{N} \cong \mathbf{n} \times \mathbf{P}$ с $\mathbf{P}$ устанавливается формулой $(k, l) \mapsto nl + k$ $(k \in \mathbf{n}, l \in \mathbf{N})$. Поэтому $n\mathbf{d} = \mathbf{d}$. Биекция $\mathbf{P} \times \mathbf{P} \leftrightarrow \mathbf{P}$ установлена в лемме 1. Поэтому $\mathbf{d}^2 = \mathbf{d}$, и простая индукция по n показывает, что $\mathbf{d}^n = \mathbf{d}$ при всех $n \geqslant 1$.

14.3. МОЩНОСТЬ КОНТИНУУМА

Рассмотрим теперь множество $\mathbf{R}$ всех вещественных чисел. Функция $e^x/(1+e^x)$ определяет биекцию R с открытым интервалом $(0, 1) = \{y \mid 0 < y < 1\}$. Поэтому справедлива

Лемма 1. *Интервал $0 < x < 1$ имеет мощность континуума.*

С другой стороны, отображение посредством ряда

$$\alpha(\mathbf{a}) = \sum_{k=1}^{\infty} a_k 2^{-k}, \qquad \mathbf{a} = a_1 a_2 a_3 \ldots, \tag{10}$$

определяет *сюръекцию* множества $2^{\mathbf{P}}$ всех функций $\mathbf{P} \to \{0, 1\}$ на замкнутый интервал $[0, 1] = \{y \mid 0 \leqslant y \leqslant 1\}$. Очевидно, α обратно справа относительно правой композиции к *инъекции* $\tilde{\alpha}$: $[0, 1] \to 2^{\mathbf{P}}$, которая определяется неравенствами

$$\sum_{k=1}^{n} a_k 2^{-k} \leqslant x < 2^{-n} + \sum_{k=1}^{n} a_k 2^{-k}. \tag{10'}$$

Отображения α и $\tilde{\alpha}$ «почти» биективны. За исключением случая, когда x — рациональное число со знаменателем степень двойки, $\tilde{\alpha}(x)$ есть единственная функция $\mathbf{a} \in 2^{\mathbf{P}}$ с $\alpha(\mathbf{a}) = x$. У рационального числа со знаменателем степень двойки есть два двоичных разложения:

$$\sum_{j=1}^{m} e_j 2^{-j} + 2^{-m-1} = \alpha(e_1 e_2 \ldots e_m 10 \ldots 0 \ldots) = e_1 e_2 \ldots e_m 011 \ldots 1 \ldots .$$

Так как множество таких рациональных чисел счетно и $2\mathbf{d} = \mathbf{d}$, нетрудно превратить α в *биекцию* β: $2^{\mathbf{P}} \leftrightarrow [0, 1]$, положив

$$\begin{aligned} \beta(e_1 e_2 \ldots e_m 10000) &= \sum_{j=0}^{m} e_j 2^{j+1} + 2^{-m-1}, \\ \beta(e_1 e_2 \ldots e_m 0111 \ldots) &= \sum_{j=0}^{m} e_j 2^{j+1} + 2^{-m-1} + \frac{1}{2} \end{aligned} \tag{11}$$

и доопределив $\beta(\mathbf{a}) = \alpha(\mathbf{a})$ для остальных последовательностей $\mathbf{a} \in 2^{\mathbf{P}}$.

Композиция этой биекции β: $2^{\mathbf{P}} \leftrightarrow [0, 1]$ с любой биекцией $[0, 1] \leftrightarrow \mathbf{R}$ доставляет биекцию $2^{\mathbf{P}} \leftrightarrow \mathbf{R}$. Аналогичную конструкцию можно провести, пользуясь n-ичными разложениями вместо двоичных для любого $n > 1$.

В итоге получаем следующий результат.

Теорема 4. *Для любого целого числа $n > 1$ мощность $\mathbf{n}^{\mathbf{P}}$ равна мощности континуума.*

В частности,

Следствие. *Множество частей $\mathcal{P}(\mathbf{P})$ имеет мощность континуума.*

Теорема 5 (Кантор). *Бесконечное множество $\mathcal{P}(\mathbf{P}) = 2^{\mathbf{P}}$ несчетно.*

Доказательство. По определению $2^{\mathbf{P}}$ есть множество бесконечных двоичных последовательностей

$$\mathbf{x} = x_1 x_2 x_3 \ldots, \quad x_i = 0, 1.$$

Предположим, что существует биекция β: $\mathbf{P} \to 2^{\mathbf{P}}$, т. е. две эти последовательности можно расположить в последовательность

$$\mathbf{x}^1 = x_1^1 x_2^1 x_3^1 \ldots$$
$$\mathbf{x}^2 = x_1^2 x_2^2 x_3^2 \ldots$$
$$\mathbf{x}^3 = x_1^3 x_2^3 x_3^3 \ldots$$
$$\ldots\ldots\ldots$$

Определим двоичную последовательность $\mathbf{y} = y_1 y_2 y_3 \ldots$ условием $y_i = 1 - x_i^i$. Иными словами, y_i есть дополнительная последовательность к *диагональной*. Тогда $\mathbf{y} \in 2^{\mathbf{P}}$, но $\mathbf{y} \neq \mathbf{x}^i$ для всех $i = 1, 2, 3, \ldots$, ибо $y_i \neq x_i^i$. Это доказывает, что не существует даже сюръекции f: $\mathbf{P} \to 2^{\mathbf{P}}$ и тем более биекции.

Этот способ доказательства называется диагональным процессом Кантора.

Из теорем 4 и 5 вытекает

Следствие. *Континуум* $\mathbf{R}$ *несчетен.*

Диагональный процесс Кантора применим также к произвольным множествам и доставляет следующий результат:

Теорема 6 (Кантор). *Пусть* X *— любое множество. Тогда не существует сюръекции* f: $X \twoheadrightarrow 2^X$.

Доказательство. Рассмотрим любую функцию f: $X \to 2^X$. Для каждого $a \in X$ обозначим через f_a: $X \to 2$ образ a относительно f. Определим дополнительную к диагональной функцию c: $X \to 2$: $c(x) = 1 - f_x(x)$. Тогда c не может совпадать ни с одной из функций f_a, ибо $c(a) = 1 - f_a(a) \neq f_a(a)$. Значит, c: $X \to 2$ отлична от всех f_a, так что семейство $\{f_a\}$ не может исчерпывать множество 2^X всех функций g: $X \to 2$.

Следствие. *Пусть* X *— любое множество. Не существует биекции* b: $X \leftrightarrow 2^X$.

Выразимость и вычислимость. Исследуя формализации математических доказательств, специалисты в математической логике обнаружили много трудностей и двусмысленностей. Несчетность $\mathbf{R}$ служит одной из фундаментальных причин неприятностей. Проиллюстрируем это следующим рассуждением.

Парадокс Ришара. Пусть A — любой конечный алфавит из ν символов (например, реализуемый на пишущей машинке или в наборном устройстве; возможно, включающий пробелы, знаки препинания и т. д.). Имеется лишь n^ν строк длины n в этом алфавите, а всего счетное множество строк конечной длины. Если в этом алфавите задан язык, часть этих строк может быть осмысленной. Оставляя только их, мы получаем, что *в любом печатном языке множество всех строк конечной длины не более чем счетно.*

В частности, множество всех определений конкретных вещественных чисел счетно. Между тем множество всех вещественных чисел несчетно. Таким образом, *лишь ничтожную долю всех вещественных чисел можно явно определить*. В этом смысле континуум не поддается полному языковому описанию (парадокс Ришара).

Понятие «определимости» или «выразимости» зависит от языка. Мы увидим далее, что «вычислимость» имеет более абсолютный смысл. Введем индуктивное понятие вычислимости для $\mathbf{R}$ и для $2^{\mathbf{P}} \cong \mathcal{P}(\mathbf{P})$.

Определение. Вещественное число x называется *вычислимым*, если существует *алгоритм*, который позволяет для каждого $n \in \mathbf{P}$ вычислить за конечное число шагов двоичную дробь $a = k/2^r$ ($k \in \mathbf{Z}$, $r \in \mathbf{N}$), такую, что $|x - a| < 2^{-n}$. Подмножество $S \subset \mathbf{P}$ называется *вычислимым*, если существует алгоритм, который для каждого $n \in \mathbf{P}$ позволяет решить, верно ли, что $n \in S$.

Множество всех вычислимых вещественных чисел и всех вычислимых подмножеств $S \subset \mathbf{P}$, разумеется, также счетно. Так как $\mathbf{R}$ и $\mathcal{P}(\mathbf{P})$ несчетны, существуют невычислимые вещественные числа и подмножества $S \subset \mathbf{P}$.

УПРАЖНЕНИЯ А

1. Подробно доказать теорему 1 и тождества, следующие из нее.

2. Доказать тождества (5)—(7), построив шесть биекций, в том числе $A \times (B \times C) \leftrightarrow (A \times B) \times C$ и $(A \sqcup B) \times C \leftrightarrow (A \times C) \sqcup (B \times C)$.

3. Доказать тождества (8), построив три биекции, в том числе $C^{A \sqcup B} \leftrightarrow \leftrightarrow C^A \times C^B$.

4. Построить биекцию $\mathbf{N}$ и множества всех конечных множеств целых чисел.

*5. Алгебраическим числом называется всякое число $z \in \mathbf{C}$, удовлетворяющее некоторому уравнению вида $z^n + \sum_{k=0}^{n-1} c_k z^k = 0$ для $n \in \mathbf{P}$, $c_0, \ldots, c_n \in \mathbf{Q}$. Доказать, что множество алгебраических чисел счетно.

Пусть x, y — кардинальные числа. Отношение $x \leqslant y$ означает, что существуют множества X, Y с $\operatorname{card} X = x$, $\operatorname{card} Y = y$ и инъекция $X \longrightarrow Y$.

6. Показать, что если $x \leqslant y$ и A, B — любые множества с $\operatorname{card} A = x$, $\operatorname{card} B = y$, то существует инъекция $A \longrightarrow B$.

7. Показать, что если $x \leqslant y$, то для любого кардинального числа z имеем $x + z \leqslant y + z$, $xz \leqslant yz$, $x^z \leqslant y^z$ и $z^x \leqslant z^y$.

8. а) Доказать, что $\mathbf{c} + n = \mathbf{c} + \mathbf{d} = \mathbf{c}$ для всех $n \in \mathbf{N}$.
   б) Доказать, что $n\mathbf{c} = \mathbf{dc} = \mathbf{c}$ для всех $n \in \mathbf{P}$.

9. а) Доказать, что $\mathbf{c}^n = \mathbf{c}^{\mathbf{d}} = \mathbf{c}$ для всех $n \in \mathbf{P}$.
   б) Доказать, что $n^{\mathbf{c}} = 2^{\mathbf{c}} = \mathbf{d}^{\mathbf{c}} = \mathbf{c}^{\mathbf{c}}$ при $n > 1$.

*10. Построить биекцию $10^{\mathbf{P}} \leftrightarrow \mathbf{R}$. (*Предупреждение*: не **забывать о неоднозначности** десятичных разложений типа $0.9999\ldots = 1.0000\ldots$.)

*11. Пусть функция f: $\mathbf{R} \longrightarrow \mathbf{R}$ принимает только положительные значения. Доказать, что для любого K можно найти такое конечное подмножество F чисел $x_i \in \mathbf{R}$, что $\sum_F f(x_i) > K$.

14.4. ВЫЧИСЛИМОСТЬ ПО ТЬЮРИНГУ

Широко распространено убеждение в том, что понятие вычислимости, сформулированное в § 14.3, не зависит от языка, выбранного для описания алгоритмов[1]). Тезис Чёрча—Тьюринга состоит в утверждении, что каждый алгоритм может быть реализован в виде конечной программы для машины Тьюринга, описанной в § 3.7. Числа и множества, вычислимые в этом смысле, мы будем называть *вычислимыми по Тьюрингу*.

Тьюринг показал, в поддержку этого тезиса, что класс вычислимых вещественных чисел является подполем $\mathbf{R}$, которое содержит все рациональные и алгебраические числа, e, π, нули функций Бесселя и вообще наиболее часто используемые в математическом анализе константы.

Аналогично, вычислимые по Тьюрингу множества $S \subset \mathbf{P}$ обладают тем свойством, что их характеристическая функция может быть получена в виде выпечатанной последовательности нулей и единиц на ленте машины Тьюринга с подходящей программой. При этом $n \in S$ в том и только том случае, если, скажем, на ленте выпечатывается 1 в n-й позиции.

Справедлив следующий результат.

Теорема 7. *Стандартная биекция* β: $\mathcal{P}(\mathbf{P}) \leftrightarrow [0, 1]$, *определенная уравнениями* (10) *и* (11), *переводит вычислимые подмножества* $S \subset P$ *в вычислимые вещественные числа* $x \in [0, 1]$.

Действительно, если S и дополнение $S' = \mathbf{P} - S$ бесконечны, то $\left| x - \sum_{k=1}^{m} e_S(k) 2^{-k} \right| < 2^{-m}$, и приближения к x в виде этих сумм алгоритмически вычислимы.

Для конечных множеств S или множеств с конечным дополнением S' величина $\beta(S)$ является рациональным числом.

Обращение теоремы 7 как будто неверно, хотя не известно конкретное вычислимое число $x \in [0, 1]$ с невычислимым $S \subset \mathbf{P}$.

Источник трудностей можно усмотреть, попытавшись доказать правдоподобную гипотезу, что число $J_0(\pi)$ не является рациональным со знаменателем степень двойки. Как бы много цифр двоичного разложения мы ни просчитали, откуда мы знаем, что, начиная с некоторого места, они будут только нулями или единицами?

---

[1]) Если этот язык достаточно богат.— *Прим. перев.*

Аналогичио, нет общей программы для машины Тьюринга, позволяющей установить *совпадение* двух вычислимых подмножеств $S \subset \mathbf{P}$ и $T \subset \mathbf{P}$. Хотя в случае $S \neq T$ это можно явно проверить за конечное время, при $S = T$ никакое конечное число проверенных уравнений $e_S(n) = e_T(n)$ не поможет установить, что $S = T$.

С этим связан следующий факт.

Теорема 8. *Свойство подмножества* **P** *быть вычислимым по Тьюрингу само не является вычислимым по Тьюрингу.*

Аналогично, множество вещественных чисел, вычислимых по Тьюрингу, само не является вычислимым по Тьюрингу[1]).

14.5. ВЫЧИСЛИМОСТЬ ПО ТЬЮРИНГУ И ПРАКТИЧЕСКАЯ ВЫЧИСЛИМОСТЬ

С практической точки зрения машины Тьюринга способны не только выпечатывать вычислимые последовательности нулей и единиц, но и совершать разнообразные операции над целыми или вычислимыми числами. Они могут также отыскивать определенный объект в списке, упорядочивать объекты и т. п.

Это приводит к фундаментальному вопросу: какие классы вычислений могут быть осуществлены машинами Тьюринга? Точка зрения Тьюринга, ставшая теперь общепринятой, состояла в том, что любая «эффективно предписанная процедура» может быть проведена машиной Тьюринга. Хотя это утверждение недоказуемо, оно подверглось большой экспериментальной проверке и легло в основу изучения «эффективных процедур» логиками, которые часто используют такую их формализацию.

Тьюринг показал также существование «универсальной машины Тьюринга», которая способна выполнить любое вычисление, выполнимое любыми машинами Тьюринга. Входные данные для вычисления на универсальной машине содержат закодированное описание конкретной машины и затем описание входных данных задачи и алгоритма вычисления. Универсальную машину Тьюринга можно смоделировать на универсальной ЭВМ. Поэтому, отвлекаясь от ограничений на объем памяти и время, утверждается, что универсальная ЭВМ способна выполнить любое вычисление, выполнимое любой машиной Тьюринга. Память ЭВМ может считаться потенциально *бесконечной*, поскольку ее можно наращивать в ходе вычислений.

---

[1]) Эти результаты неточно сформулированы и аргументы авторов неадекватны. Поэтому при переводе они опущены. Вообще весь круг вопросов о выразимости и вычислимости изложен слишком поспешно, и мы рекомендуем читателю обратиться к другому источнику, например [8].— *Прим. перев.*

Таким образом, принято считать, что универсальная ЭВМ обладает теми же возможностями, что универсальная машина Тьюринга, возможно с лентой ограниченной длины.

Практическая вычислимость. Вычислимость по Тьюрингу и ее варианты, изученные логиками, является значительно более широким понятием, чем практическая вычислимость с помощью физически реализуемого конечного автомата. Рассмотрим, например, вслед за Дэвидсоном, гипотетическую последовательностную машину, один такт работы которой требует времени, за которое свет проходит расстояние порядка радиуса ядра ($\sim 10^{-23}$ сек). Предположим, что такая машина работает с момента возникновения нашей вселенной (это время оценивается как $4 \cdot 10^{10}$ лет). За это время не пройдет $2^{28} = 2^{256}$ тактов ее работы, так что машина не успеет составить список всех булевых многочленов от семи переменных.

Такого рода принципиальные ограничения на число шагов, требуемых в вычислениях, играют важную роль, особенно, скажем, при решении дифференциальных уравнений в частных производных, где однотипные циклы многократно повторяются. На практике вычисления, требующие $10^{15} \simeq 2^{50}$ действий, уже непомерно дороги.

По сравнению с этими ограничениями на допустимое время работы теоретические ограничения, связанные с конечностью числа состояний, оказываются маловажными. Например, даже у небольшой машины PDP-8 имеется память $4096 = 2^{12}$ слов длины 35, что позволяет реализовать $2^{35 \cdot 4096} = 2^{143260}$ внутренних состояний. (Дополнительные устройства памяти могут иметь объем до 10^8 битов, что доставляет $2^{100\,000\,000}$ внутренних состояний.)

В большинстве случаев именно время, а не объем памяти ограничивает реальные возможности компьютера. Впрочем, за исключением наиболее стандартных задач, еще более существенные ограничения связаны с трудностью написания безошибочной программы из 10^4 или большего числа команд.

14.6. МАТЕМАТИЧЕСКАЯ ЛИНГВИСТИКА

Мы вернемся теперь к изучению грамматики символических языков типа АЛГОЛа, начатому в конце гл. 4. Грамматика АЛГОЛа принадлежит к классу, описываемому следующим определением[1]).

Определение. *Порождающей грамматикой* называется четверка (V, T, P, A), где V — *конечный алфавит*, состоящий из

[1]) Более четкое определение см. Гладкий А. В., Мельчук И. А., Математическая лингвистика, «Наука», М., 1969.— *Прим. перев.*

символов; $T \subset V$ — подмножество *терминальных символов*; $A \in V-T$ — выделенный нетерминальный символ, обозначающий совокупность всех порождаемых объектов; P — конечное множество *правил порождения* вида $u \to v$, где u — непустая строка нетерминальных символов, а v — некоторая строка. Множество $L=L(V,T,P,A)$ всех строк терминальных символов, которые можно получить применением правил порождения, называется *языком*. Оно является подмножеством множества T^* — всех строк символов из T.

В случае АЛГОЛа T включает все буквенные и числовые символы, символы алгебраических операций, выделенные слова, как **for**, и т. д. Множество $V-T$ включает описания типа, как, например, ⟨integer⟩, ⟨number⟩, ⟨Boolean expression⟩, используемые в метаязыке АЛГОЛа. Некоторые из многочисленных правил порождения в АЛГОЛе будут обсуждены чуть позже.

В приложениях к грамматике естественных языков терминальными символами следует считать обычные слова, тогда как символы из $V-T$ суть названия частей речи и других грамматических категорий. Правила порождения могут включать грамматические правила типа $A=$ ⟨предложение⟩ $\to$ ⟨подлежащее⟩⟨сказуемое⟩ и правила подстановки типа ⟨подлежащее⟩ $\to$ ⟨мальчик⟩ и ⟨сказуемое⟩ $\to$ ⟨улыбается⟩. Строки терминальных символов языка являются грамматически правильными *предложениями*.

В языках символической логики некоторые правила порождения $A \to v$ называются аксиомами, а другие — «правилами вывода». «Предложения», которые можно породить с помощью правил вывода и аксиом, называются *теоремами* рассматриваемой дедуктивной системы.

Приведем простые примеры порождающих грамматик с $V-T=\{A\}$, т. е. грамматик с единственной грамматической категорией.

Пример 1. Пусть $T=\{c\}$ и P состоит из двух правил порождения:

$$A \to c, \qquad A \to cA. \tag{12}$$

Эта грамматика порождает язык $L=\{c,cc,ccc,\ldots\}$, т. е. множество T^* всех конечных строк из символов c.

Пример 2. Пусть $T=\{0, 1\}$ и P состоит из следующих правил порождения:

$$A \to \Lambda, \qquad A \to 0A0, \qquad A \to 1A1, \tag{13}$$

где Λ — пустая строка (она подчиняется правилу $\mathbf{x}\Lambda\mathbf{y}=\mathbf{xy}$, т. е. исчезает в любой непустой строке). Эта грамматика порождает пустую строку и две двоичные последовательности вида $\mathbf{x}\mathbf{x}^R$, где $\mathbf{x}=x_1x_2\ \ldots\ x_n$, а $\mathbf{x}^R=x_n\ \ldots\ x_2x_1$ — *обращение* строки $\mathbf{x}$.

Пример 3. Пусть $T=\{b, c\}$ и P состоит из правил порождения

$$A \to bc, \quad A \to bAc. \tag{14}$$

Они порождают язык $\{b^n c^n\}$ $(n \in \mathbf{P})$.

Пример 4. Пусть $T=\{b, c\}$, а P состоит из правил порождения

$$A \to b, \quad A \to bA, \quad A \to cA. \tag{15}$$

Они порождают язык, состоящий из всех строк в T^*, кончающихся на b.

Примеры из АЛГОЛа. Рассмотрим правила порождения (десятичных) целых чисел без знака. Здесь $T=\{0, 1, 2, \ldots, 9\}=$ $=\langle$цифра$\rangle$ и

$$V-T=\{\langle\text{цифра}\rangle, A\}..$$

Множество $\langle$целое без знака$\rangle$ есть просто T^*; оно порождается правилами

$$A \to \langle\text{цифра}\rangle \quad (\text{или } A \to 0, A \to 1, \ldots, A \to 9) \tag{16}$$

и

$$A \to A\langle\text{цифра}\rangle \quad (\text{или } A \to A0, A \to A1, \ldots, A \to A9). \tag{16'}$$

Для порождения целых со знаком положим $T=\{\langle$цифра$\rangle, +, -\}$, где $\langle$цифра$\rangle=\{0, 1, \ldots, 9\}$, как выше, и используем правила порождения

$$A \to +\langle\text{цифра}\rangle, \quad A \to -\langle\text{цифра}\rangle, \quad A \to A\langle\text{цифра}\rangle. \tag{17}$$

Правила порождения в АЛГОЛе и в простых примерах 1—4 *контекстно-свободны* в следующем смысле.

Определение. Правило порождения $\xi \to v$ называется *контекстно-свободным*, если оно применимо к любой строке $a\xi b$ и дает строку avb. Язык, который порождается лишь контекстно-свободными правилами, сам называется *контекстно-свободным*.

Если разрешается применять правило $\xi \to v$ к строкам $a\xi b$ лишь при некоторых a, b, правило называется *зависящим от контекста*. Правила естественных языков сильно зависят от контекста.

Правила АЛГОЛа, описанные в § 4.7 в нормальной форме Бэкуса, контекстно-свободны. АЛГОЛ поэтому является контекстно-свободным языком.

УПРАЖНЕНИЯ Б

1. Показать, что если вещественные числа x, y вычислимы, то $x+y$ вычислимо.

2. То же для xy.

3. Показать, что множество вычислимых вещественных чисел образует поле.

4. Показать, что если функции f: $\mathbf{P} \longrightarrow \mathbf{P}$ и g: $\mathbf{P} \longrightarrow \mathbf{P}$ вычислимы, то $f+g$, fg, g^f и $g \circ f$ вычислимы.

5. Показать, что любой многочлен с коэффициентами из $\mathbf{P}$ является вычислимой функцией.

6. Положим $F(m, 0) = m+1$ и $F(m+1, n+1) = F(F(m, n+1), n)$.
   а) Доказать, что $F(m, 2) = 2m$ и $F(m, 3) = 2^m$.
   б) Доказать, что $F(1, 4) = 2$ и $F(m+1, 4) = 2^{F(m, 4)}$.
   в) Доказать, что $F(1, 5) = 2$ и $F(m+1, 5) = F(m, 5)^{F(m, 5)}$.

7. Доказать, что если миожества S и T в $\mathbf{P}$ вычислимы, то и $S \cup T$ вычислимо.

8. Доказать индукцией результат примера 1.

9. То же для примера 2.

14.7. АВТОМАТНЫЕ ГРАММАТИКИ

Рассмотрим теперь языки, которые «принимаются» или распознаются автоматами с конечным числом состояний специального типа. Такой автомат считывает с входной ленты строки терминальных символов некоторого языка, и после считывания символа $x \in X$ в состоянии s переходит в новое состояние $\nu(x, s)$. Считав последний символ, автомат останавливается. Если он останавливается в одном из «принимающих» состояний из отмеченного множества F, входная строка считается *принятой*. Если конечное состояние лежит в $S-F$, строка считается *отвергнутой*. Вот формальное определение.

Определение. *Акцептором с конечным числом состояний* (конечным акцептором, анализатором) называется четверка $\mathcal{A} = [X, S, \nu, s_0, F]$, где:

(1) X — конечное множество *входных символов*;

(2) S — конечное множество *внутренних состояний*;

(3) ν — функция из $X \times S$ в S;

(4) $s_0 \in S$ — *начальное состояние*;

(5) $F \subset S$ — множество *принимающих конечных состояний*.

Входные символы линейно упорядочены, скажем напечатаны на входной ленте, и последовательно считываются читающей головкой.

Пример 5. На рис. 14.3 показана диаграмма конечного акцептора. Этот акцептор принимает, в частности, входные строки

001, 0001, 100, 1010, 10110

Опишем теперь класс грамматик, порождающих в точности такие множества строк, которые принимаются подходящим конечным акцептором.

Определение. *Автоматной грамматикой* G называется четверка $[X, T, P, A]$, где:

(i) X — конечное множество входных символов,

(ii) $T \subset X$ — конечное множество терминальных символов,

(iii) P — конечное множество правил порождения вида

$$Z \to bY, \quad Z \to b, \quad \text{где} \quad Z, Y \in X - T, \; b \in T,$$

(iv) $A \in X - T$ — отмеченный *начальный символ.*

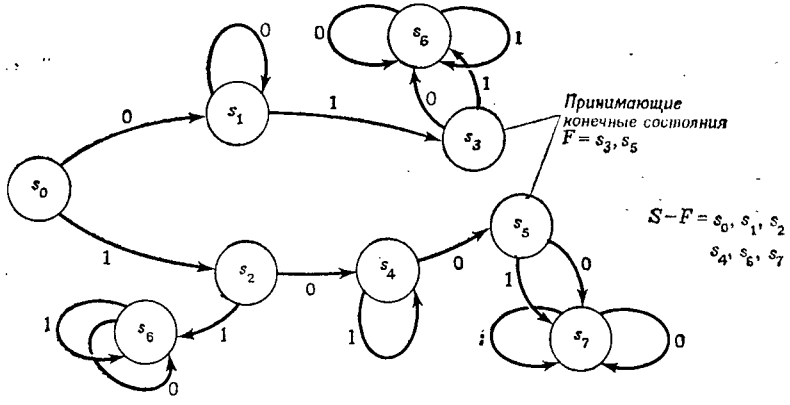


Рис. 14.3. Конечный акцептор.

Небольшие конечные акцепторы удобно описывать диаграммами состояний. Аналогичный прием применим к автоматным грамматикам. Для составления диаграммы следует произвести следующие действия:

Шаг 1. Каждому нетерминальному символу из G поставить в соответствие вершину графа.

Шаг 2. Каждому правилу $X \to bY$ поставить в соответствие дугу от X до Y, помеченную b.

Шаг 3. Каждому правилу $X \to b$ поставить в соответствие дугу от X к новой вершине, помеченной ПРИНЯТЬ.

Пример 6. На рис. 14.4 показана диаграмма состояний (помеченный ориентированный граф), отвечающая автоматной грамматике G с правилами

$$A \to cA \quad A \to bY \quad Y \to cA$$
$$Y \to b \quad Y \to bX \quad X \to c$$

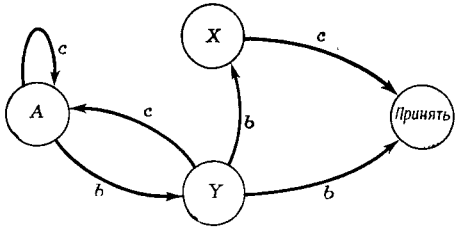


Рис. 14.4. Диаграмма грамматики.

Теорема 9. *Последовательность меток дуг вдоль любого конечного пути, начинающегося из вершины A на диаграмме грамматики G,*

является строкой, порождаемой G, и все порождаемые строки отвечают путям.

Доказательство. Это просто перефразированное описание грамматики.

Теорема 10. *Для любой автоматной грамматики G существует конечный акцептор, принимающий в точности те строки, которые порождает G.*

Доказательство. На диаграмме грамматики переобозначим A через s_0, а метки вершин — метками различных состояний. Получится диаграмма конечного акцептора.

Теорема 11. *Для любого конечного акцептора $\mathcal{A}$ существует автоматная грамматика G, порождающая в точности те строки, которые принимаются акцептором $\mathcal{A}$.*

Доказательство. На диаграмме акцептора следует заменить метки вершин нетерминальными символами и затем по меткам ребер восстановить порождающие правила.

Эти теоремы устанавливают биекцию между автоматными грамматиками и конечными акцепторами. Естественно спросить, существуют ли контекстно-свободные языки, не являющиеся автоматными языками (включение очевидно)?

Теорема 12. *Существуют контекстно-свободные языки, не являющиеся автоматными языками.*

Доказательство. Пример 3 доставляет такой язык. Действительно, пусть p — период функции, преобразующей состояния данного конечного автомата при считывании b. Он конечен. Ясно, что если автомат принимает строку $b^n c^n$, то он примет и строку b^{n+pc^n}.

Заметим, что язык примера 3 содержится в языке примера 4 (с точностью до замены b на c и наоборот), который является автоматным.

Аналогичное рассуждение показывает, что никакой конечный акцептор не может принимать только правильно расположенные последовательности скобок неограниченной длины вроде ((( )( ))( )). (Практически это не слишком серьезное ограничение; см. конец § 14.5.)

Однако с помощью конечных автоматов можно вычислять многочлены любой длины, пользуясь тождествами

$$a_0 + a_1 x + a_2 x^2 + \ldots + a_{n-1} x^{n-1} + a_n x^n = \\ = ((\ldots ((a_n) x + a_{n-1}) x + \ldots) x + a_1) x + a_0.$$

Заслав a_n в сумматор, следует затем ввести программу на подходящем языке ассемблера:

MUL x, ADD a_{n-1}, MUL x, ADD a_{n-2}, ...

..., MUL x, ADD a_0, HALT

Такие вложенные системы предложений реализуются и в естественных языках.

Их анализ использует технику синтаксического разбора, которую мы обсудим ниже.

14.8. СИНТАКСИС; МАГАЗИННЫЕ АКЦЕПТОРЫ

Заметим, что принятие строки конечным акцептором происходит в результате синтаксического разбора этой строки, поскольку состояния акцептора отвечают нетерминальным символам грамматики. Дополнительное описание семантики позволяет основывать на этом анализе трансляцию. Программа, которая сначала осуществляет синтаксический анализ строки и затем на его основе строит «перевод» ее в программу на машинном языке, называется компилятором.

С точки зрения теории автоматов процесс трансляции контекстно-свободных языков легче всего объяснить в терминах магазинной памяти, о которой мы уже говорили в гл. 4. Это обеспечивается специальной структурой памяти вычислительных машин, как бы составленной из слоев потенциально неограниченной глубины. Информация может быть запасена в ячейках и быстро извлечена оттуда (за микросекунды), либо записана на магнитных лентах или барабанах, объем которых может быть очень велик.

Синтаксический разбор может производиться компилятором разными способами. (Предполагается, что этот разбор производится однозначно.) Самый прямой и очевидный способ состоит в построении дерева разбора. Компилятор просматривает ветви дерева, пока не обнаруживает допустимый вариант разбора. Например, выражение $X \times Y + Z + U$ разбирается однозначно, и на рис. 14.5 показано конкретное дерево, соответствующее синтаксическим правилам § 4.7. Напомним, что синтаксические правила АЛГОЛа однозначно определяют порядок применения операций с учетом чтения слева направо.

Автомат с магазинной памятью есть конечный автомат, который имеет дополнительную ленту памяти. Этот автомат считывает по очереди символы с входной ленты, постоянно в одном направлении. Однако он может считывать символы с ленты памяти, впечатывать в нее новые символы и передвигать ленту за каждый такт в любом направлении. Лента памяти называется *магазином*

или *стеком*. Главное ограничение на ее использование — принцип «последним пришел — первым обслужен». Представим себе, что лента расположена вертикально. Тогда лента передвигается вниз при записи и вверх при считывании. При записи символ записывается одной ячейкой выше, чем на предыдущем такте. После

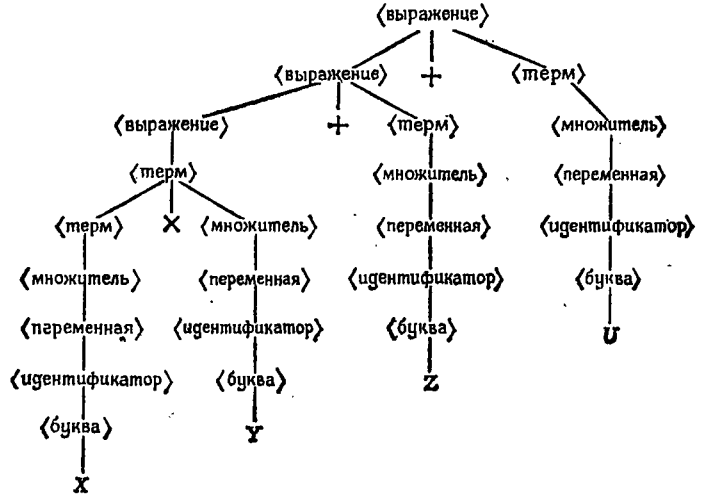


Рис. 14.5. Разбор выражения $X \times Y + Z + U$.

считывания считанный символ стирается. У ленты памяти имеется самая нижняя ячейка, в которой записан специальный символ σ — признак дна магазина. Вверх лента неограничена. По этой причине автомат с магазинной памятью не является конечным автоматом.

Акцептор с магазинной памятью — это автомат с магазинной памятью и следующим условием: если вслед за считыванием последнего символа с входной ленты автомат считывает σ из магазина, то считанная строка называется принятой; иначе — отвергнутой.

Автомат с магазинной памятью является важной абстракцией. Формальное определение таково:

Определение. *Автоматом с магазинной памятью* называется шестерка $(X, S, \Delta, \delta, \sigma, s_0)$, где:

(i) X — конечное множество входных символов,
(ii) S — конечное множество внутренних состояний,
(iii) $\Delta \supset X$ — множество входных символов, дополненное вспомогательными символами,
(iv) δ: $X \times S \times \Delta \to S \times \Delta^*$ — функция,
(v) $\sigma \in \Delta$ — конец магазинной памяти,
(vi) $s_0 \in S$ — начальное состояние.

Рассмотрим грамматику $A \longrightarrow cAb$, $A \longrightarrow cb$ и акцептор с магазинной памятью, для которого

$$X = \{c, b\}$$
$$S = \{s_0, s_1\}$$
$$\Delta = \{c, b, S\}$$

$(c, s_0, \sigma) \mapsto (s_1, c)$ $\quad$ $(b, s_2, c) \mapsto (s_2, \Lambda)$

$(b, s_0, \sigma) \mapsto (s_3, b)$ $\quad$ $(b, s_3, c) \mapsto (s_3, b)$

$(c, s_1, c) \mapsto (s_1, c)$ $\quad$ $(b, s_3, b) \mapsto (s_3, b)$

$(b, s_1, c) \mapsto (s_2, \Lambda)$ $\quad$ $(c, s_3, b) \mapsto (s_3, c)$

$(c, s_2, c) \mapsto (s_3, b)$ $\quad$ $(c, s_3, b) \mapsto (s_3, c)$

Эта грамматика порождает множество строк $\{c^n b^n: n = 1, 2, \ldots\}$, и акцептор с магазинной памятью принимает в точности такие строки. Заметим, что состояние s_3 является «ловушкой»: считав последовательность символов, не принадлежащую языку, акцептор переходит в состояние s_3 и оказывается запертым в нем. Магазин собирает символы c, пока не будет считан первый b, что заставляет автомат стереть верхний c в магазине, и т. д. пока будут считываться b. Если число считанных b совпадает с числом считанных c к концу считывания, то акцептор остановится в состоянии s_2 и в памяти останется лишь σ.

Таким образом, магазинные акцепторы могут различать вложенные друг в друга последовательности, что широко используется в случае системы программ. Однако некоторые контекстно-свободные языки не порождаются автоматами с магазинной памятью. Так, к этому классу относится язык примера 2: строки xx^R не поддаются отбору таким автоматом, ибо автомат не может найти середину строки за один проход. (Однако язык со строками $x d x^R$, где d—специальный маркер, может быть порожден магазинным акцептором.)

Все контекстно-свободные языки можно породить автоматами из следующих классов: (1) магазинные акцепторы с двумя магазинами; (2) машины Тьюринга; (3) недетерминированные магазинные автоматы.

Акцептор с двумя магазинами является непосредственным обобщением ранее введенного понятия. Вместо функции перехода $(c, s_i, b) \mapsto (s_j, d)$ он требует задания функции перехода $(c, s_i, b) \mapsto \mapsto (s_j, d, e)$, где d—символ, относящийся к первой ленте, а e—ко второй. Например, $(c, s_j, b) \mapsto (s_k, \Lambda, d)$ означает команду «стереть верхний символ на первой ленте; записать d на второй ленте».

Заметим, что наличие двух лент памяти открывает такие же возможности для хранения и обработки информации, как если

бы автомат обладал одной лентой, бесконечной в обоих направлениях и способной двигаться в любом направлении.

Поэтому автоматы с двумя магазинами по существу не отличаются от машин Тьюринга (см. гл. 3).

Недетерминированные автоматы составляют последний важный класс. Интуитивно поведение такого автомата не определено однозначно: функция δ может быть многозначна. Автомат принимает строку, если он окажется после считывания этой строки в одном из принимающих конечных состояний, независимо от того, какую из допустимых «линий поведения» он избрал. Известно, что любой контекстно-свободный язык может быть принят подходящим недетерминированным автоматом.

* 14.9. РЕКУРСИВНЫЕ ФУНКЦИИ

В оставшейся части главы мы опишем еще два круга понятий, тесно связанных с выразимостью и вычислимостью. Первое из них отвечает на следующий вопрос: какие функции f: $\mathbf{P} \mapsto \mathbf{P}$ можно конструктивно определить, исходя из описания $\mathbf{P}$ по Пеано как унарной алгебры $[\mathbf{P}, \sigma]$?

Простая рекурсия. Функции σ^m: $r \mapsto m+r$ и p_m: $r \mapsto mr$ были введены в 1.7 определениями следующего вида:

R1. Задано значение $f(1)=f_1 \in \mathbf{P}$.

R2. Задано значение $f(\sigma n)$ как функция $h(f(n))$, где h: $\mathbf{P} \longrightarrow \mathbf{P}$— ранее введенная вычислимая функция.

Так, в определении σ^m мы полагаем $f_1 = \sigma m$ и $h=\sigma$. В определении p_m полагаем $f_1 = m$ и $h=\sigma^m$.

Определение такого вида называется определением посредством *простой рекурсии*, исходя из функции h. Оно в самом деле однозначно определяет f.

Теорема 13. *Пусть функции* f: $\mathbf{P} \longrightarrow \mathbf{P}$ *и* g: $\mathbf{P} \longrightarrow \mathbf{P}$ *удовлетворяют условиям* R1 *и* R2. *Тогда* $f=g$.

Доказательство. Пусть S—множество всех $n \in \mathbf{P}$ с $f(n)= = g(n)$. Тогда $1 \in \mathbf{P}$ в силу R1. Кроме того, в силу R2 из $n \in S$ следует, что

$$f(\sigma n) = h(f(n)) = h(fg(n)) = g(\sigma n).$$

Значит, из $n \in S$ вытекает, что $\sigma n \in S$. По аксиоме 3 индукции Пеано $S=\mathbf{P}$. По определению функции имеем $f=g$.

Эту теорему можно рассматривать как единую аксиому, эквивалентную всем трем аксиомам Пеано из § 1.7.

Аксиома Пеано—Ловера. Для любого числа $f_1 \in \mathbf{P}$ и любой функции h: $\mathbf{P} \longrightarrow \mathbf{P}$ существует единственная функция f: $\mathbf{P} \longrightarrow \mathbf{P}$, удовлетворяющая условиям R1 и R2.

Чтобы определить класс рекурсивных функций, рассмотрим следующее множество $\mathcal{B}$ основных функций:

(i) Постоянные функции $\varphi(x_1, \ldots, x_n) = m$, где $m \in \mathbf{P}$, $x = (x_1, \ldots, x_n) \in \mathbf{P}^n$ — любой вектор.

(ii) Декартовы проекции $\psi_i(x_1, \ldots, x_n) = x_i$ $(i \in \mathbf{n})$.

(iii) Функция следования Пеано σ: $\sigma(n) = n + 1$.

Рассмотрим еще операцию композиции:

(iv) $h(x_1, \ldots, x_n) = f(g_1(x_1, \ldots, x_n), \ldots, g_k(x_1, \ldots, x_n))$.

Класс *примитивно рекурсивных* функций, по определению, есть минимальный класс функций, содержащий функции (i)—(iii), замкнутый относительно композиции (iv) и замкнутый относительно обобщения операции простой рекурсии R1—R2, применяемой к любому аргументу функции от многих аргументов. Рекурсивные функции получаются из него применением еще операции «минимизации».

Одна глубокая теорема (см. книги Роджерса и Дэвиса в списке литературы) утверждает, что класс *рекурсивных функций* f: $\mathbf{P}^n \longrightarrow \mathbf{P}$ совпадает с классом *вычислимых* по Тьюрингу функций.

Напомним, что класс вычислимых функций счетен, тогда как класс всех функций несчетен.

* 14.10. НЕВЫЧИСЛИМОСТЬ И ПРОБЛЕМЫ ТОЖДЕСТВА СЛОВ

Формальные языки (см. 14.4) можно рассматривать как «абстрактные теории», определяемые своими правилами порождения. Трудности распознавания приемлемых строк связаны с принципиальными проблемами оснований математики. Поговорим об этом подробнее.

Назовем *(рекурсивно) перечислимым* множеством $S \subset \mathbf{P}$ множество значений рекурсивной функции:

$$n \in S \text{ означает, что } \exists m \in \mathbf{P},\ f(m) = n \tag{18}$$

для подходящей f.

Всякое вычислимое подмножество $S \subset \mathbf{P}$, а также его дополнение (которое также вычислимо) является перечислимым. Вычислимые подмножества $\{S\}$ замкнуты относительно объединений, пересечений и дополнений, а перечислимые — только относительно объединений и пересечений. Однако дополнение перечислимого множества может не быть перечислимым.

Эти проблемы связаны со структурой множеств, определенных формулами, которые, кроме логических связок, содержат также *кванторы* $\forall$ (для всех) и $\exists$ (существует). Не существует канонического способа распознавать, когда такие выражения определяют одно и то же множество.

Проблемы тождества. Класс задач, состоящих в выяснении, эквивалентны ли два данных алгебраических выражения относительно некоторой системы тождеств, называется *проблемами тождества* (слов). Рассмотрим, например, систему равенств.

$$abc = adef, \quad abe = abc, \quad bdf = bde. \tag{19}$$

Естественно спросить, имеется ли система инструкций, которая позволяет для любых двух данных строк X, Y из символов a, b, c, d, e, f за конечное число шагов решить, можно ли перевести X в Y последовательностью подстановок из списка (19). Можно уточнить этот вопрос, поставив задачу о существовании акцептора Тьюринга, который примет строку $X \equiv Y$, если ответ положителен, и не примет ее в противном случае.

Ранее мы показали, как решить эту задачу для булевых многочленов. Проблема тождества разрешима также для абелевых групп с конечным числом образующих. К сожалению, существуют некоммутативные группы, для которых она неразрешима. Еще более простые примеры имеются среди полугрупп.

Тем более общая задача распознавания истинности или эквивалентности высказываний в формальных теориях неразрешима.

Проблема остановки. В качестве простейшего примера невычислимости упомянем еще проблему остановки для машин Тьюринга. Задача состоит в отыскании процедуры, которая для каждой машины Тьюринга T и входной ленты с символами $\mathbf{t}$ позволила бы установить, существует ли j, для которого $\delta(s^j, a^j) =$ = ОСТАНОВ. Иными словами, существует ли акцептор Тьюринга, который принимал бы в точности такие пары и отвергал остальные? Рассуждение, подобное диагональному процессу Кантора, показывает, что ответ на этот вопрос отрицателен. Проблема остановки машин Тьюринга неразрешима.

За более подробным обсуждением проблем разрешимости и вычислимости мы отсылаем читателя к списку литературы в конце этой главы.

УПРАЖНЕНИЯ В

1. Определить следующие функции (простой) рекурсией:
   а) $f(n) = 3n$, б) $g(n) = 3n^2$, в) $h(n) = n^3$.
2. Определить (простой) рекурсией функцию $n!$
3. Определить (простой) рекурсией функцию $q(n) = [\sqrt{n}]$ — наибольшее целое число q, такое, что $q^2 \leqslant n$.
4. Показать, что класс $WFPS$ правильных строк из скобок задается определением в нормальной форме Бэкуса:

   $WFPS ::= \langle\,\rangle \mid (\langle WFPS\rangle\langle WFPS\rangle) \mid (\langle WFPS\rangle) \mid WFPS\,(\,)$.

5. Построить акцептор с входным алфавитом $A=\{(,)\}$, который принимает правильные строки из скобок и никакие другие.

*6. Построить акцептор, который принимает в точности пары $(m, n)\in P$ с $m|n$.

*7. Показать, что если группа G порождена образующими $a_1,\ldots, a_n$ и $x^2=1$ для всех $x\in G$, то G — абелева группа порядка 2^n для некоторого n.

*8. Показать, что полугруппа с образующими a, b и соотношениями $aba=b$, $bab=a^r$, состоит из $5r+3$ элементов (R. C. Buck, *Am. Math. Monthly*, 75 (1968), 852—856).

СПИСОК ЛИТЕРАТУРЫ

1. Chomsky N., Formal Properties of Grammars, in Handbook of Mathematical Psychology, Wiley, 1963.
2. Davis M., Computability and Unsolvability, McGraw-Hill, 1958.
3 Ginsburg S., The Mathematical Theory of Context-Free Languages, McGraw-Hill, 1966.
4. Halmos P. R., Naive Set Theory, Van Nostrand, 1960.
5. Hays D. G., Introduction to Computational Linguistics, Elsevier, 1967.
6. Kleene S. C., Mathematical Logic, Willey 1967. (Русский перевод: Клини С., Математическая логика, М., «Мир», 1973.)
7. Minsky M., Computation-Finite and Infinite Machines; Prentice-Hall, 1967.
8. Rogers H., Theory of Recursive Functions and Effective Computability, McGraw-Hill, 1968. (Русский перевод: Роджерс Х., Теория рекурсивных функций и эффективная вычислимость, М., «Мир», 1972.)
9. Трахтенброт Б., Алгоритмы и автоматы, М., Наука, 1960.

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

ОГЛАВЛЕНИЕ

УВАЖАЕМЫЙ ЧИТАТЕЛЬ!

Ваши замечания о содержании книги, ее оформлении, качестве перевода и другие просим присылать по адресу: 129820, Москва И-110, ГСП, 1-й Рижский пер., д. 2, издательство «Мир».

Г. Биркгоф, Т. Барти

СОВРЕМЕННАЯ ПРИКЛАДНАЯ АЛГЕБРА

Редакторы Борисова Д. Ф., Маховая И. А.
Художник Антонова А. Г.
Художественный редактор Шаповалов В. И.
Технический редактор Манохина Н. И. Корректор Подгорная Т. М.

Сдано в набор 6/IV 1976 г. Подписано к печати 7/VII 1976 г. Бумага кн. журн. 60×90¹/₁₆.=12,5 бум. л. Печ. л. 25. Уч.-изд. л. 23,89. Изд. № 1/8547. Цена 1 р. 86 к. Заказ № 96

ИЗДАТЕЛЬСТВО «МИР», Москва, 1-й Рижский пер., 2

Ордена Трудового Красного Знамени Первая Образцовая типография имени А. А. Жданова Союзполиграфпрома при Государственном комитете Совета Министров СССР по делам издательств, полиграфии и книжной торговли. Москва, М-54, Валовая, 28