В.Н.НЕФЕДОВ
В.А.ОСИПОВА

КУРС ДИСКРЕТНОЙ МАТЕМАТИКИ

В. Н. НЕФЕДОВ
В. А. ОСИПОВА

КУРС ДИСКРЕТНОЙ МАТЕМАТИКИ

Допущено
Государственным комитетом СССР
по народному образованию
в качестве учебного пособия
для студентов вузов,
обучающихся по специальности
«Прикладная математика»

МОСКВА
ИЗДАТЕЛЬСТВО МАИ
1992

ББК 16.2.12
Н58
УДК 519.1(075.8)

Рецензенты:
доктор технических наук Л. Т. КУЗИН,
кандидат физико-математических наук В. В. МОРОЗОВ

Нефедов В. Н., Осипова В. А.

Н58 Курс дискретной математики: Учеб. пособие.— М.: Изд-во МАИ, 1992.— 264 с.: ил.
ISBN 5-7035-0157-X

Излагаются основы современной дискретной математики. Рассматриваются вопросы, связанные с математической логикой, теорией алгебраических систем, комбинаторикой, теорией графов. Приводится ряд практических задач и даются алгоритмы их решения.

Учебное пособие предназначено для студентов, обучающихся по специальности «Прикладная математика», но может оказаться полезным также и студентам экономических и технических факультетов, изучающих курс «Дискретная математика»

1205120000-10
Н ——————— 6-90
094(02)-92

ББК 16.2.12

ISBN 5-7035-0157-X

ПРЕДИСЛОВИЕ

В последние годы инженеры-математики, занимающиеся прикладными исследованиями, все больше используют аппарат дискретной математики. Это объясняется необходимостью создания и эксплуатации современных электронных вычислительных машин, средств передачи и обработки информации, автоматизированных систем управления и проектирования. Наконец, в современной математической науке исследования в областях, традиционно относящихся к дискретной математике (математической логике, теории алгебраических систем, теории графов и сетей и т. д.), занимают все более заметное место.

Цель создания учебного пособия — научить студентов основам дискретной математики, где дискретность понимается как антипод непрерывности. В настоящее время наряду с такими классическими разделами математики, как математический анализ, дифференциальные уравнения, в учебных планах специальности «Прикладная математика» и многих других технических и экономических специальностей появились разделы по математической логике, алгебре, комбинаторике и теории графов. В связи с этим представляется целесообразным создание учебного пособия для студентов младших курсов, в котором основы перечисленных разделов математики излагались бы в доступной форме, но достаточно полно и строго.

Данное издание включает в себя не только основные понятия и теоретические результаты, но также методы и алгоритмы решения ряда прикладных задач.

Во введении рассматриваются начальные понятия математики: множества, отношения и функции.

В первой главе излагаются основы логики высказываний и логики предикатов. Аппарат логики вы-

сказываний используется при изучении булевых функций. На примере логики высказываний осуществляется знакомство со строгой формализацией математической теории — строится исчисление высказываний. Затрагиваются вопросы, связанные с эффективной вычислимостью. С помощью машин Тьюринга и частично рекурсивных функций уточняются понятия алгоритма и вычислимости.

Во второй главе рассматриваются основные понятия теории групп, колец и полей. В качестве приложения рассматриваются элементы алгебраической теории кодирования.

Третья глава посвящена комбинаторике и перечислительной теории Пойа, использующей ряд результатов и методов теории групп.

В четвертой главе излагаются основы теории графов (ориентированных и неориентированных). Приводятся задачи теории графов и сетей, являющиеся математическими моделями ряда прикладных задач. Методы их решения доведены до уровня простых алгоритмов, реализуемых на ЭВМ. Обсуждается понятие эффективности комбинаторных алгоритмов. Рассматривается приложение теории графов к расчету электрических цепей.

Учебное пособие ориентировано на лекционный курс, читаемый одним из авторов на факультете «Прикладная математика» Московского авиационного института. В книгу не включены некоторые специальные вопросы (методы минимизации булевых функций, раскраски графов и т. д.), которые студенты изучают при выполнении курсовых работ, а также при прохождении ими лабораторного практикума.

Введение, гл. 1, 2 и разд. 4.6 написаны В. А. Осиповой, гл. 3 и разд. 4.1—4.5 — В. Н. Нефедовым.

ВВЕДЕНИЕ

Рассмотрим понятия «множество», «отношение», «функция», с помощью которых строится, по существу, любая математическая дисциплина.

0.1. НАЧАЛЬНЫЕ ПОНЯТИЯ ТЕОРИИ МНОЖЕСТВ

Под *множеством* S будем понимать любое собрание определенных и различимых между собой объектов, мыслимое как единое целое. Эти объекты называются *элементами* множества S.

В этом интуитивном определении, принадлежащем немецкому математику Г. Кантору, существенным является то обстоятельство, что собрание предметов само рассматривается как один предмет, мыслится как единое целое. Что касается самих предметов, которые могут входить в множество, то относительно них существует значительная свобода. Это может быть множество студентов, присутствующих на лекции, множество целых чисел, множество точек плоскости, множество всех людей, живущих на Земле. Заметим, что канторовская формулировка позволяет рассматривать множества, элементы которых по той или иной причине нельзя точно указать (например, множество простых чисел, множество русских воинов, погибших в битве на Куликовом поле, и т. д.).

Символом $\in$ обозначается *отношение принадлежности*. Запись $x \in S$ означает, что элемент x принадлежит множеству S. Если элемент x не принадлежит множеству S, то пишут $x \notin S$.

Г. Кантором сформулировано несколько интуитивных принципов, которые естественно считать выполняющимися для произвольных множеств.

Интуитивный принцип объемности. *Множества A и B считаются равными, если они состоят из одних и тех же элементов.*

Записывают $A = B$, если A и B равны, и $A \neq B$ — в противном случае.

Пример 0.1. Проиллюстрируем принцип объемности. Множество A всех положительных четных чисел равно множеству B положительных целых чисел, представимых в виде суммы двух положительных нечетных чисел. Действительно, если $x \in A$, то для некоторого целого положительного числа m

$x = 2m$; тогда $x = (2m-1)+1$, т. е. $x \in B$. Если $x \in B$, то для некоторых целых положительных p и q $x = (2p-1)+(2q-1) = 2(p+q-1)$, т. е. $x \in A$.

Множество, элементами которого являются объекты $a_1, \ldots, a_n$ и только они, обозначают $\{a_1, \ldots, a_n\}$.

Пример 0.2. В силу принципа объемности $\{2, 4, 6\} = \{4, 2, 6\} = \{2, 4, 4, 6\}$; $\{\{1, 2\}\} \neq \{1, 2\}$, так как единственным элементом множества $\{\{1, 2\}\}$ является множество $\{1, 2\}$, а множество $\{1, 2\}$ состоит из двух элементов: чисел 1 и 2.

При рассмотрении способов задания множеств возникает проблема их эффективного описания. Ее решение обычно основано на интуитивном понятии «формы от x». Под *формой от x* будем понимать конечную последовательность, состоящую из слов и символа x, такую, что если каждое вхождение x в эту последовательность заменить одним и тем же именем некоторого предмета соответствующего рода, то в результате получится истинное или ложное предложение. Например, формами от x являются следующие предложения: «3 делит x», «$x^2 - 2x - 1 > x$», «$x^2 = 4$», «x — родственник Иванова». Напротив, предложения «для всех x $x^2 - 4 = (x-2)(x+2)$» и «существует такое x, что $x > 0$» не являются формами от x.

Обозначим форму от x через $P(x)$.

Интуитивный принцип абстракции. *Любая форма $P(x)$ определяет некоторое множество A, а именно множество тех и только тех предметов a, для которых $P(a)$ — истинное предложение.*

Для множества A, определяемого формой $P(x)$, принято обозначение $A = \{x \mid P(x)\}$.

Пример 0.3.

1. $\{x \mid x$ — положительное число, меньшее $9\} = \{1, 2, 3, 4, 5, 6, 7, 8\}$.

2. $\{x \mid x$ — четное число$\}$ — множество четных чисел.

Описанные выше понятия теории множеств с успехом могут быть использованы в началах анализа, алгебры, математической логики и т. д. Однако надо иметь в виду, что при более строгих рассмотрениях такое интуитивное восприятие может оказаться неудовлетворительным.

Несовершенство интуитивных представлений о множествах, их недостаточность иллюстрируются, например, известным парадоксом Б. Рассела. Приведем этот парадокс. Можно указать такие множества, которые принадлежат самим себе как элементы, например множество всех множеств, и такие множества, которые не являются элементами самих себя, например множество $\{1, 2\}$, элементами которого являются числа 1 и 2. Рассмотрим теперь множество A всех таких множеств X, что X не есть элемент X. Тогда, если A не есть элемент A, то, по определению, A также есть и элемент A. С другой стороны, если

А есть элемент А, то А — одно из тех множеств X, которые не есть элементы самих себя, т. е. А не есть элемент А. В любом случае А есть элемент А и А не есть элемент А.

Этот парадокс свидетельствует о том, что широко используемая теория множеств в ее интуитивном, «наивном» изложении является противоречивой. Формализация теории множеств, связанная, в частности, с устранением парадоксов, способствовала развитию не только методов теории множеств, но и такой науки, как математическая логика.

Через $\subseteq$ обозначим *отношение включения* между множествами, т. е. $A \subseteq B$, если каждый элемент множества A есть элемент множества B. Тогда говорят, что A есть *подмножество* множества B. Если $A \subseteq B$ и $A \neq B$, то говорят, что A есть *собственное* подмножество B, и пишут $A \subset B$.

Пример 0.4. Множество четных чисел есть подмножество множества целых чисел; множество рациональных чисел есть подмножество множества действительных чисел; $\{1, 2\} \subseteq \{1, 2, 3, 4\}$.

Заметим, что: а) $X \subseteq X$; б) если $X \subseteq Y$, $Y \subseteq Z$, то $X \subseteq Z$; в) если $X \subseteq Y$ и $Y \subseteq X$, то $X = Y$.

Не надо смешивать отношения принадлежности и включения. Хотя $1 \in \{1\}$, $\{1\} \notin \{\{1\}\}$, не верно, что $1 \in \{\{1\}\}$, так как единственным элементом множества $\{\{1\}\}$ является $\{1\}$.

Множество, не содержащее элементов, называется *пустым* и обозначается $\varnothing$. Пустое множество есть подмножество любого множества.

Множество всех подмножеств A называется *множеством-степенью* и обозначается $P(A)$.

Пример 0.5. Если $A = \{1, 2, 3\}$, то $P(A) = \{\varnothing, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, A\}$.

В дальнейшем неоднократно будем пользоваться утверждением, что если множество A состоит из n элементов, то множество $P(A)$ состоит из 2^n элементов (см. задачу 3, с. 9).

Рассмотрим методы получения новых множеств из уже существующих.

Объединением множеств A и B называется множество $A \cup B$, все элементы которого являются элементами множества A или B:

$$A \cup B = \{x \mid x \in A \text{ или } x \in B\}.$$

Пересечением множеств A и B называется множество $A \cap B$, элементы которого являются элементами обоих множеств A и B:

$$A \cap B = \{x \mid x \in A \text{ и } x \in B\}.$$

Очевидно, что выполняются включения $A \cap B \subseteq A \subseteq A \cup B$ и $A \cap B \subseteq B \subseteq A \cup B$.

Относительным дополнением множества A до множества X называется множество $X \setminus A$ всех тех элементов множества X, которые не принадлежат множеству A:

$$X \setminus A = \{x | x \in X \text{ и } x \notin A\}.$$

Симметрической разностью множеств A и B называется множество

$$A + B = (A \setminus B) \cup (B \setminus A).$$

Если все рассматриваемые в ходе данного рассуждения множества являются подмножествами некоторого множества U, то это множество U называется универсальным для данного рассуждения.

Абсолютным дополнением множества A называется множество $\overline{A}$ всех тех элементов x, которые не принадлежат множеству A:

$$\overline{A} = U \setminus X.$$

Заметим, что $X \setminus A = X \cap \overline{A}$.

Для наглядного представления отношений между подмножествами какого-либо универсального множества используют диаграммы Эйлера — Венна. Само универсальное множество U изображают в виде прямоугольника, а его подмножества — в виде кругов, расположенных внутри прямоугольника. На рис. 0,1, *а* подмножество A универсального множества U изображено в виде заштрихованного круга. На рис. 0.1, *б* — *е* изображены соответственно объединение, пересечение, относительное дополнение, симметрическая разность, абсолютное дополнение.

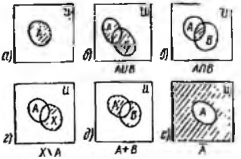


Рис. 0.1

Утверждение 0.1. *Для любых подмножеств A, B и C универсального множества U выполняются следующие тождества (основные тождества алгебры множеств):*

1. $A \cup B = B \cup A$ *(коммутативность* $\cup$*);*
2. $A \cup (B \cup C) = (A \cup B) \cup C$ *(ассоциативность* $\cup$*);*
3. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ *(дистрибутивность* $\cup$ *относительно* $\cap$*);*
4. $A \cup \varnothing = A$;
5. $A \cup \overline{A} = U$;

1′. $A \cap B = B \cap A$ *(коммутативность* $\cap$*);*

2′. $A \cap (B \cap C) = (A \cap B) \cap C$ *(ассоциативность* $\cap$*);*

3′. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ *(дистрибутивность* $\cap$ *относительно* $\cup$*);*

4′. $A \cap U = A$;

5′. $A \cap \overline{A} = \varnothing$;

6. $A \cup A = A$;
7. $A \cup U = U$;
8. $\overline{A \cup B} = \overline{A} \cap \overline{B}$ *(закон де Моргана)*;
9. $A \cup (A \cap B) = A$ *(закон поглощения)*;

6′. $A \cap A = A$;
7′. $A \cap \varnothing = \varnothing$;
8′. $\overline{A \cap B} = \overline{A} \cup \overline{B}$ *(закон де Моргана)*;
9′. $A \cap (A \cup B) = A$ *(закон поглощения)*.

Докажем тождество 3. Сначала покажем, что $A \cup (B \cap C) \subseteq (A \cup B) \cap (A \cup C)$. Действительно, если $x \in A \cup (B \cap C)$, то $x \in A$ или $x \in B \cap C$. Если $x \in A$, то $x \in A \cup B$ и $x \in A \cup C$. Следовательно, $x \in (A \cup B) \cap (A \cup C)$. Если $x \in B \cap C$, то $x \in B$ и $x \in C$. Отсюда $x \in B \cup A$ и $x \in C \cup A$, а значит, $x \in (A \cup B) \cap (A \cup C)$. Теперь покажем, что $(A \cup B) \cap (A \cup C) \subseteq A \cup (B \cap C)$. Если $x \in (A \cup B) \cap (A \cup C)$, то $x \in A \cup B$ и $x \in A \cup C$. Следовательно, $x \in A$ или $x \in B$ и $x \in C$, т. е. $x \in B \cap C$. Отсюда $x \in A \cup (B \cap C)$.

Докажем тождество 8. Пусть $x \in \overline{A \cup B}$. Тогда $x \in U$ и $x \notin A \cup B$. Следовательно, $x \notin A$ и $x \notin B$. Отсюда $x \in \overline{A}$ и $x \in \overline{B}$, а значит, $x \in \overline{A} \cap \overline{B}$. Итак, $\overline{A \cup B} \subseteq \overline{A} \cap \overline{B}$. Пусть теперь $x \in \overline{A} \cap \overline{B}$. Тогда $x \in \overline{A}$ и $x \in \overline{B}$. Следовательно, $x \in U$ и $x \notin A$ и $x \notin B$. Значит, $x \notin A \cup B$, т. е. $x \in \overline{A \cup B}$. Итак, $\overline{A} \cap \overline{B} \subseteq \overline{A \cup B}$.

Остальные тождества доказываются аналогично.

Утверждение 0.2. *Предложения о произвольных множествах A и B попарно эквивалентны:*

1) $A \subseteq B$; 2) $A \cap B = A$; 3) $A \cup B = B$.

Докажем, что из первого предложения следует второе. Действительно, так как $A \cap B \subseteq A$, то достаточно показать, что в этом случае $A \subseteq A \cap B$. Но если $x \in A$, то $x \in B$, так как $A \subseteq B$, и, следовательно, $x \in A \cap B$.

Докажем, что из второго предложения следует третье. Так как $A \cap B = A$, то $A \cup B = (A \cap B) \cup B$. По закону поглощения (см. тождество 9) $B \cup (A \cap B) = B$. Отсюда, используя закон коммутативности, получаем $A \cup B = B$.

Докажем, что из третьего предложения следует первое. Так как $A \subseteq A \cup B$, а по условию третьего предложения $A \cup B = B$, то $A \subseteq B$.

Задачи и упражнения

1. Доказать, что $\{\{1, 2\}, \{2, 3\}\} \neq \{1, 2, 3\}$.
2. Существуют ли такие множества A, B и C, что $A \cap B \neq \varnothing$, $A \cap C = \varnothing$, $(A \cap B) \setminus C = \varnothing$?
3. Доказать, что если множество A состоит из n элементов, то множество $P(A)$ состоит из 2^n элементов.
4. Доказать следующие тождества:

а) $(A \cap B) \cup (A \cap \overline{B}) = (A \cup B) \cap (A \cup \overline{B}) = A$;

б) $(\overline{A \cup B}) \cap A = A \cap B$;
в) $(A \setminus B) \setminus C = (A \setminus C) \setminus (B \setminus C)$;
г) $A \setminus (B \cup C) = (A \setminus B) \setminus C$;
д) $A \setminus (B \setminus C) = (A \setminus B) \cup (A \cap C)$;
е) $A + (B + C) = (A + B) + C$;
ж) $A \cap (B + C) = (A \cap B) + (A \cap C)$.

5. Доказать, что:
а) $(A \cup B) \subseteq C$ тогда и только тогда, когда $A \subseteq C$ и $B \subseteq C$;
б) $A \subseteq B \cap C$ тогда и только тогда, когда $A \subseteq B$ и $A \subseteq C$;
в) $A \cap B \subseteq C$ тогда и только тогда, когда $A \subseteq \overline{B} \cup C$;
г) $A \subseteq B \cup C$ тогда и только тогда, когда $A \cap \overline{B} \subseteq C$.

6. Доказать, что $P(A \cap B) = P(A) \cap P(B)$.

7. Какие из утверждений верны для всех A, B и C:
а) если $A \in B$ и $B \in C$, то $A \in C$;
б) если $A \cap B \subseteq C$ и $A \cup B \subseteq C$, то $A \cap C = \varnothing$;
в) если $A \neq B$ и $B \neq C$, то $A \neq C$;
г) если $A \subseteq \overline{B \cup C}$ и $B \subseteq \overline{A \cup C}$, то $B = \varnothing$?

8. Решить систему уравнений

$$\begin{cases} A \cap X = B; \\ A \cup X = C, \end{cases}$$

где A, B и C — данные множества; $B \subseteq A \subseteq C$.

0.2. ОТНОШЕНИЯ И ФУНКЦИИ

Упорядоченная пара $\langle x, y \rangle$ интуитивно определяется как совокупность, состоящая из двух элементов x и y, расположенных в определенном порядке. Две пары $\langle x, y \rangle$ и $\langle u, v \rangle$ считаются равными тогда и только тогда, когда $x = u$ и $y = v$.

Упорядоченная n-ка элементов $x_1, \ldots, x_n$ обозначается $\langle x_1, \ldots, x_n \rangle$ и, по определению, есть $\langle\langle x_1, \ldots, x_{n-1} \rangle, x_n \rangle$.

Бинарным (или двуместным) отношением ρ называется множество упорядоченных пар. Если ρ есть некоторое отношение и пара $\langle x, y \rangle$ принадлежит этому отношению, то наряду с записью $\langle x, y \rangle \in \rho$ употребляется запись $x \rho y$. Элементы x и y называются *координатами* (или *компонентами*) отношения ρ. *n-арным отношением* называется множество упорядоченных n-ок.

Областью определения бинарного отношения ρ называется множество $D_\rho = \{x \mid$ существует такое y, что $x \rho y\}$.

Областью значений бинарного отношения ρ называется множество $R_\rho = \{y \mid$ существует такое x, что $x \rho y\}$.

Пример 0.6.

1. Множество $\{<1, 2>, <2, 4>, <3, 3>, <2, 1>\}$ — бинарное отношение.

2. Отношение равенства на множестве действительных чисел есть множество $\{<x, y> | x$ и y — действительные числа и x равно $y\}$. Для этого отношения существует специальное обозначение $=$. Область определения D_ρ совпадает с областью значений R_ρ и является множеством действительных чисел.

3. Отношение «меньше, чем» на множестве целых чисел есть множество $\{<x, y> |$ для целых чисел x и y найдется положительное число z такое, что $x + z = y\}$. Для этого отношения есть специальное обозначение $<$. Здесь D_ρ совпадает с R_ρ и является множеством целых чисел.

Прямым произведением множеств X и Y называется совокупность всех упорядоченных пар $<x, y>$ таких, что $x \in X$ и $y \in Y$. Обозначается прямое произведение множеств X и Y через $X \times Y$.

Каждое отношение ρ есть подмножество прямого произведения некоторых множеств X и Y таких, что $D_\rho \subseteq X$ и $R_\rho \subseteq Y$. Если $X = Y$, то говорят, что ρ есть отношение на множестве X.

Прямым произведением множеств $X_1, \ldots, X_n$ называется совокупность всех упорядоченных n-ок $<x_1, \ldots, x_n>$ таких, что $x_i \in X_i$, $i = 1, \ldots, n$. Обозначается прямое произведение множеств $X_1, \ldots, X_n$ через $X_1 \times X_2 \times \ldots \times X_n$. Если $X_1 = X_2 = \ldots = X_n = X$, то пишут $X_1 \times X_2 \times \ldots \times X_n = X^n$. Любое n-местное отношение есть подмножество прямого произведения некоторых множеств $X_1, \ldots, X_n$.

Пример 0.7.

1. Пусть $X = \{1, 2, 3\}$, $Y = \{0, 1\}$. Тогда

$X \times Y = \{<1, 0>, <1, 1>, <2, 0>, <2, 1>, <3, 0>, <3, 1>\}$;

$Y \times X = \{<0, 1>, <0, 2>, <0, 3>, <1, 1>, <1, 2>, <1, 3>\}$.

Мы указали, кроме того, такие множества X и Y, что $X \times Y \neq Y \times X$.

2. Пусть X — множество точек отрезка $[0, 1]$, а Y — множество точек отрезка $[1, 2]$. Тогда $X \times Y$ — множество точек квадрата $[0, 1] \times [1, 2]$ с вершинами в точках $(0, 1)$, $(0, 2)$, $(1, 1)$, $(1, 2)$.

Для бинарных отношений обычным образом определены теоретико-множественные операции объединения, пересечения и т. д.

Обратным отношением для ρ называется отношение

$$\rho^{-1} = \{<x, y> | <y, x> \in \rho\}.$$

Композицией отношений ρ_1 и ρ_2 называется отношение $\rho_2 \circ \rho_1 = \{<x, z> |$ существует y такое, что $<x, y> \in \rho_1$ и $<y, z> \in \rho_2\}$.

Для любых бинарных отношений выполняются следующие свойства:

1°. $(\rho^{-1})^{-1} = \rho$;

2°. $(\rho_2 \circ \rho_1)^{-1} = \rho_1^{-1} \circ \rho_2^{-1}$.

Первое свойство очевидно. Для доказательства второго свойства покажем, что множества, записанные в левой и правой частях равенства, состоят из одних и тех же элементов. Действительно, $\langle x, y\rangle \in (\rho_2 \circ \rho_1)^{-1} \Leftrightarrow \langle y, x\rangle \in \rho_2 \circ \rho_1 \Leftrightarrow$ существует z такое, что $\langle y, z\rangle \in \rho_1$ и $\langle z, x\rangle \in \rho_2 \Leftrightarrow$ существует z такое, что $\langle z, y\rangle \in \rho_1^{-1}$ и $\langle x, z\rangle \in \rho_2^{-1} \Leftrightarrow \langle x, y\rangle \in \rho_1^{-1} \circ \rho_2^{-1}$.

Бинарное отношение f называется *функцией,* если из $\langle x, y\rangle \in f$ и $\langle x, z\rangle \in f$ следует, что $y = z$.

Поскольку функции являются бинарными отношениями, то к ним применим интуитивный принцип объемности, т. е. две функции f и g равны, если они состоят из одних и тех же элементов. Область определения функции обозначается D_f, а область ее значений — R_f. Определяются они так же, как и для бинарных отношений. Часто приходится сталкиваться с трудностями при определении области значений функции. Поэтому, если $D_f = X$ и $R_f \subseteq Y$, то говорят, что функция f задана *на* множестве X со значениями *во* множестве Y и осуществляет *отображение* множества X во множество Y (или устанавливает *соответствие* между множествами X и Y). Это отображение обозначается таким образом: $f : X \to Y$.

Если f — функция, то вместо $\langle x, y\rangle \in f$ пишут $y = f(x)$ и говорят, что y — значение, соответствующее аргументу x, или y — образ элемента x при отображении f. При этом x называют прообразом элемента y.

Пример 0.8. Отношение $\{\langle 1, 2\rangle, \langle 2, 3\rangle, \langle \square, \triangle\rangle\}$ — функция; отношение $\{\langle 1, 2\rangle, \langle 1, 3\rangle, \langle 2, 4\rangle\}$ не является функцией; отношение $\{\langle x, x^2 + 2x + 1\rangle \mid x$ — действительное число$\}$ — функция, которую обычно обозначают $y = x^2 + 2x + 1$.

Назовем f n-местной функцией из X в Y, если $f : X^n \to Y$. Тогда пишем $y = f(x_1, \ldots, x_n)$ и говорим, что y — значение функции при значении аргументов $x_1, \ldots, x_n$.

Пусть $f : X \to Y$. Функция (отображение) f называется *инъективной (инъективным),* если для любых x_1, x_2, y из $y = f(x_1)$ и $y = f(x_2)$ следует, что $x_1 = x_2$ (или, иначе, из $\langle x_1, y\rangle \in f$ и $\langle x_2, y\rangle \in f$ следует, что $x_1 = x_2$). Функция (отображение) f называется *сюръективной (сюръективным),* если для любого элемента $y \in Y$ существует элемент $x \in X$ такой, что $y = f(x)$. Функция (отображение) f называется *биективной (биективным),* если f одновременно сюръективна и инъективна. Если существует биективная функция $f : X \to Y$, то говорят, что f осуществляет *взаимно однозначное соответствие* между множествами X и Y.

Пример 0.9. Рассмотрим три функции, отображающие множество действительных чисел R во множество действительных чисел, $f_i : R \to R$, $i = 1, 2, 3$:

1) функция $f_1(x) = e^x$ инъективна, но не сюръективна;

2) функция $f_2(x) = x^3 - x$ сюръективна, но не инъективна;

3) функция $f_3(x) = 2x + 1$ биективна.

Композиция двух функций f и g есть отношение $g \circ f = \{\langle x, z\rangle \mid$ существует y такое, что xfy и $ygz\}$.

Утверждение 0.3. *Композиция двух функций есть функция. При этом, если* $f : X \to Y$, $g : Y \to Z$, то $g \circ f : X \to Z$.

Действительно, если $\langle x, y\rangle \in g \circ f$ и $\langle x, z\rangle \in g \circ f$, то существует такое u, что xfu, ugy, и существует такое v, что xfv, vgz. Поскольку f — функция, то $u = v$; поскольку g — функция, то $y = z$ и, следовательно, $g \circ f$ — функция. Вторая часть утверждения очевидна.

Верно также и следующее утверждение.

Утверждение 0.4. *Композиция двух биективных функций есть биективная функция.*

Тождественным отображением множества X в себя называется отображение $e_X : X \to X$ такое, что для любого $x \in X$ $e_X(x) = x$. Тогда, если $f : X \to Y$, то $e_Y \circ f = f$, $f \circ e_X = f$.

Пусть f^{-1} — отношение, обратное f. Выясним, при каких условиях отношение f^{-1} будет функцией. Его называют тогда *обратной функцией* или, если f осуществляет отображение множества X во множество Y, *обратным отображением*.

Утверждение 0.5. *Отображение* $f : X \to Y$ *имеет обратное отображение* $f^{-1} : Y \to X$ *тогда и только тогда, когда* f *— биекция.*

Если f — биекция, то, поскольку f сюръективно, f^{-1} определено на множестве Y. Кроме того, f — функция, так как если $\langle y, x_1\rangle \in f^{-1}$ и $\langle y, x_2\rangle \in f^{-1}$, то $\langle x_1, y\rangle \in f$ и $\langle x_2, y\rangle \in f$, а в силу инъективности f имеем $x_1 = x_2$.

Пусть теперь отображение f имеет обратное отображение f^{-1}, определенное на множестве Y со значениями во множестве X. Тогда f сюръективно, поскольку любой элемент $y \in Y$ имеет прообраз $x \in X$. При этом f инъективно, так как если $\langle x_1, y\rangle \in f$ и $\langle x_2, y\rangle \in f$, то $\langle y, x_1\rangle \in f^{-1}$, $\langle y, x_2\rangle \in f^{-1}$, а поскольку f^{-1} — функция, то $x_1 = x_2$.

Заметим, что, для того чтобы обратное отношение f^{-1} было функцией, достаточно инъективности функции f.

Поскольку функция есть бинарное отношение, то выполняются следующие свойства инъективных функций f и g:

1°. $(f^{-1})^{-1} = f$;

2°. $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$.

Если $f : X \to Y$ — биекция, то

3°. $(f^{-1} \circ f) = e_X$;

4°. $(f \circ f^{-1}) = e_Y$.

Задачи и упражнения

1. Найти D_ρ, R_ρ, ρ^{-1}, $\rho \circ \rho$, $\rho^{-1} \circ \rho$, $\rho \circ \rho^{-1}$ отношений:
а) $\rho = \{\langle x, y\rangle \mid x, y$ — натуральные числа и x делит $y\}$;
б) $\rho = \{\langle x, y\rangle \mid x, y$ — действительные числа и $x + y \leqslant 0\}$;
в) $\rho = \{\langle x, y\rangle \mid x, y \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ и $y \geqslant \sin x\}$.

2. Найти геометрическую интерпретацию множества $A \times B$, где A — множество точек отрезка $[0, 1]$; B — множество точек квадрата с вершинами в точках $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$.

3. Доказать, что $(A \times B) \cup (C \times D) \subseteq (A \cup C) \times (B \cup D)$. При каких A, B, C, D включение можно заменить равенством?

4. Доказать, что для произвольных множеств A, B, C, D:
а) $(A \cap B) \times (C \cap D) = (A \times C) \cap (B \times D)$;
б) $(A \cup B) \times C = (A \times C) \cup (B \times C)$;
в) $(A \setminus B) \times C = (A \times C) \setminus (B \times C)$;
г) $A \times (B \setminus C) = (A \times B) \setminus (A \times C)$.

5. Пусть A, $B \neq \varnothing$ и $(A \times B) \cup (B \times A) = C \times D$. Доказать, что в этом случае $A = B = C = D$.

6. Доказать, что число бинарных отношений на n-элементном множестве равно 2^{n^2}.

7. Пусть X — конечное множество и отображение $f: X \to X$ инъективно. Доказать, что тогда f биективно.

8. *Характеристической* функцией множества A называется функция

$$\chi_A(x) = \begin{cases} 1, & \text{если } x \in A; \\ 0, & \text{если } x \notin A. \end{cases}$$

Пусть известны характеристические функции множеств A и B. Доказать, что:
а) $\chi_{A \cap B}(x) = \chi_A(x) \cdot \chi_B(x)$;
б) $\chi_{A \cup B}(x) = \chi_A(x) + \chi_B(x) - \chi_A(x) \cdot \chi_B(x)$;
в) $\chi_{\overline{A}}(x) = 1 - \chi_A(x)$;
г) $\chi_{A \setminus B}(x) = \chi_A(x) - \chi_A(x) \cdot \chi_B(x)$.

9. Доказать, что для любой функции f и произвольных множеств A и B:
а) $f(A \cup B) = f(A) \cup f(B)$;
б) $f(A \cap B) \subseteq f(A) \cap f(B)$. В каком случае включение можно заменить равенством?
в) $f(A) \setminus f(B) \subseteq f(A \setminus B)$.

0.3. СПЕЦИАЛЬНЫЕ БИНАРНЫЕ ОТНОШЕНИЯ

В математике важную роль играют два вида специальных бинарных отношений: отношение эквивалентности и отношение порядка.

Отношение ρ на множестве X называется *рефлексивным*, если для любого элемента $x \in X$ выполняется $x\rho x$.

Отношение ρ на множестве X называется *симметричным*, если для любых $x, y \in X$ из $x\rho y$ следует $y\rho x$.

Отношение ρ на множестве X называется *транзитивным*, если для любых $x, y, z \in X$ из $x\rho y$, $y\rho z$ следует $x\rho z$.

Рефлексивное, симметричное и транзитивное отношение на множестве X называется *отношением эквивалентности* на множестве X.

Пример 0.10.

1. Отношение равенства на множестве целых чисел есть отношение эквивалентности.

2. Отношение подобия на множестве треугольников есть отношение эквивалентности.

3. Отношение $x < y$ на множестве действительных чисел не рефлексивно, не симметрично, но транзитивно на этом множестве.

4. Отношение сравнимости по модулю натурального числа n на множестве целых чисел $Z: x \equiv y \pmod n$ тогда и только тогда, когда $x - y$ делится на n. Это отношение рефлексивно на Z, так как для любого $x \in Z$ $x - x = 0$, и, следовательно, делится на n. Это отношение симметрично, так как если $x - y$ делится на n, то $y - x$ делится на n. Это отношение транзитивно, так как если $x - y$ делится на n, то для некоторого целого t_1 имеем $x - y = t_1 n$, а если $y - z$ делится на n, то для некоторого целого t_2 имеем $y - z = t_2 n$. Отсюда $x - z = (t_1 + t_2)n$, т. е. $x - z$ делится на n.

5. Отношение принадлежности к одной студенческой группе на множестве студентов института — отношение эквивалентности.

6. На множестве $N \times N$, где N — множество натуральных чисел, определим отношение $\rho: \langle x, y\rangle \rho \langle u, v\rangle \Leftrightarrow xv = yu$. Это отношение рефлексивно: $\langle x, y\rangle \rho \langle x, y\rangle$, так как $xy = yx$; симметрично: если $\langle x, y\rangle \rho \langle u, v\rangle$, то $\langle u, v\rangle \rho \langle x, y\rangle$, так как из $xv = yu$ следует, что и $uy = vx$; транзитивно: если $\langle x, y\rangle \rho \langle u, v\rangle$, $\langle u, v\rangle \rho \langle w, z\rangle$, то $\langle x, y\rangle \rho \langle w, z\rangle$, так как, перемножая левые и правые части равенств $xv = yu$ и $uz = vw$, после сокращения получаем $xz = yw$.

Пусть ρ — отношение эквивалентности на множестве X.

Классом эквивалентности, порожденным элементом x, называется подмножество множества X, состоящее из тех элементов

$y \in X$, для которых $x\rho y$. Класс эквивалентности, порожденный элементом x, обозначается через $[x]$:

$$[x] = \{y \mid y \in X \text{ и } x\rho y\}.$$

Пример 0.11.

1. Отношение равенства на множестве целых чисел порождает следующие классы эквивалентности: для любого элемента $x \in \mathbf{Z}$ $[x] = \{x\}$, т. е. каждый класс эквивалентности состоит только из одного элемента — числа x.

2. Отношение сравнимости по модулю числа n на множестве целых чисел $\mathbf{Z}$ порождает следующие классы эквивалентности: вместе с любым числом $a \in \mathbf{Z}$ в этом же классе эквивалентности содержатся все числа вида $a + kn$, где k — целое. Очевидно, что числа $0, 1, 2, \ldots, n-1$ порождают различные классы эквивалентности, которые обозначим $\{0\}, \{1\}, \{2\}, \ldots, \{n-1\}$. Они называются классами вычетов по модулю n. Все остальные классы эквивалентности для этого отношения совпадают с ними, так как любое число $a \in \mathbf{Z}$ можно представить в виде $a = qn + r$, где $0 \leqslant r < n$.

3. Для отношения принадлежности к одной студенческой группе классом эквивалентности является множество студентов одной группы.

4. Класс эквивалентности, порожденный парой $\langle x, y\rangle$ для отношения ρ из примера 0.10, п. 6, определяется соотношением

$$[\langle x, y\rangle] = \left\{\langle u, v\rangle \mid \frac{x}{y} = \frac{u}{v}\right\}.$$

Каждый класс эквивалентности в этом случае определяет одно положительное рациональное число.

Утверждение 0.6. *Пусть ρ — отношение эквивалентности на множестве X. Тогда:* 1) *если $x \in X$, то $x \in [x]$;* 2) *если $x, y \in X$ и $x\rho y$, то $[x] = [y]$ (т. е. класс эквивалентности порождается любым своим элементом).*

Для доказательства первой части утверждения достаточно воспользоваться рефлексивностью отношения ρ: $x\rho x$ и, следовательно, $x \in [x]$. Докажем вторую часть утверждения. Пусть $z \in [y]$. Тогда $y\rho z$ и в силу транзитивности отношения ρ $x\rho z$, т. е. $z \in [x]$. Отсюда $[y] \subseteq [x]$. Аналогично в силу симметричности ρ можно показать, что $[x] \subseteq [y]$, а значит, $[x] = [y]$.

Разбиением множества X называется совокупность попарно непересекающихся подмножеств X таких, что каждый элемент множества X принадлежит одному и только одному из этих подмножеств.

Пример 0.12.

1. $X = \{1, 2, 3, 4, 5\}$. Тогда $\{\{1, 2\}, \{3, 5\}, \{4\}\}$ — разбиение множества X.

2. Пусть X — множество студентов института. Тогда разбиением этого множества является, например, совокупность студенческих групп.

Утверждение 0.7. *Всякое разбиение множества X определяет на X отношение эквивалентности ρ: $x\rho y$ тогда и только тогда, когда x и y принадлежат одному подмножеству разбиения.*

Рефлексивность и симметричность ρ очевидны. Пусть теперь $x\rho y$ и $y\rho z$. Тогда $x, y \in X_1$, $y, z \in X_2$, где X_1 и X_2 — подмножества из разбиения X. Поскольку $y \in X_1$, $y \in X_2$, то $X_1 = X_2$. Следовательно, $x, z \in X_1$ и $x\rho z$.

Утверждение 0.8. *Всякое отношение эквивалентности ρ определяет разбиение множества X на классы эквивалентности относительно этого отношения.*

Докажем, что совокупность классов эквивалентности определяет разбиение множества X. В силу утверждения 0.6 $x \in [x]$, а следовательно, каждый элемент множества X принадлежит некоторому классу эквивалентности. Из утверждения 0.6 вытекает также, что два класса эквивалентности либо не пересекаются, либо совпадают, так как если $z \in [x]$ и $z \in [y]$, то $x\rho z$, откуда $[x] = [z]$, и $y\rho z$, откуда $[y] = [z]$. Следовательно, $[x] = [y]$.

Совокупность классов эквивалентности элементов множества X по отношению эквивалентности ρ называется *фактор-множеством* множества X по отношению ρ и обозначается X/ρ.

Рассмотрим еще один тип специальных бинарных отношений. Отношение ρ на множестве X называется *антисимметричным*, если для любых $x, y \in X$ из $x\rho y$ и $y\rho x$ следует $x = y$.

Рефлексивное, антисимметричное и транзитивное отношение называется отношением *частичного порядка* на множестве X и обозначается символом $\prec$.

Пример 0.13.

1. Отношение $x \leqslant y$ на множестве действительных чисел есть отношение частичного порядка.

2. Во множестве подмножеств некоторого универсального множества U отношение $A \subseteq B$ есть отношение частичного порядка.

3. Схема организации подчинения в учреждении есть отношение частичного порядка на множестве должностей.

Отношение частичного порядка на множестве X, для которого любые два элемента сравнимы, т. е. для любых $x, y \in X$ $x \prec y$ или $y \prec x$, называется отношением *линейного порядка*.

Пример 0.14. В примере 0.13 отношение, определенное в п. 1, есть отношение линейного порядка, а отношение, определенное в п. 2, таковым не является.

Заметим, что мы определили тип отношений, прообразом которых служит интуитивное понятие отношения порядка (предшествования, предпочтения).

Пусть на множестве X задано отношение частичного порядка ρ. Как можно задать отношение частичного порядка на множестве $X \times X$, т. е. как сравнивать пары элементов из множест-

ва X? Один из возможных способов состоит в следующем: на множестве $X \times X$ определяем отношение Π условием $\langle a, b\rangle \Pi \langle c, d\rangle \Leftrightarrow apc$ и bpd. Отношение Π есть отношение частичного порядка. Оно называется отношением Парето.

Множество X с заданным на нем частичным (линейным) порядком называется *частично (линейно) упорядоченным*.

Рассмотрим непустое конечное множество X, на котором задано отношение частичного порядка $\prec$. Запишем $x \prec y$, если $x \prec y$ и $x \neq y$. Говорят, что элемент y покрывает элемент x, если $x \prec y$ и не существует такого элемента u, что $x \prec u \prec y$. Тогда $x \prec y$ равносильно тому, что существуют такие элементы $x_1, \ldots, x_n$, что $x = x_1 \prec x_2 \prec \ldots \prec x_n = y$, где x_{i+1} покрывает x_i.

Любое частично упорядоченное множество можно представить в виде схемы, в которой каждый элемент изображается точкой на плоскости, и если y покрывает x, то точки x и y соединяют отрезком, причем точку, соответствующую x, располагают ниже y. Такие схемы называются диаграммами Хассе. На рис. 0.2, *а*, *б* показаны две диаграммы Хассе, причем вторая соответствует линейно упорядоченному множеству.

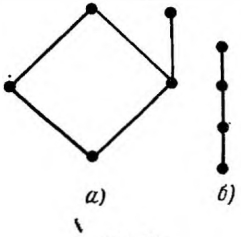


Рис 0.2

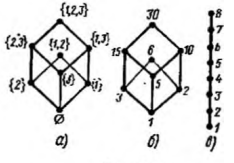


Рис. 0.3

Пример 0.15. Рассмотрим три отношения частичного порядка и построим для них диаграммы Хассе:

1. Пусть $A = \{1, 2, 3\}$. На множестве $P(A)$ рассмотрим отношение «быть подмножеством». Множество $P(A)$ содержит восемь элементов: $\{\varnothing, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}$. Его диаграмма Хассе изображена на рис. 0.3, *а* (отрезок, соединяющий точки $\{1, 2\}$ и $\{1, 2, 3\}$, на рисунке не показан).

2. Пусть $X = \{1, 2, 3, 5, 6, 10, 15, 30\}$. На этом восьмиэлементном множестве рассмотрим отношение частичного порядка «$x \prec y \Leftrightarrow y$ делится на x». Диаграмма Хассе для этого отношения изображена на рис. 0.3, *б* (отрезок, соединяющий точки 6 и 30, на рисунке не показан).

3. На множестве $X = \{1, 2, 3, 4, 5, 6, 7, 8\}$ рассмотрим отношение линейного порядка $\leqslant$. Его диаграмма Хассе изображена на рис. 0.3, *в*.

Заметим, что диаграммы Хассе первых двух отношений совпадают. Это означает, что эти частично упорядоченные множества имеют одинаковую структуру, причем отличную от структуры третьего частично упорядоченного множества, хотя оно тоже содержит восемь элементов. Более точно такая общность структуры определяется понятием изоморфизма.

Два частично упорядоченных множества X и Y называются *изоморфными*, если существует биекция $\varphi: X \to Y$, сохраняющая отношение частичного порядка. Иными словами, $x_1 \underset{X}{\prec} x_2$ тогда и только тогда, когда $\varphi(x_1) \underset{Y}{\prec} \varphi(x_2)$, где $\underset{X}{\prec}$ и $\underset{Y}{\prec}$ — отношения частичного порядка, заданные на множествах X и Y соответственно.

Частично упорядоченные множества, рассмотренные в примере 0.15 (пункты 1 и 2), изоморфны.

Утверждение 0.9. *Всякое частично упорядоченное множество X изоморфно некоторой системе подмножества множества X, частично упорядоченной отношением включения.*

Для каждого элемента $a \in X$ рассмотрим множество $S_a = \{x \in X \mid x \prec a\}$. Тогда $S_a \subseteq X$ и $\{S_a \mid a \in X\}$ — совокупность всех таких подмножеств. Докажем, что эта система подмножеств, частично упорядоченная отношением включения, изоморфна X. Рассмотрим отображение $\varphi: X \to \{S_a \mid a \in X\}$ такое, что $\varphi(a) = S_a$. Тогда φ — биекция. Действительно, если $S_a = S_b$, то, поскольку $a \in S_a$ в силу рефлексивности $\prec$, имеем $a \in S_b$ и $a \prec b$. Аналогично получаем $b \prec a$ и в силу антисимметричности $\prec$ имеем $a = b$, т. е. отображение φ инъективно. Кроме того, φ сюръективно, так как у любого подмножества S_a есть прообраз a. Докажем, что отображение сохраняет отношение частичного порядка. Пусть $a \prec b$. Тогда из $x \prec a$ в силу транзитивности отношения $\prec$ следует $x \prec b$, а значит, и $S_a \subseteq S_b$. Если $S_a \subseteq S_b$, то, поскольку $a \in S_a$, имеем $a \in S_b$, откуда $a \prec b$.

Задачи и упражнения

1. Привести примеры отношений:
   а) не рефлексивного, но симметричного и транзитивного;
   б) не симметричного, но рефлексивного и транзитивного;
   в) не транзитивного, но рефлексивного и симметричного.
2. На множестве прямых на плоскости рассмотрим отношения:

а) параллельности прямых;

б) перпендикулярности прямых.

Определить, будут ли эти отношения отношениями эквивалентности на этом множестве.

3. На множестве $N \times N$, где N — множество натуральных чисел $\{1, 2, 3 \ldots\}$, определим отношение $\rho: \langle x, y\rangle \rho \langle u, v\rangle \Leftrightarrow x + v = y + u$. Доказать, что ρ — отношение эквивалентности на этом множестве.

4. Доказать, что пересечение отношений эквивалентности на множестве X есть отношение эквивалентности на этом множестве.

5. Доказать, что объединение $\rho_1 \cup \rho_2$ двух отношений эквивалентности ρ_1 и ρ_2, заданных на множестве X, является отношением эквивалентности тогда и только тогда, когда $\rho_1 \cup \rho_2 = \rho_1 \circ \rho_2$.

6. Доказать, что если ρ — частичный порядок, то ρ^{-1} — также частичный порядок.

7. Привести пример линейного порядка на множестве $N \times N$, где N — множество натуральных чисел.

8. Доказать, что всякий частичный порядок на конечном множестве может быть продолжен до линейного порядка.

0.4. АЛГЕБРАИЧЕСКИЕ ОПЕРАЦИИ

Систематизируем понятие алгебраической операции, с которым мы уже встречались в различных разделах курса математики.

Пусть дано множество M. Говорят, что на M определена *бинарная алгебраическая операция*, если всякой упорядоченной паре элементов множества M по некоторому закону ставится в соответствие вполне определенный элемент этого же множества.

Примерами бинарных операций на множестве целых чисел являются сложение и умножение. Однако нашему определению не удовлетворяют, например, множество отрицательных целых чисел относительно умножения и множество действительных чисел относительно деления из-за невозможности деления на нуль.

Среди известных бинарных операций, производимых не над числами, отметим векторное умножение векторов пространства, умножение квадратных матриц порядка n, композицию отображений множества X в себя, теоретико-множественное объединение и пересечение множеств.

Как видим, фактическое задание алгебраической операции на множестве может быть произведено различными методами. Возможно также непосредственное перечисление всех результатов операции для конечных множеств. Его удобно описать с помощью так называемой таблицы Кэли. Слева и сверху квадратной таблицы выписывают все элементы множества. На пере-

Таблица 0.1

	X_1	X_2	X_3	X_4
X_1	X_1	X_2	X_3	X_4
X_2	X_2	X_3	X_4	X_1
X_3	X_3	X_4	X_1	X_2
X_4	X_4	X_1	X_2	X_3

Таблица 0.2

	И	Л
И	И	Л
Л	Л	Л

сечении строки, соответствующей элементу a, и столбца, соответствующего элементу b, записывают результат операции над a и b. Из двух приведенных таблиц Кэли (табл. 0.1 и 0.2) вторая — таблица для операции конъюнкции на множестве (И, Л), о которой будет говориться в гл. 1.

Будем употреблять следующую терминологию и символику: операцию называть умножением, а результат применения операции к элементам a и b — произведением ab. Это мультипликативная терминология. Иногда естественнее и удобнее использовать аддитивную терминологию и символику: операцию называть сложением, а результат ее выполнения — суммой $a+b$ элементов a и b.

Если для любых элементов a и b множества M справедливо равенство $ab = ba$, то операцию называют *коммутативной*. Коммутативны, например, сложение и умножение чисел, сложение матриц одного порядка и т. д. Некоммутативными операциями являются векторное произведение векторов, произведение матриц порядка n при $n \geqslant 2$ и др.

Если для любых элементов a, b, c множества M справедливо равенство $a(bc) = (ab)c$, то операцию называют *ассоциативной*. Ассоциативны, например, сложение и умножение целых чисел, умножение матриц, композиция отображений, а также операции, определенные таблицами Кэли. Неассоциативной операцией является векторное умножение векторов пространства.

В ряде случаев множество M, на котором определена алгебраическая операция, обладает *единичным элементом*, т. е. таким элементом e, что $ae = ea = a$ для всех a из M. Единичный элемент единственен, так как если существует еще один элемент e' с этим же свойством, то $ee' = e$ и $ee' = e'$, откуда $e = e'$. При аддитивной записи единичный элемент называется нулевым и обозначается 0.

На множестве квадратных матриц порядка n единичным

элементом относительно операции умножения является единичная матрица, на множестве отображений множества X в себя единичным элементом относительно композиции отображений является тождественное отображение. Число 1 есть единичный элемент множества целых чисел относительно операции умножения, а множество четных чисел не имеет единичного элемента относительно этой операции.

Если операция ассоциативна, то можно однозначно говорить о произведении любого конечного числа элементов, взятых в определенном порядке. Пусть дана упорядоченная система из n элементов множества $M: a_1, a_2, \ldots, a_n$, в которой некоторым образом расставлены скобки, указывающие на порядок выполнения операции.

Теорема 0.1. *Если операция, определенная на M, ассоциативна, то результат ее последовательного применения к n элементам множества не зависит от расстановки скобок.*

Доказательство проведем индукцией по числу множителей n. Для $n = 3$ утверждение теоремы следует из закона ассоциативности. Докажем это для n множителей, предполагая, что для меньшего числа множителей утверждение верно. В этом случае достаточно доказать, что для любых k и l, где $1 \leqslant k, l \leqslant n-1$, $(a_1 \ldots a_k)(a_{k+1} \ldots a_n) = (a_1 \ldots a_l)(a_{l+1} \ldots a_n)$, так как внутри скобок расстановка их несущественна по индуктивному предположению. Для этого покажем, что обе части равенства совпадают с произведением элементов $a_1, \ldots, a_n$, взятых в следующем фиксированном порядке: $(\ldots((a_1a_2)a_3)\ldots a_{n-1})a_n$ (это произведение называется левонормированным произведением элементов $a_1, \ldots, a_n$). Действительно, при $k = n-1$ имеем $(a_1 \ldots a_{n-1})a_n = (\ldots(a_1a_2)\ldots a_{n-1})a_n$, т. е. левонормированное произведение. При $k < n-1$ ввиду ассоциативности получаем $(a_1 \ldots a_k)(a_{k+1} \ldots a_n) = (a_1 \ldots a_k)((a_{k+1} \ldots a_{n-1})a_n) = ((a_1 \ldots a_k)(a_{k+1} \ldots a_{n-1}))a_n = (\ldots(\ldots(a_1a_2)\ldots a_k)a_{k+1})\ldots a_{n-1})a_n$, т. е. снова имеем левонормированное произведение. К такому же виду приводится и правая часть доказываемого равенства.

В силу теоремы 0.1 при записи и вычислении произведения $a_1 \ldots a_n$ скобки не ставят, а следят только за порядком множителей, и то лишь в случае, если операция некоммутативна. В частности, при $a_1 = a_2 = \ldots = a_n = a$ произведение $aa \ldots a$ обозначают символом a^n и называют n-й степенью элемента. Если множество M обладает единичным элементом, то полагают $a^0 = e$.

Из теоремы 0.1 вытекают соотношения

$$a^m a^n = a^{m+n};\ (a^m)^n = a^{mn},\ m,\ n \in N. \tag{0.1}$$

В аддитивной символике степеням соответствуют кратные $na = a + a + \ldots + a$ и выполняются соотношения

$$ma + na = (m+n)a;\ n(ma) = (nm)a,\ m,\ n \in N.$$

ЭЛЕМЕНТЫ МАТЕМАТИЧЕСКОЙ ЛОГИКИ

ГЛАВА 1

Математическая логика — современный вид формальной логики, т. е. науки, изучающей умозаключения с точки зрения их формального строения. Рассмотрим следующие два умозаключения:

1) «Все люди смертны. Сократ — человек. Следовательно, Сократ смертен»;

2) «Все граждане России имеют право на образование. Иванов — гражданин России. Следовательно, Иванов имеет право на образование».

Легко заметить, что они составлены, в сущности, по одной и той же формальной схеме: «Все M суть P; S есть M. Следовательно, S есть P».

Содержание терминов S, P, M для справедливости этих умозаключений безразлично. Умозаключения, составленные по этой схеме, ученые-схоласты назвали силлогизмами первой фигуры по модусу *Barbara*.

Вплоть до начала XIX в. формальная логика практически не выходила за рамки такого рода силлогических умозаключений. Однако, начиная с работ Дж. Буля, можно говорить о превращении ее в математическую логику. Особенности математической логики заключаются в ее математическом аппарате, в преимущественном внимании к умозаключениям, применяемым в самой математике.

Математическая логика — это обширная наука, которая кроме традиционной проблематики занимается вопросами оснований математики и теории алгоритмов и имеет целый ряд приложений.

1.1. ЛОГИКА ВЫСКАЗЫВАНИЙ

1.1.1. Высказывания. Логические связки. Формулы логики высказываний

Под *высказыванием* принято понимать языковое предложение, о котором имеет смысл говорить, что оно истинно или ложно.

Высказываниями являются, например, следующие предложения: «$2 \times 2 = 4$», «5 — простое число», «Волга впадает в Черное море». Первые два предложения истинны, а третье — ложно. Предложения «Вашингтон — столица США», «Нью-Йорк — столица США», «Царевич Дмитрий был убит по приказу Бориса Годунова» являются высказываниями, причем первое из них истинно, второе — ложно, а третье до сих пор обсуждается историками. Предложения «$x + y = 4$» и «Который час?» высказываниями не являются. Предложения «Город стоит на берегу реки» и «Сегодня хорошая погода» тоже не следует относить к высказываниям ввиду их недостаточной уточненности.

В логике высказываний интересуются не содержанием, а *истинностью* или *ложностью* высказываний (т. е. их *истинностным значением*). Истинностные значения — *истина* и *ложь* — будем обозначать И и Л соответственно. Множество {И, Л} называется *множеством истинностных значений*.

Грамматическими средствами в разговорном языке из нескольких простых высказываний можно составить сложное (составное) высказывание. Например, с помощью союзов «и», «или» и отрицательной частицы «не» можно из простых высказываний «Москва стоит на берегу Невы» (ложного) и «Санкт-Петербург стоит на берегу Невы» (истинного) составить следующие сложные высказывания: «Москва не стоит на берегу Невы», «Москва стоит на берегу Невы или Санкт-Петербург стоит на берегу Невы», «Москва стоит на берегу Невы и Санкт-Петербург стоит на берегу Невы». Первые два высказывания истинны, а последнее — ложно.

Рассмотрим также логические операции (связки) над высказываниями, при которых истинностные значения составных высказываний определяются только истинностными значениями составляющих высказываний, а не их смыслом.

Отрицанием высказываний P называется высказывание, истинное тогда и только тогда, когда высказывание P ложно. Отрицание P обозначается через $\neg P$ и читается как «не P». Отрицание высказывания определяется также таблицей истинности (см. табл. 1.1).

В разговорной речи отрицание соответствует составлению из высказывания P нового высказывания, например: «неверно, что P».

Таблица 1.1

P	$\neg P$
И	Л
Л	И

Таблица 1.2

P	Q	$P\&Q$
И	И	И
И	Л	Л
Л	И	Л
Л	Л	Л

Таблица 1.3

P	Q	$P \vee Q$
И	И	И
И	Л	И
Л	И	И
Л	Л	Л

Конъюнкцией двух высказываний P и Q называется высказывание, истинное тогда и только тогда, когда истинны оба высказывания. Конъюнкция высказываний P и Q обозначается через $P\&Q$ и читается как «P и Q». Конъюнкция определяется также таблицей истинности (см. табл. 1.2).

В разговорной речи конъюнкции соответствует соединение высказываний союзом «и».

Дизъюнкцией двух высказываний P и Q называется высказывание, ложное тогда и только тогда, когда оба высказывания ложны. Дизъюнкция высказываний P и Q обозначается через $P \vee Q$ и читается как «P или Q». Дизъюнкция определяется также таблицей истинности (см. табл. 1.3).

В разговорной речи дизъюнкция соответствует соединению высказываний союзом «или» в неразделительном смысле.

Из двух высказываний P и Q можно составить высказывание «P влечет Q» (или, иначе, «из P следует Q», «если P, то Q»). Не математик может признать утверждение типа «если $2 \times 2 = 5$, то Москва — столица России» ложным, поскольку для него истинность высказываний «P влечет Q» означает, что P по смыслу должно влечь за собой Q. Но тогда связка «влечет» зависит от смысла самих этих высказываний. Однако практика показывает, что можно обороты типа «P влечет Q» и «из P следует Q» использовать таким образом, чтобы под ними каждый раз подразумевалась некоторая операция, не зависящая от смысла высказываний. Рассмотрим следующие высказывания:

1) если $0 = 0$, то $1 = 1$;
2) если $0 = 1$, то $0 = 0$;
3) если $0 = 0$, то $0 = 1$;
4) если $0 = 1$, то $1 = 2$.

Первое утверждение естественно считать истинным, поскольку, используя равенство $0 = 0$, а также другие свойства чисел, можно вывести равенство $1 = 1$ (например, прибавляя 1 к обеим частям равенства $0 = 0$).

Второе утверждение также естественно считать истинным: умножая на 0 обе части равенства $0 = 1$, получаем равенство $0 = 0$.

Третье утверждение приходится считать ложным, ибо, исходя из истинного равенства, мы с помощью умозаключений никогда не придем к ложному.

Четвертое утверждение естественно считать истинным: прибавляя 1 к обеим частям равенства $0 = 1$, получаем равенство $1 = 2$.

Таким образом, используя оборот «если P, то Q» как логическую операцию (связку), определим ее следующим образом.

Импликацией двух высказываний P и Q называется высказывание, ложное тогда и только тогда, когда P истинно, а Q ложно. Импликация высказываний P и Q обозначается через $P \supset Q$ (или $P \Rightarrow Q$) и читается как «P влечет Q» (или, иначе, «если P, то Q», «из P следует Q»). Высказывание P называется *посылкой* импликации, а высказывание Q — *заключением* импликации. Импликация определяется также таблицей истинности (см. табл. 1.4).

Эквиваленцией двух высказываний P и Q называется высказывание, истинное тогда и только тогда, когда истинностные значения P и Q совпадают. Эквиваленция высказываний P и Q обозначается через $P \sim Q$ и читается как «P эквивалентно Q». Эквиваленция определяется также таблицей истинности (см. табл. 1.5).

Таблица 1.4

P	Q	$P \supset Q$
И	И	И
И	Л	Л
Л	И	И
Л	Л	И

Таблица 1.5

P	Q	$P \sim Q$
И	И	И
И	Л	Л
Л	И	Л
Л	Л	И

Определим понятие *формулы логики высказываний*.

Алфавитом называется любое непустое множество. Элементы этого множества называются *символами* данного алфавита. *Словом* в данном алфавите называется произвольная конечная последовательность символов (возможно, пустая). Слово a называется *подсловом* слова b, если $b = b_1ab_2$ для некоторых слов b_1 и b_2.

Алфавит логики высказываний содержит следующие символы: высказывательные переменные $X_1, X_2, X_3, \ldots$; логические символы $\&, \vee, \neg, \supset, \sim$; символы скобок $(,)$.

Слово в алфавите логики высказываний называется *формулой*, если оно удовлетворяет следующему определению:

1) любая высказывательная переменная — формула;

2) если A и B — формулы, то $(\neg A)$, $(A\&B)$, $(A \vee B)$, $(A \supset B)$, $(A \sim B)$ — формулы;

3) только те слова являются формулами, для которых это следует из 1) и 2).

Подформулой формулы A называется любое подслово A, само являющееся формулой.

Для упрощения записи будем в формуле опускать внешние скобки и те пары скобок, без которых можно восстановить эту формулу, если пользоваться следующим правилом: каждое вхождение знака $\neg$ относится к наикратчайшей подформуле, следующей за ним.

Пример 1.1. Слово $(X_1 \& X_2) \supset X_3 \neg X_1$ не является формулой, а слова $(\neg X_1 \supset X_2) \vee X_1$, $(X_1 \sim X_2) \supset \neg X_3$ — формулы. Слова $X_1 \sim X_2$, $\neg X_3$, X_1, X_2, X_3 — подформулы последней формулы.

Принцип математической индукции, который будем использовать в рассуждениях, формулируется следующим образом: если какое-то высказывание $P(t)$, зависящее от натурального параметра t, доказано для $t = 0$ и при произвольном t удается из истинности $P(t)$ обосновать истинность $P(t+1)$, то $P(t)$ истинно для всех t.

Будем применять также и другую формулировку этого принципа: если $P(t)$ истинно при $t = 0$ и для любого t из истинности $P(s)$ при всех $s \leqslant t$ следует истинность $P(t+1)$, то $P(t)$ истинно для всех t.

Применительно к высказывательным формулам принцип математической индукции можно сформулировать следующим образом: если какое-то утверждение $P(F)$, зависящее от параметра F, который пробегает все множество высказывательных формул, истинно для всех формул, не содержащих логических символов (т. е. формул вида X_i), и при любом натуральном n из того, что $P(F)$ истинно для всех формул F с числом логических символов меньше n, следует, что $P(F)$ истинно для всех формул с n логическими символами, то $P(F)$ истинно для всех формул F.

Иногда параметр F будет пробегать не все множество формул, а лишь часть его, скажем, все формулы, построенные с помощью фиксированного набора переменных.

Если каждой высказывательной переменной, входящей в формулу, придавать истинностные значения И и Л, то формула будет определять *истинностную функцию*, т. е. функцию, определенную на множестве {И, Л} со значениями в этом множестве. Истинностная функция может быть представлена таблицей истинности.

Пример 1.2. Формулы $(X_1 \supset X_2) \vee (X_1 \supset (X_2 \& X_1))$ и $(X_1 \supset X_2) \vee \neg X_3$ имеют соответственно таблицы истинности 1.6 и 1.7.

Упорядоченный набор высказывательных переменных $\langle X_{i_1}, \ldots, X_{i_k} \rangle$ назовем *списком* переменных формулы A, если все переменные формулы A содержатся в этом наборе. В списке

переменных формулы A часть переменных может быть фиктивной, т. е. не входить явно в формулу A. *Оценкой* списка переменных назовем сопоставление каждой переменной списка некоторого истинностного значения. При этом, если список переменных формулы A содержит k переменных, то имеется 2^k различных оценок, и, следовательно, таблица истинности этой формулы содержит 2^k строк.

Пример 1.3. Список переменных первой из формул, приведенных в примере 1.2, есть $\langle X_1, X_2\rangle$ или $\langle X_1, X_2, X_3\rangle$, $\langle X_1, X_2, X_4\rangle$ и т. д. (здесь X_3, X_4 — фиктивные переменные).

Для формул, зависящих от фиксированного списка $\langle X_{i_1}, \ldots, X_{i_k}\rangle$, определим индуктивно их *значение на данной оценке списка* (или, иначе, при данном распределении истинностных значений переменных):

1) формула есть одна из переменных $X_{i_1}, \ldots, X_{i_k}$, например, X_{i_1}; тогда значение ее на данной оценке есть то истинностное значение, которое сопоставлено переменной X_{i_1};

2) формула имеет вид $(\neg A)$, а значение A на данной оценке есть s; тогда значение $(\neg A)$ на данной оценке определяется как $\neg s$, где $\neg s$ вычисляется по табл. 1.1;

3) формула имеет вид $(A\&B)$, или $(A\vee B)$ или $(A\supset B)$, или $(A \sim B)$, а значения A, B на данной оценке есть s, t. Тогда значение $(A\&B)$ будет $s\&t$, где $s\&t$ вычисляется по табл. 1.2, а значения остальных формул определяются аналогично.

Ясно, что значение формулы является функцией оценок. Теперь можно сформулировать основные понятия.

1.1.2. Равносильность формул

Пусть A и B — две формулы, зависящие от одного и того же списка переменных $\langle X_{i_1}, \ldots, X_{i_k}\rangle$. Будем называть их *равносильными*, если на любой оценке списка $\langle X_{i_1}, \ldots, X_{i_k}\rangle$ они принимают одинаковые значения.

На первый взгляд, это определение кажется неоднозначным, так как A и B могут зависеть от бесконечного числа списков переменных. Однако эта неоднозначность только кажущаяся.

Таблица 1.6

X_1	X_2	$X_1\supset X_2$	$X_2\&X_1$	$X_1\supset(X_2\&X_1)$	$(X_1\supset X_2)\vee(X_1\supset(X_2\&X_1))$
И	И	И	И	И	И
И	Л	Л	Л	Л	Л
Л	И	И	Л	И	И
Л	Л	И	Л	И	И

Таблица 1.7

X_1	X_2	X_3	$X_1 \supset X_2$	$\neg X_2$	$(X_1 \supset X_2) \vee \neg X_2$
И	И	И	И	Л	И
И	И	Л	И	И	И
И	Л	И	Л	Л	Л
И	Л	Л	Л	И	И
Л	И	И	И	Л	И
Л	И	Л	И	И	И
Л	Л	И	И	Л	И
Л	Л	Л	И	И	И

Любой из возможных списков должен содержать все переменные A и B, причем значения A и B на любых оценках списка будут зависеть только от значений переменных, входящих в формулы A и B.

Равносильность формул имеет свой алгебраический аналог в тождественном равенстве алгебраических выражений. Равносильность формул A и B будем обозначать через $A \rightleftharpoons B$ (так же, как в алгебре; например, $a(b + c) = ab + ac$).

Нужно различать символы $\sim$ и $=$. Так, $\sim$ является символом формального языка, с помощью которого строятся формулы, а символ $=$ заменяет слово «равносильно».

Отношение равносильности есть отношение эквивалентности. Действительно, оно рефлексивно, так как для любой формулы A $A \rightleftharpoons A$; симметрично, так как для любых формул A и B, если $A = B$, то $B = A$; транзитивно, так как для любых формул A, B, C, если $A = B$ и $B = C$, то $A = C$.

Поскольку в определении равносильности формул A, B безразлично, какой список переменных берется (лишь бы он включал в себя переменные A и B), то в качестве $\langle X_{i_1}, \ldots, X_{i_k} \rangle$ возьмем список всех переменных формул A, B, C. На любой оценке этого списка формулы A, B, а равно и B, C принимают одни и те же значения. Следовательно, и формулы A, C на всех оценках принимают одинаковые значения.

Основные равносильности. *Для любых формул A, B, C справедливы следующие равносильности:*

1) $A \& B = B \& A$ *(коммутативность &);*
2) $A \& A = A$ *(идемпотентность &);*
3) $A \& (B \& C) = (A \& B) \& C$ *(ассоциативность &);*
4) $A \vee B = B \vee A$ *(коммутативность $\vee$);*
5) $A \vee A = A$ *(идемпотентность $\vee$);*
6) $A \vee (B \vee C) = (A \vee B) \vee C$ *(ассоциативность $\vee$);*
7) $A \vee (B \& C) = (A \vee B) \& (A \vee C)$ *(дистрибутивность $\vee$ относительно &);*
8) $A \& (B \vee C) = (A \& B) \vee (A \& C)$ *(дистрибутивность & относительно $\vee$);*
9) $A \& (A \vee B) = A$ *(первый закон поглощения);*

10) $A \vee (A \& B) \equiv A$ *(второй закон поглощения)*;
11) $\neg\neg A \equiv A$ *(снятие двойного отрицания)*;
12) $\neg(A \& B) \equiv \neg A \vee \neg B$ *(первый закон де Моргана)*;
13) $\neg(A \vee B) \equiv \neg A \& \neg B$ *(второй закон де Моргана)*;
14) $A \equiv (A \& B) \vee (A \& \neg B)$ *(первая формула расщепления)*;
15) $A \equiv (A \vee B) \& (A \vee \neg B)$ *(вторая формула расщепления)*.

Любая из этих равносильностей легко может быть доказана с помощью таблиц истинности. Рассмотрим, например, равносильность 7. Пусть $\langle X_{i_1}, \ldots, X_{i_k} \rangle$ — какой-либо список переменных, от которого зависят *A*, *B*, *C*. Тогда для значений *A*, *B*, *C* на какой-нибудь оценке списка переменных имеется восемь вариантов. Для каждого варианта по таблицам истинности нетрудно подсчитать значения левой и правой частей равносильности 7 и убедиться в том, что в любом из восьми случаев эти значения совпадают (см. табл. 1.8).

Таблица 18

A	*B*	*C*	*B&C*	$A \vee (B\&C)$	$A \vee B$	$A \vee C$	$(A \vee B)\&(A \vee C)$
И	И	И	И	И	И	И	И
И	И	Л	Л	И	И	И	И
И	Л	И	Л	И	И	И	И
И	Л	Л	Л	И	И	И	И
Л	И	И	И	И	И	И	И
Л	И	Л	Л	Л	И	Л	Л
Л	Л	И	Л	Л	Л	И	Л
Л	Л	Л	Л	Л	Л	Л	Л

Однако часто равносильность экономнее доказывать без составления полной таблицы, а лишь с помощью некоторого рассуждения. Рассмотрим, например, первый закон де Моргана (равносильность 12). Докажем, что если на какой-то оценке списка переменных, от которого зависят *A* и *B*, левая часть равносильности 12 получает значение Л, то и правая ее часть получит значение Л, и, наоборот, если на какой-то оценке списка переменных правая часть равносильности принимает значение Л, то и левая часть примет значение Л. Это и будет означать равносильность.

Итак, пусть на некоторой оценке списка переменных формула $\neg(A \& B)$ принимает значение Л. Тогда формула $A \& B$ принимает значение И, а поэтому обе формулы *A*, *B* принимают значение И. Но в этом случае, очевидно, и правая часть равносильности 12 примет значение Л. И наоборот, пусть формула $\neg A \vee \neg B$ принимает значение Л. Тогда формулы $\neg A$, $\neg B$ принимают значение Л, а формулы *A*, *B* — значение И. Очевидно, что и левая часть равносильности 12 примет значение Л.

Следующая группа равносильностей показывает, что одни связки могут быть выражены через другие:

16) $A \sim B \equiv (A \supset B) \,\&\, (B \supset A) \equiv (A \,\&\, B) \vee (\neg A \,\&\, \neg B)$;

17) $A \supset B \equiv \neg A \vee B \equiv \neg (A \,\&\, \neg B)$;

18) $A \vee B \equiv \neg A \supset B \equiv \neg \neg A \,\&\, \neg B)$;

19) $A \,\&\, B \equiv \neg (A \supset \neg B) \equiv (\neg (\neg A \vee \neg B)$.

В силу транзитивности отношения равносильности, если $A_1 \equiv A_2$, $A_2 \equiv A_3, \ldots, A_{k-1} \equiv A_k$, то $A_1 \equiv A_k$. В таком случае для простоты будем записывать цепочку $A_1 \equiv A_2 \equiv \ldots \equiv A_{k-1} \equiv A_k$.

Приведем правило, с помощью которого можно переходить от одних равносильностей к другим.

Лемма 1.1. *Пусть $A \equiv B$ и C — произвольная формула.* Тогда $\neg A \equiv \neg B$, $A \,\&\, C \equiv B \,\&\, C$, $C \,\&\, A \equiv C \,\&\, B$, $A \vee C \equiv B \vee C$, $C \vee A \equiv C \vee B$, $A \supset C \equiv B \supset C$, $C \supset A \equiv C \supset B$. $A \sim C \equiv B \sim C$, $C \sim A \equiv C \sim B$.

Докажем, например, равносильность $A \supset C \equiv B \supset C$. На произвольной оценке списка переменных, от которого зависят A, B, C, формулы A и B принимают одинаковое значение (скажем, s). Пусть t — значение C на этой оценке. Обе части рассматриваемой равносильности принимают одно и то же значение $s \supset t$.

Лемма 1.2. *Пусть $A \equiv B$ и C — формула, в которой выделено одно вхождение переменной X_i. Пусть C_A получается из C заменой этого вхождения X_i на A, а C_B — из C заменой того же вхождения X_i на B. Тогда $C_A \equiv C_B$.*

Докажем это индукцией по числу n логических символов C. Если $n = 0$, то формула C должна совпадать с X_i (так как в ней имеется вхождение переменной X_i). В этом случае C_A есть A, C_B есть B, а $C_A \equiv C_B$ — не что иное, как $A \equiv B$.

Пусть лемма доказана для числа логических символов меньше n и пусть C — формула с n логическими символами. Она имеет вид $\neg D$, или $D \,\&\, E$, или $D \vee E$, или $D \supset E$, или $D \sim E$, причем в первом случае выделенное вхождение X_i содержится в D, а в остальных случаях — либо в D, либо в E, но не в D и E сразу. Рассмотрим, например, случай, когда C имеет вид $D \supset E$, а выделенное вхождение X_i содержится в D. Заменяя X_i в этом вхождении в D на A и B, получаем соответственно формулы D_A и D_B. Ясно, что C_A есть $D_A \supset E$, а C_B есть $D_B \supset E$. Так как в формуле D меньше логических символов чем в C, то $D_A \equiv D_B$. Применим теперь лемму 1.1 в случае $A \supset C \equiv B \supset C$, где в роли A выступает D_A, в роли B — D_B, в роли C — E. В результате получаем $C_A \equiv C_B$. Другие случаи рассматриваются аналогично.

Утверждение 1.1 *(правило равносильных преобразований). Пусть C_A — формула, содержащая A в качестве своей подформулы. Пусть C_B получается из C_A заменой A в этом вхождении на B. Тогда, если $A \equiv B$, то $C_A \equiv C_B$.*

Рассмотрим произвольную переменную X_i и получим формулу C из C_A заменой A на X_i. Будем считать это вхождение X_i в C выделенным. Тогда C, A, B, C_A, C_B удовлетворяют условиям леммы 1.2, а значит, $C_A = C_B$.

Заметим, что алгебраический аналог этого правила достаточно очевиден (впрочем, как и само правило) и обычно применяется без особого обоснования. Например, пользуясь тождеством $x + x = 2x$, получают $y(x + x) = y(2x)$.

Утверждение 1.2 *(правило устранения логических символов $\supset$ и $\sim$). Для каждой формулы можно указать равносильную ей формулу, не содержащую логических символов $\sim$ и $\supset$.*

В самом деле, опираясь на правило равносильных преобразований, можно в исходной формуле каждую подформулу вида $A \sim B$ заменить на $(A \& B) \vee (\neg A \& \neg B)$, а каждую подформулу вида $A \supset B$ — на $\neg A \vee B$ (см. равносильности 16 и 17).

Можно дать и более строгое доказательство, применив индукцию по числу логических символов.

Пример 1.4. Формула $(X_1 \supset (X_2 \supset X_3)) \sim \neg (X_2 \supset X_1)$ преобразуется следующим образом: $(X_1 \supset (X_2 \supset X_3)) \sim \neg (X_2 \supset X_1) \equiv (X_1 \supset (\neg X_2 \vee X_3)) \sim (\neg(\neg X_2 \vee X_1)) \equiv (\neg X_1 \vee (\neg X_2 \vee X_3)) \sim \neg(\neg X_2 \vee X_1) \equiv ((\neg X_1 \vee (\neg X_2 \vee X_3)) \& \neg(\neg X_2 \vee X_1)) \vee (\neg(\neg X_1 \vee (\neg X_2 \vee X_3)) \& \neg \neg(\neg X_2 \vee X_1))$.

1.1.3. Тождественно-истинные формулы. Правильные рассуждения

Пусть формула A зависит от списка переменных $<X_{i_1}, \ldots, X_{i_k}>$.

Формула A называется *тавтологией* (или *тождественно-истинной* формулой), если на любых оценках списка переменных $<X_{i_1}, \ldots, X_{i_k}>$ она принимает значение И.

Формула A называется *выполнимой*, если на некоторой оценке списка переменных $<X_{i_1}, \ldots, X_{i_k}>$ она принимает значение И.

Формула A называется *тождественно-ложной*, если на любых оценках списка переменных $<X_{i_1}, \ldots, X_{i_k}>$ она принимает значение Л.

Формула A называется *опровержимой*, если на некоторой оценке списка переменных $<X_{i_1}, \ldots, X_{i_k}>$ она принимает значение Л.

Как и в определении равносильности, здесь не имеет значения, будут ли в списке фиктивные переменные.

Приведем утверждения, которые являются очевидными следствиями данных определений:

1) A — тавтология тогда и только тогда, когда A не является опровержимой;

2) A тождественно-ложна тогда и только тогда, когда A не является выполнимой;

3) A — тавтология тогда и только тогда, когда $\neg A$ тождественно-ложна;

4) A тождественно-ложна тогда и только тогда, когда $\neg A$ — тавтология;

5) $A \sim B$ — тавтология тогда и только тогда, когда A и B равносильны.

С точки зрения логики тавтологии суть не что иное, как логические законы, ибо при любой подстановке вместо переменных тавтологии конкретных высказываний в результате получим истинные высказывания. Перечислим наиболее важные тавтологии (A, B, C — произвольные формулы):

1) $A \vee \neg A$ *(закон исключенного третьего или tertium non-datur);*

2) $A \supset A$;

3) $A \supset (B \supset A)$;

4) $(A \supset B) \supset ((B \supset C) \supset (A \supset C))$ *(цепное рассуждение);*

5) $(A \supset (B \supset C)) \supset ((A \supset B) \supset (A \supset C))$;

6) $(A \,\&\, B) \supset A$, $(A \,\&\, B) \supset B$;

7) $A \supset (B \supset (A \,\&\, B))$;

8) $A \supset (A \vee B)$, $B \supset (A \vee B)$;

9) $(\neg B \supset \neg A) \supset ((\neg B \supset A) \supset B)$;

10) $((A \supset B) \supset A) \supset A$ *(закон Пирса).*

Каждую из этих тавтологий можно обосновать, например, составив таблицу и вычислив по ней значение формулы при произвольных значениях A, B, C.

При доказательстве утверждений различных математических теорий обычно используют рассуждения, которые на языке логики можно выразить формулами.

Рассуждение называется *правильным*, если из конъюнкции посылок следует заключение, т. е. всякий раз, когда все посылки истинны, заключение тоже истинно.

Пусть $P_1, \ldots, P_n$ — посылки, D — заключение. Тогда для определения правильности рассуждения по схеме $\dfrac{P_1, \ldots, P_n}{D}$, т. е. утверждения о том, что из данных посылок $P_1, \ldots, P_n$ следует заключение D, требуется установить тождественную истинность формулы $(P_1 \,\&\, \ldots \,\&\, P_n) \supset D$ *.

Так как речь идет лишь о правильности рассуждения, истинность заключения не является ни необходимым, ни достаточным условием правильности рассуждения.

---

* Вообще говоря, выражение $(P_1 \,\&\, \ldots \,\&\, P_n) \supset D$ не удовлетворяет определению формулы. Однако далее (см. разд. 1 1.5) мы убедимся, что такая запись корректна.

Пример 1.5. Рассмотрим два рассуждения:

1. Если число 5 простое, то оно нечетное. Число 5 нечетное. Следовательно, число 5 простое. Заключение истинно, но рассуждение неправильное. Это рассуждение по схеме $\frac{A \supset B,\ B}{A}$. Легко проверить, что формула $((A \supset B)\ \&\ B) \supset A$ не является тождественно-истинной.

2. Если Петр занимается спортом, то Петр никогда не болеет. Петр занимается спортом. Следовательно, Петр никогда не болеет. Это рассуждение по схеме $\frac{A \supset B,\ A}{B}$. Формула $((A \supset B)\ \&\ A) \supset B$ тождественно-истинна, и, значит, рассуждение правильное.

Распространенными схемами правильных рассуждений являются следующие схемы: $\frac{A \supset B,\ A}{B}$ и $\frac{A \supset B,\ \neg B}{\neg A}$.

Рассмотрим условное высказывание вида $A \supset B$, где A — конъюнкция посылок, B — заключение. Иногда удобнее вместо доказательства истинности этого условного высказывания установить логическую истинность некоторого другого высказывания, равносильного исходному. Такие формы доказательства называются *косвенными методами доказательства.*

Одним из них является способ доказательства от противного. Предположим, что утверждение $A \supset B$ ложно. Тогда, исходя из этого предположения, приходим к противоречию, т. е. доказываем, что некоторое утверждение (соответствующее высказыванию С) выполняется и не выполняется (одновременно). Применимость этой формы косвенного метода доказательства оправдывается равносильностью

$$A \supset B \equiv \neg (A \supset B) \supset (C\ \&\ \neg C) \equiv (A\ \&\ \neg B) \supset (C\ \&\ \neg C).$$

Существуют и другие схемы доказательства от противного:

$$A \supset B \equiv (A\ \&\ \neg B) \supset \neg A,$$
$$A \supset B \equiv (A\ \&\ \neg B) \supset B.$$

Еще одной формой косвенного метода доказательства является доказательство по закону контрапозиции, основанное на равносильности

$$A \supset B \equiv \neg B \supset \neg A,$$

когда вместо истинности $A \supset B$ мы доказываем истинность $\neg B \supset \neg A$.

1.1.4. Двойственность. Закон двойственности

Будем рассматривать формулы, содержащие только логические символы &, $\vee$, $\neg$.

Символы &, $\vee$ называются *двойственными друг другу*. Формула A^* называется *двойственной* формуле A, если она получена из A одновременной заменой всех символов &, $\vee$ на двойственные. Например, формула $X_1 \& (X_2 \vee \neg X_1)$ двойственна формуле $X_1 \vee (X_2 \& \neg X_1)$. Очевидно, что $(A^*)^*$ совпадает с A.

Пусть $\langle X_{i_1}, \ldots, X_{i_k} \rangle$ — некоторый список переменных, а $\langle s_1, \ldots, s_k \rangle$ — его оценка. Назовем оценку $\langle t_1, \ldots, t_k \rangle$ *двойственной* оценке $\langle s_1, \ldots, s_k \rangle$, если $\langle t_1, \ldots, t_k \rangle$ получается из $\langle s_1, \ldots, s_k \rangle$ заменой всех И на Л и всех Л на И.

Лемма 1.3. *Пусть A — формула, а $\langle X_{i_1}, \ldots, X_{i_k} \rangle$ — список переменных, от которого она зависит. Тогда A принимает значение* И *на оценке $\langle s_1, \ldots, s_k \rangle$ в том и только в том случае, если A^* принимает значение* Л *на оценке $\langle t_1, \ldots, t_k \rangle$, двойственной оценке $\langle s_1, \ldots, s_k \rangle$.*

Докажем лемму 1.3 для всех формул A, зависящих от списка переменных $\langle X_{i_1}, \ldots, X_{i_k} \rangle$, индукцией по числу n логических символов A.

Если $n = 0$, то A совпадает с одной из переменных X_{i_l} $(1 \leqslant l \leqslant k)$. В этом случае A^* тоже совпадает с X_{i_l}. То, что X_{i_l} принимает значение И на оценке $\langle s_1, \ldots, s_k \rangle$, означает, что s_l есть И. Но это равносильно тому, что t_l есть Л, а также тому, что X_{i_l} принимает значение Л на оценке $\langle t_1, \ldots, t_k \rangle$.

Пусть утверждение леммы 1.3 справедливо при числе логических символов, меньшем n. Докажем, что оно остается справедливым и при числе символов, равном n. Тогда формула A должна иметь вид $\neg B$, $B \& C$ или $B \vee C$. В соответствии с этим утверждением различают три случая:

1. Формула A есть $\neg B$. Тогда, очевидно, A^* совпадает с $\neg (B^*)$. Пусть A истинна на оценке $\langle s_1, \ldots, s_k \rangle$. Тогда B будет ложной на ней. В формуле B^* число логических символов меньше n. Так как $(B^*)^*$ есть B, то из ложности $(B^*)^*$ на оценке $\langle s_1, \ldots, s_k \rangle$ следует истинность B^* на оценке $\langle t_1, \ldots, t_k \rangle$ (ибо $\langle s_1, \ldots, s_k \rangle$ и $\langle t_1, \ldots, t_k \rangle$ взаимно двойственны). Отсюда следует, что формула A^* (или $\neg (B^*)$) ложна на $\langle t_1, \ldots, t_k \rangle$. Аналогично из ложности A^* на $\langle t_1, \ldots, t_k \rangle$ выводится истинность A на $\langle s_1, \ldots, s_k \rangle$.

2. Формула A есть $B \& C$. В этом случае A^* есть $B^* \vee C^*$. Пусть A принимает значение И на оценке $\langle s_1, \ldots, s_k \rangle$. Тогда и B, C будут истинны на $\langle s_1, \ldots, s_k \rangle$. Так как в формулах B, C число логических символов меньше n, то B^*, C^* примут значение Л на $\langle t_1, \ldots, t_k \rangle$, а значит, и $B^* \vee C^*$ принимает значение Л на этой оценке. И наоборот, если $B^* \vee C^*$ принимает значение Л на $\langle t_1, \ldots, t_k \rangle$, то формулы B^*, C^* будут ложны на $\langle t_1, \ldots, t_k \rangle$, и, далее, в силу индуктивного предположения B, C будут истинны на $\langle s_1, \ldots, s_k \rangle$, а значит, и $B \& C$ будет истинна на этой оценке.

3. Формула A есть $B \vee C$. В этом случае A^* есть $B^* \& C^*$. Пусть A принимает значение И на $\langle s_1, \ldots, s_k \rangle$. Тогда либо B, либо C будет истинной на $\langle s_1, \ldots, s_k \rangle$. Если это, например, будет B, то, поскольку в B логических символов меньше, чем n, B^* будет ложной на $\langle t_1, \ldots, t_k \rangle$. Но тогда и $B^* \& C^*$ будет ложной на $\langle t_1, \ldots, t_k \rangle$. Аналогично рассуждаем и в случае истинности C. Доказательство обратного утверждения также несложно.

Теорема 1.1 (*принцип двойственности*). *Если* $A \equiv B$, *то* $A^* \equiv B^*$.

Пусть $A \equiv B$. Если $\langle X_{i_1}, \ldots, X_{i_k} \rangle$ — список переменных, от которого зависят формулы A, B (и, очевидно, A^*, B^*), а $\langle t_1, \ldots, t_k \rangle$ — оценка этого списка, то A^* принимает значение И на этой оценке в том и только в том случае, если $(A^*)^*$ (т. е. A) принимает значение Л на оценке $\langle s_1, \ldots, s_k \rangle$, двойственной оценке $\langle t_1, \ldots, t_k \rangle$. Но последнее в силу равносильности A и B имеет место тогда и только тогда, когда B принимает значение Л на оценке $\langle s_1, \ldots, s_k \rangle$. С другой стороны, B^* принимает значение И на оценке $\langle t_1, \ldots, t_k \rangle$ только тогда, когда $(B^*)^*$ (т. е. B) принимает значение Л на $\langle s_1, \ldots, s_k \rangle$. Итак, видим, что истинность A^* на $\langle t_1, \ldots, t_k \rangle$ равносильна истинности B^* на $\langle t_1, \ldots, t_k \rangle$. Поскольку оценка $\langle t_1, \ldots, t_k \rangle$ произвольна, получаем, что $A^* \equiv B^*$.

Принцип двойственности можно применить для нахождения новых равносильностей. Например, используя следующий частный случай дистрибутивности $\&$ относительно $\vee$

$$X_i \& (X_j \vee X_k) \equiv (X_i \& X_j) \vee (X_i \& X_k),$$

получаем равносильность

$$X_i \vee (X_j \& X_k) \equiv (X_i \vee X_j) \& (X_i \vee X_k).$$

Другие приложения принципа двойственности будут приведены ниже.

1.1.5. Нормальные формы формул

Заметим, что в силу ассоциативности операций $\&$ и $\vee$ (см. теорему 0.1), как бы мы не расставляли скобки в выражениях $A_1 \& A_2 \& \ldots \& A_k$, $A_1 \vee A_2 \vee \ldots \vee A_k$ $(k > 3)$, всегда будем приходить к равносильным формулам. Допуская некоторую вольность, каждое из этих выражений будем считать формулой и называть соответственно *многочленной конъюнкцией* и *дизъюнкцией* формул $A_1, \ldots, A_k$. Для этих формул, используя, например, индукцию по паре (k, l), нетрудно получить равносильности, выражающие обобщенную дистрибутивность:

$$(A_1 \& A_2 \& \ldots \& A_k) \vee (B_1 \& B_2 \& \ldots \& B_l) \equiv (A_1 \vee B_1) \&$$

$$\&\,(A_1 \vee B_2)\,\&\,\ldots\&\,(A_1 \vee B_l)\,\&\,(A_2 \vee B_1)\,\&\,(A_2 \vee B_2)\,\&\,\ldots$$
$$\ldots\,\&\,(A_2 \vee B_l)\,\&\,\ldots\,\&\,(A_k \vee B_1)\,\&\,(A_k \vee B_2)\,\&\,\ldots\,\&\,(A_k \vee B_l);$$
$$(A_1 \vee A_2 \vee \ldots \vee A_k)\,\&\,(B_1 \vee B_2 \vee \ldots \vee B_l) \equiv (A_1 \& B_1) \vee$$
$$\vee\,(A_1 \& B_2) \vee \ldots \vee (A_1 \& B_l) \vee (A_2 \& B_1) \vee (A_2 \& B_2) \vee \ldots$$
$$\ldots \vee (A_2 \& B_l) \vee \ldots \vee (A_k \& B_1) \vee (A_k \& B_2) \vee \ldots \vee (A_k \& B_l),$$

а также обобщенные законы де Моргана:

$$\neg\,(A_1 \& \ldots \& A_k) \equiv \neg A_1 \vee \ldots \vee \neg A_k;\ \neg\,(A_1 \vee \ldots \vee A_k) \equiv$$
$$\equiv \neg A_1 \& \ldots \& \neg A_k.$$

Определим теперь некоторые канонические виды формул.

Формулу называют *элементарной конъюнкцией*, если она является конъюнкцией (быть может, одночленной) переменных и отрицаний переменных. Например, формулы $X_2, \neg X_3, X_2 \& \neg X_3, X_2 \& X_3, X_1 \& \neg X_2 \& X_4 \& X_2$ являются элементарными конъюнкциями.

Говорят, что формула находится в *дизъюнктивной нормальной форме* (ДНФ), если она является дизъюнкцией (быть может, одночленной) элементарных конъюнкций. Например, формулы $X_2, \neg X_3, X_1 \& X_2 \& X_3, (X_1 \& X_2 \& \neg X_3) \vee (X_1 \& \neg X_2 \& X_3)$ находятся в ДНФ.

Теорема 1.2 *(о приведении к ДНФ). Для любой формулы A можно найти такую формулу B, находящуюся в ДНФ, что $A = B$.*

Формула B называется *дизъюнктивной нормальной формой формулы A*.

Доказательство теоремы распадается на три этапа:

1-й этап. Для формулы A строим такую формулу A_1, что $A \equiv A_1$ и в A_1 не содержится символов $\sim$, $\supset$ (см. утверждение 1.2)

2-й этап. Докажем теперь, что для формулы A_1 можно найти формулу A_2 такую, что $A_1 \equiv A_2$ и в A_2 отрицание находится только перед переменными. Такая формула называется формулой с «тесными» отрицаниями. Докажем это утверждение индукцией по числу n логических символов формулы A_1.

Если $n = 0$, то A_1 есть какая-то переменная X_l. В качестве A_2 нужно взять X_l.

Пусть утверждение выполняется для всех формул A_1 с числом символов меньше n. Пусть в формуле A_1 содержится точно n логических символов. Рассмотрим следующие случаи:

1) A_1 имеет вид $B_1 \vee C_1$. Тогда в B_1, C_1 логических символов меньше, чем n. Поэтому существуют формулы B_2, C_2 такие, что $B_1 \equiv B_2$, $C_1 \equiv C_2$ и в B_2, C_2 отрицание встречается только перед переменными. Ясно, что $B_2 \vee C_2$ равносильна A_1 и является формулой с «тесными» отрицаниями;

2) A_1 имеет вид $B_1 \& C_1$. Доказательство аналогично предыдущему случаю;

3) A_1 имеет вид $\neg\neg B_1$. Тогда $A_1 \equiv B_1$ и в B_1 логических символов меньше, чем n. Поэтому к B_1 применимо индуктивное предположение;

4) A_1 имеет вид $\neg(B_1 \vee C_1)$. Тогда $A_1 \equiv \neg B_1 \& \neg C_1$ и в $\neg B_1$, $\neg C_1$ логических символов меньше, чем n. Поэтому существуют такие формулы B_2, C_2, что $\neg B_1 \equiv B_2$, $\neg C_1 \equiv C_2$ и в B_2, C_2 отрицание встречается только перед переменными. Ясно, что $A_1 \equiv B_2 \& C_2$ и $B_2 \& C_2$ является формулой с «тесными» отрицаниями;

5) A_1 имеет вид $\neg(B_1 \& C_1)$. Тогда $A_1 \equiv \neg B_1 \vee \neg C_1$, и далее поступаем, как в предыдущем случае.

При практическом преобразовании встречающиеся в формуле отрицания просто передвигают к переменным, используя законы де Моргана и уничтожая пары стоящих рядом отрицаний, если таковые встречаются.

Пример 1.6. Преобразуем к формуле с «тесными» отрицаниями:

$$\neg(\neg\neg(X_1 \& \neg X_2) \vee (X_2 \& \neg X_1)) \equiv \neg\neg\neg(X_1 \& \neg X_2) \& \neg(X_2 \& \neg X_1) \equiv \neg(X_1 \& \neg X_2) \& (\neg X_2 \vee \neg\neg X_1) \equiv (\neg X_1 \vee \neg\neg X_2) \& (\neg X_2 \vee X_1) \equiv (\neg X_1 \vee X_2) \& (\neg X_2 \vee X_1).$$

3-й этап. Полученную формулу A_7 можно считать построенной из переменных и их отрицаний с помощью многочленных конъюнкций и дизъюнкций. Применив теперь обобщенную дистрибутивность & относительно $\vee$, последовательно преобразуем формулу (аналогично приведению алгебраического выражения, составленного из переменных, с помощью сложений и умножений к виду многочлена). Заметим, что при этом $\vee$ будет аналогична сложению, а & — умножению. Полученная в результате преобразований формула B будет удовлетворять требованиям доказываемой теоремы.

Пример 1.7. Применим преобразования 3-го этапа к формуле с «тесными» отрицаниями, полученной в примере 1.6: $(\neg X_1 \vee X_2) \& (\neg X_2 \vee X_1) \equiv (\neg X_1 \& \neg X_2) \vee (\neg X_1 \& X_1) \vee (X_2 \& \neg X_2) \vee (X_2 \& X_1)$. В результате мы получили формулу, находящуюся в ДНФ.

Говорят, что формула A находится в *конъюнктивной нормальной форме (КНФ)*, если формула A^* определена (т. е. в A нет символов $\sim$ и $\supset$) и находится в ДНФ.

КНФ можно дать и другое равносильное определение. Формулу называют *элементарной дизъюнкцией*, если она является дизъюнкцией (возможно, одночленной) переменных и отрицаний переменных. Формула находится в КНФ, если она является конъюнкцией (возможно, одночленной) элементарных дизъюнкций.

Теорема 1.3 *(о приведении к КНФ). Для любой формулы*

А можно найти такую формулу В, что А находится в КНФ и $A \equiv B$.

Формула B называется *конъюнктивной нормальной формой формулы* A.

Первое доказательство. Пусть $A \equiv A_1$ и A_1 не содержит символов $\sim$, $\supset$. Пусть B_1 — дизъюнктивная нормальная форма для формулы A^*_1. Тогда B^*_1 находится в КНФ и, кроме того, по принципу двойственности $B^*_1 \equiv (A^*_1)^* \equiv A_1 \equiv A$. Значит, B^*_1 удовлетворяет требованиям теоремы.

Второе доказательство. Применив первые два этапа из доказательства теоремы 1.2 о ДНФ, получим формулу A_2, равносильную A, не содержащую символов $\sim$, $\supset$ и содержащую отрицания только перед переменными. Преобразуем теперь A_2 как алгебраическое выражение, считая на этот раз & аналогом сложения, а $\vee$ — аналогом умножения и применяя дистрибутивность $\vee$ относительно &. Приведение формулы A_2 к виду многочлена дает на этот раз КНФ.

Пример 1.8. Приведем к КНФ формулу

$(X_1 \& X_2) \sim (\neg X_1 \& X_3) \equiv ((X_1 \& X_2) \& (\neg X_1 \& X_3)) \vee$
$\vee (\neg (X_1 \& X_2) \& \neg (\neg X_1 \& X_3)) \equiv (X_1 \& X_2 \& \neg X_1 \& X_3) \vee$
$\vee ((\neg X_1 \vee \neg X_2) \& (\neg \neg X_1 \vee \neg X_3)) \equiv (X_1 \& X_2 \& \neg X_1 \& X_3) \vee$
$\vee ((\neg X_1 \vee \neg X_2) \& (X_1 \vee \neg X_3)) \equiv (X_1 \vee \neg X_1 \vee \neg X_2) \&$
$\& (X_1 \vee X_1 \vee \neg X_3) \& (X_2 \vee \neg X_1 \vee \neg X_2) \& (X_2 \vee X_1 \vee \neg X_3) \&$
$\& (\neg X_1 \vee \neg X_1 \vee \neg X_2) \& (\neg X_1 \vee X_1 \vee \neg X_3) \& (X_3 \vee \neg X_1 \vee$
$\vee \neg X_2) \& (X_3 \vee X_1 \vee \neg X_3)$.

Заметим, что первое преобразование основано на равносильности 16.

Нетрудно видеть, что ДНФ не является однозначно определенной. Рассмотрим, например, формулу $X_1 \vee (X_2 \& X_3)$. Она уже находится в ДНФ. В то же время преобразование $X_1 \vee$ $\vee (X_2 \& X_3) \equiv (X_1 \vee X_2) \& (X_1 \vee X_3) \equiv (X_1 \& X_1) \vee (X_1 \& X_3) \vee$ $\vee (X_2 \& X_1) \vee (X_2 \& X_3)$ дает для этой формулы другую ДНФ. Конечно, все ДНФ данной формулы равносильны. Выделим среди ДНФ формулы так называемую совершенную дизъюнктивную нормальную форму.

Пусть формула A зависит от списка переменных $\langle X_{i_1}, \ldots, X_{i_k} \rangle$. **Говорят, что** A находится в *совершенной дизъюнктивной нормальной форме* (СДНФ) относительно этого списка, если выполняются следующие условия:

1) A находится в ДНФ;

2) каждый дизъюнктивный член A является k-членной конъюнкцией, причем на l-м месте $(1 \leqslant l \leqslant k)$ этой конъюнкции обязательно стоит либо переменная X_{i_l}, либо ее отрицание $\neg X_{i_l}$;

3) все дизъюнктивные члены A попарно различны.

Пример 1.9. Пусть $\langle X_1, X_2, X_3 \rangle$ — список переменных. Тогда формулы $X_1 \& \neg X_2 \& \neg X_3$, $(X_1 \& X_2 \& X_3) \vee (\neg X_1 \& X_2 \&$

$\& X_3) \vee (X_1 \& X_2 \& \neg X_3)$ находятся в СДНФ относительно этого списка переменных.

Теорема 1.4. *Пусть формула A зависит от списка переменных $<X_{i_1}, \ldots, X_{i_k}>$ и A не тождественно-ложная формула. Тогда существует такая формула B, что $A \equiv B$ и B находится в СДНФ относительно списка переменных $<X_{i_1}, \ldots, X_{i_k}>$.*

Согласно теореме о приведении к ДНФ существует формула A_1 такая, что $A \equiv A_1$ и A_1 находится в ДНФ. При этом можно считать, что A_1 зависит от списка переменных $<X_1, \ldots, X_{i_k}>$ (в процессе приведения формулы к ДНФ новые переменные не появляются). Будем исходить из этой формулы и просматривать ее дизъюнктивные члены, т. е. элементарные конъюнкции:

1. Пусть в элементарную конъюнкцию одновременно входят какая-нибудь переменная X_{i_1} и ее отрицание. Если это единственная элементарная конъюнкция формулы, то она на всех оценках принимает значение Л, а следовательно, и вся формула, что невозможно, так как предполагается, что формула не тождественно-ложная.

Значит, имеются другие элементарные конъюнкции, и формула (после некоторых перестановок) будет иметь вид

$$(X_{i_1} \& \neg X_{i_1} \& C) \vee D,$$

где C — остальные члены нашей элементарной конъюнкции; D — остальные дизъюнктивные члены всей формулы. Но $(X_{i_1} \& \& \neg X_{i_1} \& C) \vee D \equiv D$, поскольку первый дизъюнктивный член левой части при всех оценках принимает значение Л. Следовательно, наша формула равносильна D, а рассматриваемую элементарную конъюнкцию можно отбросить. Поскольку Л не тождественно-ложная, то после всех таких шагов всегда останутся какие-то неотброшенные элементарные конъюнкции.

2. Пусть в некоторой элементарной конъюнкции переменная X_{i_1} (или $\neg X_{i_1}$) встречается несколько раз. Тогда в силу идемпотентности (равносильности 2) можно оставить только одно вхождение X_{i_1} (или $\neg X_{i_1}$).

3. После проведенной обработки каждая элементарная конъюнкция C будет содержать какую-нибудь переменную не более одного раза (включая ее вхождение под знаком отрицания); при этом возможны только следующие варианты:

а) конъюнкция C содержит в качестве своего конъюнктивного члена один раз X_{i_1} и не содержит ни разу $\neg X_{i_1}$;

б) конъюнкция C содержит один раз $\neg X_{i_1}$ и не содержит ни разу X_{i_1};

в) конъюнкция C не содержит ни X_{i_1}, ни $\neg X_{i_1}$.

В последнем случае мы заменяем C на $(C \& X_{i_1}) \vee (C \& \neg X_{i_1})$ по основной равносильности 14 (первой формуле ра

щепления). Эту операцию следует проводить до тех пор, пока для каждой элементарной конъюнкции и каждой переменной X_{i_j} не будут выполнены условия «а» или «б».

4. Переупорядочим в каждой элементарной конъюнкции ее конъюнктивные члены таким образом, чтобы на l-м месте в ней стояла X_{i_l} или $\neg X_{i_l}$.

5. Если в преобразованной формуле несколько раз повторяется одна и та же элементарная конъюнкция, то, пользуясь основной равносильностью 5 (идемпотентностью), выбрасываем все ее вхождения, кроме одного.

Пример 1.10. Формула $(X_1 \& \neg X_1 \& X_3) \vee (X_1 \& \neg X_3 \& X_1) \vee \vee X_2$ зависит от списка переменных $<X_1, X_2, X_3>$. Приведем ее к СДНФ относительно этого списка. Первую элементарную конъюнкцию можно отбросить, во второй оставить только одно вхождение X_1, а затем провести преобразования по пп. 3, 4, 5:

$$(X_1 \& \neg X_3) \vee X_2 = (X_1 \& \neg X_3 \& X_2) \vee (X_1 \& \neg X_3 \& \neg X_2) \vee$$
$$\vee (X_2 \& X_1) \vee (X_2 \& \neg X_1) = (X_1 \& \neg X_3 \& X_2) \vee (X_1 \& \neg X_3 \&$$
$$\& \neg X_2) \vee (X_2 \& X_1 \& X_3) \vee (X_2 \& X_1 \& \neg X_3) \vee (X_2 \& \neg X_1 \& X_3) \vee$$
$$\vee (X_2 \& \neg X_1 \& \neg X_3) \equiv (X_1 \& X_2 \& \neg X_3) \vee (X_1 \& \neg X_2 \& \neg X_3) \vee$$
$$\vee (X_1 \& X_2 \& X_3) \vee (\neg X_1 \& X_2 \& X_3) \vee (\neg X_1 \& X_2 \& \neg X_3).$$

Как уже отмечалось, СДНФ обладает некоторой однозначностью, а именно справедлива следующая теорема.

Теорема 1.5 (*о единственности СДНФ*). *Если B_1 и B_2 — совершенные дизъюнктивные нормальные формы формулы A относительно списка переменных $<X_{i_1}, \ldots, X_{i_l}>$, то B_1 и B_2 могут отличаться только порядком своих дизъюнктивных членов.*

Докажем эту теорему несколько позднее (см. *замечание 1.2*, с. 49).

Следует отметить, что если расширить список переменных $<X_{i_1}, \ldots, X_{i_l}>$, от которого зависит формула A, новыми переменными (реально в A не входящими), то относительно нового списка A будем иметь другую СДНФ. Например, СДНФ формулы X_1 относительно списка переменных $<X_1>$ совпадает с самой формулой, а относительно списка переменных $<X_1, X_2>$ является формулой $(X_1 \& X_2) \vee (X_1 \& \neg X_2)$.

Аналогично определяется совершенная конъюнктивная нормальная форма.

Пусть формула A зависит от списка переменных $<X_{i_1}, \ldots, X_{i_k}>$. Тогда говорят, что A находится в *совершенной конъюнктивной нормальной форме* (СКНФ) относительно этого списка, если формула A^* находится в СДНФ относительно списка переменных $<X_{i_1}, \ldots, X_{i_k}>$ или (эквивалентное определение) если выполняются следующие условия:

1) A находится в КНФ;

2) каждый конъюнктивный член A является k-членной дизъ-

юнкцией, причем на l-м месте $(1 \leqslant l \leqslant k)$ этой дизъюнкции обязательно стоят либо переменная X_{i_l}, либо ее отрицание $\neg X_{i_l}$;

3) все конъюнктивные члены A попарно различны.

Пример 1.11. Формулы

$X_1 \vee \neg X_2 \vee \neg X_3$; $(X_1 \vee X_2 \vee X_3) \& (\neg X_1 \vee X_2 \vee X_3) \& (X_1 \vee \vee \neg X_2 \vee \neg X_3)$ находятся в СКНФ относительно списка переменных $<X_1, X_2, X_3>$.

Теорема 1.6. *Пусть формула A зависит от списка переменных $<X_{i_1}, \ldots, X_{i_k}>$ и A не тождественно-истинная. Тогда существует такая формула B, что $A = B$ и B находится в СКНФ относительно списка переменных $<X_{i_1}, \ldots, X_{i_k}>$.*

Пусть A уже находится в КНФ. По условию A на какой-то оценке принимает значение Л. Тогда A^* на двойственной оценке принимает значение И и по теореме о СДНФ существует такая формула B_1, что $A^* \equiv B_1$ и B_1 находится в СДНФ. По принципу двойственности $B^*_1 = A$ и B^*_1 находится в СКНФ.

Можно доказать эту теорему по аналогии с доказательством теоремы о СДНФ. При этом применяются равносильности $(X_{i_l} \vee \neg X_{i_l} \vee C) \& D \equiv D$, $(C \vee X_{i_l}) \& (C \vee \neg X_{i_l}) \equiv C$ и законы идемпотентности.

Теорема 1.7. (*о единственности СКНФ*). *Если B_1 и B_2 — совершенные конъюнктивные нормальные формы формулы A относительно списка переменных $<X_{i_1}, \ldots, X_{i_k}>$, то B_1 и B_2 могут отличаться только порядком своих конъюнктивных членов.*

В самом деле, B^*_1 и B^*_2 в условиях теоремы будут СДНФ для формулы A^* и могут отличаться (по теореме о единственности СДНФ) только порядком дизъюнктивных членов. Отсюда следует утверждение теоремы.

СДНФ и СКНФ можно использовать для распознавания равносильности двух формул.

Критерий равносильности: *две формулы A_1 и A_2, зависящие от списка переменных $<X_{i_1}, \ldots, X_{i_k}>$ и не равные тождественно Л (И), равносильны в том и только в том случае, если они приводятся к СДНФ (СКНФ), отличающимся лишь порядком своих дизъюнктивных (конъюнктивных) членов.*

Если A_1 и A_2 приводятся к одной СДНФ B, то $A_1 = B = A_2$. С другой стороны, если $A_1 \equiv A_2$ и B_1 — СДНФ для A_1, а B_2 — СДНФ для A_2, то $B_1 \equiv A_1 \equiv A_2$, т. е. B_1 будет СДНФ и для A_2, и в силу теоремы о единственности СДНФ B_1 должна отличаться от B_2 только порядком своих дизъюнктивных членов.

Поскольку приведение формул к СДНФ (СКНФ) можно проводить автоматически путем последовательных преобразований, то критерий позволяет устанавливать равносильность, не обращаясь к определению и к оценкам.

1.1.6. Разрешимость

Проблемой разрешимости для логики высказываний называют следующую проблему: существует ли такая процедура, которая позволяла бы для произвольной формулы в конечное число шагов определить, является ли она тавтологией?

Ясно, что эта проблема разрешима, поскольку всегда можно перебрать все оценки списка переменных и вычислить на них значения формулы. Опишем теперь другую, более экономичную процедуру распознавания, связанную с приведением формулы к КНФ.

Утверждение 1.3. *Формула является тавтологией в том и только в том случае, если в ее КНФ в любую из элементарных дизъюнкций в качестве дизъюнктивных членов входят какая-нибудь переменная и ее отрицание.*

Это утверждение является тривиальным следствием двух нижеследующих лемм, первая из которых очевидна.

Лемма 1.4. *Конъюнкция является тавтологией в том и только в том случае, если каждый ее конъюнктивный член является тавтологией.*

Лемма 1.5. *Элементарная дизъюнкция является тавтологией в том и только в том случае, если в ней одновременно присутствуют какая-нибудь переменная и ее отрицание.*

Действительно, если в элементарную дизъюнкцию входят какая-либо переменная и ее отрицание, то это — тавтология в силу тавтологичности $A \vee \neg A$. И наоборот, пусть ни одна переменная не входит в элементарную дизъюнкцию A вместе с отрицанием. Пусть также $\langle X_{i_1}, \ldots, X_{i_k} \rangle$ — список переменных, от которого зависит A. Определим следующую оценку этого списка: переменная X_{i_l} принимает значение Л, если она входит в A как дизъюнктивный член, и значение И — в противном случае. На этой оценке все дизъюнктивные члены A примут значение Л. В самом деле, либо такой дизъюнктивный член есть переменная X_{i_l} (и тогда он принимает значение Л), либо отрицание переменной $\neg X_{i_l}$. В последнем случае X_{i_l} по условию не может быть дизъюнктивным членом и, следовательно, принимает значение И, а $\neg X_{i_l}$ — значение Л. Таким образом, формула A опровержима.

Пример 1.12.

1. Докажем, что $(X_1 \supset X_2) \supset ((X_3 \supset X_1) \supset (X_3 \supset X_2))$ — тавтология:

$$
\begin{aligned}
&(X_1 \supset X_2) \supset ((X_3 \supset X_1) \supset (X_3 \supset X_2)) \equiv \\
&\equiv \neg(\neg X_1 \vee X_2) \vee (\neg(\neg X_3 \vee X_1) \vee \neg X_3 \vee X_2) \equiv \\
&\equiv (X_1 \,\&\, \neg X_2) \vee (X_3 \,\&\, \neg X_1) \vee \neg X_3 \vee X_2 \equiv \\
&\equiv (X_1 \vee X_3 \vee \neg X_3 \vee X_2) \,\&\, (X_1 \vee \neg X_1 \vee \neg X_3 \vee \\
&\vee X_2) \,\&\, (\neg X_2 \vee X_3 \vee \neg X_3 \vee X_2) \,\& \\
&\& \,(\neg X_2 \vee \neg X_1 \vee \neg X_3 \vee X_2).
\end{aligned}
$$

В первую дизъюнкцию входят X_1 и $\neg X_3$, во вторую — X_1 и $\neg X_1$, в третью — X_3 и $\neg X_3$, в четвертую — X_2 и $\neg X_2$.

2. Рассмотрим формулу

$$(X_1 \& X_2 \& X_3) \vee (\neg X_1 \& \neg X_2 \& X_3) \vee (X_1 \& \neg X_2 \neg X_3).$$

При приведении ее к КНФ среди элементарных дизъюнкций будет дизъюнкция $X_1 \vee \neg X_2 \vee \neg X_3$, которая не удовлетворяет нужным условиям. Поэтому данная формула не является тавтологией.

Двойственное утверждение справедливо и для тождественно-ложной формулы.

Утверждение 1.4. *Формула является тождественно-ложной в том и только в том случае, если в ее ДНФ каждая элементарная конъюнкция одновременно содержит в качестве конъюнктивных членов какую-нибудь переменную и ее отрицание.*

Это доказать несложно, исходя из принципа двойственности или используя леммы, аналогичные 1.4 и 1.5.

Задачи и упражнения

1. Сколькими способами можно расставить скобки в слове $X_1 \supset \neg X_2 \vee X_2 \vee X_2 \& X_3$, чтобы получилась формула?

2. Составить таблицы истинности для формул $(X_1 \supset \neg X_2) \& (\neg X_1 \vee X_2)$ и $(X_1 \supset (X_2 \supset X_3)) \subset ((X_1 \supset X_2) \supset (X_1 \supset X_3))$.

3. Записать составные высказывания в виде формул, употребляя высказывательные переменные для обозначения простых высказываний:

а) для того чтобы x было нечетным, достаточно, чтобы x было простым;

б) если идет дождь, то дует ветер;

в) если дует ветер, то идет дождь;

г) ветер дует тогда и только тогда, когда идет дождь;

д) неверно, что ветер дует тогда и только тогда, когда нет дождя;

е) Таня ходит в театр только тогда, когда там показывают пьесу из современной жизни.

4. Доказать равносильности 1—6, 8—11 и 13—15.

5. Доказать следующие равносильности:

а) $\neg (A \supset B) \equiv A \& \neg B$;

б) $A \supset \neg A \equiv \neg A$;

в) $(A \vee B) \& (A \vee C) \& (B \vee D) \& (C \vee D) \equiv (A \& D) \vee (B \& C)$;

г) $A \& (A \vee C) \& (B \vee C) \equiv (A \& B) \vee (A \& C)$;

д) $(A \& B) \vee (A \& C) \vee (B \& D) \vee (C \& D) \equiv (A \vee D) \& (B \vee C)$;

е) $A \vee (\neg A \& B) \equiv A \vee B$.

6. Доказать тождественную истинность формул 1—10 (см. с. 33) и тождественную истинность следующих формул:

а) $(\neg A \supset \neg B) \supset (B \supset A)$;

б) $(\neg B \supset \neg A) \supset ((\neg B \supset A) \supset B)$;

в) $(A \supset B) \supset ((C \vee A) \supset (C \vee B))$;

г) $((A \supset B) \supset A) \supset A$;

д) $\neg A \supset (A \supset B)$.

7. Доказать, что если $A \vee B$ и $\neg A \vee C$ тождественно-истинны, то и $B \vee C$ тождественно-истинна.

8. Проверить правильность следующего рассуждения. Если Джонс не встречал этой ночью Смита, то либо Смит был убийцей, либо Джонс лжет. Если Смит не был убийцей, то Джонс не встречал Смита этой ночью, и убийство имело место после полуночи. Если убийство было совершено после полуночи, то либо Смит был убийцей, либо Джонс лжет. Следовательно, Смит был убийцей.

9. Обосновать метод доказательства «разбором случаев»:

$$(A_1 \vee A_2 \vee \ldots \vee A_n) \supset B \equiv (A_1 \supset B) \,\&\, (A_2 \supset B) \,\&\, \ldots \,\&\, (A_n \supset B).$$

Привести пример такого доказательства.

10. Привести к ДНФ и КНФ следующие формулы:

а) $(X_1 \,\&\, X_2) \supset (\neg X_2 \,\&\, X_3)$;

б) $\neg(\neg (X_1 \vee \neg X_2) \supset (X_2 \,\&\, \neg X_1))$;

в) $(\neg X_1 \supset \neg X_2) \sim (X_2 \sim X_3)$;

г) $((X_1 \supset X_2) \supset (X_3 \supset \neg X_1)) \supset (\neg X_2 \supset \neg X_3)$.

11. Привести к СДНФ и СКНФ следующие формулы:

а) $(X_1 \,\&\, \neg X_2 \,\&\, X_3) \vee (X_1 \,\&\, \neg X_1 \,\&\, \neg X_3) \vee (X_1 \,\&\, X_2 \,\&\, X_3)$;

б) $(X_1 \,\&\, X_2) \supset (\neg X_1 \,\&\, X_3)$;

в) $(\neg X_1 \vee X_2) \,\&\, (X_1 \supset X_2)$;

г) $(\neg X_1 \supset \neg X_2) \sim (X_2 \sim X_3)$;

д) $\neg (X_1 \vee X_3) \,\&\, (X_1 \supset X_2)$.

12. Пусть формула A не содержит никаких логических символов, кроме $\sim$. Доказать, что A является тождественно-истинной тогда и только тогда, когда каждая переменная входит в A четное число раз.

13. Пусть формула A не содержит никаких логических символов, кроме $\sim$ и $\neg$. Доказать, что A является тождественно-истинной тогда и только тогда, когда каждая переменная и символ $\neg$ входят в нее четное число раз.

1.2. БУЛЕВЫ ФУНКЦИИ

1.2.1. Представление булевой функции формулой логики высказываний

Булевой функцией $f(X_1, \ldots, X_n)$ называется произвольная n-местная функция из множества {0, 1} во множество {0, 1}. Итак, как аргументы булевой функции принимают значения из множества {0, 1}, так и сама функция принимает значения из этого же множества.

Всякую булеву функцию от n переменных можно задать таблицей из 2^n строк, в которой в каждой строке записывают одну из оценок списка переменных, принимающих значение 0 или 1. Например, для $n = 3$ булеву функцию можно задать табл. 1.9.

Таблица 1.9

X_1	X_2	X_3	$f(X_1, X_2, X_3)$
1	1	1	$f(1, 1, 1)$
1	1	0	$f(1, 1, 0)$
1	0	1	$f(1, 0, 1)$
1	0	0	$f(1, 0, 0)$
0	1	1	$f(0, 1, 1)$
0	1	0	$f(0, 1, 0)$
0	0	1	$f(0, 0, 1)$
0	0	0	$f(0, 0, 0)$

Так как длина каждого столбца равна 2^n, а различных столбцов из 0 и 1 длины 2^n имеется 2^{2^n}, то существует точно 2^{2^n} различных n-местных булевых функций (или булевых функций от n переменных). Удобно константы 0 и 1 считать нуль-местными булевыми функциями.

Пусть истинностному значению И соответствует 1, а истинностному значению Л — 0. Тогда каждой формуле логики высказываний F можно поставить в соответствие булеву функцию f. При этом, если формуле F_1 соответствует функция f_1, а формуле F_2 — функция f_2 и $F_1 \equiv F_2$, то $f_1 = f_2$.

Составим теперь в новых обозначениях табл. 1.10 для булевых функций, соответствующих основным логическим операциям.

Докажем, что формулы дают нам, по сути дела, все булевы функции, а именно имеет место следующая теорема.

Теорема 1.8 *(первая теорема о представлении булевой функции). Пусть $f(X_1, \ldots, X_k)$ — k-местная булева функция ($k \geqslant 1$).*

Таблица 1.10

X_1	X_2	$\neg X_1$	$X_1 \& X_2$	$X_1 \vee X_2$	$X_1 \supset X_2$	$X_1 \sim X_2$
1	1	0	1	1	1	1
1	0	0	0	1	0	0
0	1	1	0	1	1	0
0	0	1	0	0	1	1

Если f не равна тождественно 0, то существует такая формула F, зависящая от списка переменных $<X_1, \ldots, X_k>$ и находящаяся в СДНФ относительно этого списка, что F выражает собой функцию f. Формула F определена однозначно с точностью до перестановки дизъюнктивных членов.

Докажем сначала вспомогательное утверждение. При $s = 1$ под A^s будем подразумевать формулу A, а при $s = 0$ — формулу $\neg A$. С каждой оценкой списка переменных $<s_1, \ldots, s_k>$ будем ассоциировать элементарную конъюнкцию $X_1^{s_1} \& \ldots \ldots \& X_k^{s_k}$.

Лемма 1.6. *Конъюнкция $X_1^{s_1} \& \ldots \& X_k^{s_k}$, ассоциированная с оценкой $<s_1, \ldots, s_k>$, обращается в 1 на одной и только на одной оценке списка переменных $<X_1, \ldots, X_k>$, а именно на оценке $<s_1, \ldots, s_k>$.*

В самом деле, на оценке $<s_1, \ldots, s_k>$ формула $X_1^{s_1} \& \ldots \ldots \& X_k^{s_k}$ принимает значение 1, так как каждый ее конъюнктивный член $X_l^{s_l}$ принимает значение 1. Действительно, возможны два случая: либо $s_l (1 \leqslant l \leqslant k)$ есть 1, и тогда $X_l^{s_l}$ есть X_l, либо s_l есть 0, и тогда $X_l^{s_l}$ есть $\neg X_l$. В каждом из этих случаев $X_l^{s_l}$ на оценке $<s_1, \ldots, s_k>$ принимает значение 1.

С другой стороны, пусть оценка $<t_1, \ldots, t_k>$ не совпадает с оценкой $<s_1, \ldots, s_k>$. Тогда при некотором $l (1 \leqslant l \leqslant k)$ $s_l \neq t_l$.

Возможны два случая:

1) $s_l = 1$, $t_l = 0$;

2) $s_l = 0$, $t_l = 1$.

В первом случае $X_l^{s_l}$ есть X_l, а во втором — $X_l^{s_l}$ есть $\neg X_l$. В любом случае $X_l^{s_l}$ на оценке $<t_1, \ldots, t_k>$ принимает значение 0, а значит, и вся элементарная конъюнкция $X_1^{s_1} \& \ldots \ldots \& X_k^{s_k}$ принимает значение 0. Лемма доказана.

Пусть теперь функция $f(X_1, \ldots, X_k)$ задана таблицей. Выберем из таблицы все строки, соответствующие оценкам, на которых f принимает значение 1 (поскольку $f \not\equiv 0$, то такие строки найдутся).

Для оценки списка переменных в каждой выбранной строке построим ассоциированную с ней элементарную конъюнкцию и составим дизъюнкцию всех таких конъюнкций. Полученная формула и будет искомой.

Для этого нам нужно доказать следующие два утверждения:

1) если $f(s_1, \ldots, s_k) = 1$, то F на оценке $<s_1, \ldots, s_k>$ принимает значение 1;

2) если $f(s_1, \ldots, s_k) = 0$, то F на оценке $<s_1, \ldots, s_k>$ принимает значение 0.

Пусть $f(s_1, \ldots, s_k) = 1$. Тогда в таблице для f строка, соответствующая оценке $<s_1, \ldots, s_k>$, находится среди вы-

бранных строк, а значит, элементарная конъюнкция $X_1^{s_1} \& \ldots \ldots \& X_k^{s_k}$ находится среди дизъюнктивных членов F. В силу леммы 1.6 конъюнкция $X_1^{s_1} \& \ldots \& X_k^{s_k}$ принимает на оценке $<s_1, \ldots, s_k>$ значение 1, а вместе с ней и вся формула.

Пусть $f(s_1, \ldots, s_k) = 0$. Любой дизъюнктивный член F имеет вид $X_1^{t_1} \& \ldots \& X_k^{t_k}$, причем оценка $<t_1, \ldots, t_k>$ каждый раз отличается от оценки $<s_1, \ldots, s_k>$, так как строка, соответствующая оценке $<s_1, \ldots, s_k>$, не могла быть выбранной. В силу леммы 1.6 каждая такая элементарная конъюнкция обращается в 0 на оценке $<s_1, \ldots, s_k>$, а значит, и вся формула F обращается в 0 на этой оценке, что и требовалось доказать.

Пример 1.13. Для функции, заданной табл. 1.11, соответствующей формулой будет

$$(X_1 \& X_2 \& X_3) \vee (X_1 \& \neg X_2 \& \neg X_3) \vee (\neg X_1 \& X_2 \& X_3) \vee \vee (\neg X_1 \& \neg X_2 \& X_3).$$

Докажем единственность. Пусть F_1 и F_2 — две формулы, выражающие функцию f, находящиеся в СДНФ относительно списка $<X_1, \ldots, X_k>$ и существенно различные, т. е. либо в F_1 есть элементарная конъюнкция, не содержащаяся в F_2, либо в F_2 есть элементарная конъюнкция, не содержащаяся в F_1.

Рассмотрим, например, первый случай. Если $X_1^{s_1} \& \ldots \ldots \& X_k^{s_k}$ — элементарная конъюнкция, содержащаяся в F_1, но не в F_2, то она будет ассоциирована с оценкой $<s_1, \ldots, s_k>$. Любая элементарная конъюнкция, содержащаяся в F_2, будет ассоциирована с некоторой другой оценкой. Поэтому на оценке $<s_1, \ldots, s_k>$ любая такая конъюнкция примет значение 0, а следовательно, и вся формула F_2 примет значение 0. С другой стороны, на этой оценке $X_1^{s_1} \& \ldots \& X_k^{s_k}$ примет значение 1, а поэтому и формула F_1 примет значение 1. Тогда F_1 и F_2 будут выражать собой различные булевы функции, откуда и следует единственность формулы.

Таблица 1.11

X_1	X_2	X_3	$f(X_1, X_2, X_3)$
1	1	1	1
1	1	0	0
1	0	1	0
1	0	0	1
0	1	1	1
0	1	0	0
0	0	1	1
0	0	0	0

Замечание 1.1. Из теоремы 1.8 следует доказанное ранее (см. теорему 1.4) утверждение, что для любой не тождественно-ложной формулы A существует равносильная ей формула F в СДНФ. Однако мы доказали тогда более сильное утверждение, а именно, что F может быть получена из A с помощью равносильных преобразований.

Замечание 1.2. Теперь легко доказать утверждение о единственности СДНФ для некоторой формулы A. В самом деле, если A выражает булеву функцию $f(X_1, \ldots, X_k)$, то и любая СДНФ для A должна (в силу равносильности) выражать собой ту же функцию. Поэтому все эти СДНФ должны совпадать с точностью до порядка элементарных конъюнкций.

Аналогично можно доказать, что булевы функции представимы формулами в СКНФ.

Теорема 1.9 *(вторая теорема о представлении булевой функции). Пусть $f(X_1, \ldots, X_k)$ — k-местная булева функция ($k \geqslant 1$), не равная тождественно 1. Существует такая формула F, зависящая от списка переменных $\langle X_1, \ldots, X_k \rangle$ и находящаяся в СКНФ относительно этого списка, что F выражает собой $f(X_1, \ldots, X_k)$. Формула F определена однозначно с точностью до перестановки конъюнктивных членов.*

Перечислим все булевы функции от двух переменных. Таких функций существует $2^{2^2} = 16$. Две функции сохраняют постоянное значение: $f_1(X, Y) = 0$; $f_2(X, Y) = 1$. Четыре функции существенно зависят только от одного из аргументов: $f_3(X, Y) = X$; $f_4(X, Y) = Y$; $f_5(X, Y) = \neg X$; $f_6(X, Y) = \neg Y$. Следующие четыре функции принимают значение 1 точно на одной оценке; в СДНФ этих функций должна быть только одна элементарная конъюнкция: $f_7(X, Y) = X \,\&\, Y$ (оценка (1,1)); $f_8(X, Y) = \neg X \,\&\, Y$ (оценка (0,1)); $f_9(X, Y) = X \,\&\, \neg Y$ (оценка (1,0)); $f_{10}(X, Y) = \neg X \,\&\, \neg Y$ (оценка (0,0)). Последняя функция называется иногда функцией Шеффера и обозначается $X|Y$. Двойственными этим четырем функциям являются функции $f_{11}(X, Y) = X \vee Y$; $f_{12}(X, Y) = \neg X \vee Y = X \supset Y$; $f_{13}(X, Y) = X \vee \neg Y = Y \supset X$; $f_{14}(X, Y) = \neg X \vee \neg Y$. Все эти функции на трех оценках принимают значение 1 и только на одной оценке — значение 0, а именно: на (0,0), (1,0), (0,1), (1,1) соответственно. Последняя из них называется функцией Вебба и обозначается $X \circ Y$. Наконец, имеются две функции, существенно зависящие от каждого из аргументов и принимающие на двух оценках значение 0 и на двух — значение 1: $f_{15}(X, Y) = (X \,\&\, Y) \vee (\neg X \,\&\, \neg Y) = X \sim Y$ (на оценках (1,1) (0,0) функция принимает значение 1); $f_{16}(X, Y) = (\neg X \,\&\, Y) \vee (X \,\&\, \neg Y)$ (на оценках (0,1), (1,0) функция принимает значение 1). Последняя функция соответствует разделительному «или». Она называется также сложением по модулю 2 и обозначается $X + Y$.

Как видим, рассмотренные функции попарно различны, так что наше перечисление является исчерпывающим.

1.2.2. Полные системы булевых функций

Система булевых функций $\{f_1, \ldots, f_m\}$ называется *полной*, если любая булева функция может быть выражена через функции $f_1, \ldots, f_m$ с помощью *суперпозиций* (т. е. составления сложных функций).

Дадим определение суперпозиции функций. Пусть

$K^0 = \{f_1(X_1, \ldots, X_{k_1}), f_2(X_1, \ldots, X_{k_2}), \ldots, f_m(X_1, \ldots, X_{k_m})\}$ —

конечная система булевых функций. Функция f называется суперпозицией ранга 1 (или элементарной суперпозицией) функций $f_1, \ldots, f_m$, если f может быть получена одним из следующих способов: а) переименованием некоторой переменной X_l какой-нибудь функции f_i, т. е. $f = f_i(X_1, \ldots, X_{l-1}, Y, X_{l+1}, \ldots, X_{k_i})$, где Y может совпасть с любой переменной; б) подстановкой некоторой функции $f_i (1 \leqslant i \leqslant m)$ вместо какой-либо переменной X_j любой из функций $f_i \in K^0$, т. е. $f = f_i(X_1, \ldots, X_{j-1}, f_i(X_1, \ldots, X_{k_i}), X_{j+1}, \ldots, X_{k_i})$.

Суперпозиции ранга 1 образуют класс функций K^1. Класс функций, получающийся из функций класса K^{r-1} суперпозиций ранга $r-1$ с помощью элементарных суперпозиций, называется классом функций K^r суперпозиций ранга r. *Суперпозициями* функций из K^0 называются функции, входящие в какой-либо из классов K^r.

Несложно доказывается следующее

Утверждение 1.5. *Пусть система $\{f_1, \ldots, f_m\}$ — полная и любая из функций $f_1, \ldots, f_m$ может быть выражена с помощью суперпозиций через функции $g_1, \ldots, g_l$. Тогда система $\{g_1, \ldots, g_l\}$ тоже полная.*

Пример 1.14. Докажем полноту следующих систем функций: а) $\{\neg, \&, \vee\}$; б) $\{\neg, \vee\}$; в) $\{\neg, \&\}$; г) $\{\neg, \supset\}$.

Полнота системы $\{\neg, \&, \vee\}$ непосредственно следует из теорем 1.8 и 1.9.

Для доказательства полноты системы $\{\neg, \vee\}$ воспользуемся полнотой системы $\{\neg, \&, \vee\}$ и утверждением 1.5, где в роли функций f_1, f_2, f_3 выступают соответственно $\neg$, $\&$, $\vee$, а в роли функций g_1, g_2 — $\neg$, $\vee$. Тогда $f_1 = g_1$ и $f_2 = X_1 \& X_2 = \neg(\neg X_1 \vee \vee \neg X_2)$, т. е. функция f_2 выражена через g_1 и g_2, а $f_3 = g_2$.

Полнота системы $\{\neg, \&\}$ доказывается аналогично предыдущему случаю с использованием равносильности $X_1 \vee X_2 = \neg(\neg X_1 \& \neg X_2)$.

Для доказательства полноты системы $\{\neg, \supset\}$ воспользуемся полнотой системы $\{\neg, \vee\}$ и утверждением 1.5, где в роли функций f_1 и f_2 выступают соответственно $\neg$, $\vee$, а в роли функций g_1, g_2 — $\neg$, $\supset$. Тогда $f_1 = g_1$, $f_2 = X_1 \vee X_2 = \neg X_1 \supset X_2$.

В разд. 1.1.1 была определена функция $X + Y$ — сложение по модулю 2. Запишем таблицу истинности для этой функции (см. табл. 1.12). Иногда удобнее вместо символа $\&$ писать символ $\cdot$ или вообще опускать его, как это делается в арифметике. Тогда функцию $X \& Y$ можно записать как $X \cdot Y$ или просто XY. Таблица истинности для этой функции также содержится в табл. 1.12.

Рассмотрим теперь систему функций $\{+, \cdot, 1\}$. Это полная система, что следует из утверждения 1.5, полноты системы $\{\neg, \cdot\}$ и равносильности $\neg X = X + 1$. По таблицам истинности нетрудно проверить, что выполняются тождества (часть из них уже проверена при доказательстве основных равносильностей):

1) $X+Y=Y+X$; 1′) $X \cdot Y = Y \cdot X$;

2) $(X+Y)+Z=X+(Y+Z)$; 2′) $(X \cdot Y) \cdot Z = X \cdot (Y \cdot Z)$;

3) $X+X=0$; 3′) $X \cdot X = X$;

Таблица 1.12

X	Y	$X+Y$	XY
1	1	0	1
1	0	1	0
0	1	1	0
0	0	0	0

4) $X \cdot (Y+Z) = X \cdot Y + X \cdot Z$; (1.1)

5) $0 + X = X$;

6) $0 \cdot X = 0$;

7) $1 \cdot X = X$.

Заметим, что все тождества, кроме 3 и 3′, выражают свойства, аналогичные свойствам арифметического сложения и умножения. Следовательно, в силу полноты системы $\{+, \cdot, 1\}$ и тождеств $1 - 2$, $1' - 2'$, $4 - 7$ любую булеву функцию можно представить в виде многочлена от своих переменных.

Многочленом Жегалкина называется многочлен, являющийся суммой константы 0 или 1 и различных одночленов, в которые все переменные входят не выше чем в первой степени:

$$\Sigma X_{i_1} \dots X_{i_k} + a_l,$$

причем на каждом наборе $\langle i_1, \dots, i_k \rangle$ все i_j $(j = 1, \dots, k)$ различны, $a_l \in \{0, 1\}$.

Воспользовавшись тождествами 3 и 3′, можно доказать, что каждая булева функция может быть представлена многочленом Жегалкина. Поскольку число различных булевых функций от n переменных равно 2^{2^n} и число различных многочленов Жегалкина от n переменных также равно 2^{2^n}, то представление булевой функции многочленом Жегалкина единственно.

Пример 1.15. Представим многочленами Жегалкина функции $X \vee Y$ и $X \vee Y \vee Z$:

$$X \vee Y = \neg(\neg X \& \neg Y) = (X+1)(Y+1) + 1 =$$
$$= XY + X + Y + 1 + 1 = XY + X + Y;$$
$$X \vee Y \vee Z = XYZ + XY + XZ + YZ + X + Y + Z.$$

Утверждение 1.5 дает лишь достаточное условие полноты системы булевых функций. Перейдем теперь к установлению критерия полноты системы булевых функций.

Класс (множество) K булевых функций называется *функционально замкнутым*, если вместе с функциями из этого класса он содержит и все их суперпозиции.

Очевидно, что для доказательства замкнутости класса достаточно показать, что элементарные суперпозиции не выводят из этого класса. Функционально замкнутыми являются: класс всех булевых функций, класс, содержащий только тождественные функции вида $f(X) = X$, класс функций от одной переменной.

Утверждение 1.6. *Никакая полная система булевых функций не может содержаться в функционально замкнутом классе, отличном от класса K_1 всех булевых функций.*

Действительно, в противном случае для полной системы $\{f_1, \ldots, f_m\}$ найдется замкнутый класс K такой, что $\{f_1, \ldots, f_m\} \subset K \subset K_1$ и $K \neq K_1$. Следовательно, найдется функция f такая, что $f \notin K$, т. е. f не может быть выражена через $f_1, \ldots, f_m$, что противоречит полноте этой системы.

Рассмотрим некоторые функционально замкнутые классы функций.

Через T_0 обозначается класс *функций, сохраняющих* 0, т. е. функций, удовлетворяющих условию $f(0, 0, \ldots, 0) = 0$.

Через T_1 обозначается класс *функций, сохраняющих* 1, т. е. функций, удовлетворяющих условию $f(1, 1, \ldots, 1) = 1$.

Заметим, что $X_1 \sim X_2 \notin T_0$, а $X_1 \sim X_2 \in T_1$, $X_1 + X_2 \in T_0$, а $X_1 + X_2 \notin T_1$, т. е. эти классы функций различны и не совпадают с классом всех булевых функций.

Поскольку элементарные суперпозиции не выводят из классов T_0 и T_1 соответственно, то они функционально замкнуты. Например, если $f_1(X_1, \ldots, X_{k_1}) \in T_0$ и $f_2(X_1, \ldots, X_{k_2}) \in T_0$, то функция $f_1(X_1, \ldots, X_{i-1}, f_2(X_i, \ldots, X_{k_2}), X_{i+1}, \ldots, X_{k_1})$ на нулевой оценке, очевидно, также принимает значение 0, т. е. принадлежит классу T_0. Переименование переменной также не выводит из T_0.

Пусть $f(X_1, \ldots, X_n)$ — булева функция. Функция $f^*(X_1, \ldots, X_n)$ называется *двойственной* $f(X_1, \ldots, X_n)$, если $f^*(X_1, \ldots, X_n) = \neg f(\neg X_1, \ldots, \neg X_n)$. Очевидно, двойственной $X_1 \vee X_2$ является функция $X_1 \& X_2$ и наоборот; двойственной $f(X_1) = 0$ является функция $f(X_2) = 1$ и наоборот.

Утверждение 1.7 *(принцип двойственности). Пусть*

$$\Phi = f_1(X_1, \ldots, X_{i-1}, f_2(X_i, \ldots, X_{k_2}), X_{i+1}, \ldots, X_{k_1}).$$

Тогда

$$\Phi^* = f^*_1(X_1, \ldots, X_{i-1}, f^*_2(X_i, \ldots, X_{k_2}), X_{i+1}, \ldots, X_{k_1}).$$

Действительно, без ограничения общности можно считать, что $k_1 > k_2$. Тогда

$$\Phi^*(X_1, \ldots, X_{k_1}) = \neg\Phi(\neg X_1, \ldots, \neg X_{k_1}) = \neg f_1(\neg X_1, \ldots$$

$$\ldots, \neg X_{i-1}, \neg f_2(\neg X_1, \ldots, \neg X_{k_1}), \neg X_{i+1}, \ldots, \neg X_{k_2}) =$$
$$= \neg f_1(\neg X_1, \ldots, \neg X_{i-1}, \neg f^*_2(X_1, \ldots, X_{k_2}), \neg X_{i+1}, \ldots, \neg X_{k_1}) =$$
$$= f^*_1(X_1, \ldots, X_{i-1}, f^*_2(X_1, \ldots, X_{k_2}), X_{i+1}, \ldots, X_{k_1}).$$

Функция $f(X_1, \ldots, X_n)$ называется *самодвойственной*, если $f(X_1, \ldots, X_n) = f^*(X_1, \ldots, X_n)$. Класс самодвойственных функций обозначается через S.

Пример 1.16. Функция $f(X_1, X_2, X_3) = X_1 + X_2 + X_3$ самодвойственная. Действительно,

$$f^*(X_1, X_2, X_3) = \neg f(\neg X_1, \neg X_2, \neg X_3) = (X_1 + 1) +$$
$$+ (X_2 + 1) + (X_3 + 1) + 1 = X_1 + X_2 + X_3.$$

Класс S самодвойственных функций функционально замкнут. Это следует из утверждения 1.7, поскольку если $f_1, f_2 \in S$, т. е. $f^*_1 = f_1$, $f^*_2 = f_2$, то $\Phi = \Phi^*$. Переименование переменной также не выводит из S.

Функция $f(X_1, \ldots, X_n)$ называется *линейной*, если $f(X_1, \ldots, X_n) = a_0 + a_1X_1 + \ldots + a_nX_n$, где $a_i \in \{0, 1\}$. Класс линейных функций обозначается через L. Очевидно, $X + Y \in L$, $0 \in L$, $XY \notin L$.

Класс линейных функций L функционально замкнут. Если $f_1 = a_0 + a_1X_1 + \ldots + a_{k_1}X_{k_1} \in L$ и $f_2 = b_0 + b_1x_1 + \ldots + b_{k_2}X_{k_2} \in L$, то $a_0 + \ldots + a_{i-1}X_{i-1} + a_i(b_0 + b_1X_1 + \ldots + b_{k_2}X_{k_2}) + a_{i+1}X_{i+1} + \ldots + a_{i+1}X_{k_1} \in L_1$, что следует из тождеств 3 и 5 (см. (1.1)). Переименование переменной также не выводит из L.

Введем отношение частичного порядка на множестве оценок списка переменных $<X_1, \ldots, X_n>$. Говорят, что оценка $\alpha = <a_1, \ldots, a_n>$ предшествует оценке $\beta = <\beta_1, \ldots, \beta_n>$, где $a_i \in \{0, 1\}$, $\beta_i \in \{0, 1\}$, $i = 1, \ldots, n$, если $a_i \leqslant \beta_i$ для любого i (обозначаем $\alpha < \beta$). Например, $<0, 1, 0, 1> < <0, 1, 1, 1>$, а оценки $<0, 1, 1>$ и $<1, 0, 1>$ несравнимы. Введенное отношение $<$ есть отношение частичного порядка (см. введение, разд. 0.3).

Функция $f(X_1, \ldots, X_n)$ называется *монотонной*, если для любых оценок α и β списка переменных таких, что $\alpha < \beta$, имеем $f(\alpha) \leqslant f(\beta)$. Класс монотонных функций обозначается через M.

Пример 1.17. Функции $f(X_1) = X_1$, $f(X_1, X_2) = X_1 \& X_2$, $f(X_1, X_2) = X_1 \vee X_2$ монотонные. Функция $f(X_1, X_2) = X_1 \supset X_2$ не монотонная, так как $<0, 0> < <1, 0>$, а для этой функции $f(0, 0) = 1$, $f(1, 0) = 0$.

Класс M монотонных функций функционально замкнут. Действительно, элементарные суперпозиции не выводят из класса M монотонных функций. В частности, если: $f_1(X_1, \ldots,$

$X_{k_1}) \in M$ и $f_2(X_1, \ldots, X_{k_2}) \in M$, то $f_1(X_1, \ldots, X_{i-1}, f_2(X_1, \ldots, X_{k_2}), X_{i+1}, \ldots, X_{k_1}) \in M$. Без ограничения общности можно считать, что $k_1 \geqslant k_2$. Если $\alpha = <\alpha_1, \ldots, \alpha_{k_1}>$, $\beta = <\beta_1, \ldots, \beta_{k_1}>$ и $\alpha \prec \beta$, то $f_2(\alpha_1, \ldots, \alpha_{k_2}) \leqslant f_2(\beta_1, \ldots, \beta_{k_2})$. Отсюда $f_1(\alpha_1, \ldots, \alpha_{i-1}, f_2(\alpha_1, \ldots, \alpha_{k_2}), \alpha_{i+1}, \ldots, \alpha_{k_1}) \leqslant f_1(\beta_1, \ldots, \beta_{i-1}, f_2(\beta_1, \ldots, \beta_{k_2}), \beta_{i+1}, \ldots, \beta_{k_1})$, так как $<\alpha_1, \ldots, \alpha_{i-1}, f_2(\alpha_1, \ldots, \alpha_{k_2}), \alpha_{i+1}, \ldots, \alpha_{k_1}> \prec <\beta_1, \ldots, \beta_{i-1}, f_2(\beta_1, \ldots, \beta_{k_2}), \beta_{i+1}, \ldots, \beta_{k_1}>$.

Классы T_0, T_1, S, L, M неполные и попарно различные, поскольку можно привести примеры булевых функций, не принадлежащих ни одному из этих классов, и примеры функций, принадлежащих одному из любых двух классов, но не принадлежащих другому. Кроме перечисленных можно указать и много других функционально замкнутых классов. Оказывается, что для проверки полноты системы булевых функций можно ограничиться рассмотренными пятью функционально замкнутыми классами.

Теорема 1.10 (*теорема Поста*). *Для того чтобы система булевых функций* $\{f_1, \ldots, f_m\}$ *была полной, необходимо и достаточно, чтобы для каждого из классов* T_0, T_1, L, M *и* S *нашлась функция* f_i *из системы, не принадлежащая этому классу.*

Докажем только необходимость этого условия. Классы T_0, T_1, L, M и S попарно различны и не совпадают с классом K_1 всех булевых функций. Если бы все функции $f_1, \ldots, f_m$ принадлежали какому-либо из этих классов, то, как следует из утверждения 1.6, в силу его замкнутости система $\{f_1, \ldots, f_m\}$ не была бы полной.

Пример 1.18. Доказать полноту системы $\{+, \vee, 1\}$. Составим таблицу Поста. В клетках табл. 1.13 будем писать знак «+» или «—» в зависимости от того, принадлежит рассматриваемая функция данному функционально замкнутому классу или нет. В силу теоремы Поста для полноты системы необходимо и достаточно, чтобы в каждом столбце был хотя бы один минус.

Таблица 1.13

f	T_0	T_1	S	L	M
$X+Y$	+	—	—	+	—
$X \vee Y$	+	+	—	—	+
1	—	+	—	+	+

Принадлежность функции $X+Y$ классам T_0, L и непринадлежность классу T_1 очевидна. Эта функция несамодвойственная, так как $(X+Y)^*=\neg(\neg X+\neg Y)=((X+1)+(Y+1))+1=X+Y+1$.

Функция $X+Y$ немонотонная, так как $<1, 0><<1, 1>$, а $1+0=1$ и $1+1=0$.

Функция $X \vee Y$, очевидно, принадлежит классам T_0 и T_1. Она несамодвойственная, так как двойственная ей функция есть $X \& Y$. Она нелинейная, так как многочлен Жегалкина этой функции имеет вид $X \vee Y = XY + X + Y$. Монотонность функции $X \vee Y$ легко проверяется по ее таблице истинности. Для функции, тождественно-равной 1, таблица Поста заполняется тривиально. Так как каждый столбец таблицы Поста содержит по меньшей мере один знак «—», то система $\{+, \vee, 1\}$ — полная.

Система функций G называется *независимой*, если никакая функция $f \in G$ не представлена суперпозициями функций из $G \setminus \{f\}$. Независимая система функций называется *базисом* функционально замкнутого класса K, если всякая функция из K есть суперпозиция функций из G.

Пример 1.19.

1. Система функций $\{\&, \neg\}$ независимая, так как $\& \in T_1$, $\neg \in T_1$, $\& \in L$, $\neg \in L$.

2. Система функций $\{\&, \vee, \neg\}$ не является независимой, так как $X \vee Y = \neg(\neg X \& \neg Y)$.

3. Система функций $\{+, \cdot, 1\}$ независимая, так как $+ \notin T_1$, $\cdot \in T_1$, $1 \in T_1$; $+ \in L$, $\cdot \notin L$, $1 \in L$; $+ \in T_0$, $\cdot \in T_0$, $1 \notin T_0$.

4. Системы функций $\{\&, \neg\}$, $\{+, \cdot, 1\}$ — базисы класса K_1 всех булевых функций.

5. Система функций $\{0, \sim\}$ — базис функционально замкнутого класса L линейных функций. Это независимая система, так как $0 \in T_0$, $\sim \notin T_0$, $0 \notin T_1$, $\sim \in T_1$. С другой стороны, каждая линейная функция выражается суперпозициями функций $+$, $\neg$, поскольку $\neg X = X + 1$. Но эти функции в свою очередь выражаются через 0 и $\sim$: $\neg X = X \sim 0$, $X + Y = \neg(X \sim Y)$.

1.2.3. Минимизация в классе дизъюнктивных нормальных форм

Выше было доказано, что произвольная булева функция может быть представлена формулой в дизъюнктивной и конъюнктивной нормальной форме. Равносильными преобразованиями можно получить формулу, содержащую меньшее, чем исходная, число переменных. Например:

а) $(X_1 \& X_2) \vee (X_1 \& \neg X_2) = X_1$;

б) $(X_1 \& X_2) \vee X_1 = X_1$;

в) $(X_1 \& X_2) \vee (X_1 \& X_3) = X_1 \& (X_2 \vee X_3)$.

Заметим, что последнее преобразование выводит формулу из класса дизъюнктивных нормальных форм. Будем минимизировать формулы в классе ДНФ.

Дизъюнктивная нормальная форма называется *минимальной*, если она содержит наименьшее общее число вхождений высказывательных переменных по сравнению со всеми равносильными ей дизъюнктивными нормальными формами.

Следовательно, минимальную ДНФ данной формулы можно найти, перебрав конечное число равносильных ей ДНФ и выбрав среди них ту, которая содержит минимальное число переменных. Однако при большом числе переменных такой перебор практически невыполним. Существуют эффективные способы нахождения минимальной ДНФ (см., например, [11]). Рассмотрим только один из них — метод минимизирующих карт. Хотя этот метод и не отличается большой эффективностью, зато он прост для изложения и не требует введения дополнительных понятий.

Пусть булева функция задана таблицей истинности или СДНФ. При записи формул будем опускать символ & и вместо $\neg X_i$ писать $\overline{X}_i$, например $X_1\overline{X}_2X_3 \vee X_1\overline{X}_3$ вместо $(X_1 \& \neg X_2 \& X_3) \vee (X_1 \& \neg X_3)$. Составим следующую карту (см. табл. 1.14).

Утверждение 1.8. *Если конъюнкция $X_1^{\varepsilon_1} \dots X_n^{\varepsilon_n}$, $\varepsilon_i \in \{0, 1\}$, $i = 1, \dots, n$, принадлежащая j-й строке табл. 1.14, не входит в СДНФ, выражающую функцию $f(X_1, \dots, X_n)$, то любая конъюнкция j-й строки не входит ни в какую ДНФ, выражающую исходную функцию.*

Действительно, если конъюнкция $X_1^{\varepsilon_1} \dots X_n^{\varepsilon_n}$ не входит в СДНФ, выражающую функцию $f(X_1, \dots, X_n)$, то по алгоритму построения СДНФ (см. теорему 1.8) $f(\varepsilon_1, \dots, \varepsilon_n) = 0$. Если бы какая-то конъюнкция j-й строки вошла в некоторую ДНФ функции f, то $f(\varepsilon_1, \dots, \varepsilon_n) = 1$.

Опишем способ построения минимальной ДНФ:

1. Отметим в табл. 1.14 строки, в которых соответствующая им конъюнкция $X_1^{\varepsilon_1} \dots X_n^{\varepsilon_n}$ не принадлежит СДНФ, выражающей функцию $f(X_1, \dots, X_n)$. Вычеркнем все конъюнкции в этих строках.

2. Вычеркнутые в этих строках конъюнкции вычеркнем также во всех остальных строках таблицы.

3. В каждой строке выберем из оставшихся конъюнкций лишь конъюнкции с наименьшим числом сомножителей, а остальные вычеркнем.

4. В каждой строке выберем по одному оставшемуся элементу и составим из них ДНФ.

5. Из всех ДНФ, полученных в п. 4, выберем минимальную.

Заметим, что п. 5 предусматривает перебор различных ДНФ для нахождения минимальной из них. Однако при этом число

Таблица 1.14

X_1	X_2	...	X_{n-1}	X_n	X_1X_2	X_1X_3	...	$X_{n-2}X_n$	$X_{n-1}X_n$	...	$X_1X_2 \ldots X_{n-1}X_n$
X_1	X_2	...	X_{n-1}	$\overline{X}_n$	X_1X_2	X_1X_3	...	$X_{n-2}\overline{X}_n$	$X_{n-1}\overline{X}_n$	...	$X_1X_2 \ldots X_{n-1}\overline{X}_n$
X_1	X_2	...	$\overline{X}_{n-1}$	X_n	X_1X_2	X_1X_3	...	$X_{n-2}X_n$	$\overline{X}_{n-1}X_n$	...	$X_1X_2 \ldots \overline{X}_{n-1}X_n$
...	...	...	...	...	...	...	...	...	...	...	...
$\overline{X}_1$	$\overline{X}_2$	...	$\overline{X}_{n-1}$	$\overline{X}_n$	$\overline{X}_1\overline{X}_2$	$\overline{X}_1\overline{X}_3$	...	$\overline{X}_{n-2}\overline{X}_n$	$\overline{X}_{n-1}\overline{X}_n$	...	$\overline{X}_1\overline{X}_2 \ldots \overline{X}_{n-1}\overline{X}_n$

вариантов перебора, как правило, меньше, чем в случае, когда перебираются все равносильные ДНФ.

Действуя в соответствии с пп. 1—5, получим минимальную ДНФ, выражающую функцию $f(X_1, \ldots, X_n)$. Пусть F — формула в ДНФ, полученная в п. 5. Тогда, если $f(\varepsilon_1, \ldots, \varepsilon_n) = 1$, то $F(\varepsilon_1, \ldots, \varepsilon_n) = 1$. Действительно, если конъюнкция $X_1^{\varepsilon_1} \ldots X_n^{\varepsilon_n}$ соответствует j-й строке табл. 1.14, то эта строка осталась невычеркнутой; любая конъюнкция в этой строке на оценке $<\varepsilon_1, \ldots, \varepsilon_n>$ принимает значение 1. Следовательно, и формула F, содержащая одну из таких конъюнкций в качестве дизъюнктивного члена, принимает на оценке $<\varepsilon_1, \ldots, \varepsilon_n>$ значение 1, т. е. $F(\varepsilon_1, \ldots, \varepsilon_n) = 1$. Если $f(\varepsilon_1, \ldots, \varepsilon_n) = 0$, то на оценке $<\varepsilon_1, \ldots, \varepsilon_n>$ все невычеркнутые в табл. 1.14 конъюнкции принимают значение 0, так как все конъюнкции, которые на этой оценке принимают значение 1, вычеркнуты. Следовательно, составленная из этих конъюнкций ДНФ принимает на этой оценке значение 0, т. е. $F(\varepsilon_1, \ldots, \varepsilon_n) = 0$.

Пример 1.20. Пусть $f(X_1, X_2, X_3) = X_1X_2\overline{X}_3 \vee X_1\overline{X}_2X_3 \vee \vee X_1\overline{X}_2\overline{X}_3 \vee \overline{X}_1\overline{X}_2X_3$. Составим табл. 1.15.

Отметим знаком * вычеркнутые строки. После применения пп. 1—3 в табл. 1.15 останутся конъюнкции, приведенные в табл. 1.16. После применения пп. 4 и 5 получим минимальную ДНФ $X_1\overline{X}_2 \vee \overline{X}_2X_3$.

Поскольку алгоритмы нахождения минимальной ДНФ до-

Таблица 1.15

X_1	X_2	X_3	X_1X_2	X_1X_3	X_2X_3	$X_1X_2X_3$	*
X_1	X_2	$\overline{X}_3$	X_1X_2	$X_1\overline{X}_3$	$X_2\overline{X}_3$	$X_1X_2\overline{X}_3$	
X_1	$\overline{X}_2$	X_3	$X_1\overline{X}_2$	X_1X_3	$\overline{X}_2X_3$	$X_1\overline{X}_2X_3$	
X_1	$\overline{X}_2$	$\overline{X}_3$	$X_1\overline{X}_2$	$X_1\overline{X}_3$	$\overline{X}_2\overline{X}_3$	$X_1\overline{X}_2\overline{X}_3$	
$\overline{X}_1$	X_2	X_3	$\overline{X}_1X_2$	$\overline{X}_1X_3$	X_2X_3	$\overline{X}_1X_2X_3$	*
$\overline{X}_1$	X_2	$\overline{X}_3$	$\overline{X}_1X_2$	$\overline{X}_1\overline{X}_3$	$X_2\overline{X}_3$	$\overline{X}_1X_2\overline{X}_3$	*
$\overline{X}_1$	$\overline{X}_2$	X_3	$\overline{X}_1\overline{X}_2$	$\overline{X}_1X_3$	$\overline{X}_2X_3$	$\overline{X}_1\overline{X}_2X_3$	
$\overline{X}_1$	$\overline{X}_2$	$\overline{X}_3$	$\overline{X}_1\overline{X}_2$	$\overline{X}_1\overline{X}_3$	$\overline{X}_2\overline{X}_3$	$\overline{X}_1\overline{X}_2\overline{X}_3$	*

				$X_1\overline{X}_3$		
			$X_1\overline{X}_2$		$\overline{X}_2X_3$	
			$X_1\overline{X}_2$	$X_1\overline{X}_3$		
					$\overline{X}_2X_3$	

вольно сложны, иногда применяют алгоритмы упрощения, получая некоторые приемлемые для дальнейшего использования ДНФ, среди которых содержатся и минимальные. К таким типам ДНФ относятся и сокращенные ДНФ, которые мы будем использовать в гл. 4. Сокращенная ДНФ формулы может быть получена, например, по следующему алгоритму.

Рассмотрим произвольную конъюнктивную нормальную форму исходной формулы. Воспользуемся законом обобщенной дистрибутивности $(A_1 \vee \ldots \vee A_k) \& (B_1 \vee \ldots \vee B_l) = (A_1 \& B_1) \vee \ldots \vee (A_1 \& B_l) \vee \ldots \vee (A_k \& B_l)$ и «раскроем» скобки. Затем произведем упрощения полученной формулы, пользуясь равносильностями $(A \& B) \vee A = A$, $A \vee A = A$ и удаляя тождественно-ложные дизъюнктивные члены. В результате получим ДНФ, которая называется *сокращенной*.

1.2.4. Контактные схемы

Рассмотрим одно из приложений логики высказываний — применение ее к теории электрических цепей, а именно к контактным схемам.

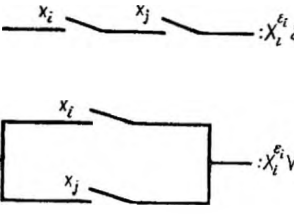
Пусть $x_1, \ldots, x_n$ — набор контактов в электрической схеме. Контакты могут быть размыкающими и замыкающими. Контакт называется *размыкающим*, если он размыкается при подаче напряжения на обмотку реле, к которому он подключен, а когда напряжение не подается, контакт замкнут. Контакт называется

замыкающим, если он замыкается при подаче напряжения на обмотку реле, к которому он подключен, а когда напряжение не подается, контакт разомкнут. В схеме один и тот же контакт может неоднократно быть как размыкающим, так и замыкающим. Будем считать, что $x_i = 1$, если на обмотку контакта x_i подается напряжение, и $x_i = 0$ — в противном случае.

Каждой последовательно-параллельной схеме с контактами $x_1, \ldots, x_n$ поставим в соответствие ее *функцию проводимости*

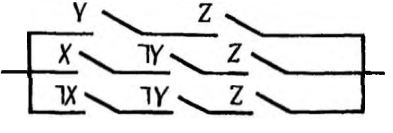
$$f(x_1, \ldots, x_n) = \begin{cases} 1, \text{ если схема проводит ток;} \\ 0, \text{ если схема не проводит ток.} \end{cases}$$

Тогда функция проводимости схемы, состоящей из одного контакта x_i, есть $X_i^{\varepsilon_i}$, где $\varepsilon_i = 1$, если контакт замыкающий, и $\varepsilon_i = 0$, если контакт размыкающий. Функция проводимости схемы из последовательно соединенных контактов x_i и x_j есть $X_i^{\varepsilon_i} \& X_j^{\varepsilon_j}$, а функция проводимости схемы из параллельно соединенных контактов — $X_i^{\varepsilon_i} \vee X_j^{\varepsilon_j}$:



Следовательно, каждой последовательно-параллельной контактной схеме можно поставить в соответствие формулу логики высказываний, реализующую функцию проводимости этой схемы. Две схемы считаются эквивалентными, если они одновременно проводят (или не проводят) ток при подаче напряжения на одноименные реле, т. е. если они имеют одинаковую функцию проводимости. Применяя равносильности логики высказываний, можно упрощать контактные схемы, заменив их эквивалентными, содержащими меньшее число контактов.

Пример 1.21. Запишем функцию проводимости и упростим электрическую схему:



Функция проводимости: $(Y \& Z) \vee (X \& \neg Y \& Z) \vee (\neg X \& \& \neg Y \& Z) = (Y \& Z) \vee ((\neg Y \& Z) \& (X \vee \neg X)) = Z \& (Y \vee \vee \neg Y) = Z$.

Эквивалентная схема:

Задачи и упражнения

1. Указать формулы в СДНФ и СКНФ, выражающие следующие функции:

а) $f(X_1, X_2, X_3)$, равную 1 тогда и только тогда, когда большинство переменных равно 1;

б) $f(X_1, X_2, X_3, X_4)$, равную 1 тогда и только тогда, когда $X_1 + X_2 + X_3 + X_4 \geqslant 4$.

2. Выразить с помощью суперпозиции:

а) $\supset$ через 1, $+$, $\cdot$;

б) $\&$, $\vee$ через $\supset$, $\neg$;

в) $\neg$ через 0, $\supset$;

г) $\neg$, $\vee$, $\&$ через $|$ (здесь $|$ — функция Шеффера $X|Y = \neg X \& \neg Y$);

д) $\neg$, $\vee$, $\&$ через $\circ$ (здесь $\circ$ — функция Вебба $X \circ Y = \neg X \vee \neg Y$).

3. Представить многочленом Жегалкина: а) $X \supset Y$; б) $X \sim Y$.

4. Доказать, что самодвойственная функция на противоположных наборах значений переменных принимает противоположные значения.

5. Доказать, что число самодвойственных функций от n переменных равно $2^{2^{n-1}}$.

6. Доказать, что число линейных функций от n переменных равно 2^{n+1}.

7. Доказать, что пересечение двух функционально замкнутых классов есть функционально замкнутый класс.

8. Доказать, что совокупность функций, двойственных функциям из функционально замкнутого класса, образует функционально замкнутый класс.

9. Доказать, что следующие классы функций не являются функционально замкнутыми:

а) класс функций от двух переменных;

б) класс функций, сохраняющих нуль, но не сохраняющих единицу.

10. Доказать, что объединение функционально замкнутых классов может не быть функционально замкнутым.

11. Доказать, что нельзя выразить с помощью суперпозиций: а) $\supset$ через $\neg$ и $\sim$; б) $\neg$ через $\supset$.

12. Проверить с помощью теоремы Поста полноту следующих систем булевых функций: а) {+, ~}; б) {¬}; в) {0, ⊃}; г) {~, ∨, 0}; д) {|}; е) {o}. Будут ли эти системы функций независимыми?

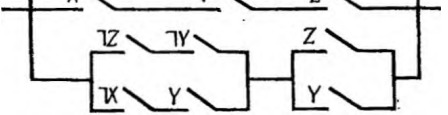
13. Доказать, что единственными полными системами из одной двуместной булевой функции являются системы {|} и {o}.

14. Доказать, что для класса K_1 всех булевых функций базисами являются: а) {∨, ¬}; б) {o}; в) {|}.

15*. Доказать, что {&, ⊃} — базис для класса T_1.

16. Методом минимизирующих карт найти минимальную ДНФ для булевой функции $f(X_1, X_2, X_3)$, которая равна 1 только на оценках <1, 1, 1>, <1, 0, 1>, <0, 0, 1>, <0, 0, 0>.

17. Записать функцию проводимости и упростить схему:



1.3. ИСЧИСЛЕНИЕ ВЫСКАЗЫВАНИЙ

1.3.1. Аксиоматические теории

Таблицы истинности позволяют ответить на многие вопросы, касающиеся формул логики высказываний, например на вопрос о тавтологичности (тождественной истинности) формулы или на вопрос о равносильности двух формул. Однако более сложные вопросы логики высказываний уже не могут быть решены с помощью таблиц истинности. Поэтому рассмотрим другой метод — метод формальных аксиоматических теорий.

Формальная аксиоматическая теория T считается определенной, если:

1) задано некоторое счетное множество символов — символов теории *T*; конечные последовательности символов теории *T* называются выражениями теории *T*;

2) имеется подмножество выражений теории *T*, называемых формулами теории *T*;

3) выделено некоторое множество формул, называемых аксиомами теории T;

4) имеется конечное множество $R_1, \ldots, R_m$ отношений между формулами, называемых *правилами вывода*.

Если формула A и формулы $A_1, \ldots, A_l$ находятся в некотором отношении R_i, то A называется *непосредственным следствием* из формул $A_1, \ldots, A_l$, полученным по правилу R_i.

Выводом в теории T называется всякая последовательность $A_1, \ldots, A_n$ формул такая, что для любого i формула A_i есть либо аксиома теории T, либо непосредственное следствие каких-либо предыдущих формул.

Формула A называется *теоремой* теории T, если в ней существует вывод, в котором последней формулой является A. Этот вывод называется выводом формулы A. Иными словами, теоремы аксиоматической теории — это формулы, которые могут быть выведены по определенным правилам.

Формула A называется *следствием множества формул* Γ тогда и только тогда, когда существует такая последовательность формул $A_1, \ldots, A_n$, что A_n есть A, и для любого i, $1 \leqslant i \leqslant n$, A_i есть либо аксиома, либо формула из Γ, либо непосредственное следствие некоторых предыдущих формул. Эта последовательность называется выводом A из Γ. Элементы Γ называются *посылками* вывода или *гипотезами*. Сокращенно можно записать $\Gamma \vdash A$ (т. е. A есть следствие Γ). Если множество Γ состоит из формул $B_1, \ldots, B_k$, то пишут $B_1, \ldots, B_k \vdash A$. Если Γ — пустое множество, то $\Gamma \vdash A$ тогда и только тогда, когда A есть теорема. В этом случае принято писать $\vdash A$.

Приведем несколько простых свойств понятия выводимости из системы гипотез. Пусть Γ — произвольное множество формул, а A, B, C — произвольные формулы:

I. $\Gamma, A \vdash A$. Действительно, вывод формулы A из системы гипотез Γ, A состоит из одной формулы A.

II. Если $\Gamma, A, B \vdash C$, то $\Gamma, B, A \vdash C$.

III. Если $\Gamma \vdash A$ и B — произвольная формула, то $\Gamma, B \vdash A$. Действительно, в качестве вывода формулы A из системы гипотез Γ, B можно взять вывод формулы A из системы гипотез Γ.

IV. Если $\Gamma \vdash A$, $\Gamma \vdash B$ и $A, B \vdash C$, то $\Gamma \vdash C$. Пусть $A_1, \ldots, A_n$ — вывод формулы A из Γ; $B_1, \ldots, B_m$ — вывод формулы B из Γ; $C_1, \ldots, C_k$ — вывод формулы C из A, B. Тогда очевидно, что $A_1, \ldots, A_n, B_1, \ldots, B_m, C_1, \ldots, C_k$ есть вывод формулы C из Γ.

V. Если $\Gamma, A \vdash B$ и $\Gamma \vdash A$, то $\Gamma \vdash B$ (правило удаления выводимой гипотезы). Пусть $B_1, \ldots, B_n$ — вывод формулы B из Γ, A. Если в этом выводе не встречается формула A, то имеем вывод B из Γ. Если в этом выводе встречается формула A, то пусть $B_{i_1}, \ldots, B_{i_k}$ — все вхождения формулы A в вывод. Пусть также $A_1, \ldots, A_m$ — вывод A из Γ. Тогда $B_1, \ldots, B_{i_1-1}, A_1, \ldots,$

A_m, $B_{i_1+1}, \ldots, B_{i_2-1}$, $A_1, \ldots, A_m$, $B_{i_{k+1}}, \ldots, B_{i_k-1}$, $A_1, \ldots, A_m$, $B_{i_k+1}, \ldots, B_m$ — вывод формулы B из Γ.

1.3.2. Исчисление высказываний: определение, свойства. Теорема дедукции

Рассмотрим один из возможных способов формализации логики высказываний — исчисление высказываний. Оно является простым примером формальной аксиоматической теории. Определим эту формальную аксиоматическую теорию следующим образом:

1. Символы исчисления высказываний: $\neg$, $\supset$, (,) и буквы X_i с целыми положительными числами в качестве индексов: X_1, X_2, X_3 ... Символы $\neg$ и $\supset$ — логические символы, символы X_1, X_2, X_3 ...— переменные.

2. Формулы исчисления высказываний: а) все переменные X_i — формулы; б) если A и B — формулы, то $(\neg A)$ и $(A \supset B)$ тоже формулы.

Пример 1.22. Последовательность символов $\neg X_1 \supset X_2 \supset \supset X_1$ — выражение, но не формула.

Пример 1.23. Пусть A, B, C — формулы. Тогда

$$(C \supset (A \supset B)), \quad (((\neg A) \supset B) \supset (\neg C)) \tag{1.2}$$

тоже формулы.

Для сокращения записи опустим в формуле внешние скобки и те пары скобок, без которых можно восстановить формулу по следующему правилу: каждое вхождение знака $\neg$ относится к наикратчайшей подформуле, следующей за этим знаком. Тогда формулы (1.2) примут вид

$$C \supset (A \supset B), \quad (\neg A \supset B) \supset \neg C.$$

3. Аксиомы исчисления высказываний. Каковы бы ни были формулы A, B и C, следующие формулы являются аксиомами:

A1. $A \supset (B \supset A)$;

A2. $(A \supset (B \supset C)) \supset ((A \supset B) \supset (A \supset C))$;

A3. $(\neg B \supset \neg A) \supset ((\neg B \supset A) \supset B)$.

Выражения A1 — A3 называются схемами аксиом, поскольку каждое из них порождает бесконечное множество формул, являющихся аксиомами исчисления высказываний. Например, формула $X_1 \supset (X_2 \supset X_1)$ есть аксиома, полученная по схеме A1, формула $(\neg A \supset \neg A) \supset ((\neg A \supset A) \supset A)$ (где A — любая формула) — аксиома, полученная по схеме A3.

4. Единственным правилом вывода формулы служит правило *modus ponens* (сокращенно *m. p.*). Пусть имеются три формулы: A, $A \supset B$ и B. Про формулу B будем говорить, что она

получается по правилу вывода *m. p.* из формул A и $A \supset B$. Это правило вывода записывается обычно в виде $\frac{A,\ A \supset B}{B}$.

Хотя для исчисления высказываний мы выбрали только два логических символа $\neg$ и $\supset$, с помощью следующих определений можно ввести и остальные операции: $A \,\&\, B$ означает $\neg(A \supset \neg B)$; $A \vee B$ означает $\neg A \supset B$; $A \sim B$ означает $\neg((B \supset A) \supset \neg(A \supset B))$.

Докажем несколько утверждений о выводимости формул в исчислении высказываний.

Наряду со свойствами выводимости I—V (см. с. 63) в исчислении высказываний выполняется также следующее свойство:

VI. Если $\Gamma \vdash A \supset B$ и $\Gamma \vdash A$, то $\Gamma \vdash B$. Пусть $A_1, \ldots, A_n$ — вывод формулы A из Γ, где A_n совпадает с A. Пусть $B_1, \ldots, B_m$ — вывод формулы $A \supset B$ из Γ, где B_m совпадает с $A \supset B$. Тогда $A_1, \ldots, A_n, B_1, \ldots, B_m, B$ — вывод формулы B из Γ. Последняя формула в этом выводе получена применением правила *m. p.* к формулам A_n и B_m.

Утверждение 1.9. $\vdash A \supset A$ *для любой формулы* A.

Построим вывод формулы $A \supset A$:

(1) $(A \supset ((A \supset A) \supset A)) \supset ((A \supset (A \supset A)) \supset (A \supset A))$

(подстановка в схему аксиом A2 формулы $A \supset A$ вместо B и формулы A вместо C);

(2) $A \supset ((A \supset A) \supset A)$

(подстановка в схему аксиом A1 формулы $A \supset A$ вместо B);

(3) $(A \supset (A \supset A)) \supset (A \supset A)$

(из (1) и (2) по *m. p.*);

(4) $A \supset (A \supset A)$

(подстановка в схему аксиом A1 формулы A вместо B);

(5) $A \supset A$

(из (3) и (4) по *m. p.*).

Утверждение 1.10. *Если* $\Gamma \vdash A$ *и* B *— любая формула, то* $\Gamma \vdash B \supset A$.

Пусть $A_1, \ldots, A_n$ — вывод формулы A из Γ, где A_n совпадает с A. Тогда последовательность $A_1, \ldots, A_n$, $A \supset (B \supset A)$, $B \supset A$ — вывод формулы $B \supset A$ из Γ.

В математических рассуждениях часто какое-то утверждение B доказывают в предположении верности некоторого другого утверждения A, после чего заключают, что верно утверждение «если A, то B». В исчислении высказываний этот прием обосновывается следующей теоремой.

Теорема 1.11 (*о дедукции*). *Пусть* Γ — *множество формул,* A *и* B — *формулы и* $\Gamma, A \vdash B$. *Тогда* $\Gamma \vdash A \supset B$.

Пусть

$$B_1, \ldots, B_n \tag{1.3}$$

есть вывод формулы B из Γ и A. Доказательство проведем индукцией по n — длине вывода (1.3). При $n = 1$ формула B совпадает с B_1. Согласно определению вывода возможны три случая:

1) B_1 — аксиома;

2) B_1 — формула из множества Γ;

3) B_1 совпадает с A.

В первых двух случаях имеем $\Gamma \vdash B_1$. Тогда согласно утверждению 1.10 $\Gamma \vdash A \supset B_1$, т. е. $\Gamma \vdash A \supset B$. В третьем случае формула $A \supset B$ имеет вид $A \supset A$. Согласно утверждению 1.9 $\vdash A \supset A$, откуда $\Gamma \vdash A \supset A$.

Допустим теперь, что если длина вывода формулы B из Γ, A меньше n, то утверждение теоремы верно. Докажем его для случая, когда длина вывода (1.3) равна n. При этом возможно, что:

1) B_n — аксиома;

2) B_n — формула из Γ;

3) B_n совпадает с A;

4) B_n получена по *т. р.* из B_i и B_j, где $i < n$, $j < n$.

В первых трех случаях доказательство проводится так же, как при $n = 1$. В четвертом случае либо формула B_j имеет вид $B_i \supset B_n$, либо формула B_i имеет вид $B_j \supset B_n$. Ограничимся рассмотрением случая, когда B_j имеет вид $B_i \supset B_n$. Отбрасывая последние $n - i$ формулы из (1.3), получаем вывод B_i из Γ и A, а отбрасывая последние $n - j$ формулы из (1.3), — вывод B_j из Γ и A; длины этих выводов меньше n. По индуктивному предположению

(1) $\Gamma \vdash A \supset B_i$;

(2) $\Gamma \vdash A \supset (B_i \supset B_n)$.

По схеме аксиом A2 имеем

(3) $\Gamma \vdash (A \supset (B_i \supset B_n)) \supset ((A \supset B_i) \supset (A \supset B_n))$.

Применяя свойство VI к (2) и (3), получаем

(4) $\Gamma \vdash (A \supset B_i) \supset (A \supset B_n)$.

Применяя свойство VI к (1) и (4), получаем

$\Gamma \vdash A \supset B_n$.

Теорема доказана.

Следствие 1 (*правило силлогизма*): $A \supset B$, $B \supset C \vdash A \supset C$.

Построим вывод:

(1) $A \supset B$ — гипотеза;

(2) $B \supset C$ — гипотеза;

(3) A — гипотеза;

(4) B — применяем правило *т. р.* к (1) и (3);
(5) C — применяем правило *т. р.* к (2) и (4).

Тогда $A \supset B$, $B \supset C$, $A \vdash C$. По теореме о дедукции $A \supset B$, $B \supset C \vdash A \supset C$.

Следствие 2: $A \supset (B \supset C)$, $B \vdash A \supset C$.

В результате двукратного применения правила *т. р.* получаем A, $A \supset (B \supset C)$, $B \vdash C$. Отсюда по теореме о дедукции $A \supset (B \supset C)$, $B \vdash A \supset C$.

Для любых формул A и B в исчислении высказываний верны следующие утверждения (приведем их без доказательств).

Утверждение 1.11. $\vdash \neg\neg A \supset A$.
Утверждение 1.12. $\vdash A \supset \neg\neg A$.
Утверждение 1.13. $\vdash \neg A \supset (A \supset B)$.
Утверждение 1.14. $\vdash (\neg B \supset \neg A) \supset (A \supset B)$.
Утверждение 1.15. $\vdash (A \supset B) \supset (\neg B \supset \neg A)$.
Утверждение 1.16. $\vdash A \supset (\neg B \supset \neg (A \supset B))$.
Утверждение 1.17. $\vdash (A \supset B) \supset ((\neg A \supset B) \supset B)$.

Следствие. Если Γ, $A \vdash B$ и Γ, $\neg A \vdash B$, то $\Gamma \vdash B$.

По теореме о дедукции $\Gamma \vdash A \supset B$ и $\Gamma \vdash \neg A \supset B$. Отсюда в силу утверждения 1.17 $\Gamma \vdash B$.

1.3.3. Полнота и непротиворечивость исчисления высказываний. Независимость аксиом

Мы формализовали логику высказываний и построили исчисление высказываний как аксиоматическую теорию. Покажем, что множество теорем исчисления высказываний совпадает с множеством тождественно-истинных формул логики высказываний.

Докажем сначала следующее утверждение.

Теорема 1.12. *Всякая выводимая (из пустой системы гипотез) формула исчисления высказываний тождественно-истинна.*

Непосредственной проверкой убедимся в том, что аксиомы исчисления высказываний — тождественно-истинные формулы. В силу свойств импликации формула, получающаяся из тождественно-истинных формул по правилу *т. р.*, тождественно-истинна. Следовательно, любая выводимая формула тождественно-истинна.

Докажем теперь обратное теореме 1.12 утверждение о том, что любая тождественно-истинная формула выводима в исчислении высказываний, т. е. является теоремой.

Символы X^ε_i и A^ε имеют здесь тот же смысл, что и в разд. 1.2.1:

$$X^\varepsilon_i = \begin{cases} X_i, \text{ если } \varepsilon = 1; \\ \neg X_i, \text{ если } \varepsilon = 0; \end{cases} \qquad A^\varepsilon = \begin{cases} A, \text{ если } \varepsilon = 1; \\ \neg A, \text{ если } \varepsilon = 0. \end{cases}$$

Вместо символа И употребляем символ 1, вместо символа Л — символ 0.

Пусть $X_1, \ldots, X_n$ — все переменные формулы A и задана некоторая оценка списка переменных $\langle \varepsilon_1, \ldots, \varepsilon_n \rangle$, состоящая

из нулей и единиц. Через $A(\varepsilon_1, \ldots, \varepsilon_n)$ будем обозначать значение формулы A на этой оценке.

Пример 1.24. Пусть формула A имеет вид $(X_1 \supset X_2) \supset \supset \neg X_1$, где X_1 и X_2 — переменные. Тогда $X^0{}_1$ — это $\neg X_1$, $X^1{}_2$ — это X_2, формула A^1 — это A, формула A^0 — это $\neg((X_1 \supset X_2) \supset \supset \neg X_1)$.

Рассмотрим все четыре возможные оценки, соответствующие различным распределениям истинностных значений переменных: $\langle 0, 0\rangle$, $\langle 0, 1\rangle$, $\langle 1, 0\rangle$, $\langle 1, 1\rangle$. В нашем случае $A(0, 0) = 1$, $A(0, 1) = 1$, $A(1, 0) = 1$, $A(1, 1) = 0$.

Лемма 1.7. *Пусть A — произвольная формула, $X_1, \ldots, X_n$ — все переменные, входящие в A (для удобства мы обозначили их так, как если бы они были первыми переменными), и пусть также задана произвольная оценка списка переменных $\langle \varepsilon_1, \ldots, \varepsilon_n\rangle$. Тогда $X^{\varepsilon_1}{}_1, \ldots, X^{\varepsilon_n}{}_n \vdash A^{\varepsilon}$, где ε — значение формулы A на оценке $\langle \varepsilon_1, \ldots, \varepsilon_n\rangle$.*

Пример 1.25. Пусть A — формула, рассмотренная в примере 1.24. Тогда из леммы 1.7 следуют выводимости $X_1, \neg X_2 \vdash (X_1 \supset X_2) \supset \neg X_1$; $X_1, X_2 \vdash \neg((X_1 \supset X_2) \supset \neg X_1))$.

Докажем лемму 1.7. Под длиной формулы A будем понимать число вхождений логических символов в A. Доказательство проведем индукцией по k — длине формулы A. При $k = 0$ формула A представляет собой переменную X_i, и утверждение леммы сводится к $X^{\varepsilon}{}_i \vdash X^{\varepsilon}{}_i$.

Пусть для формул, длина которых меньше k, утверждение леммы справедливо. Докажем его для формул длины k. Возможны два случая.

Случай 1. Формула A имеет вид $\neg B$. Длина формулы B равна $k - 1$, множество переменных формул B совпадает с множеством переменных формулы A. Пусть $B(\varepsilon_1, \ldots, \varepsilon_n) = \varepsilon'$. Тогда $\varepsilon = \neg\varepsilon'$.

Если $\varepsilon' = 0$, то $\varepsilon = 1$. По индуктивному предположению

$$X^{\varepsilon_1}{}_1, \ldots, X^{\varepsilon_n}{}_n \vdash B^0.$$

Но B^0 — это $\neg B$ или в данном случае A^1. Следовательно,

$$X^{\varepsilon_1}{}_1, \ldots, X^{\varepsilon_n}{}_n \vdash A^{\varepsilon}.$$

Если $\varepsilon' = 1$, то $\varepsilon = 0$. По индуктивному предположению

$$X^{\varepsilon_1}{}_1, \ldots, X^{\varepsilon_n}{}_n \vdash B^1.$$

Согласно утверждению 1.12 и правилу *т. р.*

$$X^{\varepsilon_1}{}_1, \ldots, X^{\varepsilon_n}{}_n \vdash \neg\neg B.$$

Но $\neg\neg B$ — это и есть A^0. Следовательно,

$$X^{\varepsilon_1}{}_1, \ldots, X^{\varepsilon_n}{}_n \vdash A^{\varepsilon}.$$

Случай 2. Формула A имеет вид $B \supset C$. Длина формул

и C меньше k. Пусть значение формулы B на оценке $\langle \varepsilon_1, \ldots, \varepsilon_n \rangle$ равно ε', значение формулы C на оценке $\langle \varepsilon_1, \ldots, \varepsilon_n \rangle$ равно ε''. Тогда $\varepsilon = \varepsilon' \supset \varepsilon''$.

Если $\varepsilon' = 0$, то $\varepsilon = 1$. По индуктивному предположению

$$X^{\varepsilon_1}_1, \ldots, X^{\varepsilon_n}_n \vdash B^{\varepsilon'},$$

т. е.

$$X^{\varepsilon_1}_1, \ldots, X^{\varepsilon_n}_n \vdash \neg B.$$

Согласно утверждению 1.13 $\vdash \neg B \supset (B \supset C)$. Применяя правило *m. p.*, получаем

$$X^{\varepsilon_1}_1, \ldots, X^{\varepsilon_n}_n \vdash B \supset C.$$

Следовательно,

$$X^{\varepsilon_1}_1, \ldots, X^{\varepsilon_n}_n \vdash A^{\varepsilon}.$$

Пусть теперь $\varepsilon' = 1$, $\varepsilon'' = 0$. Тогда $\varepsilon = 0$ и A^{ε} — это $\neg (B \supset C)$. По индуктивному предположению

$$X^{\varepsilon_1}_1, \ldots, X^{\varepsilon_n}_n \vdash B, \; X^{\varepsilon_1}_1, \ldots, X^{\varepsilon_n}_n \vdash \neg C.$$

Согласно утверждению 1.16 $\vdash B \supset (\neg C \supset \neg (B \supset C))$. В результате двукратного применения правила *m. p.* получаем

$$X^{\varepsilon_1}_1, \ldots, X^{\varepsilon_n}_n \vdash \neg (B \supset C).$$

Если $\varepsilon' = 1$, $\varepsilon'' = 1$, то $\varepsilon = 1$ и A^{ε} — это $B \supset C$. По индуктивному предположению

$$X^{\varepsilon_1}_1, \ldots, X^{\varepsilon_n}_n \vdash C.$$

По схеме аксиом A1 имеем $\vdash C \supset (B \supset C)$. Применяя правило *m. p.*, получаем

$$X^{\varepsilon_1}_1, \ldots, X^{\varepsilon_n}_n \vdash B \supset C.$$

Теорема 1.13. *Если формула A исчисления высказываний является тождественно-истинной, то она выводима.*

Пусть A — тождественно-истинная формула, а $X_1, \ldots, X_n$ — все ее переменные. Тогда в силу леммы 1.7 на любой оценке $\langle \varepsilon_1, \ldots, \varepsilon_n \rangle$, состоящей из нулей и единиц, получаем

$$X^{\varepsilon_1}_1, \ldots, X^{\varepsilon_n}_n \vdash A.$$

Поэтому в случае, когда $\varepsilon_n = 1$, имеем

$$X^{\varepsilon_1}_1, \ldots, X^{\varepsilon_{n-1}}_{n-1}, \; X_n \vdash A,$$

а в случае, когда $\varepsilon = 0$, имеем

$$X^{\varepsilon_1}_1, \ldots, X^{\varepsilon_{n-1}}_{n-1}, \; \neg X_n \vdash A.$$

Применяя следствие утверждения 1.17, получаем

$$X^{\varepsilon_1}_1, \ldots, X^{\varepsilon_{n-1}}_{n-1} \vdash A.$$

Точно так же, рассматривая случаи, когда ε_{n-1} принимает значения 1 и 0, получаем

$$X_1^{\varepsilon_1}, \ldots, X_{n-2}^{\varepsilon_{n-2}} \vdash A$$

и т. д. Наконец, приходим к $\vdash A$.

Из теорем 1.12 и 1.13 непосредственно вытекает следующее утверждение.

Теорема 1.14. *Формула A исчисления высказываний выводима тогда и только тогда, когда она тождественно-истинна.*

Свойство аксиоматической теории, заключающееся в том, что если формула A выражает логический закон (как, например, тождественно-истинная формула), то она выводима в этой теории, называется *полнотой* аксиоматической теории (или полнотой в широком смысле). Из теоремы 1.13 следует, что исчисление высказываний есть полная аксиоматическая теория.

Формальную аксиоматическую теорию называют *непротиворечивой*, если не существует формулы A такой, что одновременно выводимы формулы A и $\neg A$.

Теорема 1.15. *Исчисление высказываний непротиворечиво.*

Действительно, согласно теореме 1.12 всякая выводимая формула тождественно-истинна. Отрицание этой формулы не является тождественно-истинной формулой. Следовательно, ни для какой формулы A невозможно, чтобы одновременно $\vdash A$ и $\vdash \neg A$.

Наряду с полнотой аксиоматической теории в широком смысле рассматривают ее полноту в узком смысле. Формальную аксиоматическую теорию называют *полной в узком смысле*, если добавление любой невыводимой формулы в качестве схемы аксиом приводит к противоречивой теории.

Теорема 1.16. *Исчисление высказываний полно в узком смысле.*

Пусть F — произвольная невыводимая формула (согласно теореме 1.15 в качестве F можно взять любую не тождественно-истинную формулу), $X_1, \ldots, X_n$ — список ее переменных, а $<\varepsilon_1, \ldots, \varepsilon_n>$ — такая оценка списка переменных, на которой формула F принимает ложное значение (т. е. 0). Пусть также

$$B_i = \begin{cases} A \supset A, \text{ если } \varepsilon_i = 1; \\ \neg (A \supset A), \text{ если } \varepsilon_i = 0, \end{cases}$$

где A — произвольная формула. Тогда формула $F(B_1, \ldots, B_n)$ будет тождественно-ложной.

Рассмотрим исчисление, в котором к схемам аксиом А1 — А3 в качестве еще одной схемы аксиом добавлена формула F:

А4. $F(A_1, \ldots, A_n)$.

Выводимость в этом исчислении обозначим символом $\vdash_F$. Поскольку $F(B_1, \ldots, B_n)$ получена по схеме А4, то $\vdash_F F(B_1, \ldots$

B_n). Пусть C — произвольная формула. Тогда $F(B_1, \ldots, B_n) \supset \supset C$ — тождественно-истинная формула в силу свойств импликации. Следовательно, $\vdash F(B_1, \ldots, B_n) \supset C$, а значит, $\underset{F}{\vdash} F(B_1, \ldots, B_n) \supset C$. Применив правило *m. p.*, получим, что $\underset{F}{\vdash} C$. Если вместо формулы C рассмотреть формулу $\neg C$, то также получим, что $\underset{F}{\vdash} \neg C$, т. е. расширенная теория оказалась противоречивой.

Во всякой формальной теории возникает вопрос о независимости ее аксиом, т. е. вопрос о том, можно ли какую-нибудь аксиому вывести из остальных, применяя правила вывода данной системы.

Оказывается, что система аксиом А1 — А3 исчисления высказываний независима. Установим независимость аксиомы А3 от остальных. Будем считать, что переменные принимают значения из множества $\{a, b\}$. Операции $\neg$ и $\supset$ зададим табл. 1.17 и 1.18 соответственно.

Легко проверить, что аксиомы А1 и А2 при такой интерпретации примут значение a. Из определения $\supset$ вытекает, что применение правила вывода *m. p.* к формулам, тождественно-равным a, дает формулу, также тождественно-равную a. Следовательно, формулы, выводимые из аксиом А1 и А2 по правилу *m. p.*, тождественно-равны a.

С другой стороны, аксиома А3

$$(\neg B \supset \neg A) \supset ((\neg B \supset A) \supset B)$$

не равна тождественно a, так как, например, при $A = a$, $B = b$ принимает значение b.

Можно установить независимость каждой из аксиом А1 и А2 от остальных (при этом удобно использовать трехзначные логики).

Следует заметить, что исчисление высказываний можно описать системами аксиом, отличными от А1 — А3. Кроме того, вместо символов $\neg$ и $\supset$ можно использовать и другие наборы символов, лишь бы с их помощью можно было выразить все остальные логические операции.

Таблица 1.17

X	$\neg X$
a	?
b	

Таблица 1.18

X	Y	$X \supset Y$
a	a	a
a	b	b
b	a	a
b	b	a

1.4. ЛОГИКА И ИСЧИСЛЕНИЕ ПРЕДИКАТОВ

Мы рассмотрели два различных способа описания логики высказываний — содержательное и аксиоматическое.

Логика высказываний — очень узкая логическая система. Есть такие типы логических рассуждений, которые не могут быть осуществлены в рамках логики высказываний, например:

1. Всякий друг Ивана есть друг Петра. Сидор не есть друг Петра. Следовательно, Сидор не есть друг Ивана.

2. Простое число два — четное. Следовательно, существуют простые четные числа.

Корректность этих умозаключений основана на внутренней структуре самих предложений и на смысле слов «всякий» и «существуют».

1.4.1. Предикаты, кванторы. Формулы логики предикатов

Рассмотрим предложения, зависящие от параметров, например: «x — четное число», «x меньше y», «$x + y = z$», «x — отец y», «x и y — братья» и т. п. Если x, y, z в первых трех предложениях заменить некоторыми числами, то получим определенные высказывания, которые могут быть истинными или ложными. Например: «3 — четное число», «2 меньше 5», «$3 + 2 = 7$». Последние два предложения выражают родственные отношения между членами семьи и также превращаются в определенные высказывания, истинные или ложные, при замене x и y именами членов этой семьи: «Иван — отец Петра», «Иван и Олег — братья».

Предложения такого типа называются предикатами. Точнее, *предикатом* $P(x_1, \ldots, x_n)$ называется функция, переменные которой принимают значения из некоторого множества M, а сама она принимает два значения: И (истинное) и Л (ложное), т. е. $P(x_1, \ldots, x_n) : M_n \to \{И, Л\}$.

Предикат от n аргументов называют n-местным предикатом. Множество M значений переменных определяется обычно математическим контекстом. Например, основное соотношение элементарной геометрии на плоскости — точки x, y, z лежат на одной прямой — выражается предикатом $L(x, y, z)$, где в качестве значений x, y и z рассматриваются конкретные точки.

Предикаты обозначаются большими буквами латинского алфавита. Иногда бывает удобно указывать число переменных у предикатов. В таких случаях у символов предикатов пишут верхний индекс, который и указывает число аргументов, например: $P^{(n)}(x_1, \ldots, x_n)$ — n-местный предикат. Высказывания считаются нуль-местными предикатами.

Над предикатами можно производить обычные логические операции. В результате получаются новые предикаты.

Пример 1.26.

1. Пусть $P^{(1)}(x)$ означает предикат «x делится на два», $Q^{(1)}(x)$ — предикат «x делится на три». Тогда выражение $P^{(1)}(x) \,\&\, Q^{(1)}(x)$ означает предикат «x делится на два и x делится на три», т. е. определяет предикат делимости на 6.

2. Пусть $S^{(2)}(x, y)$ означает предикат «$x=y$». Он принимает значение И тогда и только тогда, когда $x=y$. В этом случае выражение $\neg S^{(2)}(x, x) \supset S^{(2)}(x, y)$ определяет предикат, принимающий значение И при любых x и y.

Кроме операций логики высказываний будем применять еще операции связывания квантором.

Квантор общности. Пусть $P(x)$ — некоторый предикат, принимающий значение И или Л для каждого элемента x множества M. Тогда под выражением $(\forall x)P(x)$ будем подразумевать высказывание истинное, когда $P(x)$ истинно для каждого элемента x из множества M, и ложное — в противном случае. Читается это выражение так: «для всех x $P(x)$». Это высказывание уже не зависит от x. Символ $\forall x$ называется квантором общности.

Квантор существования. Пусть $P(x)$ — некоторый предикат. Под выражением $(\exists x)P(x)$ будем понимать высказывание истинное, когда существует элемент множества M, для которого $P(x)$ истинно, и ложное — в противном случае. Читается это выражение так: «существует x такое, что $P(x)$» или «существует x, для которого $P(x)$». Символ $\exists x$ называется квантором существования.

Операцию связывания квантором можно применять и к предикатам от большого числа переменных (подробнее об этом будет сказано позже).

Пример 1.27. Для предикатов, приведенных в примере 1.26, имеем: $(\exists x)(P^{(1)}(x) \,\&\, Q^{(1)}(x))$ — истинное высказывание; $(\forall x)(P^{(1)}(x) \,\&\, Q^{(1)}(x))$ — ложное высказывание.

На языке предикатов можно составить гораздо более сложные предложения, чем на языке логики высказываний.

Определим понятие формулы логики предикатов. Алфавит логики предикатов содержит следующие символы:

1) символы предметных переменных: $x_1, x_2, \ldots, x_n, \ldots$;

2) символы предикатов: $A^{(t)}_1, A^{(t)}_2, \ldots, A^{(t)}_k, \ldots$, где $t=0, 1, 2, \ldots$;

3) логические символы: $\neg$, $\&$, $\vee$, $\supset$, $\sim$;

4) символы кванторов: $\exists$, $\forall$;

5) скобки и запятую:), (.

Во избежание нагромождения индексов часто символы предметных переменных будем обозначать через x, y, z, а символы предикатов — через P, S, Q, R и т. д.

Слово в алфавите логики предикатов называется *формулой*, если оно удовлетворяет следующему индуктивному определению

(одновременно определяется понятие свободной и связанной переменной формулы):

1. Если $A^{(n)}_l$ — символ предиката, $x_{l_1}, x_{l_2}, \ldots, x_{l_n}$ — символы предметных переменных, не обязательно различные, то $A^{(n)}_l(x_{l_1}, x_{l_2}, \ldots, x_{l_n})$ — формула. Такая формула называется *атомарной*. Все предметные переменные атомарных формул свободные, связанных переменных нет.

2. Пусть A — формула. Тогда $(\neg A)$ тоже формула. Свободные и связанные переменные формулы $(\neg A)$ — это соответственно свободные и связанные переменные формулы A.

3. Пусть A и B — формулы, причем нет таких предметных переменных, которые были бы связаны в одной формуле и свободны — в другой. Тогда

$$(A \vee B),\ (A \& B),\ (A \supset B],\ (A \sim B) \tag{1.4}$$

есть формулы, в которых свободные переменные формул A и B остаются свободными, а связанные переменные формул A и B остаются связанными.

4. Пусть A — формула, содержащая свободную переменную x. Тогда

$$(\forall x)A,\ (\exists x)A \tag{1.5}$$

тоже формулы. Переменная x в них связана. Остальные же переменные, которые в формуле A свободны, остаются свободными и в формулах (1.5). Переменные, которые в формуле A связаны, остаются связанными и в формулах (1.5). В первой из формул (1.5) формула A называется *областью действия* квантора $\forall x$, а во второй — *областью действия* квантора $\exists x$.

5. Слово в алфавите логики предикатов 1—5 является формулой только в том случае, если это следует из правил 1—4.

Заметим, что по определению формулы никакая переменная не может быть одновременно свободной и связанной.

Оставим в силе принятое в разд. 1.1.1 соглашение об опускании скобок.

Пример 1.28.

1. Следующие выражения являются формулами логики предикатов: $A^{(3)}_5(x_1, x_5, x_7)$ — атомарная формула, в которой x_1, x_5, x_7 — свободные переменные; $(\forall x_1)(\exists x_2)A^{(3)}_1(x_1, x_2, x_3) \supset (\forall x_1)A^{(2)}_1(x_1, x_4)$ — формула, в которой x_1, x_2 — связанные, а x_3, x_4 — свободные переменные.

2. Выражение $(\exists x_1)(\forall x_2)A^{(2)}_1(x_1, x_3) \& A^{(2)}_2(x_1, x_2)$ не является формулой.

Значение формулы определено лишь тогда, когда задана какая-нибудь интерпретация входящих в нее символов.

Под *интерпретацией* понимают систему $M = \langle M, f \rangle$, состоящую из непустого множества M и соответствия f, сопоставляющего каждому предикатному символу $A^{(n)}_l$ определенный

l-местный предикат (будем обозначать предикаты, поставленные в соответствие предикатным символам, теми же символами).

При заданной интерпретации считают, что предметные переменные пробегают множество M, а символы $\neg$, $\vee$, $\&$, $\supset$, $\sim$ и символы кванторов имеют свой обычный смысл. Для данной интерпретации каждая формула без свободных переменных представляет собой высказывание, которое истинно или ложно, а всякая формула со свободными переменными выражает некоторый предикат на множестве M, который истинен при одних значениях переменных из этого множества и ложен при других.

Определим значение формулы в данной интерпретации, следуя индуктивным шагам определения формулы. Значение формулы F на наборе $<a_1, \ldots, a_n>$, $a_i \in M$, своих свободных переменных $x_{i_1}, \ldots, x_{i_n}$ обозначим символом $F|_{<a_1, \ldots, a_n>}$.

1. Формула F — атомарная формула $A^{(l)}_j(x_{i_1}, \ldots, x_{i_l})$. Пусть $x_{j_1}, \ldots, x_{j_s}$ — все различные свободные переменные этой формулы, выписанные в определенном порядке. Значением формулы F на наборе $<a_1, \ldots, a_s>$, $a_i \in M$, называется значение l-местного предиката, сопоставленного символу $A^{(l)}_j$ при соответствующем замещении его переменных элементами $a_1, \ldots, a_s$.

2. Формула F имеет вид $\neg A$. Пусть значение формулы A на наборе $<a_1, \ldots, a_n>$, $a_i \in M$, есть ε. Тогда $F|_{<a_1, \ldots, a_n>} = \neg \varepsilon$.

3. Формула F имеет вид $A \vee B$, $A \& B$, $A \supset B$ или $A \sim B$. Значение формулы F на наборе $<a_1, \ldots, a_n>$ значений своих свободных переменных есть соответственно $\varepsilon_1 \vee \varepsilon_2$, $\varepsilon_1 \& \varepsilon_2$, $\varepsilon_1 \supset \varepsilon_2$, $\varepsilon_1 \sim \varepsilon_2$, где ε_1 — значение формулы A, а ε_2 — значение формулы B на этом наборе.

4. Формула F имеет вид $(\forall x)A$. Если $x_{i_1}, \ldots, x_{i_n}$ — совокупность всех свободных переменных формулы F, то $x, x_{i_1}, \ldots, x_{i_n}$ — все свободные переменные формулы A. Значение $(\forall x)A|_{<a_1, \ldots, a_n>} = И$ тогда и только тогда, когда для любого $a \in M$ $A|_{<a, a_1, \ldots, a_n>} = И$.

5. Формула F имеет вид $(\exists x)A$. Если $x_{i_1}, \ldots, x_{i_n}$ — совокупность всех свободных переменных формулы F, то $x, x_{i_1}, \ldots, x_{i_n}$ — все свободные переменные формулы A. Значение $(\exists x)A|_{<a_1, a_2, \ldots, a_n>} = И$ тогда и только тогда, когда для некоторого $a \in M$ $A|_{<a, a_1, \ldots, a_n>} = И$.

Пример 1.29. Рассмотрим три формулы:

1) $A^{(2)}_1(x_1, x_2)$;

2) $(\forall x_2)A^{(2)}_1(x_1, x_2)$;

3) $(\exists x_2)(\forall x_1)A^{(2)}_1(x_2, x_1)$.

Возьмем в качестве области интерпретации множество целых положительных чисел и интерпретируем $A^{(2)}_1(x, y)$ как

$x \leqslant y$. Тогда первая формула — это предикат $x_1 \leqslant x_2$, который принимает истинное значение для всех пар a, b целых положительных чисел таких, что $a \leqslant b$. Вторая формула выражает свойство: «для каждого целого положительного числа y $x \leqslant y$», которое выполняется только при $x = 1$. Наконец, третья формула — это истинное высказывание о существовании наименьшего целого положительного числа. Если бы в качестве области интерпретации мы рассматривали множество целых чисел, то третья формула была бы ложным высказыванием.

Пример 1.30. Пусть $M = \langle N, f \rangle$, где N — множество натуральных чисел, f — соответствие, сопоставляющее предикатным символам $S^{(3)}(x, y, z)$, $P^{(3)}(x, y, z)$ следующие предикаты: $S^{(3)}(x, y, z)$; $x + y = z$; $P^{(3)}(x, y, z) : x \cdot y = z$.

Запишем формулы, истинные в M тогда и только тогда, когда выполнены следующие условия:

а) $x = 0$; б) $x = 1$; в) x — четное число; г) x — простое число; д) $x = y$; е) $x \leqslant y$; ж) x делит y; з) коммутативность сложения.

Ответы: а) $F_1(x) = (\forall y) S^{(3)}(x, y, y)$, так как $x + y = y$ для любого y тогда и только тогда, когда $x = 0$; б) $F_2(x) = (\forall y) P^{(3)}(x, y, y)$; в) $F_3(x) = (\exists y) S^{(3)}(y, y, x)$; г) $F_4(x) = \neg F_1(x) \,\&\, (\forall y)(\forall z)(P^{(3)}(y, z, x) \supset (F_2(y) \vee F_2(z)))$, где F_1, F_2 — формулы, определенные в пп. «а» и «б»; д) $F_5(x) = (\forall z)(\forall u)(S^{(3)}(x, z, u) \supset S^{(3)}(y, z, u))$; е) $F_6(x, y) = (\exists z) S^{(3)}(x, z, y)$; ж) $F_7(x, y) = (\exists z) P^{(3)}(x, z, y)$; з) $(\forall x) \times (\forall y)(\forall z)(S^{(3)}(x, y, z) \supset S^{(3)}(y, x, z))$.

Пример 1.31. Пусть $f(x)$ — произвольная фиксированная функция, заданная на отрезке $[a, b]$.

1. Рассмотрим интерпретацию $M = \langle M, f_1 \rangle$, где M — множество действительных чисел; f_1 — соответствие, сопоставляющее предикатным символам $P(x, \delta)$, $Q(x, \varepsilon)$ и $R(\varepsilon)$ предикаты $P(x, \delta) : |x - x_0| < \delta$, $Q(x, \varepsilon) : |f(x) - A| < \varepsilon$; $R(\varepsilon) : \varepsilon > 0$. Здесь x_0 — фиксированный элемент отрезка $[a, b]$; A — некоторое фиксированное действительное число. Тогда утверждение о том, что число A — предел функции $f(x)$ при $x \to x_0$, записывается формулой

$$(\forall \varepsilon)(\exists \delta)(\forall x)((R(\varepsilon) \,\&\, P(x, \delta)) \supset Q(x, \varepsilon)).$$

2. Рассмотрим интерпретацию $M = \langle M, f_2 \rangle$, где M — множество действительных чисел; f_2 — соответствие, сопоставляющее предикатным символам $P(x, \delta)$, $R(\varepsilon)$, $S(x, \varepsilon)$ предикаты $P(x, \delta) : |x - x_0| < \delta$, $R(\varepsilon) : \varepsilon > 0$, $S(x, \varepsilon) : |f(x) - f(x_0)| < \varepsilon$. Здесь x_0 — произвольный фиксированный элемент отрезка $[a, b]$. Тогда утверждение о том, что функция $f(x)$ непрерывна в точке x_0, записывается формулой

$$(\forall \varepsilon)(\exists \delta)(\forall x)((R(\varepsilon) \,\&\, P(x, \delta)) \supset S(x, \varepsilon)).$$

3. Рассмотрим интерпретацию $M = \langle M, f_3 \rangle$, где M — множество действительных чисел; f_3 — соответствие, сопоставляющее предикатным символам $P_1(x, x_1, \delta)$, $R(\varepsilon)$, $S_1(x, x_1, \delta)$, $D(x)$

предикаты $P_1(x, x_1, \delta) : |x - x_1| < \delta$; $R(\varepsilon) : \varepsilon > 0$; $S_1(x, x_1, \varepsilon) : |f(x) - f(x_1)| < \varepsilon$; $D(x) : x \in [a, b]$. Тогда утверждение о том, что функция $f(x)$ непрерывна на отрезке $[a, b]$, записывается формулой

$$(\forall x_1)\,(\forall \varepsilon)\,(\exists \delta)\,(\forall x)\,((D(x_1) \& R(\varepsilon) \& P_1(x, x_1, \delta)) \supset S_1(x, x_1, \varepsilon)).$$

1.4.2. Равносильность формул

Пусть формулы F и G имеют одно и то же множество свободных переменных (в частности, пустое).

Формулы F и G *равносильны в данной интерпретации*, если на любом наборе значений свободных переменных они принимают одинаковые значения (т. е. если формулы выражают в данной интерпретации один и тот же предикат).

Формулы F и G *равносильны на множестве* M, если они равносильны во всех интерпретациях, заданных на множестве M.

Формулы F и G *равносильны* (в логике предикатов), если они равносильны на всех множествах (тогда будем писать $F \equiv G$).

Пример 1.32.

1. На множестве $M = \{a, b\}$ зададим предикаты $P_1(x, y)$ и $P_2(x, y)$ (см. табл. 1.19 и 1.20).

Рассмотрим две формулы:

$$A^{(2)}_1(x_1, x_2) \& A^{(2)}_1(x_1, x_3); \tag{1.6}$$

$$A^{(2)}_1(x_1, x_2) \& A^{(2)}_1(x_2, x_3). \tag{1.7}$$

Таблица 1.19

x	y	$P_1(x, y)$
a	a	И
a	b	Л
b	a	Л
b	b	И

Таблица 1.20

x	y	$P_2(x, y)$
a	a	И
a	b	И
b	a	Л
b	b	Л

Если областью интерпретации служит множество M и формула $A^{(2)}_1$ интерпретируется как предикат P_1, то формулы (1.6) и (1.7) равносильны в этой интерпретации, так как принимают значение И только на двух наборах свободных переменных $<a, a, a>$ и $<b, b, b>$.

Если областью интерпретации является то же множество M, но формула $A^{(2)}_1$ интерпретируется как предикат P_2, то формулы (1.6) и (1.7) не равносильны, так как на наборе $<a, b, b>$ формула (1.6) принимает значение И, а формула (1.7) — значение Л.

2. Формулы $(\forall x_1)A^{(1)}_1(x_1)$ и $(\exists x_1)A^{(1)}_1(x_1)$ равносильны на одноэлементном множестве. Действительно, если область интерпретации является одноэлементным множеством, то какой бы предикат мы не взяли в качестве интерпретации $A^{(1)}_1$ на этом множестве, он принимает только одно значение: И или Л. В первом случае обе формулы принимают значение И, во втором — Л, и, следовательно, они равносильны на этом множестве.

С другой стороны, на двухэлементном множестве $\{a, b\}$ эти формулы не равносильны. Достаточно в качестве интерпретации $A^{(1)}_1$ рассмотреть предикат P такой, что $P(a) = $ И, $P(b) = $ Л.

Укажем несколько правил перехода от одних формул к другим, им равносильным (во всех интерпретациях).

Очевидно, что для формул логики предикатов сохраняются все равносильности и правила равносильных преобразований логики высказываний. Кроме того, можно доказать следующие правила:

1. *Перенос квантора через отрицание.* Пусть A — формула, содержащая свободную переменную x. Тогда справедливы равносильности

$$\neg(\forall x)A(x) \equiv (\exists x)\neg A(x); \tag{1.8}$$

$$\neg(\exists x)A(x) \equiv (\forall x)\neg A(x). \tag{1.9}$$

Докажем сначала равносильность (1.8). Пусть $x_{i_1}, \ldots, x_{i_n}$ — множество (быть может, пустое) всех свободных переменных формулы A, отличных от x. Пусть $M = \langle M, f\rangle$ — произвольная интерпретация. Докажем, что на любом наборе значений своих свободных переменных $\langle a_1, \ldots, a_n\rangle$, $a_i \in M$, формулы $\neg(\forall x)A(x)$ и $(\exists x)\neg A(x)$ принимают одинаковые истинностные значения.

Возможны два случая:

1) для любого элемента $a \in M$ $A(x)|_{\langle a, a_1, \ldots, a_n\rangle} = $ И;

2) для некоторого элемента $a_0 \in M$ $A(x)|_{\langle a_0, a_1, \ldots, a_n\rangle} = $ Л.

В первом случае для любого элемента $a \in M$ $\neg A(x)|_{\langle a, a_1, \ldots, a_n\rangle} = $ Л. Отсюда по определению $(\exists x)\neg A(x)|_{\langle a_1, \ldots, a_n\rangle} = $ Л. С другой стороны, в этом случае $(\forall x)A(x)|_{\langle a_1, \ldots, a_n\rangle} = $ И. Отсюда $\neg(\forall x)A(x)|_{\langle a_1, \ldots, a_n\rangle} = $ Л.

Во втором случае для элемента $a_0 \in M$ $\neg A(x)|_{\langle a_0, a_1, \ldots, a_n\rangle} = $ И. Отсюда $(\exists x)\neg A(x)|_{\langle a_1, \ldots, a_n\rangle} = $ И. С другой стороны, в этом случае $(\forall x)A(x)|_{\langle a_1, \ldots, a_n\rangle} = $ Л. Отсюда $\neg(\forall x)A(x)|_{\langle a_1, \ldots, a_n\rangle} = $ И. Равносильность (1.8) доказана.

Докажем теперь равносильность (1.9). Применим равносильность (1.8) к формуле $\neg A(x)$. Тогда $\neg(\forall x)\neg A(x) \equiv$

$\equiv (\exists x) \neg\neg A(x) \equiv (\exists x) A(x)$, и применив равносильность *11* основных равносильностей логики высказываний, получим $\neg(\exists x) A(x) \equiv \neg\neg(\forall x) \neg A(x) \equiv (\forall x) \neg A(x)$.

2. *Вынос квантора за скобки.* Пусть формула $A(x)$ содержит свободную переменную x, формула B не содержит переменной x и обе они удовлетворяют п. 3 определения формул (см. с. 74). Тогда

$$\begin{array}{l}(\exists x)(A(x) \& B) \equiv (\exists x) A(x) \& B;\\ (\forall x)(A(x) \& B) \equiv (\forall x) A(x) \& B;\\ (\exists x)(A(x) \vee B) \equiv (\exists x) A(x) \vee B;\\ (\forall x)(A(x) \vee B) \equiv (\forall x) A(x) \vee B.\end{array} \qquad (1.10)$$

Докажем первую из этих равносильностей (остальные доказываются аналогично).

Пусть $x_{i_1}, \ldots, x_{i_n}$ — все свободные переменные формулы $(\exists x)(A(x) \& B)$. Тогда они же и все свободные переменные формулы $(\exists x) A(x) \& B$.

Рассмотрим произвольную интерпретацию с множеством M. Пусть $<a_1, \ldots, a_n>$, $a_i \in M$, — произвольный набор значений свободных переменных $x_{i_1}, \ldots, x_{i_n}$. Так как формула B не содержит переменной x, то можно определить значение этой формулы на наборе $<a_1, \ldots, a_n>$ (точнее, на его части, относящейся к свободным переменным формулы B). Если $B|_{<a_1, \ldots, a_n>} = Л$, то $((\exists x) A(x) \& B)|_{<a_1, \ldots, a_n>} = Л$, и для любого элемента a из множества M на наборе значений $<a, a_1, \ldots, a_n>$ своих свободных переменных $x, x_{i_1}, \ldots, x_{i_n}$ формула $A(x) \& B$ принимает значение Л. Отсюда $(\exists x)(A(x) \& B)|_{<a_1, \ldots, a_n>} = Л$. Если $B_{<a_1, \ldots, a_n>} = И$, то для любого элемента a из множества M на наборе $<a, a_1, \ldots, a_n>$ формулы $A(x) \& B$ и $A(x)$ принимают одинаковые истинностные значения. Отсюда $((\exists x) A(x) \& B)|_{<a_1, \ldots, a_n>} = (\exists x) A(x)|_{<a_1, \ldots, a_n>} = (\exists x)(A(x) \& B)|_{<a_1, \ldots, a_n>}$.

Заметим, что если не требовать, чтобы формула B не содержала переменной x, то будут выполняться только две равносильности:

$$(\forall x)(A(x) \& B(x)) \equiv (\forall x) A(x) \& (\forall x) B(x);$$
$$(\exists x)(A(x) \vee B(x)) \equiv (\exists x) A(x) \vee (\exists x) B(x).$$

3. *Перестановка одноименных кванторов:*

$$(\forall y)(\forall x) A(x, y) \equiv (\forall x)(\forall y) A(x, y);$$
$$(\exists y)(\exists x) A(x, y) \equiv (\exists x)(\exists y) A(x, y).$$

4. *Переименование связанных переменных.* Заменяя связанную переменную формулы A другой переменной, не входящей в эту формулу, в кванторе и всюду в области действия квантора получаем формулу, равносильную A.

Это утверждение легко доказывается индукцией по длине формулы.

Рассмотрим способ упрощения формул, опирающийся на приведенные равносильности. Под длиной формулы здесь и далее будем понимать общее число входящих в нее символов предикатов, логических символов и символов кванторов. Так, формула $(\forall x_1)A^{(2)}{}_1(x_1, x_2) \,\&\, (\exists x_3)A^{(1)}{}_2(x_3)$ имеет длину 5.

Формулы, в которых из логических символов имеются только символы $\&$, $\vee$ и $\neg$, причем символ $\neg$ встречается лишь перед символами предикатов, будем называть *приведенными* формулами.

Пример 1.33.

1. $A^{(1)}{}_1(x_1) \vee A^{(2)}{}_2(x_1, x_2)$.
2. $(\forall x_1)A^{(1)}{}_1(x_1) \,\&\, (\exists x_2) \neg A^{(2)}{}_2(x_2, x_3)$.
3. $\neg(A^{(1)}{}_2(x_1) \vee A^{(1)}{}_1(x_2))$.
4. $(\forall x)A^{(1)}{}_1(x_1) \supset (\exists x_2) \neg A^{(1)}{}_1(x_2)$.
5. $\neg\,((\exists x_2)A^{(1)}{}_1(x_2) \supset A^{(1)}{}_2(x_1))$.

Первые две формулы — приведенные, а остальные не являются приведенными.

Теорема 1.17. *Для любой формулы существует равносильная ей приведенная формула, причем множества свободных и связанных переменных этих формул совпадают.*

Такая приведенная формула называется *приведенной формой* данной формулы.

Пользуясь равносильностями логики высказываний, легко указать формулу, равносильную данной и содержащую из логических символов только символы $\&$, $\vee$ и $\neg$. Поэтому будем считать, что рассматриваемые формулы содержат только эти логические символы.

Докажем это утверждение индукцией по длине формулы. При $n = 1$ формула атомарная и, следовательно, приведенная.

Предположим, что для формул длины меньше n утверждение теоремы верно. Докажем его для формулы F длины n. Обозначим длину формулы F через $\partial(F)$.

Формула F имеет вид $G \vee H$, $G \,\&\, H$, $\neg G$, $(\forall x)G(x)$ или $(\exists x)G(x)$.

Случай 1. Так как $\partial(G) \leqslant n-2$, $\partial(H) \leqslant n-2$, то по индуктивному предположению существуют приведенные формы G_1 и H_1 формул G и H соответственно, и множества свободных и связанных переменных формул G_1 и G, H_1 и H совпадают. Тогда $G_1 \vee H_1$ — формула, $G_1 \vee H_1 \equiv G \vee H$. Отсюда $G_1 \vee H_1$ — приведенная форма формулы $G \vee H$ с тем же множеством свободных и связанных переменных, что и $G \vee H$.

Случай 2 аналогичен случаю 1.

Случай 3. Здесь формула G может иметь вид: 3.1) $G_1 \vee H_1$; 3.2) $G_1 \,\&\, H_1$; 3.3) $\neg G_1$; 3.4) $(\forall x)G_1(x)$ или 3.5) $(\exists x)G_1(x)$.

Случай 3.1: $\neg G \equiv \neg G_1 \& \neg H_1$, где $\partial(\neg G_1) \leqslant n-2$, $\partial(\neg H_1) \leqslant n-2$. По индуктивному предположению существуют приведенные формулы G_2 и H_2, равносильные $\neg G_1$ и $\neg H_1$ соответственно, и множества свободных и связанных переменных формул G_1 и G_2, H_1 и H_2 совпадают. Тогда $G_2 \& H_2$ — приведенная форма формулы $\neg G$ с тем же множеством свободных и связанных переменных.

Случай 3.2 аналогичен случаю 3.1.

Случай 3.3: $\neg G \equiv \neg\neg G_1 = G_1$, где $\partial(G_1) = n-2$. Применяя индуктивное предположение к формуле G_1, получаем приведенную форму формулы $\neg G$ с тем же множеством свободных и связанных переменных.

Случай 3.4: $\neg G \equiv (\exists x)\neg G_1(x)$, где $\partial(\neg G_1(x)) = n-1$. По индуктивному предположению существует приведенная формула $G_2(x)$, равносильная формуле $\neg G_1(x)$ с тем же множеством связанных и свободных переменных. Тогда $(\exists x)G_2(x)$ — требуемая приведенная форма формулы $\neg G$.

Случай 3.5 аналогичен случаю 3.3.

Случай 4. Формула $G(x)$ имеет длину $n-1$. По индуктивному предположению существует приведенная форма $G_1(x)$ этой формулы с тем же множеством свободных и связанных переменных. Тогда $(\forall x)G_1(x)$ — требуемая приведенная форма формулы $(\forall x)G(x)$.

Случай 5 аналогичен случаю 4. Теорема доказана.

Приведенная формула называется *нормальной*, если она не содержит символов кванторов или все символы кванторов стоят впереди (т. е. логические символы и символы предикатов стоят в области действия каждого квантора).

Пример 1.34.

1. $(\forall x_1)(\exists x_2)(\neg A^{(1)}_1(x_1) \vee A^{(2)}_1(x_1, x_2))$ — нормальная формула.

2. $(\forall x_1)\neg A^{(1)}_1(x_1) \& (\exists x_2)A^{(1)}_2(x_2)$ — приведенная формула, не являющаяся нормальной.

Теорема 1.18. *Для любой приведенной формулы существует равносильная ей нормальная формула той же длины.*

Такая формула называется *нормальной формой* данной приведенной формулы.

Доказательство проведем индукцией по длине n формулы. При $n = 1$ формула атомарная и, следовательно, нормальная.

Предположим, что утверждение теоремы верно для всех формул длины меньше n. Докажем его для формул длины n. Пусть F — формула длины n. Тогда формула F имеет вид $G \vee H$, $G \& H$, $\neg G$, $(\forall x)G(x)$ или $(\exists x)G(x)$.

Случай 1. По условию $G \vee H$ — приведенная формула. Тогда G и H — тоже приведенные формулы длины меньше n. По индуктивному предположению существуют равносильные им нормальные формулы G_1 и H_1 соответственно, где $\partial(G_1) = \partial(G)$; $\partial(H_1) = \partial(H)$. Если формулы G_1 и H_1 не содержат символов

кванторов, то $G_1 \vee H_1$ — нормальная форма длины n формулы $G \vee H$.

Пусть, например, формула G_1 содержит символ квантора. Тогда G_1 имеет вид $(\forall x)G_2(x)$, где $\partial(G_2(x)) = \partial(G_1) - 1$ (случай, когда G_1 имеет вид $(\exists x)G_2(x)$, аналогичен). Если переменная x входит в формулу H_1, то только как связанная переменная (иначе нарушается определение формулы). Применив правило переименования связанных переменных, перейдем к формуле H_2, равносильной H_1 и не содержащей переменной x. Очевидно, H_2 — приведенная формула той же длины, что и H_1. По правилу выноса квантора за скобки $(\forall x)G_2(x) \vee H_2 \equiv (\forall x)(G_2(x) \vee H_2)$. Так как $\partial(G_2(x) \vee H_2) = \partial(G_1) - 1 + 1 + \partial(H_1) = n - 1$, то по индуктивному предположению существует равносильная ей нормальная формула $F_1(x)$ той же длины. Тогда $(\forall x)F_1(x)$ — нормальная форма формулы $G \vee H$ той же длины.

Случай 2 аналогичен случаю 1.

Случай 3. Так как формула $\neg G$ приведенная, то G — атомарная формула, и тогда $\neg G$ — нормальная формула.

Случай 4. Здесь $G(x)$ — приведенная формула, $\partial(G(x)) = n - 1$. По индуктивному предположению существует равносильная ей нормальная формула $G_1(x)$ той же длины. Тогда $(\forall x)G_1(x)$ — нормальная форма формулы $(\forall x)G(x)$ длины n.

Случай 5 аналогичен случаю 4. Теорема доказана.

Из теорем 1.17 и 1.18 вытекает следующее утверждение.

Теорема 1.19. *Для любой формулы существует равносильная ей нормальная формула.*

1.4.3. Выполнимость. Общезначимость

Рассмотрим некоторую интерпретацию с множеством M.

Говорят, что формула A *выполнима в данной интерпретации*, если существует набор $<a_1, \ldots, a_n>$, $a_i \in M$, значений свободных переменных $x_1, \ldots, x_i$ формулы A такой, что $A|_{<a_1, \ldots, a_n>} = \text{И}$.

Говорят, что формула A *истинна в данной интерпретации*, если она принимает значение И на любом наборе $<a_1, \ldots, a_n>$, $a_i \in M$, значений своих свободных переменных $x_i, \ldots, x_{i_n}$.

Говорят, что формула A *общезначима* или тождественно истинна (в логике предикатов), если она истинна в каждой интерпретации.

Говорят, что формула A *выполнима* (в логике предикатов), если существует интерпретация, в которой A выполнима.

Формула A общезначима тогда и только тогда, когда формула $\neg A$ не является выполнимой, и формула A выполнима тогда и только тогда, когда формула $\neg A$ не является общезначимой.

Очевидно, что если F и G — равносильные (в логике предикатов) формулы, то $F \sim G$ — общезначимая формула.

Применив это утверждение к формулам, равносильность которых доказана в разд. 1.4.2, получим общезначимые формулы. Докажем общезначимость некоторых других формул.

Утверждение 1.18. *Формула*

$$(\forall x)A(x) \supset A(y), \tag{1.11}$$

где переменная y не входит в формулу $A(x)$, общезначима.

Пусть $x, x_{i_1}, \ldots, x_{i_n}$ — все свободные переменные формулы $A(x)$. Тогда $y, x_{i_1}, \ldots, x_{i_n}$ — перечень свободных переменных формулы (1.11). Рассмотрим произвольную интерпретацию с множеством M.

Пусть $<b, a_1, \ldots, a_n>$, где $b \in M$, $a_i \in M (1 \leqslant i \leqslant n)$ — произвольный набор значений свободных переменных формулы (1.11). Докажем, что

$$(\forall x)A(x) \supset A(y) \,|_{<b,\, a_1, \ldots, a_n>} = \text{И}.$$

Действительно, для формулы $A(x)$ либо существует элемент $a_0 \in M$ такой, что на наборе $<a_0, a_1, \ldots, a_n>$ значений свободных переменных $x, x_{i_1}, \ldots, x_{i_n}$

$$A(x) \,|_{<a_0,\, a_1, \ldots, a_n>} = \text{Л},$$

либо для любого элемента $a \in M$ на наборе $<a, a_1, \ldots, a_n>$ значений свободных переменных $x, x_{i_1}, \ldots, x_{i_n}$

$$A(x) \,|_{<a,\, a_1, \ldots, a_n>} = \text{И}.$$

В первом случае

$$(\forall x)A(x) \,|_{<a_1, \ldots, a_n>} = \text{Л},$$

и тогда

$$(\forall x)A(x) \supset A(y) \,|_{<b,\, a_1, \ldots, a_n>} = \text{И}.$$

Во втором случае

$$(\forall x)A(x) \,|_{<a_1, \ldots, a_n>} = \text{И};\ A(y) \,|_{<b,\, a_1, \ldots, a_n>} = \text{И},$$

и тогда

$$(\forall x)A(x) \supset A(y) \,|_{<b,\, a_1, \ldots, a_n>} = \text{И}.$$

Утверждение 1.19. *Формула $A(y) \supset (\exists x)A(x)$, где переменная y не входит в формулу $A(x)$, общезначима.*

В силу утверждения 1.18 формула $(\forall x) \neg A(x) \supset \neg (\beta) \forall$ общезначима. Имеем

$$(\forall x) \neg A(x) \supset \neg A(y) \equiv \neg (\forall x) \neg A(x) \vee \neg A(y) \equiv$$

$$\equiv (\exists x)\,\neg\neg A(x) \vee \neg A(y) \equiv (\exists x)A(x) \vee \neg A(y) \equiv$$
$$\equiv A(y) \supset (\exists x)A(x).$$

Следовательно, формула $A(y) \supset (\exists x)A(x)$ общезначима.

Как говорилось выше, одноименные кванторы можно переставлять, следовательно, формулы

$$(\exists x)(\exists y)A(x, y) \sim (\exists y)(\exists x)A(x, y),$$
$$(\forall x)(\forall y)A(x, y) \sim (\forall y)(\forall x)A(x, y)$$

общезначимы.

Общезначимой является также формула $(\exists x)(\forall y)A(x, y) \supset \supset (\forall y)(\exists x)A(x, y)$ (доказательство будет приведено в конце разд. 1.4.4). Однако формула $(\forall x)(\exists y)A(x, y) \supset (\exists y)(\forall x)A(x, y)$ не является общезначимой. Действительно, пусть формула $A(x, y)$ — атомарная формула $A^{(2)}_1(x, y)$. Рассмотрим интерпретацию, областью которой является множество целых чисел; символу $A^{(2)}_1$ поставим в соответствие предикат $x<y$. Тогда формула $(\forall x)(\exists y)A^{(2)}_1(x, y)$ истинна в этой интерпретации, а формула $(\exists y)(\forall x)A^{(2)}_1(x, y)$ ложна.

Утверждение 1.20. *Пусть A — тождественно-истинная формула логики высказываний, $X_{i_1}, \ldots, X_{i_n}$ — список ее переменных. Подставив вместо каждой переменной X_{i_k}, $k=1, 2, \ldots, n$, формулы логики предикатов B_k (так, чтобы при этом не нарушались пп. 1—4 определения формулы, см. с. 74), получим общезначимую формулу логики предикатов.*

Задача распознавания общезначимости формул логики предикатов существенно сложнее, чем формул логики высказываний. Так же, как и в логике высказываний, она называется *проблемой разрешимости* и ставится следующим образом: указать эффективный способ (алгоритм) распознавания общезначимости формул (т. е. является ли данная формула общезначимой или нет). (Подробно понятие алгоритма будет обсуждаться в разд. 1.5.)

В общем случае эта проблема в логике предикатов неразрешима. Приведем это утверждение без доказательства.

Теорема Черча. *Не существует алгоритма, который для любой формулы логики предикатов устанавливает, общезначима она или нет.*

Однако в некоторых частных случаях проблема разрешимости решается. Например, если рассматривать формулы логики предикатов, содержащие только одноместные предикатные символы, то такой алгоритм существует. Логика, в которой употребляются только одноместные предикаты, соответствует логике, которая была описана еще Аристотелем.

Алгоритм проверки общезначимости формул, содержащих только одноместные предикатные символы, основан на следующем утверждении.

Утверждение 1.21. *Пусть F — формула, содержащая ровно n одноместных предикатных символов. Для того чтобы формула F была выполнимой, необходимо и достаточно, чтобы она была выполнимой во всех интерпретациях $<M, f>$ с множеством M, содержащим не более 2^n элементов.*

Приведем схему доказательства этого утверждения.

Пусть в интерпретациях $M_1=<M_1, f_1>$ и $M_2=<M_2, f_2>$ одноместным предикатным символам $A^{(1)}_j$ формулы поставлены в соответствие предикаты P_j и Q_j, $j=1, 2, \ldots$. Говорят, что интерпретации M_1 и M_2 гомоморфны, если существует сюръекция $\varphi: M_1 \to M_2$ такая, что для любого $a \in M_1$ и для любого $j = 1, 2, \ldots$ $P_j(a) = Q_j(\varphi(a))$. Тогда, как следует из индуктивного определения, формула, содержащая только одноместные предикатные символы, в гомоморфных интерпретациях одновременно либо выполнима, либо невыполнима.

Покажем, что если формула F, содержащая ровно n одноместных предикатных символов $A^{(1)}_1, \ldots, A^{(1)}_n$, выполнима, то она выполнима и в некоторой интерпретации с конечным множеством M, содержащим не более 2^n элементов. Пусть $M_1=<M_1, f_1>$ — интерпретация, в которой выполнима формула F, и пусть в этой интерпретации предикатным символам $A^{(1)}_j$ соответствуют предикаты P_j, $j = 1, 2, \ldots, n$. Для любого $a \in M_1$ рассмотрим подмножество $M_a \subseteq M_1$, состоящее из таких элементов b, что $<P_1(a), P_2(a), \ldots, P_n(a)>=<P_1(b), P_2(b), \ldots, P_n(b)>$. Число таких подмножеств M_a не превышает числа наборов из И и Л длины n, т. е. не превышает 2^n. Выберем в каждом из этих подмножеств по одному представителю и составим из них множество M. Тогда интерпретации $M_1=<M_1, f_1>$ и $M_2=<M, f>$, где f — ограничение функции f_1 на $M \subseteq M_1$, гомоморфны, и, следовательно, формула F выполнима в интерпретации M_2 с множеством M, содержащим не более 2^n элементов. Отсюда следует утверждение 1.21.

1.4.4. Исчисление предикатов

Как было сказано выше, в логике предикатов, в отличие от логики высказываний, нет эффективного способа для распознавания общезначимости формул. Поэтому аксиоматический метод становится существенным при изучении формул, содержащих кванторы. Выделение общезначимых формул так же, как и в исчислении высказываний, осуществляется путем указания некоторой совокупности формул, которые называются аксиомами, и указания правил вывода, позволяющих из общезначимых формул получать общезначимые.

В отличие от общепринятого изложения, рассмотрим некоторую узкую аксиоматическую теорию, которую также будем называть исчислением предикатов.

Исчисление предикатов — это аксиоматическая теория, символами которой являются, по существу, те же символы, что и в логике предикатов (см. разд. 1.4.1):

1) символы предметных переменных: $x_1, x_2, \ldots, x_n, \ldots$;

2) символы предикатов: $A^{(t)}_1, A^{(t)}_2, \ldots, A^{(t)}_n, \ldots$ $(t = 0, 1, 2, \ldots)$;

3) логические символы: $\neg$, $\supset$;

4) символы кванторов: $\forall$, $\exists$;

5) скобки и запятая: (,).

Сформулированное в разд. 1.4.1 определение формулы остается в силе и для исчисления предикатов (с той лишь разницей, что мы употребляем только два логических символа: $\neg$ и $\supset$; остальные связки можно ввести, например, так, как это сделано в исчислении высказываний, см. с. 64).

Аксиомы исчисления предикатов. Каковы бы ни были формулы A и B, следующие формулы являются аксиомами *:

A1. $A \supset (B \supset A)$;

A2. $(A \supset (B \supset C)) \supset ((A \supset B) \supset (A \supset C))$;

A3. $\neg B \supset \neg A) \supset ((\neg B \supset A) \supset B)$;

A4. $(\forall x_i) A(x_i) \supset A(x_j)$, где формула $A(x_i)$ не содержит переменной x_j;

A5. $A(x_i) \supset (\exists x_j) A(x_j)$, где формула $A(x_i)$ не содержит переменной x_j.

Правила вывода исчисления предикатов *:

1. Правило *m. p*: $\dfrac{A,\ A \supset B}{B}$.

2. Правило связывания квантором общности: $\dfrac{B \supset A(x_i)}{B \supset (\forall x_i) A(x_i)}$, где формула B не содержит переменной x_i.

3. Правило связывания квантором существования: $\dfrac{A(x_i) \supset B}{(\exists x_i) A(x_i) \supset B}$, где формула B не содержит переменной x_i.

4. Правило переименования связанной переменной. Связанную переменную формулы A можно заменить (в кванторе и во всех вхождениях в области действия квантора) другой переменной, не являющейся свободной в A.

Понятия вывода, теоремы, вывода из системы гипотез определяются в исчислении предикатов так же, как и в любой аксиоматической теории (см. разд. 1.3.1).

Теорема 1.20 *(ослабленная теорема о дедукции). Если* $\Gamma, A \vdash B$ *и существует вывод в исчислении предикатов, построенный с применением только правила I, то* $\Gamma \vdash A \supset B$.

Эта теорема доказывается аналогично теореме о дедукции в исчислении высказываний.

---

* При этом не должно нарушаться определение формулы.

Можно показать, что класс всех теорем исчисления предикатов совпадает с классом общезначимых формул.

Утверждение 1.22. *Аксиомы исчисления предикатов — общезначимые формулы.*

Для аксиом А1—А3 это следует из утверждения 1.20, для аксиом А4 и А5 — из утверждений 1.18 и 1.19.

Утверждение 1.23. *Формула, получающаяся из общезначимой формулы по любому из правил вывода 1—4, является общезначимой.*

Для правила вывода 1 это утверждение следует из свойств импликации.

Рассмотрим правило вывода 2. Пусть $B \supset A(x_i)$ — общезначимая формула. Докажем, что формула $B \supset (\forall x_i)A(x_i)$ тоже общезначима. Пусть $x_{k_1}, \ldots, x_{k_n}$ — все свободные переменные этой формулы. Тогда $x_i, x_{k_1}, \ldots, x_{k_n}$ — перечень свободных переменных формулы $B \supset A(x_i)$ (формула B не содержит переменной x_i).

Пусть задана произвольная интерпретация с множеством M и пусть $\langle a_1, \ldots, a_n \rangle$, $a_i \in M$, $1 \leqslant i \leqslant n$, — произвольный набор значений свободных переменных формулы $B \supset (\forall x_i)A(x_i)$.

Возможны два случая: 1) $B|_{\langle a_1, \ldots, a_n \rangle} = И$;

2) $B|_{\langle a_1, \ldots, a_n \rangle} = Л$.

В первом случае из общезначимости формулы $B \supset A(x_i)$ следует, что для любого элемента a из множества M $A(x_i)|_{\langle a, a_1, \ldots, a_n \rangle} = И$. Тогда $(\forall x_i)A(x_i)|_{\langle a_1, \ldots, a_n \rangle} = И$, откуда

$B \supset (\forall x_i)A(x_i)|_{\langle a_1, \ldots, a_n \rangle} = И$.

Во втором случае в силу свойства импликации

$B \supset (\forall x_i)A(x_i)|_{\langle a_1, \ldots, a_n \rangle} = И$.

Рассмотрим правило вывода 3. Пусть $A(x_i) \supset B$ — общезначимая формула. Докажем, что формула $(\exists x_i)A(x_i) \supset B$ тоже общезначима. Пусть $x_{k_1}, \ldots, x_{k_n}$ — все свободные переменные этой формулы. Тогда $x_i, x_{k_1}, \ldots, x_{k_n}$ — перечень свободных переменных формулы $A(x_i) \supset B$ (формула B не содержит переменной x_i).

Пусть задана произвольная интерпретация с множеством M и пусть $\langle a_1, \ldots, a_n \rangle$, $a_i \in M$, $1 \leqslant i \leqslant n$, — произвольный набор значений свободных переменных формулы $(\exists x_i)A(x_i) \supset B$.

Возможны два случая: 1) $B|_{\langle a_1, \ldots, a_n \rangle} = И$;

2) $B|_{\langle a_1, \ldots, a_n \rangle} = Л$.

В первом случае в силу свойств импликации

$(\exists x_i)A(x_i) \supset B|_{\langle a_1, \ldots, a_n \rangle} = И$.

Во втором случае из общезначимости формулы $A(x_i) \supset B$ следует, что для любого элемента a из множества M

$A(x_i)|_{<a, a_1, \ldots, a_n>} = Л$. Тогда $(\exists x_i)A(x_i)|_{<a_1, \ldots, a_n>} = Л$, откуда

$$(\exists x_i)A(x_i) \supset B|_{<a_1, \ldots, a_n>} = И.$$

Справедливость доказываемого утверждения для правила вывода 4 почти очевидна.

Теорема 1.21. *Любая выводимая в исчислении предикатов формула общезначима.*

Эта теорема вытекает из утверждений 1.22 и 1.23.

Пример 1.35. Докажем общезначимость формулы

$$(\exists x_i)(\forall x_j)A(x_i, x_j) \supset (\forall x_j)(\exists x_i)A(x_i, x_j).$$

Для этого покажем, что эта формула является теоремой исчисления предикатов, т. е.

$$\vdash (\exists x_i)(\forall x_j)A(x_i, x_j) \supset (\forall x_j)(\exists x_i)A(x_i, x_j).$$

Построим этот вывод:

(1) $\vdash (\forall x_j)A(x_i, x_j) \supset A(x_i, x_i)$ (аксиома 4);

(2) $\vdash A(x_i, x_i) \supset (\exists x_k)A(x_k, x_i)$ (аксиома 5);

(3) $A \supset B, B \supset C \vdash A \supset C$ (по теореме 1.20);

(4) $\vdash (\forall x_j)A(x_i, x_j) \supset (\exists x_k)A(x_k, x_i)$ ((3) применено к (1) и (2));

(5) $\vdash (\exists x_i)(\forall x_j)A(x_i, x_j) \supset (\exists x_k)A(x_k, x_i)$ (правило вывода 3 применено к (4));

(6) $\vdash (\exists x_i)(\forall x_j)A(x_i, x_j) \supset (\forall x_j)(\exists x_k)A(x_k, x_j)$ (правило вывода 2 примсено к (5));

(7) $\vdash (\exists x_i)(\forall x_j)A(x_i, x_j) \supset (\forall x_j)(\exists x_k)A(x_k, x_j)$ (правило вывода 4 применено к (6));

(8) $\vdash (\exists x_i)(\forall x_j)A(x_i, x_j) \supset (\forall x_j)(\exists x_i)A(x_i, x_j)$ (правило вывода 4 применено к (7)).

Теорема 1.22. *Исчисление предикатов непротиворечиво.*

Действительно, в силу теоремы 1.21 невозможно одновременно $\vdash A$ и $\vdash \neg A$.

Приведем без доказательства формулировку теоремы, обратной теореме 1.21.

Теорема Геделя *(о полноте исчисления предикатов). Всякая общезначимая формула выводима в исчислении предикатов.*

Задачи и упражнения

1. Будут ли следующие выражения формулами, и если это формулы, то какие переменные в них являются свободными, а какие связанными:

а) $(\forall x_1)(\exists x_2)(\forall x_3)A^{(1)}_1(x_1, x_2, x_3, x_4)$;

б) $(\forall x_1)A^{(2)}_1(x_1, x_2) \supset (\exists x_2)A^{(2)}_1(x_1, x_2)$;

в) $(\exists x_1)((\exists x_2)(A^{(2)}_2(x_1, x_3) \& A^{(2)}_3(x_2, x_4)))$?

2. В той же интерпретации M, что и в примере 4.5, записать формулы, выражающие следующие утверждения: а) x — нечетное число; б) $x < y$; в) z — наибольший общий делитель x и y; г) ассоциативность сложения; д) бесконечность множества простых чисел; е) конечность множества простых чисел; ж) существование наибольшего натурального числа. Будут ли формулы «е» и «ж» истинны в данной интерпретации?

3. Пусть M — множество точек прямых и плоскостей трехмерного евклидова пространства. Рассмотрим интерпретацию $M = \langle M, f\rangle$, где f — соответствие, сопоставляющее предикатным символам $P_1(x)$, $P_2(x)$, $P_3(x)$, $Q_1(x, y)$, $R(x, y)$ следующие предикаты: $P_1(x)$: «x — точка»; $P_2(x)$: «x — прямая»; $P_3(x)$: «x — плоскость»; $Q_1(x, y)$: «x лежит на y»; $R(x, y)$: «x совпадает с y». Записать в этой интерпретации формулы, выражающие следующие утверждения:

а) через каждые две точки можно провести прямую и притом единственную, если эти две точки различны;

б) через каждые три точки, не лежащие на одной прямой, можно провести единственную плоскость.

4. В интерпретации M_3 (см. пример 1.31, п. 3) записать утверждение о том, что функция $f(x)$ равномерно-непрерывна на отрезке. При каком условии, наложенном на функцию $f(x)$, в этой интерпретации истинна формула $(\exists\delta)(\forall\varepsilon)(\forall x_2)(\forall x_1) D_1(x) \& D(x) \& R(\varepsilon) \& (P_1(x_1, x_2, \delta) \supset \supset S_1(x_1, x_2, \varepsilon))$?

5. В интерпретации $M = \langle M, f\rangle$, где $M = P(A)$, A — некоторое множество, f — соответствие, сопоставляющее предикатному символу $P(x, y)$ предикат $x \subseteq y$, записать, что: а) x — пересечение y и z; б) x — объединение y и z; в) $x = \varnothing$; г) $x = A$; д) x — дополнение y.

6. Доказать или опровергнуть следующие равносильности (формула B не содержит вхождений переменной x):

а) $(\forall x)(A(x) \supset B) \equiv (\forall x)A(x) \supset B$;

б) $(\exists x)(A(x) \supset B) \equiv (\exists x)A(x) \supset B$;

в) $(\forall x)(B \supset A(x)) \equiv B \supset (\forall x)A(x)$;

г) $(\exists x)(B \supset A(x)) \equiv B \supset (\exists x)A(x)$.

7. Доказать, что следующие формулы равносильны:

а) $(\forall x)(A(x) \& C(x)) \equiv (\forall x)A(x) \& (\forall x)C(x)$;

б) $(\exists x)(A(x) \vee C(x)) \equiv (\exists x)A(x) \vee (\exists x)C(x)$.

8. Для следующих формул указать приведенную нормальную форму:

а) $(\exists x)(\forall y)A^{(2)}_1(x, y) \supset (\forall x)(\exists y)A^{(2)}_2(x, y)$;

б) $(\forall x)A^{(1)}_1(x) \supset (\forall x)(\exists y)A^{(2)}_1(x, y)$.

9. В интерпретациях примера 1.31 записать утверждение о том, что: а) число A не есть предел функции $f(x)$ при $x \to x_0$; б) функция $f(x)$ разрывна в точке x_0.

10. Привести примеры формул, равносильных в интерпретациях $\langle M, f\rangle$ с двухэлементным множеством M.

11. Выполнимы ли следующие формулы:
а) $(\forall x)A^{(1)}_1(x)$;
б) $(\exists x)(\forall y)(A^{(2)}_1(x, x) \& \neg A^{(2)}_1(x, y))$;
в) $(\exists x)A^{(1)}_1(x) \supset A^{(1)}_1(y)$?
12. Будут ли общезначимыми следующие формулы:
а) $(\forall x)A(x) \supset A(y)$;
б) $A(x) \supset (\forall y)A(y)$;
в) $(\exists x)A(x) \supset (\forall x)A(x)$;
г) $(\exists x)(\forall y)A(x, y) \sim (\forall y)(\exists x)A(x, y)$?
13. Выполнимы ли следующие формулы:
а) $(\forall x)(\exists y)(A^{(1)}(x) \sim \neg A^{(1)}(y))$;
б) $(\exists y)(\forall x)(A^{(1)}(x) \sim \neg A^{(1)}(y))$?

1.5. ЭФФЕКТИВНАЯ ВЫЧИСЛИМОСТЬ

Для решения ряда однотипных задач иногда целесообразно использовать чисто механические вычислительные процессы. С их помощью искомые величины вычисляются последовательно из данных величин по определенным правилам. Описание таких процессов принято называть алгоритмами. Вообще говоря, под алгоритмом интуитивно понимается некоторое формальное предписание, действуя согласно которому можно получить решение задачи.

Типичными примерами алгоритмов служат решения следующих задач:

1) нахождение наибольшего общего делителя двух положительных натуральных чисел;

2) извлечение квадратного корня из рационального числа с заданной степенью точности;

3) вычисление ранга целочисленной матрицы;

4) определение тождественной истинности формулы логики высказываний.

Перечисленные алгоритмы имеют ряд общих черт, которые естественно считать присущими любому алгоритму:

1) элементарность шагов алгоритма: решение задачи разбивается на этапы, каждый из которых должен быть простым и локальным;

2) детерминированность алгоритма: после выполнения очередного этапа однозначно определено, что делать на следующем этапе;

3) направленность алгоритма: должно быть указано, что считать результатом применения алгоритма;

4) массовость алгоритма: алгоритм служит для решения не какой-то одной задачи, а целого класса однотипных задач.

Интуитивное понятие алгоритма, хотя и не строго, но настолько ясно, что всегда можно определить, является ли данный

процесс алгоритмом. Поэтому удовлетворительным доказательством существования алгоритма считается описание фактического процесса решения какой-либо задачи.

Однако для некоторых известных проблем (например, для проблемы установления общезначимости формул логики предикатов) не удавалось найти разрешающего алгоритма. Безуспешные попытки найти такие алгоритмы привели к предположению, что их не существует. Но для того чтобы доказать несуществование алгоритма, надо точно знать, что такое алгоритм.

Начиная с 30-х годов было предложено несколько уточнений понятия алгоритма. Считается, что все они достаточно полно отражают основные черты интуитивного понятия алгоритма. Действительно, все формальные определения алгоритма в некотором смысле эквивалентны друг другу, все алгоритмы в точном смысле являются алгоритмами в интуитивном смысле, и, как показывает опыт, все известные алгоритмы можно задать алгоритмами в точном смысле.

Для алгоритмических проблем типичной является, например, ситуация, когда требуется найти алгоритм для вычисления числовой функции $f(x_1, \ldots, x_n)$, зависящей от целочисленных значений аргументов $x_1, \ldots, x_n$. Числовые функции, значения которых можно вычислять посредством некоторого (единого для данной функции) алгоритма, называются *вычислимыми функциями*. Принято считать, что $f(x_1, \ldots, x_n)$ *эффективно вычислима*, если имеется какая-нибудь механическая процедура (алгоритм), следуя которой можно найти значения $f(k_1, \ldots, k_n)$, когда известны значения $k_1, \ldots, k_n$ аргументов. Поскольку понятие алгоритма интуитивно, то и понятие вычислимой функции интуитивно.

Ниже мы рассмотрим некоторое уточнение понятия вычислимой функции — частично рекурсивные функции. Затем определим уточнение понятия алгоритма, связанное с машинами Тьюринга, что позволит нам доказать неразрешимость некоторых алгоритмических проблем.

1.5.1. Рекурсивные функции

Будем рассматривать только числовые функции, т. е. функции, аргументы и значения которых принадлежат множеству натуральных чисел N (в теории рекурсивных функций полагают $N=0, 1, 2, \ldots$). Иначе говоря, *частичной числовой* n-*местной функцией* $f(x_1, \ldots, x_n)$ называется функция, определенная на некотором подмножестве $M \subseteq N^n$ с натуральными значениями. Если область определения совпадает с множеством N^n, т. е. $f: N^n \to N$, то говорят, что функция f всюду определенная, в противном случае — частично определенная.

Например, $f(x, y) = x + y$ — всюду определенная двуместная функция, $f(x, y) = x - y$ — частично определенная функция (она определена, когда $x \geqslant y$).

Рекурсивным определением функции принято называть такое ее определение, при котором значения функции для данных аргументов определяются значениями той же функции для более простых аргументов или же значениями более простых функций. Простейшим примером рекурсивного определения являются числа Фиббоначи, представляющие собой последовательность чисел $f(n)$, удовлетворяющих условиям $f(0) = 1$, $f(1) = 1$, $f(n + 2) = f(n) + f(n + 1)$, т. е. 1, 1, 2, 3, 5, 8, 13 ...

Опишем класс числовых функций, играющих весьма важную роль в математической логике.

Следующие всюду определенные числовые функции назовем *простейшими:*

1) $S(x) = x + 1$ — прибавление единицы;

2) $O(x) = 0$ — нуль-функция;

3) $I_m(x_1, \ldots, x_n) = x_m$, $m = 1, \ldots, n$, — проектирующая функция.

Все эти функции вычислимы в интуитивном смысле.

Определим теперь операторы. Они обладают тем свойством, что, применяя их к функциям, вычислимым в интуитивном смысле, получаем функции, также вычислимые в интуитивном смысле.

1. *Оператор суперпозиции.* Говорят, что n-местная функция $\psi(x_1, \ldots, x_n)$ получена с помощью оператора суперпозиции из m-местной функции $\varphi(x_1, \ldots, x_m)$ и n-местных функций $f_1(x_1, \ldots, x_n), \ldots, f_m(x_1, \ldots, x_n)$, если

$$\psi(x_1, \ldots, x_n) = \varphi(f_1(x_1, \ldots, x_n), \ldots, f_m(x_1, \ldots, x_n)). \quad (1.12)$$

2. *Оператор примитивной рекурсии.* Говорят, что $(n + 1)$-местная функция $f(x_1, \ldots, x_n, y)$ получена из n-местной функции $\varphi(x_1, \ldots, x_n)$ и $(n + 2)$-местной функции $\psi(x_1, \ldots, x_n, y, z)$ с помощью оператора примитивной рекурсии, если ее значения можно вычислять по следующей схеме (схеме примитивной рекурсии):

$$\begin{cases} f(x_1, \ldots, x_n, 0) = \varphi(x_1, \ldots, x_n); \\ f(x_1, \ldots, x_n, y + 1) = \psi(x_1, \ldots, x_n, y, f(x_1, \ldots, x_n, y)). \end{cases} \quad (1.13)$$

При $n = 0$ схема примитивной рекурсии имеет вид

$$\begin{cases} f(0) = a; \\ f(y + 1) = \psi(y, f(y)), \end{cases}$$

где a — константа.

Заметим, что схема примитивной рекурсии (1.13) полностью определяет функцию f.

Функция $f(x_1, \ldots, x_n)$ называется *примитивно-рекурсивной*, если она может быть получена из простейших функций с помощью конечного числа применений операторов суперпозиции (1.12) и примитивной рекурсии (1.13).

Пример 1.36. Следующие функции примитивно-рекурсивны:

а) $S(x)$, $O(x)$, $I_m(x_1, \ldots, x_n)$;

б) функция $O(x_1, \ldots, x_n) = 0$, так как $O(x_1, \ldots, x_n) = O(I_1(x_1, \ldots, x_n))$, т. е. она получена из простейших функций $O(x)$ и $I_1(x_1, \ldots, x_n)$ с помощью оператора суперпозиции;

в) $f(x) = x + n$, т. е. $f(x) = \underbrace{S(S(\ldots S(x))\ldots)}_{n \text{ раз}}$;

г) $f(x, y) = x + y$. Действительно,

$$\begin{cases} f(x, 0) = x + 0 = x = I_1(x); \\ f(x, y + 1) = x + (y + 1) = (x + y) + 1 = S(x + y), \end{cases}$$

т. е. $f(x, y) = x + y$ получена из примитивно-рекурсивных функций $\varphi(x) = I_1(x)$ и $\psi(x, y, z) = z + 1 = S(z)$ с помощью оператора примитивной рекурсии.

3. *Оператор минимизации* (μ-оператор). Говорят, что функция $f(x_1, \ldots, x_n)$ получена из функции $g(x_1, \ldots, x_n, y)$ с помощью *оператора минимизации* (или просто μ-оператора) и обозначают $f(x_1, \ldots, x_n) = \mu y[g(x_1, \ldots, x_n, y) = 0]$, если $f(x_1, \ldots, x_n)$ определена и равна y тогда и только тогда, когда $g(x_1, \ldots, x_n, 0), \ldots, g(x_1, \ldots, x_n, y - 1)$ определены и не равны 0, а $g(x_1, \ldots, x_n, y) = 0$.

Функция $f(x_1, \ldots, x_n)$ называется *частично рекурсивной*, если она может быть получена из простейших функций с помощью конечного числа применений операторов суперпозиции, примитивной рекурсии и минимизации. Всюду определенные частично рекурсивные функции называются *общерекурсивными*.

Пример 1.37. Рассмотрим функцию

$$f(x, y) = \begin{cases} x - y, \text{ если } x \geqslant y; \\ \text{не определена в остальных случаях.} \end{cases}$$

Покажем, что она частично рекурсивна. Введем вспомогательные функции:

а) $x \dot{-} 1 = \begin{cases} x - 1, \text{ если } x \geqslant 1; \\ 0, \text{ если } x = 0; \end{cases}$

б) $x \dot{-} y = \begin{cases} x - y, \text{ если } x \geqslant y; \\ 0, \text{ если } x < y. \end{cases}$

Эти функции примитивно-рекурсивны. Действительно,

$$\begin{cases} 0 \dot{-} 1 = 0 = O(x); \\ (y + 1) \dot{-} 1 = y = I_2(x, y, z); \end{cases}$$

$$\begin{cases} x \dot{-} 0 = x; \\ x \dot{-} (y + 1) = (x \dot{-} y) \dot{-} 1. \end{cases}$$

Тогда функция $|x - y| = (x \dot{-} y) + (y \dot{-} x)$ также примитивно-рекурсивна. Пусть $g(x, y, z) = |x - (z + y)|$. Эта функция примитивно-рекурсивна. Функция $f(x, y)$ может быть получена из нее с помощью оператора минимизации: $f(x, y) = \mu z[|x - (z + y)| = 0]$, и, следовательно, она частично рекурсивна.

А. Черч выдвинул следующую гипотезу.

Тезис Черча. *Всякая эффективно вычислимая функция является частично рекурсивной.*

В формулировку тезиса Черча входит интуитивное понятие эффективной вычислимости, поэтому его нельзя ни доказать, ни опровергнуть в общепринятом математическом смысле. Это некоторый факт, в пользу которого говорит многолетняя математическая практика.

Частичная рекурсивность — это уточнение понятия вычислимой функции. С помощью этого точного определения можно доказывать или опровергать вычислимость. Наряду с частичной рекурсивностью имеются и другие уточнения понятия вычислимой функции (например, понятие вычислимости по Тьюрингу). Можно доказать их эквивалентность.

1.5.2. Машины Тьюринга

Машина Тьюринга — это математическая модель идеализированного вычислительного устройства. Приведем сначала неформальное ее описание:

1. Пусть имеется лента, т. е. полоса, разбитая на конечное число ячеек. В каждой ячейке ленты в каждый момент времени записан один из символов

$$a_0, a_1, \ldots, a_n. \tag{1.14}$$

Совокупность этих символов называется внешним алфавитом машины. Символ a_0 — пустой (его обычно обозначают символом 0).

В процессе работы машины к существующим ячейкам могут пристраиваться новые пустые ячейки, так что ленту можно считать неограниченной в обе стороны.

2. Каждая машина обладает внутренней памятью, которая может находиться в конечном числе состояний. Состояния внутренней памяти обозначают символами

$$q_0, q_1, \ldots, q_m, \tag{1.15}$$

отличными от символов (1.14), и называют внутренними состояниями машины. Одно из таких состояний (обычно q_0) называют заключительным, а другое (обычно q_1) — начальным.

3. Имеется управляющая головка, которая может перемещаться вдоль ленты таким образом, что в каждый момент времени она находится в определенной ячейке ленты. Принято говорить, что машина воспринимает эту ячейку. Машина действует не непрерывно, а лишь в дискретные моменты времени.

В зависимости от внутреннего состояния и от символа, записанного на воспринимаемой ячейке, в следующий момент времени машина переходит в новое внутреннее состояние (возможно, в то же самое), записывает новый символ в ту же ячейку (возможно, тот же самый) и либо сдвигает управляющую головку на одну клетку влево или вправо, либо оставляет ее на месте. Если управляющая головка воспринимает самую правую (левую) ячейку, а машина по ходу работы должна сдвинуть головку в отсутствующую ячейку справа (слева), то она пристраивает недостающую ячейку. Попав в заключительное состояние, машина прекращает работу.

Конфигурацией на ленте (или машинным словом) называется совокупность, образованная:

1) последовательностью $a_{i_1}, a_{i_2}, \ldots, a$ символов, записанных в ячейках ленты, где a_{i_1} — символ, записанный в первой ячейке слева, a_{i_2} — символ, записанный во второй ячейке слева, и т. д. (любая такая последовательность называется словом в алфавите (1.14));

2) состоянием внутренней памяти q_l;

3) номером k воспринимаемой ячейки.

Конфигурацию машины будем записывать так:

$$a_{i_1} \ldots . a_{i_{k-1}} \overset{a_{i_k}}{q_l} a_{i_{k+1}} \ldots a_{i_t} .$$

Так как каждая машина имеет конечный алфавит и конечное число внутренних состояний, то из описания ее работы видно, что она может выполнять конечное число действий.

Если машина, находясь во внутреннем состоянии q_l и воспринимая ячейку с символом a_j, записывает к следующему моменту времени в эту ячейку символ a_r, переходит во внутреннее состояние q_s и сдвигается по ленте, то говорят, что машина выполняет команду

$$q_l a_j \to q_s a_r S, \tag{1.16}$$

где S — сдвиг. Вместо S будем писать букву L, если сдвиг осуществляется влево, букву R, если сдвиг происходит вправо, и C, если головка остается на месте. При этом говорят, что машина переводит конфигурацию $a_{i_1} \ldots a_{i_k} \overset{a_j}{q_s} a_{i_{k+1}} \ldots a_{i_t}$, где $a_{i_1} \ldots a_{i_t}$ и $a_{i_{k+1}} \ldots a_{i_t}$ — произвольные слова в алфавите (1.14), в конфигурацию $a_{i_1} \ldots \overset{a_{i_k}}{q_s} a_r a_{i_{k+1}} \ldots a_{i_t}$,

$a_{i_1} \ldots a_{i_k} a_r \overset{a_{i_{k+1}}}{q_s} \ldots a_{i_t}$ или $a_{i_1} \ldots a_{i_{k-1}} \overset{a_r}{q_s} a_{i_{k-1}} \ldots a_{i_t}$ в зависимости от того, какое значение принимает сдвиг в команде.

Совокупность всех команд, которые может выполнить машина, называется ее программой. Программа машины должна содержать одну и только одну команду, начинающуюся словом $q_i a_j$, $i=1, \ldots, m$, $j=0, 1, \ldots, n$. Каждая машина Тьюринга определяется своим алфавитом, состояниями внутренней памяти и программой.

Чтобы полностью определить работу машины, надо указать ее конфигурацию для начального момента времени. Будем считать, что в начальной конфигурации головка воспринимает самую правую непустую ячейку.

Итак, *машина Тьюринга* есть, по определению, набор

$$M = \langle A, Q, \Pi \rangle, \tag{1.17}$$

где A — внешний алфавит (1.14) с выделенным пустым символом a_0; Q — алфавит внутренних состояний (1.15) с выделенными символами конечного и начального состояний q_0 и q_1; Π — программа, т. е. конечная последовательность упорядоченных пятерок символов (1.16).

Если машина, начав работу с некоторым словом, записанным на ленте, придет в заключительное состояние, то она называется *применимой* к этому слову. Результатом ее работы считается слово, записанное на ленте в заключительном состоянии. Если же машина ни в какой момент времени не придет в заключительное состояние, то она называется *не применимой* к данному слову, и результат ее работы не определен.

Пример 1.38. Рассмотрим машину M_1 с внешним алфавитом $\{0, |\}$, двумя внутренними состояниями $\{q_0, q_1\}$ и программой

$q_1| \to q_1|R;$
$q_10 \to q_0|C.$

Машина M_1 выполняет следующую операцию: к любому слову, состоящему из символов |, она прибавляет еще один символ | и останавливается. Если, например, в начальном состоянии на ленте записано слово $|||$, то в процессе работы машины появятся следующие конфигурации:

$\underset{q_1}{|||}$ — начальная конфигурация (для краткости здесь и далее в записи конфигурации опускаем все пустые символы, расположенные левее первого непустого и правее последнего непустого);

$\underset{q_1}{||}|0$ — следующая конфигурация;

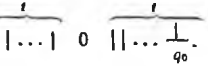
$\underset{q_1}{|||}|$ — заключительная конфигурация.

Машина M_1 применима к любому слову в алфавите $\{0, |\}$.

Пример 1.39 (удвоение слова). Построим машину с алфавитом $\{0, |\}$, которая по любому слову в алфавите $\{|\}$ строит два

таких слова; точнее, эта машина переводит конфигурацию вида

в конфигурацию вида



Можно указать несколько таких машин. Одна из них имеет следующую программу:

$$
\begin{array}{ll}
q_1| \to q_2|L & q_6| \to q_6|L \\
q_10 \to q_30C & q_60 \to q_70L \\
q_2| \to q_2|L & q_7| \to q_7|L \\
q_20 \to q_30R & q_70 \to q_8|C \\
q_3| \to q_40C & q_8| \to q_8|R \\
q_30 \to q_{10}0L & q_80 \to q_90R \\
q_40 \to q_50L & q_9| \to q_9|R \\
q_4| \to q_4|C & q_90 \to q_20C \\
q_50 \to q_6|C & q_{10}0 \to q_00L \\
q_5| \to q_5|C & q_{10}| \to q_{10}|C
\end{array}
$$

Команда $q_90 \to q_20C$ зацикливает программу, и машина перерабатывает слово до тех пор, пока не придет в состояние q_0, воспринимая при этом пустой символ.

Пусть M_1 и M_2 — две машины Тьюринга с общим алфавитом $\{a_0, a_1, \ldots, a_n\}$ и $\{q_0, q_1, \ldots, q_m\}$ — внутренние состояния машины M_1, $\{q_0, q'_1, \ldots, q'_l\}$ — внутренние состояния машины M_2.

Композицией M машин M_1 и M_2 называется машина с алфавитом $\{a_0, a_1, \ldots, a_n\}$, внутренними состояниями $\{q_0, q_1, \ldots, q_m, q_{m+1}, \ldots, q_{m+l}\}$ и программой, получающейся следующим образом. Во всех командах машины M_1 заменим заключительный символ q_0 символом q_{m+1}, а остальные символы оставим без изменения. Во всех командах машины M_2 символ q_0 оставим неизменным, а все остальные символы $q'_i (i = 1, \ldots, l)$ заменим соответственно символами q_{m+i}. Совокупность всех команд машин M_1 и M_2, измененная указанным способом, и будет программой композиции M машин M_1 и M_2. «Работа» машины M равносильна последовательной «работе» машин M_1 и M_2.

А. Тьюринг выдвинул следующую гипотезу.

Тезис Тьюринга. *Всякий интуитивный алгоритм может быть реализован с помощью некоторой машины Тьюринга.*

Выше уже приводились доводы в пользу этого тезиса. Как показывает опыт, любые действия, которые может выполнить вычислитель — человек, могут быть разложены в последовательность действий некоторой машины Тьюринга.

Часто в алгоритмических проблемах речь идет не о построе-

нии алгоритма, а о вычислимости некоторых специальным образом построенных функций.

Воспользуемся специальным кодированием натуральных чисел в алфавите $\{|\}$: каждое натуральное число представим $(n+1)$ символом $|$, т. е. числа 0, 1, 2, ... кодируем словами $|$, $||$, $|||$, ...

Частичная числовая n-местная функция $f(x_1, \ldots, x_n)$ называется *вычислимой по Тьюрингу*, если существует машина M, вычисляющая ее в следующем смысле:

1. Если набор значений аргументов $\langle x_1, \ldots, x_n \rangle$ принадлежит области определения функции f, то машина M, начиная работу в конфигурации

$$0\,|^{x_1+1}\;0\,|^{x_2+1}\ldots 0\frac{|^{x_n+1}}{q_1} \tag{1.18}$$

(где $|^x = \underbrace{|\ldots|}_{x \text{ раз}}$ и воспринимается самый правый символ $|$), останавливается, заканчивая работу в конфигурации

$$0\frac{|^{y+1}}{q_0},$$

где $y = f(x_1, \ldots, x_n)$.

2. Если набор значений аргументов $\langle x_1, \ldots, x_n \rangle$ не принадлежит области определения функции f, то машина M, начиная работу в конфигурации (1.18), работает бесконечно, т. е. не приходит в заключительное состояние.

Пример 1.40. Простейшие функции $S(x) = x+1$ и $O(x) = 0$ вычислимы по Тьюрингу. Первую из них вычисляет машина M_1, описанная в примере 1.38, а вторую — машина с теми же алфавитами и программой

$$q_1| \to q_1 0L,$$
$$q_1 0 \to q_0| C.$$

Можно доказать, что любая частично рекурсивная функция вычислима по Тьюрингу и, наоборот, любая вычислимая по Тьюрингу функция частично рекурсивна.

Теперь, обладая точным формальным понятием алгоритма мы можем доказать неразрешимость некоторых алгоритмических проблем.

Каждая машина Тьюринга по определению есть набор (1.17), т. е. задается внешним алфавитом A (1.14), алфавитом внутренних состояний Q (1.15) и программой Π:

$$q_{i_\alpha} a_{j_\alpha} \to q_{r_\alpha} a_{s_\alpha} S_{t_\alpha},\ \alpha = 1, \ldots, k, \tag{1.19}$$

где S_1 — это R; S_2 — это L; S_3 — это C. При этом можно считать, что существуют некоторые обширные алфавиты A_0 и Q_0, в которых записываются символы (1.14) и (1.15) для всех M

шин Тьюринга. Тогда символы q_{i_k}, a_{j_s}, η_{r_u}, a_{λ_μ} из (1.19) есть символы алфавитов A_0 и Q_0.

Укажем способ нумерации всех машин Тьюринга (геделева нумерация). Пусть p_1, p_2, p_3, ... — последовательность всех простых чисел, расположенных в порядке возрастания (т. е. последовательность 2, 3, 5, 7, 11 ...). Номером машины Тьюринга M с программой (1.19) назовем число

$$n(M) = p^{i_1}_1 p^{j_1}_2 p^{r_1}_3 p^{\nu_1}_4 p^{i_1}_5 p^{j_2}_6 p^{i_2}_7 p^{r_2}_8 p^{\nu_2}_9 p^{i_3}_{10} \ldots$$

$$\ldots p^{i_k}_{5k-4}\, p^{j_s}_{5k-3}\, p^{r_u}_{5k-2}\, p^{\nu_\mu}_{5k-1}\, p^{i_t}_{5k}.$$

Пример 1.41. Номер машины M_1, описанной в примере 1.38, — число

$$n(M_1) = 2^1\, 3^1\, 5^1\, 7^1\, 11^1\, 13^1\, 17^0\, 19^0\, 23^1\, 29^3.$$

Естественно, что не все натуральные числа являются номерами машин Тьюринга. Если n — номер некоторой машины в алфавитах (1.14), (1.15), то ее программу можно однозначно восстановить по этому номеру.

Как и раньше, кодируем натуральные числа символом $|$. Будем рассматривать машины Тьюринга, алфавит которых содержит символ $|$.

Напомним, что машина называется применимой к начальному слову, если она, начав работать с этим словом на ленте, придет в заключительное состояние.

Машина M, применимая к слову $n(M)$ (т. е. к коду своего собственного номера), называется *самоприменимой*. Машина, не применимая к слову $n(M)$, называется *несамоприменимой*.

Машина M_1 (см. пример 1.38) самоприменима. Примером несамоприменимой машины является машина, в правых частях команд которой не встречаются заключительные состояния. Такая машина не применима ни к какому слову.

Следуя О. Б. Лупанову *, рассмотрим алгоритмическую *проблему самоприменимости*. Она состоит в следующем: указать алгоритм, который по любой машине Тьюринга устанавливал бы, самоприменима она или нет.

Согласно тезису Тьюринга такой алгоритм следует искать в виде машины Тьюринга, т. е. требуется построить машину Тьюринга, которая была бы применима к кодам номеров всех машин Тьюринга и в зависимости от того, самоприменима она или нет, имела бы различные заключительные конфигурации. Пусть, например, в случае самоприменимой машины заключительная конфигурация имеет вид

$$\ldots \underset{q_0}{1} \ldots$$

* См.: Лупанов О. Б. Конспект лекций по математической логике — М., Изд-во МГУ, 1970.

где вне обозреваемой ячейки могут быть записаны какие угодно символы алфавита, а в случае несамоприменимой машины заключительная конфигурация имеет вид

$$\ldots\frac{0}{q_0}\ldots$$

Теорема 1.23. *Проблема самоприменимости алгоритмически неразрешима, т. е. не существует машины Тьюринга, решающей эту проблему в указанном выше смысле.*

Допустим, что такая машина A существует. Тогда можно построить машину B, которая 1) применима ко всем кодам номеров несамоприменимых машин и 2) не применима ко всем кодам номеров самоприменимых машин.

Действительно, машина B получается из машины A следующим образом: алфавит сохраняется неизменным, заключительное состояние q_0 машины A считается незаключительным состоянием машины B, а заключительным состоянием машины B считается новое состояние q'_0. Программа машины B состоит из всех команд машины A и еще двух команд: $q_0| \to q_0|C$; $q_00 \to q'_00C$. Очевидно, что машина B удовлетворяет требованиям 1 и 2.

Сама машина B либо самоприменима, либо несамоприменима. Если она самоприменима, т. е. применима к коду своего номера, то она применима и к коду самоприменимой машины, но тогда не выполняется требование 2.

Если машина B несамоприменима, т. е. не применима к коду своего номера, то она не применима и к коду несамоприменимой машины, но тогда не выполняется требование 1.

Итак, мы пришли к противоречию, основываясь на допущении о том, что существует машина A, решающая проблему самоприменимости. Следовательно, такой машины не существует.

Заметим, что неразрешима именно массовая проблема: не существует единого алгоритма, который решал бы проблему самоприменимости. Мы же приводили примеры конкретных машин Тьюринга, для которых такой алгоритм существует.

Используя результат теоремы 1.23, можно доказать неразрешимость других алгоритмических проблем. Рассмотрим, например, *проблему применимости к начальному слову*. Она состоит в следующем: указать алгоритм, который по машине M и слову X устанавливал бы, применима машина M к слову X или нет. В терминах машин Тьюринга эту проблему можно сформулировать так: можно ли построить машину, которая была бы применима ко всем словам вида $n(M)0X$, где M — произвольная машина, X — произвольное слово, и в случае, если машина M применима к слову X, приводила бы к заключительной конфигурации

$$\ldots\frac{1}{q_0}\ldots$$

а в случае, если машина M не применима к слову X, приводила бы к заключительной конфигурации

$$\ldots\underset{q_0}{0}\ldots$$

Теорема 1.24. *Проблема применимости к начальному слову алгоритмически неразрешима, т. е. не существует машины Тьюринга, решающей эту проблему в указанном выше смысле.*

Допустим, что такая машина D существует. Пусть E — машина, удваивающая слова (см. пример 1.38). Рассмотрим машину G, которая является композицией машин E и D. Если в начальный момент работы машины G на ленте будет слово $n(M)$, то после работы машины E на ленте будет слово $n(M)0n(M)$. Машина D применима ко всем таким словам и остановится в конфигурации

$$\ldots\underset{q_0}{1}\ldots,$$

если машина M применима к слову $n(M)$, и в конфигурации

$$\ldots\underset{q_0}{0}\ldots,$$

если машина M не применима к слову $n(M)$. Но тогда машина G решает проблему самоприменимости, что невозможно в силу теоремы 1.23.

Итак, мы пришли к противоречию, исходя из допущения, что существует машина, решающая проблему применимости. Следовательно, такой машины не существует.

Задачи и упражнения

1. Доказать примитивную рекурсивность следующих функций:

а) $f(x) = x + k$; б) $f(x, y) = xy$; в) $f(x, y) = \min(x, y)$; г) $f(x, y) = \max(x, y)$.

2. Доказать, что всякая примитивно-рекурсивная функция общерекурсивна.

3. Доказать, что следующие функции частично рекурсивны:

а) нигде не определенная функция ω, т. е. функция ω с пустой областью определения;

б) $f(x \div y) = \begin{cases} x/y, \text{ если } y \text{ делит } x; \\ \text{не определена в остальных случаях.} \end{cases}$

4. Построить машины Тьюринга, вычисляющие функции: а) $f(x) = x + k$; б) $f(x, y) = x \div y$; в) $f(x) = x \div 1$.

АЛГЕБРАИЧЕСКИЕ СТРУКТУРЫ

ГЛАВА 2

Приведем краткое изложение основных понятий теории алгебраических структур (групп, колец и полей). Проиллюстрируем также приложение теоретико-групповых методов к одной из задач теории информации: порождению, кодированию и декодированию групповых кодов.

2.1. ГРУППЫ

2.1.1. Полугруппы и моноиды

Множество Π с заданной в нем бинарной ассоциативной операцией называется *полугруппой*. Полугруппа с единичным элементом называется *моноидом* (или просто полугруппой с единицей).

Пример 2.1.

1. Пусть X — произвольное множество, $M(X)$ — множество всех отображений X в себя. Тогда относительно операции композиции отображений $M(X)$ — моноид, он некоммутативный. Обозначим его $M(X, \circ, e_X)$.

2. Множество квадратных матриц порядка n относительно умножения матриц — некоммутативный моноид с единичным элементом — единичной матрицей, а относительно сложения матриц — коммутативный моноид с единичным элементом — нулевой матрицей. Обозначим их соответственно $(M_n(\mathbf{R}), \cdot, E)$ $(M_n(\mathbf{R}), +, 0)$.

3. Множество целых чисел — коммутативный моноид как относительно сложения, так и умножения. Обозначим их соответственно $(\mathbf{Z}, +, 0)$, $(\mathbf{Z}, \cdot, 1)$. Множество целых чисел, делящихся на $n (n>1)$, — коммутативная полугруппа без единицы относительно умножения: Обозначим ее $(n\mathbf{Z}, \cdot)$.

4. Пусть $A=\{a_1, \ldots, a_n\}$ — конечное множество символов алфавит. Конечная последовательность символов называется словом в алфавите A. На множестве Π слов в алфавите A вве-

дем бинарную операцию — «приписывание», т. е. если $a=a_{i_1}\ldots a_{i_k}$, $b=a_{j_1}\ldots a_{j_l}$, то $ab=a_{i_1}\ldots a_{i_k}a_{j_1}\ldots a_{j_l}$. Тогда П — полугруппа. Она называется свободной полугруппой, порожденной множеством A.

5. Множество $\{X_1, X_2, X_3, X_4\}$ относительно операции, заданной таблицей Кэли (см. табл. 0.1),— коммутативный моноид, единичный элемент которого X_1.

Подмножество П′ полугруппы П называется *подполугруппой*, если $ab\in\Pi'$ для всех $a, b\in\Pi'$. В этом случае говорят также, что подмножество П′ замкнуто относительно рассматриваемой операции. Очевидно, подполугруппа П′ сама является полугруппой относительно операции в П. Если M — моноид и подмножество M' не только замкнуто относительно операции, но и содержит единичный элемент, то M' называется *подмоноидом* M.

Например, множество чисел, кратных n,— подполугруппа в полугруппе целых чисел относительно умножения $(\mathbf{Z}, \cdot, 1)$ и подмоноид в $(\mathbf{Z}, +, 0)$. В полугруппе П слов в алфавите A подмножество слов в алфавите $A'\subseteq A$ также подполугруппа.

Элемент a моноида M с единичным элементом e называется *обратимым*, если для некоторого элемента b выполняется равенство $ab=ba=e$. Элемент b называется *обратным* a и обозначается a^{-1}. Обратный элемент единственен. Действительно, если еще и $ab'=b'a=e$, то $b'=eb'=(ba)b'=b(ab')=be=b$.

Нетрудно видеть, что $(a^{-1})^{-1}=a$. Кроме того, если a, b обратимы, то $(ab)^{-1}=b^{-1}a^{-1}$, так как $(ab)(b^{-1}a^{-1})=a(bb^{-1})a^{-1}=aea^{-1}=aa^{-1}=e$, и аналогично $(b^{-1}a^{-1})(ab)=e$, т. е. элемент ab тоже обратим.

Множество всех обратимых элементов моноида образует подмоноид, так как содержит единичный элемент и замкнуто относительно операции: если a и b обратимы, то $b^{-1}a^{-1}$ — элемент, обратный ab.

Рассмотрим проблему тождества слов в полугруппах.

Пусть $S=\{s_1, \ldots, s_n\}$ — подмножество элементов полугруппы П такое, что любой элемент из П может быть представлен как произведение элементов из S. Тогда множество S называется системой образующих полугруппы П. Например, для свободной полугруппы П, порожденной алфавитом $A=\{a_1, \ldots, a_n\}$, само множество A является системой образующих; для полугруппы целых чисел относительно сложения $(\mathbf{Z}, +, 0)$ системой образующих является множество $\{-1, 1, 0\}$, а для полугруппы целых чисел относительно умножения $(\mathbf{Z}, \cdot, 1)$ образующими являются все простые числа и единица.

Следует заметить, что далеко не все произведения элементов множества S будут различными элементами полугруппы П. В общем случае вопрос о равенстве таких произведений довольно сложен.

Будем рассматривать полугруппы, порожденные конечным множеством своих элементов. Они называются конечно-порожденными.

Можно указать некоторый способ задания полугрупп без использования индивидуальных свойств элементов множества, в котором определена полугрупповая операция, а именно задание полугруппы образующими и определяющими соотношениями.

Каждая полугруппа П может быть задана образующими

$$a_1;\ a_2,\ \ldots,\ a_n \tag{2.1}$$

(алфавит полугруппы) и определяющими соотношениями

$$A_i = B_i\ (i=1,\ 2,\ \ldots), \tag{2.2}$$

где A_i и B_i — слова в алфавите (2.1).

Элемент полугруппы, т. е. слово в алфавите (2.1), называют словом полугруппы П.

Элементарным преобразованием полугруппы П называется переход от слова вида XA_iY к слову XB_iY (или обратно), где X, Y — произвольные слова полугруппы П, а $A_i = B_i$ — одно из определяющих соотношений (2.2).

Элементарные преобразования представляются в виде схем

$$XA_iY \to XB_iY;\ XB_iY \to XA_iY.$$

К схемам элементарных преобразований относятся также тавтологические переходы вида $X \to X$. Графическое совпадение двух слов X и Y обозначают $X \equiv Y$.

Соотношения (2.2) определяют равенство слов в полугруппе П, которое связано с элементарными преобразованиями полугруппы П следующим образом. Два слова X и Y полугруппы П равны в П тогда и только тогда, когда можно указать последовательность элементарных преобразований полугруппы П

$$X \equiv X_0 \to X_1 \to \ldots \to X_i \to X_{i+1} \to \ldots \to X_k \equiv Y,$$

переводящую слово X в слово Y.

Для свободной полугруппы с алфавитом (2.1) множество определяющих соотношений пусто; два слова равны тогда и только тогда, когда они графически совпадают.

Полугруппу $(\mathbf{Z},\ +,\ 0)$ целых чисел относительно сложения можно задать образующими $\{-1,\ 1,\ 0\}$ и определяющими (в аддитивной записи) соотношениями:

$$1+(-1)=0;\ (-1)+1=0.$$

Проблема тождества слов в полугруппе заключается в следующем: указать алгоритм, который по любым двум словам устанавливал бы, равны они в полугруппе П или нет. Доказано, что эта проблема алгоритмически неразрешима. Простым примером полугруппы с неразрешимой проблемой тождества слов

является полугруппа с образующими a, b, c, d, e и определяющими соотношениями $ac=ca$, $ad=da$, $bc=cb$, $bd=db$, $eca=ce$, $edb=de$, $cca=ccae$.

2.1.2. Группы: определение и примеры

Непустое множество G с одной бинарной алгебраической операцией называется *группой*, если выполняются следующие условия:

1) операция в G ассоциативна;

2) в G существует единичный элемент e: $ea=ae=a$ для всех $a\in G$;

3) для каждого элемента a существует обратный ему элемент a^{-1}: $aa^{-1}=a^{-1}a=e$.

Иными словами, моноид G, все элементы которого обратимы, называется группой.

Если операция в G коммутативна, то группа называется *коммутативной* или *абелевой*.

Подмножество $H\subseteq G$ называется *подгруппой* в G, если ему принадлежит единичный элемент e, для любых элементов h_1, $h_2\in H$ выполняется $h_1h_2\in H$, т. е. H замкнуто относительно операции, и для любого $h\in H$ выполняется $h^{-1}\in H$. Подгруппа $H\subset G$ называется *собственной*, если $H\neq e$ и $H\neq G$.

Пример 2.2.

1. Множество целых чисел образует группу целых чисел относительно операции сложения $(\mathbf{Z}, +, 0)$. Эта группа коммутативна. Аналогично множество рациональных и действительных чисел образует соответственно группы относительно сложения $(\mathbf{Q}, +, 0)$ и $(\mathbf{R}, +, 0)$. Подмножество четных чисел образует подгруппу. Подмножество нечетных чисел не образует подгруппу, так как операция сложения выводит из этого множества.

2. Множество целых чисел не образует группу относительно умножения, так как может не существовать обратного элемента. Все отличные от нуля рациональные числа и действительные числа образуют группы относительно умножения, причем коммутативные. Положительные рациональные и положительные действительные числа образуют подгруппы этих групп.

3. Пусть X — произвольное множество, $S(X)$ — множество всех биективных отображений X в себя. Тогда $S(X)$ — группа относительно операции композиции $\circ$. Она называется группой преобразований.

4. Рассмотрим множество M_n квадратных матриц порядка n с определителем, отличным от нуля. Это некоммутативная группа $(M_n, \cdot, E)$ относительно операции умножения матриц, поскольку каждый элемент имеет обратный — обратную матрицу. Подмножество матриц с определителем, равным 1, образует подгруппу, так как

$\det E=1$; $\det A=1$; $\det B=1 \Rightarrow \det AB=1$; $\det A=1 \Rightarrow$
$\Rightarrow \det A^{-1}=1$.

5. Множество элементов $\{x_1, x_2, x_3, x_4\}$ относительно операции, определенной таблицей Кэли (см. табл. 0.1),— группа. Для элемента x_2, например, обратным является элемент x_4 (см. введение, разд. 0.4).

6. Рассмотрим множество классов вычетов по модулю n, т. е. классов эквивалентности по отношению ρ сравнимости по модулю числа n на множестве целых чисел ($a\rho b$ тогда и только тогда, когда $a=b \pmod n$). Обозначим эти классы через $C_0, \ldots, C_{n-1}$, где $a \in C_k$ тогда и только тогда, когда $a=k \pmod n$, $0 \leqslant k<n$ (см. введение, разд. 0.2). Множество классов вычетов образует группу относительно операции сложения классов, определяемой по следующему закону: $C_k+C_l=C_r$, где $k+l=r \pmod n$, $0 \leqslant r<n$, и $e=C_0$, $C^{-1}{}_k=C_{n-k}$. Например, при $n=3$ эту группу можно задать таблицей Кэли (см. табл. 2.1).

7. Рассмотрим свободную полугруппу с алфавитом $\{a_1, \ldots, a_n\}$. Сопоставим символам $a_1, \ldots, a_n$ символы $a_1^{-1}, \ldots, a_n^{-1}$. Пустое слово обозначим символом 1. Тогда *свободной группой* с образующими $a_1, \ldots, a_n$ называется полугруппа с единицей 1, которая задается определяющими соотношениями $a_i a_i^{-1}=1$, $a_i^{-1} a_i=1$, $i=1, \ldots, n$.

Таблица 2.1

	C_0	C_1	C_2
C_0	C_0	C_1	C_2
C_1	C_1	C_2	C_0
C_2	C_2	C_0	C_1

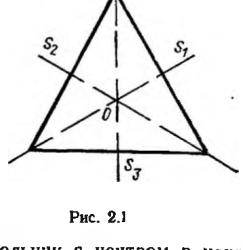


Рис. 2.1

8. Рассмотрим правильный n-угольник с центром в начале координат 0. Тогда множество преобразований плоскости, переводящее многоугольник в себя, образует некоммутативную группу, называемую группой самосовмещений многоугольника. Можно показать, что любое такое преобразование есть композиция вращений вокруг точки 0 и осевых симметрий. Например, при $n=3$ группа самосовмещений правильного треугольника состоит из вращений φ_0, φ_1, φ_2 против часовой стрелки на углы $0, \frac{2\pi}{3}$ и $\frac{4\pi}{3}$, которые переводят треугольник в себя, и трех симметрий ψ_1, ψ_2, ψ_3 относительно осей симметрии S_1, S_2, S_3 (рис. 2.1). Из геометрических соображений понятно, что $\varphi_1{}^3=$

$=e=\varphi_0$, $\psi^2_1=e$, $(\varphi_1\psi_1)^2=e$, а элементы φ_2, ψ_2, ψ_3 выражаются через φ_1 и ψ_1: $\varphi_2=\varphi^2_1$, $\psi_2=\psi_1\varphi_1$, $\psi_3=\psi_1\varphi^2_1$.

Если группа состоит из конечного числа элементов, то она называется *конечной*, а число ее элементов — *порядком* группы. В противном случае группа называется *бесконечной*.

2.1.3. Циклические группы. Группы подстановок

Пусть G — группа, H и F — ее подгруппы. Тогда пересечение $D=H\cap F$ непустое, поскольку оно содержит единичный элемент. D также является подгруппой группы G. Действительно, если элементы a и b принадлежат D, то их произведение и обратные им элементы содержатся как в H, так и в F, и поэтому также принадлежат D. Аналогично доказывается и следующее утверждение

Теорема 2.1. *Пересечение любого множества подгрупп группы G само является подгруппой этой группы.*

Пусть S — произвольное непустое подмножество группы G. Рассмотрим всевозможные подгруппы G, которые содержат S в качестве подмножества. Одной из них будет, в частности, сама группа G. В силу теоремы 2.1 пересечение всех таких подгрупп будет подгруппой G, которая называется подгруппой, *порожденной* множеством S, и обозначается $<S>$.

Если множество S состоит из одного элемента a, то порожденная им подгруппа $<a>$ называется *циклической* подгруппой, порожденной элементом a.

Обозначим $(a^{-1})^k=a^{-k}$.

Теорема 2.2. *Циклическая подгруппа $<a>$, порожденная элементом a, состоит из всех степеней элемента a.*

Заметим, что все степени элемента a принадлежат подгруппе $<a>$ и для любого целого k $(a^{-1})^{-k}=a^k$. С другой стороны, эти степени сами составляют подгруппу, так как $a^m a^n=a^{m+n}$, $a^0=e$, а обратным элементу a^n является элемент a^{-n}. Действительно, нетрудно доказать, что для любых целых m и n

$$a^m a^n=a^{m+n};\quad (a^m)^n=a^{mn}.$$

Для натуральных m и n это следует из соотношений (0.1). Если $m<0$, $n<0$, то

$$a^m a^n=(a^{-1})^{-m}(a^{-1})^{-n}=(a^{-1})^{-(m+n)}=a^{m+n}.$$

Если $m<0$, $n>0$, то

$$a^m a^n=(a^{-1})^{-m}a^n=(a^{-1}\ldots a^{-1})\,(a\ldots a)=$$

$$=\begin{cases}a^{n-(-m)}, & \text{если } n\geqslant -m\\ (a^{-1})^{-m-n}, & \text{если } n<-m\end{cases}=a^{m+n}.$$

($-m$ раз, n раз)

Случай $m>0$, $n>0$ аналогичен предыдущему. Доказательство второго равенства предлагается провести самостоятельно.

Группа, совпадающая с одной из своих циклических подгрупп (т. е. состоящая из степеней одного из своих элементов), называется *циклической*, а элемент, из степеней которого состоит циклическая группа,— ее образующим. Всякая циклическая группа коммутативна.

Пример 2.3.

1. Группа $(Z, +, 0)$ — циклическая. Ее образующий элемент — число 1. Это бесконечная группа. В качестве ее образующего можно взять и число -1.

2. Рассмотрим множество квадратных матриц второго порядка с целыми элементами и определителем, равным 1. Это группа относительно операции умножения матриц (покажите сами). Тогда матрица $A=\begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix}$ порождает бесконечную циклическую подгруппу, при этом $A^n=\begin{bmatrix}1 & n\\ 0 & 1\end{bmatrix}$.

3. Для группы самосовмещений правильного n-угольника (см. пример 2.2, п. 8) подгруппа вращений относительно центра — точки 0 состоит из поворотов на углы $\varphi_0=0$, $\varphi_1=\frac{2\pi}{n}$, ..., $\varphi_{n-1}=(n-1)\frac{2\pi}{n}$ против часовой стрелки. Это циклическая подгруппа порядка n, порожденная элементом φ_1. Из геометрических соображений ясно, что $\varphi_s=\varphi_1^s$, $\varphi_s^{-1}=\varphi_1^{n-s}$ и $\varphi_1^n=\varphi_0$ — единичное преобразование.

Теорема 2.3. *Всякая подгруппа циклической группы сама циклическая.*

Действительно, пусть $G=<a>$ — циклическая группа с образующим элементом a и H — подгруппа G, отличная от единичной. Предположим, что наименьшая положительная степень элемента a, содержащаяся в H, есть a^k. Тогда $<a^k>\subseteq H$. Допустим, что в H содержится элемент a^l, где $l\neq 0$ и l не делится на k. Тогда, если d — наибольший общий делитель чисел k и l, существуют такие целые числа u и v, что $ku+lv=d$, и, следовательно, в H должен содержаться элемент $(a^k)^u(a^l)^v=a^d$. Но так как $d<k$, то приходим к противоречию относительно выбора элемента a^k. Следовательно, $H=<a^k>$.

Пусть G — произвольная группа, a — некоторый ее элемент. Если все степени элемента a различны, то говорят, что элемент a имеет *бесконечный порядок*. Если для некоторых m и n, где $m\neq n$, $a^m=a^n$, то $a^{|m-n|}=e$, т. е. существуют положительные степени элемента a, равные единичному элементу. Пусть q — наименьшее положительное число, для которого $a^q=e$. Тогда говорят, что a — элемент *конечного* порядка q.

Рассмотрим еще один важный класс групп.

Пусть X — конечное множество из n элементов. Группа всех биекций множества X в себя называется *симметрической* группой степени n. Без ограничения общности можно считать, что множество X состоит из элементов $\{1, 2, ..., n\}$. Каждая биекция

$\varphi: X \to X$ называется *подстановкой* и записывается символом $\begin{pmatrix} 1 & 2 & \dots & n \\ i_1 & i_2 & \dots & i_n \end{pmatrix}$, где под элементом k, $1 \leqslant k \leqslant n$, записан его образ $\varphi(k) = i_k$. Произведением подстановок является композиция отображений $(\varphi\psi)(k) = \varphi(\psi(k))$. Например, для подстановок $\varphi = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{pmatrix}$ и $\psi = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \end{pmatrix}$ имеем $\varphi\psi = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 1 & 4 \end{pmatrix}$. В то же время $\psi\varphi = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 4 & 3 & 2 \end{pmatrix}$, так что $\varphi\psi \neq \psi\varphi$. Единичную (тождественную) подстановку обозначаем $e = \begin{pmatrix} 1 & 2 & \dots & n \\ 1 & 2 & \dots & n \end{pmatrix}$. Симметрическая группа степени n обозначается S_n и содержит $n!$ элементов.

Пример 2.4. Группа S_3 состоит из следующих шести элементов:

$a_1 = e = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}$, $a_2 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$, $a_3 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}$, $a_4 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}$,

$a_5 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}$, $a_6 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$.

Для подстановки $\pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 4 & 3 & 5 & 2 \end{pmatrix}$ имеем

$\pi^2 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 5 & 3 & 2 & 4 \end{pmatrix}$, $\pi^3 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 2 & 3 & 4 & 5 \end{pmatrix} = e$.

Тогда $\pi(2) = 4$, $\pi^2(2) = 5$, $\pi^3(2) = 2$.

В подстановке π элементы 1 и 3 остаются на месте, элемент 2 переходит в 4, элемент 4 — в 5, а элемент 5 — снова в 2. Такая подстановка называется циклом (245) длины 3. Этот же цикл можно записать и так: (452), (524).

В общем случае подстановка π, перемещающая элементы $j_1, j_2, \dots, j_k$ так, что $\pi(j_1) = j_2, \dots, \pi(j_{k-1}) = j_k, \pi(j_k) = j_1$ (т. е. $\pi^k(j_1) = j_1$, где k — наименьшее число, обладающее этим свойством), и оставляющая на месте остальные элементы, называется *циклом* длины k и обозначается $(j_1, \dots, j_k)$. Циклы называются *независимыми*, если любые два из них не имеют общих переставляемых элементов.

Теорема 2.4. *Каждая подстановка в S_n является произведением независимых циклов. Разложение подстановки в произведение циклов длины $\geqslant 2$ определено однозначно с точностью до порядка циклов.*

Два элемента i и j множества X называются эквивалентными относительно подстановки π, если $j = \pi^s(i)$ для некоторого целого числа s. Введенное отношение есть отношение эквивалентности на множестве X. Оно разбивает множество X на классы эквивалентности по этому отношению: $X = X_1 \cup X_2 \cup \dots \cup X_p$. Каждый элемент $i \in X$ принадлежит одному и только одному классу X_t, причем множество X_t состоит из образов элемента i при действии степеней подстановки π: $X_t = \{i, \pi(i), \pi^2(i), \dots, \pi^{k_t - 1}(i)\}$, где k_t — количество элементов в X_t. Множества X_t часто называют π-*орбитами*. В каждом классе эквивалентности

X_r выберем по одному представителю i_r и поставим ему в соответствие цикл $\pi_r=(i_r, \pi(i_r), \ldots, \pi^{k_r-1}(i_r))$ соответствующей длины k_r. Любой элемент, не принадлежащий X_r, остается на месте при действии степеней π_r. Тогда подстановка π есть произведение циклов

$$\pi=\pi_1\pi_2\ldots\pi_p. \tag{2.3}$$

Замечание 2.1. Если цикл $\pi_r=(i_r)$ имеет длину 1, то он действует как тождественная подстановка. Такие циклы в записи (2.3) можно опускать, например:

$$\begin{pmatrix}1&2&3&4&5&6&7&8\\5&2&8&7&6&1&4&3\end{pmatrix}=(156)\,(38)\,(47)\,(2)=(156)\,(38)\,(47).$$

Докажем единственность. Пусть

$$\pi=\alpha_1\alpha_2\ldots\alpha_r \tag{2.4}$$

есть разложение, отличное от (2.3); i — символ, не остающийся на месте при действии подстановки π. Тогда для одного и только одного цикла π_s из разложения (2.3) $\pi_s(i)\neq i$ и для одного и только одного цикла α_t из разложения (2.4) $\alpha_t(i)\neq i$. Для каждого $k=0, 1, 2, \ldots$ имеем $\pi^k{}_s(i)=\pi^k(i)=\alpha^k{}_t(i)$. Поскольку цикл однозначно определяется действием подстановки на символ i, не остающийся на месте, то $\pi_s=\alpha_t$. Аналогично доказываются совпадения и остальных циклов разложений (2.3) и (2.4).

Цикл длины 2 называется транспозицией. Любой цикл можно записать в виде произведения транспозиций, например:

$$(1\ 2\ldots t{-}1\ t)=(1\ t)(1\ t{-}1)\ldots(1\ 3)(1\ 2).$$

Тогда из теоремы 2.4 вытекает

Следствие. Каждая подстановка в S_n является произведением транспозиций.

Пример 2.5. В группе S_4 $(123)=(13)(12)=(23)(13)=$ $=(13)(24)(12)(14)$.

Разложение в произведение транспозиций не является единственным.

Можно доказать, что если $\pi=\tau_1\ldots\tau_k$ — разложение π в произведение транспозиций, то число $\varepsilon_\pi=(-1)^k$, называемое четностью подстановки π, не зависит от способа разложения и

$$\varepsilon_{\pi\sigma}=\varepsilon_\pi\,\varepsilon_\sigma \tag{2.5}$$

для любых двух подстановок π и σ.

Подстановка $\pi\in S_n$ называется четной, если $\varepsilon_\pi=1$, и нечетной, если $\varepsilon_\pi=-1$. Все транспозиции — нечетные подстановки.

Множество четных подстановок степени n образует подгруппу A_n, которая называется знакопеременной. Действительно, согласно (2.5) $\varepsilon_{\pi\sigma}=1$, если $\varepsilon_\pi=\varepsilon_\sigma=1$, и $\varepsilon_{\pi\tau^{-1}}=\varepsilon_\pi$, поскольку $\varepsilon_\tau=1$. Множество нечетных подстановок не образует группу, так как произведение двух нечетных подстановок есть четная подстановка.

2.1.4. Изоморфизм групп

Группы G и H называются изоморфными, если существует биекция $\varphi: G \to H$, сохраняющая групповую операцию, т. е.

$$\varphi(g_1 g_2) = \varphi(g_1)\varphi(g_2) \text{ для любых } g_1,\ g_2 \in G. \qquad (2.6)$$

Группа самосовмещений правильного треугольника (см. пример 2.2, п. 8) изоморфна симметрической группе S_3. Действительно, группа самосовмещений состоит из трех вращений φ_0, φ_1, φ_2 против часовой стрелки на углы $0, \frac{2\pi}{3}$ и $\frac{4\pi}{3}$ и трех симметрий ψ_1, ψ_2, ψ_3 относительно осей симметрии (см. рис. 2.1). Если перенумеровать множество вершин, то каждому из этих преобразований будет соответствовать подстановка на множестве $\{1, 2, 3\}$:

$$\varphi_0 \to \begin{pmatrix}1&2&3\\1&2&3\end{pmatrix},\ \varphi_1 \to \begin{pmatrix}1&2&3\\2&3&1\end{pmatrix},\ \varphi_2 \to \begin{pmatrix}1&2&3\\3&1&2\end{pmatrix},\ \psi_1 \to \begin{pmatrix}1&2&3\\1&3&2\end{pmatrix},$$
$$\psi_2 \to \begin{pmatrix}1&2&3\\3&2&1\end{pmatrix},\ \psi_3 \to \begin{pmatrix}1&2&3\\2&1&3\end{pmatrix}.$$

Это соответствие есть биекция, причем из геометрических соображений вытекает справедливость условия (2.6), т. е. групповая операция сохраняется.

Группа классов вычетов по модулю 3 изоморфна циклической группе порядка 3, порожденной элементом a, что следует из таблицы Кэли (см. табл. 2.1).

Мультипликативная группа положительных действительных чисел изоморфна аддитивной группе всех действительных чисел. Биекция $\varphi(a) = \ln a$ устанавливает изоморфизм, что следует из равенства

$$\ln(ab) = \ln a + \ln b.$$

Докажем некоторые свойства изоморфизма:

1. Единичный элемент переходит в единичный.

Поскольку для любого $g \in G$ $\varphi(e)\varphi(g) = \varphi(eg) = \varphi(g) = \varphi(ge) = \varphi(g)\varphi(e)$, то $\varphi(e) = e'$ — единичный элемент группы H.

2. Обратный элемент переходит в обратный: $\varphi(g^{-1}) = \varphi(g)^{-1}$.

Действительно, $\varphi(g)\varphi(g^{-1}) = \varphi(gg^{-1}) = \varphi(e) = e'$, аналогично $\varphi(g^{-1})\varphi(g) = e'$, откуда $\varphi(g)^{-1} = \varphi(g^{-1})$.

3. Обратное отображение $\varphi^{-1}: H \to G$ (существующее в силу того, что φ — биекция) также является изоморфизмом.

Пусть $h_1,\ h_2 \in H$. Покажем, что $\varphi^{-1}(h_1 h_2) = \varphi^{-1}(h_1)\varphi^{-1}(h_2)$. Пусть g_1 и g_2 — элсменты G такие, что $\varphi(g_1) = h_1$, $\varphi(g_2) = h_2$. Тогда $\varphi(g_1 g_2) = \varphi(g_1)\varphi(g_2) = h_1 h_2$ и, следовательно, $g_1 g_2 = \varphi^{-1}(h_1 h_2)$ или $\varphi^{-1}(h_1)\varphi^{-1}(h_2) = \varphi^{-1}(h_1 h_2)$.

Теорема 2.5. *Все циклические группы одного порядка изоморфны.*

Действительно, если $G=<g>$ — бесконечная группа, то все степени g^m различны и G изоморфна аддитивной группе целых чисел $<\mathbb{Z}, +, 0>$, поскольку биекция $\varphi(g^m)=m$ удовлетворяет условию (2.6). Пусть теперь G — конечная группа порядка q. Тогда она изоморфна группе классов вычетов по модулю q (см. пример 2.2, п. 6). Биекция φ переводит элемент g^m в класс C_m, $m=0, 1, \ldots, q-1$. Полагая $m+n=lq+r$, $0\leqslant r<q$, для любых m и n имеем

$$\varphi(g^{m+n})=\varphi(g^r)=C_r=C_m+C_n=\varphi(g^m)+\varphi(g^n).$$

Разумеется, группы одного порядка могут не быть изоморфными. Например, можно показать, что существуют ровно две изоморфные группы четвертого порядка: циклическая группа четвертого порядка $Z_4=<a>$ и так называемая четвертая группа Клейна V_4 с таблицами Кэли (см. соответственно табл. 2.2*а* и 2.2*б*).

Таблица 2.2*а*

	e	a	b	c
e	e	a	b	c
a	c	b	c	e
b	b	c	e	a
c	c	e	a	b

$b=a^2$, $c=a^3$

Таблица 2.2*б*

	e	a	b	c
e	e	a	b	c
a	a	e	c	b
b	b	c	e	a
c	c	b	a	e

$a^2=b^2=c^2=e$

Оказывается, что с точностью до изоморфизма симметрические группы описывают все конечные группы.

Теорема 2.6 *(теорема Кэли). Всякая конечная группа порядка n изоморфна некоторой подгруппе симметрической группы S_n.*

Пусть G — исходная группа порядка n, $g_1=e$, g_2, ..., g_n — ее элементы. S_n — симметрическая группа порядка n, которую можно считать группой всех биекций множества G в себя, так как природа элементов, составляющих отображаемое множество, несущественна.

Для любого элемента $a\in G$ рассмотрим отображение $L_a: G\to G$, состоящее в умножении всех элементов из G слева на a:

$$L_a(g_i)=ag_i.$$

Тогда a, ag_2, ..., ag_n — различные элементы группы G, так как $ag_i = ag_j \Rightarrow a^{-1}(ag_i) = a^{-1}(ag_j) \Rightarrow (a^{-1}a)g_i = (a^{-1}a)g_j \Rightarrow g_i = g_j$, и, следовательно, снова составляют всю группу G, отличаясь от $g_1, g_2, ..., g_n$ лишь расположением элементов. Значит, L_a — биективное отображение (подстановка); обратным ему будет отображение $L_a{}^{-1} = L_{a^{-1}}$, а единичным отображением является L_e. Вследствие ассоциативности умножения в G имеем

$$L_{ab}(g) = (ab)g = a(bg) = L_a(L_b(g)).$$

Отсюда следует, что множество $H = \{L_e, L_{g_2}, ..., L_{g_n}\}$ образует подгруппу в группе всех биективных отображений множества G в себя, т. е. в S_n. Тогда отображение $\varphi: G \to H \subset S_n$ такое, что $\varphi(a) = L_a$ для любого $a \in G$ есть изоморфизм, поскольку

$$\varphi(ab) = L_{ab} = L_a L_b = \varphi_a \varphi_b.$$

2.1.5. Смежные классы по подгруппе. Нормальные делители. Фактор-группы

Пусть H — подгруппа группы G. *Левым смежным классом* G по H называется множество gH всех элементов вида gh, где g — фиксированный элемент из G, а h пробегает все элементы подгруппы H, т. е.

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

Правый смежный класс определяется аналогично:

$$Hg = \{hg \mid h \in H\}.$$

Заметим, что одним из смежных классов является сама подгруппа H:

$$H = He = eH.$$

Теорема 2.7. *Два левых смежных класса G по H либо не пересекаются, либо совпадают, и множество левых смежных классов образует разбиение G.*

Действительно, пусть классы g_1H и g_2H имеют общий элемент $a = g_1h_1 = g_2h_2$. Тогда $g_2 = g_1h_1h_2^{-1}$, и любой элемент g_2h класса g_2H имеет вид $g_1h_1h_2^{-1}h$, где $h_1h_2^{-1}h \in H$. Значит, $g_2H \subseteq g_1H$. Аналогично доказывается, что $g_1H \subseteq g_2H$, и, следовательно, эти классы совпадают.

Так как любой элемент $g \in G$ содержится в смежном классе gH, то множество левых смежных классов образует разбиение G: $G = \bigcup gH$.

Поскольку каждое разбиение порождает отношение эквивалентности, то из теоремы 2.7 вытекает

Следствие. Отношение принадлежности к одному левому смежному классу есть отношение эквивалентности.

Два элемента $g_1, g_2 \in G$ лежат в одном левом смежном классе G по H тогда и только тогда, когда $g_2{}^{-1}g_1 \in H$. Действительно,

пусть $g_1 \in gH$, $g_2 \in gH$. Тогда $g_1 = gh_1$, $g_2 = gh_2 \Rightarrow g_2^{-1} g_1 = h_2^{-1} g^{-1} g_1 = h_2^{-1} g^{-1} g h_1 = h_2^{-1} h_1 \in H$.

Если $g_2^{-1} g_1 \in H$, то $g_1 = g_2 (g_2^{-1} g_1) \in g_2 H$. Отсюда согласно следствию из теоремы 2.7 соотношение $x \rho y \Leftrightarrow y^{-1} x \in H$ задает отношение эквивалентности на G.

Множество левых смежных классов G по H обозначается G/H. Это есть множество классов эквивалентности по отношению ρ, т. е. фактор-множество G/ρ. Мощность множества G/H называется *индексом* подгруппы H в группе G.

Аналогичные утверждения выполняются и для правых смежных классов.

Одной из важных теорем теории групп является

Теорема 2.8. *(теорема Лагранжа). Порядок конечной группы делится на порядок каждой своей подгруппы.*

Пусть порядок группы G равен n, порядок подгруппы H равен k. Из теоремы 2.7 вытекает, что G есть объединение непересекающихся левых смежных классов G по H. Пусть j — число левых смежных классов, т. е. индекс подгруппы H. Тогда $n = kj$.

Множество левых и правых смежных классов группы по одной и той же подгруппе, вообще говоря, различно. Пусть S_3 — симметрическая группа порядка 3, H — подгруппа, порожденная элементом (12), $H = <(12)>$. Тогда S_3 разбивается на следующие левые смежные классы по H: $\{e, (12)\}$, $\{(13), (123)\}$, $\{(23), (132)\}$ и следующие правые смежные классы по H: $\{e, (12)\}$, $\{(13), (132)\}$, $\{(23), (123)\}$.

Подгруппа H называется *нормальным делителем* группы G, если множество левых смежных классов G по H совпадает с множеством правых смежных классов. Это означает, что для всякого элемента $g \in G$ $gH = Hg$, т. е. для всякого элемента $g \in G$ и для всякого элемента $h \in H$ можно подобрать такие элементы $h', h'' \in H$, что $gh = h'g$ и $hg = gh''$.

Очевидно, что если G — коммутативная группа, то любая ее подгруппа является нормальным делителем G.

Под произведением двух подмножеств A и B группы G принято понимать множество всех элементов группы G вида ab, где $a \in A$, $b \in B$. Тогда, если H — нормальный делитель, произведение двух смежных классов является смежным классом, т. е. $g_1 H g_2 H = g_1 g_2 H$ в силу ассоциативности и равенства $g_2 H = H g_2$.

Таким образом, во множестве смежных классов группы G по нормальному делителю H определена операция умножения. Для того чтобы найти произведение двух смежных классов, надо произвольным образом выбрать в них по одному представителю (каждый смежный класс порождается любым своим элементом) и взять тот смежный класс, в котором лежит произведение этих представителей.

Если H — нормальный делитель G, то фактор-множество G/H, т. е. множество смежных классов G по H, является группой, которая называется *фактор-группой*.

Действительно, введенная выше операция умножения смежных классов ассоциативна, роль единицы играет сама подгруппа H: $gHH=HgH=gH$; для смежного класса gH обратным будет смежный класс $g^{-1}H$, так как $gHg^{-1}H=eH=H$.

Пример 2.6.

1. Пусть $G=(\mathbf{Z}, +, 0)$ — аддитивная группа целых чисел, H — подгруппа чисел, кратных n. Тогда G/H — циклическая группа порядка n. Она изоморфна группе классов вычетов по модулю числа n (см. пример 2.2, п. 6).

2. В симметрической группе S_n знакопеременная подгруппа A_n является нормальным делителем, фактор-группа S_n/A_n — циклическая группа порядка 2.

Заметим, что фактор-группа G/H абелевой группы — абелева: $g_1H\cdot g_2H=g_1g_2H=g_2g_1H=g_2H\cdot g_1H$. Фактор-группа циклической группы — циклическая. Действительно, пусть $G=<a>$, gH — смежный класс. Тогда для некоторого k имеем $g=a^k$ и $gH=(aH)^k$.

Задачи и упражнения

1. Доказать, что $(\{0, 1\}, \&)$ и $(\{0, 1\}, \sim)$ — моноиды. Указать единичные элементы.

2. Доказать, что $(\mathbf{Z}, *)$, где $x*y=xy+x+y$, — моноид. Найти все обратимые элементы.

3. Написать таблицу Кэли и выяснить, являются ли группой: а) вращения квадрата; б) симметрии квадрата; в) симметрии ромба; г) симметрии прямоугольника.

4. Пусть a и b — произвольные элементы группы G. Доказать, что каждое из уравнений $ax=b$ и $ya=b$ имеет и притом единственное решение в данной группе.

5. Пусть $aa=e$ для любого элемента a группы G. Доказать, что G коммутативна.

6. Доказать, что любая группа порядка 3 является коммутативной.

7. Пусть a — элемент конечного порядка q. Тогда порядок циклической подгруппы $<a>$ равен q.

8. Разложить подстановки

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

в произведение независимых циклов.

9. Найти порядок подгруппы, порожденной подстановкой $\begin{pmatrix}1&2&3&4&5\\2&3&4&5&1\end{pmatrix}$.

10. Пусть H — множество подстановок группы S_4. Будет ли оно подгруппой в следующих случаях:

а) $H=\left\{\begin{pmatrix}1&2&3&4\\1&2&3&4\end{pmatrix}, \begin{pmatrix}1&2&3&4\\2&1&4&3\end{pmatrix}, \begin{pmatrix}1&2&3&4\\3&4&1&2\end{pmatrix}, \begin{pmatrix}1&2&3&4\\4&3&2&1\end{pmatrix}\right\}$;

6) $S=\left\{\begin{pmatrix}1&2&3&4\\2&3&1&4\end{pmatrix},\ \begin{pmatrix}1&2&3&4\\1&2&3&4\end{pmatrix},\ \begin{pmatrix}1&2&3&4\\3&1&2&4\end{pmatrix},\ \begin{pmatrix}1&2&3&4\\2&4&1&3\end{pmatrix}\right\}$?

11. Описать все подгруппы симметрической группы S_3.

12. Найти все (с точностью до изоморфизма) группы, содержащие два и три элемента.

13. Какие из следующих групп изоморфны: а) группа вращений квадрата; б) группа самосовмещений ромба; в) группа самосовмещений прямоугольника; г) группа классов вычетов по модулю 4?

14. Описывая самосовмещения следующих геометрических фигур подстановками на множестве вершин, указать:

а) группу вращений тетраэдра;

б) подгруппы этой группы, изоморфные циклической группе второго и третьего порядка;

в) группу вращений куба.

15. Доказать, что пересечение нормальных делителей есть нормальный делитель.

16. Доказать, что любая подгруппа индекса 2 есть нормальный делитель.

17. Определить множество правых и левых смежных классов симметрической группы S_3 по $H=<(132)>$. Доказать, что H — нормальный делитель.

18. Доказать, что множество Z всех элементов группы G, каждый из которых перестановочен со всеми элементами этой группы, является нормальным делителем (центр группы G).

19. Доказать, что группа Клейна V_4 есть нормальный делитель симметрической группы S_4.

2.2. КОЛЬЦА И ПОЛЯ

Ознакомимся еще с двумя важнейшими понятиями алгебры — кольцом и полем.

2.2.1. Кольца: определение, свойства, примеры

Непустое множество K, на котором заданы две бинарные операции — сложение $(+)$ и умножение $(\cdot)$, удовлетворяющие условиям:

1) относительно операции сложения K — коммутативная группа;

2) относительно операции умножения K — полугруппа;

3) операции сложения и умножения связаны законом дистрибутивности, т. е. $(a+b)c=ac+bc$, $c(a+b)=ca+cb$ для всех $a, b, c\in K$, называется *кольцом* $(K, +, \cdot)$.

Структура $(K, +)$ называется *аддитивной группой* кольца. Если операция умножения коммутативна, т. е. $ab=ba$ для всех $a, b\in K$, то кольцо называется *коммутативным*.

Если относительно операции умножения существует единичный элемент, который в кольце принято обозначать единицей 1, то говорят, что K есть *кольцо с единицей*.

Подмножество L кольца называется *подкольцом*, если L — подгруппа аддитивной группы кольца и L замкнуто относительно операции умножения, т. е. для всех a, $b \in L$ выполняется $a-b \in L$ и $ab \in L$.

Пересечение подколец будет подкольцом. Тогда, как и в случае групп, подкольцом, *порожденным* множеством $S \subset K$, называется пересечение всех подколец K, содержащих S.

Пример 2.7.

1. Множество целых чисел относительно операций умножения и сложения $(\mathbb{Z}, +, \cdot)$ — коммутативное кольцо. Множество $n\mathbb{Z}$ целых чисел, делящихся на n, будет подкольцом без единицы при $n>1$.

Аналогично множество рациональных и действительных чисел — коммутативные кольца с единицей.

2. Множество квадратных матриц порядка n относительно операций сложения и умножения матриц есть кольцо с единицей E — единичной матрицей. При $n>1$ оно некоммутативное.

3. Пусть K — произвольное коммутативное кольцо. Рассмотрим всевозможные многочлены

$$a_0+a_1x+a_2x^2+\ldots+a_nx^n, \quad n \geqslant 0,$$

с переменной x и коэффициентами $a_0, a_1, a_2, \ldots, a_n$ из K. Относительно алгебраических операций сложения и умножения многочленов — это коммутативное кольцо. Оно называется *кольцом многочленов* $K[x]$ от переменной x над кольцом K (например, над кольцом целых, рациональных, действительных чисел). Аналогично определяется кольцо многочленов $K[x_1, \ldots, x_n]$ от n переменных как кольцо многочленов от одной переменной x_n над кольцом $K[x_1, \ldots, x_{n-1}]$.

4. Пусть X — произвольное множество, K — произвольное кольцо. Рассмотрим множество всех функций $f: X \to K$, определенных на множестве X со значениями в K. Определим сумму и произведение функций, как обычно, равенствами

$$(f+g)(x)=f(x)+g(x); \quad (fg)(x)=f(x)g(x),$$

где $+$ и $\cdot$ — операции в кольце K.

Нетрудно проверить, что все условия, входящие в определение кольца, выполняются, и построенное кольцо будет коммутативным, если коммутативно исходное кольцо K. Оно называется *кольцом функций* на множестве X со значениями в кольце K.

5. Рассмотрим множество $C_0, C_1, \ldots, C_{n-1}$ классов вычетов по модулю n (см. пример 2.2, п. 6). Оно образует коммутативную группу относительно операции сложения:

$C_k+C_l=C_r$, где $k+l=r \pmod n$, $0 \leqslant r < n$.

Определим операцию умножения классов вычетов:
$C_k C_l = C_r$, где $kl = r \pmod n$, $0 \leqslant r < n$.

Так как эти операции сводятся к соответствующим операциям над числами из классов вычетов, то множество классов вычетов также есть коммутативное кольцо с единицей C_1, которое обозначается Z_n. Это важный пример кольца, состоящего из конечного числа элементов.

Приведем таблицы сложения и умножения в Z_4 (см. соответственно табл. 2.3*а* и 2.3*б*).

Таблица 2.3*а*

+	C_0	C_1	C_2	C_3
C_0	C_0	C_1	C_2	C_3
C_1	C_1	C_2	C_3	C_0
C_2	C_2	C_3	C_0	C_1
C_3	C_3	C_0	C_1	C_2

Таблица 2.3*б*

·	C_0	C_1	C_2	C_3
C_0	C_0	C_0	C_0	C_0
C_1	C_0	C_1	C_2	C_3
C_2	C_0	C_2	C_0	C_2
C_3	C_0	C_3	C_2	C_1

Многие свойства колец — это переформулированные соответствующие свойства групп и полугрупп, например: $a^m a^n = a^{m+n}$, $(a^m)^n = a^{mn}$ для всех m, $n \in N$ и всех $a \in K$.

Другие специфические свойства колец моделируют свойства чисел:

1) для всех $a \in K$ $a \cdot 0 = 0 \cdot a = 0$;

2) $(-a)b = a(-b) = -(ab)$;

3) $-a = (-1)a$.

Действительно:

1) $a + 0 = a \Rightarrow a(a+0) = aa \Rightarrow a^2 + a \cdot 0 = a^2 \Rightarrow a^2 + a \cdot 0 = a^2 + 0 \Rightarrow a \cdot 0 = 0$ (аналогично $0 \cdot a = 0$);

2) $0 = a \cdot 0 = a(b - b) = ab + a(-b) \Rightarrow a(-b) = -(ab)$ (аналогично $(-a)b = -(ab)$);

3) используя второе свойство, имеем
$-a = (-a)1 = a(-1) = (-1)a$.

2.2.2. Поле

В кольцах целых, рациональных и действительных чисел из того, что произведение $ab = 0$, следует, что либо $a = 0$, либо $b = 0$. Но в кольце квадратных матриц порядка $n > 1$ это свойство уже не выполняется, так как, например, $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$.

Если в кольце K $ab=0$ при $a\neq 0$, $b\neq 0$, то a называется левым, а b — правым *делителем нуля*. Если в K нет делителей нуля (кроме элемента 0, который является тривиальным делителем нуля), то K называется кольцом *без делителей нуля*.

Пример 2.8.

1. В кольце функций $f: \mathbf{R}\to\mathbf{R}$ на множестве действительных чисел $\mathbf{R}$ рассмотрим функции $f_1(x)=|x|+x$; $f_2(x)=|x|-x$. Для них $f_1(x)=0$ при $x\leqslant 0$ и $f_2(x)=0$ при $x\geqslant 0$, а поэтому произведение $f_1(x)f_2(x)$ — нулевая функция, хотя $f_1(x)\not\equiv 0$ и $f_2(x)\not\equiv 0$. Следовательно, в этом кольце есть делители нуля.

2. Рассмотрим множество пар целых чисел (a, b), в котором заданы операции сложения и умножения:

$$(a_1, b_1)+(a_2, b_2)=(a_1+a_2,\ b_1+b_2);$$
$$(a_1, b_1)\,(a_2, b_2)=(a_1a_2,\ b_1b_2).$$

Это множество образует коммутативное кольцо с единицей $(1,1)$ и делителями нуля, так как $(1,0)\,(0,1)=(0,0)$.

3. В кольце Z_4 элемент C_2 — делитель нуля (см. табл. 2.3*а*).

Если в кольце нет делителей нуля, то в нем выполняется закон сокращения, т. е. $ab=ac$, $a\neq 0\Rightarrow b=c$. Действительно, $ab-ac=0\Rightarrow a(b-c)=0\Rightarrow (b-c)=0\Rightarrow b=c$.

Пусть K — кольцо с единицей. Элемент a называется *обратимым*, если существует такой элемент a^{-1}, для которого $aa^{-1}=a^{-1}a=1$.

Обратимый элемент не может быть делителем нуля, так как если $ab=0$, то $a^{-1}(ab)=0\Rightarrow (a^{-1}a)b=0\Rightarrow 1b=0\Rightarrow b=0$ (аналогично $ba=0\Rightarrow b=0$).

Теорема 2.9. *Все обратимые элементы кольца K с единицей образуют группу относительно умножения.*

Действительно, умножение в K ассоциативно, единица содержится во множестве обратимых элементов и произведение не выводит из множества обратимых элементов, так как если a и b обратимы, то $(ab)^{-1}=b^{-1}a^{-1}$.

Важную алгебраическую структуру образуют коммутативные кольца K, в которых каждый ненулевой элемент обратим, т. е. относительно операции умножения множество $K\setminus\{0\}$ образует группу. В таких кольцах определены три операции: сложение, умножение и деление.

Коммутативное кольцо P с единицей $1\neq 0$, в котором каждый ненулевой элемент обратим, называется *полем*.

Относительно умножения все отличные от нуля элементы поля образуют группу, которая называется *мультипликативной группой* поля.

Произведение ab^{-1} записывается в виде дроби $\frac{a}{b}$ и имеет смысл лишь при $b\neq 0$. Элемент $\frac{a}{b}$ является единственным ре-

шением уравнения $bx=a$. Действия с дробями подчиняются привычным для нас правилам:

$$\frac{a}{b}=\frac{c}{d}\Leftrightarrow ad=bc,\ b,\ d\neq 0;\ \frac{a}{b}+\frac{c}{d}=\frac{ad+bc}{bd},\ b,\ d\neq 0;$$

$$-\frac{c}{b}=\frac{-a}{b}=\frac{a}{-b},\ b\neq 0;\ \frac{a}{b}\cdot\frac{c}{d}=\frac{ac}{bd},\ b,\ d\neq 0;$$

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

Докажем, например, второе из них. Пусть $x=\frac{a}{b}$ и $y=\frac{c}{d}$ — решения уравнений $bx=a$, $dy=c$. Из этих уравнений следует $dbx=da$, $bdy=bc\Rightarrow bd(x+y)=da+bc\Rightarrow t=\frac{da+bc}{bd}$ — единственное решение уравнения $bdt=da+bc$.

Пример 2.9.

1. Кольцо целых чисел не образует поля. Полем является множество рациональных и множество действительных чисел.

2. Рассмотрим кольцо классов вычетов Z_n (см. пример 2.7, п. 5). Покажем, что Z_n является полем тогда и только тогда, когда $n=p$ — простое число. Если $n=p$ — простое число, то Z_p — множество из p элементов C_0, C_1, ..., C_{p-1}. Докажем, что любой элемент C_s, кроме C_0, обратим. Числа s и p взаимно просты. Следовательно, существуют такие целые числа l и m, что имеет место равенство $sl+pm=1$, причем l можно выбрать так, что $1\leqslant l<p$. Отсюда следует, что в кольце вычетов Z_p выполняется $C_sC_l=C_1$, т. е. C_l — элемент, обратный C_s. Если n — составное число ($n=st$), то $C_sC_t=C_0$. Следовательно, C_s — делитель нуля и Z_n — не поле.

Поле вычетов Z_p — это пример поля, состоящего из конечного числа элементов. Следовательно, существуют конечные поля.

Рассмотрим теперь аддитивную группу поля $(P, +)$. Единичный элемент поля 1 есть элемент этой группы. Рассмотрим подгруппу, порожденную 1. Она состоит из всех кратных 1:

$$n1=1+...+1,\ (-n)\cdot 1=-(n\cdot 1)=n(-1);\ 0\cdot 1=0.$$

Так как $1\neq 0$, то ее порядок не меньше двух.

Характеристикой поля P называется число, равное 0, если элемент 1 порождает подгруппу бесконечного порядка, и порядку p этой подгруппы, если он конечен.

Покажем, что если характеристика поля P не равна 0, то p — простое число. Действительно, пусть p — составное число ($p=a\cdot b$). По определению, $p1=0$. Тогда $(ab)1=(a1)(b1)=0$. Но в поле нет делителей нуля. Следовательно, $a1=0$ или $b1=0$. Но это противоречит тому, что p — порядок подгруппы, порожденной 1.

Понятие характеристики есть одно из важных структурных понятий поля.

Задачи и упражнения

1. Пусть $P(X)$ — множество всех подмножеств множества X с операциями

$$A+B=(A\cup B)\setminus(A\cap B),\ AB=A\cap B,\ A,\ B\subseteq X.$$

Доказать, что оно является кольцом с единицей, все элементы аддитивной группы которого имеют порядок 2.

2. Доказать, что матрицы вида $\begin{bmatrix} a & b \\ 2b & a \end{bmatrix}$ с действительными a и b образуют кольцо относительно обычных операций сложения и умножения матриц.

3. Будут ли следующие множества подкольцами указанных колец:

а) множество $\mathbf{Z}$ целых чисел в кольце $\mathbf{Z}[x]$ целочисленных многочленов;

б) множество $n\mathbf{Z}[x]$ многочленов, коэффициенты которых кратны числу $n>1$, в кольце $Z[x]$ целочисленных многочленов;

в) множество $\mathbf{Z}$ целых чисел в кольце A целых гауссовых чисел, т. е. чисел вида $a+bi$ с целыми a, b?

4. Образуют ли поле относительно сложения и умножения чисел:

а) комплексные числа;

б) комплексные числа вида $a+bi$ с целыми a и b;

в) комплексные числа вида $a+bi$ с рациональными a и b?

5. Образует ли поле множество матриц вида $\begin{bmatrix} a & b \\ 2b & a \end{bmatrix}$:

а) с рациональными a, b;

б) с действительными a, b?

2.3. ЭЛЕМЕНТЫ ТЕОРИИ КОДИРОВАНИЯ

В теории передачи информации чрезвычайно важным является решение проблемы кодирования и декодирования, обеспечивающее надежную передачу по каналам связи с «шумом».

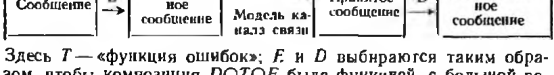
Передача информации сводится к передаче по некоторому каналу связи символов некоторого алфавита. Однако в реальных ситуациях сигналы при передаче практически всегда могут искажаться, и переданный символ будет восприниматься неправильно. Например, в системе ЭВМ — ЭВМ одна из вычислительных машин может быть связана с другой через спутник. Канал связи в этом случае физически реализуется электромагнитным полем между поверхностью Земли и спутником. Электромагнитные сигналы, накладываясь на внешнее поле, могут исказиться и ослабиться. Для обеспечения надежности передачи информа-

ции в таких системах разработаны эффективные методы, использующие коды различных типов.

Рассмотрим одну из таких моделей, связанную с *групповыми кодами*.

Алфавит, в котором записываются сообщения, считаем состоящим из двух символов {0, 1}. Он называется двоичным алфавитом. Тогда сообщение есть конечная последовательность символов этого алфавита. Сообщение, подлежащее передаче, кодируется по определенной схеме более длинной последовательностью символов в алфавите {0, 1}. Эта последовательность называется *кодом* или *кодовым словом*. При приеме можно исправлять или распознавать ошибки, возникшие при передаче по каналу связи, анализируя информацию, содержащуюся в дополнительных символах. Принятая последовательность символов декодируется по определенной схеме в сообщение, с большой вероятностью совпадающее с переданным.

Блочный двоичный (m, n)-код определяется двумя функциями: $E: \{0, 1\}^m \to \{0, 1\}^n$ и $D: \{0, 1\}^n \to \{0, 1\}^m$, где $m \leqslant n$ и $\{0, 1\}^n$ — множество всех двоичных последовательностей длины n. Функция E определяет схему кодирования, а функция D — схему декодирования. Математическую модель системы связи можно представить в виде схемы



Здесь T — «функция ошибок»; E и D выбираются таким образом, чтобы композиция $DOTOE$ была функцией, с большой вероятностью близкой к тождественной. Двоичный (m, n)-код содержит 2^m кодовых слов.

Коды делятся на два больших класса: коды с обнаружением ошибок, которые с большой вероятностью определяют наличие ошибки в принятом сообщении, и коды с исправлением ошибок, которые с большой вероятностью могут восстановить посланное сообщение.

Пример 2.10.

1. Код с проверкой четности. Это пример простого кода, с большой вероятностью указывающего на наличие ошибки.

Пусть $a = a_1 \ldots a_m$ — сообщение длины m.

Схема кодирования определяется таким образом:

$$E(a) = b = b_1 \ldots b_{m+1},$$

где $b_i = a_i$ при $i = 1, \ldots, m$;

$$b_{m+1} = \begin{cases} 0, \text{ если } \sum\limits_{i=1}^{m} b_i \text{ — четное число;} \\ 1, \text{ если } \sum\limits_{i=1}^{m} b_i \text{ — нечетное число.} \end{cases}$$

Схема декодирования определяется таким образом:

$D(b)=c=c_1...c_m$,

где $c_i=b_i$ при $i=1, ..., m$.

Например, при $m=2$ $E(00)=000$, $E(01)=011$, $E(10)=101$, $E(11)=110$. Очевидно, подразрядная сумма любой кодированной последовательности $\sum_{i=1}^{m+1} b_i$ должна быть четной, т. е. $b_1+...+b_{m+1}=0 (\mathrm{mod} 2)$. Если $\sum_{i=1}^{m+1} b_i$ нечетная, то это означает, что при передаче сообщения произошла ошибка. Однако, если $\sum_{i=1}^{m+1} b_i$ четная, то это еще не означает, что ошибки не было. Поразрядная сумма остается четной при двух ошибках и вообще любом четном их числе.

2. Код с тройным повторением. Это пример весьма неэкономного кода с исправлением ошибок.

Схема кодирования, т. е. функция E, определяется таким образом: каждое сообщение разбивается на блоки длины m и каждый блок передается трижды. Схема декодирования, т. е. функция D, определяется следующим образом: принятое сообщение разбивается на блоки длины $3m$ и в каждом блоке из трех символов b_i, b_{i+m}, b_{i+2m}, принимающих значение 0 или 1, выбирается тот, который встретился большее число раз (два или три раза). Этот символ считается i-м символом в декодированном сообщении.

Можно определить также $(1, r)$-код с r-кратным повторением, в котором каждый символ 0 или 1 кодируется последовательностью из r таких символов. Множество кодовых слов состоит из двух слов: $\underbrace{00...0}_{r \text{ раз}}$ и $\underbrace{11...1}_{r \text{ раз}}$. При декодировании решение принимается «большинством голосов», т. е. переданным считается символ, встречающийся большее число раз. При больших r мы практически обезопасим себя от ошибок, однако передача сообщений будет идти чрезвычайно медленно.

2.3.1. Расстояние Хемминга

На множестве двоичных слов длины m расстоянием $d(a, b)$ между словами a и b называют число несовпадающих позиций этих слов, например: расстояние между словами $a=01101$ и $b=00111$ равно 2.

Определенное таким образом понятие называется *расстоянием Хемминга*. Оно удовлетворяет следующим *аксиомам расстояний*:

1) $d(a, b)\geqslant 0$ и $d(a, b)=0$ тогда и только тогда, когда $a=b$;

2) $d(a, b)=d(b, a)$;

3) $d(a, b)+d(b, c)\geqslant d(a, c)$ (неравенство треугольника).

Весом $w(a)$ слова a называют число единиц среди его координат. Тогда расстояние между словами a и b есть вес их суммы $a+b$: $d(a, b)=w(a+b)$, где символом $+$ обозначена операция покоординатного сложения по модулю 2.

Интуитивно понятно, что код тем лучше приспособлен к обнаружению и исправлению ошибок, чем больше различаются кодовые слова. Понятие расстояния Хемминга позволяет это уточнить.

Теорема 2.10. *Для того чтобы код позволял обнаруживать ошибки в k (или менее) позициях, необходимо и достаточно, чтобы наименьшее расстояние между кодовыми словами было* $\geqslant k+1$.

Доказательство этой теоремы аналогично доказательству следующего утверждения.

Теорема 2.11. *Для того чтобы код позволял исправлять все ошибки в k (или менее) позициях, необходимо и достаточно, чтобы наименьшее расстояние между кодовыми словами было* $\geqslant 2k+1$.

Действительно, если наименьшее расстояние между кодовыми словами $\geqslant 2k+1$, то для любых кодовых слов a и b имеем $d(a, b)\geqslant 2k+1$. Пусть при передаче некоторого слова a произошло $r\leqslant k$ ошибок, в результате чего было принято слово c. Тогда $d(a, c)=r\leqslant k$. Из неравенства треугольника (аксиомы 3) следует, что $d(a, c)+d(c, b)\geqslant d(a, b)\geqslant 2k+1$. Отсюда расстояние $d(c, b)$ от слова c до любого другого кодового слова b больше k. Значит, для декодирования посланного слова надо найти кодовое слово a, ближайшее к принятому слову c в смысле расстояния Хемминга.

Если наименьшее расстояние между кодовыми словами меньше, чем $2k+1$, то найдутся такие два кодовых слова a и b, расстояние между которыми будет $d(a, b)\leqslant 2k$. Тогда, если в кодовом слове a будет k ошибок, принятое слово c находится от другого кодового слова b на расстоянии, не большем, чем от a. Поэтому нельзя определить, какое из слов (a или b) было передано.

В математической модели кодирования и декодирования удобно рассматривать строки ошибок. Данное сообщение $a=a_1a_2...a_m$ кодируется кодовым словом $b=b_1b_2...b_n$. При передаче канал связи добавляет к нему строку ошибок $e=e_1e_2...e_n$, так что приемник принимает сигнал $c=c_1c_2...c_n$, где $c_i=b_i+e_i$. Система, исправляющая ошибки, переводит слово $c_1c_2...c_n$ в ближайшее кодовое слово $b_1b_2...b_n$. Система, обнаруживающая ошибки, только устанавливает, является ли принятое слово кодовым или нет. Последнее означает, что при передаче произошла ошибка.

Пример 2.11.

1. Рассмотрим (2, 3)-код с проверкой четности. Тогда (см. пример 2.10, п. 1) множество кодовых слов есть 000, 101, 011, 110. Минимальное расстояние между кодовыми словами равно двум. Этот код способен обнаруживать однократную ошибку.

2. Рассмотрим (2, 5)-код со схемой кодирования $E(00) = 00000 = b^1$; $E(01) = 01011 = b^2$; $E(10) = 10101 = b^3$; $E(11) = 11110 = b^4$. Минимальное расстояние между кодовыми словами равно трем. Этот код способен исправлять однократную ошибку. Однократная ошибка приводит к приему слова, находящегося на расстоянии 1 от единственного кодового слова, которое и было передано.

2.3.2. Матричное кодирование

При явном задании схемы кодирования в (m, n)-коде следует указать 2^m кодовых слов, что весьма неэффективно.

Одним из экономных способов описания схемы кодирования является методика матричного кодирования.

Пусть $G = \|g_{ij}\|$ — матрица порядка $m \times n$ с элементами g_{ij}, равными 0 или 1. Символ $+$ обозначает сложение по модулю 2. Тогда схема кодирования задается системой уравнений

$$b_j = a_1 g_{1j} + a_2 g_{2j} + \ldots + a_m g_{mj} = \sum_{i=1}^{m} a_i g_{ij}, \quad j = 1, \ldots, n,$$

или в матричной форме

$$b = aG,$$

где $a = a_1 \ldots a_m$ — вектор, соответствующий передаваемому сообщению; $b = b_1 \ldots b_n$ — вектор, соответствующий кодированному сообщению; G — порождающая матрица кода.

Порождающая матрица кода определена неоднозначно. Код не должен приписывать различным словам-сообщениям одно и то же кодовое слово. Можно доказать, что этого не произойдет, если первые m столбцов матрицы G образуют единичную матрицу.

Заметим, что вместо 2^m кодовых слов достаточно знать m слов, являющихся строками матрицы G.

Пример 2.12.

1. Порождающей матрицей (1, r)-кода с повторением является матрица

$$G = [1 \ldots 1],$$

так как $1 \ldots 1 = 1G$, $0 \ldots 0 = 0G$.

2. Порождающей матрицей (2, 3)-кода с проверкой четности является матрица

$$G = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

3. Рассмотрим матрицу G порядка 3×6:

$$G=\begin{bmatrix}1&0&0&1&1&0\\0&1&0&0&1&1\\0&0&1&1&1&1\end{bmatrix}.$$

Сообщения $a^1=100$, $a^2=010$, $a^3=001$ кодируются соответственно первой, второй и третьей строками матрицы G. Полный список кодирования следующий:

$a^0=000\to 000000$;
$a^1=100\to 100110$;
$a^2=010\to 010011$;
$a^3=110\to 110101$;
$a^4=001\to 001111$;
$a^5=101\to 101001$;
$a^6=011\to 011100$;
$a^7=111\to 111010$.

2.3.3. Групповые коды

Двоичный (m, n)-код называется *групповым*, если его кодовые слова образуют группу.

Заметим, что множество всех двоичных слов длины m образует коммутативную группу с операцией покоординатного сложения по модулю 2, в которой выполняется соотношение $a+a=0$. Следовательно, множество слов-сообщений a длины m есть коммутативная группа.

Пусть G — порождающая матрица кода порядка $m\times n$. Тогда множество кодовых слов $b=aG$ есть группа, так как если $b^1=a^1G$, $b^2=a^2G$, то $b^1+b^2=a^1G+a^2G=(a^1+a^2)G$, т. е. матричные коды являются групповыми.

В групповом коде наименьшее расстояние между кодовыми словами равно наименьшему весу ненулевого кодового слова, что видно из соотношения $d(b^i,b^j)=w(b^i+b^j)$. Следовательно, код, рассмотренный в примере 2.11, п. 2, способен исправлять однократную ошибку и обнаруживать двойную, так как наименьший вес кодового слова равен 3.

Пусть задан групповой код с порождающей матрицей G и кодирование происходит по схеме $b=aG$.

При передаче кодовые слова могут исказиться, в результате чего будет принято сообщение $c=c_1\ldots c_n$, где $c=b+e$ и $e=e_1\ldots e_n$ — вектор (строка) ошибок.

Предложим схему декодирования, при которой вероятность того, что $D(aG)\neq a$, будет минимальной.

Обозначим через C множество всех слов, которые могут быть приняты. Это есть множество всех двоичных слов длины n. Оно образует коммутативную группу. Множество B всех кодовых слов есть подгруппа C. Рассмотрим множество смежных

классов C по B, т. е. фактор-группу C/B. *Лидером* смежного класса назовем слово, имеющее наименьший вес. Поскольку смежные классы либо не пересекаются, либо совпадают, то любой элемент $c \in C$ однозначно представляется в виде суммы $c=e+b$ лидера e и кодового слова b. Декодирование слова c состоит в выборе кодового слова b в качестве переданного и в последующем переходе к слову a, где $b=E(a)$.

Данный метод кодирования удобно реализовать с помощью таблицы, первая строка которой представляет собой множество кодовых слов, т. е. смежный класс $0+B$, состоящий из элементов $0, b^1, \ldots, b^{2^m-1}$, а остальные строки соответствуют остальным смежным классам по B, причем первый столбец этой таблицы есть столбец лидеров.

Для примера 2.12 (см. п. 3) таблица декодирования выглядит таким образом:

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

Чтобы декодировать принятое слово $c=e^i+b^l$, где e^i — лидер, следует отыскать его в таблице и выбрать в качестве переданного кодовое слово b^l, находящееся в первой строке того же столбца, что и c. Например, если принято слово 110011, то считается, что передано слово 010011; если принято слово 100101, то передано слово 110101, а если принято слово 110101, то считается, что оно и было передано.

Покажем, что при таком способе декодирования:

1) исправляются все строки ошибок, являющиеся лидерами;

2) кодовое слово, стоящее в данном столбце, является ближайшим кодовым словом ко всем словам этого столбца.

Действительно:

1. Если при передаче слова b произошла ошибка e, где e — лидер, то $c=b+e$ и есть искомое представление слова c, и при декодировании считается, что передано кодовое слово b, т. е. ошибка e исправляется.

2. Пусть c — слово, стоящее в том же столбце, что и кодовое слово b^i. Тогда $c=b^i+e$, где e — лидер соответствующего смежного класса. Имеем $d(c, b^i)=w(e)$. Если b^j — другое кодовое слово, то $c=b^j+e'$ и, поскольку $b^i-b^j \in B$, $e'=b^i-b^j+e$ лежит в том же смежном классе, что и e. Следовательно, $d(c, b^j)=w(e') \geqslant w(e)$.

2.3.4. Коды Хемминга

Опишем один из классов групповых кодов — коды Хемминга, которые исправляют однократную ошибку, поскольку минимальный вес кодового слова равен 3. Это (n, n)-коды, где $n=2^r-1-r$, $n=2^r-1$ для любого $r\geqslant 2$.

Схема кодирования:

1. Сообщения — слова длины 2^r-1-r, где $r\geqslant 2$; кодовые слова имеют длину 2^r-1.

2. В каждом кодовом слове $b=b_1 \ldots b_{2^r-1}$ символы, индексы которых являются степенью двойки, т. е. $b_{2^0}, b_{2^1}, \ldots, b_{2^{r-1}}$ — контрольные, а остальные — символы сообщения, расположенные в том же порядке. Например, при $r=4$ b_1, b_2, b_4, b_8 — контрольные символы, $b_3, b_5, b_6, b_7, b_9, b_{10}, b_{11}, b_{12}, b_{13}, b_{14}, b_{15}$ — символы сообщения.

3. Рассмотрим матрицу M порядка $r\times(2^r-1)$ такую, что в i-м столбце этой матрицы стоят символы двоичного разложения числа i. Тогда матрицы M при $r=2, 3, 4$ имеют соответственно вид

$$M_{2,3}=\begin{bmatrix}011\\101\end{bmatrix};\quad M_{3,7}=\begin{bmatrix}0001111\\0110011\\1010101\end{bmatrix};$$

$$M_{4,15}=\begin{bmatrix}000000011111111\\000111100001111\\011001100110011\\101010101010101\end{bmatrix}.$$

4. Запишем систему уравнений

$$bM^T=0. \tag{2.7}$$

Например, при $r=3$ эта система имеет вид

$$\begin{aligned}&b_1+b_5+b_6+b_7=0;\\ &b_2+b_3+b_6+b_7=0;\\ &b_4+b_5+b_6+b_7=0.\end{aligned} \tag{2.8}$$

Заметим, что по построению матрицы M в каждое из уравнений системы (2.7) входит один и только один символ b_{2^i}, индекс которого есть степень двойки.

5. При кодировании сообщения значения контрольных символов $b_{2^0}, b_{2^1}, \ldots, b_{2^{r-1}}$ получим из системы (2.7).

Схема декодирования.

Пусть принято слово $c=b+e$, где b — кодовое слово, e — ошибка. Тогда $bM^T=0$, и, следовательно, $(b+e)M^T=bM^T+eM^T=eM^T$.

Если $eM^T=0$, то считается, что ошибки не было. Это действительно так при $e=0$.

Если произошла ошибка ровно в одной позиции, т. е. вектор ошибок e имеет только одну единицу в i-й позиции, то eM^T

есть вектор, совпадающий с i-м столбцом матрицы M, являющийся двоичным разложением числа i. В этом случае в переданном слове $c=b+e$ надо изменять символ в i-й позиции и вычеркнуть контрольные символы. Тогда полученное слово будет результатом декодирования.

Если ошибка допущена более чем в одной позиции, декодирование дает неверный результат. Например, если строка ошибок e будет кодовым словом, то $(b+e)M^T=0$, и, следовательно, в результате декодирования слово не изменится.

Пример 2.13. Найдем порождающую матрицу для (4,7)-кода Хемминга. Определим фундаментальную систему решений системы уравнений (2.8):

$a_1=1110000$, $a_2=1001100$, $a_3=0101010$, $a_4=1101001$.

Порождающей будет матрица, составленная из этих векторов:

$$G=\begin{bmatrix} 1110000 \\ 1001100 \\ 0101010 \\ 1101001 \end{bmatrix}.$$

Пример 2.14. Рассмотрим (4,7)-код Хемминга. Пусть $c=0011111$. Тогда $c=b+e$, где $b=0001111$ — кодовое слово, $e=0010000$ — строка ошибок, и $cM^T=(b+e)M^T=eM^T=011$. Число 011 есть двоичное разложение числа 3. Следовательно, ошибка совершена в третьей позиции. Исправляя ее, получаем кодовое слово 0001111, декодируя которое, получаем слово $a=0111$.

Задачи и упражнения

1. Определить положение одиночной ошибки в искаженном слове 1100011 (4,7)-кода Хемминга. Какое слово было передано?

2. Пусть 11010011 и 11001111 — искаженные слова (4,7)-кода Хемминга с проверкой на четность. Какое из слов содержит одиночную, а какое двойную ошибку? Определить положение одиночной ошибки.

КОМБИНАТОРИКА

ГЛАВА 3

Основная задача комбинаторики — пересчет и перечисление элементов в конечных множествах. Если нас интересует, сколько элементов, принадлежащих заданному конечному множеству, обладает некоторым свойством или заданным набором свойств, то это задача *пересчета*. Если для каких-либо целей необходимо выделить все элементы множества, удовлетворяющие заданным свойствам, то это задача *перечисления*. В некоторых задачах на исходном конечном множестве элементов определена некоторая целевая функция, причем нас интересуют элементы множества, доставляющие минимальное (или максимальное) значение этой функции. В этом случае имеем задачу *оптимизации*. При этом под решением задачи оптимизации *в сильном смысле* понимается совокупность всех элементов, доставляющих минимальное (или максимальное) значение целевой функции, а под решением задачи *в слабом смысле* — произвольный элемент, доставляющий минимальное (или максимальное) значение целевой функции. Иногда интересуются лишь минимальным (или максимальным) значением функции. Перечисленные задачи тесно связаны друг с другом. Например, при решении задач оптимизации обычно предполагается, что мы располагаем методом перечисления элементов исходного множества (которое, как правило, является совокупностью элементов некоторого более широкого множества, удовлетворяющих заданным свойствам), а для того чтобы оценить эффективность методов перечисления или оптимизации, часто целесообразно решить задачу пересчета элементов в исходном множестве или в некоторых его подмножествах.

3.1. КОМБИНАТОРНЫЕ СХЕМЫ

Приведем некоторые начальные сведения из комбинаторики.

3.1.1. Правила суммы, произведения

Пусть X - конечное множество, состоящее из n элементов. Тогда говорят, что объект x из X может быть выбран n способами, и пишут $|X| = n$. Пусть $X_1, \ldots, X_k$ — попарно непересекающиеся множества, т. е. $X_i \cap X_j = \emptyset$ при $i \neq j$. Тогда, очевидно, выполняется равенство

$$\left| \bigcup_{i=1}^{k} X_i \right| = \sum_{i=1}^{k} |X_i|.$$

В комбинаторике этот факт называется *правилом суммы*. Для $k=2$ оно формулируется следующим образом. Если объект x может быть выбран m способами, а объект y — другими n способами, то выбор «либо x, либо y» может быть осуществлен $m+n$ способами.

Другим часто применяемым в комбинаторике правилом является *правило произведения*. Если объект x может быть выбран m способами и после каждого из таких выборов объект y в свою очередь может быть выбран n способами, то выбор упорядоченной пары $\langle x, y \rangle$ может быть осуществлен mn способами.

Докажем правило произведения, используя правило суммы. Пусть $\{a_1, \ldots, a_m\}$ — множество элементов, из которых выбирается объект x, и для каждого $i \in \{1, \ldots, m\}$ X_i — множество пар $\langle x, y \rangle$ при $x = a_i$. Тогда множества X_i попарно не пересекаются, $|X_i| = n$, $i = 1, 2, \ldots, m$, множество всех пар $\langle x, y \rangle$ есть $\bigcup_{i=1}^{m} X_i$, и по правилу суммы имеем

$$\left| \bigcup_{i=1}^{m} X_i \right| = \sum_{i=1}^{m} |X_i| = mn.$$

Мы сформулировали и доказали правило произведения для последовательности из двух объектов. В общем случае правило произведения формулируется следующим образом. Если объект x_1 может быть выбран n_1 способами, после чего объект x_2 может быть выбран n_2 способами и для любого i, где $2 \leqslant i \leqslant m-1$, после выбора объектов $x_1, \ldots, x_i$ объект x_{i+1} может быть выбран n_{i+1} способами, то выбор упорядоченной последовательности из m объектов $\langle x_1, x_2, \ldots, x_m \rangle$ может быть осуществлен $n_1 n_2 \ldots n_m$ способами.

Обобщенное правило произведения является следствием правила произведения для упорядоченной пары объектов и доказывается методом математической индукции.

3.1.2. Размещения и сочетания

Набор элементов $x_{i_1}, \ldots, x_{i_r}$ из множества $X = \{x_1, \ldots, x_n\}$ называется *выборкой* объема r из n элементов или, иначе, (n, r)-выборкой.

Выборка называется *упорядоченной*, если порядок следования элементов в ней задан. Две упорядоченные выборки, различающиеся лишь порядком следования элементов, считаются различными.

Если порядок следования элементов в выборке не является существенным, то такая выборка называется *неупорядоченной*.

В выборках могут допускаться или не допускаться повторения элементов. Упорядоченная (n, r)-выборка, в которой элементы могут повторяться, называется (n, r)-*размещением с повторениями*. Если элементы упорядоченной (n, r)-выборки попарно различны, то она называется (n, r)-*размещением без повторений* или просто (n, r)-*размещением*. Будем, кроме того, (n, n)-размещения без повторений называть *перестановками* множества X. Неупорядоченная (n, r)-выборка, в которой элементы могут повторяться, называется (n, r)-*сочетанием с повторениями*. Если элементы неупорядоченной (n, r)-выборки попарно различны, то она называется (n, r)-*сочетанием без повторений* или просто (n, r)-*сочетанием*. Заметим, что любое (n, r)-сочетание можно рассматривать как r-элементное подмножество n-элементного множества.

Пример 3.1. Пусть $X=\{1, 2, 3\}$. Тогда: 1) $\langle 1,1\rangle$, $\langle 1,2\rangle$, $\langle 1,3\rangle$, $\langle 2,1\rangle$, $\langle 2,2\rangle$, $\langle 2,3\rangle$, $\langle 3,1\rangle$, $\langle 3,2\rangle$, $\langle 3,3\rangle$ — (3,2)-размещения с повторениями; 2) $\langle 1,2\rangle$, $\langle 1,3\rangle$, $\langle 2,1\rangle$, $\langle 2,3\rangle$, $\langle 3,1\rangle$, $\langle 3,2\rangle$ — (3,2)-размещения; 3) $\{1,1\}$, $\{1,2\}$, $\{1,3\}$, $\{2,2\}$, $\{2,3\}$, $\{3,3\}$ — (3,2)-сочетания с повторениями; 4) $\{1,2\}$, $\{1,3\}$, $\{2,3\}$ — (3,2)-сочетания.

Число (n, r)-размещений с повторениями обозначаем через $\overline{A}^r_n$, а без повторений — через A^r_n. Число перестановок n-элементного множества обозначаем через P_n (т. е. $P_n=A^n_n$). Число (n, r)-сочетаний с повторениями обозначаем через $\overline{C}^r_n$, а без повторений — через C^r_n.

Утверждение 3.1. $\overline{A}^r_n=n^r$.

Действительно, каждое (n, r)-размещение с повторениями является упорядоченной последовательностью длины r, причем каждый член этой последовательности может быть выбран любым из n способов, откуда по обобщенному правилу произведения и получаем требуемую формулу.

В дальнейшем для большей общности формул будем считать, что $0!=1$.

Утверждение 3.2. $A^r_n=n(n-1)\ldots(n-r+1)=\dfrac{n!}{(n-r)!}$ *при* $r\leqslant n$ *и* $A^r_n=0$ *при* $r>n$.

Рассмотрим случай, когда $r\leqslant n$ (случай $r>n$ очевиден). Каждое (n, r)-размещение без повторений является упорядоченной последовательностью длины r, члены которой попарно различны и выбираются из множества с n элементами. Тогда первый член этой последовательности может быть выбран n способами, после каждого выбора первого члена последовательности второй член может быть выбран $n-1$ способами и т. д. Соответственно после каждого выбора первого и т. д. $(r-1)$-го членов последовательности r-й член может быть выбран

$n-(r-1)=n-r+1$ способами, откуда по обобщенному правилу произведения и получаем требуемую формулу.

Следствие. $P_n=A^n_n=n(n-1)\ldots1=n!$

Утверждение 3.3. $C^r_n=\dfrac{A^r_n}{r!}=\dfrac{n!}{(n-r)!r!}$ при $r\leqslant n$, и $C^r_n=0$ при $r>n$.

Рассмотрим нетривиальный случай, когда $r\leqslant n$. Каждое (n, r)-сочетание можно упорядочить $r!$ способами. Объединение получаемых таким образом попарно непересекающихся множеств (n, r)-размещений для всех возможных (n, r)-сочетаний, очевидно, даст все (n, r)-размещения. Тогда по правилу суммы имеем $A^r_n=\Sigma r!=C^r_n r!$ (здесь суммирование производится по всем (n, r)-сочетаниям без повторений), откуда $C^r_n=\dfrac{A^r_n}{r!}$.

Утверждение 3.4. $\overline{C}^r_n=C^r_{n+r-1}$.

Каждому (n, r)-сочетанию B с повторениями, составленному из элементов множества $X=\{x_1, \ldots, x_n\}$, поставим в соответствие вектор $\alpha(B)$ длины $n+r-1$ из r нулей и $n-1$ единиц такой, что число нулей, находящихся между $(i-1)$-й и i-й единицами, где $2\leqslant i\leqslant n-1$, будет равно числу элементов x_i, входящих в сочетание B, а число нулей, стоящих перед первой единицей (после $(n-1)$-й единицы), равно числу элементов x_1 (соответственно x_n), входящих в сочетание B. Это соответствие между (n, r)-сочетаниями с повторениями и векторами с $n-1$ единицами и r нулями взаимно однозначно (см. ниже пример 3.2). С другой стороны, число векторов с $n-1$ единицами и r нулями равно числу r-элементных множеств (номеров нулевых компонент в векторах), являющихся подмножествами $(n+r-1)$-элементного множества $\{1, 2, \ldots, n+r-1\}$ (множества всех номеров компонент в векторах), т. е. числу $(n+r-1, r)$-сочетаний без повторений. Таким образом, $\overline{C}^r_n=C^r_{n+r-1}$.

Пример 3.2. Пусть $n=4$, $r=6$, $X=\{1, 2, 3, 4\}$, $B=\{2, 2, 3, 3, 3, 4\}$ — (4,6)-сочетание с повторениями. Тогда $\alpha(B)=100100010$. С другой стороны, если, например, $\alpha(B)=110010000$, то однозначно получаем, что $B=\{3, 3, 4, 4, 4, 4\}$.

Замечание 3.1. При определении выборки предполагалось, что она содержит по крайней мере один элемент. Однако для общности рассуждений в число выборок часто включают и пустую выборку, не содержащую элементов. Она единственна для всех рассмотренных нами случаев, т. е. $\overline{A}^0_n=A^0_n=\overline{C}^0_n=C^0_n=1$; при этом формулы, приведенные в утверждениях 3.1—3.4, остаются справедливыми.

3.1.3. Размещения и функциональные отображения

Обозначим через Y^X множество всех отображений $f: X\to Y$.

Утверждение 3.5. *Пусть* $|X|=r$, $|Y|=n$. *Тогда* $|Y^X|=\overline{A^r_n}=n^r=|Y|^{|X|}$.

Пусть $X=\{x_1, \ldots, x_r\}$. Тогда любое отображение $f: X \to Y$ можно представить в виде упорядоченной последовательности $\langle f(x_1), \ldots, f(x_r)\rangle$ значений функции f в точках из X, где $f(x_i) \in Y$. Очевидно, что мы тем самым установили взаимно однозначное соответствие между множеством функциональных отображений и множеством упорядоченных выборок с повторениями объема r из множества Y объема n (т. е. множеством (n, r)-размещений с повторениями), откуда и следует справедливость доказываемого утверждения.

Аналогично доказывается

Утверждение 3.6. *Пусть* $|X|=r$, $|Y|=n$. *Тогда число всех инъективных отображений вида* $f: X \to Y$ *равно* A^r_n.

Следствие. Для S_n — множества всех биективных отображений n-элементного множества в себя имеем $|S_n|=A^n_n=P_n=n!$

3.1.4. Примеры применения формул

Пример 3.3. Сколькими способами можно раскрасить квадрат, разделенный на четыре части (рис. 3.1), пятью цветами: а) допуская окрашивание разных частей в один цвет; б) если различные части окрашиваются разными цветами?

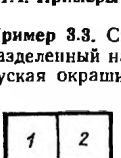


Рис. 3.1

Будем рассматривать каждое раскрашивание как функциональное отображение множества номеров частей квадрата $X=\{1, 2, 3, 4\}$ в множество цветов Y, где $|Y|=5$. Тогда, используя утверждения 3.5 и 3.6, имеем: а) $5^4=625$; б) $5!/(5-4)!=5!=120$.

Пример 3.4. Сколькими способами можно выбрать 5 номеров из 36?

Нас интересует количество неупорядоченных (36, 5)-выборок без повторений, т. е. (36, 5)-сочетаний. Используя утверждение 3.3, получаем, что требуемое число способов равно $C^5_{36}=376992$.

Пример 3.5. В скольких случаях при игре в «Спортлото» (угадывание 5 номеров из 36) будут правильно выбраны: а) ровно 3 номера; б) ровно 4 номера; в) ровно 5 номеров; г) не менее 3 номеров?

Решение: а) 3 из 5 «правильных» номеров можно выбрать C^3_5 способами, а 2 оставшихся «неправильных» номера — C^2_{31} способами. Далее по правилу произведения получаем, что искомое число равно $C^3_5 \cdot C^2_{31}=\frac{5!}{3!2!}\frac{31!}{29!2!}=4650$. В остальных случаях соответственно имеем: б) $C^4_5 \cdot C^1_{31}=155$; в) 1; г) $4650+155+1=4806$.

Пример 3.6. В скольких случаях при выборе из колоды в 52 карты 10 карт среди них окажутся все 4 туза?

Исключив из рассмотрения тузы, получим, что выбираются 6 карт из 48, а такой выбор можно осуществить C^6_{48} способами.

Пример 3.7. Имеется 30 монет достоинством 1, 2 и 3 копейки. Сколько существует различных комбинаций монет (например, 3 монеты по 1 копейке, 17 — по 2 копейки, 10 — по 3 копейки)?

По условиям задачи требуется определить количество неупорядоченных выборок с повторениями объема 30 из множества объема 3, т. е. число (3, 30)-сочетаний с повторениями. Используя утверждение 3.4, получаем, что искомое число равно $\overline{C}^{30}_3 = C^{30}_{3+30-1} = C^{30}_{32} = 496$.

Пример 3.8. Выведем формулу для количества целочисленных решений системы

$$x_1+x_2+...+x_n=r,\ x_i \geqslant a_i,\ i=1, 2, ..., n, \tag{3.1}$$

где $n \geqslant 1$; a_i — целые числа.

Прежде всего заметим, что при любых целых $n \geqslant 1$, $r \geqslant 0$ величина $\overline{C}^r_n$ выражает количество решений в целых неотрицательных числах уравнения $x_1+x_2+...+x_n=r$. Действительно, каждому решению $\langle x_1, x_2, ..., x_n \rangle$ этого уравнения можно поставить в соответствие неупорядоченную выборку с повторениями объема r из множества $A=\{a_1, a_2, ..., a_n\}$ такую, что в ней содержатся x_1 элементов вида a_1, x_2 элементов вида a_2 и т. д., x_n элементов вида a_n. Нетрудно видеть, что указанное соответствие является взаимно однозначным, откуда и следует справедливость доказываемого утверждения. Теперь для определения количества целочисленных решений системы (3.1) сделаем замену переменных. Пусть $u_i=x_i-a_i$. Тогда $x_i=u_i+a_i$, а следовательно, из (3.1) получаем

$$u_1+u_2+...+u_n=r-\sum_{i=1}^{n} a_i,\ u_i \geqslant 0,\ i=1, 2, .., n, \tag{3.2}$$

т. е. мы свели исходную задачу к уже рассмотренному случаю (понятно, что количество целочисленных решений систем (3.1) и (3.2) совпадает). Но тогда, используя доказанное ранее утверждение, получаем, что количество целочисленных решений системы (3.1) (или (3.2)) в случае $r \geqslant \sum_{i=1}^{n} a_i$ равно

$$\overline{C}^{\,r-\sum_{i=1}^{n} a_i}_n = C^{\,r-\sum_{i=1}^{n} a_i}_{n+r-\sum_{i=1}^{n} a_i-1} \tag{3.3}$$

Если же $r < \sum_{i=1}^{n} a_i$, то множество целочисленных решений системы (3.1) является пустым, и количество решений в этом случае равно 0.

3.1.5. Разбиения

Подсчитаем число разбиений конечного множества X, где $|X|=n$, на k подмножеств $X_1, X_2, \ldots, X_k$ ($k \geqslant 1$) таких, что каждое X_i содержит n_i элементов, т. е.

$$\bigcup_{i=1}^{k} X_i = X,\ X_i \cap X_j = \varnothing \text{ при } i \neq j,\ |X_i| = n_i,\ i = 1, 2, \ldots, k. \quad (3.4)$$

Очевидно, что при этом $\sum_{i=1}^{k} n_i = n$. Отметим, что для некоторых номеров i возможно $X_i = \varnothing$. Число указанных разбиений при фиксированных n_i обозначим через $C_n^{n_1, \ldots, n_k}$.

Замечание 3.2. В данном случае набор подмножеств множества X в разбиении является упорядоченным (т. е. $X_1, \ldots, X_k$ — упорядоченная последовательность множеств). Ниже, кроме того, рассматривается случай, когда набор подмножеств в разбиении не является упорядоченным.

Утверждение 3.7. $C_n^{n_1, \ldots, n_k} = \dfrac{n!}{n_1! \ldots n_k!}$.

Как отмечалось ранее, каждое из множеств X_i можно рассматривать как сочетание без повторений. Предварительно докажем справедливость формулы

$$C_n^{n_1, \ldots, n_k} = C_n^{n_1} C_{n-n_1}^{n_2} \ldots C_{n-n_1-\ldots-n_{k-1}}^{n_k}. \quad (3.5)$$

Действительно, для образования сочетания, соответствующего множеству X_1, могут быть использованы все элементы множества X, т. е. множество X_1 может быть выбрано $C_n^{n_1}$ способами. После выбора X_1 множество X_2 может быть выбрано $C_{n-n_1}^{n_2}$ способами (так как X_2 является подмножеством множества $X \setminus X_1$ и $|X \setminus X_1| = n - n_1$), и для любого i, где $2 \leqslant i \leqslant k$, после выбора множеств $X_1, \ldots, X_{i-1}$ множество X_i может быть выбрано $C_{n-n_1-\ldots-n_{i-1}}^{n_i}$ способами. Но тогда по правилу произведения выбор упорядоченной последовательности множеств $X_1, \ldots, X_k$, удовлетворяющих (3.4), можно осуществить $C^{n_1}{}_n C_{n-n_1}^{n_2} \ldots C_{n-n_1-\ldots-n_{k-1}}^{n_k}$ способами, т. е. формула (3.5) доказана. Используя теперь формулу (3.5), а также утверждение 3.3 и производя необходимые сокращения, получаем, что доказываемое утверждение справедливо.

Утверждение 3.8. *Число* $C_n^{n_1, \ldots, n_k}$, где $n_i \geqslant 0$, $\sum_{i=1}^{k} n_i = n$, *равно числу* (k, n)*-размещений с повторениями, среди элементов которых содержится* n_1 *элементов 1-го типа,* n_2 *элементов 2-го типа и т. д.,* n_k *элементов* k*-го типа.*

Каждому размещению указанного типа поставим в соответствие разбиение множества $X=\{1, 2, \ldots, n\}$ номеров элементов в выборке на подмножества $X_1, \ldots, X_k$, где X_i — множество номеров элементов i-го типа в выборке. Очевидно, что при этом выполняется (3.4). Указанное соответствие между размещениями заданного типа и разбиениями, удовлетворяющими (3.4), является взаимно однозначным, откуда в силу утверждения 3.7 и вытекает справедливость доказываемого утверждения.

Пример 3.9. В студенческой группе, состоящей из 25 человек, при выборе профорга за выдвинутую кандидатуру проголосовали 12 человек, против — 10, воздержались — 3. Сколькими способами могло быть проведено такое голосование?

Пусть X — множество студентов в группе, X_1 — множество студентов, проголосовавших за выдвинутую кандидатуру, X_2 — множество студентов, проголосовавших против, X_3 — множество студентов, воздержавшихся от голосования. Тогда $|X|=25$, $|X_1|=12$, $|X_2|=10$, $|X_3|=3$, $X=X_1 \cup X_2 \cup X_3$, $X_i \cap X_j = \varnothing$ при $i \neq j$, а следовательно, искомое число равно $C_{25}^{12, 10, 3}$. Используя утверждение 3.7, получаем

$$C_{25}^{12, 10, 3} = \frac{25!}{12!10!3!} = 1\,487\,285\,800.$$

Пример 3.10. Сколькими способами можно раскрасить квадрат, разделенный на девять частей (рис. 3.2), четырьмя цветами таким образом, чтобы в первый цвет были окрашены 3 части, во второй — 2, в третий — 3, в четвертый — 1?

1	2	3
6	5	4
7	8	9

Рис. 3.2

Пусть X — множество цветов, где $|X|=4$. Тогда каждое раскрашивание, рассматриваемое как последовательность цветов, в которые окрашиваются пронумерованные части квадрата, является упорядоченной выборкой с повторениями объема 9 из множества X, т. е. $(4,9)$-размещением с повторениями. При этом нас интересуют размещения с заданной комбинацией элементов (3 элемента — первый цвет, 2 — второй, 3 — третий, 1 — четвертый). Но тогда, используя утверждение 3.8, получаем, что искомое число равно $C_9^{3, 2, 3, 1} = \frac{9!}{3!2!3!1!} = 5040$.

Подсчитаем теперь, сколькими способами можно разбить множество

X, где $|X|=n$, на подмножества, среди которых для каждого $i=1, 2, ..., n$ имеется $m_i \geqslant 0$ подмножеств с i элементами, где $\sum_{i=1}^{n} i m_i = n$. При этом в отличие от рассмотренного ранее случая в нашем случае набор подмножеств в разбиении не является упорядоченным (т. е. порядок подмножеств в разбиении не является существенным). Так, например, разбиения множества $X = \{1, 2, 3, 4, 5\}$ вида

$\{1,3\}, \{4\}, \{2,5\};$
$\{4\}, \{2,5\}, \{1,3\};$
$\{2,5\}, \{4\}, \{1,3\}$

считаются одинаковыми. Обозначим число указанных неупорядоченных разбиений множества X через $N(m_1, ..., m_n)$.

Утверждение 3.9. $N(m_1, ..., m_n) = \dfrac{n!}{m_1! \ldots m_n! (1!)^{m_1} \ldots (n!)^{m_n}}$.

Заметим, что каждое из неупорядоченных разбиений, рассмотренных при определении величины $N(m_1, ..., m_n)$, можно, нумеруя множества в этом разбиении, привести $m_1! \ldots m_n!$ способами к упорядоченным разбиениям вида

$$X_1, ..., X_{m_1}, X_{m_1+1}, ..., X_{m_1+m_2}, ..., X_{m_1+\ldots+m_{n-1}+1}, ..., X_{m_1+\ldots+m_n}, \tag{3.6}$$

где

$$|X_1| = \ldots = |X_{m_1}| = 1, \ |X_{m_1+1}| = \ldots = |X_{m_1+m_2}| = 2, \ldots,$$
$$\ldots, \ |X_{m_1+\ldots+m_{n-1}+1}| = \ldots = |X_{m_1+\ldots+m_n}| = n. \tag{3.7}$$

При этом объединение получаемых таким образом попарно непересекающихся множеств разбиений вида (3.6), (3.7) для всех рассматриваемых неупорядоченных разбиений, очевидно, даст совокупность всех разбиений вида (3.6), (3.7), а следовательно, по правилу суммы, используя утверждение 3.7, имеем

$$\frac{n!}{(1!)^{m_1} \ldots (n!)^{m_n}} = \Sigma m_1! \ldots m_n! = N(m_1, ..., m_n) m_1! \ldots m_n!$$

(где суммирование производится по всем рассматриваемым неупорядоченным разбиениям), откуда и следует справедливость доказываемого утверждения.

Пример 3.11. Сколькими способами из группы в 25 человек можно сформировать 5 коалиций по 5 человек?

Пусть X — множество людей в группе, m_i — число коалиций по i человек, где $i=1, 2, ..., 25$. Тогда по условиям задачи $|X| = 25$, $m_5=5$, $m_i=0$, $i \in \{1, 2, ..., 25\} \setminus \{5\}$, а следовательно, искомое число будет равно $N(0, 0, 0, 0, 5, 0, ..., 0)$, где в силу утверждения 3.9 $N(0, 0, 0, 0, 5, 0, ..., 0) = \dfrac{25!}{5!(5!)^5} = \dfrac{25!}{(5!)^6}$.

Пример 3.12. Сколькими способами можно задать отношение эквивалентности на множестве $X=\{1, 2, ..., 25\}$ с тремя классами эквивалентности?

Используя тот факт, что множество классов эквивалентности является разбиением множества X, получаем, что искомое число выражается формулой

$$N(m_1, \ldots, m_{25}) = \sum_{\substack{m_1 + 2m_2 + \ldots + 25m_{25} = 25 \\ m_1 + m_2 + \ldots + m_{25} = 3}} - \sum_{\substack{m_1 + 2m_2 + \ldots + 25m_{25} = 25 \\ m_1 + m_2 + \ldots + m_{25} = 3}} \frac{25!}{m_1! \ldots m_{25}! (11)^{m_1} \ldots (25!)^{m_{25}}},$$

где под «$m_1 + 2m_2 + \ldots + 25m_{25} = 25$, $m_1 + m_2 + \ldots + m_{25} = 3$» понимается множество всех решений этой системы уравнений в целых неотрицательных числах (например, $m_1 = 1$, $m_{12} = 2$, $m_i = 0$, $i \neq 1$, $i \neq 12$ — одно из таких решений).

3.1.6. Полиномиальная формула

Определим коэффициенты $c_{n_1, \ldots, n_k}$ в формуле

$$(x_1 + \ldots + x_k)^n = \sum_{n_1 + \ldots + n_k = n} c_{n_1, \ldots, n_k} x_1^{n_1} \ldots x_k^{n_k},$$ называемой

полиномиальной, где суммирование производится по всем решениям уравнения $n_1 + \ldots + n_k = n$ в целых неотрицательных числах.

Утверждение 3.10. $c_{n_1, \ldots, n_k} = C_n^{n_1, \ldots, n_k} = \dfrac{n!}{n_1! \ldots n_k!}$.

Введем обозначения для сомножителей в $(x_1 + \ldots + x_k)^n$. Обозначим $A_i = (x_1 + \ldots + x_k)$, $i = 1, 2, \ldots, n$. Тогда $(x_1 + \ldots + x_k)^n = A_1 \ldots A_n$. Пусть $A = \{a_1, \ldots, a_n\}$. Пересчитаем все одночлены, полученные в результате перемножения $a_1 \ldots a_n$, в которых x_1 встречается n_1 раз, x_2 — n_2 раз и т. д., x_k — n_k раз, т. е. одночлены вида

$$x_1^{n_1} \ldots x_k^{n_k}, \text{ где } n_1 + \ldots + n_k = n. \tag{3.8}$$

Рассмотрим любой из таких одночленов. Для каждого $i \in \{1, 2, \ldots, k\}$ обозначим через A_i подмножество множества A такое, что в этот одночлен войдут переменные x_i из тех и только тех сомножителей, которые перечислены в A_i. Тем самым мы поставили в соответствие одночлену разбиение множества A на подмножества $A_1, \ldots, A_k$ такие, что

$$|A_i| = n_i, \ i = 1, \ldots, k, \ \bigcup_{i=1}^{k} A_i = A, \ A_i \cap A_j = \varnothing \text{ при } i \neq j. \tag{3.9}$$

Понятно, что указанное соответствие между одночленами (3.8) и разбиениями вида (3.9) является взаимно однозначным. Но тогда $c_{n_1, \ldots, n_k}$ (количество одночленов вида (3.8)) равно

$$c_{n_1, \ldots, n_k} = C_n^{n_1, \ldots, n_k} = \frac{n!}{n_1! \ldots n_k!}.$$

Пример 3.13. Определим коэффициент c в одночлене $cx_1^3x_2^4x_3^3$ многочлена (с приведенными подобными членами), получаемого из выражения $(x_1+x_2+x_3)^{10}$.

В силу утверждения 3.10 имеем $c=C_{10}^{3,4,3}=\frac{10!}{3!4!3!}=4200$.

3.1.7. Формула включений и исключений

Пусть X_1, X_2 — два конечных множества. Тогда, если $X_1\cap X_2=\varnothing$, то $|X_1\cup X_2|=|X_1|+|X_2|$. Пусть теперь $X_1\cap X_2\neq\varnothing$. Тогда в $|X_1|+|X_2|$ каждый элемент из $X_1\cap X_2$ будет учтен два раза. Поэтому

$$|X_1\cup X_2|=|X_1|+|X_2|-|X_1\cap X_2|, \tag{3.10}$$

т. е. мы выразили количество элементов в объединении множеств через количество элементов в их пересечении. Получим соответствующую формулу для произвольного числа множеств, которая называется *формулой включений и исключений*.

Утверждение 3.11. *Пусть X_i — конечные множества, $i=1, 2, ..., n$, $n\geqslant 2$. Тогда*

$$|X_1\cup...\cup X_n|=(|X_1|+...+|X_n|)-(|X_1\cap X_2|+|X_1\cap X_3|+...+|X_{n-1}\cap X_n|)+(|X_1\cap X_2\cap X_3|+...+|X_{n-2}\cap X_{n-1}\cap X_n|)-...+(-1)^{n+1}|X_1\cap...\cap X_n|. \tag{3.11}$$

Доказательство будем проводить индукцией по n. При $n=2$ формула (3.11) совпадает с (3.10). Предположим, что доказываемая формула верна для случая $n-1$ подмножеств, где $n\geqslant 3$. Докажем ее справедливость для n подмножеств. Разобьем множества X_1, ..., X_n на две группы: X_1, ..., X_{n-1}; X_n. Тогда согласно (3.10) получаем

$$\begin{aligned}|X_1\cup...\cup X_n|&=|(X_1\cup...\cup X_{n-1})\cup X_n|=\\&=|X_1\cup...\cup X_{n-1}|+|X_n|-|(X_1\cup...\cup X_{n-1})\cap X_n|=\\&=|X_1\cup...\cup X_{n-1}|+|X_n|-|A_1\cup...\cup A_{n-1}|,\end{aligned} \tag{3.12}$$

где $A_i=X_i\cap X_n$, $i=1, ..., n-1$.

Используя индуктивное предположение, имеем:

а) $|X_1\cup...\cup X_{n-1}|=(|X_1|+...+|X_{n-1}|)-(|X_1\cap X_2|+...+|X_{n-2}\cap X_{n-1}|)+(|X_1\cap X_2\cap X_3|+...+|X_{n-3}\cap X_{n-2}\cap X_{n-1}|)-...+(-1)^n|X_1\cap...\cap X_{n-1}|$;

б) $|A_1\cup...\cup A_{n-1}|=(|A_1|+...+|A_{n-1}|)-(|A_1\cap A_2|+|A_1\cap A_3|+...+|A_{n-2}\cap A_{n-1}|)+...+(-1)^n|A_1\cap...\cap A_{n-1}|=$
$=(|X_1\cap X_n|+...+|X_{n-1}\cap X_n|)-(|X_1\cap X_2\cap X_n|+|X_1\cap X_3\cap X_n|+...+|X_{n-2}\cap X_{n-1}\cap X_n|)+...+(-1)^n|X_1\cap...\cap X_n|$.

Из (3.12), учитывая «а», «б», получаем (3.11).

Следствие. Пусть X — конечное множество, X_1, ..., X_n — подмножества X. Тогда

$$|X\setminus(X_1\cup\ldots\cup X_n)| = |X| - (|X_1|+\ldots+|X_n|) + (|X_1\cap X_2|+\ldots + |X_{n-1}\cap X_n|) - \ldots + (-1)^n |X_1\cap\ldots\cap X_n|. \tag{3.13}$$

Действительно,

$$[X\setminus(X_1\cup\ldots\cup X_n)]\cup(X_1\cup\ldots\cup X_n) = X,$$
$$[X\setminus(X_1\cup\ldots\cup X_n)]\cap(X_1\cup\ldots\cup X_n) = \varnothing,$$

откуда

$$|X\setminus(X_1\cup\ldots\cup X_n)| + |X_1\cup\ldots\cup X_n| = |X|,$$

а следовательно,

$$|X\setminus(X_1\cup\ldots\cup X_n)| = |X| - |X_1\cup\ldots\cup X_n|. \tag{3.14}$$

Для получения (3.13) остается только в (3.14) применить формулу (3.11).

Приведем теперь еще одну (наиболее распространенную) форму записи формулы включений и исключений. Пусть X — конечное множество, состоящее из N элементов, α_1, ..., α_n — некоторые свойства (одноместные предикаты, определенные на X), которыми могут обладать или не обладать элементы из X. Обозначим $\forall i \in \{1, 2, \ldots, n\}$ $X_i = \{x \in X \mid \alpha_i(x)\}$ — множество элементов в X, обладающих свойством α_i. Обозначим также $N(\alpha_{i_1}, \ldots, \alpha_{i_k}) = |X_{i_1}\cap\ldots\cap X_{i_k}| = |\{x\in X \mid \alpha_{i_1}(x)\,\&\ldots\&\,\alpha_{i_k}(x)\}|$ — количество элементов в X, обладающих одновременно свойствами α_{i_1}, ..., α_{i_k}; $N_0 = |X\setminus(X_1\cup\ldots\cup X_n)|$ — количество элементов в X, не обладающих ни одним из свойств α_1, ..., α_n. Тогда по формуле (3.13) получим

$$N_0 = N - S_1 + S_2 - \ldots + (-1)^n S_n, \tag{3.15}$$

где

$$S_k = \sum_{1\leqslant i_1<\ldots<i_k\leqslant n} N(\alpha_{i_1}, \ldots, \alpha_{i_k}),\quad k = 1, 2, \ldots, n.$$

Пример 3.14. Пусть $n = 3$, т. е. имеются три свойства: α_1, α_2, α_3. Тогда из (3.15) получаем

$$N_0 = N - N(\alpha_1) - N(\alpha_2) - N(\alpha_3) + N(\alpha_1, \alpha_2) + N(\alpha_1, \alpha_3) + N(\alpha_2, \alpha_3) - N(\alpha_1, \alpha_2, \alpha_3). \tag{3.16}$$

Пример 3.15. Пусть $X = \{0, 1, \ldots, 10\}$; $\alpha_1(x)$: «x четное»; $\alpha_2(x)$: «$x > 6$»; $\alpha_3(x)$: «$2 < x < 8$». Подсчитаем количество N_0 элементов в X, не обладающих свойствами α_1, α_2, α_3. Используя формулу (3.16), получаем $N_0 = 11 - 6 - 4 - 5 + 2 + 2 + 1 - 0 = 1$ (нетрудно видеть, что единственным элементом в X, не обладающим свойствами α_1, α_2, α_3, является число 1).

Пример 3.16. Применяя формулу включений и исключений, задачу определения количества целочисленных решений системы

$$x_1+x_2+\ldots+x_n=r,\ a_i\leqslant x_i\leqslant b_i,\ i=1,\ 2,\ \ldots,\ n, \qquad (3.17)$$

где a_i, b_i, r — целые числа; $a_i\leqslant b_i$, $i=1, 2, \ldots, n$, легко свести к совокупности задач определения количества целочисленных решений систем вида (3.1). Для этого необходимо воспользоваться свойствами $\alpha_i(x)$: «$x_i\geqslant b_i+1$», где $x=\langle x_1, \ldots, x_n\rangle$, $i=1, 2, \ldots, n$, а за исходное множество X взять совокупность целочисленных решений x системы (3.1). Пусть N — количество целочисленных решений системы (3.1). Тогда N_0, определяемое по формуле (3.15), и будет выражать количество целочисленных решений системы (3.17).

Пример 3.17. Определим количество трехзначных чисел, в которых сумма цифр равняется 20.

Если через x_1, x_2, x_3 обозначить соответственно первую, вторую и третью цифры в произвольном трехзначном числе $a=x_1\cdot 10^2+x_2\cdot 10+x_3$, то для решения задачи достаточно определить количество целочисленных решений системы

$$x_1+x_2+x_3=20,\ 1\leqslant x_1\leqslant 9,\ 0\leqslant x_2\leqslant 9,\ 0\leqslant x_3\leqslant 9. \qquad (3.18)$$

Пусть X — множество целочисленных решений $x=\langle x_1, x_2, x_3\rangle$ системы

$$x_1+x_2+x_3=20,\ x_1\geqslant 1,\ x_2\geqslant 0,\ x_3\geqslant 0,$$

N — количество элементов в X. Введем следующие три свойства:

$\alpha_1(x)$: «$x_1\geqslant 10$»; $\alpha_2(x)$: «$x_2\geqslant 10$»; $\alpha_3(x)$: «$x_3\geqslant 10$».

Используя теперь формулу включений и исключений (3.16), а также формулу (3.3) для подсчета количества целочисленных решений в системе вида (3.1), определим число N_0 целочисленных решений системы (3.18):

$$N_0=\overline{C}_3^{20-1}-\overline{C}_3^{20-10}-\overline{C}_3^{20-10-1}-\overline{C}_3^{20-10-1}+\overline{C}_3^{20-10-10}+\overline{C}_3^{20-10-10}+0-0=C_{21}^{19}-C_{12}^{10}-2C_{11}^{9}+2=$$
$$=210-66-110+2=36.$$

Пример 3.18 (задача о беспорядках). Имеется n различных предметов $a_1, a_2, \ldots, a_n$ и n различных ячеек $b_1, b_2, \ldots, b_n$. Сколькими способами можно разместить предметы по ячейкам так, чтобы никакой предмет a_i не попал в ячейку b_i?

В качестве исходного множества X возьмем совокупность всевозможных расположений предметов по ячейкам. Тогда $N=|X|=n!$ Введем свойства α_i: «a_i находится в ячейке b_i», $i=1, 2, \ldots, n$. Число $N(\alpha_{i_1}, \ldots, \alpha_{i_k})$ расположений, при которых предмет a_{i_ν} находится в ячейке b_{i_ν}, $\nu=1, \ldots, k$, равно $(n-k)!$ Но тогда

$$S_k = \sum_{1 \leqslant i_1 < \ldots < i_k \leqslant n} N(a_{i_1}, \ldots, a_{i_k}) = C^k_n (n-k)! = \frac{n!}{k!}.$$

Используя теперь формулу включений и исключений (3.15), получаем, что число N_0 расположений, при которых ни одно из свойств не выполняется (т. е. ни один из предметов a_i не попал в ячейку b_i), равно

$$N + \sum_{k=1}^{n} (-1)^k S_k = n! + \sum_{k=1}^{n} (-1)^k \frac{n!}{k!} = n! \sum_{k=0}^{n} (-1)^k \frac{1}{k!}.$$

Задачи и упражнения

1. Сколькими способами можно дать клички четырем щенкам, имея семь возможных вариантов (щенки названы по-разному)?

2. Сколькими способами можно раскрасить квадрат, разделенный на девять частей (см. рис. 3.2), пятью цветами, допуская окрашивание разных частей в один цвет?

3. Сколькими способами можно разместить 10 различных шаров по трем различным урнам?

4. Обосновать (без использования формул) $C^k_n = C^k_{n-1} + C^{k-1}_{n-1}$.

5. Определить количество прямоугольных матриц размерности $m \times n$ с элементами из $\{0,1\}$ с попарно различными строками ($n \leqslant 2^n$).

6. Из колоды, состоящей из 52 карт, выбрали 10 карт. Определить, в скольких случаях среди них окажутся: а) пиковая дама; б) все четыре дамы; в) все карты одной масти; г) ни одного туза; д) ровно один туз; е) хотя бы один туз; ж) ровно два туза.

7. Имеются монеты по 1, 2 и 3 копейки; всего 20 монет. Сколько существует различных комбинаций монет?

8. Сколькими способами можно разложить 20 одинаковых шаров по четырем различным урнам?

9. Определить количество целочисленных решений системы $x_1 + x_2 + x_3 = 40$, $x_1 \geqslant 3$, $x_2 \geqslant 0$, $x_3 \geqslant 2$.

10. Сколькими способами можно разложить 20 различных шаров по трем различным урнам так, чтобы в первой, второй и третьей урнах находилось соответственно 5, 3 и 12 шаров?

11. Сколькими способами группу из 25 человек можно разбить на семь коалиций: 2 — по 5 человек, 1 — 7 человек, 4 — по 2 человека?

12. Определить коэффициент k в следующих членах многочлена (с приведенными подобными членами), получаемого из алгебраического выражения $(a+b+c)^2(a^2+b^2+c^2)^4$:
а) $ka^3b^3c^4$; б) $ka^2b^4c^4$; в) ka^5b^5; г) $ka^2b^2c^6$.

13. Определить количество целочисленных решений системы $x_1 + x_2 + x_3 = 40$, $4 \leqslant x_1 \leqslant 15$, $9 \leqslant x_2 \leqslant 18$, $5 \leqslant x_3 \leqslant 16$.

14. Определить количество шестизначных чисел, в которых сумма первых трех цифр совпадает с суммой последних трех цифр.

15 Найти количество целых положительных чисел, не превосходящих 200 и не делящихся ни на одно из простых чисел: а) 2, 3, 5; б) 7, 11, 13.

3.2. РЕШЕНИЕ ЗАДАЧ ПЕРЕСЧЕТА МЕТОДОМ ПОЙА

3.2.1. Реализация группы

Пусть X — некоторое конечное множество элементов, $S(X)$ — группа всех биективных отображений множества X в себя, G — некоторая группа. Под *реализацией* G в $S(X)$ будем понимать любой гомоморфизм $\tau: G \to S(X)$. Для простоты обозначений вместо $\tau(g)$, где $g \in G$, пишем τ_g, а вместо $\tau_g(x)$, где $x \in X$, — gx.

Пример 3.19. Пусть $X = \mathbf{R}^2$ — множество точек на плоскости, $S(\mathbf{R}^2)$ — группа биективных отображений $\mathbf{R}^2$ в себя, G — группа вращений на плоскости вокруг начальной точки $O(0,0)$. Элементы группы G будем обозначать углами вращений, при этом в качестве положительного выберем направление против часовой стрелки. Пусть, например, $G = \{\alpha \in \mathbf{R} \mid 0 \leqslant \alpha < 2\pi\}$. Бинарной операцией на G является последовательное выполнение вращений. Обозначим указанную операцию символом $\oplus$. Тогда, очевидно, выполняется равенство

$$\alpha \oplus \beta = \begin{cases} \alpha+\beta, \text{ если } \alpha+\beta < 2\pi; \\ \alpha+\beta-2\pi, \text{ если } \alpha+\beta \geqslant 2\pi. \end{cases}$$

Единичным элементом в группе G является вращение на угол 0 и $\forall \alpha \in G \setminus \{0\}\ \alpha^{-1} = 2\pi - \alpha$. Под реализацией введенной группы G будем понимать отображение, ставящее в соответствие каждому вращению α биекцию $\tau_\alpha: \mathbf{R}^2 \to \mathbf{R}^2$ такую, что для любой точки $M \in \mathbf{R}^2$ точка $\tau_\alpha(M)$ есть результат перемещения точки M после поворота вектора $\overrightarrow{OM}$ на угол α против часовой стрелки. Тогда, например, $\pi/2(1, 0) = \tau_{\pi/2}[(1, 0)] = (0, 1)$ (рис. 3.3).

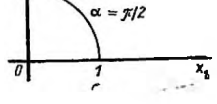


Рис. 3.3

3.2.2. Действие группы на множестве

Пусть сохраняются условия, описанные в разд. 3.2.1, и τ — некоторая реализация G в $S(X)$. Заметим, что поскольку при гомоморфизме единичный элемент отображается на единичный элемент, то

$$\forall x \in X \quad ex = x, \tag{3.19}$$

где e — единичный элемент из G. Кроме того, по определению гомоморфизма

$$\forall x \in X, \ \forall g, \ h \in G \quad \tau_{gh}(x) = \tau_g[\tau_h(x)] = g(hx),$$

откуда

$$\forall x \in X, \ \forall g, \ h \in G \quad (gh)x = g(hx). \tag{3.20}$$

Всякий раз, когда имеется отображение $\langle g, x\rangle \to gx$ прямого произведения $G \times X$ в X, удовлетворяющее свойствам (3.19), (3.20), будем говорить, что группа G *действует* на множестве X.

В определении действия группы G на множестве X явным образом не используется гомоморфизм τ, а следовательно, действие группы на множестве можно вводить непосредственно, без предварительного указания гомоморфизма τ. Тем не менее в последнем случае можно по формуле

$$\tau_g(x) = gx, \quad x \in X, \tag{3.21}$$

для каждого $g \in G$ однозначно определить отображение $\tau_g : X \to X$, и при этом отображение $\tau : g \to \tau_g$ будет гомоморфизмом группы G в $S(X)$. Действительно, используя (3.19), (3.20), получаем, что для любого элемента $g \in G$ выполняются условия

$$\forall x \in X \quad \tau_{g^{-1}}(\tau_g(x)) = g^{-1}(gx) = (g^{-1}g)x = ex = x; \tag{3.22}$$

$$\forall x \in X \quad \tau_g(\tau_{g^{-1}}(x)) = g(g^{-1}x) = (gg^{-1})x = ex = x, \tag{3.23}$$

откуда следует, что $\forall g \in G$ τ_g — биекция из X в X (из (3.23) получаем сюръективность τ_g, а из (3.22) — инъективность: $x_1, x_2 \in X$, $\tau_g(x_1) = \tau_g(x_2) \Rightarrow x_1 = \tau_{g^{-1}}(\tau_g(x_1)) = \tau_{g^{-1}}(\tau_g(x_2)) = x_2$) и, кроме того, в силу (3.20) $\forall g, h \in G$ имеем

$$\forall x \in X \quad (\tau_g\tau_h)(x) = \tau_g(\tau_h(x)) = g(hx) = (gh)x = \tau_{gh}(x),$$

а следовательно,

$$\forall g, h \in G \quad \tau_{gh} = \tau_g\tau_h,$$

т. е. мы показали, что τ — гомоморфизм.

Таким образом, в тех случаях, когда это удобно, будем определять действие группы на множестве заданием гомоморфизма τ, а в других случаях — непосредственным заданием отображения $\langle g, x\rangle \to gx$, удовлетворяющего условиям (3.19), (3.20).

Заметим далее, что из (3.19), (3.20) следует

$$\forall g \in G,\ \forall x_1,\ x_2 \in X\quad x_2 = gx_1 \Rightarrow x_1 = g^{-1}x_2 \tag{3.24}$$

(так как, используя $x_2 = gx_1$, получаем $g^{-1}x_2 = g^{-1}(gx_1) = (g^{-1}g)x_1 = ex_1 = x_1$).

3.2.3. Орбиты. Лемма Бернсайда о числе орбит

Две точки $x_1,\ x_2 \in X$ называются *эквивалентными относительно группы* G, действующей на множестве X (или G-*эквивалентными*), если $\exists g \in G \mid x_2 = gx_1$.

Из (3.19), (3.20), (3.24) непосредственно получаем, что G-эквивалентность является отношением эквивалентности на X ((3.19) показывает рефлексивность этого отношения, (3.24) — симметричность, а из (3.20) следуст транзитивность: $x_2 = g_1x_1$, $x_3 = g_2x_2 \Rightarrow x_3 = g_2(g_1x_1) = (g_2g_1)x_1$), и, следовательно, указанное отношение разбивает X на классы эквивалентности, которые называют G-*орбитами*. Орбиту, содержащую элемент $x_0 \in X$, будем обозначать через $G(x_0)$, т. е. $G(x_0) = \{gx_0 \mid g \in G\}$.

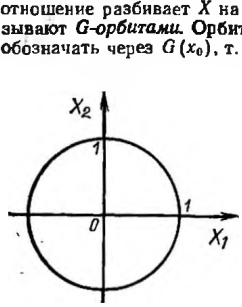


Рис. 3.4

Пример 3.20. Орбитой точки (0,1) в примере 3.19 является окружность с центром в точке O радиуса 1, проходящая через точку (0, 1) (рис. 3.4).

Пусть теперь X — конечное множество элементов, G — конечная группа, действующая на X. Рассмотрим задачу об определении числа N G-орбит. Обозначим через $N(g)$, где $g \in G$, число элементов из X, остающихся на месте при действии g, т. е. $N(g) = |\{x \in X \mid gx = x\}|$.

Лемма 3.1. (*лемма Бернсайда*). $N = \dfrac{1}{|G|} \sum\limits_{g \in G} N(g)$.

Для доказательства леммы потребуются вспомогательные утверждения.

Утверждение 3.12. *Пусть* $x_0 \in X$, $St(x_0) = \{g \in G \mid gx_0 = x_0\}$. *Тогда* $St(x_0)$ *— подгруппа группы* G *(стационарная подгруппа точки* x_0*).*

В силу $ex_0 = x_0$ (см. (3.19)) имеем $e \in St(x_0)$. Пусть $g_1,\ g_2 \in St(x_0)$. Тогда, используя (3.20), получаем $(g_1g_2)x_0 = g_1(g_2x_0) = g_1x_0 = x_0$, т. е. $g_1g_2 \in St(x_0)$. Пусть $g \in St(x_0)$. Тогда $gx_0 = x_0$, откуда, используя (3.24), получаем $x_0 = g^{-1}x_0$, а следовательно, $g^{-1} \in St(x_0)$.

Утверждение 3.13. *Пусть* $x_0 \in X$. *Тогда* $|G(x_0)| = |G|/|St(x_0)|$.

Рассмотрим левые смежные классы группы G по подгруппе $H = St(x_0)$. Поставим в соответствие каждой точке $x \in G(x_0)$ смежный класс $gH \in G/H$, где g — элемент из G такой, что $x = gx_0$. Покажем, что каждому элементу x_0 поставлен в соответствие единственный смежный класс и тем самым задано некоторое отображение $\varphi : G(x_0) \to G/H$. Действительно, пусть для некоторого элемента $g_1 \in G$ выполняется $x = g_1 x_0$. Тогда $g_1 x_0 = gx_0$, откуда $(g^{-1}g_1)x_0 = x_0$, а следовательно, $g^{-1}g_1 \in St(x_0) = H$. Но тогда $g_1 \in gH$, а значит, $g_1 H = gH$. Покажем теперь, что указанное отображение $\varphi : G(x_0) \to G/H$ является взаимно однозначным, откуда и будет следовать, что $|G(x_0)| = |G/H| = |G|/|H|$. Пусть $g_1, g_2 \in G$, $x_1 = g_1 x_0$, $x_2 = g_2 x_0$, $x_1 \neq x_2$. Покажем, что $g_1 H \neq g_2 H$. Действительно, если $g_1 H = g_2 H$, то $g_2^{-1} \times g_1 H = H$, откуда $g_2^{-1} g_1 \in H$. Но тогда $(g_2^{-1}g_1)x_0 = x_0$, а следовательно, $g_1 x_0 = g_2 x_0$, что противоречит условию $x_1 \neq x_2$. Таким образом, доказана инъективность отображения φ. Для доказательства его сюръективности осталось заметить, что $\forall gH \in G/H$ gH является образом точки $gx_0 \in G(x_0)$.

Утверждение 3.14. *Пусть* $x_0 \in X$. *Тогда* $\forall x \in G(x_0)$ $|St(x)| = |St(x_0)|$.

Пусть $x \in G(x_0)$. Тогда найдется $g_0 \in G : x = g_0 x_0$. Рассмотрим произвольный элемент $g \in St(x)$. Тогда $gx = x$, а следовательно, $(gg_0)x_0 = g(g_0 x_0) = gx = x = g_0 x_0$, откуда $(g_0^{-1}gg_0)x_0 = x_0$, т. е. $g_0^{-1}gg_0 \in St(x_0)$. Таким образом, $g_0^{-1} St(x) g_0 \subseteq St(x_0)$, откуда

$$g_0 St(x_0) g_0^{-1} \supseteq St(x). \tag{3.25}$$

Пусть теперь g — произвольный элемент из $St(x_0)$. Тогда $gx_0 = x_0$, а значит, в силу $x_0 = g_0^{-1}x$ имеем $(gg_0^{-1})x = g(g_0^{-1}x) = gx_0 = x_0 = g_0^{-1}x$, откуда $(g_0 g g_0^{-1})x = x$, т. е. $g_0 g g_0^{-1} \in St(x)$. Таким образом, $g_0 St(x_0) g_0^{-1} \subseteq St(x)$ и, используя (3.25), получаем $g_0 St(x_0) g_0^{-1} = St(x)$, откуда и следует, что $|St(x)| = |St(x_0)|$.

Теперь докажем **лемму Бернсайда.**

Обозначим через X_G множество G-орбит. Пусть также $\forall g \in G$, $\forall x \in X$

$$\alpha(gx = x) = \begin{cases} 1, & \text{если } gx = x; \\ 0, & \text{если } gx \neq x. \end{cases}$$

Тогда, используя утверждения 3.13 и 3.14, имеем

$$\sum_{g \in G} N(g) = \sum_{g \in G} \sum_{x \in X} \alpha(gx = x) = \sum_{x \in X} \sum_{g \in G} \alpha(gx = x) =$$

$$= \sum_{G(x_0) \in X_G} \sum_{x \in G(x_0)} \sum_{g \in G} \alpha(gx = x) = \sum_{G(x_0) \in X_G} \sum_{x \in G(x_0)} |St(x)| =$$

$$= \sum_{G(x_0)\in X_G} \sum_{x\in G(x_0)} |St(x_0)| = \sum_{G(x_0)\in X_G} |G(x_0)|\,|St(x_0)| =$$

$$= \sum_{G(x_0)\in X_G} (|G|/|St(x_0)|)\,|St(x_0)| = |G|\,|X_G| = |G|N,$$

откуда и следует справедливость доказываемой формулы.

3.2.4. Примеры применения леммы Бернсайда для решения комбинаторных задач

Пример 3.21. Составляются слова длины 3 из букв a и b. Слова считаются эквивалентными, если получаются одно из другого переменой местами крайних букв (например, $abb \sim bba$). Определим число N классов эквивалентности.

Пусть $X=\{c=c_1c_2c_3 \mid c_i\in\{a, b\},\ i=1, 2, 3\}=\{a, b\}^3$. Рассмотрим группу $G=\{e, \sigma\}$, где e — единичный элемент группы G, $\sigma\neq e$, $\sigma^2=e$. Определим отображение $<g, c>\to gc$ прямого произведения $G\times X$ в X. Пусть $\forall c\in X\ ec=c$, $\sigma c=\sigma c_1c_2c_3=c_3c_2c_1$ (т. е. e оставляет слово c на месте, а σ меняет в слове c крайние буквы местами). Покажем, что введенное отображение удовлетворяет (3.19), (3.20). Условие (3.19) выполняется по определению. Покажем выполнение условия (3.20). Рассмотрим нетривиальный случай с $g=h=\sigma$ (случаи с $g=e$ или $h=e$ очевидны). Тогда

$$\forall c\in X\ (\sigma\sigma)c=\sigma^2c=ec=c,\ \sigma(\sigma c)=\sigma c_3c_2c_1=c_1c_2c_3=c,$$

откуда и вытекает справедливость (3.20). По условиям задачи искомое число N совпадает с числом G-орбит, а значит, используя лемму Бернсайда, имеем $N=\frac{1}{2}(N(e)+N(\sigma))$. Заметим, что

$$N(e)=|X|=2^3=8;$$

$$N(\sigma)=|\{c\in X \mid \sigma c_1c_2c_3=c_1c_2c_3\}|=|\{c\in X \mid c_3c_2c_1=c_1c_2c_3\}|=|\{c\in X \mid c_1=c_3\}|=2^2=4,$$

а следовательно, $N=\frac{1}{2}(8+4)=6$.

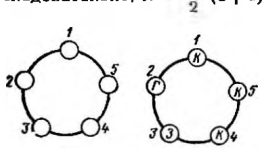


Рис. 3.5

Пример 3.22. Составляются ожерелья из плоских бусин трех цветов, при этом окрашена только одна сторона бусин. Каждое ожерелье состоит из пяти бусин. Определим число N различных ожерелий.

Пронумеруем бусины в ожерелье, начиная с некоторой бу-

сины и увеличивая номера в процессе обхода ожерелья в направлении, обратном движению часовой стрелки (рис. 3.5, *а*). Каждой бусине можно поставить в соответствие один из трех возможных цветов, например, к (красный), г (голубой), з (зеленый). Тогда любое раскрашивание ожерелья с проиумерованными бусинами можно описать упорядоченным набором цветов длины 5, предполагая, что t-й элемент этого набора соответствует цвету i-й бусины. Например, упорядоченная пятерка $c=$ $=<$к, г, з, к, к$>$ означает, что первая бусина окрашена в красный цвет, вторая — в голубой и т. д. (рис. 3.5, *б*). Рассмотрим множество X раскрашиваний ожерелья с пронумерованными бусинами, т. е.

$$X=\{c=<c_1, c_2, c_3, c_4, c_5>\,|\,c_i\in\{\text{к, г, з}\}\}=\{\text{к, г, з}\}^5.$$

Очевидно, что $|X|=|\{\text{к, г, з}\}|^5=3^5=243$. Рассмотрим также группу G вращений ожерелья на плоскости (вокруг центра этого ожерелья), совмещающих его с самим собой. При этом под композицией $\alpha\circ\beta$ двух произвольных вращений α, $\beta\in G$ будем понимать результат последовательно применяемых к ожерелью вращений β, α (т. е. сначала выполняется вращение β, а затем к полученному результату применяется вращение α). Очевидно, что группа G состоит из пяти элементов — вращений (против часовой стрелки) на углы $2\pi i/5$, где $i=0, 1, 2, 3, 4$. Обозначим элементы группы G величинами соответствующих углов.

Заметим, что рассматриваемая в этой задаче группа G является коммутативной, что делает излишним указание, в какой последовательности надо осуществлять вращения в композиции $\alpha\circ\beta$. Однако в других задачах группа вращений может оказаться некоммутативной, и тогда указание последовательности вращений при определении композиции $\alpha\circ\beta$ является необходимым (мы, естественно, сохраним выбранное определение композиции и при решении всех последующих задач).

Поставим в соответствие каждому вращению $\alpha\in G$ подстановку τ_α на множестве номеров бусин $\{1, 2, 3, 4, 5\}$ такую, что для любого $i\in\{1, 2, 3, 4, 5\}$ $\tau_\alpha(i)$ будет номером i-й бусины ожерелья, полученного в результате вращения α, относительно исходного положения ожерелья (это можно представить себе так: одно ожерелье закреплено на плоскости и с ним совмещается некоторое подвижное ожерелье, к которому применяется вращение α). Тогда $\tau_0=e$, $\tau_{2\pi/5}=(1\ 2\ 3\ 4\ 5)$, $\tau_{4\pi/5}=(1\ 3\ 5\ 2\ 4)$, $\tau_{6\pi/5}=(1\ 4\ 2\ 5\ 3)$, $\tau_{8\pi/5}=(1\ 5\ 4\ 3\ 2)$. Обозначим $T=\{e, \tau_{2\pi/5}, \tau_{4\pi/5}, \tau_{6\pi/5}, \tau_{8\pi/5}\}$. Непосредственной проверкой убеждаемся, что отображение $\alpha\to\tau_\alpha$ группы G на T является биекцией. Заметим далее, что (по определению композиции $\circ$ и отображения $\alpha\to\tau_\alpha$) выполняется

$$\forall\alpha,\ \beta\in G\quad \tau_{\alpha\circ\beta}=\tau_\alpha\,\tau_\beta. \tag{3.26}$$

откуда следует, что группа G изоморфна группе $T \subset S_5$ (для доказательства того, что T — группа, можно воспользоваться формулой (3.26), а также тем, что $\tau_0 = e$). Но тогда группы G и T для нас неразличимы и для простоты будем считать, что $G = T$.

Определим отображение $\langle g, c\rangle \to gc$ прямого произведения $G \times X$ в X. Пусть

$$\forall c \in X,\ \forall g \in G\ \ gc = g\langle c_1, c_2, c_3, c_4, c_5\rangle = \langle c_{g^{-1}(1)}, c_{g^{-1}(2)}, c_{g^{-1}(3)}, c_{g^{-1}(4)}, c_{g^{-1}(5)}\rangle,$$

т. е. $\forall c \in X,\ \forall \tau_\alpha \in G\ \ \tau_\alpha c$ — раскрашивание, получающееся в результате вращения α ожерелья с раскрашиванием c. Например, $\tau_{2\pi/5}\langle к, к, ч, г, з\rangle = \langle з, к, к, к, г\rangle$ (рис. 3,6, $а$ — раскрашивание $c = \langle к, к, к, г, з\rangle$, рис. 3.6, $б$ — раскрашивание $\tau_{2\pi/5} c$).

Покажем, что введенное отображение $\langle g, c\rangle \to gc$ прямого произведения $G \times X$ в X удовлетворяет (3.20) (выполнение (3.19) очевидно). Докажем справедливость (3.20). Пусть $\tau_\alpha, \tau_\beta \in G$. Тогда $\tau_\alpha(\tau_\beta c) = \tau_\alpha \langle c_{\tau^{-1}_\beta(1)}, \ldots, c_{\tau^{-1}_\beta(5)}\rangle$. Обозначим $c'_i = c_{\tau^{-1}_\beta(i)}$, $i = 1, 2, 3, 4, 5$. Используя введенные обозначения, получаем

$$\tau_\alpha(\tau_\beta c) = \tau_\alpha \langle c'_1, \ldots, c'_5\rangle = \langle c'_{\tau^{-1}_\alpha(1)}, \ldots, c'_{\tau^{-1}_\alpha(5)}\rangle =$$
$$= \langle c_{\tau^{-1}_\beta(\tau^{-1}_\alpha(1))}, \ldots, c_{\tau^{-1}_\beta(\tau^{-1}_\alpha(5))}\rangle = \langle c_{(\tau_\alpha \tau_\beta)^{-1}(1)}, \ldots$$
$$\ldots, c_{(\tau_\alpha \tau_\beta)^{-1}(5)}\rangle = (\tau_\alpha \tau_\beta) c,$$

т. е. условие (3.20) выполняется.

Вернемся к исходным ожерельям с непронумерованными бусинами. Их число N равно количеству орбит при действии группы G на множестве X, и в силу леммы Бернсайда имеем

$$N = \frac{1}{|G|} \sum_{g \in G} N(g) = \frac{1}{5}[N(e) + N(\tau_{2\pi/5}) + N(\tau_{4\pi/5}) + N(\tau_{6\pi/5}) + N(\tau_{8\pi/5})].$$

Заметим, что вращение на угол 0 оставляет на месте любое раскрашивание, а следовательно, $N(e) = |X| = 3^5 = 243$. Вращение на угол $2\pi/5$ оставляет на месте раскрашивания, при которых бусина с номером i окрашена так же, как и бусина с номером $\tau^{-1}_{2\pi/5}(i)$, $i = 1, 2, 3, 4, 5$, т. е. если $c_1 = c_2 = c_3 = c_4 = c_5 = c$ (все бусины окрашены в один цвет). Очевидно, что возможны три таких раскрашивания, а значит, $N(\tau_{2\pi/5}) = 3$. Аналогично по-

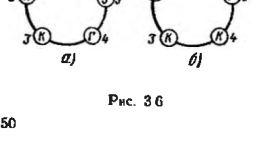


Рис. 3.6

лучаем $N_1(\tau_{2\pi/5})=3$, $i=2, 3, 4$, и, следовательно, $N=\frac{1}{5}(243+$ $+4\cdot 3)=255/5=51$.

Пример 3.23. Пусть $\Omega=\{1, 2, ..., n\}$ — множество номеров элементов некоторой фигуры Φ; R — конечное множество цветов, в которые могут быть окрашены элементы фигуры Φ; $X=$ $=\{<c_1, ..., c_n>\mid c_i\in R,\ i=1, 2, ..., n\}=R^n$ — множество раскрашиваний фигуры Φ. Пусть, далее, σ — некоторая подстановка из S_n (т. е. $\sigma:\Omega\to\Omega$). Определим действие подстановки σ на произвольное раскрашивание $c\in X$ по формуле

$$\sigma c=\sigma<c_1, ..., c_n>=<c_{\sigma^{-1}(1)}, ..., c_{\sigma^{-1}(n)}>. \tag{3.27}$$

Выведем формулу для $N(\sigma)$ — числа раскрашиваний, остающихся на месте при действии на них подстановки σ. Разложим σ в произведение независимых элементарных циклов

$$\sigma=\sigma_1...\sigma_k, \tag{3.28}$$

где k — количество циклов (в (3.28) учитываются и все циклы длины 1).

Напомним, что каждому элементарному циклу $(i_1\ i_2\ ...\ i_m)$ соответствует σ-орбита $\{i_1, i_2, ..., i_m\}$, и при этом совокупность всех σ-орбит является разбиением множества $\{1, 2, ..., n\}$. Пусть c — некоторое раскрашивание фигуры Φ, которое остается на месте при действии на него подстановки σ. Для любого цикла $\sigma_j=$ $=(i_1\ i_2\ ...\ i_m)$, входящего в разложение (3.28), имеем

$$i_1=\sigma^{-1}(i_2),\ i_2=\sigma^{-1}(i_3),\ ...,\ i_{m-1}=\sigma^{-1}(i_m),$$

откуда, используя то обстоятельство, что $c=\sigma c$, в силу (3.27) получаем

$$c_{i_m}=c_{\sigma^{-1}(i_m)}=c_{i_{m-1}}=c_{\sigma^{-1}(i_{m-1})}=c_{i_{m-2}}=...=c_{\sigma^{-1}(i_2)}=c_{i_1},$$

а следовательно, элементы фигуры Φ, входящие в одну и ту же σ-орбиту, должны быть окрашены одинаково. Поскольку σ-орбиты попарно не пересекаются, то они могут быть окрашены независимо друг от друга. Но тогда по правилу произведения $N(\sigma)=|R|^k$.

Пример 3.24. Составляются ожерелья из плоских бусин трех цветов, окрашенных одинаково с обеих сторон. Каждое ожерелье состоит из пяти бусин. Определим число N различных ожерелий.

По сравнению с примером 3.22 к группе вращений G фигуры, изображенной на рис. 3.5, *а*, добавляются осевые преобразования симметрии. У этой фигуры, очевидно, пять осей симметрии: l_1, l_2, l_3, l_4, l_5, где $\forall i\in\{1, 2, 3, 4, 5\}$ l_i — прямая, проходящая через i-ю бусину и центр окружности, на которой расположены бусины в ожерелье. Тогда, например, преобразова-

нию симметрии относительно оси l_1 соответствует подстановка $\sigma_1=(1)(2\ 5)(3\ 4)$. Используя пример 3.23, получаем $N(\sigma_1)=$ $=3^3=27$. Аналогично рассматриваются преобразования симметрии относительно других осей. Но тогда в силу леммы Бернсайда имеем $N=\frac{1}{10}(3^5+4\cdot 3+5\cdot 3^3)=39$.

3.2.5. Запас. Производящая функция запаса

Множество исследуемых объектов в комбинаторике часто называют запасом. Пусть A — запас, K — коммутативное кольцо с единицей (например, кольцо многочленов), а w — отображение вида $w: A\to K$. Будем $w(a)$, где $a\in A$, называть *весом* элемента a, а отображение w — *весовой функцией*. Сумма весов всех элементов запаса A называется *производящей функцией запаса A и обозначается* $W(A)$, т. е. $W(A)=\sum\limits_{a\in A} w(a)$.

Приведем примеры некоторых производящих функций:

1. Пусть запас — множество всех подмножеств некоторого конечного множества X, где $|X|=n$. Поставим в соответствие каждому множеству $Y\in P(X)$ вес $w(Y)=t^k$, где $k=|Y|$. Так как число k-элементных подмножеств n-элементного множества равняется C^k_n, то

$$W(P(X))=\sum_{k=0}^{n} C^k_n t^k$$

или с учетом утверждения 3.10

$$W(P(X))=\sum_{k=0}^{n} C^k_n t^k=(1+t)^n.$$

С помощью полученной производящей функции можно выводить различные свойства чисел C^k_n. Положив, например, $t=1$, имеем

$$\sum_{k=0}^{n} C^k_n=2^n.$$

Заметим далее, что для любых натуральных m, n выполняется цепочка равенств

$$\sum_{k=0}^{m+n} C^k_{m+n}t^k=(1+t)^{m+n}=(1+t)^m(1+t)^n=$$
$$=\sum_{i=0}^{m} C^i_m t^i\sum_{j=0}^{n} C^j_n t^j=\sum_{k=0}^{m+n}\sum_{s=0}^{k} C^s_m C^{k-s}_n t^k,$$

откуда получаем тождество Коши

$$C^k_{m+n}=\sum_{s=0}^{k} C^s_m C^{k-s}_n,\ k=0,\ 1,\ \ldots,\ m+n.$$

2. Пусть X — конечное множество, $|X|=n$, а запасом A является совокупность различных разбиений множества X на k

подмножеств в предположении, что подмножества в каждом разбиении расположены в строго определенном порядке. Поставим в соответствие каждому разбиению $X_1, \ldots, X_k$ множества X вес $w(X_1, \ldots, X_k)$:

$$w(X_1, \ldots, X_k) = t_1^{n_1} t_2^{n_2} \ldots t_k^{n_k}, \text{ где } n_i = |X_i|, \ i = 1, 2, \ldots, k.$$

В этом случае, используя утверждение 3.10, для производящей функции запаса A имеем

$$W(A) = \sum_{n_1+\ldots+n_k=n} C_n^{n_1, \ldots, n_k} t_1^{n_1} \ldots t_k^{n_k} = (t_1+\ldots+t_k)^n,$$

откуда, полагая, например, $t_1 = \ldots = t_k = 1$, получаем тождество

$$\sum_{n_1+\ldots+n_k=n} C_n^{n_1, \ldots, n_k} = k^n.$$

3. Пусть $X = \{x_1, \ldots, x_n\}$. Поставим в соответствие произвольной неупорядоченной выборке с повторениями объема k из множества X вес t^k. Пусть также для каждого номера $i \in \{1, 2, \ldots, n\}$ имеются ограничения на число повторений элемента x_i в выборке, задаваемые рядом

$$\sum_{j=0}^{\infty} c_{ij} t^j,$$

где $c_{ij} = \begin{cases} 1, & \text{если в выборку могут войти ровно } j \text{ элементов } x_i; \\ 0 & \text{— в противном случае.} \end{cases}$

Очевидно, что производящая функция запаса всех допустимых сочетаний с повторениями равняется

$$\prod_{i=1}^{n} \sum_{j=0}^{\infty} c_{ij} t^j.$$

Например, если в выборке допустимо не более чем двукратное повторение каждого элемента x_i, то производящая функция имеет вид $(1+t+t^2)^n$.

3.2.6. Цикловой индекс группы, действующей на множестве

Пусть G — конечная группа, X — конечное множество, где $|X| = n$. Пусть также τ — некоторая реализация группы G в $S(X)$, определяющая действие группы G на множестве X (т. е. задающая отображение $\langle g, x \rangle \to gx$ прямого произведения $G \times X$ в X, удовлетворяющее (3.19), (3.20)). Для любого элемента $g \in G$ обозначим через $j_k(g)$ количество циклов длины k в разложении подстановки $\tau_g \in S(X)$ в произведение независимых

элементарных циклов, где $k=1, 2, ..., n$. Каждому элементу $g \in G$ подставим в соответствие вес

$$w_G(g) = t_1^{j_1(g)} \; t_2^{j_2(g)} \ldots t_n^{j_n(g)} \tag{3.29}$$

(т. е. элемент кольца $\mathbf{Z}[t_1, ..., t_n]$). Тогда *цикловой индекс* $P(G, X, t_1, ..., t_n)$ группы G, действующей на X, есть многочлен от переменных $t_1, ..., t_n$, определяемый формулой

$$P(G, X, t_1, \ldots, t_n) = \frac{1}{|G|} \sum_{g \in G} w_G(g) = \frac{1}{|G|} \sum_{g \in G} t_1^{j_1(g)} \; t_2^{j_2(g)} \ldots t_n^{j_n(g)}. \tag{3.30}$$

Пример 3.25. Пусть $X=\{1, 2, ..., 10\}$, G — группа, действующая на X, и для некоторого элемента $g \in G$ выполняется равенство

$$\tau_g = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 9 & 6 & 1 & 8 & 5 & 2 & 7 & 10 & 3 & 4 \end{pmatrix}.$$

Определим вес $w(g)$ по формуле (3.29). Имеем $\tau_g = (1\,9\,3)\,(2\,6)\,(4\,8\,10)\,(5)\,(7)$, откуда $w_G(g) = t_1^2 t_2^1 t_3^2 t_4^0 \ldots t_{10}^0 = t_1^2 t_2 t_3^2$.

Пример 3.26. Пусть G — группа вращений на плоскости фигуры Φ (рис. 3.7) вокруг центра этой фигуры, совмещающих ее с самой собой, $X=\{1, 2, 3, 4, 5\}$ — множество номеров элементов фигуры. При этом (как и в примере 3.22) под композицией $\alpha \circ \beta$ двух произвольных вращений $\alpha, \beta \in G$ будем понимать вращение, являющееся результатом последовательно применяемых к фигуре Φ вращений β, α (т. е. сначала выполняется вращение β, а затем к полученной в результате вращения β фигуре применяется вращение α). Очевидно, что группа G состоит из четырех элементов — вращений (против часовой стрелки) на углы $2\pi i/4$, $i=0, 1, 2, 3$. Обозначим элементы группы G величинами соответствующих углов. Поставим в соответствие каждому вращению $\alpha \in G$ подстановку $\tau_\alpha \in S(X) = S_5$ такую, что для любого $i \in X$ $\tau_\alpha(i)$ будет номером i-го элемента фигуры, полученной в результате вращения α, относительно исходного положения фигуры. Тогда $\tau_0 = e$, $\tau_{\pi/2} = (1\,4\,3\,2)\,(5)$, $\tau_\pi = (1\,3)\,(2\,4)\,(5)$, $\tau_{3\pi/2} = (1\,2\,3\,4)\,(5)$.

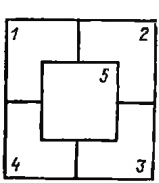


Рис. 3.7

Заметим, что по определению композиции вращений $\circ$ и отображения $\alpha \mapsto \tau_\alpha$ группы G в S_5 выполняется $\forall \alpha, \beta \in G$ $\tau_{\alpha \circ \beta} = \tau_\alpha \tau_\beta$, откуда следует, что отображение $\tau: G \to S_5$, ставящее в соответствие каждому элементу $\alpha \in G$ подстановку $\tau_\alpha \in S_5$, является гомоморфизмом, т. е. τ — реализация группы G в $S(X)$. При этом в соответствии с (3.30) цикловой индекс группы G, действующей на множестве X, равен

$$P(G, X, t_1, t_2, t_3, t_4, t_5) = \frac{1}{|G|} \sum_{g \in G} w_G(g) = \frac{1}{4} (w_G(0) + w_G(\pi/2) + w_G(\pi) +$$

$$+ w_G(3\pi/2)) = \frac{1}{4}(t_1^6 + t_1t_4 + t_1t_2^2 + t_1t_4) = \frac{1}{4}(t_1^6 + 2t_1t_4 + t_1t_2^2). \qquad (3.31)$$

3.2.7. G-эквивалентные отображения

Предположим, что сохраняются условия, описанные в разд. 3.2.6. Пусть, далее, R — некоторое конечное множество. Две функции $f_1, f_2 \in R^X$ будем называть G-эквивалентными и обозначать $f_1 \sim f_2$, если $\exists g \in G$: $\forall x \in X$ $f_1(x) = f_2(gx)$, или, что то же самое, если $\exists g \in G$: $\forall x \in X$ $f_1(x) = f_2(g^{-1}x)$.

Очевидно, что введенное бинарное отношение $\sim$ на множестве R^X является эквивалентностью, а следовательно, оно порождает разбиение множества R^X на классы эквивалентности $F \in R^X/\sim$.

Пусть каждому элементу $r \in R$ придан некоторый вес $w_R(r) \in K$, где K — кольцо многочленов над $\mathbf{Z}$ от некоторых переменных. Тогда вес функции $f \in R^X$ есть, по определению, $w(f) = \prod_{x \in X} w_R(f(x))$. Если $f_1 \sim f_2$, то очевидно, что $w(f_1) = w(f_2)$, поэтому можно определить вес класса эквивалентности $w(F)$, где $F \in R^X/\sim$, как вес $w(f)$ любого элемента $f \in F$.

Пример 3.27. Пусть $R = \{$красный, голубой$\}$ — множество цветов, $X = \{1, 2, 3, 4, 5\}$ — множество элементов фигуры Φ, рассмотренной в примере 3.26. Тогда любое отображение $f : X \to R$ можно рассматривать как раскрашивание элементов фигуры Φ в цвета из R. Придадим красному цвету вес a, а голубому — вес b (заметим, что a, b — элементы кольца $\mathbf{Z}[a, b]$). Тогда, например, любое раскрашивание f такое, что два элемента фигуры Φ окрашены в красный цвет, а остальные три элемента — в голубой, имеет вес $w(f) = a^2b^3$. Пусть далее на множестве X действует группа G, взятая из примера 3.26. Тогда в соответствии с введенным выше определением раскрашивания f_1, f_2 являются эквивалентными, если $\exists \alpha \in G$: $\forall i \in X$ $f_1(i) = f_2(\alpha^{-1}i)$, т. е. если найдется такое вращение $\alpha \in G$, что при совмещении фигуры Φ с раскрашиванием f_1 с фигурой, получаемой из фигуры Φ с раскрашиванием f_2 вращением на угол α (против часо-

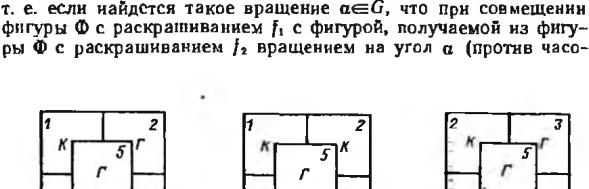


Рис. 3.8

вой стрелки), совмещенные друг с другом элементы будут окрашены одинаково. Например, раскрашивания f_1, f_2, показанные на рис. 3.8, *а*, *б*, эквивалентны, так как при совмещении фигуры Φ с раскрашиванием f_1 с фигурой, получаемой из фигуры Φ с раскрашиванием f_2 вращением на угол $\pi/2$ (против часовой стрелки), совмещаемые элементы раскрашены одинаково (см. на рис. 3.8, *в* результат вращения фигуры Φ с раскрашиванием f_2 на угол $\pi/2$; сравните с рис. 3.8, *а*). При этом $\forall i \in X\ f_1(i) = f_2((\pi/2)^{-1} i)$.

3.2.8. Производящая функция запаса классов эквивалентности

Предположим, что сохраняются условия, описанные в разд. 3.2.6 и 3.2.7. Пусть запас — множество классов эквивалентности $F \in R^X/\sim$. Возникает вопрос: какова производящая функция запаса $\sum_F w(F)$?

Зная производящую функцию запаса, можно решать некоторые достаточно сложные комбинаторные задачи. Действительно, пусть, например, $R = \{r_1, r_2\}$, $w(r_1) = a$, $w(r_2) = b$ и известна производящая функция запаса, представленная в виде

$$\sum_F W(F) = \sum_{m+k=n} c_{m,k} a^m b^k$$

(т. е. приведенная к виду многочлена, где суммирование в выражении справа производится по всем целым неотрицательным m, k, удовлетворяющим равенству $m+k=n$). Но тогда известны и все числа $c_{m,k}$, выражающие количество классов эквивалентности веса $a^m b^k$. Эти числа дают ответы на многие прикладные комбинаторные задачи, в частности, если r_1, r_2 — цвета, позволяют определить число различных (попарно неэквивалентных) раскрашиваний при заданной комбинации цветов. Так, в условиях примера 3.27 число $c_{3,2}$ выражает количество различных раскрашиваний фигуры Φ таких, что 3 элемента окрашены в красный цвет, а 2 — в голубой; $c_{3,2}+c_{2,3}+c_{1,4}+c_{0,5}$ — количество различных раскрашиваний фигуры Φ таких, что, по крайней мере, 2 элемента окрашены в голубой цвет.

Для практического построения производящей функции запаса классов эквивалентности воспользуемся следующей теоремой.

Теорема 3.1 *(теорема Пойа). Производящая функция запаса классов эквивалентности удовлетворяет равенству*

$$\sum_F w(F) = P(G, X, \sum_{r\in R} w_R(r), \sum_{r\in R} [w_R(r)]^2, \ldots, \sum_{r\in R} [w_R(r)]^n), \quad (3.32)$$

где $P(G, X, t_1, t_2, \ldots, t_n)$ — цикловой индекс группы G, действующей на X.

Следствие. Если веса $w_R(r)$, $r \in R$, выбраны равными 1, то число классов эквивалентности равно

$$P(G, X, |R|, \ldots, |R|). \quad (3.33)$$

Пример 3.28. Используя примеры 3.26 и 3.27, определим производящую функцию запаса классов эквивалентности. В силу теоремы Пойа, применяя формулу (3.31), имеем

$$\sum_{F} w(F) = P(G, X, a+b, a^2+b^2, a^3+b^3, a^4+b^4, a^5+b^5) = \frac{1}{4}[(a+b)^5 + 2(a+b)(a^4+b^4) + (a+b)(a^2+b^2)^2]. \quad (3.34)$$

Воспользовавшись равенством (3.34), определим, сколькими попарно неэквивалентными способами можно раскрасить фигуру Ф так, чтобы три ее элемента были окрашены в красный цвет, а два — в голубой. Для этого найдем коэффициент c при члене ca^3b^2 в производящей функции запаса классов эквивалентности, приведенной к виду многочлена. Очевидно, что $c = \frac{1}{4}(C^3_5 + 2) = \frac{1}{4}(10+2) = 3$, т. е. существуют три попарно неэквивалентных способа раскрашивания с заданной комбинацией цветов. Используя теперь следствие из теоремы Пойа, получаем, что общее число попарно неэквивалентных способов раскрашивания, выражаемое формулой (3.33), равно

$$\frac{1}{4}(2^5 + 2\cdot 2^2 + 2^3) = 12.$$

3.2.9. Обоснование теоремы Пойа

Для доказательства теоремы Пойа потребуются некоторые вспомогательные утверждения. Для любого натурального числа l обозначим $I_l = \{1, 2, \ldots, l\}$.

Утверждение 3.15. *Пусть K — коммутативное кольцо; $a_{ij} \in K$, $i = 1, \ldots, k$, $j = 1, \ldots, m$. Тогда*

$$\prod_{i=1}^{k} \sum_{j=1}^{m} a_{ij} = \sum_{\varphi \in I_m^{I_k}} \prod_{i=1}^{k} a_{i,\varphi(i)}, \quad (3.35)$$

где $I_m^{I_k}$ — множество всех отображений φ: $I_k \to I_m$.

Доказательство проведем индукцией по k. Очевидно, что при $k = 1$ равенство (3.35) выполняется. Пусть оно справедливо и при $k = l$ (где $k \geqslant 2$). Покажем его выполнение при k:

$$\prod_{i=1}^{k} \sum_{j=1}^{m} a_{ij} = \left(\prod_{i=1}^{k-1} \sum_{j=1}^{m} a_{ij}\right)\left(\sum_{j=1}^{m} a_{kj}\right) = \left(\sum_{\varphi \in I_m^{I_{k-1}}} \prod_{i=1}^{k-1} a_{i,\varphi(i)}\right)\left(\sum_{j=1}^{m} a_{kj}\right) =$$

$$= \sum_{\varphi \in I_m^{I_{k-1}}} \sum_{j=1}^{m} a_{kj} \prod_{i=1}^{k-1} a_{i,\varphi(i)} = \sum_{\varphi \in I_m^{I_k}} \prod_{i=1}^{k} a_{i,\varphi(i)}.$$

Утверждение 3.16. *Пусть $X_1, \ldots, X_k$ — непустые подмножества множества X, образующие разбиение множества X, т. е.*

$$\bigcup_{i=1}^{k} X_i = X, \quad X_i \cap X_j = \varnothing \quad \text{при } i \neq j.$$

Пусть далее $n_i = |X_i|$, $i = 1, \ldots, k$. *Тогда производящая функция запаса* функций из R^X, принимающих постоянные значения на каждом из X_i, равна

$$\prod_{i=1}^{k} \sum_{r \in R} [w_R(r)]^{n_i}. \tag{3.36}$$

Рассмотрим класс функций $R^{\overline{X}}$, где $\overline{X} = \{X_1, \ldots, X_k\}$. Определим веса $w(\varphi)$ функций $\varphi \subseteq R^{\overline{X}}$ по формуле

$$w(\varphi) = \prod_{i=1}^{k} [w_R(\varphi(X_i))]^{n_i}.$$

Очевидно, что между функциями из рассматриваемого запаса и функциями из $R^{\overline{X}}$ можно установить взаимно однозначное соответствие. Для этого каждой функции $f : X \to R$ рассматриваемого запаса поставим в соответствие функцию $\varphi_f \in R^{\overline{X}}$ такую, что $\forall i \in \{1, \ldots, k\}$ $\varphi_f(X_i) = f(x_i)$, где $x_i \in X_i$. При таком соответствии, очевидно, $w(\varphi_f) = w(f)$, а следовательно, производящая функция рассматриваемого запаса совпадает с производящей функцией запаса $R^{\overline{X}}$. Но тогда для доказательства утверждения 3.16 осталось заметить, что в силу утверждения 3.15 производящая функция запаса $R^{\overline{X}}$ выражается формулой (3.36) (здесь $a_{ij} = [w_R(r_j)]^{n_i}$, где $R = \{r_1, \ldots, r_m\}$).

Теперь докажем **теорему Пойа.**

Пусть D — множество значений весов элементов $f \in R^X$ и $H(d) = \{f \in R^X \mid w(f) = d\}$, где $d \in D$. Пусть группа G действует на $H(d)$ и при этом $\forall f \in H(d)$ gf определяется по правилу

$$(gf)(x) = f(g^{-1}x), \quad x \in X.$$

Заметим, что $\forall f \in H(d)$, $\forall g \in G$ $gf \sim f$, а следовательно, $\forall f \in H(d)$, $\forall g \in G$ $gf \in H(d)$. Покажем, что введенное отображение $\langle g, f \rangle \to gf$ прямого произведения $G \times H(d)$ в $H(d)$ действительно удовлетворяет условиям (3.19), (3.20). Условие (3.19) очевидным образом выполняется. Покажем справедливость (3.20). Действительно, $\forall g, h \in G$, обозначив $\tilde{f} = hf$, имеем

$$[(gh)f](x) = f((gh)^{-1}x) = f(h^{-1}(g^{-1}x)) = \tilde{f}(g^{-1}x) =$$
$$= (g\tilde{f})(x) = [g(hf)](x), \quad x \in X.$$

Используя лемму Бернсайда, получаем

$$\forall d \in D \quad |H(d)/\sim| = \frac{1}{|G|} \sum_{g \in G} |\{f \in H(d) \mid gf = f\}|. \tag{3.37}$$

Заметим, что $\forall F \in R^X/\sim$ $F \subseteq H(d)$, где $d = w(F)$, откуда

$$\forall F \in R^X/\sim \; F \in H(w(F))/\sim. \tag{3.38}$$

Пусть $d \in D$. В силу (3.38) имеем $\{F \in R^X/\sim \mid w(F) = d\} \subseteq H(d)/\sim$. Покажем также противоположное включение. Пусть $F_1 \in H(d)/\sim$. Рассмотрим произвольную функцию $f \in F_1$, а также класс эквивалентности $F \in R^X/\sim$ такой, что $f \in F$. Тогда $w(F) = w(f) = d$, а следовательно, в силу (3.38) $F \in H(d)/\sim$. Таким образом, $F, F_1 \in H(d)/\sim$, $F \cap F_1 \neq \varnothing$, откуда в силу утверждения 0.6 $F = F_1$, т. е. $F_1 \in \{F \in R^X/\sim \mid w(F) = d\}$. Итак, мы доказали, что

$$\forall d \in D \quad \{F \in R^X/\sim \mid w(F) = d\} = H(d)/\sim. \tag{3.39}$$

Используя теперь тот факт, что $H(d_1) \cap H(d_2) = \varnothing$ при $d_1 \neq d_2$, в силу (3.39) получаем, что $\{H(d)/\sim \mid d \in D\}$ — разбиение множества $R^X/\sim$, откуда, воспользовавшись (3.37), а также тем, что $\forall F \in H(d)/\sim \; w(F) = d$, имеем

$$\begin{aligned}
&\sum_{F \in R^X/\sim} w(F) = \sum_{d \in D} \sum_{F \in H(d)/\sim} w(F) = \sum_{d \in D} d \mid H(d)/\sim \mid = \\
&= \sum_{d \in D} \frac{1}{|G|} \sum_{g \in G} d \mid \{f \in H(d) \mid gf = f\} \mid = \\
&= \sum_{d \in D} \frac{1}{|G|} \sum_{g \in G} \sum_{f \in \{f \in H(d) \mid gf = f\}} w(f) = \\
&= \frac{1}{|G|} \sum_{g \in G} \sum_{d \in D} \sum_{f \in \{f \in H(d) \mid gf = f\}} w(f) = \\
&= \frac{1}{|G|} \sum_{g \in G} \sum_{f \in \{f \in R^X \mid gf = f\}} w(f) .
\end{aligned} \tag{3.40}$$

Заметим, далее, что $\forall g \in G$ подстановка τ_g разбивает множество X на τ_g-орбиты $X^{(g)}_1, \ldots, X^{(g)}_{k(g)}$, где $k(g)$ — число орбит. Обозначим $n_i(g) = |X^{(g)}_i|$, где $i = 1, \ldots, k(g)$. Нетрудно видеть, что

$$gf = f \Leftrightarrow \forall i \in \{1, \ldots, k(g)\} \; \exists c_i \in R: \; \forall x \in X^{(g)}_i \; f(x) = c_i.$$

Но тогда, используя утверждение 3.16, в силу (3.40) получаем

$$\begin{aligned}
&\sum_{F \in R^X/\sim} w(F) = \frac{1}{|G|} \sum_{g \in G} \sum_{f \in \{f \in R^X \mid gf = f\}} w(f) = \\
&= \frac{1}{|G|} \sum_{g \in G} \prod_{i=1}^{k(g)} \sum_{r \in R} [w_R(r)]^{n_i(g)} = \\
&= \frac{1}{|G|} \sum_{g \in G} \prod_{i=1}^{n} \{\sum_{r \in R} [w_R(r)]^i\}^{j_i(g)} = \\
&= P(G, \sum_{r \in R} w_R(r), \sum_{r \in R} [w_R(r)]^2, \ldots, \sum_{r \in R} [w_R(r)]^n).
\end{aligned}$$

Задачи и упражнения

1. На квадратных листах бумаги пишут пятизначные числа от 0 до 99999. Если в некотором числе менее пяти значащих цифр, то его дополняют слева нулями (например, пишут 00536 вместо 536). Считается, что после вращения листа на угол π цифры 0, 1, 8 не меняются, а цифры 6, 9 переходят друг в друга. Сколько нужно иметь листов бумаги для записи всех чисел с учетом возможности вращения листов?

2. Составляются ожерелья из плоских бусин трех цветов, окрашенных одинаково с обеих сторон. Каждое ожерелье состоит из шести бусин. Определить число различных ожерелий.

3. Определить, сколькими способами можно раскрасить тремя цветами пронумерованные элементы фигуры Ф. Два способа раскрашивания считаются одинаковыми, если один получается из другого вращением фигуры на плоскости (вокруг центра Ф). Варианты фигуры Ф показаны на рис. 3.9, *а*, *б*.

4. Определить, сколькими способами можно раскрасить тремя цветами (красным, голубым, зеленым) пронумерованные элементы фигуры Ф, предполагая, что n_1 элементов должны быть окрашены в красный цвет, n_2 — в голубой, n_3 — в зеленый. Два способа раскрашивания считаются одинаковыми, если один получается из другого вращением фигуры на плоскости (вокруг центра Ф). Варианты фигуры Ф показаны на рис. 3.9, *а*, *б*. Соответствующие им наборы значений n_1, n_2, n_3 равны $n_1=3$, $n_2=4$, $n_3=2$ или $n_1=1$, $n_2=4$, $n_3=4$.

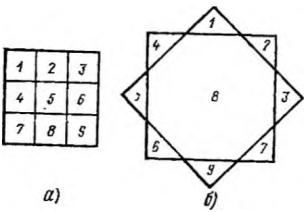


Рис. 3.9

5. Решить задачи 3 и 4 в предположении, что два способа раскрашивания считаются одинаковыми, если один получается из другого вращением фигуры Ф на плоскости (вокруг центра Ф) или в результате осевого преобразования симметрии.

КОНЕЧНЫЕ ГРАФЫ И СЕТИ

ГЛАВА 4

4.1. ОСНОВНЫЕ ПОНЯТИЯ И ОПРЕДЕЛЕНИЯ

Прежде чем определить понятие конечного графа в наиболее общей форме, представим себе на плоскости (или в вещественном аффинном пространстве произвольной размерности k) некоторое конечное множество V точек и конечный набор X линий, соединяющих некоторые пары точек из V. Указанной геометрической конфигурацией описывается, например, схема автомобильных дорог, связывающих города некоторой области (рис. 4.1).

Для многих задач оказывается несущественным, соединены ли точки конфигурации отрезками прямых или криволинейными дугами (например, при решении задачи о нахождении маршрута движения по дорогам, связывающего два заданных города и проходящего через минимальное число дорог). Важно лишь то, что каждая линия соединяет какие-либо две из заданного набора точек. При рассмотрении подобных задач достаточно ограничиться исследованием совокупности двух конечных множеств V, X, где V — непустое множество, X — некоторый набор пар элементов из V вида (v, w). Введенная пара множеств (V, X) допускает также многочисленные другие интерпретации и является предметом детального изучения в математике. Элементы множества V будем называть *вершинами*, а элементы набора X — *ребрами*. В общем случае в наборе X могут встречаться пары с одинаковыми элементами вида (v, v), а также одинаковые пары. Ребра вида (v, v) называются *петлями*. Одинаковые пары в X называются *кратными* (или *параллельными*) ребрами. Количество одинаковых пар (v, w) в X называется *кратностью* ребра (v, w). Про множество V и набор X будем говорить, что они определяют *граф с кратными ребрами и петлями* (или *псевдограф*) $G = (V, X)$. Псевдограф без петель на-

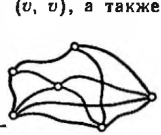


Рис. 4.1

зывается графом с кратными ребрами (или *мультиграфом*). Если в наборе X ни одна пара не встречается более одного раза, то мультиграф $G=(V, X)$ называется *графом*. Если пары в наборе X являются упорядоченными, то граф называется *ориентированным* (кратко — *орграфом*). Ребра орграфа называются *дугами*. Если пары в наборе X являются неупорядоченными, то граф называется *неориентированным* графом (или просто графом). Ребра в неориентированном графе (в отличие от дуг в орграфе) будем обозначать (v, w). Неориентированные графы будем обозначать буквой G или G с индексами (например, G_0, G_1, ...), а орграфы — буквой D или D с индексами (например, D_0, D_1, ...). Кроме того, договоримся обозначать вершины буквами v, u, w (без индексов или с индексами), а ребра и дуги—буквами x, y, z (без индексов или с индексами).

Всюду далее будем соответствующую некоторому графу геометрическую конфигурацию, в которой вершины изображены кружочками, а ребра — линиями, соединяющими соответствующие вершины, называть *изображением* этого графа. При изображении орграфа направления дуг будем отмечать стрелками, примыкающими к их концам.

Пример 4.1. Пусть $V=\{v_1, v_2, v_3, v_4\}$, $X=\{x_1=(v_1, v_2)$, $x_2=(v_1, v_2)$, $x_3=(v_2, v_2)$, $x_4=(v_2, v_3)\}$. Тогда $D=(V, X)$ — ориентированный псевдограф, изображение которого приведено на рис. 4.2.

Пример 4.2. Пусть $V=\{v_1, v_2, v_3, v_4, v_5\}$, $X=\{x_1=(v_1, v_2)$, $x_2=\{v_2, v_3\}$, $x_3=\{v_2, v_4\}$, $x_4=\{v_3, v_4\}\}$. Тогда $G=(V, X)$ — неориентированный граф (или просто граф), изображение которого дано на рис. 4.3.

←Рис. 4.2

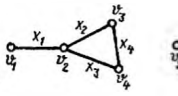


Рис. 4.3→

Приведем ряд понятий и определений для ориентированных и неориентированных графов. Там, где это не оговорено особо, те же понятия и определения переносятся без изменений на ориентированные и неориентированные псевдографы.

4.1.1. Смежность, инцидентность, степени

Если $x=(v, w)$ — ребро графа, то вершины v, w называются *концами* ребра x; в этом случае также говорят, что ребро x *соединяет* вершины v, w.

Если $x=(v, w)$ — дуга орграфа, то вершина v называется *началом*, а вершина w — *концом* дуги x; в этом случае также говорят, что дуга x *исходит* из вершины v и *заходит* в вершину w.

Если вершина v является концом (началом или концом) ребра (дуги) x, то говорят, что v и x *инцидентны*.

Вершины v, w графа $G=(V, X)$ называются *смежными*, если $\{v, w\}\in X$. Два ребра называются смежными, если они имеют общую вершину.

Степенью вершины v графа G называется число $\delta(v)$ ребер графа G, инцидентных вершине v. Вершина графа, имеющая степень 0, называется *изолированной*, а степень 1 — *висячей*.

Замечание 4.1. В случае неориентированного псевдографа обычно считается, что вклад каждой петли, инцидентной некоторой вершине v, в $\delta(v)$ равен 2 (тогда как вклад любого другого ребра, инцидентного вершине v, равен 1).

Полустепенью исхода (захода) вершины v орграфа D называется число $\delta^+(v)$ ($\delta^-(v)$) дуг орграфа D, исходящих из вершины v (заходящих в вершину v).

Замечание 4.2. В случае ориентированного псевдографа вклад каждой петли, инцидентной некоторой вершине v, равен 1, как в $\delta^+(v)$, так и в $\delta^-(v)$.

Пример 4.3.

1. В графе G (см. пример 4.2) концами ребра x_1 являются вершины v_1, v_2; вершина v_2 инцидентна ребрам x_1, x_2, x_3; степень вершины v_2 равна 3, т. е. $\delta(v_2)=3$; вершины v_1, v_2 смежные, ребра x_1, x_2 смежные; вершина v_4 висячая; вершина v_5 изолированная.

2. В ориентированном псевдографе (см. пример 4.1) дуга x_1 исходит из вершины v_1 и заходит в вершину v_2; вершина v_2 инцидентна дугам x_1, x_2, x_3, x_4; $\delta^+(v_2)=2$, $\delta^-(v_2)=3$.

Количество вершин и ребер в графе G обозначим соответственно через $n(G)$ и $m(G)$, а количество вершин и дуг в орграфе D — через $n(D)$ и $m(D)$.

Утверждение 4.1 *Для любого псевдографа G выполняется равенство*

$$\sum_{v\in V}\delta(v)=2m(G). \tag{4.1}$$

Равенство (4.1) является очевидным следствием того, что вклад каждого ребра в сумму из левой части (4.1) равен 2.

Приведем также соответствующее утверждение для орграфа.

Утверждение 4.2. *Для любого ориентированного псевдографа D выполняется равенство*

$$\sum_{v\in V}\delta^+(v)=\sum_{v\in V}\delta^-(v)=m(D). \tag{4.2}$$

Доказательство (4.2) очевидно.

4.1.2. Изоморфизм, гомеоморфизм

Графы $G_1=(V_1, X_1)$ и $G_2=(V_2, X_2)$ называются *изоморфными*, если существует биективное (взаимно однозначное) отображение $\varphi: V_1\to V_2$, сохраняющее смежность, т. е. $\{v, w\}\in X_1 \Leftrightarrow \{\varphi(v), \varphi(w)\}\in X_2$. Соответственно орграфы $D_1=(V_1, X_1)$ и

$D_2=(V_2, X_2)$ называются изоморфными, если существует биективное отображение $\varphi: V_1 \to V_2$ такое, что

$$(v, w) \in X_1 \Leftrightarrow (\varphi(v), \varphi(w)) \in X_2. \quad (4.3)$$

Замечание 4.3. Из определений следует, что изоморфные графы (орграфы) отличаются лишь обозначением вершин.

Приведем следующие очевидные свойства изоморфизма.

1) если графы $G_1=(V_1, X_1)$, $G_2=(V_2, X_2)$ изоморфны и $\varphi: V_1 \to V_2$ — биективное отображение, сохраняющее смежность, то: а) $\forall v \in V_1 \ \delta(v)=\delta(\varphi(v))$; б) $m(G_1)=m(G_2)$, $n(G_1)=n(G_2)$;

2) если орграфы $D_1=(V_1, X_1)$, $D_2=(V_2, X_2)$ изоморфны и $\varphi: V_1 \to V_2$ — биективное отображение, удовлетворяющее (4.3), то: а) $\forall v \in V_1 \ \delta^+(v)=\delta^+(\varphi(v))$, $\delta^-(v)=\delta^-(\varphi(v))$; б) $m(D_1)=m(D_2)$, $n(D_1)=n(D_2)$.

Замечание 4.4. Для псевдографов определение изоморфизма несколько усложняется. Псевдографы $G_1=(V_1, X_1)$, $G_2=(V_2, X_2)$ называются изоморфными, если существуют два таких биективных отображения $\varphi: V_1 \to V_2$, $\psi: X_1 \to X_2$, что $\forall x=\{v, w\} \in X_1$ $\psi(x)=\{\varphi(v), \varphi(w)\}$. Это означает дополнительное требование на сохранение кратностей соответствующих ребер. Аналогично ориентированные псевдографы $D_1=(V_1, X_1)$, $D_2=(V_2, X_2)$ называются изоморфными, если существуют такие биективные отображения $\varphi: V_1 \to V_2$, $\psi: X_1 \to X_2$, что $\forall x=(v, w) \in X_1$ $\psi(x)=(\varphi(v), \varphi(w))$.

Нетрудно показать эквивалентность приведенных определений изоморфизма для случая, когда в G_1, G_2 (в D_1, D_2) отсутствуют кратные ребра (дуги).

Пример 4.4. Графы, изображенные на рис. 4.4, *а*, *б*, изоморфны, но они не изоморфны графу, изображенному на рис. 4.4, *в* (почему?).

Отметим, что изоморфизм графов (орграфов) является отношением эквивалентности на множестве графов (орграфов).

Операция подразбиения (измельчения) дуги (u, v) в орграфе $D=(V, X)$ состоит в удалении из X дуги (u, v), добавлении к V новой вершины w и добавлении к $X \setminus \{(u, v)\}$ двух дуг (u, w), (w, v). Аналогично определяется операция подразбиения ребра в графах.

Орграф D_1 называется *подразбиением* орграфа D_2, если орграф D_1 можно получить из D_2 путем последовательного применения операции подразбиения дуг. Аналогично определяется подразбиение графа.

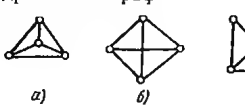


Рис. 4.4

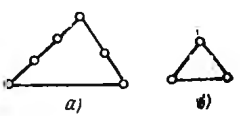


Рис. 4.5

Пример 4.5. Граф, изображенный на рис. 4.5, *а*, является подразбиением графа, изображенного на рис. 4.5, *б*.

Орграфы D_1, D_2 (графы G_1, G_2) называются *гомеоморфными*, если существуют их подразбиения, являющиеся изоморфными.

Пример 4.6. Графы, изображенные на рис. 4.5, гомеоморфны.

4.1.3. Маршруты, пути

Введем понятие *маршрута* для графа $G=(V, X)$ (и соответственно понятие *пути* для орграфа $D=(V, X)$). Последовательность

$$v_1x_1v_2x_2v_3\ldots x_kv_{k+1} \tag{4.4}$$

(где $k\geqslant 1$, $v_i\in V$, $i=1, \ldots, k+1$, $x_j\in X$, $j=1, \ldots, k$), в которой чередуются вершины и ребра (дуги) и для каждого $j=1, \ldots, k$ ребро (дуга) x_j имеет вид $\{v_j, v_{j+1}\}$ ((v_j, v_{j+1})), называется маршрутом, соединяющим вершины v_1, v_{k+1} (путем из v_1 в v_{k+1}). При этом v_1 называется *начальной*, v_{k+1} — *конечной* вершинами маршрута (пути) (4.4), а остальные вершины — *внутренними*. Одна и та же вершина может одновременно оказаться начальной, конечной и внутренней. Последовательность вершин в маршруте определяет на ребрах, входящих в маршрут, ориентацию. Заметим в этой связи, что ориентацию некоторого ребра $x=\{v, w\}$ всегда можно указать при записи его как пары вершин. Например, запись $\{v, w\}$ указывает на то, что ребро x ориентировано от вершины v к вершине w.

Пример 4.7.

1. Последовательность $v_1x_1v_2x_3v_4x_4v_3$ — маршрут, соединяющий вершины v_1, v_3 в графе G (см. пример 4.2).

2. Последовательность $v_1x_2v_2x_3v_2x_4v_3$ — путь из v_1 в v_3 в ориентированном псевдографе D (см. пример 4.1).

Замечание 4.5. Последовательность (4.4) можно однозначно восстановить по последовательности

$$x_1x_2\ldots x_k \tag{4.5}$$

(если (4.4) — маршрут, предполагается, что в последовательности (4.5) дополнительно указывается ориентация ребер, определяемая маршрутом (4.4)), а следовательно, вместо (4.4) можно использовать более короткую запись (4.5). Отметим далее, что в случае, когда в последовательности (4.4) $x_1, \ldots, x_k$ имеют кратности, равные 1, ее можно однозначно восстановить по последовательности вершин

$$v_1v_2\ldots v_{k+1}, \tag{4.6}$$

а следовательно, вместо (4.4) также можно использовать более короткую запись (4.6). В общем случае вместо последовательности (4.4) можно использовать сокращенную последовательность, в которой опущены все x_i кратности 1.

Пример 4.8.

1. Последовательности $x_1x_3x_4$, $v_1v_2v_4v_1$ — сокращенные записи маршрута, приведенного в примере 4.7, п. 1.

2. Последовательности $x_2x_3x_4$, $v_1x_2v_2v_2v_3$ — сокращенные записи пути, приведенного в примере 4.7, п. 2.

Пусть $x_1x_2...x_k$ — маршрут в графе G (см. замечание 4.5) и для некоторой последовательности номеров i_1, ..., i_r, где $r \geqslant 1$, $1 \leqslant i_1 < i_2 < ... < i_r \leqslant k$, $x_{i_1}x_{i_2}...x_{i_r}$ снова является маршрутом в графе G. Тогда $x_{i_1}x_{i_2}...x_{i_r}$ называется *подмаршрутом* маршрута $x_1x_2...x_k$. При этом будем говорить, что маршрут $x_{i_1}x_{i_2}...x_{i_r}$ *выделен* из маршрута $x_1x_2...x_k$. Аналогично определяется *подпуть*, выделенный из пути орграфа.

Число ребер (дуг) в маршруте (пути) называется *длиной* маршрута (пути).

Маршрут (путь) (4.4) называется *замкнутым*, если его начальная вершина совпадает с конечной, т. е. если $v_1 = v_{k+1}$.

Замечание 4.6. Далее всюду при подсчете числа вхождений вершин в замкнутый маршрут (путь) начальную и конечную вершины будем считать за одно вхождение этой вершины в маршрут (путь).

Замечание 4.7. При замкнутом маршруте (пути) $x_1x_2...x_k$ обычно считается, что последовательности $x_1x_2...x_k$, $x_2x_3...x_kx_1$, ..., $x_kx_1x_2...x_{k-1}$ — различные записи одного и того же маршрута (пути).

Незамкнутый маршрут (путь), в котором все ребра (дуги) попарно различны, называется *цепью*. Цепь, в которой все вершины попарно различны, называется *простой цепью*.

Замкнутый маршрут (путь), в котором все ребра (дуги) попарно различны, называется *циклом (контуром)*. Цикл (контур), в котором все вершины попарно различны (см. замечание 4.6), называется *простым*.

Пример 4.9. Рассмотрим граф G из примера 4.2. Тогда:

1) $v_1x_1v_2x_3v_4x_4v_3$ — маршрут длины 3, соединяющий v_1, v_3; это простая цепь, так как все ребра и вершины попарно различны;

2) $v_2x_2v_3x_4v_4x_3v_2$ — замкнутый маршрут длины 3; это простой цикл, так как все ребра и вершины попарно различны;

3) $v_1x_1v_2x_2v_3x_4v_4x_3v_2$ — маршрут длины 4, соединяющий вершины v_1, v_2; это цепь, которая не является простой, так как вершина v_2 встречается дважды;

4) $v_1x_1v_2x_2v_3x_2v_2$ — маршрут длины 3, соединяющий вершины v_1, v_2 и не являющийся цепью, так как ребро $x_2 = \{v_2, v_3\}$ встречается дважды.

Пример 4.10. Рассмотрим ориентированный псевдограф D из примера 4.1. Тогда:

1) $v_1x_1v_2x_4v_3$ — путь длины 2 из v_1 в v_3; это простая цепь, так как все дуги и вершины попарно различны;

2) $v_3x_3v_2$ — простой контур длины 1;

3) $v_1x_2v_2x_3v_2x_4v_3$ — цепь из v_1 в v_3 длины 3, которая не является простой.

Нетрудно показать, что в псевдографах, мультиграфах и графах минимальная длина простого цикла равна соответственно 1, 2 и 3 (каковы минимальные длины простых контуров в ориентированных псевдографах, мультиграфах и графах?).

Утверждение 4.3. *В псевдографе G (в ориентированном псевдографе D) из всякого цикла (замкнутого пути) можно выделить простой цикл (простой контур).*

Доказательство будем проводить для G (для D доказательство аналогично) индукцией по k — количеству ребер в цикле. При $k=1$ всякий цикл является простым. Пусть при некотором $k\geqslant 2$ доказываемое утверждение справедливо для любого цикла длины $\leqslant k-1$. Покажем его справедливость для произвольного цикла $\mu=v_1x_1...v_kx_kv_1$ длины k. Рассмотрим любые два номера i, j, где $1\leqslant i<j\leqslant k$, такие, что $v_i=v_j$. Если таких номеров нет, то цикл μ является простым (по определению). Если же указанные номера нашлись, то рассматриваем цикл $v_ix_i...v_{j-1}x_{j-1}v_j$ длины $j-i\leqslant k-1$, а из него в силу индуктивного предположения можно выделить простой цикл.

Утверждение 4.4. *Из всякого незамкнутого маршрута (пути) можно выделить простую цепь с теми же начальной и конечной вершинами.*

Доказательство будем проводить для маршрута (для пути доказательство аналогично) индукцией по k — количеству ребер в маршруте. При $k=1$ всякий маршрут является простой цепью. Пусть при некотором $k\geqslant 2$ доказываемое утверждение справедливо для любого маршрута длины $\leqslant k-1$. Покажем его справедливость для произвольного маршрута $\eta=v_1x_1v_2...x_kv_{k+1}$, где $v_1\neq v_{k+1}$, длины k. Рассмотрим два номера i, j, где $1\leqslant i<j\leqslant k+1$, такие, что $v_i=v_j$. Если таких номеров нет, то маршрут η является простой цепью. Если же указанные номера нашлись, то рассматриваем подмаршрут $\eta'=v_1x_1v_2...x_{i-1}v_ix_jv_{j+1}...$ $...x_kv_{k+1}$ (т. е. предполагаем, что $i\neq 1$, $j\neq k+1$; случаи, когда $i=1$ или $j=k+1$, рассмотрите самостоятельно) длины $\leqslant k-1$, а из него в силу индуктивного предположения можно выделить простую цепь, соединяющую вершины v_1, v_{k+1}.

Введем понятие *композиции* путей (маршрутов). Пусть $\pi_1=v_1x_1v_2...x_{k-1}v_k$, $\pi_2=v_kx_kv_{k+1}...x_{l-1}v_l$ — пути в орграфе D, где $k\geqslant 2$, $l\geqslant k+1$. Назовем путь

$\pi_1\bigcirc\pi_2=v_1x_1v_2...x_{k-1}v_kx_kv_{k+1}...x_{l-1}v_l$

(очевидно, что $\pi_1\bigcirc\pi_2$ — путь в D) композицией путей π_1, π_2. Аналогично определяется композиция маршрутов.

4.1.4. Матричное задание графов. Матрицы смежности, инцидентности

Пусть $D=(V, X)$ — орграф, где $V=\{v_1, ..., v_n\}$, $X=\{x_1, ..., x_m\}$.

Матрицей смежности орграфа D называется квадратная матрица $A(D)=[a_{ij}]$ порядка n, у которой

$$a_{ij}=\begin{cases}1, \text{ если } (v_i, v_j)\in X;\\ 0, \text{ если } (v_i, v_j)\subset X.\end{cases}$$

Матрицей инцидентности (или матрицей инциденций) орграфа D называется $(n\times m)$-матрица $B(D)=[b_{ij}]$, у которой

$$b_{ij}=\begin{cases}1, \text{ если вершина } v_i \text{ является концом дуги } x_j;\\ -1, \text{ если вершина } v_i \text{ является началом дуги } x_j;\\ 0, \text{ если вершина } v_i \text{ не инцидентна дуге } x_j.\end{cases}$$

Введем также матрицы смежности и инцидентности для неориентированных графов. Пусть $G=(V, X)$ — граф, где $V=\{v_1, \ldots, v_n\}$, $X=\{x_1, \ldots, x_m\}$.

Матрицей смежности графа G называется квадратная матрица $A(G)=[a_{ij}]$ порядка n, у которой

$$a_{ij}=\begin{cases}1, \text{ если } \{v_i, v_j\}\in X;\\ 0, \text{ если } \{v_i, v_j\}\in X.\end{cases}$$

Матрицей инцидентности графа G называется $(n\times m)$-матрица $B(G)=[b_{ij}]$, у которой

$$b_{ij}=\begin{cases}1, \text{ если вершина } v_i \text{ инцидентна ребру } x_j;\\ 0, \text{ если вершина } v_i \text{ не инцидентна ребру } x_j.\end{cases}$$

Пример 4.11. Для орграфа D, изображенного на рис. 4.6, матрица $A(D)$ приводится в табл. 4.1*а*, а матрица $B(D)$ — в табл. 4.1*б*.

Пример 4.12. Для графа G, изображенного на рис. 4.7, матрица $A(G)$ приводится в табл. 4.2*а*, а матрица $B(G)$ — в табл. 4.2*б*.

Замечание 4.8. Матрицу смежности можно определить и для псевдографов. Тогда в случае ориентированного (неориентированного) псевдографа $a_{ij}=k$, где k — кратность дуги (v_i, v_j) (ребра $\{v_i, v_j\}$) в этом псевдографе. Определение матрицы инцидентности без изменений переносится и на произвольные мультиграфы (ориентированные и неориентированные) и даже на неориентированные псевдографы.

Пример 4.13. Для ориентированного псевдографа D, изображенного на рис. 4.8, матрица $A(D)$ приводится в табл. 4.3.

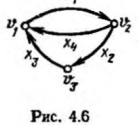


Рис. 4.6

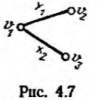


Рис. 4.7

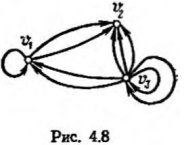


Рис. 4.8

Таблица 4.1а

	v_1	v_2	v_3
v_1	0	1	0
v_2	1	0	1
v_3	1	0	0

Таблица 4.1б

	x_1	x_2	x_3	x_4
v_1	−1	0	1	1
v_2	1	−1	0	−1
v_3	0	1	−1	0

Таблица 4.2а

	v_1	v_2	v_3
v_1	0	1	1
v_2	1	0	0
v_3	1	0	0

Таблица 4.2б

	x_1	x_2
v_1	1	1
v_2	1	0
v_3	0	1

Таблица 4.3

	v_1	v_2	v_3
v_1	1	2	0
v_2	0	0	0
v_3	2	3	2

Нетрудно видеть, что матрица $A(G)$ является симметричной для любого неориентированного графа G. Матрица $A(D)$, где D — орграф, в общем случае не является симметричной (см. примеры 4.11 и 4.13).

По матрице смежности графа (орграфа) всегда можно определить ребра графа (дуги орграфа) как пары инцидентных им вершин, а для псевдографов, кроме того, и кратности ребер (дуг). Однако, если ребра (дуги) были пронумерованы, то восстановить их номера по матрице смежности невозможно. В этом смысле матрица инцидентности оказывается более информативной, чем матрица смежности, поскольку позволяет получить полную информацию о ребрах (дугах), включая их нумерацию.

С помощью введенных матриц удобно задавать графы (орграфы) для обработки на ЭВМ. Однако следует отметить, что при большом количестве вершин матрица смежности оказывается громоздкой и число элементов в ней может превысить допустимый объем оперативной памяти ЭВМ. То же можно сказать и о матрице инцидентности, причем ее размеры зависят, кроме того, и от количества ребер (дуг).

Приведем очевидные свойства матриц смежности и инцидентности:

1. Сумма элементов матрицы $A(G)$, где $G=(V, X)$ — мультиграф, $V=\{v_1, ..., v_n\}$, по i-й строке (или по i-му столбцу) равна $\delta(v_i)$.

2. Суммы элементов матрицы $A(D)$, где $D=(V, X)$ — ориентированный псевдограф, $V=\{v_1, ..., v_n\}$, по i-й строке и по i-му столбцу соответственно равны $\delta^+(v_i)$, $\delta^-(v_i)$.

3. Пусть D — ориентированный мультиграф с непустым множеством дуг. Тогда

а) сумма строк матрицы $B(D)$ является нулевой строкой;

б) любая строка матрицы $B(D)$ является линейной комбинацией остальных строк;

в) ранг матрицы $B(D)$ не превосходит $n(D)-1$;

г) для любого контура в D сумма столбцов матрицы $B(D)$, соответствующих дугам, входящим в этот контур, равна нулевому столбцу.

4. Пусть G — мультиграф с непустым множеством ребер. Тогда при покоординатном сложении по модулю 2:

а) сумма строк матрицы $B(G)$ является нулевой строкой;

б) любая строка матрицы $B(G)$ является суммой остальных строк;

в) для любого цикла в G сумма столбцов матрицы $B(G)$, соответствующих ребрам, входящим в этот цикл, равна нулевому столбцу.

Обозначим через $A^k=[a^{(k)}_{ij}]$ k-ю степень матрицы смежности $A=A(D)$ орграфа D (аналогичное обозначение вводим и для графа G).

Утверждение 4.5. *Элемент $a^{(k)}_{ij}$ матрицы $A^{(k)}$ ориентированного псевдографа $D=(V, X)$ (псевдографа $G=(V, X)$), где $V=\{v_1, ..., v_n\}$, равен числу всех путей (маршрутов) длины k из v_i в v_j (соединяющих v_i, v_j).*

Доказательство проведем для D (для G оно аналогично) индукцией по k. При $k=1$ справедливость доказываемого утверждения следует непосредственно из определения матрицы $A=A(D)$. Предположим, что доказываемое утверждение справедливо при $k=k'$, где $k'\geqslant 1$. Покажем его справедливость и при $k=k'+1$. Обозначим через $\Pi(D, v_i, v_j, r)$, где $r\geqslant 1$, множество путей длины r из v_i в v_j в ориентированном псевдографе D. Разобьем множество путей $\Pi(D, v_i, v_j, k'+1)$ на n групп. Первая группа — это множество $\Pi(D, v_i, v_1, v_j, k'+1)$ путей с предпоследней вершиной v_1, вторая группа — множество $\Pi(D, v_i, v_2, v_j, k'+1)$ путей с предпоследней вершиной v_2 и т. д., n-я группа — множество $\Pi(D, v_i, v_n, v_j, k'+1)$ путей с предпоследней вершиной v_n. Очевидно, что совокупность указанных групп является разбиением множества $\Pi(D, v_i, v_j, k'+1)$, а следовательно,

$$|\Pi(D, v_i, v_j, k'+1)|=\sum_{l=1}^{n}|\Pi(D, v_i, v_l, v_j, k'+1)|. \tag{4.7}$$

Заметим, что по правилу произведения

$$|\Pi(D, v_i, v_l, v_j, k'+1)|=|\Pi(D, v_i, v_l, k')||\Pi(D, v_l, v_j, 1)|=$$
$$=|\Pi(D, v_i, v_l, k')|a_{lj},\ l=1, 2, ..., n. \tag{4.8}$$

Из (4.7), (4.8), используя то, что в силу индуктивного предположения выполняется равенство
$|\Pi(D, v_i, v_l, k')| = a_{il}^{(k')}$, $l = 1, 2, \ldots, n$,

получаем

$$|\Pi(D, v_i, v_j, k'+1)| = \sum_{l=1}^{n} a_{il}^{(k')} a_{lj} = a_{ij}^{(k'+1)}.$$

Утверждение 4.5 имеет многочисленные следствия. Приведем, например, утверждение, позволяющее по степеням матрицы смежности определять наличие контуров в орграфе D.

Утверждение 4.6. *Для того чтобы n-вершинный орграф D с матрицей смежности $A = A(D)$ имел хотя бы один контур, необходимо и достаточно, чтобы матрица $K = A^2 + A^3 + \ldots + A^n$ имела ненулевые диагональные элементы.*

Достаточность. Пусть $K = [k_{ij}]$ и для некоторого номера i выполняется $k_{ii} > 0$. В этом случае для некоторого $r \in \{2, \ldots, n\}$ справедливо $a_{ii}^{(r)} > 0$, а следовательно, в силу утверждения 4.5 найдется путь в D из v_i в v_i. Но тогда в силу утверждения 4.3 в орграфе D найдется простой контур.

Необходимость. Пусть в орграфе D имеется некоторый контур. В утверждении 4.3 было показано, что из всякого контура можно выделить простой контур. Нетрудно видеть, что длина простого контура не превышает числа вершин n. Но тогда в силу утверждения 4.5 для любой вершины v_i, принадлежащей некоторому простому контуру длины l, где $2 \leqslant l \leqslant n$, элемент $a_{ii}^{(l)}$ матрица A^l отличен от нуля, а следовательно, и элемент k_{ii} матрицы K отличен от нуля.

Замечание 4.9. В случае ориентированного n-вершинного псевдографа D для существования в D контура необходимо и достаточно, чтобы матрица $K = A + A^2 + \ldots + A^n$ имела ненулевые диагональные элементы. Доказательство аналогично.

4.1.5. Булевы матрицы. Операции над булевыми матрицами

Будем $(m \times n)$-матрицу $C = [c_{ij}]$, у которой $c_{ij} \in \{0, 1\}$, $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n$, называть *булевой матрицей*.

Заметим, что в случае, когда G — псевдограф без кратных ребер, матрица смежности $A(G)$ состоит из нулей и единиц, т. е. является булевой (то же имеет место и для ориентированного псевдографа D без кратных дуг). У псевдографа G, кроме того, булевой является и матрица $B(G)$.

Над булевыми матрицами одинаковой размерности будем производить обычные логические операции. Например, если $C = [c_{ij}]$, $D = [d_{ij}]$ — булевы $(m \times n)$-матрицы, то $F = [f_{ij}] = C \vee D$ есть булева $(m \times n)$-матрица, у которой $f_{ij} = c_{ij} \vee d_{ij}$, $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n$. Кроме того, введем операцию $*$ логического умножения булевых матриц. Пусть $C = [c_{ij}]$ — булева $(m \times k)$-матрица и $D = [d_{ij}]$ — булева $(k \times n)$-матрица. Тогда

$F=[f_{ij}]=C*D$ — булева $(m\times n)$-матрица, у которой $f_{ij}=\bigvee_{r=1}^{k}(c_{ir}\&d_{rj})$, $i=1, 2, ..., m$, $j=1, 2, ..., n$. Если $D=C_1*C_2*\ ...*C_k$ и $C_1=C_2=\ ...\ =C_k=C$, где C — квадратная булева матрица, то будем писать $D=C^k$.

Введем теперь операцию sign перехода от произвольной $(m\times n)$-матрицы $D=[d_{ij}]$ с неотрицательными элементами к булевой $(m\times n)$-матрице $C=[c_{ij}]=\operatorname{sign} D$, у которой $c_{ij}=\operatorname{sign} d_{ij}$, $i=1, 2, ..., m$, $j=1, 2, ..., n$, где для любого числа $t\geqslant 0$

$$\operatorname{sign} t=\begin{cases}1, \text{ если } t>0;\\ 0, \text{ если } t=0.\end{cases}$$

Нетрудно показать, что для любых матриц D_1, D_2 (подходящих размерностей) с неотрицательными элементами выполняются равенства

$$\operatorname{sign}(D_1+D_2)=\operatorname{sign} D_1\vee\operatorname{sign} D_2,\ \operatorname{sign}(D_1D_2)=\operatorname{sign} D_1*\operatorname{sign} D_2. \tag{4.9}$$

Отметим, что булевы матрицы более экономичны в вычислительном отношении, чем целочисленные. Действительно, запоминание булевой матрицы требует меньшего объема оперативной памяти ЭВМ, чем целочисленной матрицы той же размерности. Кроме того, выполнение на ЭВМ логических операций над булевыми матрицами требует меньшего объема вычислений, чем над целочисленными матрицами тех же размерностей. В связи с этим представляют интерес методы решения задач теории графов, основанные на выполнении логических операций над матрицей смежности как над булевой матрицей.

Вернемся к рассмотренной ранее задаче о выяснении наличия контуров в орграфе. Из утверждения 4.6, используя (4.9), получаем, что справедливо

Утверждение 4.7. *Для того чтобы n-вершинный орграф D с матрицей смежности $A=A(D)$ имел хотя бы один контур, необходимо и достаточно, чтобы матрица $A^2\vee A^3\vee\ ...\ \vee A^n$ имела ненулевые диагональные элементы.*

4.1.6. Объединение, пересечение графов. Подграфы

Объединением графов $G_1=(V_1, X_1)$, $G_2=(V_2, X_2)$ называется граф $G_1\cup G_2=(V_1\cup V_2, X_1\cup X_2)$.

Пересечением графов $G_1=(V_1, X_1)$, $G_2=(V_2, X_2)$, где $V_1\cap V_2\neq\varnothing$, называется граф $G_1\cap G_2=(V_1\cap V_2, X_1\cap X_2)$.

Подграфом графа G называется граф, все вершины и ребра которого содержатся среди вершин и ребер графа G. Подграф называется *собственным*, если он отличен от самого графа. Подграфом графа $G=(V, X)$, *порожденным* подмножеством $V_1\subseteq V$, где $V_1\neq\varnothing$, называется граф $G_1=(V_1, X_1)$, множество X_1 ребер

которого состоит из тех и только тех ребер графа G, оба конца которых лежат в V_1.

Все приведенные определения распространяются на орграфы.

Для дальнейших рассуждений понадобится следующее простое утверждение.

Утверждение 4.8. *Пусть $G=(V, X)$ — некоторый граф, $V_1 \subseteq V$, $V_1 \neq \varnothing$, G_1 — подграф графа G, порожденный множеством V_1. Тогда $A(G_1)$ является подматрицей матрицы $A(G)$, находящейся на пересечении строк и столбцов, соответствующих вершинам из V_1.*

Замечание 4.10. Аналогичное утверждение справедливо и для орграфов. Более того, оно остается справедливым и для произвольных псевдографов (ориентированных и неориентированных).

4.1.7. Связность. Компоненты связности

Говорят, что вершина w орграфа D (графа G) *достижима* из вершины v, если либо $w=v$, либо существует путь из v в w (маршрут, соединяющий v, w).

Граф (орграф) называется *связным (сильно связным)*, если для любых двух его вершин v, w существует маршрут (путь), соединяющий v, w (из v в w).

Орграф называется *односторонне связным*, если для любых двух его вершин по крайней мере одна достижима из другой.

Псевдографом, *ассоциированным* с ориентированным псевдографом $D=(V, X)$, называется псевдограф $G=(V, X_0)$, в котором X_0 получается из X заменой всех упорядоченных пар (v, w) на неупорядоченные $\{v, w\}$ (см. рис. 4.9, где *а* — ориентированный псевдограф, *б* — ассоциированный с ним псевдограф).

Орграф называется *слабо связным*, если связным является ассоциированный с ним псевдограф.

Если граф (орграф) не является связным (слабо связным), то он называется *несвязным*.

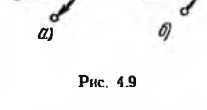


Рис. 4.9

Компонентой связности (сильной связности) графа G (орграфа D) называется его связный (сильно связный) подграф, не являющийся собственным подграфом никакого другого связного (сильно связного) подграфа графа G (орграфа D).

Пример 4.14. У графа, изображенного на рис. 4.10, три компоненты связности.

Пример 4.15. У орграфа, изображенного на рис. 4.11, три компоненты сильной связности, показанные на рис. 4.12, *а* — *в*.

Из определения компоненты связности (сильной связности) заключаем, что справедливо

Утверждение 4.9.

1. *Пусть* $G_1=(V_1, X_1)$ — *компонента связности графа* G. *Тогда* G_1 — *подграф графа* G, *порожденный множеством* V_1.

2. *Пусть* $D_1=(V_1, X_1)$ — *компонента сильной связности орграфа* D. *Тогда* D_1 — *подграф орграфа* D, *порожденный множеством* V_1.

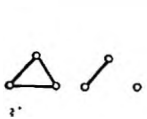


Рис. 4.10

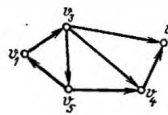


Рис. 4.11

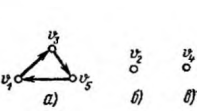


Рис. 4.12

Замечание 4.11. Утверждение 4.9 остается в силе и для произвольных псевдографов (ориентированных и неориентированных).

Нетрудно показать, что справедливы следующие утверждения.

Утверждение 4.10. *Пусть* $G=(V, X)$ — *псевдограф с* p *компонентами связности:* $G_1=(V_1, X_1)$, ..., $G_p=(V_p, X_p)$. *Тогда*

1) $V=V_1\cup...\cup V_p$, $X=X_1\cup...\cup X_p$, *т. е.* $G=G_1\cup...\cup G_p$;
2) $V_i\cap V_j=\varnothing$, $X_i\cap X_j=\varnothing$ *при* $i\neq j$;
3) $n(G_1)+...+n(G_p)=n(G)$, $m(G_1)+...+m(G_p)=m(G)$.

Утверждение 4.11. *Пусть* $D=(V, X)$ — *ориентированный псевдограф с* p *компонентами сильной связности:* $D_1=(V_1, X_1)$, ..., $D_p=(V_p, X_p)$. *Тогда*

1) $V=V_1\cup...\cup V_p$, $X\supseteq X_1\cup...\cup X_p$;
2) $V_i\cap V_j=\varnothing$, $X_i\cap X_j=\varnothing$ *при* $i\neq j$;
3) $n(D_1)+...+n(D_p)=n(D)$, $m(D_1)+...+m(D_p)\leqslant m(D)$.

Утверждение 4.12. *Пусть* ρ — *отношение достижимости на множестве* V *вершин псевдографа* G, *т. е.* $v\rho w$ *тогда и только тогда, когда либо* $v=w$, *либо существует маршрут, соединяющий* v, w. *Тогда:*

1) ρ — *эквивалентность на* V;

2) $v\rho w$ *тогда и только тогда, когда вершины* v, w *принадлежат одной компоненте связности псевдографа* G;

3) *для любого класса эквивалентности* $V_1\in V/\rho$ *псевдограф* G_1, *порожденный множеством* V_1, *является компонентой связности псевдографа* G;

4) *для любой компоненты связности* $G_1=(V_1, X_1)$ *псевдографа* G *выполняется* $V_1\in V/\rho$.

Утверждение 4.13. *Пусть* ρ_1 — *отношение достижимости на множестве* V *вершин ориентированного псевдографа* D, *т. е.* $v\rho_1 w$ *тогда и только тогда, когда вершина* w *достижима из* v.

Пусть также ρ_2 — *отношение двусторонней достижимости на* V, *т. е.* $\rho_2=\rho_1\cap\rho_1^{-1}$. *Тогда:*

1) ρ_1 *рефлексивно, транзитивно;*

2) ρ_2 — *эквивалентность на* V;

3) $v\rho_2 w$ *тогда и только тогда, когда вершины* v, w *принадлежат одной компоненте сильной связности ориентированного псевдографа* D;

4) *для любого класса эквивалентности* $V_1\in V/\rho_2$ *ориентированный псевдограф* D_1, *порожденный множеством* V_1, *является компонентой сильной связности ориентированного псевдографа* D;

5) *для любой компоненты сильной связности* $D_1=(V_1, X_1)$ *ориентированного псевдографа* D *выполняется* $V_1\in V/\rho_2$.

В дальнейшем количество компонент связности графа G будем обозначать через $p(G)$. Аналогично через $p(D)$ будем обозначать количество компонент сильной связности орграфа D.

Под операцией *удаления вершины* из графа (орграфа) будем понимать операцию, заключающуюся в удалении некоторой вершины вместе с инцидентными ей ребрами (дугами).

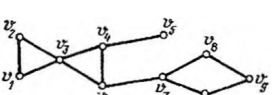


Рис. 4.13

Вершина графа, удаление которой увеличивает число компонент связности, называется *разделяющей* (или *точкой сочленения*).

Пример 4.16. Точками сочленения графа, изображенного на рис. 4.13, являются вершины v_3, v_4, v_6, v_7.

Следующее утверждение очевидно.

Утверждение 4.14. *Если* D' — *орграф, полученный в результате удаления нескольких вершин из орграфа* D, *то* $A(D')$ *получается из* $A(D)$ *в результате удаления строк и столбцов, соответствующих удаленным вершинам.*

Замечание 4.12. Аналогичное утверждение справедливо и для произвольных псевдографов (ориентированных и неориентированных).

4.1.8. Матрицы связности

Пусть $D=(V, X)$ — орграф, где $V=\{v_1, \ldots, v_n\}$. *Матрицей достижимости* орграфа D называется квадратная матрица $T(D)=[t_{ij}]$ порядка n, у которой $t_{ij}=1$, если вершина v_j достижима из v_i, и $t_{ij}=0$ — в противном случае. *Матрицей сильной связности* орграфа D называется квадратная матрица $S(D)=[s_{ij}]$ порядка n, у которой $s_{ij}=1$, если вершина v_i достижима из v_j и одновременно v_j достижима из v_i, и $s_{ij}=0$ — в противном случае (т. е. $s_{ij}=1$ тогда и только тогда, когда вершины v_i, v_j принадлежат одной компоненте сильной связности орграфа D; см. утверждение 4.13, п. 3).

Пусть $G=(V, X)$ — граф, где $V=\{v_1, \ldots, v_n\}$. *Матрицей связности* графа G называется квадратная матрица $S(G)=[s_{ij}]$ порядка n, у которой $s_{ij}=1$, если $i=j$ или существует маршрут, соединяющий v_i, v_j, и $s_{ij}=0$ — в противном случае (т. е. $s_{ij}=1$ тогда и только тогда, когда вершины v_i, v_j принадлежат одной компоненте связности графа G; см. утверждение 4.12, п. 2).

Воспользовавшись утверждением 4.5, равенствами (4.9), а также тем фактом, что в силу утверждения 4.4 из любого незамкнутого маршрута или пути можно выделить простую цепь с теми же начальной и конечной вершинами, получим справедливость следующих утверждений.

Утверждение 4.15. *Пусть $G=(V, X)$, где $V=\{v_1, \ldots, v_n\}$, — граф с матрицей смежности $A=A(G)$. Тогда*

$$S(G)=\operatorname{sign}(E+A+A^2+\ldots+A^{n-1})=E\vee A\vee A^2\vee\ldots\vee A^{n-1},$$

где E — единичная матрица порядка n.

Утверждение 4.16. *Пусть $D=(V, X)$, где $V=\{v_1, \ldots, v_n\}$, — орграф с матрицей смежности $A=A(D)$. Тогда:*

1) $T(D)=\operatorname{sign}(E+A+A^2+\ldots+A^{n-1})=$
$=E\vee A\vee A^2\vee\ldots\vee A^{n-1}$;

2) $S(D)=T(D)\&[T(D)]^{\tau}$,

где τ — обозначение операции транспонирования матрицы.

Утверждения 4.15 и 4.16 дают простые, легко реализуемые на ЭВМ методы вычисления матриц $S(G)$, $T(D)$, $S(D)$. Существуют и более экономичные методы вычисления этих матриц. Опишем, например, метод Уоршелла, основанный на следующем утверждении.

Утверждение 4.17. *Пусть A — матрица смежности графа $G=(V, X)$ (орграфа $D=(V, X)$), где $V=\{v_1, \ldots, v_n\}$. Рассмотрим последовательность булевых квадратных матриц $B^{(l)}$ порядка n, где $l=0, 1, \ldots, n$, $B^{(0)}=A\vee E$, элементы $b^{(l)}_{ij}$ которых вычисляются по следующей итерационной формуле:*

$$b^{(l)}_{ij}=b^{(l-1)}_{ij}\vee(b^{(l-1)}_{il}\&b^{(l-1)}_{lj}),$$

где $l=1, 2, \ldots, n$. Тогда $S(G)=B^{(n)}$ (и соответственно $T(D)=B^{(n)}$, $S(D)=T(D)\&[T(D)]^{\tau}$).

Доказательство будем проводить для G (для D оно аналогично). Покажем индукцией по l, что $b^{(l)}_{ij}=1$ тогда и только тогда, когда либо $i=j$, либо существует маршрут, соединяющий v_i, v_j, внутренние вершины которого принадлежат множеству $\{v_1, \ldots, v_l\}$. Из этого утверждения при $l=n$ следует справедливость утверждения 4.17. При $l=0$ элементы $b^{(0)}_{ij}$ матрицы $B^{(0)}=A\vee E$ очевидным образом удовлетворяют требуемому условию. Предположим, что при некотором l, где $1\leqslant l\leqslant n$, элементы $b^{(l-1)}_{ij}$ также удовлетворяют требуемому условию. Покажем

выполнение этого условия и для элементов $b^{(0)}_{ij}$. Пусть $i \neq j$ (случай $i=j$ очевиден) и существует маршрут, соединяющий v_i, v_j, внутренние вершины которого принадлежат множеству $\{v_1, ..., v_l\}$. Докажем, что тогда $b^{(l)}_{ij}=1$. Пусть η — один из таких маршрутов. Будем считать, что η — простая цепь (иначе, следуя утверждению 4.4, выделим из η простую цепь, соединяющую v_i, v_j). Если v_l не является внутренней вершиной цепи η, то в силу индуктивного предположения $b^{(l-1)}_{ij}=1$, откуда $b^{(l)}_{ij}=1 \vee (b^{(l-1)}_{il} \& b^{(l-1)}_{lj})=1$. Пусть теперь v_l является внутренней вершиной цепи η, т. е. цепь η имеет вид $\eta=\eta_1 \circ \eta_2$, где η_1, η_2 — простые цепи, соединяющие вершины v_i, v_l и v_l, v_j соответственно, внутренние вершины которых принадлежат множеству $\{v_1, ..., v_{l-1}\}$. Но тогда в силу индуктивного предположения $b^{(l-1)}_{il}=b^{(l-1)}_{lj}=1$, откуда $b^{(l)}_{ij}=b^{(l-1)}_{ij} \vee (1 \& 1)=1$.

Пусть теперь $b^{(l)}_{ij}=1$, $i \neq j$. Покажем, что существует маршрут, соединяющий v_i, v_j, внутренние вершины которого принадлежат множеству $\{v_1, ..., v_l\}$. В силу того, что $b^{(l)}_{ij}=b^{(l-1)}_{ij} \vee \vee (b^{(l-1)}_{il} \& b^{(l-1)}_{lj})=1$, выполняется либо $b^{(l-1)}_{ij}=1$, либо $b^{(l-1)}_{ij}=0$, $b^{(l-1)}_{il}=b^{(l-1)}_{lj}=1$. Если $b^{(l-1)}_{ij}=1$, то в силу индуктивного предположения существует маршрут, соединяющий v_i, v_j, внутренние вершины которого принадлежат множеству $\{v_1, ..., v_{l-1}\}$. Если $b^{(l-1)}_{ij}=0$, $b^{(l-1)}_{il}=b^{(l-1)}_{lj}=1$, то в силу индуктивного предположения существуют маршруты η_1, η_2, соединяющие v_i, v_l и v_l, v_j соответственно, внутренние вершины которых принадлежат множеству $\{v_1, ..., v_{l-1}\}$. Но тогда маршрут $\eta_1 \circ \eta_2$ и будет искомым.

Замечание 4.13. Если в утверждении 4.17 заменить $B^{(0)}=A \vee E$ на $B^{(0)}=E \vee \operatorname{sign} A$, то оно останется справедливым и для произвольных псевдографов (ориентированных и неориентированных).

4.1.9. Выделение компонент связности

Опишем алгоритм нахождения числа компонент сильной связности орграфа, а также выделения этих компонент. Аналогичным образом решается задача нахождения количества компонент связности, а также выделения компонент связности неориентированного графа. Однако для определенности приводим рассуждения для орграфа.

Воспользуемся следующими утверждениями.

Утверждение 4.18. *Пусть D — орграф с $p \geqslant 2$ компонентами сильной связности: $D_1, ..., D_p$. Тогда в результате удаления из D вершин, содержащихся в D_1, получаем орграф с $p-1$ компонентами сильной связности: $D_2, ..., D_p$.*

Воспользуемся тем очевидным фактом, что если D' — компонента сильной связности орграфа D, то D' является компонентой сильной связности и любого подграфа орграфа D, содержащего все вершины и дуги орграфа D'. Используя утверждение

4.11, п. 2, заключаем, что после удаления из D вершин, содержащихся в D_1, имеем орграф D, подграфами которого являются D_2, ..., D_p, а следовательно, D_2, ..., D_p являются компонентами сильной связности орграфа D. Кроме того, в силу утверждения 4.11, пп. 1, 2, получаем, что объединение множеств вершин орграфов D_2, ..., D_p дает множество вершин орграфа D, а значит, D_2, ..., D_p — все компоненты сильной связности орграфа D.

Утверждение 4.19. *Пусть D' — компонента сильной связности орграфа D. Пусть также $p(D) \geqslant 2$ и D'' — орграф, получаемый в результате удаления из D вершин, содержащихся в D'. Тогда матрицами $A(D'')$, $S(D'')$ являются подматрицы матриц $A(D)$, $S(D)$, получаемые в результате удаления из них строк и столбцов, соответствующих вершинам орграфа D'.*

Утверждение 4.19 является следствием утверждений 4.18 и 4.14.

Из определения матрицы сильной связности вытекает, что справедливо

Утверждение 4.20. *Единицы i-й строки или i-го столбца матрицы сильной связности орграфа $D=(V, X)$, где $V=\{v_1, ..., v_n\}$, соответствуют вершинам компоненты сильной связности орграфа D, содержащей вершину v_i.*

Из утверждений 4.18—4.20 следует справедливость алгоритма определения числа компонент сильной связности орграфа D, а также матриц смежности этих компонент.

Алгоритм 4.1:

Шаг 1. Полагаем $p=1$, $S_1=S(D)$.

Шаг 2. Включаем в множество вершин V_p очередной компоненты сильной связности D_p орграфа D вершины, соответствующие единицам первой строки матрицы S_p. В качестве $A(D_p)$ берем подматрицу матрицы $A(D)$, находящуюся на пересечении строк и столбцов, соответствующих вершинам из V_p.

Шаг 3. Вычеркиваем из S_p строки и столбцы, соответствующие вершинам из V_p. Если в результате такого вычеркивания не остается ни одной строки (и соответственно ни одного столбца), то p — количество компонент сильной связности и $A(D_1)$, ..., $A(D_p)$ — матрицы смежности компонент сильной связности D_1, ..., D_p орграфа D. В противном случае обозначаем оставшуюся после вычеркивания из S_p соответствующих строк и столбцов матрицу через S_{p+1}, присваиваем $p:=p+1$ и переходим к шагу 2.

Замечание 4.14. После изменений в обозначениях и терминологии алгоритм 4.1 можно применить для определения числа компонент связности графа G, а также матриц смежности этих компонент. Для обоснования этого достаточно воспользоваться утверждениями, аналогичными утверждениям 4.18—4.20, но сформулированными для неориентированного графа G. Более того, алгоритм 4.1 остается справедливым и для произвольных

псевдографов (ориентированных и неориентированных). Доказательство аналогично.

Задачи и упражнения

1. Показать, что в любом графе количество вершин нечетной степени четно.

2. Показать, что из всякого замкнутого маршрута нечетной длины можно выделить простую цепь.

3. Показать, что ребро, входящее в цикл графа, входит в некоторый его простой цикл.

4. Показать, что любая вершина, входящая в цикл, не является висячей.

5. Доказать, что в связном графе, содержащем, по крайней мере, две вершины, найдется вершина, не являющаяся точкой сочленения.

6. Доказать, что если в орграфе D отсутствуют вершины с нулевой полустепенью исхода (захода), то в D имеется простой контур.

7. Доказать, что удаление из орграфа вершины v с $\delta^+(v) \leqslant 1$ ($\delta^-(v) \leqslant 1$) приводит к орграфу, контуры которого совпадают с контурами исходного орграфа.

8. Определить, имеют ли контуры орграфы с матрицами смежности:

а) $\begin{bmatrix} 0111 \\ 0000 \\ 0101 \\ 0110 \end{bmatrix}$; б) $\begin{bmatrix} 0110 \\ 0000 \\ 0101 \\ 1100 \end{bmatrix}$; в) $\begin{bmatrix} 0101 \\ 0000 \\ 1101 \\ 0100 \end{bmatrix}$; г) $\begin{bmatrix} 0010 \\ 0001 \\ 0100 \\ 1000 \end{bmatrix}$; д) $\begin{bmatrix} 0010 \\ 0011 \\ 0000 \\ 1010 \end{bmatrix}$.

9. Определить матрицы достижимости и сильной связности для орграфов с матрицами смежности из задачи 8.

10. Пусть орграф D задан матрицей смежности. Определить матрицу сильной связности $S(D)$. Используя алгоритм 4.1, найти количество компонент сильной связности орграфа D и определить матрицы смежности этих компонент. Построить изображения орграфа D и его компонент сильной связности. Рассмотреть случаи:

а) $\begin{bmatrix} 00100 \\ 00000 \\ 01011 \\ 01000 \\ 10010 \end{bmatrix}$; б) $\begin{bmatrix} 000011 \\ 001100 \\ 000000 \\ 011000 \\ 010001 \\ 111000 \end{bmatrix}$; в) $\begin{bmatrix} 000001 \\ 101011 \\ 100000 \\ 001001 \\ 011100 \\ 001000 \end{bmatrix}$.

4.2. ЗАДАЧИ ПОИСКА МАРШРУТОВ (ПУТЕЙ) В ГРАФЕ (ОРГРАФЕ)

4.2.1. Поиск маршрута в графе

При решении некоторых прикладных задач нередко возникает необходимость найти маршрут, соединяющий заданные вершины в графе G. Приведем алгоритм решения этой задачи.

Алгоритм 4.2 (алгоритм Тэрри) поиска маршрута в связном графе $G=(V, X)$, соединяющего заданные вершины v, $w \in V$, где $v \neq w$.

Если, исходя из вершины v и осуществляя последовательный переход от каждой достигнутой вершины к смежной ей вершине, руководствоваться следующими правилами:

1) идя по произвольному ребру, всякий раз отмечать направление, в котором оно было пройдено;

2) исходя из некоторой вершины v', всегда следовать только по тому ребру, которое не было пройдено или было пройдено в противоположном направлении;

3) для всякой вершины v', отличной от v, отмечать первое заходящее в v' ребро, если вершина v' встречается в первый раз;

4) исходя из некоторой вершины v', отличной от v, по первому заходящему в v' ребру идти лишь тогда, когда нет других возможностей, то всегда можно найти маршрут в связном графе G, соединяющий две заданные вершины v, w.

Пример 4.17. Используя алгоритм Тэрри, найти маршрут, соединяющий v_1, v_5, в графе G, изображенном на рис. 4.14.

Поиск вершины v_5 в G будем осуществлять так, как будто мы ничего не знаем об этом графе (можно себе представить, что граф G — это схема лабиринта, где v_5 — выход из него, а v_1 — развилка, из которой мы начинаем поиск выхода). На рис. 4.15 показан один из возможных вариантов движения по графу G согласно алгоритму Тэрри. Пронумерованными пунктирными

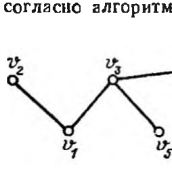


Рис. 4.14

Рис. 4.15

дугами показана схема движения по графу G. Знаками $\times$ помечены первые заходящие в вершины ребра (пометка делается ближе к той вершине, в которую ребро заходит). Указанная на рис. 4.15 схема движения соответствует маршруту $v_1v_2v_1v_3v_4v_3v_5$.

Отметим, что после того, как из вершины v_1 зашли в вершину v_3 (см. дугу *3*), в силу правила 4 мы не можем вернуться в v_1, так как имеются другие возможности, а $\{v_1, v_3\}$ является первым заходящим в v_3 ребром. Далее, после того, как из вершины v_4 зашли в вершину v_3 (см. дугу *5*), в силу правила 2 мы не можем вернуться в вершину v_4, а в силу правила 4 не можем идти к вершине v_1, и тем самым остается единственная возможность — идти к вершине v_5.

Обоснование алгоритма Тэрри. Допустим, что, руководствуясь этим алгоритмом, мы остановимся в некоторой вершине u (не достигнув вершины w), и все ребра, инцидентные u, уже пройдены в направлении из u (тогда в силу правила 2 мы уже не сможем выйти из u). Покажем, что в этом случае: а) вершина u совпадает с v; б) все вершины графа G являются пройденными.

Докажем сначала справедливость утверждения «а». Если u не совпадает с v, то пусть в вершине u мы побывали k раз (включая последний). Тогда ребра, инцидентные u, были пройдены k раз по направлению к u и $k-1$ раз в направлении из u (так как число заходов в u, за исключением последнего, соответствует числу исходов из этой вершины). Таким образом, используя то, что по предположению были пройдены все ребра, инцидентные u, в направлении из u, а также то, что в силу правила 2 по каждому ребру, инцидентному u, разрешается идти не более одного раза в направлении из u, имеем $\delta(u) = k-1$, а это противоречит тому, что по направлению к u было пройдено k различных (см. снова правило 2) ребер, а следовательно, $\delta(u) \geqslant k$. Полученное противоречие подтверждает, что $u=v$.

Докажем теперь справедливость утверждения «б». Пусть (см. утверждение «а»)

$$v_1 v_2 \ldots v_k, \text{ где } v_1 = v_k = v, \tag{4.10}$$

есть последовательность вершин, расположенных в том же порядке, в каком мы их проходили, действуя согласно алгоритму. Очевидно, что (4.10) является маршрутом в графе G (точнее, сокращенной записью маршрута). Покажем, что маршрут (4.10) содержит все вершины графа G. Предварительно докажем, что каждое ребро, инцидентное любой вершине v_j, где $1 \leqslant j \leqslant k$, было пройдено по одному разу в обоих направлениях. Доказательство проведем индукцией по j.

Базис индукции. Поскольку в замкнутом маршруте для каждой содержащейся в нем вершины число исходов из этой вершины равно числу заходов в нее, то в силу того, что согласно утверждению «а» и правилу 2 все ребра, инцидентные вершине $v = v_1$, были пройдены по разу в направлении из v (т. е. мы $\delta(v)$ раз исходили из v), получаем, что ровно $\delta(v)$ раз мы заходили в v, а поскольку в силу правила 2 каждый раз мы заходили в v по новому ребру, то в результате все ребра, инцидент-

ные вершине $v_1=v$, были пройдены по разу в обоих направлениях.

Индуктивный шаг. Допустим, что при некотором j, где $2\leqslant j\leqslant k$, доказываемое утверждение верно для всех вершин v_1, ..., v_{j-1}. Докажем его для вершины v_j. Если при некотором $i<j$ выполняется $v_i=v_j$, то справедливость доказываемого утверждения для вершины v_j вытекает из того, что по индуктивному предположению оно верно для вершины v_i. Пусть теперь $\forall i\in\{1, 2, ..., j-1\}$ $v_i\neq v_j$, т. е. вершина v_j встретилась в первый раз. Тогда $\{v_{j-1}, v_j\}$ — первое заходящее в вершину v_j ребро, и по индуктивному предположению оно будет пройдено в обоих направлениях, что в силу правила 4 возможно лишь в случае, когда все остальные ребра, инцидентные v_j, будут пройдены в направлении из v_j. Далее, поскольку в замкнутом маршруте, как уже отмечалось ранее, для каждой вершины, содержащейся в этом маршруте, число исходов из этой вершины равно числу заходов в нее, то, используя правило 2, получаем, что все ребра, инцидентные v_j, будут пройдены по разу в обоих направлениях.

Итак, каждую вершину в маршруте (4.10) мы проходим вместе со всеми смежными ей вершинами, откуда в силу связности графа G следует, что маршрут (4.10) проходит через все вершины графа G, а это противоречит исходному предположению, что вершина w не была достигнута.

Замечание 4.15. Алгоритм 4.2 и его обоснование остаются в силе и для случая, когда G — связный псевдограф.

Замечание 4.16. Если псевдограф $G=(V, X)$ не является связным, то с помощью алгоритма 4.2, исходя из произвольной вершины $v\subset V$ и помечая пройденные вершины и ребра, можно выделить компоненту связности псевдографа G, содержащую вершину v. Алгоритм закончит свою работу в тот момент, когда в первый раз невозможно будет удовлетворить правилу 2 (т. е. мы пришли в вершину u, и все ребра, инцидентные этой вершине, пройдены в направлении из u; при этом, как показано при обосновании алгоритма 4.2, $u=v$).

Замечание 4.17. Из полученного с помощью алгоритма 4.2 маршрута всегда можно выделить простую цепь, соединяющую v, w (см. утверждение 4.4).

4.2.2. Поиск путей (маршрутов) с минимальным числом дуг (ребер)

Путь в орграфе D из вершины v в вершину w, где $v\neq w$, называется *минимальным*, если он имеет минимальную длину среди всех путей орграфа D из v в w. Аналогично определяется и минимальный маршрут в графе G.

Рассмотрим некоторые свойства минимальных путей (маршрутов).

Утверждение 4.21. *Любой минимальный путь (маршрут) является простой цепью.*

Доказательство проведем для пути (для маршрута оно аналогично). Предположим, что в некотором орграфе D нашелся минимальный путь $\pi=v_1v_2...v_k$, где $v_i\neq v_k$, не являющийся простой цепью. Тогда найдутся номера i, j такие, что $1\leqslant i<j\leqslant k$ и $v_i=v_j$. Пусть $i>1$, $j<k$. Рассмотрим путь $v_1...v_iv_{j+1}...v_k$. Его длина равна $(i-1)+(k-j)=k+i-j-1<k-1$, что противоречит минимальности π. Случаи $i=1$ или $j=k$ доказываются аналогично (случая $i=1$, $j=k$ быть не может в силу $v_1\neq v_k$).

Утверждение 4.22. *(о минимальности подпути минимального пути). Пусть $\pi=v_1v_2...v_k$, где $v_1\neq v_k$ — минимальный путь (маршрут) в орграфе D (в графе G). Тогда для любых номеров i, j таких, что $1\leqslant i<j\leqslant k$, путь (маршрут) $\pi_0=v_iv_{i+1}...v_j$ также является минимальным.*

Доказательство проведем для орграфа D (для графа G оно аналогично). Заметим, что в силу $i\neq j$ выполняется $v_i\neq v_j$ (см. утверждение 4.21). Будем считать, что $i>1$, $j<k$ (случаи, когда $i=1$ или $j=k$, рассмотрите самостоятельно). Тогда $\pi=\pi_1\cup\pi_0\cup\pi_2$, где $\pi_1=v_1v_2...v_i$, $\pi_2=v_jv_{j+1}...v_k$ — пути в D. Предполагая, что путь π_0 не является минимальным, получаем, что в D существует путь $\tilde{\pi}_0$ из v_i в v_j меньшей длины. Но тогда длина пути $\tilde{\pi}=\pi_1\cup\tilde{\pi}_0\cup\pi_2$, равная сумме длин путей π_1, $\tilde{\pi}_0$, π_2, меньше длины пути π, равной сумме длин путей π_1, π_0, π_2, что противоречит минимальности π, поскольку $\tilde{\pi}$, так же как и π, является путем в D из v_1 в v_k.

Рассмотрим теперь задачу поиска минимального пути (маршрута). Введем некоторые обозначения. Пусть $D=(V, X)$ — орграф, $v\in V$, $V_1\subseteq V$. Обозначим $D(v)=\{w\in V\mid (v, w)\in X\}$ — образ вершины v; $D^{-1}(v)=\{w\in V\mid (w, v)\in X\}$ — прообраз вершины v; $D(V_1)=\bigcup\limits_{v\in V_1}D(v)$ — образ множества вершин V_1; $D^{-1}(V_1)=\bigcup\limits_{v\in V_1}D^{-1}(v)$ — прообраз множества вершин V_1. Пусть $G=(V, X)$ — граф, $v\in V$, $V_1\subseteq V$. Обозначим $G(v)=\{w\in V\mid \{v, w\}\in X\}$ — образ вершины v; $G(V_1)=\bigcup\limits_{v\in V_1}G(v)$ — образ множества вершин V_1.

Пусть $D=(V, X)$ — орграф с $n\geqslant 2$ вершинами и v, w — заданные вершины из V, где $v\neq w$. Опишем алгоритм поиска минимального пути из v в w в орграфе D *(алгоритм фронта волны)*.

Алгоритм 4.3:

Шаг 1. Помечаем вершину v индексом 0. Затем помечаем вершины, принадлежащие образу вершины v, индексом 1. Множество вершин с индексом 1 обозначаем $FW_1(v)$. Полагаем $k=1$.

Шаг 2. Если $FW_k(v)=\varnothing$ или выполняется $k=n-1$, $w\in FW_k(v)$, то вершина w не достижима из v, и работа алгорит-

ма на этом заканчивается. В противном случае переходим к шагу 3.

Шаг 3. Если $w \in FW_k(v)$, то переходим к шагу 4. В противном случае существует путь из v в w длины k, и этот путь является минимальным. Последовательность вершин

$$vw_1w_2 \ldots w_{k-1}w,$$

где

$$\begin{aligned} &w_{k-1} \in FW_{k-1}(v) \cap D^{-1}(w);\\ &w_{k-2} \in FW_{k-2}(v) \cap D^{-1}(w_{k-1});\\ &\ldots\ldots\ldots\ldots\ldots\\ &w_1 \in FW_1(v) \cap D^{-1}(w_2), \end{aligned} \tag{4.11}$$

и есть искомый минимальный путь из v в w. На этом работа алгоритма заканчивается.

Шаг 4. Помечаем индексом $k+1$ все непомеченные вершины, которые принадлежат образу множества вершин с индексом k. Множество вершин с индексом $k+1$ обозначаем $FW_{k+1}(v)$. Присваиваем $k := k+1$ и переходим к шагу 2.

Замечание 4.18. Множество $FW_k(v)$ в алгоритме 4.3 обычно называют *фронтом волны k-го уровня.*

Замечание 4.19. Вершины w_1, ..., w_{k-1} из (4.11), вообще говоря, могут быть выделены неоднозначно. Эта неоднозначность соответствует случаям, когда существует несколько различных минимальных путей из v в w. Нетрудно описать алгоритм, позволяющий находить все минимальные пути из v в w в орграфе D (опишите этот алгоритм самостоятельно).

Пример 4.18. Используя алгоритм 4.3, определим минимальный путь из v_1 в v_6 в орграфе D, заданном матрицей смежности, представленной в табл. 4.4.

Действуя согласно алгоритму 4.3, последовательно определяем $FW_1(v_1) = \{v_4, v_5\}$; $FW_2(v_1) = D(FW_1(v_1)) \setminus \{v_1, v_4, v_5\} = \{v_2, v_3\}$; $FW_3(v_1) = D(FW_2(v_1)) \setminus \{v_1, v_2, v_3, v_4, v_5\} = \{v_6\}$. Таким образом, $v_6 \in FW_3(v_1)$, а значит (см. шаг 3), существует путь из v_1 в v_6 длины 3, и этот путь является минимальным. Найдем теперь минимальный путь из v_1 в v_6. Определим множество

$$FW_2(v_1) \cap D^{-1}(v_6) = \{v_2, v_3\} \cap \{v_2, v_3\} = \{v_2, v_3\}.$$

Выберем любую вершину из найденного множества, например вершину v_3. Определим далее множество

$$FW_1(v_1) \cap D^{-1}(v_3) = \{v_4, v_5\} \cap \{v_4, v_5, v_6\} = \{v_4, v_5\}.$$

Выберем любую вершину из найденного множества, например вершину v_5. Тогда $v_1v_5v_3v_6$ — искомый минимальный путь из v_1 в v_6 (длины 3) в орграфе D.

	v_1	v_2	v_3	v_4	v_5	v_6
v_1	0	0	0	1	1	0
v_2	1	0	0	1	1	1
v_3	1	1	0	1	1	1
v_4	0	1	1	0	1	0
v_5	1	1	1	1	0	0
v_6	1	1	1	1	1	0

Обоснование алгоритма 4.3. Введем для каждого $k \in \{1, \ldots, n-1\}$ множество $W_k(v)$, состоящее из всех вершин $u \in V$, достижимых из v, и таких, что длина минимального пути из v в u равна k. Кроме того, полагаем $W_0(v) = \{v\}$. Докажем, что справедлива рекуррентная формула

$$W_{k+1}(v) = D(W_k(v)) \setminus \bigcup_{i=0}^{k} W_i(v), \quad k = 0, 1, \ldots, n-2. \qquad (4.12)$$

Пусть $k \in \{0, 1, \ldots, n-2\}$. Покажем сначала, что множество из левой части равенства (4.12) содержится в множестве из правой его части. Пусть $u \in W_{k+1}(v)$. Рассмотрим минимальный путь η из v в u в орграфе D. По определению множества $W_{k+1}(v)$ такой путь существует, и его длина равна $k+1$. Пусть v' — вершина в этом пути, непосредственно предшествующая вершине u. Тогда $v' \in W_k(v)$ (см. утверждение 4.22), откуда в силу $u \in D(v')$ получаем $u \in D(W_k(v))$. Используя теперь то, что $u \in W_{k+1}(v)$, а следовательно, $u \notin W_i(v)$, $i = 0, \ldots, k$, имеем $u \in D(W_k(v)) \setminus \bigcup_{i=0}^{k} W_i(v)$.

Покажем теперь, что множество из правой части равенства (4.12) содержится в множестве из левой его части. Пусть $u \in D(W_k(v)) \setminus \bigcup_{i=0}^{k} W_i(v)$. Тогда в силу $u \in D(W_k(v))$ вершина u достижима из v и при этом существует путь длины $k+1$ из v в u. Покажем, что этот путь является минимальным (откуда и будет следовать, что $u \in W_{k+1}(v)$). Действительно, предположив, что минимальный путь из v в u в орграфе D имеет длину j, где $1 \leqslant j \leqslant k$, получаем $u \in W_j(v)$, а это противоречит тому, что $u \notin \bigcup_{i=0}^{k} W_i(v)$.

Таким образом, формула (4.12) полностью доказана.

Обозначим далее через $W(v)$ множество всех вершин орграфа D, достижимых из v. Используя утверждение 4.21 (а также тот очевидный факт, что

длина любой простой цепи в n-вершинном орграфе D не превосходит $n-1$), получаем, что справедлива формула

$$W(v) = \bigcup_{k=0}^{n-1} W_k(v). \qquad (4.13)$$

Перейдем теперь непосредственно к обоснованию алгоритма 4.3. Прежде всего заметим, что $FW_1(v) = W_1(v)$ (см. шаг 1) и последовательное нахождение множеств $FW_k(v)$ при $k \geqslant 2$ по алгоритму 4.3 (см. шаг 4) производится аналогично последовательному построению множеств $W_k(v)$ по рекуррентной формуле (4.12), т. е. выделяемые в алгоритме 4.3 множества $FW_k(v)$ совпадают с введенными выше множествами $W_k(v)$. Но тогда из определения множеств $W_k(v)$ следует, что минимальное число k, при котором $w \in W_k(v)$, и будет длиной минимального пути из v в w в орграфе D, что соответствует шагу 3 (очевидно, что может существовать лишь единственное число k, при котором $w \in W_k(v)$). Если же $w \notin W_k(v)$, $k = 1, 2, \ldots, n-1$, то в силу (4.13) вершина w не достижима из v, что также соответствует логике алгоритма (см. шаг 2). Осталось обосновать существование вершин $w_1, \ldots, w_{k-1}$, удовлетворяющих (4.11). Эти вершины найдутся, так как в силу (4.12) выполняется $W_i(v) \subseteq D(W_{i-1}(v))$, $i = 1, 2, \ldots, n-1$, а следовательно, $\forall i \in \{1, 2, \ldots, n-1\}$, $\forall u \in W_i(v)$ найдется вершина $u' \in W_{i-1}(v)$ такая, что $(u', u) \in X$, т. е. $u' \in W_{i-1}(v) \cap D^{-1}(u)$, откуда $\forall i \in \{1, 2, \ldots, n-1\}$, $\forall u \in W_i(v)$ $W_{i-1}(v) \cap D^{-1}(u) \neq \varnothing$.

Замечание 4.20. Если выражения $D(\cdot)$, $D^{-1}(\cdot)$ заменить на $G(\cdot)$, то при соответствующем изменении терминологии алгоритм 4.3 и его обоснование переносятся и на поиск минимального маршрута в неориентированном графе G.

4.2.3. Расстояния в графе

Пусть $G = (V, X)$ — граф (в общем случае — псевдограф). Заметим, что если вершины $v, w \in V$, где $v \neq w$, можно соединить маршрутом в G, то обязательно существует минимальный маршрут, соединяющий вершины v, w. Действительно, если некоторый маршрут η_1 длины k_1 не является минимальным, то существует маршрут η_2 длины $k_2 < k_1$. Если и маршрут η_2 не является минимальным, то имеется маршрут η_3 длины $k_3 < k_2$ и т. д. Очевидно, что такое уменьшение длины маршрута можно повторить не более $k_1 - 1$ раз (так как минимальная длина маршрута равна 1). Поэтому в G обязательно найдется минимальный маршрут, соединяющий вершины v, w. Обозначим длину этого маршрута через $d(v, w)$. Положим также $d(v, v) = 0$ для любой вершины $v \in V$. Кроме того, пусть $\forall v, w \in V$ $d(v, w) = +\infty$ (далее для краткости вместо $+\infty$ будем писать ∞), если $v \neq w$ и в G не существует маршрута, соединяющего v, w. Тем самым мы

определили величину $d(v, w)$ для любых вершин $v, w \in V$. Величину $d(v, w)$ (конечную или бесконечную) будем называть *расстоянием между вершинами* v, w. Расстояние $d(v, w)$ удовлетворяет аксиомам метрики:

1) $d(v, w) \geqslant 0$, причем $d(v, w) = 0$ тогда и только тогда, когда $v = w$;

2) $d(v, w) = d(w, v)$;

3) $d(v, w) \leqslant d(v, u) + d(u, w)$,

где v, u, w — произвольные вершины графа G. Кроме того, в связном графе G для любых его вершин v, w выполняется

4) $d(v, w) < \infty$.

Свойства 1, 2 и 4 очевидны. Докажем справедливость свойства 3. В случае $d(v, u) = \infty$ или $d(u, w) = \infty$ свойство 3, очевидно, выполняется. Пусть теперь $d(v, u) < \infty$, $d(u, w) < \infty$, а следовательно, в G существует минимальный маршрут η_1, соединяющий v, u, и минимальный маршрут η_2, соединяющий u, w. Рассмотрим маршрут $\eta_1 \cup \eta_2$, соединяющий v, w. Его длина равна $d(v, u) + d(u, w)$, откуда и следует справедливость свойства 3.

Замечание 4.21. Используя понятие расстояния, можно теперь рассмотренные ранее множества $W_i(v)$ ввести для неориентированного графа $G = (V, X)$ следующим образом: $W_i(v) = \{w \in V \mid d(v, w) = i\}$, где $i = 0, 1, \ldots, n(G) - 1$, $v \in V$.

4.2.4. Диаметр, радиус и центр графа

Пусть $G = (V, X)$ — связный граф (или в общем случае — псевдограф). Тогда величина $d(G) = \max\limits_{v, w \in V} d(v, w)$ является конечной и называется *диаметром* графа G.

Пусть v — произвольная вершина из V. Величина $r(v) = \max\limits_{w \in V} d(v, w)$ (очевидно, конечная) называется *максимальным удалением* (эксцентриситетом) в графе G от вершины v.

Радиусом графа G называется величина $r(G) = \min\limits_{v \in V} r(v)$.

Любая вершина $v \in W$ такая, что $r(v) = r(G)$, называется *центром* графа G.

Пример 4.19. Для графа G, изображенного на рис. 4.16, имеем: $d(G) = 3$, $r(v_1) = 3$, $r(v_2) = 2$, $r(v_3) = 2$, $r(v_4) = 2$, $r(v_5) = 3$, $r(G) = 2$; v_2, v_3, v_4 — центры графа G.

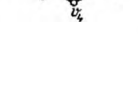


Рис. 4.16

4.2.5. Минимальные пути (маршруты) в нагруженных орграфах (графах)

Назовем орграф $D = (V, X)$ *нагруженным*, если на множестве дуг X определена некоторая функция $l: X \to \mathbf{R}$, которую часто называют *весовой* функцией. Тем самым в нагруженном орграфе D каждой дуге $x \in X$ поставлено в соответствие некоторое действительное число $l(x)$. Значение $l(x)$ будем называть *длиной дуги* x. Для любого пути π нагруженного орграфа D обозначим через $l(\pi)$ сумму длин входящих в π дуг, при этом каждая дуга учитывается столько раз, сколько она входит в путь. Величину $l(\pi)$ будем называть *длиной пути* π в нагруженном орграфе D. Ранее так называлось количество дуг в пути π. В связи с этим заметим, что если длины дуг выбраны равными 1, то $l(\pi)$ выражает введенную ранее длину пути π в ненагруженном орграфе. Следовательно, любой ненагруженный орграф можно считать нагруженным с длинами дуг, равными 1. Аналогично определяется и нагруженный граф, а также длина маршрута в нем.

Путь в нагруженном орграфе D из вершины v в вершину w, где $v \neq w$, называется *минимальным*, если он имеет минимальную длину среди всех путей орграфа D из v в w. Аналогично определяется и минимальный маршрут в нагруженном графе G.

Если в нагруженном орграфе D имеются замкнутые пути отрицательной длины, то для заданных вершин v, w орграфа D, где $v \neq w$, минимального пути из v в w может не быть. Действительно, если в D имеется замкнутый путь σ отрицательной длины и существует путь из v в w, проходящий хотя бы через одну вершину, содержащуюся в σ, то очевидно, что в D найдется путь π из v в w вида $\pi = \pi_1 \circ \sigma \circ \pi_2$, где π_1, π_2 — пути в D (возможно также, что либо π_1, либо π_2 является пустой последовательностью; при этом предполагаем, что $\varnothing \circ \sigma = \sigma \circ \varnothing = \sigma$). Но тогда $\pi_1 \circ \sigma \circ \sigma \circ \pi_2$, $\pi_1 \circ \sigma \circ \sigma \circ \sigma \circ \pi_2$, ... также пути в D из v в w, и длина каждого следующего пути в этой последовательности отличается от длины предыдущего на $l(\sigma) < 0$, а значит, длины путей из v в w могут принимать сколь угодно малые отрицательные значения. Аналогичная ситуация имеет место в случае, когда в нагруженном графе G вершины v, w находятся в одной компоненте связности, содержащей хотя бы одно ребро отрицательной длины. В таких случаях имеют смысл лишь задачи поиска минимальных путей (маршрутов) среди путей (маршрутов), число дуг (ребер) в которых ограничено сверху.

Приведем некоторые свойства минимальных путей (маршрутов) в нагруженном орграфе $D = (V, X)$ (графе $G = (V, X)$):

1) если $\forall x \in X \; l(x) > 0$, то любой минимальный путь (маршрут) является простой цепью;

2) если $v_1 v_2 \ldots v_k$ — минимальный путь (маршрут), то для любых номеров i, j таких, что $1 \leqslant i < j \leqslant k$, путь (маршрут) $v_i v_{i+1} \ldots v_j$ также является минимальным;

3) если $v...uw$ — минимальный путь (маршрут) среди путей из v в w (среди маршрутов, соединяющих v, w), содержащих не более $k+1$ дуг (ребер), то $v...u$ — минимальный путь (маршрут) среди путей из v в u (среди маршрутов, соединяющих v, u), содержащих не более k дуг (ребер).

Свойства 1—3 доказываются аналогично утверждениям 4.21 и 4.22.

Рассмотрим теперь задачу поиска минимальных путей (маршрутов) в нагруженном орграфе (графе). При этом для определенности рассуждения будем проводить для орграфа (для графа они аналогичны).

Замечание 4.22. При решении некоторых практических задач возникает необходимость поиска максимальных путей в нагруженном орграфе. Такая задача легко сводится к исследуемой ниже задаче поиска минимальных путей заменой знаков при длинах дуг на противоположные.

Пусть $D = (V, X)$ — нагруженный орграф, $V = \{v_1, ..., v_n\}$, $n \geqslant 2$. Введем величины $\lambda_i^{(k)}$, где $i=1, ..., n$, $k=1, 2, ...$ Для каждых фиксированных i и k величина $\lambda_i^{(k)}$ равна длине минимального пути среди путей из v_1 в v_i, содержащих не более k дуг; если же таких путей нет, то $\lambda_i^{(k)}=\infty$. Кроме того, если произвольную вершину $v \in V$ считать путем из v в v нулевой длины, то величины $\lambda_i^{(k)}$ можно ввести также и для $k=0$, при этом

$$\lambda_1^{(0)}=0,\ \lambda_i^{(0)}=\infty,\ i=2, ..., n. \tag{4.14}$$

Введем также в рассмотрение квадратную матрицу $C(D) = [c_{ij}]$ порядка n с элементами

$$c_{ij} = \begin{cases} l(v_i, v_j), & \text{если } (v_i, v_j) \in X; \\ \infty, & \text{если } (v_i, v_j) \overline{\in} X, \end{cases}$$

которую будем называть *матрицей длин дуг* нагруженного орграфа D.

Следующее утверждение дает простые формулы для вычисления величин $\lambda_i^{(k)}$.

Утверждение 4.23. *При $i = 2, ..., n$, $k \geqslant 0$ выполняется равенство*

$$\lambda_i^{(k+1)} = \min_{1 \leqslant j \leqslant n} \{\lambda_j^{(k)} + c_{ji}\}, \tag{4.15}$$

а при $i = 1$, $k \geqslant 0$ справедливо равенство

$$\lambda_1^{(k+1)} = \min\{0;\ \min_{1 \leqslant j \leqslant n} \{\lambda_j^{(k)} + c_{j1}\}\}. \tag{4.16}$$

Пусть $i \in \{2, ..., n\}$, $k \geqslant 0$. Докажем справедливость (4.15) (доказательство (4.16) проводится аналогично с учетом того, что $\lambda_1^{(0)}=0$). Обозначим правую часть (4.15) через $\widetilde{\lambda}_i^{(k+1)}$.

Покажем сначала, что

$$\lambda_i^{(k+1)} \geqslant \widetilde{\lambda}_i^{(k+1)}. \tag{4.17}$$

В случае $\lambda_i^{(k+1)} = \infty$ неравенство (4.17), очевидно, выполняется. Пусть теперь $\lambda_i^{(k+1)} < \infty$ и $\pi = v_i \dots v_k v_l$ — путь из v_i в v_l, содержащий не более $k+1$ дуг, такой, что $l(\pi) = \lambda_i^{(k+1)}$. Тогда (см. свойство 3) $\pi' = v_i \dots v_{i'}$ — минимальный путь среди путей из v_i в $v_{i'}$, содержащих не более k дуг, и следовательно,

$$\lambda_i^{(k+1)} = l(\pi) = l(\pi') + l(v_{i'}, v_l) = \lambda_{i'}^{(k)} + c_{i'l} \geqslant \min_{1 \leqslant j \leqslant n} (\lambda_j^{(k)} + c_{jl}) = \tilde{\lambda}_l^{(k+1)},$$

т. е. неравенство (4.17) выполняется и в этом случае.

Покажем далее, что

$$\lambda_i^{(k+1)} \leqslant \tilde{\lambda}_l^{(k+1)}. \tag{4.18}$$

В случае $\tilde{\lambda}_l^{(k+1)} = \infty$ неравенство (4.18), очевидно, выполняется. Пусть теперь $\tilde{\lambda}_l^{(k+1)} < \infty$ и $i' \in \{1, 2, \dots, n\}$ — номер такой, что

$$\lambda_{i'}^{(k)} + c_{i'l} = \min_{1 \leqslant j \leqslant n} (\lambda_j^{(k)} + c_{jl}) = \tilde{\lambda}_l^{(k+1)}.$$

Тогда в силу $\tilde{\lambda}_l^{(k+1)} < \infty$ имеем $\lambda_{i'}^{(k)} < \infty$, $c_{i'l} < \infty$, а следовательно, $(v_{i'}, v_l) \in X$, $c_{i'l} = l(v_{i'}, v_l)$ и существует путь π', содержащий не более k дуг, такой, что $l(\pi') = \lambda_{i'}^{(k)}$. Но тогда для пути $\pi = \pi' \cup v_{i'} v_l$, содержащего не более $k+1$ дуг, выполняется

$$\lambda_i^{(k+1)} \leqslant l(\pi) = \lambda_{i'}^{(k)} + c_{i'l} = \tilde{\lambda}_l^{(k+1)},$$

т. е. (4.18) справедливо и в этом случае.

Используя утверждение 4.23, нетрудно описать алгоритм нахождения таблицы значений величин $\lambda_i^{(k)}$ (будем записывать ее в виде матрицы, где i — номер строки, $k+1$ — номер столбца). Действительно, используя рекуррентные соотношения (4.15), (4.16) и исходя из (4.14), последовательно определяем набор величин $\lambda_1^{(k)}, \dots, \lambda_n^{(k)}$ ($(k+1)$-й столбец матрицы), начиная с $k=0$, а затем шаг за шагом увеличивая значение k до любой необходимой величины.

Будем теперь предполагать, что в D отсутствуют простые контуры отрицательной длины (ниже в утверждении 4.27 приводится простой метод проверки этого условия). Для дальнейших рассуждений нам понадобятся дополнительные утверждения.

Утверждение 4.24. *Из всякого замкнутого пути отрицательной длины можно выделить простой контур отрицательной длины.*

Утверждение 4.24 доказывается аналогично утверждению 4.3. Из утверждения 4.24 следует, что справедливо

Утверждение 4.25. *Если в нагруженном орграфе отсутствуют простые контуры отрицательной длины, то в нем нет и замкнутых путей отрицательной длины.*

Утверждение 4.26. *В нагруженном орграфе, в котором отсутствуют простые контуры отрицательной длины, можно из всякого незамкнутого пути выделить простую цепь с теми же начальной и конечной вершинами, длина которой не превышает длины исходного пути.*

Доказательство будем проводить индукцией по k — количеству дуг в пути. При $k=1$ утверждение 4.26 выполняется, поскольку всякий незамкнутый путь с одной дугой является простой цепью. Предположим, что при некотором $k\geqslant 2$ утверждение 4.26 выполняется для всякого незамкнутого пути, содержащего не более $k-1$ дуг. Покажем его справедливость и для каждого незамкнутого пути π, содержащего ровно k дуг. Пусть $\pi = u_1u_2 \ldots u_{k+1}$. Рассмотрим любые две вершины u_i, u_j, где $1\leqslant i< j\leqslant k+1$, такие, что $u_i = u_j$. Если таких вершин нет, то π — простая цепь, и тогда доказываемое утверждение справедливо. Если же указанные вершины нашлись, то рассматриваем путь $\pi' = u_1 \ldots u_{i-1}u_j \ldots u_{k+1}$ (т. е. предполагаем, что $i\geqslant 2$; случай $i=1$ разберите самостоятельно), а также путь $\pi'' = u_iu_{i+1} \ldots u_j$. Очевидно, что

$$l(\pi) = l(\pi') + l(\pi''). \tag{4.19}$$

По условиям доказываемого утверждения (см. также утверждение 4.25) выполняется $l(\pi'')>0$, откуда в силу (4.19) $l(\pi')\leqslant l(\pi)$. Заметим, что π' — путь, содержащий не более $k-1$ дуг, а следовательно, по индуктивному предположению из него можно выделить простую цепь σ из u_1 в u_{k+1} такую, что $l(\sigma)\leqslant l(\pi')\leqslant l(\pi)$.

Воспользовавшись тем, что число дуг в простой цепи не превосходит $n-1$, из утверждения 4.26 получаем, во-первых, что

$$\lambda_i^{(k)} = \lambda_i^{(n-1)},\ i=2,\ \ldots,\ n,\ \text{при всех } k\geqslant n-1. \tag{4.20}$$

Во-вторых, если $\lambda_i^{(n-1)} = \infty$ (где $i\in\{2, \ldots, n\}$), то вершина v_i не достижима из v_1, а если $\lambda_i^{(n-1)}<\infty$, то v_i достижима из v_1 и при этом $\lambda_i^{(n-1)}$ — длина минимального пути из v_1 в v_i. Таким образом, по $\lambda_i^{(n-1)}$ можно судить о достижимости вершин v_i, $i=2, \ldots, n$, из v_1, а также определять длины минимальных путей из v_1 в достижимые вершины. Учитывая (4.20), ограничимся рассмотрением величин $\lambda_i^{(k)}$ при $k=0, 1, \ldots, n-1$. Заметим, что

$$\lambda_1^{(k)}=0,\ k=1, 2, \ldots \tag{4.21}$$

Равенства (4.21) являются следствием (4.16), а также того, что в D отсутствуют замкнутые пути отрицательной длины (см. утверждение 4.25).

Замечание 4.23. При определении нагруженного орграфа мы предполагали, что величины $l(x)$, $x\in X$, конечны. Между тем величины $\lambda_i^{(k)}$ ($i=1, 2, \ldots, n$, $k=0, 1, \ldots, n-1$) можно аналогичным образом определить и для случая, когда для некоторых $x\in X$ выполняется $l(x)=\infty$. При этом сохраняет силу утверждение 4.23, а следовательно, и алгоритм построения таблицы величин $\lambda_i^{(k)}$. Отметим, однако, что теперь могут существовать вершины, достижимые из v_1, с длинами минимальных путей из v_1 в эти вершины, равными ∞, т. е. мы уже не можем судить по $\lambda_i^{(n-1)}$ о достижимости вершин v_i из v_1.

Замечание 4.24. Если при некотором k_0, где $0 \leqslant k_0 \leqslant n-2$, выполняются равенства $\lambda_i^{(k_0)} = \lambda_i^{(k_0+1)}$, $i=1, 2, \ldots, n$, то в силу (4.15), (4.16) $\forall k \geqslant k_0$ $\lambda_i^{(k)} = \lambda_i^{(k_0)}$, $i=1, 2, \ldots, n$, и, в частности, $\lambda_i^{(n-1)} = \lambda_i^{(k_0)}$, $i=1, 2, \ldots, n$, т. е. в данном случае значения $\lambda_i^{(k)}$ при $k > k_0$ не несут никакой дополнительной информации, и тогда имеет смысл оборвать процесс последовательного определения наборов величин $\lambda_1^{(k)}, \ldots, \lambda_n^{(k)}$ на значении $k=k_0$.

Приведем алгоритм, который позволяет по таблице величин $\lambda_i^{(k)}$, $i=1, 2, \ldots, n$, $k=0, 1, \ldots, n-1$, определить минимальный путь в нагруженном орграфе D из v_1 в любую достижимую вершину, причем из всех возможных путей он выделяет путь с наименьшим числом дуг.

Алгоритм 4.4 (алгоритм Форда — Беллмана) нахождения минимального пути в нагруженном орграфе D из v_1 в v_{i_1} ($i_1 \neq 1$):

Шаг 1. Пусть мы уже составили таблицу величин $\lambda_i^{(k)}$, $i=1, 2, \ldots, n$, $k=0, 1, \ldots, n-1$. Если $\lambda_{i_1}^{(n-1)} = \infty$, то вершина v_{i_1} не достижима из v_1 (предполагаем, что все величины $l(x)$, $x \in X$, конечны). В этом случае работа алгоритма заканчивается.

Шаг 2. Пусть $\lambda_{i_1}^{(n-1)} < \infty$. Тогда число $\lambda_{i_1}^{(n-1)}$ выражает длину любого минимального пути из v_1 в v_{i_1} в нагруженном орграфе D. Определим минимальное число $k_1 \geqslant 1$, при котором выполняется равенство $\lambda_{i_1}^{(k_1)} = \lambda_{i_1}^{(n-1)}$. По определению чисел $\lambda_i^{(k)}$ получаем, что k_1 — минимальное число дуг в пути среди всех минимальных путей из v_1 в v_{i_1} в нагруженном орграфе D.

Шаг 3. Последовательно определяем номера $i_2, \ldots, i_{k_1+1}$ такие, что

$$\begin{aligned}
&\lambda_{i_2}^{(k_1-1)} + c_{i_2, i_1} = \lambda_{i_1}^{(k_1)};\\
&\lambda_{i_3}^{(k_1-2)} + c_{i_3, i_2} = \lambda_{i_2}^{(k_1-1)};\\
&\ldots\ldots\ldots\ldots\\
&\lambda_{i_{k_1+1}}^{(0)} + c_{i_{k_1+1}, i_{k_1}} = \lambda_{i_{k_1}}^{(1)}
\end{aligned} \tag{4.22}$$

(эти номера найдутся в силу (4.15); докажите, что $i_2 \neq 1, \ldots, i_{k_1} \neq 1$). Из (4.22) с учетом того, что $\lambda_{i_1}^{(k_1)} = \lambda_{i_1}^{(n-1)} < \infty$, имеем $c_{i_2, i_1} < \infty, \ldots, c_{i_{k_1+1}, i_{k_1}} < \infty$, $\lambda_{i_{k_1+1}}^{(0)} < \infty$, откуда, используя (4.14), получаем

$$(v_{i_2}, v_{i_1}), \ldots, (v_{i_{k_1+1}}, v_{i_{k_1}}) \in X,\ l(v_{i_2}, v_{i_1}) = c_{i_2, i_1}, \ldots, l(v_{i_{k_1+1}}, v_{i_{k_1}}) = c_{i_{k_1+1}, i_{k_1}};\ \lambda_{i_{k_1+1}}^{(0)} = 0,\ i_{k_1+1} = 1,\ v_{i_{k_1+1}} = v_1. \tag{4.23}$$

Складывая равенства (4.22) и учитывая (4.23), имеем

$$l(v_1 v_{i_{k_1}} \ldots v_{i_2} v_{i_1}) = \lambda_{i_1}^{(k_1)},$$

т. е. $v_1 v_{i_{k_1}} \ldots v_{i_2} v_{i_1}$ — искомый минимальный путь из v_1 в v_{i_1} в нагруженном орграфе D. Заметим, что в этом пути ровно k_1 дуг. Следовательно, мы определили путь с минимальным числом дуг среди всех минимальных путей из v_1 в v_{i_1} в нагруженном орграфе D.

Замечание 4.25. Номера $i_2, i_3, \ldots, i_{k_1}$, удовлетворяющие (4.22), вообще говоря, могут быть выделены неоднозначно. Эта неоднозначность соответствует случаям, когда существует несколько различных путей из v_1 в v_{i_1} с минимальным числом дуг среди минимальных путей из v_1 в v_{i_1} в нагруженном орграфе D.

Замечание 4.26. Алгоритм 4.4 можно модифицировать, с тем чтобы определить минимальный путь из v_1 в заданную вершину v_{i_1} среди путей из v_1 в v_{i_1}, содержащих не более k_0 дуг, где k_0 — заданное число, $k_0 \geqslant 1$. Для этого в алгоритме 4.4 вместо $\lambda_{i_1}^{(n-1)}$ следует воспользоваться $\lambda_{i_1}^{(k_0)}$. Отметим, что при использовании указанной модификации алгоритма 4.4 предположение об отсутствии простых контуров отрицательной длины излишне. При этом, если в орграфе D имеются простые контуры отрицательной длины, может выполняться неравенство $\lambda_{i_1}^{(k_0)} < 0$.

Замечание 4.27. Алгоритм 4.4 и его модификация, описанная выше, после соответствующего изменения в терминологии и обозначениях применимы и к неориентированному графу G. При этом условие отсутствия простых контуров отрицательной длины заменяется условием отсутствия ребер отрицательной длины.

Пример 4.20. Определим минимальный путь из v_1 в v_6 в нагруженном орграфе D, изображенном на рис. 4.17, около каждой дуги которого указана ее длина.

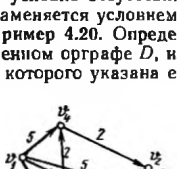


Рис. 4.17

Таблица 4.5

	α_1^0	α_2^0	α_3^0	α_4^0	α_5^0	α_6^0	$\lambda_1^{(0)}$	α_1^1	α_2^1	α_3^1	α_4^1	α_5^1	α_6^1	$\lambda_1^{(1)}$	α_1^2
v_1	0	∞	15	5	7	12	(0)	0	0	0	0	0	0	(0)	0
v_2	∞	∞	∞	∞	∞	2	∞	∞	∞	7	2	5	5	(2)	5
v_3	∞	2	∞	∞	∞	∞	∞	∞	5	3	3	3	3	(3)	3
v_4	∞	2	∞	∞	∞	∞	∞	∞	5	4	4	4	4	(4)	4
v_5	∞	∞	7	2	∞	∞	∞	∞	7	2	2	2	2	(2)	2
v_6	∞	∞	∞	∞	∞	∞	∞	∞	19	12	9	9	7	(7)	7

Длины дуг орграфа D положительны, следовательно, у него нет простых контуров отрицательной длины. Составим (6×6)-матрицу $C(D)$ длин дуг нагруженного орграфа D. Эта матрица представлена в табл. 4.5. Справа от матрицы $C(D)$ припишем шесть столбцов $(\lambda_1^{(k)}, \lambda_2^{(k)}, \ldots, \lambda_6^{(k)})^{\mathrm{T}}$, где $k=0, 1, 2, 3, 4, 5$, которые будем определять, используя (4.21), рекуррентное соотношение (4.15) и исходя из (4.14). Величина $\lambda_6^{(5)}=7$ выражает длину минимального пути из v_1 в v_6 в нагруженном орграфе D. Най-

дем минимальное число $k_1 \geqslant 1$, при котором выполняется равенство $\lambda_6(k_1) = \lambda_6^{(5)}$. Из табл. 4.5 получаем, что $k_1 = 4$. Таким образом, минимальное число дуг в пути среди всех минимальных путей из v_1 в v_6 в нагруженном орграфе D равняется 4. Определим теперь последовательность номеров i_1, i_2, i_3, i_4, i_5, где $i_1 = 6$, удовлетворяющих (4.22) (для этого используем формулу (4.15)). Из табл. 4.5 получаем, что в качестве такой последовательности надо взять номера 6, 2, 3, 5, 1, так как

$$\lambda_2^{(3)} + c_{26} = 5+2=7=\lambda_6^{(4)};$$
$$\lambda_3^{(2)} + c_{32} = 3+2=5=\lambda_2^{(3)};$$
$$\lambda_5^{(1)} + c_{53} = 2+1=3=\lambda_3^{(2)};$$
$$\lambda_1^{(0)} + c_{15} = 0+2=2=\lambda_5^{(1)}.$$

Тогда $v_1v_5v_3v_2v_6$ — искомый минимальный путь из v_1 в v_6 в нагруженном орграфе D, причем он содержит минимальное число дуг среди всех возможных минимальных путей из v_1 в v_6.

Пример 4.21. Определим путь из v_1 в v_6 в нагруженном орграфе D (см. пример 4.20) минимальной длины среди путей из v_1 в v_6, содержащих не более трех дуг.

Из табл. 4.5 получаем $\lambda_6^{(3)} = 9$, т. е. искомый путь имеет длину, равную 9. Находим теперь минимальное число k_1, при котором $\lambda_6^{(k_1)} = \lambda_6^{(3)}$. Из табл. 4.5 следует, что $k_1 = 3$. Далее определяем последовательность номеров i_1, i_2, i_3, i_4, где $i_1 = 6$, удовлетворяющих (4.22). Из табл. 4.5 видно, что в качестве такой последовательности можно взять номера 6, 2, 4, 1. Тогда $v_1v_4v_2v_6$ — искомый путь.

Пример 4.22. Определим путь из v_1 в v_1 в нагруженном орграфе D, изображенном на рис. 4.18, *а*, минимальной длины среди путей из v_1 в v_1, содержащих не более пяти дуг.

В табл. 4.6 указаны матрицы $C(D)$ длин дуг и шесть столбцов $(\lambda_1^{(k)}, \lambda_2^{(k)}, \lambda_3^{(k)})^T$ для $k = 0, 1, 2, 3, 4, 5$, которые определены

Таблица 4.6

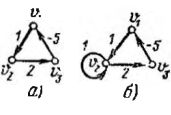


Рис. 4.18

	v_1	v_2	v_3		(0) λ_i	(1) λ_i	(2) λ_i	(3) λ_i	(4) λ_i	(5) λ_i
v_1	∞	1	∞		0	0	0	2	−2	−2
v_2	∞	∞	2		∞	1	1	1	−1	−1
v_3	5	∞	∞		∞	∞	3	3	3	1

по рекуррентным формулам (4.15), (4.16), исходя из (4.14). Из таблицы получаем $\lambda_1^{(5)} = -2$, т. е. искомый путь имеет длину, равную −2. Находим теперь минимальное число k_1, при котором $\lambda_1^{(k_1)} = \lambda_1^{(5)}$. Из таблицы следует, что $k_1 = 3$. Далее определяем последовательность номеров i_1, i_2, i_3, i_4, где $i_1 = 1$, удовлетворяющих (4.22). Из таблицы видно, что в качестве такой по-

следовательности надо взять номера 1, 3, 2, 1. Тогда $v_1v_2v_3v_1$ — искомый путь.

Замечание 4.28. Нетрудно показать, что утверждение 4.23, а следовательно, и алгоритм 4.4 вместе с его модификацией остаются справедливыми и для нагруженного ориентированного псевдографа. При этом в случае кратных дуг следует учитывать лишь дуги минимальной длины, а при наличии петель — лишь петли отрицательной длины. Это замечание переносится и на неориентированные псевдографы.

Пример 4.23. Если в орграф D (см. пример 4.22) добавить петлю (v_2, v_2) длины -1 (см. рис. 4.18, *б*), то решением примера 4.22 будет путь $v_1v_2v_2v_2v_3v_1$ (см. табл. 4.7).

Таблица 4.7

	v_1	v_2	v_3	$\lambda_i^{(0)}$	$\lambda_i^{(1)}$	$\lambda_i^{(2)}$	$\lambda_i^{(3)}$	$\lambda_i^{(4)}$	$\lambda_i^{(5)}$
v_1	∞	1	∞	0	0	0	−2	−3	−4
v_2	∞	−1	2	∞	1	0	−1	−2	−3
v_3	−5	∞	∞	∞	∞	3	2	1	0

Приведем теперь утверждение, дающее нам легко проверяемое необходимое и достаточное условие того, что в D отсутствуют простые контуры отрицательной длины. При этом будем рассматривать случай, когда из вершины v_1 достижимы все другие вершины орграфа D. Этого всегда можно добиться удалением вершин, не достижимых из v_1.

Утверждение 4.27. *Пусть в орграфе $D = (V, X)$, где $V = \{v_1, ..., v_n\}$, любая вершина v_i, $i \in \{2, ..., n\}$, достижима из v_1. Тогда, для того чтобы в D отсутствовали простые контуры отрицательной длины, необходимо и достаточно, чтобы выполнялось условие*

$$\lambda_i^{(n-1)} = \lambda_i^{(n)}, \quad i = 1, ..., n. \tag{4.24}$$

Необходимость следует из уже доказанных для этого случая равенств (4.20), (4.21).

Достаточность. Используя утверждение 4.23, из (4.24) получаем равенства (4.20) (в том числе и для $i = 1$), что может быть только при отсутствии в D простых контуров отрицательной длины. Действительно, если в D существует простой контур с r дугами, проходящий через вершину v_i и имеющий длину l_0, где $l_0 < 0$, то $\lambda_i^{(kr+n-1)} \leqslant \lambda_i^{(n-1)} + kl_0$, $k = 1, 2, ...$, что противоречит (4.20).

4.2.6. Специальные пути (маршруты) в орграфах (графах)

Для определенности будем рассматривать орграфы (наши рассуждения справедливы и для графов). Нередко при решении прикладных задач необходимо найти не обязательно минимальный путь из v в w в орграфе D, где v, w — заданные вершины,

а путь (или множество путей), обладающий некоторым свойством α, где α — одноместный предикат, определенный на множестве П путей в орграфе D. Свойство α называется *латинским*, если из того, что путь $\pi = \pi_1 \circ \pi_2$, где π_1, $\pi_2 \in \Pi$, обладает свойством α, следует, что пути π_1, π_2 также обладают свойством α.

Пример 4.24. Приведем некоторые латинские свойства путей в орграфе:

1) не проходить через данную вершину (или через заданное множество вершин);

2) не проходить через данную дугу (или через заданное множество дуг);

3) быть простой цепью;

4) быть простой цепью или простым контуром;

5) быть цепью или контуром;

6) не проходить через каждую вершину более k раз (где k — заданное число, $k \geqslant 1$).

Опишем матричный способ перечисления путей в орграфе, обладающих заданным латинским свойством α, называемый *методом латинской композиции*.

Предварительно введем некоторые обозначения. Пусть D — орграф, α — латинское свойство путей в D, Π_α — множество путей в D, обладающих свойством α. Введем бинарную операцию $\overset{\alpha}{\circ}$ на $\overline{\Pi}_\alpha = \Pi_\alpha \cup \{\varnothing\}$. Пусть $\pi_1 = u_1 u_2 \ldots u_k \in \Pi_\alpha$, $\pi_2 = w_1 w_2 \ldots w_l \in \Pi_\alpha$. Положим

$$\pi_1 \overset{\alpha}{\circ} \pi_2 = \begin{cases} \pi_1 \circ \pi_2, \text{ если } u_k = w_1 \text{ и путь } \pi_1 \circ \pi_2 \text{ обладает} \\ \text{свойством } \alpha; \\ \varnothing \text{ — в противном случае.} \end{cases}$$

Кроме того, для любого пути $\pi = \Pi_\alpha$ положим

$$\varnothing \overset{\alpha}{\circ} \pi = \pi \overset{\alpha}{\circ} \varnothing = \varnothing \overset{\alpha}{\circ} \varnothing = \varnothing.$$

Очевидно, что введенная бинарная операция является ассоциативной, т. е. $(\overline{\Pi}_\alpha, \overset{\alpha}{\circ})$ — полугруппа. Будем применять бинарную операцию $\overset{\alpha}{\circ}$ и к множествам элементов из $\overline{\Pi}_\alpha$. Пусть $\Pi_1 \subseteq \overline{\Pi}_\alpha$, $\Pi_2 \subseteq \overline{\Pi}_\alpha$. Тогда по определению

$$\Pi_1 \overset{\alpha}{\circ} \Pi_2 = \bigcup_{\substack{\pi_1 \in \Pi_1 \\ \pi_2 \in \Pi_2}} \pi_1 \overset{\alpha}{\circ} \pi_2.$$

Пусть $D = (V, X)$ — орграф, где $V = \{v_1, \ldots, v_n\}$. Введем для любого целого $k \geqslant 1$ *латинскую матрицу* $L^{(k)}_\alpha = [l^{(k)}_{ij}] = L_\alpha{}^{(k)}(D)$ размерности $n \times n$ такую, что $l^{(k)}_{ij}$ — множество путей длины k из v_i в v_j, обладающих свойством α (в частности, $l^{(k)}_{ij} = \varnothing$, если таких путей нет).

Определим теперь композицию $L_\alpha^{(k)} \overset{a}{\circ} L_\alpha^{(m)}$ латинских матриц $L_\alpha^{(k)}$, $L_\alpha^{(m)}$, где $k \geqslant 1$, $m \geqslant 1$. Под результатом композиции $C = L_\alpha^{(k)} \overset{a}{\circ} L_\alpha^{(m)}$ будем понимать квадратную матрицу $C = [c_{ij}]$ порядка n с элементами

$$c_{ij} = \bigcup_{r=1}^{n} l_{ir}^{(k)} \overset{a}{\circ} l_{rj}^{(m)}.$$

Утверждение 4.28. *При любом $k \geqslant 1$ выполняется равенство*

$$L_\alpha^{(k)} = L_\alpha^{(1)} \overset{a}{\circ} L_\alpha^{(1)} \overset{a}{\circ} \ldots \overset{a}{\circ} L_\alpha^{(1)},$$

где при $k \geqslant 2$ справа берется композиция k матриц, а при $k = 1$ выражение справа вырождается в $L_\alpha^{(1)}$.

Доказательство будем проводить индукцией по k. При $k = 1$ справедливость утверждения 4.28 очевидна. Предположим, что данное утверждение выполняется при некотором $k \geqslant 1$. Покажем его справедливость при $k+1$. Заметим, что согласно определению $l_{ij}^{(k+1)}$, $l_{ir}^{(k)}$, $l_{rj}^{(1)}$, $\overset{a}{\circ}$ выполняется равенство

$$l_{ij}^{(k+1)} = (l_{i1}^{(k)} \overset{a}{\circ} l_{1j}^{(1)}) \cup \ldots \cup (l_{in}^{(k)} \overset{a}{\circ} l_{nj}^{(1)}),$$

а следовательно, $L_\alpha^{(k+1)} = L_\alpha^{(k)} \overset{a}{\circ} L_\alpha^{(1)}$, откуда, используя индуктивное предположение, получаем справедливость доказываемого утверждения.

Следствие. Для любых $k \geqslant 1$, $m \geqslant 1$ выполняется равенство

$$L_\alpha^{(k)} \overset{a}{\circ} L_\alpha^{(m)} = L_\alpha^{(k+m)}. \qquad (4.25)$$

Из утверждения 4.28 вытекает, что задача перечисления путей в орграфе D заданной длины $k \geqslant 2$, обладающих свойством α, сводится к выполнению следующих простых действий. Составляем матрицу $L_\alpha^{(1)}$ (по определению). Теперь для перечисления путей длины 2, обладающих свойством α, достаточно найти $L_\alpha^{(2)} = L_\alpha^{(1)} \overset{a}{\circ} L_\alpha^{(1)}$. Для перечисления путей длины 3, обладающих свойством α, находим $L_\alpha^{(3)} = L_\alpha^{(2)} \overset{a}{\circ} L_\alpha^{(1)}$ и т. д. Применяя формулу (4.25), можно при больших k ускорить решение задачи. Так, например, можно найти $L_\alpha^{(4)}$ сразу после определения $L_\alpha^{(2)}$ по формуле $L_\alpha^{(4)} = L_\alpha^{(2)} \overset{a}{\circ} L_\alpha^{(2)}$.

Пример 4.25. Найти все простые цепи длины 3 в орграфе D, изображенном на рис. 4.19.

Воспользуемся методом латинской композиции. Последовательность результатов композиции $L_\alpha^{(1)}$, $L_\alpha^{(2)} = L_\alpha^{(1)} \overset{a}{\circ} L_\alpha^{(1)}$, $L_\alpha^{(3)} = L_\alpha^{(2)} \overset{a}{\circ} L_\alpha^{(1)}$ представлена соответственно в табл. 4.8*а*—4.8*в*, где α — свойство маршрутов «быть простой цепью».

Замечание 4.29. При больших n, k, где n — количество вершин в орграфе D, k — число дуг в перечисляемых путях орграфа D, практическое применение метода латинской композиции оказывается затруднительным (поскольку требуется большой объем вычислений и оперативной памяти ЭВМ), что накладывает ограничения на область его использования.

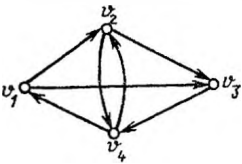


Рис. 4.19

Замечание 4.30. Как уже отмечалось, метод латинской композиции (при соответствующем изменении в терминологии) применим и к неориентированным графам и даже к произвольным псевдографам (ориентированным и неориентированным).

Таблица 4.8а

$\varnothing$	v_1v_2	v_1v_3	$\varnothing$
$\varnothing$	$\varnothing$	v_2v_3	v_2v_4
$\varnothing$	$\varnothing$	$\varnothing$	v_3v_4
v_4v_1	v_4v_2	$\varnothing$	$\varnothing$

Таблица 4.8б

$\varnothing$	$\varnothing$	$v_1v_2v_3$	$v_1v_2v_4$ $v_1v_3v_4$
$v_2v_4v_1$	$\varnothing$	$\varnothing$	$v_2v_3v_4$
$v_3v_4v_1$	$v_3v_4v_2$	$\varnothing$	$\varnothing$
$\varnothing$	$v_4v_1v_2$	$v_4v_1v_3$ $v_4v_2v_3$	$\varnothing$

Таблица 4.8в

$\varnothing$	$v_1v_3v_4v_2$	$\varnothing$	$v_1v_2v_3v_4$
$v_2v_3v_4v_1$	$\varnothing$	$v_2v_4v_1v_3$	$\varnothing$
$\varnothing$	$v_3v_4v_1v_2$	$\varnothing$	$\varnothing$
$\varnothing$	$\varnothing$	$v_4v_1v_2v_3$	$\varnothing$

4.2.7. Эйлеровы цепи и циклы

Классической в теории графов является следующая задача. В городе Кенигсберге имеется два острова, соединенных семью мостами с берегами реки Преголь и друг с другом так, как показано на рис. 4.20. Задача состоит в следующем: осуществить про-

гулку по городу таким образом, чтобы, пройдя по одному разу по каждому мосту, вернуться обратно. Решение этой задачи сводится к нахождению некоторого специального маршрута в графе.

Пусть G — псевдограф. Цепь (цикл) в G называется *эйлеровой* (*эйлеровым*), если она (он) проходит по одному разу через каждое ребро псевдографа G.

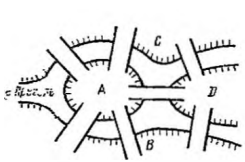
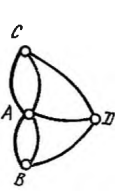


Рис. 4.20

Рис. 4.21



Поставим в соответствие схеме, приведенной на рис. 4.20, мультиграф G, изображенный на рис. 4.21, в котором каждой части суши соответствует вершина, а каждому мосту — ребро, соединяющее соответствующие вершины. На языке теории графов задача звучит следующим образом: найти эйлеров цикл в мультиграфе G (решение этой задачи было дано Л. Эйлером и будет приведено в примере 4.26).

Очевидно, что свойство маршрутов α: «проходить через каждое ребро не более одного раза» является латинским, а следовательно, все эйлеровы цепи и циклы псевдографа G можно получить, применяя к G метод латинской композиции. Они будут перечислены в матрице $L_\alpha^{(m)}(G)$, где $m = m(G)$ — количество ребер в G. При этом все эйлеровы циклы псевдографа G полностью перечисляются в любом диагональном элементе этой матрицы.

Прежде чем решать задачу о выделении эйлеровой цепи или эйлерова цикла в псевдографе G, надо выяснить, существуют ли они. Простейшее необходимое условие их существования, очевидно, заключается в связности G. Исчерпывающий ответ на вопрос об их существовании дают приводимые ниже теоремы 4.1 и 4.2.

Нам понадобится следующее вспомогательное утверждение.

Утверждение 4.29. *Если в псевдографе G имеется хотя бы одно ребро и отсутствуют висячие вершины, то G содержит хотя бы один простой цикл.*

Если в G имеется хотя бы одна петля $x = \{v, v\}$, то простым циклом является vxv. Пусть теперь в G нет петель, т. е. G — мультиграф. Если в G имеются кратные ребра $x_1 = \{v, w\}$, $x_2 = \{v, w\}$, то простым циклом является vx_1wx_2v. Пусть теперь в G нет кратных ребер, т. е. G — граф, и v_1, v_2 — произвольные смежные вершины в G (они найдутся, так как по условиям доказываемого утверждения в G имеется ребро). Рассмотрим по-

последовательность $v_1, v_2, v_3, \ldots$ вершин графа G такую, что для любого $i \geqslant 3$ вершины v_i, v_{i-1} смежны и $v_i \neq v_{i-2}$ (см. рис. 4.22). Поскольку в G висячих вершин нет, то такую последовательность можно продолжать неограниченно. Используя конечность множества вершин в G, получаем, что обязательно произойдет совпадение $v_i = v_j$, где $1 \leqslant i < j - 2$. Пусть это будет первое совпадение, т. е. совпадение с наименьшим номером j. Тогда $v_i v_{i+1} \ldots v_j$ — простой цикл в G.

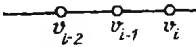


Рис. 4.22

Замечание 4.31. Доказательство утверждения 4.29, по существу, представляет собой алгоритм выделения простого цикла из псевдографа G с непустым множеством ребер и без висячих вершин.

Введем следующие обозначения. Если μ_1, μ_2 — циклы некоторого псевдографа G, имеющие хотя бы одну общую вершину и не имеющие общих ребер, то, очевидно, существует цикл, проходящий через все ребра, входящие в μ_1 и μ_2. Обозначим через $\mu_1 + \mu_2$ любой из таких циклов. Кроме того, для любого цикла μ обозначим через $V(\mu)$, $X(\mu)$ множества вершин и ребер, входящих в μ.

Утверждение 4.30. *Пусть μ — цикл без петель. Тогда $\forall v \in V(\mu)$ количество ребер в $X(\mu)$, инцидентных v, четно.*

Пусть $v \in V(\mu)$. Поскольку в цикле μ отсутствуют петли и все ребра попарно различны (по определению), то с каждым новым вхождением в μ вершины v в этот цикл войдут также два новых инцидентных ей ребра, а следовательно, общее число ребер в $X(\mu)$ инцидентных v, четно.

Следствие. Если вершина входит в некоторый цикл, то она не может быть висячей.

Теперь докажем, что справедлива

Теорема 4.1. *Для того чтобы связный псевдограф G обладал эйлеровым циклом, необходимо и достаточно, чтобы степени его вершин были четными.*

Необходимость. Пусть G обладает эйлеровым циклом. Покажем, что степени его вершин четны. Удалим из G все петли. В результате получим мультиграф G', который, очевидно, также обладает эйлеровым циклом. Поскольку эйлеров цикл мультиграфа G' содержит все ребра из G', а следовательно, и все вершины из G', то в силу утверждения 4.30 степени всех вершин мультиграфа G' четны, откуда, учитывая то, что вклад петли в степень вершины, инцидентной этой петле, равен 2 (см. замечание 4.1), получаем четность степеней всех вершин из G.

Достаточность будем доказывать индукцией по m — количеству ребер в G. При $m = 1$ связный псевдограф G с вершинами четной степени может выглядеть только следующим образом: $G = (V, X)$, где $V = \{v\}$, $X = \{x = \{v, v\}\}$, а в таком псевдографе существует эйлеров цикл. Предположим, что для некото-

рого целого $m \geqslant 2$ достаточность доказана для всякого псевдографа с $\leqslant m-1$ ребрами. Докажем ее справедливость для псевдографов с m ребрами. Пусть в связном псевдографе G с m ребрами степени вершин четны. Покажем, что в нем существует эйлеров цикл. В силу утверждения 4.29 в G имеется простой цикл μ_0. Если μ_0 содержит все ребра из G, то искомый эйлеров цикл найден. В противном случае удаляем из G все ребра, содержащиеся в μ_0. В результате получаем псевдограф G', каждая компонента связности которого является либо изолированной вершиной, либо псевдографом, степень каждой вершины которого четна (см. утверждение 4.30). Пусть G_i, $i=1, ..., p$, — компоненты связности псевдографа G', отличные от изолированных вершин. По индуктивному предположению для каждого псевдографа G_i можно построить эйлеров цикл μ_i. В силу связности G цикл μ_0 имеет общие вершины с любым из циклов μ_i, $i=1, ..., p$. Но тогда искомым эйлеровым циклом в G, очевидно, является цикл

$$\mu = (...((\mu_0 + \mu_1) + \mu_2) + ... + \mu_p).$$

Теорема 4.2. *Для того чтобы связный псевдограф G обладал эйлеровой цепью, необходимо и достаточно, чтобы он имел ровно две вершины нечетной степени.*

Необходимость. Пусть G имеет эйлерову цепь, соединяющую v, w. Добавим к G дополнительное ребро $\{v, w\}$. В результате получим псевдограф G', обладающий эйлеровым циклом, а следовательно (см. теорему 4.1), степени вершин псевдографа G' четны. Но тогда четны и степени вершин псевдографа G, за исключением вершин v, w.

Достаточность. Пусть G имеет ровно две вершины v, w нечетной степени. Добавим к G новое ребро $\{v, w\}$. В результате получим связный псевдограф G' со всеми вершинами четной степени. Но тогда в G' существует эйлеров цикл (см. теорему 4.1). Исключив из этого цикла ребро $\{v, w\}$, получим эйлерову цепь в псевдографе G, соединяющую v, w.

Замечание 4.32. Из доказательства теоремы 4.2 следует, что если мы находимся в условиях этой теоремы, то любая эйлерова цепь псевдографа G соединяет вершины нечетной степени.

Таким образом, мы имеем легко проверяемые необходимые и достаточные условия существования в произвольном псевдографе G эйлеровой цепи или эйлерова цикла.

Рассмотрим теперь задачу построения реализуемых на ЭВМ алгоритмов выделения эйлеровой цепи или эйлерова цикла в псевдографе G. Для построения и обоснования таких алгоритмов нам потребуется

Утверждение 4.31. *Пусть $G = (V, X)$ — связный псевдограф, $\mu_1, ..., \mu_l$ — циклы в G такие, что $l > 1$ и*

$$X(\mu_1) \cup ... \cup X(\mu_l) = X,\ X(\mu_i) \cap X(\mu_j) = \varnothing \text{ при } i \neq j, \qquad (4.26)$$

т. е. $X(\mu_1)$, ..., $X(\mu_l)$ — разбиение множества X. Тогда для цикла μ_i найдется цикл μ_i, где $i \neq 1$, такой, что $V(\mu_1) \cap V(\mu_i) \neq \varnothing$.

Рассмотрим произвольные вершины v, w такие, что $v \in V(\mu_1)$, $w \in V(\mu_2)$. Если $w \in V(\mu_1)$, то $i=2$. Пусть теперь $w \notin V(\mu_1)$. Тогда $v \neq w$, и в силу связности G найдется маршрут $v_1x_1v_2 \ldots$ $\ldots x_{k-1}v_k$, где $k \geqslant 2$, $v_1=v$, $v_k=w$, соединяющий v, w. Пусть j — номер такой, что $1 \leqslant j < k-1$, $v_j \in V(\mu_1)$, $v_{j+1} \notin V(\mu_1)$ (номер j найдется, так как $v_1 \in V(\mu_1)$, $v_k \notin V(\mu_1)$). Тогда для ребра $x_j = \{v_j, v_{j+1}\}$ имеем $x_j \in X(\mu_i)$, а следовательно, в силу (4.26) $\exists i \in \{2, \ldots, l\} : x_j \in X(\mu_i)$. Но тогда $v_j \in V(\mu_1) \cap V(\mu_i)$, т. е. утверждение полностью доказано.

Алгоритм 4.5 выделения эйлерова цикла в связном мультиграфе $G = (V, X)$, где $X \neq \varnothing$, с четными степенями вершин:

Шаг 1. Выделим из G цикл μ_1 (в силу утверждения 4.29 цикл μ_1 найдется, так как $X \neq \varnothing$ и в G отсутствуют висячие вершины). Полагаем $l=1$, $G' = G$.

Шаг 2. Удаляем из G' ребра, принадлежащие множеству $X(\mu_l)$. Полученный псевдограф снова обозначаем через G' (в силу утверждения 4.30 в G' все вершины всегда имеют четные степени). Если в G' отсутствуют ребра, то переходим к шагу 4. В противном случае выделяем из G' цикл μ_{l+1} (см. утверждение 4.29 о существовании такого цикла) и переходим к шагу 3.

Шаг 3. Присваиваем $l := l+1$ и переходим к шагу 2.

Шаг 4. По построению μ_1, ..., μ_l — циклы, удовлетворяющие условию (4.26). Если $l=1$, то μ_1 — искомый эйлеров цикл, и на этом работа алгоритма заканчивается. В противном случае находим цикл μ_i такой, что $V(\mu_1) \cap V(\mu_i) \neq \varnothing$, где $2 \leqslant i \leqslant l$ (в силу утверждения 4.31 цикл μ_i найдется). Переходим к шагу 5.

Шаг 5. Присваиваем $l := l-1$, $\mu_1 := \mu_1 + \mu_i$, $\mu_j := \mu_{j+1}$, $j = i, \ldots, l$, и переходим к шагу 4.

Применим алгоритм 4.5 к задаче выделения эйлерова цикла в самом общем случае, когда $G = (V, X)$ — связный псевдограф, где $X \neq \varnothing$, степени вершин которого четны. Удалим из G петли. В результате получим связный мультиграф G', степени вершин которого четны. Рассмотрим нетривиальный случай, когда в G' имеется хотя бы одно ребро. Тогда, применяя к G' алгоритм 4.5, находим эйлеров цикл в G'. Добавляя очевидным образом в этот цикл удаленные петли, получаем эйлеров цикл в G.

Рассмотрим теперь задачу о выделении эйлеровой цепи в связном псевдографе $G = (V, X)$, где $X \neq \varnothing$, имеющем ровно две вершины v, w нечетной степени. Добавляя к G ребро $\{v, w\}$, получаем псевдограф G' с четными степенями вершин. Выделив из G' эйлеров цикл и удалив из него ребро $\{v, w\}$, получим эйлерову цепь, соединяющую v, w.

Пример 4.26 (решение задачи о кенигсбергских мостах). Воспользуемся теоремой 4.1. Заметим, что для мультиграфа G, изображенного на рис. 4.22, имеем $\delta(A)=5$, $\delta(B)=3$, $\delta(C)=3$, $\delta(D)=3$, т. е. необходимое и достаточное условие существования

эйлерова цикла не выполняется, а следовательно, мультиграф G не обладает эйлеровым циклом.

4.2.8. Гамильтоновы цепи и циклы

Пусть G — псевдограф. Цепь (цикл) в G называется *гамильтоновой* (*гамильтоновым*), если она (он) проходит через каждую вершину псевдографа G ровно один раз.

С понятием гамильтоновых циклов тесно связана так называемая задача коммивояжера: в нагруженном графе G определить гамильтонов цикл минимальной длины (иными словами, коммерсант должен совершить поездку по городам и вернуться обратно, побывав в каждом городе ровно один раз, и при этом стоимость такой поездки должна быть минимальной).

Таблица 49

	1	2
1		
2		

На первый взгляд, понятие гамильтонова цикла сходно с понятием эйлерова цикла. А между тем графы, приведенные в табл. 4.9, где столбцы соответствуют случаям существования (столбец 1) и несуществования (столбец 2) гамильтоновых циклов, а строки — случаям существования (строка 1) и несуществования (строка 2) эйлеровых циклов, показывают независимость этих понятий.

Гамильтоновы цепи и циклы относятся к числу специальных маршрутов в графах. Очевидно, что свойство маршрутов α: «проходить через каждую вершину не более одного раза» является латинским, а следовательно, все гамильтоновы цепи и циклы псевдографа G можно получить, применяя к G метод латинской композиции. Пусть G является n-вершинным псевдографом. Используя тот очевидный факт, что длина любой гамильтоновой цепи равна $n-1$, а длина любого гамильтонова цикла равна n, получаем, что все гамильтоновы цепи будут перечислены в непустых элементах матрицы $L_\alpha^{(n-1)}(G)$, за исключением элементов на главной диагонали, а все гамильтоновы циклы — в каждом диагональном элементе матрицы $L_\alpha^{(n)}(G)$. Однако, как отмечалось выше, метод латинской композиции в данном случае практически применим лишь при достаточно малых n, и поэтому представляет интерес разработка более экономичных методов.

Заметим, что (как и в случае с эйлеровыми цепями и циклами) было бы полезно иметь сравнительно простые необходимые и достаточные условия существования гамильтоновых цепей и

циклов. Однако этот вопрос является весьма сложным и здесь не обсуждается.

Вопросы существования и нахождения гамильтоновых цепей и циклов в псевдографах очевидным образом сводятся к аналогичным вопросам для графов, и поэтому в дальнейшем будем говорить только о графах.

Рассмотрим простой класс графов, в которых заведомо существуют гамильтоновы цепи и циклы. Граф G называется *полным*, если каждая его вершина смежна со всеми остальными вершинами. Очевидно, что в полном графе всегда существуют гамильтонов цикл, а также гамильтоновы цепи, соединяющие две произвольные вершины этого графа. Таким образом, простейшим достаточным условием существования гамильтоновых цепей и циклов в графе является его полнота. Приведем также простейшие необходимые условия. Очевидным необходимым условием существования гамильтоновых цепей и циклов в графе G является связность G. Более тонким необходимым условием существования гамильтонова цикла в графе G является

Утверждение 4.32. *Если граф G обладает гамильтоновым циклом, то в нем отсутствуют точки сочленения.*

Пусть в G имеется точка сочленения v_1. Докажем, что G не может обладать гамильтоновым циклом. Предположим противное, т. е. что G обладает гамильтоновым циклом $\mu = v_1v_2 \dots v_nv_1$, где $n = n(G)$. Поскольку v_1 — точка сочленения, то в результате удаления v_1 из G получаем граф G' с компонентами связности $G_1, \dots, G_p$, где $p \geqslant 2$. Пусть компоненты связности пронумерованы таким образом, что $v_2 \in G_1 = (V_1, X_1)$. Заметим, что по определению гамильтонова цикла выполняется $v_2 \neq v_1$, $v_3 \neq v_1$, откуда $\{v_2, v_3\} \in X_1$, а следовательно, $v_3 \in V_1$. Аналогично получаем, что $v_4, \dots, v_n \in V_1$, а это противоречит тому, что $p \geqslant 2$.

Приведем некоторые наиболее простые методы выделения гамильтоновых цепей и циклов в графе $G = (V, X)$, где $V = \{v_1, \dots, v_n\}$. Пожалуй, самым простым является метод перебора всевозможных перестановок $v_{i_1}, v_{i_2}, \dots, v_{i_n}$ множества V. Для каждой из них проверяем, является ли $v_{i_1}v_{i_2} \dots v_{i_n}$ маршрутом в G. Если является, то $v_{i_1}v_{i_2} \dots v_{i_n}$ — гамильтонова цепь в G, в противном случае переходим к следующей перестановке. Тогда по окончании перебора будут выделены все гамильтоновы цепи в графе G. Аналогично для выделения гамильтоновых циклов перебираем всевозможные перестановки $v_1, v_{i_1}, \dots, v_{i_{n-1}}$ множества V, для каждой из которых проверяем, является ли $v_1v_{i_1} \dots v_{i_{n-1}}v_1$ маршрутом в G. Если является, то $v_1v_{i_1} \dots v_{i_{n-1}}v_1$ — гамильтонов цикл в G, в противном случае переходим к следующей перестановке. Тогда по окончании перебора будут выделены все гамильтоновы циклы в графе G. Очевидно, что при выделении всех гамильтоновых цепей нам придется перебрать $n!$ перестановок, а при выделении всех гамильтоновых циклов — $(n-1)!$ перестановок. При этом

случае полного графа ни одна из перестановок не окажется отброшенной, т. е. данный метод является эффективным для графов, близких к полным.

Отметим, что описанный метод не учитывает информации об исследуемом графе G и является как бы ориентированным на самый «худший» случай, когда G — полный граф.

Рассмотрим метод, аналогичный предыдущему, но использующий информацию о G. Составим всевозможные последовательности вершин $v_{i_1}, v_{i_2}, \ldots, v_{i_r}$, где $v_{i_1} \in V$, $v_{i_2} \in G(v_{i_1}) \setminus \{v_{i_1}\}$, ..., $v_{i_r} \in G(v_{i_{r-1}}) \setminus \{v_{i_1}, \ldots, v_{i_{r-1}}\}$, $G(v_{i_r}) \setminus \{v_{i_1}, \ldots, v_{i_r}\} = \varnothing$. Тогда в каждом случае, когда $r = n$, последовательность $v_{i_1} v_{i_2} \ldots v_{i_n}$ есть гамильтонова цепь в графе G. Соответственно в каждом случае, когда $r = n$, $v_{i_1} \in G(v_{i_n})$, последовательность $v_{i_1} v_{i_2} \ldots v_{i_n} v_{i_1}$ есть гамильтонов цикл в графе G. При этом будут выделены все гамильтоновы цепи и циклы в графе G. Как и в предыдущем методе, при выделении гамильтоновых циклов можно предполагать, что $i_1 = 1$.

Задачи и упражнения

1. Используя алгоритм Тэрри, определить замкнутый маршрут в каждом из графов, изображенных на рис. 4.14 и 4.16, проходящий ровно два раза (по одному разу в каждом направлении) через каждое ребро графа.

2. Доказать, что в сильно связном орграфе с симметричной матрицей смежности существует контур, проходящий по одному разу через каждую дугу орграфа.

3. Найти минимальный путь из v_1 в v_7 в орграфах, заданных матрицами смежности:

а) $\begin{bmatrix} 0&0&0&1&1&1&0&0 \\ 1&0&1&1&0&1&1&1 \\ 1&1&0&1&1&1&0&1 \\ 0&0&0&0&1&1&1&0 \\ 1&1&0&1&1&0&1&0 \\ 0&1&1&1&0&0&0&0 \\ 1&1&1&0&0&1&1&0 \end{bmatrix}$; б) $\begin{bmatrix} 0&0&0&0&1&1&1&0 \\ 1&0&1&0&1&0&1&1 \\ 1&1&0&1&1&1&0&0 \\ 0&0&1&0&1&1&1&0 \\ 1&1&0&1&0&1&1&0 \\ 0&0&1&1&0&1&0&0 \\ 1&1&0&1&1&0&1&0 \end{bmatrix}$; в) $\begin{bmatrix} 0&0&0&1&1&0&1&0 \\ 1&0&1&1&0&1&0&1 \\ 0&1&0&1&1&1&1&0 \\ 1&0&0&0&1&1&0&0 \\ 1&0&1&1&0&1&1&0 \\ 0&0&0&1&1&0&0&0 \\ 1&0&1&1&0&1&1&0 \end{bmatrix}$.

4. Определить минимальный путь из v_1 в v_7 в нагруженных орграфах с заданными матрицами длин дуг:

а) $\begin{bmatrix} \infty&\infty&5&4&2&2&9 \\ \infty&\infty&1&1&\infty&1&1 \\ 2&\infty&\infty&4&1&\infty&3 \\ \infty&2&1&\infty&4&\infty&\infty \\ \infty&\infty&2&2&\infty&3&6 \\ 1&5&\infty&1&4&\infty&\infty \\ 2&\infty&1&\infty&1&2&\infty \end{bmatrix}$; б) $\begin{bmatrix} \infty&4&\infty&\infty&5&1&\infty \\ 3&\infty&2&1&\infty&\infty&\infty \\ 1&1&\infty&\infty&\infty&\infty&3 \\ \infty&3&1&\infty&1&\infty&\infty \\ \infty&\infty&2&\infty&\infty&1&5 \\ \infty&3&\infty&2&2&\infty&\infty \\ \infty&\infty&2&\infty&\infty&2&\infty \end{bmatrix}$; в) $\begin{bmatrix} \infty&\infty&9&\infty&\infty&2&12 \\ 1&\infty&\infty&\infty&1&2&4 \\ 2&1&\infty&\infty&\infty&1&\infty&2 \\ \infty&1&1&\infty&\infty&1&\infty \\ 1&2&\infty&2&\infty&\infty&\infty \\ \infty&\infty&\infty&\infty&\infty&1&\infty&8 \\ \infty&2&1&\infty&1&2&\infty \end{bmatrix}$.

5. Определить путь из v_1 в v_7 минимальной длины в каждом нагруженном орграфе (см. задачу 4) среди путей из v_1 в v_7, содержащих не более k дуг, где: а) $k=2$; б) $k=3$; в) $k=4$.

6. Определить, имеются ли в нагруженном орграфе D с заданной матрицей длин дуг $C(D)$ простые контуры отрицательной длины? Найти пути минимальной длины из v_1 во все остальные вершины среди путей, содержащих не более шести дуг. Рассмотреть случай

$$C(D) = \begin{bmatrix} \infty & 6 & \infty & 6 & \infty & 10 \\ \infty & \infty & 2 & \infty & -1 & 3 \\ \infty & \infty & \infty & \infty & \infty & \infty \\ \infty & -3 & \infty & \infty & \infty & \infty \\ \infty & \infty & 1 & 2 & \infty & \infty \\ \infty & \infty & 5 & \infty & \infty & \infty \end{bmatrix}.$$

7. Проверить, существуют ли в мультиграфах, заданных матрицами смежности, эйлеровы цепи и циклы? Если да, то найти их. Рассмотреть случаи:

а) $\begin{bmatrix} 0 & 1 & 0 & 0 & 2 & 1 \\ 1 & 0 & 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 2 & 0 & 1 \\ 0 & 0 & 2 & 0 & 1 & 0 \\ 2 & 0 & 0 & 1 & 0 & 1 \\ 1 & 2 & 1 & 1 & 1 & 0 \end{bmatrix}$; б) $\begin{bmatrix} 0 & 1 & 0 & 0 & 2 & 1 \\ 1 & 0 & 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 2 & 0 & 1 \\ 0 & 0 & 2 & 0 & 1 & 1 \\ 2 & 0 & 0 & 1 & 0 & 0 \\ 1 & 2 & 1 & 1 & 0 & 0 \end{bmatrix}$.

4.3. ДЕРЕВЬЯ И ЦИКЛЫ

Граф G называется *деревом*, если он является связным и не имеет циклов. Граф G, все компоненты связности которого являются деревьями, называется *лесом*.

Пример 4.27. Граф, изображенный на рис. 4.23, является деревом.

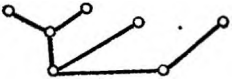


Рис. 4.23

4.3.1. Свойства деревьев

Следующие утверждения эквивалентны:

1) граф G есть дерево;

2) граф G является связным и не имеет простых циклов;

3) граф G является связным и число его ребер ровно на единицу меньше числа вершин;

4) любые две различные вершины графа G можно соединить единственной (и притом простой) цепью;

5) граф G не содержит циклов, но, добавляя к нему любое новое ребро, получаем ровно один (с точностью до направления обхода и начальной вершины обхода) и притом простой цикл (проходящий через добавляемое ребро).

Заметим, что для доказательства эквивалентности утверждений 1 и 2 достаточно воспользоваться тем фактом, что из любого

цикла можно выделить простой цикл (см. утверждение 4.3). Ниже (см. утверждения 4.33—4.40) будет обоснована эквивалентность утверждения 1 любому из утверждений 3—5, а тем самым будет доказана эквивалентность утверждений 1—5.

Утверждение 4.33. *Если у дерева G есть, по крайней мере, одно ребро, то у него обязательно найдется висячая вершина.*

Предположим, что в G нет висячих вершин. Тогда в силу утверждения 4.29 в G найдется цикл, а это противоречит тому, что G — дерево.

Утверждение 4.34. *Пусть G — связный граф, v — висячая вершина в G, G' — граф, полученный из G в результате удаления вершины v и инцидентного ей ребра. Тогда G' — связный граф.*

Предположим, что граф G' не является связным. Тогда в нем найдутся вершины v_1, $v_2 (v_1 \neq v_2)$, которые нельзя соединить маршрутом. Но в G их можно соединить маршрутом μ (в силу связности G). Выделим из маршрута μ цепь η, также соединяющую вершины v_1, v_2 (см. утверждение 4.4). Если эта цепь не проходит через v, то она является цепью и в G', что противоречит сделанному ранее предположению, а следовательно, она проходит через v. Пусть w — вершина, смежная с v. Она единственная, так как v — висячая вершина. Тогда указанная цепь η имеет вид $\eta = v_1 \ldots wvw \ldots v_2$, а значит, ребро (v, w) в этой цепи встречается более одного раза, что противоречит определению цепи. Полученное противоречие показывает, что исходное предположение неверно, т. е. G' — связный граф.

Замечание 4.33. Утверждение 4.34 остается справедливым и для произвольного псевдографа G. Доказательство аналогично.

Утверждение 4.35. *Пусть G — дерево с n вершинами и m ребрами. Тогда $m = n - 1$.*

Доказательство проведем индукцией по n — количеству вершин. При $n=1$ имеем $m=0$, т. е. $m = n - 1$. Пусть при некотором $n \geqslant 2$ доказываемое равенство справедливо для любого дерева с $n-1$ вершинами. Докажем его справедливость для любого дерева G с n вершинами. Поскольку в дереве с $n \geqslant 2$ вершинами имеется, по крайней мере, одно ребро (дерево — связный граф), то в силу утверждения 4.33 в рассматриваемом дереве G найдется висячая вершина. Удалим ее вместе с инцидентным ей ребром. В силу утверждения 4.34 оставшийся граф G' будет связным. Кроме того, он не содержит циклов (так как в противном случае и граф G, являющийся деревом, содержал бы их), т. е. G' — дерево. Заметим, что G' содержит $n-1$ вершин и $m-1$ ребер. Но тогда по индуктивному предположению выполняется $m - 1 = n - 2$, откуда $m = n - 1$.

Утверждение 4.36. *Пусть G' — граф, являющийся деревом, G — граф, полученный в результате добавления к G' новой вершины v и ребра (v, w), где w — некоторая вершина графа G'. Тогда G — дерево.*

Прежде всего заметим, что v — висячая вершина графа G. Для доказательства того, что G — дерево, сначала покажем, что граф G связный. Для этого достаточно доказать, что любую вершину u графа G' можно сое-

дить маршрутом с v (поскольку в силу связности графа G' любые две вершины графа G, отличные от v, заведомо можно соединить маршрутом). Рассмотрим нетривиальный случай, когда $u \neq w$. В силу связности графа G' существует маршрут η' в G' (а следовательно, и в G), соединяющий u, w. Но тогда $\eta \cup wv$ — маршрут в G, соединяющий u, v. Покажем теперь, что в G нет циклов. Предположим, что в G имеется цикл μ. Согласно следствию из утверждения 4.30 вершина v (являющаяся висячей) не содержится в этом цикле, а значит, μ — цикл в G', но это противоречит тому, что G' — дерево.

Утверждение 4.37. *Пусть $G = (V, X)$ — связный граф с m ребрами и n вершинами и пусть также выполняется равенство $m = n - 1$. Тогда G — дерево.*

Доказательство проведем индукцией по n — количеству вершин. Если $n=1$, то $m=n-1=0$. Граф, содержащий одну вершину и не имеющий ребер, очевидно, является деревом. Пусть при некотором $n \geqslant 2$ доказываемое утверждение справедливо для любого графа не более чем с $n-1$ вершинами. Докажем справедливость этого утверждения для произвольного графа G с n вершинами. Покажем, что в G имеется висячая вершина. Если ее нет, то $\forall v \in V \ \delta(v) \geqslant 2$, а следовательно, используя утверждение 4.1, получаем, что $2m = \sum\limits_{v \in V} \delta(v) \geqslant 2n$, откуда $m \geqslant n$, а это противоречит условию $m = n - 1$. Таким образом, в графе G имеется висячая вершина v. Удалим ее вместе с инцидентным ей ребром. В результате получим граф G' с $n-1$ вершинами и $n-2$ ребрами. Согласно утверждению 4.34 граф G' является связным, а следовательно, по индуктивному предположению G' — дерево. Но тогда в силу утверждения 4.36 и граф G является деревом.

Замечание 4.34. Утверждение 4.37 остается справедливым и для произвольного псевдографа G. Доказательство аналогично.

Из утверждений 4.35, 4.37 следует эквивалентность утверждений 1 и 3 (см. с. 206).

Утверждение 4.38. *Пусть G — дерево. Тогда любая цепь в G будет простой.*

Пусть $v_1 v_2 \ldots v_k$ — цепь в дереве G, не являющаяся простой. Тогда при некоторых $i_1, i_2 \in \{1, 2, \ldots, k\}$ выполняется $i_1 \neq i_2$, $v_{i_1} = v_{i_2}$. Пусть для определенности $i_1 < i_2$. Тогда $v_{i_1} v_{i_1+1} \ldots v_{i_2}$ — цикл в G, а это противоречит тому, что G — дерево.

Утверждение 4.39. *Для справедливости утверждения 1 необходимо и достаточно, чтобы выполнялось утверждение 4* (см. с. 206).

Необходимость. Пусть G — дерево и v, w — некоторые вершины графа G, где $v \neq w$. Тогда в силу связности G их можно соединить цепью (см. утверждение 4.4). Предположим, что найдутся две различные цепи η_1, η_2, соединяющие v, w. Согласно утверждению 4.38 цепи η_1, η_2 являются простыми. Пусть $\eta_1 = v_1 v_2 \ldots v_k$, $\eta_2 = w_1 w_2 \ldots w_l$, где $v_1 = w_1 = v$, $v_k = w_l = w$, $k \geqslant 2$, $l \geqslant 2$. Поскольку $\eta_1 \neq \eta_2$, найдется номер $k_1 \geqslant 1$ такой, что $v_i = w_i$, ..., $v_{k_1} = w_{k_1}$, $v_{k_1+1} \neq w_{k_1+1}$. Пусть k_2 — первый среди номеров $k_1 + 1$, ..., k такой, что вершина v_{k_2} встречается среди вершин w_{k_1+1}, ..., w_l (обязательно найдется, так как $v_k = w_l$). Пусть, далее, k_3 — первый номер среди $k_1 + 1$, ..., l такой, что $w_{k_3} = w_{k_2}$. В силу неравенства $v_{k_1+1} \neq w_{k_1+1}$ не может выполняться равенство $k_2 = k_3 = k_1 + 1$. Но тогда

$v_{k_1} v_{k_1+1} \dots v_{k_2} w_{k_3-1} \dots w_{k_1+1} w_{k_1}$

есть простой цикл в графе G (см. рис. 4.24), а это противоречит тому, что G — дерево.

Достаточность. Пусть в графе G любые две вершины можно соединить, и притом единственной цепью. Докажем, что G — дерево. Очевидно, что граф U связный. Покажем, что в G нет циклов. Пусть в G имеется цикл $v_1 v_2 \dots v_k v_1$, где $k \geqslant 3$ (так как при $k=2$ маршрут $v_1 v_2 v_1$ не является циклом). Но тогда вершины v_1, v_2 можно соединить двумя различными цепями: $v_1 v_2$, $v_1 v_k v_{k-1} \dots v_2$, что противоречит исходному предположению.

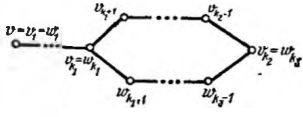


Рис. 4.24

Замечание 4.35. Нетрудно показать, что условие достаточности в утверждении 4.39 остается справедливым и для произвольного псевдографа G.

Утверждение 4.40. *Для справедливости утверждения 1 необходимо и достаточно, чтобы выполнялось утверждение 5* (см. с. 206).

Необходимость. Пусть $G = (V, X)$ — дерево. Тогда в G нет циклов. Пусть v, w — любые вершины из V такие, что $v \neq w$, $\{v, w\} \notin X$. Рассмотрим граф $G' = (V, X')$, где $X' = X \cup \{v, w\}$. Используя утверждение 4.39, получаем, что в G найдется простая цепь η_1, соединяющая v, w. Но тогда $\mu_1 = wv \circ \eta_1$ — простой цикл в графе G', проходящий через ребро $\{v, w\}$. Предположим, что в G' существует некоторый другой цикл μ_2. Тогда μ_2 обязательно должен проходить через ребро $\{v, w\}$ (так как в противном случае μ_2 был бы циклом в дереве G), а следовательно, он имеет вид (с точностью до направления обхода и выбора начальной вершины обхода) $\mu_2 = wv \circ \eta_2$, где η_2 — цепь в G', соединяющая v, w. Заметим, что поскольку μ_2 — цикл, то цепь η_2 не проходит через ребро $\{v, w\}$, а значит, является цепью в G. Но тогда в силу утверждения 4.39 получаем $\eta_1 = \eta_2$, а следовательно, $\mu_1 = \mu_2$.

Достаточность. Пусть для графа $G = (V, X)$ выполняется утверждение 5 (см. с. 206). Предположим, что G не является деревом. Согласно утверждению 5 граф G не содержит циклов, а поскольку в силу сделанного предположения граф G не является деревом, то он не может быть связным. Но тогда найдутся вершины v, $w \in V$, $v \neq w$, такие, что их нельзя соединить маршрутом в G. Добавим к графу G ребро $\{v, w\}$. В результате получим граф G', содержащий (см. утверждение 5) некоторый цикл μ. Очевидно, что μ проходит через ребро $\{v, w\}$ (так как в противном случае μ — цикл в G, а G не содержит циклов), а следова-

тельно, он имеет вид (с точностью до направления обхода и выбора начальной вершины обхода) $\mu = wv\cup\eta$, где η — цепь в G', соединяющая v, w. В силу того, что μ — цикл, η не содержит ребро $\{v, w\}$, а значит, является цепью в G, что противоречит сделанному ранее предположению.

4.3.2. Остовное дерево связного графа

Остовным деревом связного графа G называется любой его подграф, содержащий все вершины графа G и являющийся деревом.

Пусть G — связный граф. Тогда в силу утверждения 4.35 остовное дерево графа G (если оно существует) должно содержать $n(G) - 1$ ребер. Таким образом, любое остовное дерево графа G есть результат удаления из G ровно $m(G) - (n(G) - 1) = m(G) - n(G) + 1$ ребер.

Число $m(G) - n(G) + 1$ называется *цикломатическим числом* связного графа G и обозначается через $\nu(G)$.

Замечание 4.36. Понятия остовного дерева и цикломатического числа аналогичным образом определяются и для произвольного связного псевдографа G.

Покажем существование остовного дерева для произвольного связного псевдографа $G = (V, X)$, описав алгоритм его выделения.

Алгоритм 4.6:

Шаг 1. Выбираем в G произвольную вершину u_1, которая образует подграф G_1 псевдографа G, являющийся деревом. Полагаем $i = 1$.

Шаг 2. Если $i = n$, где $n = n(G)$, то задача решена, и G_i — искомое остовное дерево псевдографа G. В противном случае переходим к шагу 3.

Шаг 3. Пусть уже построено дерево G_i, являющееся подграфом псевдографа G и содержащее некоторые вершины u_1, ..., u_i, где $1 \leqslant i \leqslant n - 1$. Строим граф G_{i+1}, добавляя к графу G_i новую вершину $u_{i+1} \in V$, смежную в G с некоторой вершиной u_j графа G_i, и новое ребро $\{u_{i+1}, u_j\}$ (в силу связности G и того обстоятельства, что $i < n$, указанная вершина u_{i+1} обязательно найдется). Согласно утверждению 4.36 граф G_{i+1} также является деревом. Присваиваем $i := i+1$ и переходим к шагу 2.

Пример 4.28. Используя алгоритм 4.6, выделим остовное дерево графа G, изображенного на рис. 4.25. На рис. 4.26 приведена последовательность графов G_i, $i = 1, 2, 3, 4, 5$, получаемых в результате выполнения алгоритма 4.6. При этом в силу того, что $n(G) = 5$. G_5 — остовное дерево графа G.

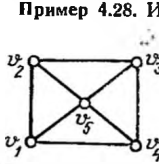


Рис. 4.25

Замечание 4.37. Остовное дерево связного графа может быть выделено, вообще говоря, не единственным способом (тем более

это верно для связного псевдографа). Общее число остовных деревьев связного графа может оказаться весьма большим. Например, для полного графа с n вершинами оно равно n^{n-2}.

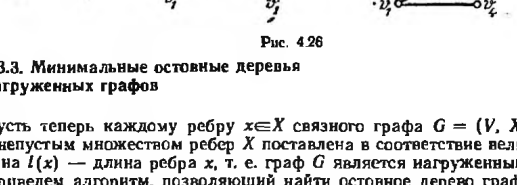


Рис. 4.26

4.3.3. Минимальные остовные деревья нагруженных графов

Пусть теперь каждому ребру $x \in X$ связного графа $G = (V, X)$ с непустым множеством ребер X поставлена в соответствие величина $l(x)$ — длина ребра x, т. е. граф G является нагруженным. Приведем алгоритм, позволяющий найти остовное дерево графа G с минимальной суммой длин содержащихся в нем ребер (по сравнению со всеми другими остовными деревьями графа G). Остовное дерево связного нагруженного графа G с минимальной суммой длин содержащихся в нем ребер будем называть *минимальным остовным деревом* (МОД) графа G.

Алгоритм 4.7 (выделения МОД нагруженного связного графа G):

Шаг 1. Выберем в графе G ребро минимальной длины. Вместе с инцидентными ему вершинами оно образует подграф G_2 графа G. Положим $i=2$.

Шаг 2. Если $i = n$, где $n = n(G)$, то задача решена, и G_i — искомое МОД графа G. В противном случае переходим к шагу 3.

Шаг 3. Строим граф G_{i+1}, добавляя к графу G_i новое ребро минимальной длины, выбранное среди всех ребер графа G, каждое из которых инцидентно какой-нибудь вершине графа G_i и одновременно инцидентно какой-нибудь вершине графа G, не содержащейся в G_i. Вместе с этим ребром включаем в G_{i+1} и инцидентную ему вершину, не содержащуюся в G_i. Присваиваем $i := i + 1$ и переходим к шагу 2.

Пример 4.29. Определим МОД нагруженного графа G, изображенного на рис. 4.27, используя алгоритм 4.7. На рис. 4.28 приведена последовательность графов G_i, $i=2$, 3, 4, 5, получаемых в результате выполнения алгоритма 4.7. При этом в силу того, что $n(G)=5$, G_5 — МОД графа G.

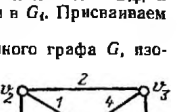


Рис. 4.27

Обоснование алгоритма 4.7. Для любого дерева T подграф T_1 графа T, в свою очередь являющийся деревом, будем

называть поддеревом дерева T. Покажем, что в результате применения к G алгоритма 4.7 получим МОД графа G. Предварительно докажем, что справедлива

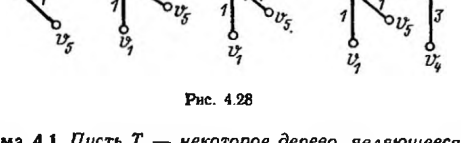


Рис. 4.28

Лемма 4.1. *Пусть T — некоторое дерево, являющееся поддеревом некоторого МОД G' связного нагруженного графа $G = (V, X)$. Пусть далее x — ребро минимальной длины, выбранное среди всех ребер графа G, каждое из которых инцидентно какой-нибудь вершине дерева T и одновременно инцидентно какой-нибудь вершине графа G, не содержащейся в T. Тогда, добавляя к T ребро x вместе с инцидентной ему вершиной, не содержащейся в T, получаем граф T', снова являющийся поддеревом некоторого МОД графа G.*

Предварительно заметим, что в силу утверждения 4.36 T — дерево. Пусть $T = (V_1, X_1)$, $G' = (V, X')$. Если $x \in X'$, то T' — поддерево дерева G', являющегося МОД графа G, т. е. в этом случае доказываемое утверждение справедливо. Пусть теперь $x \notin X'$. Рассмотрим граф G'', получаемый из G' в результате добавления к G' ребра x. Тогда в силу утверждения 4.40 (об эквивалентности утверждений I, 5 на с. 206) в G'' найдется простой цикл μ, проходящий через ребро x. Пусть $x = \{v_1, v_2\}$, $\mu = v_1 v_2 \ldots v_k$, где $k \geqslant 4$, $v_1 \in V_1$, $v_2 \notin V_1$, $v_k = v_1 \in V_1$. Пусть также k_1 — произвольный номер из $\{2, \ldots, k-1\}$ такой, что $v_{k_1} \notin V_1$, $v_{k_1+1} \in V_1$ (k_1 найдется, поскольку $v_2 \notin V_1$, $v_k \in V_1$). Так как μ — цикл, ребро $x' = \{v_{k_1}, v_{k_1+1}\}$ отлично от x, а следовательно, является ребром графа G. По условиям выбора x выполняется неравенство $l(x') \geqslant l(x)$, а поэтому, исключив из дерева G' ребро x' и включив вместо него ребро x, получим новое дерево G''' (граф G'' связный, так как получается в результате удаления из связного графа G'' ребра x', содержащегося в цикле μ этого графа, и, кроме того, количество ребер графа G''' ровно на единицу меньше числа его вершин, а следовательно, согласно утверждению 4.37, G''' — дерево), общая сумма длин ребер которого не увеличится, а значит, G''' снова будет МОД графа G, и при этом T' является поддеревом дерева G''', т. е. лемма доказана.

Для окончательного обоснования алгоритма 4.7 заметим, что любая вершина графа G является поддеревом любого МОД графа G, а следовательно, в силу леммы 4.1 граф G_2 (см. алго-

ритм 4.7) является поддеревом некоторого МОД графа G. Но тогда, снова используя лемму 4.1, последовательно находим, что каждый из графов G_2, G_3, ..., G_n, полученных в результате применения к G алгоритма 4.7, является поддеревом некоторого МОД графа G, а так как G_n содержит все вершины графа G, заключаем, что G_n — искомое МОД графа G.

Замечание 4.38. Для выделения МОД нагруженного псевдографа G следует предварительно удалить из G петли, из кратных ребер оставить лишь ребра минимальной длины, а затем применить к полученному таким образом графу алгоритм 4.7.

4.3.4. Вектор-циклы

Пусть $G = (V, X)$ — некоторый мультиграф, $X = \{x_1, ..., x_m\}$, $m \geqslant 1$. Введем произвольно ориентацию для ребер из X, т. е. каждое ребро $x = \{v, w\} \in X$ превратим в дугу x' такую, что либо $x' = (v, w)$, либо $x' = (w, v)$. В результате из исходного мультиграфа $G = (V, X)$ получим ориентированный мультиграф $D = (V, X')$.

Всюду в этом разделе будем пользоваться термином «цикл» в более широком смысле, а именно: под циклом будем понимать произвольный замкнутый маршрут. Как и ранее, замкнутый маршрут, в котором все ребра и вершины попарно различны, будем называть простым циклом.

Рассмотрим некоторый цикл

$$\mu = v_1 y_1 v_2 y_2 \ldots v_k y_k v_1 \tag{4.27}$$

в мультиграфе G, где $y_i = \{v_i, v_{i+1}\} \in X$, $i = 1, ..., k$, $k \geqslant 2$ (при этом $v_{k+1} = v_1$). Запись (4.27) задает направление прохода цикла μ через каждое ребро $y_i = \{v_i, v_{i+1}\}$ (от v_i к v_{i+1}), которое является для нас весьма существенным, и поэтому далее всюду будем рассматривать циклы с фиксированным направлением прохода через ребра, входящие в эти циклы. Говорят, что цикл μ проходит через ребро $y_i = \{v_i, v_{i+1}\}$ *в направлении выбранной ориентации*, если $y'_i = (v_i, v_{i+1})$; в противном случае (т. е. если $y'_i = (v_{i+1}, v_i)$) говорят, что цикл μ проходит через ребро y_i *в направлении, противоположном выбранной ориентации*.

Пусть $\mathbf{Z}^m$ — множество m-мерных векторов с целочисленными координатами и μ — некоторый цикл в мультиграфе G с выбранным направлением прохода через ребра, входящие в этот цикл. Вектор $C(\mu) \in \mathbf{Z}^m$ называется *вектор-циклом*, соответствующим циклу μ (или просто вектор-циклом μ), если i-я координата $C_i(\mu)$ вектора $C(\mu)$ равна разности $C^+_i(\mu) - C^-_i(\mu)$, где $C^+_i(\mu)$ — число проходов в цикле μ через ребро x_i в направлении выбранной ориентации, $C^-_i(\mu)$ — число проходов в цикле μ через ребро x_i в направлении, противоположном выбранной ориентации.

Пример 4.30. Пусть $G = (V, X)$ — мультиграф, изображенный на рис. 4.29, *а*. Введем в G ориентацию на ребрах. В результате получим, например, ориентированный мультиграф D, изображенный на рис. 4.29, *б*. Рассмотрим некоторые циклы в G:

$\mu_1 = v_1x_1v_2x_2v_3x_3v_1$;

$\mu_2 = v_1x_3v_3x_2v_2x_8v_4x_7v_3x_2v_2x_1v_1$;

$\mu_3 = v_1x_3v_3x_2v_2x_1v_1x_4v_5x_5v_3x_2v_2x_1v_1$;

$\mu_4 = v_1x_4v_5x_4v_1$;

$\mu_5 = v_3x_7v_4x_6v_3$.

Тогда

$C(\mu_1) = (1, 1, -1, 0, 0, 0, 0, 0)$;
$C(\mu_2) = (-1, -2, 1, 0, 0, 0, 1, 1)$;
$C(\mu_3) = (-2, -2, 1, -1, -1, 0, 0, 0)$;
$C(\mu_4) = (0, 0, 0, 0, 0, 0, 0, 0)$;
$C(\mu_5) = (0, 0, 0, 0, 0, -1, -1, 0)$.

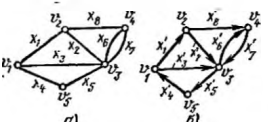


Рис. 4.29

Обозначим через $\mathbf{Z}^m{}_G$ множество всех вектор-циклов мультиграфа G.

Цикл μ называется *линейной комбинацией* циклов $\mu_1, ..., \mu_k$, если для некоторых $\alpha_1, ..., \alpha_k \in \mathbf{Q}$, где $\mathbf{Q}$ — множество рациональных чисел, выполняется равенство

$$C(\mu) = \alpha_1 C(\mu_1) + ... + \alpha_k C(\mu_k). \tag{4.28}$$

Вместо (4.28) будем использовать сокращенную запись

$$\mu = \alpha_1\mu_1 + ... + \alpha_k\mu_k. \tag{4.29}$$

Теорема 4.3 (*о независимости коэффициентов в линейной комбинации от ориентации ребер*). *Пусть для некоторой ориентации ребер выполняется соотношение (4.29). Тогда оно справедливо и для любой другой ориентации ребер.*

Запишем равенство (4.28) в координатной форме:

$$C_i(\mu) = C^+{}_i(\mu) - C^-{}_i(\mu) = \sum_{j=1}^{k} \alpha_j C_i(\mu_j) =$$
$$= \sum_{j=1}^{k} \alpha_j [C^+{}_i(\mu_j) - C^-{}_i(\mu_j)], \tag{4.30}$$

где $i \in \{1, ..., m\}$. При изменении ориентации ребра x_i из (4.30) будет следовать равенство

$$C^-{}_i(\mu) - C^+{}_i(\mu) = \sum_{j=1}^{k} \alpha_j [C^-{}_i(\mu_j) - C^+{}_i(\mu_j)]. \tag{4.31}$$

После умножения (4.31) на —1 получаем

$$C^+_i(\mu) - C^-_i(\mu) = \sum_{j=1}^{k} \alpha_j [C^+_i(\mu_j) - C^-_i(\mu_j)],$$

откуда

$$C_i(\mu) = \sum_{j=1}^{k} \alpha_j C_i(\mu_j),$$

т. е. равенство (4.30) имеет место и в этом случае.

Обозначим через $\mathbf{0}$ нулевой вектор в $\mathbf{Z}^m$ (т. е. $\mathbf{0} = (0, ..., 0) \in \in \mathbf{Z}^m$).

Система циклов $\{\mu_1, ..., \mu_k\}$ (где $k \geqslant 1$) называется *независимой*, если соответствующая ей система вектор-циклов линейно независима (т. е. если для любых $\alpha_1, ..., \alpha_k \in \mathbf{Q}$ из $\alpha_1 C(\mu_1) + ... + \alpha_k C(\mu_k) = 0$ следует $\alpha_1 = ... = \alpha_k = 0$). Используя теорему 4.3, заключаем, что выбор ориентации ребер не влияет на независимость системы циклов.

Отметим, что поскольку $\mathbf{Z}^m_G \subset \mathbf{Q}^m$, где $\mathbf{Q}^m$ — m-мерное линейное пространство над полем $\mathbf{Q}$, то максимальное количество элементов в независимой системе циклов не превышает m.

Замечание 4.39. Нетрудно видеть, что если $\{\mu_1, ..., \mu_k\}$ — независимая система циклов и цикл μ_{k+1} не является ее линейной комбинацией, то $\{\mu_1, ..., \mu_k, \mu_{k+1}\}$ — независимая система циклов.

4.3.5. Цикловой базис мультиграфа

Пусть мы находимся в условиях разд. 4.3.4. Независимая система циклов $\{\mu_1, ..., \mu_k\}$ называется *цикловым базисом* мультиграфа G, если любой цикл из G является линейной комбинацией циклов этой системы.

Рассмотрим вопрос о существовании циклового базиса мультиграфа G. Мультиграф G будем называть *ациклическим*, если либо $m = 0$, либо $\mathbf{Z}^m_G = \{0\}$ (т. е. в G отсутствуют простые циклы). Очевидно, что в ациклическом мультиграфе нет циклового базиса. С другой стороны, справедливо

Утверждение 4.41. *Если мультиграф G не является ациклическим, то в нем существует цикловой базис.*

Пусть μ_1 — произвольный цикл в G такой, что $C(\mu_1) \neq 0$. Тогда либо $\{\mu_1\}$ — цикловой базис, либо в G существует цикл μ_2 такой, что он не является линейной комбинацией цикла μ_1. В последнем случае (см. замечание 4.39) $\{\mu_1, \mu_2\}$ — независимая система циклов. При этом снова либо $\{\mu_1, \mu_2\}$ — цикловой базис, либо в G существует цикл μ_3 такой, что он не является линейной комбинацией циклов μ_1, μ_2. В последнем случае $\{\mu_1, \mu_2, \mu_3\}$ — независимая система циклов. В силу того, что количество элементов в независимой системе циклов не превышает m, указанный процесс последовательного ее расширения не является бесконеч-

ным и на некотором k-м шаге, где $k \leqslant m$, получаем систему циклов $\{\mu_1, \ldots, \mu_k\}$, являющуюся цикловым базисом мультиграфа G.

Рассмотрим теперь вопрос о количестве элементов в цикловом базисе мультиграфа G. Для этого нам понадобится.

Утверждение 4.42. *Пусть* $\{\mu_1, \ldots, \mu_k\}$, $\{\eta_1, \ldots, \eta_{k'}\}$ — *системы циклов мультиграфа* G *такие, что* $k' > k$, *каждый цикл из системы* $\{\eta_1, \ldots, \eta_{k'}\}$ *является линейной комбинацией циклов* $\mu_1, \ldots, \mu_k$. *Тогда система циклов* $\{\eta_1, \ldots, \eta_{k'}\}$ *не является независимой.*

Согласно условиям утверждения 4.42 для некоторых $\alpha_{ij} \in \mathbf{Q}$, где $i = 1, \ldots, k$, $j = 1, \ldots, k'$, имеем

$$\begin{aligned} \eta_1 &= \alpha_{11}\mu_1 + \ldots + \alpha_{1k}\mu_k; \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ \eta_{k'} &= \alpha_{k'1}\mu_1 + \ldots + \alpha_{k'k}\mu_k. \end{aligned} \tag{4.32}$$

Рассмотрим матрицу

$$A = \begin{bmatrix} \alpha_{11} & \ldots & \alpha_{1k} \\ \ldots & \ldots & \ldots \\ \alpha_{k'1} & \ldots & \alpha_{k'k} \end{bmatrix}$$

размерности $k' \times k$. Пусть r — ранг матрицы A, $A_i = (\alpha_{i1}, \ldots, \alpha_{ik})$, $i = 1, \ldots, k'$,— строки матрицы A. Тогда $r \leqslant k < k'$, а следовательно, некоторая строка в A (например, первая) является линейной комбинацией остальных строк, т. е.

$$A_1 = \alpha_2 A_2 + \ldots + \alpha_{k'} A_{k'}, \tag{4.33}$$

где $\alpha_2, \ldots, \alpha_{k'} \in \mathbf{Q}$ (с учетом того, что $A_1, \ldots, A_{k'} \in \mathbf{Q}^m$, где $\mathbf{Q}^m$ — линейное пространство над полем $\mathbf{Q}$). Из (4.32), (4.33) получаем

$$\eta_1 = \alpha_2\eta_2 + \ldots + \alpha_{k'}\eta_{k'},$$

т. е. система циклов $\eta_1, \ldots, \eta_{k'}$ не является независимой.

Из утверждения 4.42 следует, что количество элементов в цикловом базисе мультиграфа G — величина постоянная, не зависящая от выбора циклового базиса и равная максимальному числу элементов в независимых системах циклов мультиграфа G. Ниже (см. теорему 4.5) будет показано, что эта величина совпадает с геометрической характеристикой мультиграфа G, а именно: количество элементов циклового базиса равно $\nu(G) = m(G) - n(G) + p(G)$ (в частности, $\nu(G) = 0$ для ациклического мультиграфа G), при этом $\nu(G)$ называется *цикломатическим числом* мультиграфа G. Напомним, что ранее цикломатическое число было введено для связного мультиграфа G (т. е. при $p(G) = 1$).

В дальнейшем нам понадобится также следующее вспомогательное утверждение.

Утверждение 4.43. *Если* $M = \{\mu_1, ..., \mu_k\}$ — *цикловой базис мультиграфа* G, $H = \{\eta_1, ..., \eta_k\}$ — *система циклов мультиграфа* G *такая, что каждый цикл из* M *является линейной комбинацией циклов из* H, *то* H — *цикловой базис мультиграфа* G.

Заметим, что каждый цикл мультиграфа G является линейной комбинацией циклов из M, а значит, и циклов из H. Покажем теперь независимость системы H. Предположим, что эта система не является независимой. Тогда некоторый цикл η_i является линейной комбинацией остальных циклов из H. Пусть для определенности $i=k$. Тогда каждый цикл из M является линейной комбинацией циклов $\eta_1, \ldots \eta_{k-1}$, а следовательно, в силу утверждения 4.42 получаем, что система M не является независимой, а это противоречит исходным предположениям.

Покажем теперь, что если мультиграф G не является ациклическим, то в нем существует цикловой базис, состоящий из простых циклов. Предварительно докажем следующие простые утверждения.

Утверждение 4.44. *Пусть* μ — *цикл в мультиграфе* G, *в котором все вершины попарно различны. Тогда:*

1) *либо цикл* μ *имеет вид* $\mu = vxwxv$, *где* $x = \{v, w\}$, $v \neq w$, *и при этом* $C(\mu) = 0$;

2) *либо* μ — *простой цикл и при этом* $C(\mu) \neq 0$.

Если в цикле μ все ребра попарно различны, то выполняется утверждение 2. В противном случае пусть

$$\mu = v_1x_1v_2x_2 \ldots v_kx_kv_1, \tag{4.34}$$

где $k \geqslant 2$, и при некоторых i, j справедливо равенство $x_i = x_j$, где $x_i = \{v_i, v_{i+1}\}$, $x_j = \{v_j, v_{j+1}\}$, $1 \leqslant i < j \leqslant k$ (для общности обозначений в случае $j = k$ полагаем $v_{j+1} = v_{k+1} = v_1$). Предположим, что $j > i + 1$. Тогда согласно условиям утверждения 4.44 выполняется неравенство $v_{i+1} \neq v_j$, а следовательно, в силу равенства $x_i = x_j$ имеем $v_i = v_j$, $v_{i+1} = v_{j+1}$, что противоречит условиям доказываемого утверждения. Таким образом, $j = i + 1$, и тогда $x_i = \{v_i, v_{i+1}\}$, $x_j = x_{i+1} = \{v_{i+1}, v_{i+2}\}$, а значит, в силу равенства $x_i = x_j$ имеем $v_i = v_{i+2}$, откуда согласно условиям доказываемого утверждения и в силу (4.34) справедливо $i=1$, $i + 1 = k$, а следовательно, $k=2$, $\mu = v_1x_1v_2x_1v_1$, где $x_1 = \{v_1, v_2\}$, $v_1 \neq v_2$. При этом, очевидно, $C(\mu) = 0$.

Утверждение 4.45. *Пусть* μ — *цикл в мультиграфе* G, v — *произвольная вершина, содержащаяся в* μ. *Тогда* μ *является линейной комбинацией некоторых циклов мультиграфа* G, *в каждый из которых вершина* v *либо не входит, либо входит ровно один раз.*

Доказательство будем проводить индукцией по k — длине цикла μ. При $k=2$ всякая вершина, содержащаяся в цикле μ, входит в него ровно один раз. Предположим, что при некотором

$k \geqslant 3$ доказываемое утверждение справедливо для любого цикла длины $\leqslant k-1$. Докажем его для произвольного цикла $\mu = v_1x_1v_2x_2 \ldots v_kx_kv_1$ длины k. Пусть i — минимальный номер такой, что $v_i = v$. Разберем нетривиальный случай, когда вершина v_i входит в μ более одного раза. Пусть j — минимальный из номеров средн $i+2, \ldots, k$ такой, что $v_j = v_i$. Рассмотрим циклы (считаем, что $i > 1$; случай, когда $i = 1$, аналогичен)

$$\mu_1 = v_ix_i \ldots v_{i-1}x_{i-1}v_jx_j \ldots v_kx_kv_1;$$
$$\mu_2 = v_ix_i \ldots v_{j-1}x_{j-1}v_j.$$

Очевидно, что $\mu = \mu_1 + \mu_2$. При этом по условиям выбора номера j вершина $v = v_i$ встречается в μ_2 ровно один раз. Заметим, что длина цикла μ_1 равна $k + i - j$, где $k + i - j \leqslant k - 2$, а следовательно, по индуктивному предположению цикл μ_1 является линейной комбинацией некоторых циклов, в каждый из которых вершина v либо не входит, либо входит ровно один раз, откуда в силу $\mu = \mu_1 + \mu_2$ и вытекает справедливость доказываемого утверждения.

Используя утверждения 4.44, 4.45, покажем, что справедливо

Утверждение 4.46. *Пусть μ — цикл в мультиграфе G такой, что $C(\mu) \neq 0$. Тогда μ является линейной комбинацией простых циклов.*

Используя утверждение 4.45, получаем, что μ можно представить в виде линейной комбинации циклов, в каждом из которых все вершины попарно различны. Но тогда из утверждения 4.44 следует, что в указанную линейную комбинацию войдут простые циклы и μ является их линейной комбинацией.

Теперь докажем, что справедлива

Теорема 4.4. *Пусть G — мультиграф, не являющийся ациклическим. Тогда в G существует цикловой базис, элементами которого являются простые циклы.*

Согласно утверждению 4.41 в G существует цикловой базис $\{\mu_1, \ldots \mu_\nu\}$, где $\nu > 1$. Используя утверждение 4.46, получаем, что каждый цикл μ_i является линейной комбинацией некоторых простых циклов $\mu_{i1}, \ldots, \mu_{ik_i}$ $(i = 1, \ldots, \nu)$. Далее, используя процесс, описанный при доказательстве утверждения 4.41, выделяем из простых циклов μ_{ij}, $j = 1, \ldots, k_i$; $i = 1, \ldots, \nu$, независимую систему циклов такую, что любой цикл μ_{ij} является ее линейной комбинацией. Указанная система простых циклов, очевидно, будет цикловым базисом мультиграфа G.

Докажем теперь теорему о числе элементов в цикловом базисе мультиграфа.

Теорема 4.5. *Количество элементов в цикловом базисе мультиграфа $G = (V, X)$ совпадает с его цикломатическим числом $\nu(G) = m(G) - n(G) + p(G)$. В частности, для ациклического мультиграфа $\nu(G) = 0$.*

Доказательство проведем индукцией по количеству ребер в мультиграфе G. Обозначим через $\tilde{\nu}(G)$ количество элементов в цикловом базисе мультиграфа G, при этом, если G — ациклический мультиграф, то положим $\tilde{\nu}(G) = 0$. Покажем, что $\tilde{\nu}(G) = \nu(G)$.

Базис индукции. Если $m(G) = 0$, то, очевидно, выполняется равенство $p(G) = n(G)$, а следовательно, $\nu(G) = 0$. С другой стороны, поскольку $m(G) = 0$, имеем $\tilde{\nu}(G) = 0$, т. е. в этом случае равенство $\nu(G) = \tilde{\nu}(G)$ выполняется.

Индуктивный шаг. Предположим, что при некотором $m \geqslant 1$ доказываемое равенство справедливо для всякого мультиграфа с $m - 1$ ребрами. Докажем его справедливость и для произвольного мультиграфа G такого, что $m(G) = m$. Удалив из G произвольное ребро x (оставив без изменения вершины), получим мультиграф G' такой, что $m(G') = m - 1$. Возможны случаи: а) ребро x соединяет некоторые вершины v_1, v_2, принадлежащие различным компонентам связности мультиграфа G'; б) ребро x соединяет две различные вершины v_1, v_2, принадлежащие одной компоненте связности мультиграфа G'. В случае «а», очевидно, выполняется равенство $n(G') = n(G)$, $m(G') = m(G) - 1$, $p(G') = p(G) + 1$, а следовательно, $\nu(G') = \nu(G)$. Но тогда, если мы докажем, что

$$\tilde{\nu}(G) = \tilde{\nu}(G'), \tag{4.35}$$

то, воспользовавшись тем, что по индуктивному предположению $\tilde{\nu}(G') = \nu(G')$, получим $\tilde{\nu}(G) = \tilde{\nu}(G') = \nu(G') = \nu(G)$. Таким образом, осталось доказать равенство (4.35). Если G — ациклический мультиграф, то мультиграф G' тем более является ациклическим, и тогда $\tilde{\nu}(G) = 0 = \tilde{\nu}(G')$, т. е. равенство (4.35) выполняется. Пусть теперь мультиграф G не является ациклическим. Возьмем произвольный цикловой базис $\{\mu_1, \ldots, \mu_\nu\}$, где $\nu = \tilde{\nu}(G) \geqslant 1$, мультиграфа G, состоящий из простых циклов (см. теорему 4.4). Покажем, что $\mu_1, \ldots, \mu_\nu$ — циклы в G'. Предположим противное: пусть, например, μ_1 не является циклом в G'. Но тогда μ_1 проходит через ребро x, т. е., скажем, имеет вид

$$\mu_1 = v_1 x_1 v_2 x_2 \ldots v_k x_k v_1,$$

где $x = x_1 = \{v_1, v_2\}$. В силу того, что μ_1 — простой цикл, имеем $x_i \neq x$, $i = 2, \ldots, k$. Следовательно, $v_2 x_2 \ldots v_k x_k v_1$ — цепь в G', соединяющая вершины v_2, v_1, а это противоречит тому, что вершины v_1, v_2 принадлежат различным компонентам связности мультиграфа G'. Таким образом, $\{\mu_1, \ldots \mu_\nu\}$ — независимая система циклов мультиграфа G' такая, что любой цикл в G, а сле-

довательно, и в G' является линейной комбинацией циклов этой системы. Но тогда $\{\mu_1, \ldots, \mu_\nu\}$ — цикловой базис в G', откуда и следует справедливость равенства (4.35).

В случае «б» в мультиграфе G будет существовать простой цикл, проходящий через добавляемое ребро $x = \{v_1, v_2\}$. Действительно, в силу того, что вершины v_2, v_1 принадлежат одной компоненте связности мультиграфа G', их можно соединить маршрутом в G', из которого в свою очередь можно выделить простую цепь $\theta = v_2 x_2 \ldots x_k v_1$ (см. утверждение 4.4). Поскольку θ — простая цепь в G', получаем $x_i \neq x$, $i=2, \ldots, k$. Но тогда $\mu = v_1 x v_2 x_2 \ldots v_k x_k v_1$ — простой цикл в G. Рассмотрим случай, когда мультиграф G' не является ациклическим (если G' — ациклический мультиграф, то рассуждаем аналогично). Пусть $\{\mu_1, \ldots, \mu_{\nu'}\}$, где $\nu' = \nu(G')$, — цикловой базис в G'. Так как в цикл μ входит (и притом один раз) ребро x, которое не содержится ни в одном из циклов $\mu_1, \ldots, \mu_{\nu'}$, то цикл μ не является их линейной комбинацией, а следовательно (см. замечание 4.39), $\{\mu_1, \ldots, \mu_{\nu'}, \mu\}$ — независимая система циклов. Покажем, что $\{\mu_1, \ldots, \mu_{\nu'}, \mu\}$ — цикловой базис в G. Для этого докажем, что произвольный цикл $\tilde{\mu}$ в G является линейной комбинацией циклов $\mu_1, \ldots, \mu_{\nu'}, \mu$. Согласно утверждению 4.46 достаточно рассмотреть случай, когда $\tilde{\mu}$ является простым циклом. Если ребро x не входит в $\tilde{\mu}$, то $\tilde{\mu}$ — цикл в G', а следовательно, он является линейной комбинацией циклов $\mu_1, \ldots, \mu_{\nu'}$. Пусть теперь ребро x входит в $\tilde{\mu}$. В силу того, что $\tilde{\mu}$ — простой цикл, ребро x входит в $\tilde{\mu}$ один раз. Без ограничения общности можно считать, что цикл $\tilde{\mu}$ проходит через ребро x в том же направлении, что и цикл μ, т. е. имеет вид $\tilde{\mu} = v_1 x v_2 y_1 w_1 \ldots y_l w_l$, где $l \geqslant 2$, $y_i \neq x$, $i=1, \ldots, l$, $w_l = v_1$. Рассмотрим цикл

$$\eta = v_2 y_1 w_1 \ldots y_l w_l x_k v_k \ldots x_2 v_2$$

(см. рис. 4.30). Очевидно, что η — цикл в G', так как он не проходит через ребро x, а следовательно, η является линейной комбинацией циклов $\mu_1, \ldots, \mu_{\nu'}$. Заметим далее, что $\eta = \tilde{\mu} - \mu$, а значит, цикл $\tilde{\mu} = \eta + \mu$ является линейной комбинацией циклов $\mu_1, \ldots, \mu_{\nu'}, \mu$. Таким образом, $\{\mu_1, \ldots, \mu_{\nu'}, \mu\}$ — цикловой базис в G, а следовательно, $\tilde{\nu}(G) = \nu' + 1 = \tilde{\nu}(G') + 1$. Воспользовавшись тем, что по индуктивному предположению выполняется равенство $\tilde{\nu}(G') = \nu(G') = m(G') - n(G') + p(G')$, а также тем, что в нашем случае справедливо $n(G') = n(G)$, $m(G') = m(G) - 1$,

$p(G') = p(G)$, имеем $\nu(G) = \nu(G') + 1$. Но тогда $\tilde{\nu}(G) = \tilde{\nu}(G') + 1 = \nu(G') + 1 = \nu(G)$. Теорема доказана.

Рассмотрим теперь задачу о практическом нахождении циклового базиса мультиграфа. Предварительно сведем ее к задачам нахождения цикловых базисов для связных мультиграфов.

Теорема 4.6. *Пусть $G = (V, X)$ — мультиграф, не являющийся ациклическим, G_1, ..., G_p (где $p = p(G)$) — мультиграфы, являющиеся компонентами связности мультиграфа G. Пусть также $\forall i \in \{1, ..., p\}$ $M_i = \{\mu_{i1}, ..., \mu_{i\nu_i}\}$ (где $\nu_i = \nu(G_i)$) — цикловой базис мультиграфа G_i (при этом, если G_i — ациклический мультиграф, то $M_i = \varnothing$). Тогда $M = \{\mu_{ij}, j = 1, ..., \nu_i; i = 1, ..., p\}$ — цикловой базис мультиграфа G.*

Пусть μ — произвольный цикл из G. Докажем, что μ является линейной комбинацией циклов из M. Поскольку любые две различные вершины, входящие в μ, можно соединить маршрутом, то μ является циклом в некоторой компоненте связности G_i. Но тогда μ является линейной комбинацией циклов из M_i, а так как $M_i \subseteq M$, то и циклов из M. Подсчитаем теперь количество циклов в M. Используя теорему 4.5, а также утверждение 4.10, имеем

$$|M| = \sum_{i=1}^{p} |M_i| = \sum_{i=1}^{p} \nu(G_i) = \sum_{i=1}^{p} [m(G_i) - n(G_i) + 1] = m(G) - n(G) + p(G) = \nu(G),$$

откуда согласно утверждению 4.43 получаем, что M — цикловой базис мультиграфа G.

Из доказанной теоремы следует, что задача нахождения циклового базиса произвольного мультиграфа сводится к конечной совокупности задач нахождения циклового базиса связного мультиграфа. Опишем алгоритм решения этой задачи.

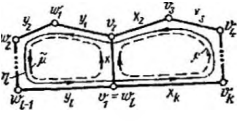


Рис. 4.30

Алгоритм 4.8 нахождения циклового базиса связного мультиграфа G. Если $\nu(G) = 0$, то G — ациклический мультиграф, а следовательно, циклового базиса не существует. Пусть теперь $\nu(G) > 0$. Выделим в G остовное дерево T. Пусть $n = n(G)$, $m = m(G)$, x_1, ..., x_{n-1} — ребра в T, а x_n, ..., x_m — остальные ребра мультиграфа G (заметим, что поскольку $\nu(G) > 0$, очевидно, выполняются неравенства $n \geqslant 2$, $m \geqslant n$). Число последних ребер равно $m - n + 1$, т. е. совпадает с $\nu(G)$. Добавляя любое из ребер x_i, $i = n$, ..., m, к дереву T, получаем некоторый подграф мультиграфа G, из которого выделяем простой цикл $\mu_{i-(n-1)}$ (см. ниже замечание 4.40), проходящий через добавляемое ребро x_i. Действуя таким образом, находим совокупность простых циклов $\{\mu_1, ..., \mu_{m-n+1}\}$. Поскольку в каждом из циклов этой системы имеется ребро, не содержащееся в других циклах, то полученная

система циклов является независимой. Итак, мы нашли независимую систему циклов с $\nu(G)$ элементами, которая является цикловым базисом мультиграфа G. Действительно, предположив, что некоторый цикл μ мультиграфа G не является линейной комбинацией циклов $\mu_1, ..., \mu_{\nu(G)}$, с учетом замечания 4.39 заключаем, что система циклов $\{\mu_1, ..., \mu_{\nu(G)}, \mu\}$ является линейно независимой, а это противоречит тому, что $\nu(G)$ — максимальное число элементов в независимых системах циклов мультиграфа G.

Замечание 4.40. Согласно утверждению 4.40 после добавления к T ребра x_i, где $i \in \{n, ..., m\}$, получаем мультиграф G_i, в котором существует единственный (с точностью до направления обхода и начальной вершины обхода) простой цикл, проходящий через добавляемое ребро x_i.

Пример 4.31. Используя алгоритм 4.8, определим цикловой базис мультиграфа G, изображенного на рис. 4.31, *а*. Имеем $\nu(G) = m(G) - n(G) + p(G) = 8 - 4 + 1 = 5 > 0$, т. е. цикловой базис мультиграфа G существует. Выделим произвольно остовное дерево T мультиграфа G (см. на рис. 4.31, *б* изображение T, где пунктирными линиями указаны ребра, удаленные из G для выделения дерева T). При выделении остовного дерева T из связного мультиграфа G было удалено всего $\nu(G) = 5$ ребер: x_3, x_4, x_6, x_7, x_8. Добавляя поочередно к T каждое из перечисленных ребер и выделяя из получаемого таким образом мультиграфа простой цикл, имеем циклы

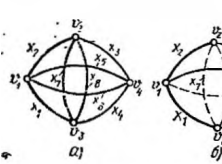


Рис. 4.31

$$\mu_1 = v_1x_2v_2x_3v_4x_5v_1;$$
$$\mu_2 = v_1x_6v_4x_4v_3x_1v_1;$$
$$\mu_3 = v_1x_6v_4x_8v_1;$$
$$\mu_4 = v_1x_2v_2x_7v_3x_1v_1;$$
$$\mu_5 = v_1x_2v_2x_8v_3x_1v_1,$$

составляющие цикловой базис мультиграфа G.

Замечание 4.41. Как следует из описания алгоритма 4.8, цикловой базис, получаемый в результате применения этого алгоритма к произвольному связному мультиграфу G, состоит из простых циклов.

4.3.6. Цикломатическая матрица мультиграфа

Пусть G — мультиграф, не являющийся ациклическим (т. е. $\nu(G) > 0$). Матрица $C(G)$ размерности $\nu(G) \times m(G)$, строками которой являются вектор-циклы циклового базиса мультиграфа G, называется *цикломатической матрицей* мультиграфа G.

Пример 4.32. Определим цикломатическую матрицу для мультиграфа G (см. пример 4.31). Введем ориентацию на ребрах мультиграфа G. В результате получим, например, ориентированный псевдограф, изображенный на рис. 4.32. Выпишем вектор-циклы, соответствующие циклам из циклового базиса мультиграфа G, найденного в примере 4.31:

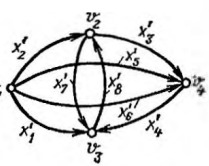


Рис. 4.32

$C(\mu_1) = (0, 1, 1, 0, -1, 0, 0, 0);$
$C(\mu_2) = (-1, 0, 0, 1, 1, 0, 0, 0);$
$C(\mu_3) = (0, 0, 0, 0, 1, -1, 0, 0);$
$C(\mu_4) = (-1, 1, 0, 0, 0, 0, 1, 0);$
$C(\mu_5) = (-1, 1, 0, 0, 0, 0, 0, -1).$

Тогда

$$C(G) = \left[\begin{array}{cccc|c|ccc} 0 & 1 & 1 & 0 & -1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 & 0 & -1 \end{array}\right].$$

Замечание 4.42. Из определения цикломатической матрицы мультиграфа G следует, что ранг матрицы $C(G)$ равен $\nu(G)$, т. е. он совпадает с количеством строк в $C(G)$.

Замечание 4.43. Если $\{\mu_1, ..., \mu_\nu\}$ — цикловой базис связного мультиграфа G, найденный по алгоритму 4.8, и $C(G)$ — цикломатическая матрица мультиграфа G, строками которой являются вектор-циклы $C(\mu_i)$, $i=1, 2, ..., \nu$, то из столбцов матрицы $C(G)$, соответствующих ребрам, не вошедшим в остовное дерево T, можно составить диагональную квадратную матрицу порядка $\nu(G)$, элементы главной диагонали которой принадлежат множеству $\{-1, 1\}$ (см. выделенные пунктирными линиями столбцы матрицы $C(G)$ из примера 4.32). Этот очевидный факт является следствием того, что каждый простой цикл μ_i имеет ребро, не входящее в другие циклы, а следовательно, соответствующий этому ребру столбец в матрице $C(G)$ содержит отличный от нуля элемент в i-й позиции (т. е. либо 1, либо -1), а все остальные элементы этого столбца равны 0.

4.3.7. Уравнения Кирхгофа для напряжений

Пусть имеется некоторая электрическая цепь S, представляющая собой набор двухполюсных элементов $a_1, ..., a_m$ (например, сопротивлений, конденсаторов, индуктивностей, источников ЭДС и

т. д.), соединенных проводниками. Поставим в соответствие электрической цепи S мультиграф $G = G(S)$. Для этого каждому j-му узлу соединений элементов цепи поставим в соответствие вершину v_j, а каждому элементу a_i — ребро x_i, соединяющее соответствующие вершины.

Введем произвольно ориентацию на ребрах мультиграфа G. В результате на мультиграфа G получим ориентированный мультиграф D с дугами $x'_1, \ldots, x'_m$. Для каждого номера $i=1, \ldots, m$ обозначим через I_i величину ток, проходящего через элемент a_i, а через U_i — напряжение между полюсами элемента a_i. Поскольку направление тока заранее предсказать не всегда возможно, то введенную ориентацию на ребрах мультиграфа G будем рассматривать как условные направления токов. Тогда после определения I_i знак этой величины подскажет нам истинное направление тока по элементу a_i (где $i \in \{1, \ldots, m\}$). Аналогично знак величины U_i также будет определяться выбранной ориентацией на ребре x_i, а именно: под U_i будем понимать величину, получаемую вычитанием из потенциала полюса элемента a_i, соответствующего началу дуги x'_i, потенциала полюса, соответствующего концу дуги x'_i (где $i \in \{1, \ldots, m\}$). Кроме того, обозначим $I = (I_1, \ldots, I_m)$, $U = (U_1, \ldots, U_m)$ — векторы токов и напряжений в электрической цепи S. Для любых векторов $A = (A_1, \ldots, A_m)$, $B = (B_1, \ldots, B_m)$ из $\mathbb{R}^m$ введем обозначение $(A, B) = \sum_{i=1}^{m} A_i B_i$ — скалярное произведение векторов A, B.

Если μ — произвольный цикл мультиграфа G, то согласно закону Кирхгофа для напряжений имеем $(C(\mu), U) = 0$. Но тогда

$$\forall C \in \mathbf{Z}^m_G \ (C, U) = 0. \tag{4.36}$$

Уравнения (4.36) относительно U называются *уравнениями Кирхгофа для напряжений*. При этом систему уравнений

$$C(G)\,U = 0 \tag{4.37}$$

будем называть *базисной системой* уравнений Кирхгофа для напряжений. Поскольку ранг матрицы $C(G)$ равен количеству уравнений в (4.37) (см. замечание 4.42), то они *линейно независимы* (т. е. ни одно из уравнений системы (4.37) не является линейной комбинацией других уравнений этой системы). С другой стороны, по определению $C(G)$ любое уравнение из (4.36) является линейной комбинацией уравнений базисной системы (4.37). Из указанных свойств базисной системы уравнений следует, что для расчета электрической цепи удобнее всего пользоваться базисной системой уравнений Кирхгофа для напряжений.

Пример 4.33. Определим базисную систему уравнений Кирхгофа для напряжений для электрической цепи S, схема которой представлена на рис. 4.33, *а*.

Электрической цепи S соответствует граф G, изображенный на рис. 4.33,*б*, где ребро x_i соответствует сопротивлению r_i при $i = 1, 2, 3, 4, 5$, а ребро x_6 — источнику ЭДС. Введем ориентацию на ребрах графа G. В результате получим, например, орграф D, показанный на рис. 4.33,*в*. Выделим остовное дерево T графа G (см. изображение T на рис. 4.33, *г*). Используя алгоритм 4.8, определим цикловой базис графа G. При выделении остовного дерева T из графа G было удалено всего три ребра: x_2, x_4, x_5, поскольку $\nu(G) = 3$ (они показаны пунктирными линии-

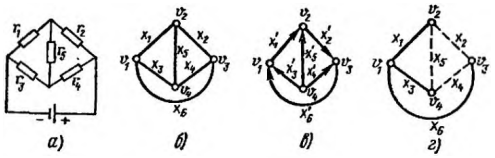


Рис. 4.33

ями на рис. 4.33, *г*). Добавляя поочередно к T каждое из перечисленных ребер и выделяя из получаемого таким образом графа простой цикл, имеем циклы

$$\mu_1 = v_1x_1v_2x_2v_3x_6v_1;$$
$$\mu_2 = v_1x_3v_4x_4v_3x_6v_1;$$
$$\mu_3 = v_1x_1v_2x_5v_4x_3v_1,$$

составляющие цикловой базис графа G. Выпишем соответствующие им вектор-циклы:

$$C(\mu_1) = (1, 1, 0, 0, 0, 1);$$
$$C(\mu_2) = (0, 0, -1, 1, 0, 1);$$
$$C(\mu_3) = (1, 0, 1, 0, -1, 0).$$

Тогда

$$C(G) = \left[\begin{array}{cc|cc|cc} 1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & -1 & 0 \end{array}\right],$$

и при этом базисная система уравнений Кирхгофа для напряжений $C(G)U=0$ имеет вид

$$U_1 + U_2 + U_6 = 0;$$
$$-U_3+U_4+U_6=0;$$
$$U_1+U_3-U_5=0.$$

Заметим, что столбцы никломатической матрицы $C(G)$, соответствующие ребрам x_2, x_4, x_5, не вошедшим в T (они выделены пунктирными линиями), образуют диагональную квадратную матрицу порядка $\nu(G)$, элементы главной диагонали которой принадлежат множеству $\{-1, 1\}$ (см. замечание 4.43), а следовательно, переменные U_2, U_4, U_5, соответствующие ребрам x_2, x_4, x_5, легко выражаются через остальные переменные:

$$\begin{aligned} U_2 &= -U_1 - U_6; \\ U_4 &= U_3 - U_6; \\ U_5 &= U_1 + U_3. \end{aligned} \qquad (4.38)$$

Переменные U_2, U_4, U_5 в системе уравнений (4.38) называются *базисными*, а U_1, U_3, U_6 — *свободными*.

Замечание 4.44. В ряде случаев при расчете электрических цепей, используя выражение базисных переменных через свободные, мы можем существенно сократить общее число неизвестных. Такая возможность имеется, например, в случае, когда элементами электрической цепи являются лишь источники ЭДС и сопротивления (см. задачу 3.10).

4.3.8. Уравнения Кирхгофа для токов

Для расчета электрической цепи одних уравнений Кирхгофа для напряжений недостаточно. Рассмотрим *уравнения Кирхгофа для токов*. Пусть, как и ранее, электрической цепи S поставлен в соответствие мультиграф G, которому в свою очередь после введения ориентации на ребрах соответствует некоторый ориентированный мультиграф D. Уравнения Кирхгофа для токов составляют относительно каждой вершины ориентированного мультиграфа D. Эти уравнения математически выражают тот факт, что при установившемся процессе в электрической цепи S сумма входящих в некоторый узел этой цепи токов равна сумме токов, выходящих из него. Но тогда с учётом условных направлений токов, определяемых ориентированным мультиграфом D, система всех уравнений Кирхгофа для токов в цепи S имеет вид

$$B(D)I = \mathbf{0},$$

где $B(D)$ — матрица инцидентности ориентированного мультиграфа D.

Пример 4.34. Определим совокупность уравнений Кирхгофа для токов в электрической цепи, приведенной в примере 4.33 (см. рис. 4.33, *а*).

В соответствии с выбором орграфа D, описанным в примере 4.33 (см. рис. 4.33, *в*), получаем

$$B(D) = \begin{bmatrix} -1 & 0 & 1 & 0 & 0 & 1 \\ 1 & -1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & -1 \\ 0 & 0 & -1 & -1 & -1 & 0 \end{bmatrix}.$$

Тогда система уравнений Кирхгофа для токов $B(D)I = 0$ имеет вид

$$\begin{aligned} -I_1 + I_3 + I_6 &= 0; \\ I_1 - I_2 + I_5 &= 0; \\ I_2 + I_4 - I_6 &= 0; \\ -I_3 - I_4 - I_5 &= 0. \end{aligned}$$

Из определения матрицы $B(D)$ следует, что для произвольного ориентированного мультиграфа D сумма всех строк матрицы $B(D)$ дает нулевую строку, а следовательно, любая строка матрицы $B(D)$ является линейной комбинацией остальных строк. Таким образом, из системы уравнений Кирхгофа $B(D)I=0$ можно исключить любое уравнение и получить при этом систему, равносильную исходной, поскольку исключенное уравнение является линейной комбинацией оставшихся. Возникает вопрос: будет ли после исключения одного уравнения из $B(D)I=0$ оставшаяся система уравнений линейно независимой? Ниже будет показано, что если $G=G(S)$ — связный мультиграф, то исключение одного уравнения из системы $B(D)I=0$ уравнений Кирхгофа для токов в электрической цепи S дает линейно независимую систему уравнений. Для доказательства этого факта нам потребуются следующие утверждения.

Утверждение 4.47. *Пусть $T=(V, X)$ — дерево, где $V=\{v_1, \ldots, v_n\}$, $X=\{x_1, \ldots, x_{n-1}\}$, $n\geqslant 2$. Тогда перенумерацией ребер и вершин в T всегда можно добиться того, чтобы подматрица матрицы $B(T)$, являющаяся результатом исключения из $B(T)$ первой строки, была квадратной матрицей порядка $n-1$ треугольного вида с нулями под главной диагональю и с единицами на главной диагонали.*

Укажем правила перенумерации вершин и ребер в T:

1) разбиваем множество вершин V на непустые подмножества $W_i(v_1)=\{v\in V \mid d(v, v_1)=i\}$, $i=0, 1, \ldots, k$, где k — последний номер такой, что $W_k(v_1)\neq\varnothing$;

2) нумеруем вершины из V: начинаем с вершины из $W_0(v_1)$, затем переходим к вершинам из $W_1(v_1)$ и т. д. и заканчиваем вершинами из $W_k(v_1)$;

3) ставим в соответствие каждой вершине v_{j+1}, где $j\in\{1, \ldots, n-1\}$, ребро x_j (относительно новой нумерации ребер) по следующему правилу. Пусть η — простая цепь в T, соединяющая v_1 с v_{j+1}, v_i — вершина, непосредственно предшествующая вершине v_{j+1} в этой цепи. В силу правила 2 имеем $i<j+1$. Обозначим $x_j=\{v_i, v_{j+1}\}$. Действуя таким образом, занумеруем $n-1$ ребро дерева T, т. е. все его ребра (очевидно, что мы не можем одно и то же ребро пронумеровать более одного раза).

В результате перенумерации вершин и ребер дерева T получаем новое дерево T', изоморфное T, такое, что каждый j-й столбец матрицы $B(T')$ (где $1<j<n-1$) имеет единицу в $(j+1)$-й строке, а также в одной из предшествующих строк (см. правило 3). Но тогда после удаления первой строки из матрицы $B(T')$ получаем квадратную матрицу порядка $n-1$ треугольного вида с нулями под главной диагональю и с единицами на главной диагонали.

Пример 4.35. На рис. 4.34, *а* изображено дерево T, а на рис. 4.34, *б* — дерево T', полученное из T перенумерацией ребер и вершин согласно правилам 1—3. В табл. 4.10*а* и 4.10*б* приведе-

ны соответственно $B(T)$ и $B(T')$. При этом, если удалим из $B(T')$ первую строку, то получим квадратную матрицу треугольного вида с нулями под главной диагональю и единицами на главной диагонали.

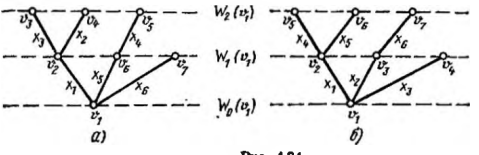


Рис. 4.34

Из утверждения 4.47, а также из того очевидного факта, что если G — мультиграф, D — ориентированный мультиграф, получаемый из G введением ориентации на ребрах, то элементы матрицы $B(G)$ являются абсолютными величинами соответствующих элементов матрицы $B(D)$, следует, что справедливо

Утверждение 4.48. *Пусть T — дерево с $n>2$ вершинами, D — орграф, получаемый из T введением ориентации на ребрах дерева T. Тогда перенумерацией вершин и дуг в D всегда можно добиться того, чтобы подматрица матрицы $B(D)$, являющаяся ре-*

Таблица 4.10а

	x_1	x_2	x_3	x_4	x_5	x_6
v_4	1	0	0	0	1	1
v_3	1	1	1	0	0	0
v_2	0	0	1	0	0	0
v_4	0	1	0	0	0	0
v_5	0	0	0	1	0	0
v_6	0	0	0	1	1	0
v_1	0	0	0	0	0	1

Таблица 4.10б

	x_1	x_2	x_3	x_4	x_5	x_6
v_1	1	1	1	0	0	0
v_2	1	0	0	1	1	0
v_3	0	1	0	0	0	1
v_4	0	0	1	0	0	0
v_5	0	0	0	1	0	0
v_6	0	0	0	0	1	0
v_7	0	0	0	0	0	1

зультатом исключения из $B(D)$ первой строки, была квадратной матрицей порядка $n-1$ треугольного вида с нулевыми элементами под главной диагональю и элементами из множества $\{-1, 1\}$ на главной диагонали.

Из утверждения 4.48 с учетом того, что ранг матрицы не меняется от перестановки строк и столбцов, следует, что справедливо

Утверждение 4.49. *Пусть T — дерево с $n \geqslant 2$ вершинами, D — орграф, полученный из T введением ориентации на ребрах дерева T. Тогда rank $B(D) = n-1$ (где rank $B(D)$ — ранг матрицы $B(D)$).*

Покажем теперь, что справедливо

Утверждение 4.50. *Пусть G — связный мультиграф, D — ориентированный мультиграф, полученный из G введением ориентации на ребрах мультиграфа G. Тогда:*

1) *rank $B(D) = n-1$;*

2) *исключение из матрицы $B(D)$ произвольной строки дает матрицу ранга $n-1$.*

Докажем сначала справедливость первого утверждения. Поскольку, как уже отмечалось ранее, сумма всех строк матрицы $B(D)$ дает нулевую строку, имеем

$$\operatorname{rank} B(D) \leqslant n-1. \tag{4.39}$$

С другой стороны, согласно утверждению 4.49 ранг подматрицы матрицы $B(D)$, составленной из столбцов матрицы $B(D)$, номера которых соответствуют номерам ребер мультиграфа G, содержащихся в некотором остовном дереве мультиграфа G, равен $n-1$, а следовательно, rank $B(D) \geqslant n-1$, откуда с учетом (4.39) получаем требуемое равенство.

Докажем теперь справедливость второго утверждения. Пусть $B'(D)$ — матрица, полученная из $B(D)$ после исключения некоторой i-й строки. В этом случае $\operatorname{rank} B'(D) \leqslant \operatorname{rank} B(D) = n-1$. Предположим что $\operatorname{rank} B'(D) < n-1$. Тогда некоторая j-я строка матрицы $B(D)$, содержащаяся в $B'(D)$ (т. е. при $j \neq i$), является линейной комбинацией остальных строк матрицы $B'(D)$. Таким образом, строки матрицы $B(D)$ с номерами i, j являются линейными комбинациями остальных строк этой матрицы, а следовательно, $\operatorname{rank} B(D) \leqslant n-2$, что противоречит первому утверждению.

Из утверждения 4.50 непосредственно следует, что если G — связный мультиграф (соответствующий некоторой электрической цепи S) и D — ориентированный мультиграф, полученный из G введением ориентации на ребрах мультиграфа G, то исключение из $B(D)I = 0$ любого уравнения дает линейно независимую систему, и при этом исключенное уравнение является линейной комбинацией оставшихся.

Задачи и упражнения

1. Показать, что у дерева G с $n(G) \geqslant 2$ найдутся, по крайней мере, две висячие вершины.

2. Показать, что дерево, содержащее ровно две висячие вершины, является простой цепью.

3. Определить любое остовное дерево графа G. Найти цикломатическое число графа G. Варианты графа G представлены на рис. 4.35, *а, б*.

4. Определить минимальное остовное дерево нагруженного графа, изображенного на рис. 4.36.

5. Пусть G — мультиграф, изображенный на рис. 4.29, *а*, и в G введена ориентация на ребрах (см. рис. 4.29, *б*). Определить вектор-циклы, соответствующие следующим циклам:

а) $\mu_1 = v_1x_3v_3x_5v_5x_4v_1x_1v_2x_8v_4x_6v_3x_5v_5x_4v_1$;

б) $\mu_2 = v_4x_8v_2x_1v_1x_4v_5x_8v_2x_2v_2x_1v_1x_4v_5x_5v_3x_7v_4$.

6. Найти в мультиграфе G (см. задачу 5) цикловой базис. Определить цикломатическую матрицу.

7. Определить цикловые базисы для графов (см. задачу 3). Введя произвольно ориентацию на ребрах, определить цикломатические матрицы для этих графов.

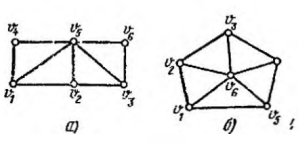


Рис. 4.35

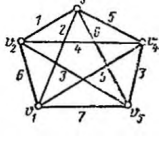


Рис. 4.36

8. Составить базисные системы уравнений Кирхгофа для напряжений в электрических цепях, схемы которых представлены на рис. 4.37, *а, б*.

9. Составить систему уравнений Кирхгофа для токов в электрических цепях (см. задачу 8).

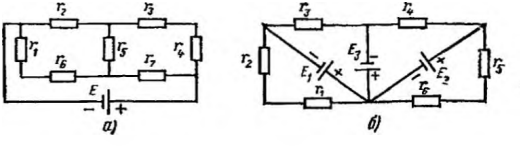


Рис. 4.37

10. В условиях задач 8, 9, используя закон Ома, составить системы уравнений для токов в электрических цепях. Считать внутренние сопротивления источников ЭДС равными r, где $r>0$.

4.4. ВНУТРЕННЯЯ И ВНЕШНЯЯ УСТОЙЧИВОСТЬ В ГРАФАХ

4.4.1. Внутренняя устойчивость в ориентированных графах

Пусть задан орграф $D=(V, X)$. Множество $U\subseteq V$ называется *внутренне устойчивым*, если

$$\forall v\in U \quad U\cap D(v)=\varnothing, \tag{4.40}$$

т. е. в орграфе D не существует дуги, связывающей какие-либо две вершины из U.

Пример 4.36. Для орграфа D, изображенного на рис. 4.38, множества $U_1=\{v_1\}$, $U_2=\{v_1, v_3\}$, $U_3=\{v_2, v_4\}$ являются внутренне устойчивыми; множество $U_4=\{v_1, v_2\}$ не является внутренне устойчивым, так как в D имеется дуга (v_1, v_2).

Внутренне устойчивое множество $U\subseteq V$ называется *максимальным*, если, добавляя к U любую вершину $v\in V\setminus U$, получаем множество, не являющееся внутренне устойчивым. Часто при решении практических задач требуется найти внутренне устойчивые множества с максимальным числом вершин. Их следует искать среди максимальных внутренне устойчивых множеств.

Замечание 4.45. Внутренне устойчивые множества, а также максимальные внутренне устойчивые множества аналогично определяются и для ориентированного псевдографа. При этом, если в ориентированном псевдографе удалить инцидентные петлям вершины вместе со всеми инцидентными им дугами и принять кратности дуг равными 1, то в полученном таким образом орграфе внутренне устойчивые множества (а следовательно, и максимальные внутренне устойчивые множества) совпадут с внутренне устойчивыми множествами (соответственно, с максимальными внутренне устойчивыми множествами) исходного ориентированного псевдографа, поэтому в дальнейшем будем рассматривать лишь орграфы без петель и кратных дуг.

Рис. 4.38

Пример 4.37. Воспользуемся множествами U_1, U_2, U_3, приведенными в примере 4.36. Тогда множество $U_1=\{v_1\}$ не является максимальным внутренне устойчивым, так как, добавляя к U_1 вершину v_3, получаем множество U_2, снова являющееся внутренне устойчивым. Напротив, множества U_2, U_3 являются максимальными внутренне устойчивыми.

Обозначим через $\sigma(D)$ совокупность всех максимальных внутренне устойчивых множеств вершин орграфа D. Тогда число

$$\alpha(D) = \max_{U \in \sigma(D)} |U|$$

называется *числом внутренней устойчивости орграфа* D.

4.4.2. Метод Магу отыскания семейства максимальных внутренне устойчивых множеств

Пусть $D = (V, X)$ — орграф, где $X \neq \varnothing$, $V = \{v_1, ..., v_n\}$. Пусть далее U — некоторое множество вершин орграфа D, т. е. $U \subseteq V$. Введем булевы переменные Y_i, соответствующие вершинам v_i, $i = 1, 2, ..., n$. Рассмотрим оценку $<\varepsilon_1, ..., \varepsilon_n>$ списка переменных $<Y_1, ..., Y_n>$, где

$$\forall i \in \{1, ..., n\} \quad \varepsilon_i = \begin{cases} 1, & \text{если } v_i \in U; \\ 0, & \text{если } v_i \notin U. \end{cases} \tag{4.41}$$

Про оценку $<\varepsilon_1, ..., \varepsilon_n>$ будем говорить, что она соответствует множеству U и, наоборот, что множество U соответствует указанной оценке. Заметим, что условие (4.40) эквивалентно тому, что $\forall i, j \in \{1, 2, ..., n\}$ выполняется

$$\{v_i \in U,\ v_j \in D(v_i)\} \Rightarrow v_j \in U,$$

а в силу (4.41) это равносильно тому, что

$$\forall i, j \in \{1, 2, ..., n\} \quad (\varepsilon_i \& a_{ij}) \supset \overline{\varepsilon_j} = 1, \tag{4.42}$$

где a_{ij} — (i, j)-й элемент матрицы смежности $A(D)$.

Преобразуя (4.42), получаем

$$\forall i, j \in \{1, 2, ..., n\} \quad \overline{a_{ij}} \vee \overline{\varepsilon_i} \vee \varepsilon_j = 1. \tag{4.43}$$

Условие (4.43) можно записать в виде

$$\bigwedge_{i=1}^{n} \bigwedge_{j=1}^{n} (\overline{a_{ij}} \vee \overline{\varepsilon_i} \vee \varepsilon_j) = 1. \tag{4.44}$$

Заметим, что при $a_{ij} = 0$ заведомо выполняется равенство $\overline{a_{ij}} \vee \overline{\varepsilon_i} \vee \varepsilon_j = 1$, а при $a_{ij} = 1$ справедливо $\overline{a_{ij}} \vee \overline{\varepsilon_i} \vee \varepsilon_j = \overline{\varepsilon_i} \vee \varepsilon_j$. Но тогда, если под записью «$a_{ij} = 1$» понимать множество всех пар $<i, j>$ таких, что $i, j \in \{1, 2, ..., n\}$, $a_{ij} = 1$ (поскольку $X \neq \varnothing$, это множество не является пустым), то условие (4.44) можно переписать в виде

$$\bigwedge_{a_{ij}=1} (\overline{\varepsilon_i} \vee \varepsilon_j) = 1,$$

или, обозначив

$$F(Y_1, ..., Y_n) = \bigwedge_{a_{ij}=1} (\overline{Y_i} \vee Y_j),$$

в виде

$$F(\varepsilon_1, ..., \varepsilon_n) = 1. \tag{4.45}$$

Таким образом, мы показали, что справедливо

Утверждение 4.51. *Необходимым и достаточным условием внутренней устойчивости множества $U \subseteq V$ является выполнение равенства (4.45), где $<\varepsilon_1, ..., \varepsilon_n>$ удовлетворяет (4.41).*

Из утверждения 4.51 следует, что внутренне устойчивые множества вершин орграфа D и только они соответствуют оценкам списка переменных $<Y_1, ..., Y_n>$, на которых выполняется равенство $F = 1$, что дает нам простой способ выделения всех внутренне устойчивых множеств.

Опишем теперь *метод Магу* выделения максимальных внутренне устойчивых множеств вершин орграфа D. Применяя обобщенную дистрибутивность & относительно $\vee$, приводим формулу логики высказываний F к ДНФ, а затем сокращаем ее до тех пор, пока это возможно, используя равносильности (см. разд. 1.2)

$$A = A \vee (A \& B),\ A \vee A = A,\ A \& A = A, \tag{4.46}$$

где A, B — произвольные формулы логики высказываний (докажите, что возможно лишь конечное число таких сокращений). В результате получаем формулу F_1 (равносильную F), находящуюся в ДНФ, каждому дизъюнктивному члену $\overline{Y}_{i_1} \& \overline{Y}_{i_k} \& ... \& \overline{Y}_{i_k}$ которой соответствует максимальное внутренне устойчивое множество $V\{v_{i_1}, v_{i_2}, ..., v_{i_k}\}$.

Замечание 4.46. Перед приведением к ДНФ часто оказывается целесообразным преобразовать F, воспользовавшись равносильностями

$$A \& A = A,\ (A \vee B) \& (A \vee C) \& ... \& (A \vee D) = A \vee (B \& C \& ... \& D), \tag{4.47}$$

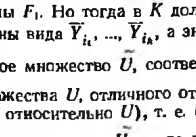
где A, B, C, ., D — произвольные формулы логики высказываний.

Обоснование метода Магу. Покажем, что любое множество, найденное методом Магу, является максимальным внутренне устойчивым, а также то, что методом Магу выделяются все максимальные внутренне устойчивые множества вершин орграфа D. Пусть $U = \{v_{i_1}, v_{i_2}, ..., v_{i_k}\}$ — некоторое множество вершин орграфа D, найденное методом Магу. Пусть далее $\{j_1, ..., j_l\} = \{1, 2, ..., n\} \setminus \{i_1, ..., i_k\}$, $k + l = n$. Тогда $K = \overline{Y}_{j_1} \& ... \& \overline{Y}_{j_l}$ — дизъюнктивный член формулы F_1, и на оценке $<\varepsilon_1, ..., \varepsilon_n>$, удовлетворяющей (4.41), очевидно, выполняется равенство $K = 1$, а следовательно, $F(\varepsilon_1, ..., \varepsilon_n) = F_1(\varepsilon_1, ..., \varepsilon_n) = 1$, откуда согласно утверждению 4.51 получаем, что U — внутренне устойчивое множество. Покажем, что U — максимальное внутренне устойчивое множество. Предположим, что это не так. Тогда найдется вершина $w \in V \setminus U$ такая, что множество $\widetilde{U} = U \cup \{w\}$ будет внутренне устойчивым. Пусть для определенности $w = v_{j_k}$. Тогда, используя утверждение 4.51 и полагая в оценке $<\varepsilon_1, ..., \varepsilon_n>$, удовлетворяющей (4.41), $\varepsilon_{j_k} = 1$ (вместо $\varepsilon_{j_k} = 0$), получаем,

что в силу внутренней устойчивости множества $\tilde{U}$ (очевидно, что теперь оценка $<\varepsilon_1, ..., \varepsilon_n>$ соответствует множеству $\tilde{U}$) выполняется равенство $F_1(\varepsilon_1, ..., \varepsilon_n) = F(\varepsilon_1, ..., \varepsilon_n) = 1$, а следовательно, по крайней мере, один из дизъюнктивных членов $\tilde{K}$ формулы F_1 должен принимать на оценке $<\varepsilon_1, ..., \varepsilon_n>$ значение 1. Но тогда в $\tilde{K}$ отсутствуют конъюнктивные члены вида $\overline{Y}_{i_1}, ..., \overline{Y}_{i_k}$, $\overline{Y}_{j_1}$, а значит, $K \vee \tilde{K} = \tilde{K}$, что противоречит определению F_1. Это противоречие показывает, что наше предположение неверно, а следовательно, U — максимальное внутренне устойчивое множество.

Покажем, теперь, что методом Магу выделяются все максимальные внутренне устойчивые множества вершин орграфа D. Предположим, что это не так, и $U = \{v_{i_1}, ..., v_{i_k}\}$ — некоторое максимальное внутренне устойчивое множество, которое не удалось выделить методом Магу. Пусть $<\varepsilon_1, ..., \varepsilon_n>$ — оценка списка переменных $<Y_1, ..., Y_n>$, удовлетворяющая (4.41). Согласно утверждению 4.51 имеем $F_1(\varepsilon_1, ..., \varepsilon_n) = F(\varepsilon_1, ..., \varepsilon_n) = 1$, а следовательно, на оценке $<\varepsilon_1, ..., \varepsilon_n>$ значение 1 должен принимать, по крайней мере, один из дизъюнктивных членов $\tilde{K}$ формулы F_1. Но тогда в $\tilde{K}$ должны отсутствовать конъюнктивные члены вида $\overline{Y}_{i_1}, ..., \overline{Y}_{i_k}$, а значит, максимальное внутренне устойчивое множество $\tilde{U}$, соответствующее $\tilde{K}$, содержит все вершины множества U, отличного от $\tilde{U}$ (см. сделанное выше предположение относительно U), т. е. U является собственным подмножеством множества $\tilde{U}$, а это противоречит тому, что U — максимальное внутренне устойчивое множество.

Пример 4.38. Используя метод Магу, определим совокупность максимальных внутренне устойчивых множеств вершин орграфа D, изображенного на рис. 4.39. Выпишем матрицу смежности



$$A(D) = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

Рис. 4.39

Тогда

$$\&_{a_{ij}=1}(\overline{Y}_i \vee \overline{Y}_j) = (\overline{Y}_1 \vee \overline{Y}_2) \,\&\, (\overline{Y}_1 \vee \overline{Y}_3) \,\&\, (\overline{Y}_1 \vee \overline{Y}_4) \,\&$$

$$\& \overline{(\overline{Y}_2 \vee \overline{Y}_1)} \& \overline{(\overline{Y}_3 \vee \overline{Y}_1)} \& \overline{(\overline{Y}_4 \vee \overline{Y}_3)} =$$

(используем (4.47) и опускаем символ &)

$$= \overline{(\overline{Y}_1 \vee \overline{Y}_2)}\,\overline{(\overline{Y}_1 \vee \overline{Y}_3)}\,\overline{(\overline{Y}_1 \vee \overline{Y}_4)}\,\overline{(\overline{Y}_4 \vee \overline{Y}_3)} =$$

$$= \overline{(\overline{Y}_1 \vee \overline{Y}_2 \overline{Y}_3 \overline{Y}_4)}\,\overline{(\overline{Y}_4 \vee \overline{Y}_3)} =$$

(применяем обобщенную дистрибутивность & относительно $\vee$ и равносильности (4.46))

$$= \overline{Y}_1 \overline{Y}_4 \vee \overline{Y}_1 \overline{Y}_3 \vee \overline{Y}_2 \overline{Y}_3 \overline{Y}_4 \overline{Y}_4 \vee \overline{Y}_2 \overline{Y}_3 \overline{Y}_4 \overline{Y}_3 =$$

$$= \overline{Y}_1 \overline{Y}_4 \vee \overline{Y}_1 \overline{Y}_3 \vee \overline{Y}_2 \overline{Y}_3 \overline{Y}_4.$$

Таким образом, мы получили формулу, находящуюся в ДНФ. Дальнейшее ее сокращение с использованием равносильностей (4.46) невозможно, а следовательно, искомыми максимальными внутренне устойчивыми множествами вершин орграфа D являются множества $\{v_2, v_3\}$, $\{v_3, v_4\}$, $\{v_1\}$, и при этом $\alpha(D) = 2$.

4.4.3. Внутренняя устойчивость в неориентированных графах

Внутренне устойчивые множества вершин аналогично можно определить и для неориентированного графа $G = (V, X)$, а именно: множество $U \subseteq V$ называется внутренне устойчивым, если

$$\forall v \in U \quad U \cap G(v) = \varnothing,$$

т. е. никакие две вершины из U не являются смежными. Нетрудно видеть, что если графу $G = (V, X)$ поставить в соответствие орграф с множеством вершин V, заменяя каждое ребро $\{v, w\}$ графа G на две дуги (v, w), (w, v), то в получаемом таким образом орграфе D совокупность внутренне устойчивых множеств вершин совпадает с совокупностью внутренне устойчивых множеств вершин графа G, поскольку в этом случае $\forall v \in V\ G(v) = D(v)$. Но тогда и совокупность максимальных внутренне устойчивых множеств вершин графа G совпадает с совокупностью максимальных внутренне устойчивых множеств вершин орграфа D, а следовательно, для их отыскания можно воспользоваться методом Магу, применяя его к орграфу D. При этом, очевидно, используемая в методе Магу матрица $A(D)$ полностью совпадает с $A(G)$. Таким образом, метод Магу без изменений переносится и на произвольные неориентированные графы. Более того, несложно показать, что если орграф D получен из $G = (V, X)$ введением ориентации на ребрах (т. е. каждое ребро $\{v, w\} \in X$ мы заменяем либо на (v, w), либо на (w, v), но не на обе эти дуги одновременно), то и в этом случае совокупность внутренне устойчивых (максимальных внутренне устойчивых) множеств вершин орграфа D совпадает с совокупностью внутренне устойчивых (максимальных внутренне устойчивых) множеств вершин графа G.

4.4.4. Внешняя устойчивость в ориентированных графах

Пусть задан орграф $D = (V, X)$. Множество $U \subseteq V$ называется *внешне устойчивым*, если

$$\forall v \in V \setminus U \quad U \cap D(v) \neq \varnothing, \tag{4.48}$$

т. е. любая вершина $v \in V$, не принадлежащая U, связана, по крайней мере, с одной вершиной из U дугой с началом в вершине v.

Пример 4.39. Для орграфа D и множеств U_1, U_2, U_3, U_4 (см. пример 4.36) множества U_2, U_3 и $U_5 = \{v_1, v_2, v_3\}$ являются внешне устойчивыми, а множество U_4 (а следовательно, и U_1) таковым не является, поскольку $v_3 \in U_4$, но при этом $(v_3, v_1) \in X$, $(v_3, v_2) \in X$.

Внешне устойчивое множество $U \subseteq V$ называется *минимальным*, если после удаления из U произвольной вершины получаем множество, не являющееся внешне устойчивым. Часто при решении практических задач требуется найти внешне устойчивые множества с минимальным числом вершин. Их следует искать среди минимальных внешне устойчивых множеств.

Замечание 4.47. Внешне устойчивые множества, а также минимальные внешне устойчивые множества аналогично определяются и для ориентированного псевдографа. При этом, если в ориентированном псевдографе удалить петли и принять кратности дуг равными 1, то в полученном таким образом орграфе внешне устойчивые множества (а следовательно, и минимальные внешне устойчивые множества) совпадают с внешне устойчивыми множествами (соответственно, с минимальными внешне устойчивыми множествами) исходного ориентированного псевдографа, и поэтому в дальнейшем будем рассматривать лишь орграфы без петель и кратных дуг.

Пример 4.40. В продолжение примеров 4.36 и 4.39 имеем: U_2, U_3 — минимальные внешне устойчивые множества, а множество U_5 таковым не является, поскольку, удалив из U_5 вершину v_2, получаем множество U_2, снова являющееся внешне устойчивым.

Обозначим через $\Theta(D)$ совокупность всех минимальных внешне устойчивых множеств вершин орграфа D. Тогда число

$$\beta(D) = \min_{U \in \Theta(D)} |U|$$

называется *числом внешней устойчивости* орграфа D.

4.4.5. Метод Магу отыскания семейства минимальных внешне устойчивых множеств

Пусть $D = (V, X)$ — орграф, где $V = \{v_1, \ldots, v_n\}$. Пусть далее U — некоторое множество вершин орграфа D, т. е. $U \subseteq V$. Сно-

ва (как и в разд. 4.4.2) воспользуемся булевыми переменными $Y_1, \ldots, Y_n$, а также оценкой $<\varepsilon_1, \ldots, \varepsilon_n>$ списка переменных $<Y_1, \ldots, Y_n>$, удовлетворяющей (4.41). Заметим, что условие (4.48) в определении внешне устойчивого множества U равносильно тому, что $\forall i \in \{1, 2, \ldots, n\}$ выполняется, по крайней мере, одно из условий:

1) $v_i \in U$;
2) $U \cap D(v_i) \neq \varnothing$ (т. е. $\exists j \in \{1, 2, \ldots, n\} : v_j \in U,\ v_j \in D(v_i)$),

что в силу (4.41) эквивалентно выполнению равенства

$$\overset{n}{\underset{i=1}{\&}} [\varepsilon_i \vee \overset{n}{\underset{j=1}{\vee}} (a_{ij} \& \varepsilon_j)] = 1, \tag{4.49}$$

где $[a_{ij}] = A(D)$. Если при каждом фиксированном i под записью «$a_{ij}=1$» понимать множество всех $j \in \{1, 2, \ldots, n\}$ таких, что $a_{ij}=1$, то условие (4.49) можно переписать в виде

$$\overset{n}{\underset{i=1}{\&}} (\varepsilon_i \vee \underset{a_{ij}=1}{\vee} \varepsilon_j) = 1,$$

или, обозначив

$$F(Y_1, \ldots, Y_n) = \overset{n}{\underset{i=1}{\&}} (Y_i \vee \underset{a_{ij}=1}{\vee} Y_j),$$

в виде

$$F(\varepsilon_1, \ldots, \varepsilon_n) = 1. \tag{4.50}$$

Таким образом, мы показали, что справедливо

Утверждение 4.52. *Необходимым и достаточным условием внешней устойчивости множества $U \subseteq V$ является выполнение равенства (4.50), где $<\varepsilon_1, \ldots, \varepsilon_n>$ удовлетворяет (4.41).*

Из утверждения 4.52 следует, что внешне устойчивые множества вершин орграфа D, и только они, соответствуют оценкам списка переменных $<Y_1, \ldots, Y_n>$, на которых выполняется равенство $F=1$, что дает простой способ выделения всех внешне устойчивых множеств вершин орграфа D.

Опишем теперь *метод Магу* выделения всех минимальных внешне устойчивых множеств вершин орграфа D. Применяя обобщенную дистрибутивность $\&$ относительно $\vee$, приводим формулу логики высказываний F к ДНФ, а затем сокращаем ее до тех пор, пока это возможно, используя равносильности (4.46). В результате получаем формулу F_1 (равносильную F), находящуюся в ДНФ, каждому дизъюнктивному члену $Y_{i_1} \& Y_{i_2} \& \ldots \& Y_{i_k}$ которой соответствует минимальное внешне устойчивое множество $\{v_{i_1}, v_{i_2}, \ldots, v_{i_k}\}$.

Замечание 4.48. Перед приведением к ДНФ часто оказывается целесообразным упростить формулу F, применяя (до тех пор, пока это возможно) закон поглощения

$$A\,\&\,(A \vee B) = A, \qquad (4.51)$$

справедливый для любых формул логики высказываний A, B. Кроме того, остается в силе замечание 4.46.

Обоснование метода Магу. Покажем, что любое множество, найденное методом Магу, является минимальным внешне устойчивым, а также то, что методом Магу выделяются все минимальные внешне устойчивые множества вершин орграфа D. Пусть $U = \{v_{i_1}, v_{i_2}, ..., v_{i_k}\}$ — некоторое множество вершин орграфа D, найденное методом Магу, $\{j_1, ..., j_l\} = \{1, 2, ... n\} \setminus \{i_1, ..., i_k\}$, $k + l = n$. Тогда $K = Y_{i_1}\,\&\,... \,\&\, Y_{i_k}$ — дизъюнктивный член формулы F_1, и на оценке $<\varepsilon_1, ..., \varepsilon_n>$, удовлетворяющей (4.41), очевидно, выполняется равенство $K = 1$, а следовательно, $F(\varepsilon_1, ... \varepsilon_n) = F_1(\varepsilon_1, ..., \varepsilon_n) = 1$, откуда согласно утверждению 4.52 получаем, что U — внешне устойчивое множество. Покажем, что U — минимальное внешне устойчивое множество. Предположим, что это не так. Тогда найдется вершина $w \in U$ такая, что множество $\tilde{U} = U \setminus \{w\}$ также будет внешне устойчивым. Пусть для определенности $w = v_{i_1}$. Тогда, используя утверждение 4.52 и полагая в оценке $<\varepsilon_1, ..., \varepsilon_n>$, удовлетворяющей (4.41), $\varepsilon_{i_1} = 0$ (вместо $\varepsilon_{i_1} = 1$), получаем, что в силу внешней устойчивости множества $\tilde{U}$ (очевидно, что теперь оценка $<\varepsilon_1, ..., \varepsilon_n>$ соответствует множеству $\tilde{U}$) выполняется равенство $F_1(\varepsilon_1, ..., \varepsilon_n) = F(\varepsilon_1, ..., \varepsilon_n) = 1$, а следовательно, по крайней мере, один из дизъюнктивных членов $\tilde{K}$ формулы F_1 должен принимать на оценке $<\varepsilon_1, ..., \varepsilon_n>$ значение 1. Но тогда в $\tilde{K}$ отсутствуют переменные $Y_{i_1}, Y_{j_1}, ..., Y_{j_l}$, а значит, $K \vee \tilde{K} = \tilde{K}$, что противоречит определению F_1. Это противоречие показывает, что наше предположение неверно, а следовательно, U — минимальное внешне устойчивое множество.

Покажем теперь, что методом Магу выделяются все минимальные внешне устойчивые множества вершин орграфа D. Предположим, что это не так, и $U = \{v_{i_1}, ..., v_{i_k}\}$ — некоторое минимальное внешне устойчивое множество, которое не удалось выделить методом Магу. Пусть $<\varepsilon_1, ..., \varepsilon_n>$ — оценка списка переменных $<Y_1, ..., Y_n>$, удовлетворяющая (4.41). Согласно утверждению 4.52 имеем $F_1(\varepsilon_1, ..., \varepsilon_n) = F(\varepsilon_1, ..., \varepsilon_n) = 1$, а следовательно, на оценке $<\varepsilon_1, ..., \varepsilon_n>$ значение 1 должен принимать, по крайней мере, один из дизъюнктивных членов $\tilde{K}$ формулы F_1. Но тогда в $\tilde{K}$ должны отсутствовать переменные $Y_{j_1}, ..., Y_{j_l}$, где $\{j_1, ..., j_l\} = \{1, ..., n\} \setminus \{i_1, ..., i_k\}$, $k + l = n$, а значит, минимальное внешне устойчивое множество $\tilde{U}$, соответствующее

$\tilde{K}$, является подмножеством множества U, отличного от $\tilde{U}$ (см. сделанное выше предположение относительно U), т. е. собственным подмножеством множества U, а это противоречит тому, что U — минимальное внешне устойчивое множество.

Пример 4.41. Используя метод Магу, определим совокупность минимальных внешне устойчивых множеств вершин орграфа D (см. пример 4.38), изображение которого приведено на рис. 4.39. Имеем

$$\bigwedge_{i=1}^{4}\Big(Y_i \vee \bigvee_{a_{ij}=1} Y_j\Big) =$$

$$= (Y_1 \vee Y_2 \vee Y_3 \vee Y_4)\,\&\,(Y_2 \vee Y_1)\,\&\,(Y_3 \vee Y_1)\,\&\,(Y_4 \vee Y_3) =$$

(применяем закон поглощения (4.51))

$$= (Y_2 \vee Y_1)\ \&\ (Y_3 \vee Y_1)\ \&\ (Y_4 \vee Y_3) =$$

(используем (4.47) и опускаем символ &)

$$= (Y_1 \vee Y_2 Y_3)\,(Y_4 \vee Y_3) =$$

(применяем обобщенную дистрибутивность & относительно $\vee$ и равносильности (4.46))

$$= Y_1 Y_4 \vee Y_1 Y_3 \vee Y_2 Y_3 Y_4 \vee Y_2 Y_3 Y_3 =$$
$$= Y_1 Y_4 \vee Y_1 Y_3 \vee Y_2 Y_3.$$

Таким образом, мы получили формулу, находящуюся в ДНФ. Дальнейшее ее сокращение с использованием равносильностей (4.46) невозможно, а следовательно, искомыми минимальными внешне устойчивыми множествами вершин орграфа D являются множества $\{v_1, v_4\}$, $\{v_1, v_3\}$, $\{v_2, v_3\}$, и при этом $\beta(D) = 2$.

4.4.6. Внешняя устойчивость в неориентированных графах

Внешне устойчивые множества вершин можно аналогично определить и для неориентированного графа $G = (V, X)$, а именно: множество $U \subseteq V$ называется внешне устойчивым, если

$$\forall v \in V \setminus U \quad U \cap G(v) \neq \varnothing,$$

т. е. любая вершина $v \in V$, не принадлежащая U, смежна, по крайней мере, с одной вершиной из U. Нетрудно видеть, что если графу $G = (V, X)$ поставить в соответствие орграф с множеством вершин V, заменяя каждое ребро $\{v, w\}$ графа G на две дуги (v, w), (w, v), то в получаемом таким образом орграфе D совокупность внешне устойчивых множеств вершин совпадает с совокупностью внешне устойчивых множеств вершин графа G (поскольку в этом случае $\forall v \in V \ \ G(v) = D(v)$). Но тогда и совокупность минимальных внешне устойчивых множеств вершин графа G совпадает с совокупностью минимальных внешне устойчивых множеств вершин орграфа D, а следовательно, для

их отыскания можно воспользоваться методом Магу (применяя его к орграфу D). При этом, очевидно, используемая в методе Магу матрица $A(D)$ полностью совпадает с $A(G)$. Таким образом, метод Магу без изменений переносится и на неориентированные графы.

4.4.7. Ядра орграфа

Пусть задан орграф $D=(V, X)$. Множество $N\subseteq V$ называется *ядром* орграфа D, если N — одновременно внутренне и внешне устойчивое множество, т. е. если выполняются условия:

1) $\forall v\in N \quad N\cap D(v)=\varnothing$;
2) $\forall v\in V\setminus N \quad N\cap D(v)\neq\varnothing$.

Замечание 4.49. Аналогично определяется ядро и для неориентированного графа G, при этом в условиях 1 и 2 $D(v)$ заменяем на $G(v)$.

Орграф может не иметь ядра, иметь одно или несколько ядер. Заметим, что если орграф D имеет ядро N, то $\beta(D)\leqslant|N|\leqslant\alpha(D)$.

Пример 4.42. Рассмотрим орграф, изображенный на рис. 4.40. Он не имеет ядер, поскольку любое множество с одной вершиной не является внешне устойчивым, а любое множество с двумя или тремя вершинами не является внутренне устойчивым.

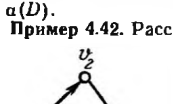


Рис 4.40

Пример 4.43. В продолжение примеров 4.36, 4.37, 4.39, 4.40 имеем: U_2, U_1 — ядра орграфа D, изображенного на рис. 4.38.

Теорема 4.7. *Для того чтобы множество $N\subseteq V$ являлось ядром орграфа $D=(V, X)$, необходимо и достаточно, чтобы оно было одновременно максимальным внутренне устойчивым и минимальным внешне устойчивым.*

Достаточность очевидна. Докажем *необходимость*. Пусть N — ядро орграфа D. Предположим, что N не является максимальным внутренне устойчивым множеством. Тогда найдется вершина $w\in V\setminus N$ такая, что множество $\overset{\vee}{N}=N\cup\{w\}$ снова будет внутренне устойчивым. В силу внешней устойчивости N, а также того, что $w\in\overset{\vee}{N}$, найдется вершина $v\in N$ такая, что $(w, v)\in X$, а это противоречит внутренней устойчивости множества $\overset{\vee}{N}$, поскольку $v,\ w\in\overset{\vee}{N}$. Предположим теперь, что N не является минимальным внешне устойчивым множеством. Тогда найдется вершина $w\in N$ такая, что множество $\tilde{N}=N\setminus\{w\}$ снова будет внешне устойчивым. Но тогда из $w\in\tilde{N}$, используя внешнюю ус-

тойчивость множества N, получаем, что найдется вершина $v \in \widetilde{N} \subset N$ такая, что $(w, v) \in X$, а это в силу $v, w \in N$ противоречит внутренней устойчивости множества N.

Из теоремы 4.7 следует, что для выделения ядер орграфа D достаточно, например, найти все минимальные внешне устойчивые множества вершин орграфа D, а затем выбрать из них внутренне устойчивые множества.

Замечание 4.50. Утверждение теоремы 4.7 остается справедливым и для неориентированного графа. Доказательство аналогично.

Пример 4.44. Определим ядра орграфа D (см. пример 4.38), изображение которого приведено на рис. 4.39. Воспользовавшись примером 4.41, выберем из минимальных внешне устойчивых множеств вершин орграфа D множества, являющиеся внутренне устойчивыми. Единственным таким множеством будет множество $\{v_2, v_3\}$. В силу теоремы 4.7 — это ядро орграфа D, и других ядер в D нет.

4.4.8. Функции на вершинах орграфа. Порядковая функция орграфа без контуров

Рассмотрим орграф $D = (V, X)$, не содержащий контуров, и определим множества $V_0, V_1, \ldots, V_r$:

$$\begin{aligned} V_0 &= \{v \in V \mid D(v) = \varnothing\}; \\ V_1 &= \{v \in V \setminus V_0 \mid D(v) \subseteq V_0\}; \\ V_2 &= \{v \in V \setminus (V_0 \cup V_1) \mid D(v) \subseteq V_0 \cup V_1\}; \end{aligned} \tag{4.52}$$

$$\ldots \ldots \ldots \ldots \ldots \ldots$$

$$V_r = \{v \in V \setminus \bigcup_{k=0}^{r-1} V_k \mid D(v) \subseteq \bigcup_{k=0}^{r-1} V_k\},$$

где r — наименьшее число такое, что

$$V \setminus \bigcup_{k=0}^{r} V_k = \varnothing$$

(при доказательстве теоремы 4.8 будет показано, что число r найдется). Множества $V_0, V_1, \ldots, V_r$ называются *уровнями орграфа D.*

Теорема 4.8. *Уровни орграфа $D = (V, X)$ без контуров являются непустыми множествами, образующими разбиение множества его вершин V.*

Предварительно докажем, что справедливы следующие утверждения:

1) $V_0 \neq \varnothing$;

2) если при некотором $i \geqslant 0$ выполняются неравенства $V_0 \neq \varnothing, \ldots, V_i \neq \varnothing$, $V \setminus \bigcup_{k=0}^{i} V_k \neq \varnothing$, то $V_{i+1} \neq \varnothing$.

Докажем сначала справедливость первого утверждения. Предположим, что $V_0 = \varnothing$. Тогда

$$\forall v \in V \; D(v) \neq \varnothing. \tag{4.53}$$

Пусть v_1 — произвольная вершина из V. Рассмотрим после-

довательность вершин $v_1, v_2, \ldots$ такую, что $\forall i \geqslant 2\ v_i \in D(v_{i-1})$. В силу (4:53) ее можно продолжать неограниченно. Поскольку количество вершин в орграфе D конечно, обязательно произойдет совпадение: $v_l = v_j$, где $1 \leqslant i < j$. Пусть это будет первым совпадением, т. е. совпадением с наименьшим номером j. Тогда $v_i v_{i+1} \ldots v_j$ — простой контур в D, а это противоречит условиям теоремы, согласно которым в D нет контуров.

Докажем теперь справедливость второго утверждения. Пусть при некотором $l \geqslant 0$ выполняются неравенства $V_0 \neq \varnothing, \ldots, V_l \neq \varnothing$, $V \setminus \bigcup_{k=0}^{l} V_k \neq \varnothing$. Предположим, что $V_{l+1} = \varnothing$. Пусть v_1 — произвольная вершина из $V \setminus \bigcup_{k=0}^{l} V_k$. В силу $V_{l+1} = \varnothing$ имеем $D(v_1) \setminus \bigcup_{k=0}^{l} V_k \neq \varnothing$, т. е. существует вершина $v_2 \in D(v_1)$ такая, что $v_2 \in V \setminus \bigcup_{k=0}^{l} V_k$. Аналогично для v_2 найдется вершина $v_3 \in D(v_2)$ такая, что $v_3 \in V \setminus \bigcup_{k=0}^{l} V_k$ и т. д. Таким образом, можно построить бесконечную последовательность $v_1, v_2, \ldots$ вершин из $V \setminus \bigcup_{k=0}^{l} V_k$ такую, что $\forall j \geqslant 2\ v_j \in D(v_{j-1})$. Но тогда аналогично доказательству первого утверждения получаем, что в D есть контур, а это противоречит исходному предположению.

В силу утверждений 1, 2, используя конечность множества V, а также тот очевидный факт, что

$$V_i \cap V_j = \varnothing \text{ при } i \neq j, \tag{4.54}$$

получаем, что существует число $r \geqslant 0$ такое, что

$$V_0 \neq \varnothing, \ldots, V_r \neq \varnothing,\ V \setminus \bigcup_{k=0}^{r} V_k = \varnothing,$$

откуда

$$V = \bigcup_{k=0}^{r} V_k. \tag{4.55}$$

Из (4.54) и (4.55) заключаем, что множества $V_0, \ldots, V_r$ образуют разбиение множества V.

Докажем также справедливость утверждения, обратного теореме 4.8.

Утверждение 4.53. *Пусть $D = (V, X)$ — орграф, $r \geqslant 0$, $V_0, \ldots, V_r$ — непустые множества, удовлетворяющие (4.52), такие, что $V = \bigcup_{k=0}^{r} V_k$. Тогда D — орграф без контуров.*

Предположим, что в D имеется некоторый контур $v_1 v_2 \ldots v_k v_1$, где $k \geqslant 2$. Пусть j — минимальный номер среди $0, \ldots, r$ такой, что $V_j \cap \{v_1, \ldots, v_k\} \neq \varnothing$. Для простоты обозначений будем считать, что $v_1 \in V_j$. Тогда $v_2 \in V_l$, где $l \geqslant j$. В силу $v_1 \in V_j$ по определению V_j имеем $D(v_1) \subseteq \subseteq \bigcup_{k=0}^{j-1} V_k$ (в случае $j = 0$ соответственно имеем $D(v_1) = \varnothing$), а это противоречит тому, что $v_2 \in D(v_1)$, $v_2 \in V_l$, $l \geqslant j$.

Функция $O(v)$, определенная на множестве вершин V орга-

фа без контуров $D = (V, X)$ и ставящая в соответствие каждой вершине $v \in V$ номер уровня, которому она принадлежит, называется *порядковой функцией* орграфа D.

Пример 4.45. Разобьем орграф D, изображенный на рис. 4.41, *а*, на уровни и определим порядковую функцию $O(v)$.

Согласно (4.52) имеем

$V_0 = \{v \in V \mid D(v) = \varnothing\} = \{v_4, v_6\}$;
$V_1 = \{v \in V \setminus V_0 \mid D(v) \subseteq V_0\} = \{v_1, v_5\}$;
$V_2 = \{v \in V \setminus (V_0 \cup V_1) \mid D(v) \subseteq V_0 \cup V_1\} = \{v_3\}$;
$V_3 = \{v \in V \setminus (V_0 \cup V_1 \cup V_2) \mid D(v) \subseteq V_0 \cup V_1 \cup V_2\} = \{v_2\}$;
$V \setminus (V_0 \cup V_1 \cup V_2 \cup V_3) = \varnothing$.

Определив множества V_0, V_1, V_2, V_3, найдем значения порядковой функции $O(v)$, $v \in V$ (на рис. 4.41, *а* они указаны при вершинах).

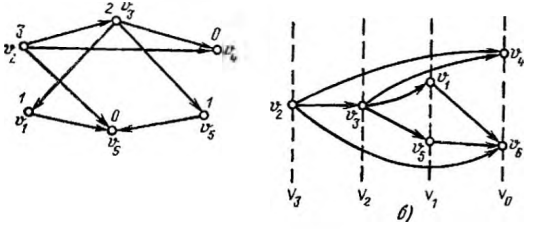


Рис. 4.41

Замечание 4.51. Для наглядности после разбиения некоторого орграфа D на уровни имеет смысл перерисовать его, последовательно расположив вершины орграфа D на вертикальных прямых, при этом вершины одного уровня располагаются на одной вертикальной прямой (см. на рис. 4.41, *б* изображение орграфа D, описанного в примере 4.45, с таким расположением вершин).

Приведем простой алгоритм выделения уровней орграфа без контуров, использующий задание орграфа матрицей смежности. Он может быть легко реализован на ЭВМ.

Алгоритм 4.9 нахождения уровней орграфа $D = (V, X)$ без контуров:

Шаг 1. Выпишем матрицу смежности $A(D)$. Образуем под матрицей $A(D)$ строку Λ_0, в i-м месте которой укажем число единиц в i-й строке матрицы $A(D)$. Уровень V_0 образуют вершины, которым в строке Λ_0 соответствует число 0. Если $V = V_0$, то задача решена и V_0 — единственный уровень орграфа D. В противном случае переходим к шагу 2.

Шаг 2. Образуем под строкой Λ_0 строку Λ_1, ставя под каждым нулем строки Λ_0 символ $\times$, а на любом другом i-м месте —

число единиц в i-й строке матрицы $A(D)$, не учитывая единицы в столбцах, находящихся над символами $\times$ в строке Λ_1. Уровень V_1 образуют вершины, которым в строке Λ_1 соответствует число 0. Полагаем $j=1$.

Шаг 3. Пусть при некотором $j \geqslant 1$ уже построены строки Λ_0, ..., Λ_j, по которым получены множества V_0, ..., V_j. Если строка Λ_j состоит из нулей и символов $\times$, то задача решена и при $r = j$ V_0, ..., V_r — уровни орграфа D. В противном случае переходим к шагу 4.

Шаг 4. Образуем под строкой Λ_j строку Λ_{j+1}, ставя под каждым нулем и символом $\times$ строки Λ_j символ $\times$, а на любом другом i-м месте — число единиц в i-й строке матрицы $A(D)$, не учитывая единицы в столбцах, находящихся над символами $\times$ в строке Λ_{j+1}. Уровень V_{j+1} образуют вершины, которым в строке Λ_{j+1} соответствует число 0. Присваиваем $j := j+1$ и переходим к шагу 3.

Пример 4.46. Применяя алгоритм 4.9, разобьем орграф D, описанный в примере 4.45 (см. рис. 4.41, *а*), на уровни. Матрица смежности орграфа D, а также строки Λ_0, Λ_1, Λ_2, Λ_3, являющиеся результатом работы алгоритма 4.9, приведены в табл. 4.11. Из таблицы следует, что $V_0 = \{v_4, v_6\}$, $V_1 = \{v_1, v_5\}$, $V_2 = \{v_3\}$, $V_3 = \{v_2\}$.

Обоснование алгоритма 4.9. Заметим, что единицы в i-й строке матрицы $A(D)$ соответствуют вершинам, принадлежащим образу i-й вершины орграфа D. Но тогда нули строки Λ_0 соответствуют вершинам, образы которых равны пустому множеству, т. е. вершинам из V_0. Далее рассмотрим нетривиальный случай, когда $V \neq V_0$, $r > 0$. По описанию алгоритма символы $\times$ в строке Λ_1 ставятся под нулями строки Λ_0, т. е. (по уже доказанному) они соответствуют вершинам из V_0, а следовательно, поскольку при определении элементов в Λ_1, не занятых символами $\times$, мы не учитываем столбцы матрицы $A(D)$, находящиеся над символами $\times$ в строке Λ_1, то нули строки Λ_1 соответствуют вершинам из $V \setminus V_0$, образы которых целиком содержатся в V_0, т. е. вершинам из V_1.

Таблица 4.11

	v_1	v_2	v_3	v_4	v_5	v_6
v_1	0	0	0	0	0	1
v_2	0	0	1	1	0	1
v_3	1	0	0	1	1	0
v_4	0	0	0	0	0	0
v_5	0	0	0	0	0	1
v_6	0	0	0	0	0	0
Λ_0	1	3	3	0	1	0
Λ_1	0	1	2	$\times$	0	$\times$
Λ_2	$\times$	1	0	$\times$	$\times$	$\times$
Λ_3	$\times$	0	$\times$	$\times$	$\times$	$\times$

Пусть теперь при некотором $j \in \{1, \ldots, r\}$ доказано, что нули строки Λ_j соответствуют вершинам из V_j, а символы $\times$ — вершинам из $V_0 \cup \ldots \cup V_{j-1}$. Тогда, если строка Λ_j целиком состоит из нулей и символов $\times$, то

$V \setminus \bigcup_{k=0}^{j} V_k = \varnothing$, а значит, $r = j$ и $V_0, \ldots, V_j$ — искомые уровни орграфа D.

В противном случае $j < r$, и согласно алгоритму образуем строку Λ_{j+1}: при этом символы $\times$ в строке Λ_{j+1} соответствуют вершинам из $V_0 \cup \ldots \cup V_j$ (так как символы $\times$ в строке Λ_{j-1} ставятся под нулями и символами $\times$ строки Λ_j). Далее, поскольку при определении элементов в Λ_{j+1}, не занятых символами $\times$, мы не учитываем столбцы матрицы $A(D)$, находящиеся над символами $\times$ в строке Λ_{j+1}, то нули строки Λ_{j+1} соответствуют вершинам из $V \setminus \bigcup_{k=0}^{j} V_k$, образы которых целиком содержатся в $\bigcup_{k=0}^{j} V_k$, т. е. вершинам из V_{j+1}.

Таким образом, мы доказали, что если действовать согласно алгоритму 4.9, то нули каждой очередной строки Λ_j, где $j \in \{0, \ldots, r\}$, будут соответствовать вершинам из V_j, и строка Λ_j будет состоять из нулей и символов $\times$ только при $j = r$. Но тогда на шаге 3 будут получены искомые уровни орграфа D. Алгоритм полностью обоснован.

Покажем, что наряду с нахождением уровней орграфа без контуров алгоритм 4.9 позволяет проверить наличие контуров у произвольного орграфа.

Утверждение 4.54. *Для того чтобы орграф $D = (V, X)$ содержал хотя бы один контур, необходимо и достаточно, чтобы в результате применения к D алгоритма 4.9 появилась строка Λ_j без нулей.*

Достаточность. Пусть после применения к орграфу D алгоритма 4.9 появилась строка Λ_j без нулей. Докажем, что в D имеется хотя бы один контур. Предположим, что D — орграф без контуров и $V_0, \ldots, V_r$ (где $r \geqslant 0$) — уровни орграфа D. Тогда согласно теореме 4.8 $V_j \neq \varnothing$, $j = 0, \ldots, r$, а следовательно, строки $\Lambda_0, \ldots, \Lambda_r$ содержат нули (см. описание алгоритма 4.9). При этом, поскольку $V = \bigcup_{k=0}^{r} V_k$ (см. теорему 4.8), то $V_r = V \setminus \bigcup_{k=0}^{r-1} V_k$ (рассматриваем нетривиальный случай, когда $r > 0$), и следовательно, строка Λ_r будет состоять из нулей и символов $\times$ (см. обоснование алгоритма 4.9). Но тогда строка Λ_r — последняя (см. шаг 3 алгоритма 4.9), а значит, не нашлось строк без нулей, что противоречит сделанному предположению.

Необходимость. Пусть орграф D содержит хотя бы один контур. Предположим, что в процессе выполнения алгоритма 4.9 все встречающиеся строки Λ_j имеют нули. Тогда множества V_j, состоящие из вершин, соответствующих нулям в Λ_j, отличны от пустых и удовлетворяют (4.52) (см. обоснование алгоритма 4.9). Из (4.52) получаем, что множества V_j попарно не пересекаются. Но тогда в силу конечности множества V последовательность множеств $\{V_j\}$, а значит, и последовательность строк $\{\Lambda_j\}$ не могут быть бесконечными. Следовательно, при некотором $r \geqslant 0$ строка Λ_r будет состоять из нулей и символов $\times$ (см. шаг 3 алгоритма 4.9), а значит (см. обоснование алгоритма 4.9), будет справедливо $V = \bigcup_{k=0}^{r} V_k$, откуда согласно утверждению 4.53 получаем,

что D — орграф без контуров, а это противоречит исходному предположению.

Пример 4.47. Проверим наличие контуров в орграфе D, заданном матрицей смежности

$$A(D) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}.$$

В табл. 4.12 приведена последовательность строк $\{\Lambda_i\}$, построенная по алгоритму 4.9. Поскольку встретилась строка Λ_3 без нулей, то в силу утверждения 4.54 в орграфе D имеется хотя бы один контур.

Таблица 4.12

Λ_0	0	4	4	2	1	4	2
Λ_1	×	3	4	1	0	3	1
Λ_2	×	3	3	0	×	3	0
Λ_3	×	1	1	×	×	1	×

4.4.9. Функция Гранди

Рассмотрим орграф $D = (V, X)$. Функция $g(v)$, ставящая в соответствие каждой вершине $v \in V$ целое число $g(v) \geqslant 0$, называется *функцией Гранди* для орграфа D, если в каждой вершине $v \in V$ числу $g(v)$ является минимальным из всех целых неотрицательных чисел, не принадлежащих множеству $\{g(w) \mid w \in D(v)\}$, и $g(v) = 0$ при $D(v) = \varnothing$.

Если для орграфа D существует функция Гранди, то говорят, что орграф D *допускает* (в противном случае — *не допускает*) функцию Гранди.

Не всякий орграф D допускает функцию Гранди (см. пример 4.49), а если и допускает, то она не обязательно единственная (см. пример 4.50).

Из определения функции Гранди следует, что справедливо

Утверждение 4.55. *Если орграф $D = (V, X)$ допускает функцию Гранди, то найдется вершина $v \in V$ такая, что $g(v) = 0$.*

Пусть в орграфе D

$$\forall v \in V \quad g(v) > 0. \tag{4.56}$$

Рассмотрим произвольную вершину $w \in V$. Тогда, с одной стороны, в силу (4.56) имеем $g(w) > 0$, а с другой стороны, используя (4.56), получаем, что либо $D(w) = \varnothing$, либо $\forall v \in D(w)$ $g(v) > 0$, а следовательно, по определению функции Гранди, $g(w) = 0$, т. е. налицо противоречие.

Пример 4.48. На рис. 4.42 приведено изображение орграфа D, допускающего функцию Гранди, около каждой вершины которого указано значение этой функции.

Пример 4.49. Покажем, что орграф D, изображенный на рис. 4.40, не допускает функцию Гранди. Предположим обратное, т. е. орграф D допускает функцию Гранди. Используя утверждение 4.55, получаем, что на некоторой вершине значение

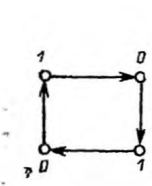


Рис. 4.42

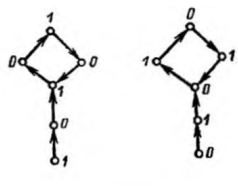


Рис. 4.43

функции Гранди равно нулю. Пусть для определенности $g(v_1) = 0$ (другие случаи рассматриваются аналогично). Тогда в силу того, что $D(v_3) = \{v_1\}$, $D(v_2) = \{v_3\}$, $D(v_1) = \{v_2\}$, последовательно получаем $g(v_3) = 1$, $g(v_2) = 0$, $g(v_1) = 1$, а это противоречит равенству $g(v_1) = 0$.

Пример 4.50. На рис. 4.43 приведены два варианта функции Гранди для одного и того же орграфа D.

Приведем достаточное условие существования функции Гранди.

Теорема 4.9. *Пусть $D = (V, X)$ — орграф без контуров. Тогда D допускает и притом единственную функцию Гранди.*

Согласно теореме 4.8 множество вершин V орграфа D можно разбить на уровни $V_0, \ldots, V_r$. По определению функции Гранди, если она допустима для D, то

$$\forall v \in V_0 \quad g(v) = 0; \quad \forall v \in V_1 \quad g(v) = 1. \tag{4.57}$$

Заметим, что значения функции Гранди на каждом уровне V_i, где $i \geqslant 2$, однозначно находятся по ее значениям на предыдущих уровнях $V_0, \ldots, V_{i-1}$ (поскольку $\forall v \in V_i \quad D(v) \subseteq \bigcup_{k=0}^{i-1} V_k$), а следовательно, исходя из (4.57), ее можно однозначно определить на всех последующих уровнях.

Замечание 4.52. Доказательство теоремы 4.9, по существу, содержит в себе алгоритм определения функции Гранди для орграфов без контуров.

Пример 4.51. Найдем функцию Гранди для орграфа, описанного в примере 4.45. Разобьем множество вершин орграфа D на уровни: $V_0 = \{v_4, v_6\}, V_1 = \{v_1, v_5\}$, $V_2 = \{v_3\}$, $V_3 = \{v_2\}$. С учетом (4.57) получаем $g(v_4) = g(v_6) = 0$, $g(v_1) = g(v_5) = 1$. Далее для вершины $v_3 \in V_2$ имеем $D(v_3) = \{v_1, v_4, v_5\}$, где $g(v_4) = 0$, $g(v_1) = g(v_5) = 1$, а следовательно, $g(v_3) = 2$. Соответственно для вершины $v_2 \in V_3$ получаем $D(v_2) = \{v_3, v_4, v_6\}$, где $g(v_4) = g(v_6) = 0$, $g(v_3) = 2$, а следовательно, $g(v_2) = 1$. На рис. 4.44 приведено изображение орграфа D, около каждой вершины которого указано значение функции Гранди.

Следующая теорема показывает, что по функции Гранди орграфа может легко определить его ядро.

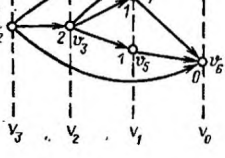
Теорема 4.10. *Если орграф $D = (V, X)$ допускает функцию Гранди $g(v)$, то множество вершин $N = \{v \in V \mid g(v) = 0\}$ является ядром этого орграфа.*

Покажем, что множество N удовлетворяет условиям (см. определение ядра):

1) $\forall v \in N \quad N \cap D(v) = \varnothing$;
2) $\forall v \in V \setminus N \quad N \cap D(v) \neq \varnothing$.

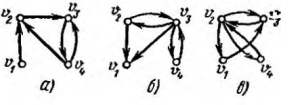
Докажем сначала, что выполняется первое условие. Пусть $v \in N$. Тогда $g(v) = 0$. Если предположить, что $N \cap D(v) \neq \varnothing$, то существует вершина $w \in N \cap D(v)$. Но тогда из $w \in N$ имеем $g(w) = 0$, а из $w \in D(v)$ получаем, что не может выполняться равенство $g(v) = 0$. Данное противоречие подтверждает справедливость первого условия.

Докажем теперь, что выполняется второе условие. Пусть $v \in V \setminus N$. Тогда $g(v) \neq 0$. Если предположить, что $N \cap D(v) = \varnothing$, то по определению функции Гранди должно выполняться



← Рис. 4.44

Рис. 4.45 ↓



равенство $g(v) = 0$, т. е. пришли к противоречию, а значит $N \cap D(v) \neq \varnothing$.

Задачи и упражнения

1. Доказать, что для любой вершины v орграфа D найдется максимальное внутренне устойчивое множество вершин орграфа D, содержащее v.

2. Привести пример орграфа с четырьмя вершинами, имеющего лишь одно минимальное внешне устойчивое множество, состоящее при этом из единственной вершины.

3. Используя метод Магу, определить максимальные внутренне устойчивые, а также минимальные внешне устойчивые множества вершин орграфов, изображенных на рис. 4.45,*а*—*в*. Найти ядра.

4. Разбить орграфы без контуров, заданные матрицами смежности, на уровни. Найти функцию Гранди и ядра орграфов. Рассмотреть случаи:

а) $\begin{bmatrix} 0&1&0&0&0&1&1 \\ 0&0&0&0&0&0&0 \\ 0&1&0&0&0&1&1 \\ 0&0&1&0&1&0&0 \\ 0&0&0&0&0&1&0 \\ 0&1&0&0&0&0&1 \\ 0&1&0&0&0&0&0 \end{bmatrix}$; б) $\begin{bmatrix} 0&1&0&0&0&1&1 \\ 0&0&1&1&0&1&0 \\ 0&0&0&0&0&0&0 \\ 0&0&1&0&0&0&0 \\ 0&0&1&0&0&1&0 \\ 0&0&0&1&0&0&0 \\ 0&1&0&0&1&1&0 \end{bmatrix}$, в) $\begin{bmatrix} 0&1&0&1&0&1&1 \\ 0&0&0&0&0&0&0 \\ 0&0&0&0&0&1&0 \\ 0&0&1&0&1&1&0 \\ 0&1&0&0&0&0&0 \\ 0&1&0&0&0&0&0 \\ 0&1&1&0&1&0&0 \end{bmatrix}$; г) $\begin{bmatrix} 0&1&0&0&1&0&1&0 \\ 0&0&1&0&0&1&0&0 \\ 0&0&0&0&0&0&0&0 \\ 0&0&1&0&0&1&0&0 \\ 0&0&0&0&0&1&0&0 \\ 0&0&0&0&0&0&0&0 \\ 0&0&0&1&1&1&0&0 \\ 0&0&1&0&0&1&1&0 \end{bmatrix}$.

5. Пусть орграф D допускает функцию Гранди. Показать, что любое множество вершин, на котором функция Гранди постоянна, является внутренне устойчивым.

4.5. ТРАНСПОРТНЫЕ СЕТИ

Транспортной сетью называется орграф $D = (V, X)$ с множеством вершин $V = \{v_1, \ldots, v_n\}$, для которого выполняются условия:

1) существует одна и только одна вершина v_1, называемая *источником*, такая, что $D^{-1}(v_1) = \varnothing$ (т. е. ни одна дуга не заходит в v_1);

2) существует одна и только одна вершина v_n, называемая *стоком*, такая, что $D(v_n) = \varnothing$ (т. е. из v_n не исходит ни одной дуги);

3) каждой дуге $x \in X$ поставлено в соответствие целое число $c(x) \geqslant 0$, называемое *пропускной способностью дуги*.

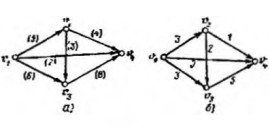


Рис. 4.46

Вершины в транспортной сети, отличные от источника и стока, называются *промежуточными*.

Пример 4.52 На рис. 4.46, *а* показана транспортная сеть, в которой v_1 — источник, v_4 — сток, v_2, v_3 — промежуточные вершины. При каждой дуге в скобках указана ее пропускная способность.

4.5.1. Поток в транспортной сети

Функция $\varphi(x)$, определенная на множестве X дуг транспортной сети D и принимающая целочисленные значения, называется *допустимым потоком* (или просто *потоком*) *в транспортной сети* D, если:

1) для любой дуги $x \in X$ величина $\varphi(x)$, называемая *потоком по дуге* x, удовлетворяет условию $0 \leqslant \varphi(x) \leqslant c(x)$;

2) для любой промежуточной вершины v выполняется равенство

$$\sum_{w \in D^{-1}(v)} \varphi(w, v) - \sum_{w \in D(v)} \varphi(v, w) = 0,$$

т. е. сумма потоков по дугам, заходящим в v, равна сумме потоков по дугам, исходящим из v.

Пример 4.53. На рис. 4.46, *б* показан допустимый поток в транспортной сети, описанной в примере 4.52. При каждой дуге указана величина потока по ней. Очевидно, что выполняются все условия, перечисленные в определении допустимого потока (проверьте выполнение второго условия для промежуточных вершин v_2, v_3).

Утверждение 4.56. *Для любого допустимого потока* φ *в транспортной сети* D *выполняется равенство*

$$\sum_{v \in D(v_1)} \varphi(v_1, v) = \sum_{v \in D^{-1}(v_n)} \varphi(v, v_n). \qquad (4.58)$$

По определению допустимого потока φ имеем

$$\sum_{v \in V \setminus \{v_1, v_n\}} \left[\sum_{w \in D^{-1}(v)} \varphi(w, v) - \sum_{w \in D(v)} \varphi(v, w) \right] = 0. \qquad (4.59)$$

Заметим, что для каждой дуги $x = (v_1, v) \in X$, где $v \neq v_n$, величина $\varphi(x)$ входит в левую часть равенства (4.59) лишь один раз и при этом со знаком плюс. Аналогично для каждой дуги $x = (v, v_n) \in X$, где $v \neq v_1$, величина $\varphi(x)$ входит в левую часть равенства (4.59) лишь один раз и при этом со знаком минус. С другой стороны, для каждой дуги $x = (u_1, u_2) \in X$, где $u_1, u_2 \in V \setminus \{v_1, v_n\}$, величина $\varphi(x)$ входит в левую часть равенства (4.59) один раз со знаком плюс (при $v = u_2$, $w = u_1 \in D^{-1}(v)$) и один раз со знаком минус (при $v = u_1$, $w = u_2 \in D(v)$), что в сумме дает нулевой вклад в левую часть равенства (4.59). Учи-

тывая сказанное, заключаем, что из равенства (4.59) следует справедливость равенства (4.58).

Величиной потока φ в транспортной сети D называется величина φ, равная сумме потоков по всем дугам, заходящим в v_n, или, что то же самое (в силу утверждения 5.1), — величина, равная сумме потоков по всем дугам, исходящим из v_1, т. е.

$$\overline{\varphi} = \sum_{v \in D^{-1}(v_n)} \varphi(v, v_n) = \sum_{v \in D(v_1)} \varphi(v_1, v).$$

Пусть φ — допустимый поток в транспортной сети D. Дуга $x \in X$ называется *насыщенной*, если поток по ней равен ее пропускной способности, т. е. если $\varphi(x) = c(x)$. Поток φ называется *полным*, если любой путь в D из v_1 в v_n содержит, по крайней мере, одну насыщенную дугу.

Поток φ называется *максимальным*, если его величина $\overline{\varphi}$ принимает максимальное значение по сравнению с другими допустимыми потоками в транспортной сети D.

Очевидно, что максимальный поток φ обязательно является полным (так как в противном случае в D существует некоторая простая цепь η из v_1 в v_n, не содержащая насыщенных дуг, а следовательно, можно увеличить на единицу потоки по всем дугам из η и тем самым увеличить на единицу $\overline{\varphi}$, что противоречит условию максимальности потока). Обратное же, вообще говоря, неверно. Как будет показано ниже (см. пример 4.57), существуют полные потоки, не являющиеся максимальными. Тем не менее полный поток можно рассматривать как некоторое приближение к максимальному потоку. В связи с этим опишем алгоритм построения полного потока в транспортной сети D.

Алгоритм 4.10:

Шаг 1. Полагаем $\forall x \in X \quad \varphi(x) = 0$ (т. е. начинаем с нулевого потока). Кроме того, полагаем $D' = D$.

Шаг 2. Удаляем из орграфа D' все дуги, являющиеся насыщенными при потоке φ в транспортной сети D. Полученный орграф снова обозначаем через D'.

Шаг 3. Ищем в D' простую цепь η из v_1 в v_n. Если такой цепи нет, то φ — искомый полный поток в транспортной сети D. В противном случае переходим к шагу 4.

Шаг 4. Увеличиваем поток $\varphi(x)$ на каждой дуге x из η на одинаковую величину $a > 0$ такую, что, по крайней мере, одна дуга из η оказывается насыщенной, а потоки по остальным дугам из η не превышают их пропускных способностей. При этом величина потока $\overline{\varphi}$ также увеличивается на a, а сам поток φ в транспортной сети D остается допустимым (поскольку в каждую промежуточную вершину, содержащуюся в η, дополнительно вошло a единиц потока и из нее вышло также a единиц потока). После этого переходим к шагу 2.

Обоснование алгоритма 4.10 несложно (проведите его самостоятельно).

Пример 4.54. Построим полный поток в транспортной сети D, изображенной на рис. 4.47, *а* (в скобках указаны пропускные способности дуг). Воспользуемся алгоритмом 4.10. Полагаем $D' = D$, $\forall x \in X\ \varphi(x) = 0$, т. е. начинаем с нулевого потока (см. на рис. 4.47, *б* изображение орграфа D с указанием величин потока φ по дугам). При нулевом потоке отсутствуют насыщенные дуги. Выделим в D' простую цепь $\eta_1 = v_1v_2v_4v_6$ и увеличим потоки по дугам из η_1 на три (до насыщения дуги (v_2, v_4)). В результате получим поток $\varphi = \varphi_1$, содержащий одну насыщенную

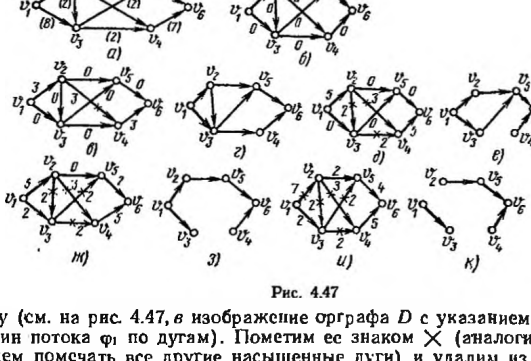


Рис. 4.47

дугу (см. на рис. 4.47, *в* изображение орграфа D с указанием величин потока φ_1 по дугам). Пометим ее знаком $\times$ (аналогично будем помечать все другие насыщенные дуги) и удалим из орграфа D'. Оставшийся орграф снова обозначаем через D' (см. изображение D' на рис. 4.47, *г*). Выделим в D' простую цепь $\eta_2 = v_1v_2v_3v_4v_6$ и увеличим потоки по дугам из η_2 на два (до насыщения дуг (v_2, v_3), (v_3, v_4)). В результате получим поток $\varphi = \varphi_2$, содержащий три насыщенные дуги (см. на рис. 4.47, *д* изображение орграфа D с указанием величин потока φ_2 по дугам). Удалим вновь полученные насыщенные дуги из D'. Оставшийся орграф снова обозначим через D' (см. его изображение на рис. 4.47, *е*). Выделим в D' простую цепь $\eta_3 = v_1v_3v_5v_6$ и увеличим потоки по дугам из η_3 на два (до насыщения дуги (v_3, v_5)). В результате получим поток $\varphi = \varphi_3$, содержащий четыре насыщенные дуги (см. на рис. 4.47, *ж* изображение орграфа D с указанием величин потока φ_3 по дугам). Удалим вновь полученную насыщенную дугу из D'. Оставшийся орграф снова обозначим через D' (см. его изображение на рис. 4.47, *з*). Выделим в D' простую цепь $\eta_4 = v_1v_2v_5v_6$ и увеличим потоки по дугам из η_4 на два (до насыщения дуги (v_1, v_2)). В результате получим поток $\varphi = \varphi_4$, содержащий пять насыщенных дуг (см. на рис. 4.47, *и*

изображение орграфа D с указанием величин потока φ_4 по дугам). Удалим вновь полученную насыщенную дугу из D'. Оставшийся орграф снова обозначим через D' (см. его изображение на рис. 4.47, *к*). Заметим, что в D' не существует пути из v_1 в v_6, а следовательно, в транспортной сети D с потоком φ_4 не существует путн из v_1 в v_6, который не содержал бы насыщенных дуг, т. е. поток φ_4 является полным. Величина φ_4 полученного полного потока равна 9. Как будет показано ниже (см. пример 4.57), этот поток не является максимальным.

4.5.2. Орграф приращений

Введем для заданной транспортной сети D и допустимого потока φ в этой сети орграф приращений $l(D, \varphi)$, имеющий те же вершины, что и сеть D. Каждой дуге $x = (v, w) \in X$ транспортной сети D в орграфе приращений $l(D, \varphi)$ соответствуют две дуги: x и $x' = (w, v)$ — дуга, противоположная по направлению дуге x. Припишем дугам $x = (v, w) \in X$, $x' = (w, v)$ орграфа приращений $l(D, \varphi)$ длину l:

$$l(x) = \begin{cases} 0, \text{ если } \varphi(x) < c(x); \\ \infty, \text{ если } \varphi(x) = c(x); \end{cases} \tag{4.60}$$

$$l(x') = \begin{cases} 0, \text{ если } \varphi(x) > 0; \\ \infty, \text{ если } \varphi(x) = 0, \end{cases} \tag{4.61}$$

т. е. орграф $l(D, \varphi)$ является нагруженным. При этом очевидно, что длина любого пути из v_1 в v_n в орграфе $l(D, \varphi)$ равна либо 0, либо ∞.

Пусть η — некоторая простая цепь в орграфе $l(D, \varphi)$. Будем говорить, что цепь η проходит через дугу $x = (v, w) \in X$, если либо x, либо $x' = (w, v)$ содержится в η. При этом, если x содержится в η, то говорим, что направления η и x совпадают, а если x' содержится в η, то говорим, что направления η и x противоположны.

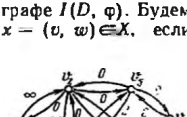


Рис. 4.48

Пример 4.55. Орграф приращений $l(D, \varphi_4)$ для транспортной сети D и потока φ_4 (см. пример 4.54) изображен на рис. 4.48.

4.5.3. Разрез. Пропускная способность разреза

Пусть D - транспортная сеть. Для любого множества $V_1 \subseteq V$ такого, что $v_1 \in V_1$, $v_n \in \overline{V}_1$, *разрезом сети* D относительно множества вершин V_1 называется множество дуг $X(V_1) = \{(w, v) \in X \mid v \in V_1, w \in \overline{V}_1\}$, т. е. множество, включающее в себя все дуги, исходящие из вершин, не принадлежа-

щих V_1, и заходящие в вершины, принадлежащие V_1. Число

$$c(X(V_1)) = \sum_{x \in X(V_1)} c(x)$$

называется *пропускной способностью* разреза $X(V_1)$. Разрез с минимальной пропускной способностью называется *минимальным*.

Пример 4.56. Для транспортной сети D (см. пример 4.52) и множества вершин $V_1 = \{v_2, v_3, v_4\}$ разрез сети D относительно V_1 представляет собой множество дуг $X(V_1) = \{(v_1, v_2), (v_1, v_3), (v_1, v_4)\}$. Его пропускная способность составляет $c(X(V_1)) = 5+6+2=13$.

Утверждение 4.57 *Для любого допустимого потока φ в транспортной сети D и любого множества $V_1 \subset V$, где $v_1 \in V_1$, $v_n \in V_1$, выполняется неравенство $\varphi \leqslant c(X(V_1))$, т. е. величина любого допустимого потока в сети D (в том числе и максимального) не превышает пропускной способности любого разреза сети D (в том числе и минимального).*

С учетом того, что все вершины в V_1, за исключением v_n, являются промежуточными, а $D(v_n) = \varnothing$, имеем

$$\begin{aligned}
\overline{\varphi} &= \sum_{w \in D^{-1}(v_n)} \varphi(w, v_n) = \sum_{w \in D^{-1}(v_n)} \varphi(w, v_n) + \\
&+ \sum_{v \in V_1 \setminus \{v_n\}} \Big[\sum_{w \in D^{-1}(v)} \varphi(w, v) - \sum_{w \in D(v)} \varphi(v, w) \Big] = \\
&= \sum_{v \in V_1} \Big[\sum_{w \in D^{-1}(v)} \varphi(w, v) - \sum_{w \in D(v)} \varphi(v, w) \Big] = \\
&= \sum_{v \in V_1} \sum_{w \in D^{-1}(v)} \varphi(w, v) - \sum_{v \in V_1} \sum_{w \in D(v)} \varphi(v, w) = \\
&= \sum_{x \in \{(w, v) \in X \mid v \in V_1\}} \varphi(x) - \sum_{x \in \{(v, w) \in X \mid v \in V_1\}} \varphi(x) \leqslant \\
&\leqslant \sum_{x \in \{(w, v) \in X \mid v \in V_1\}} \varphi(x) - \sum_{x \in \{(w, v) \in X \mid w, v \in V_1\}} \varphi(x) = \\
&= \sum_{x \in \{(w, v) \in X \mid v \in V_1,\ w \in V_1\}} \varphi(x) = \sum_{x \in X(V_1)} \varphi(x) \leqslant \\
&\leqslant \sum_{x \in X(V_1)} c(x) = c(X(V_1)).
\end{aligned}$$

Докажем следующую основную теорему.

Теорема 4.11 (*теорема Форда — Фалкерсона*). *Пусть D — транспортная сеть, φ — допустимый поток в этой сети, V_1 — множество вершин $v \in V$ таких, что длина минимального пути из v в v_n в орграфе приращений $I(D, \varphi)$ равна нулю. Тогда, если $v_1 \in V_1$, то φ — максимальный поток, величина которого равна $\overline{\varphi} = c(X(V_1))$.*

Пусть $v_1 \in V_1$. Тогда (см. доказательство утверждения 4.57) выполняется равенство

$$\overline{\varphi} = \sum_{x \in \{(w,\, v) \in X \mid v \in V_1\}} \varphi(x) - \sum_{x \in \{(v,\, w) \in X \mid v \in V_1\}} \varphi(x). \quad (4.62)$$

Если $x = (v, w) \in X$, $v \in V_1$, $w \in V_1$, то $\varphi(x) = 0$, так как в противном случае, используя (4.61), имеем $l(x') = 0$, где $x' = (w, v)$, а следовательно, в силу $v \in V_1$ в $I(D, \varphi)$ существует путь нулевой длины из w в v_n, что противоречит условию $w \in V_1$. Но тогда из (4.62) получаем

$$\overline{\varphi} = \sum_{x \in \{(w,\, v) \in X \mid v \in V_1\}} \varphi(x) - \sum_{x \in \{(v,\, w) \in X \mid v,\ w \in V_1\}} \varphi(x) = \sum_{x \in \{(w,\, v) \in X \mid v \in V_1,\ w \in V_1\}} \varphi(x) \quad (4.63)$$

Заметим, что если $x = (w, v) \in X$, $v \in V_1$, $w \in V_1$, то $l(x) = \infty$ (так как при $l(x) = 0$ в силу $v \in V_1$ в $I(D, \varphi)$ существовал бы путь нулевой длины из w в v_n, что противоречит условию $w \in V_1$), а следовательно, согласно (4.60) $\varphi(x) = c(x)$. Но тогда из (4.63) получаем $\overline{\varphi} = c(X(V_1))$. Поскольку величина любого допустимого потока в транспортной сети D не превосходит $c(X(V_1))$ (см. утверждение (4.57), то φ — максимальный поток в этой сети.

Следствие 1. Используя теорему Форда — Фалкерсона, а также утверждение 4.57, получаем, что величина максимального потока в транспортной сети равна пропускной способности минимального разреза.

Следствие 2. Пусть φ — допустимый поток в транспортной сети D. Тогда, если длина минимального пути из v_1 в v_n в орграфе приращений $I(D, \varphi)$ равна ∞, то φ — максимальный поток.

4.5.4. Алгоритм построения максимального потока в транспортной сети

Важным следствием теоремы Форда — Фалкерсона является алгоритм построения максимального потока в транспортной сети D. Опишем его.

Алгоритм 4.11:

Шаг 1. Полагаем $i = 0$. Пусть φ_0 — любой допустимый поток в транспортной сети D (например, полный; можно начинать с нулевого потока: $\varphi_0(x) = 0$, $x \in X$).

Шаг 2. По сети D и потоку φ_i строим орграф приращений $I(D, \varphi_i)$.

Шаг 3. Находим простую цепь η_i, являющуюся минимальным путем из v_1 в v_n в нагруженном орграфе $I(D, \varphi_i)$ (например, используя алгоритм Форда — Беллмана). Если длина этой цепи равна ∞, то поток φ_i максимален (см. следствие 2 теоремы 4.11), и работа алгоритма закончена. В противном случае увеличиваем поток вдоль цепи η_i на максимально допустимую величину $a_i > 0$, где $a_i \in \mathbf{Z}$ (прибавляя ее для каждой дуги $x \in X$, через которую проходит цепь η_i, к уже имеющейся величине потока по дуге x, если направления x и η_i совпадают, и вычитая, если на-

правления x и η_i противоположны), такую, что при этом сохраняется условие 1 допустимого потока (см. с. 250). В силу $l(\eta_i)=0$, используя (4.60), (4.61), получаем, что указанная величина a_i существует. В результате меняется поток в транспортной сети D, т. е. от потока φ_i мы перешли к потоку φ_{i+1} (очевидно, что при таком изменении поток остается допустимым), и при этом $\varphi_{i+1}=\varphi_i+a_i$. Присваиваем $i:=i+1$ и переходим к шагу 2.

Замечание 4.53. Поскольку последовательность величин φ_i является монотонно возрастающей, то максимальный поток будет построен за конечное число шагов.

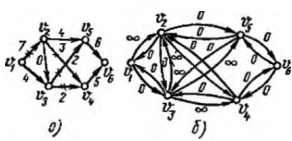


Рис. 4.49

Пример 4.57. В продолжение примеров 4.54, 4.55 построим максимальный поток в транспортной сети D. Начинаем с полного потока φ_4. Воспользуемся орграфом приращений, приведенным в примере 4.55 (см. на рис. 4.48 изображение орграфа $l(D, \varphi_4)$). В нагруженном орграфе $l(D, \varphi_4)$ имеется простая цепь $\eta_5 = v_1v_3v_2v_5v_6$, длина которой равна 0 (т. е. η_5 — минимальный путь из v_1 в v_6 в $l(D, \varphi_4)$). Увеличиваем поток вдоль цепи η_5 на максимально допустимую величину 2 (до обнуления потока по дуге (v_2, v_3)). В результате получаем поток φ_5 (см. на рис. 4.49, *а* изображение орграфа D с указанием величин потока φ_5 по дугам). Построим орграф приращений $l(D, \varphi_5)$, изображение которого приведено на рис. 4.49, *б*. Нетрудно видеть, что длина любого пути из v_1 в v_6 в нагруженном орграфе $l(D, \varphi_5)$ равна ∞, а следовательно, поток φ_5 является максимальным и при этом $\varphi_5=11$.

Задача. Найти максимальные потоки в транспортных сетях (см. рис. 4.50, *а*, *б*).

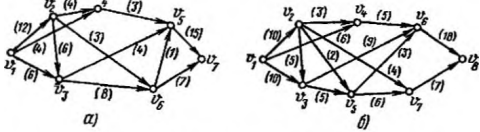


Рис. 4.50

4.6. ВЫЧИСЛИТЕЛЬНАЯ СЛОЖНОСТЬ АЛГОРИТМОВ

Мы уже говорили о том, что методы решения многих задач математики носят алгоритмический характер, что для задач, имеющих практический интерес, ищут именно разрешающие алгоритмы. Уточнение понятия алгоритма позволило выявить алгоритмически неразрешимые проблемы, к которым прежде всего относятся рассмотренные нами проблема разрешимости для логики предикатов, проблема самоприменимости для машин Тьюринга, проблема тождества слов в полугруппах.

Исследуя задачи на дискретных конечных математических структурах, как правило, можно найти комбинаторные алгоритмы для их решения, например, с помощью некоторого процесса перебора. Однако при этом число шагов быстро растет с увеличением размерности задачи, и последняя становится практически неразрешимой.

Поиск эффективных алгоритмов для решения задач дискретной математики привел к одной из важнейших ее проблем — к решению вопроса о возможности исключения перебора вариантов в комбинаторных алгоритмах.

В самом широком смысле эффективность алгоритма связана со всеми вычислительными ресурсами, необходимыми для его работы. Однако обычно под наиболее эффективным понимается наиболее «быстрый» алгоритм.

Есть целый ряд задач, представляющих практический интерес, для которых, по-видимому, не существует эффективных алгоритмов, а про имеющиеся разрешающие алгоритмы можно сказать, что они неэффективны даже при реализации на ЭВМ будущих поколений.

Оценка сложности алгоритмов и создание эффективных алгоритмов — одна из важнейших задач современной дискретной математики.

Мы будем оценивать алгоритмы лишь по временной сложности. Не приводя здесь точных понятий, заметим, что время работы алгоритма можно выразить в виде функции от «размеров» входных данных, требуемых для описания задачи. Входные и выходные данные можно кодировать некоторым «разумным» способом в виде двоичных последовательностей из нулей и единиц. Тогда алгоритм можно рассматривать как последовательность двоичных операций, работающих с памятью из двоичных же символов.

Временная сложность алгоритма отражает затраты времени (число шагов), требуемые для его работы. Это есть функция, которая каждой входной длине n ставит в соответствие минимальное время, затрачиваемое алгоритмом на решение всех однотипных индивидуальных задач этой длины.

Из курса математического анализа известно, что функция $f(n)$ есть $O(g(n))$, если существует константа c такая, что $|f(n)| \leqslant c(g(n))$ для всех $n \geqslant 0$.

Полиномиальным алгоритмом (или алгоритмом полиномиальной временной сложности) называется алгоритм, у которого временная сложность равна $O(P(n))$, где $P(n)$ — некоторая полиномиальная функция от входной длины n. Алгоритмы, для временной сложности которых не существует такой оценки, называются *экспоненциальными*.

Задача считается *труднорешаемой*, если для нее не существует разрешающего полиномиального алгоритма.

В книге [3] приведены характерные таблицы оценки зависимости времени работы алгоритмов сложности, равной n, n^3, 2^n, от размеров входной информации (табл. 4.13), а также оценки максимальной размерности задачи, решаемой за один час (табл. 4.14). Из последней таблицы видно, что повышение быстродействия машины не дает существенного роста размерности разрешимых задач, если алгоритм имеет экспоненциальную сложность.

Таблица 4.13

Функция временной сложности \ Размерность n	10	20	50	60
n	0,00001 секунд	0,00002 секунд	0,0005 секунд	0,0006 секунд
n^3	0,001 секунд	0,008 секунд	0,125 секунд	0,216 секунд
2^n	0,001 секунд	1,0 секунд	35,7 дней	366 столетий

Таблица 4.14

Функции временной сложности	Современная ЭВМ	ЭВМ, в 100 раз более быстрая	ЭВМ, в 1000 раз более быстрая
n	N_1	$100N_1$	$1000N_1$
n^3	N_2	$4{,}64N_2$	$10N_2$
2^n	N_3	$N_3+6{,}64$	$N_3+9{,}97$

Многие прикладные задачи могут быть сформулированы в терминах теории графов. Однако при решении реальных задач характеристики соответствующих графов (например, число вершин n и число ребер m) весьма велики, и их анализ возможен лишь с привлечением ЭВМ.

Задача поиска эффективных алгоритмов в теории графов имеет большое практическое значение. Поэтому в заключение дадим краткий обзор временной сложности рассмотренных нами комбинаторных алгоритмов на графах.

Задачи теории графов можно классифицировать следующим образом:

1) задачи, для которых имеются алгоритмы сложности $O(n+m)$;

2) задачи, для которых известны алгоритмы сложности $O(P(n+m))$, где P — нелинейный полином;

3) задачи, для которых известны только алгоритмы экспоненциальной сложности, но про которые нельзя сказать, что для них нет алгоримов полиномиальной сложности;

4) задачи, которые могут быть решены лишь с использованием алгоритма экспоненциальной сложности.

Разбиение задач на первые три класса зависит от мастерства математика. Однако задачи четвертого класса принципиально нельзя отнести к первым трем классам, поскольку они включают в себя экспоненциальные процедуры, например просмотр всех подграфов исходного графа.

К четвертому классу относятся задачи перечисления всех остовных деревьев графа и перечисления всех циклов графа. Это объясняется тем, что у полного n-вершинного графа число остовных деревьев равно n^{n-2} (см. разд. 4.4.3), число циклов длины i равно $C^i_n(i-1)!$, т. е. общее число циклов составляет

$$\sum_{i=2}^{n} C^i_n(i-1)! > (n-1)!.$$

К первому классу относятся следующие рассмотренные нами задачи: отыскание остовного дерева графа; выделение компоненты связности графа и сильной связности орграфа; нахождение минимального пути в орграфе (или минимального маршрута в графе); нахождение эйлерового цикла. Для решения этих задач построены алгоритмы, имеющие сложность $O(n+m)$.

Второй класс в настоящее время включает в себя такие задачи: нахождение матрицы связности (приведенный в разд. 4.1.8 алгоритм Уоршела имеет сложность $O(n^3)$); нахождение циклового базиса (сложность алгоритма $O(nm)$); нахождение минимального пути в нагруженном графе методом Форда — Беллмана (сложность алгоритма $O(n^3)$); нахождение минимального остовного дерева в нагруженном графе (сложность алгоритма $O(m^2)$).

Получение оценок сложности алгоритмов выходит за рамки данной книги. Со способами вычисления этих оценок, а также с алгоритмами, позволяющими улучшить их, можно ознакомиться в специальных монографиях, например в [9].

Третий класс содержит следующие задачи: о существовании в графе гамильтонова цикла (задача коммивояжера); об опре-

делении изоморфизма данного графа G_1 какому-нибудь подграфу графа G_2; не относящуюся непосредственно к теории графов, но фундаментальную задачу о выполнимости формулы логики высказываний, находящейся в КНФ, и многие другие задачи. Для решения этих задач не существует полиномиальных алгоритмов, хотя не кажется безусловным, что их решение возможно лишь с помощью экспоненциальных алгоритмов. Для многих задач этого класса справедливо следующее свойство сводимости: существование полиномиального алгоритма для решения одной из них дало бы полиномиальный алгоритм для решения другой. В современной дискретной математике третий класс задач является предметом пристального изучения.

ЛИТЕРАТУРА

1. *Биркгоф Г., Барти Т.* Современная прикладная алгебра. — М.: Мир, 1976.

2. *Гаврилов Г. П., Сапоженко А. А.* Сборник задач по дискретной математике. — М.: Наука, 1977.

3. *Гери М., Джонсон Д.* Вычислительные машины и трудноразрешаемые задачи. — М.: Мир, 1982.

4. *Кострикин А. И.* Введение в алгебру. — М.: Наука, 1977.

5. Комбинаторный анализ: задачи и упражнения/Под общ. ред. К. А. Рыбникова. — М.: Наука, 1982.

6. *Кофман А.* Введение в прикладную комбинаторику. — М.: Наука, 1975.

7. *Лавров И. А., Максимова Л. Л.* Задачи по теории множеств, математической логике и теории алгоритмов. — М.: Наука, 1984.

8. *Мендельсон Э.* Введение в математическую логику. — М.: Наука, 1976.

9. *Новиков П. С.* Элементы математической логики. — М.: Наука, 1973.

10. *Рейнгольд Э., Нивергельт Ю., Део Н.* Комбинаторные алгоритмы: теория и практика. — М.: Мир, 1980.

11. *Яблонский С. В.* Введение в дискретную математику. — М.: Наука, 1979.

ОГЛАВЛЕНИЕ

Учебное издание

Нефедов Виктор Николаевич
Осипова Виктория Аркадьевна
КУРС ДИСКРЕТНОЙ МАТЕМАТИКИ

Редактор *Р. М. Белозерова*
Художественный редактор *Б. В. Челноков*
Технический редактор *М. Б. Жехова*
Корректор *В. Ф. Лекшова*
Обложка художника *Ю. П. Елкина*

ИБ № 15

Сдано в набор 31.10.90. Подписано в печать 01.04.92.
Формат 60×90¹/₁₆. Бум. тип. № 1 Гарнитура Литературная.
Печать высокая. Усл. печ. л. 16,5. Уч.-изд. л. 16,38.
Тираж 20 000 экз. Заказ № 1529. С3.

Издательство МАИ
125871, Москва, Волоколамское шоссе, 4.

Арендное предприятие Московская типография № 8
107078, Москва, Каланчевский туп., д. 3/6.

УВАЖАЕМЫЙ ЧИТАТЕЛЬ!

В ИЗДАТЕЛЬСТВЕ МАИ
В БЛИЖАЙШЕЕ ВРЕМЯ ВЫХОДЯТ СЛЕДУЮЩИЕ КНИГИ:

1. КУРС НАЧЕРТАТЕЛЬНОЙ ГЕОМЕТРИИ С АЛГОРИТМАМИ ДЛЯ ЭВМ: Учебник для инж.-техн. вузов / *А. М. Тевлин, Л. Г. Нартова, В. С. Полозов, В. И. Якунин:* Под ред. А. М. Тевлина и Л. Г. Нартовой.— М.: Изд-во МАИ, 1992.— 16 л.

2. МЕТОДИЧЕСКОЕ ПОСОБИЕ ПО МАТЕМАТИКЕ ДЛЯ ПОСТУПАЮЩИХ В ВУЗЫ / *В. А. Васильева, Т. Д. Кудрина, Р. Н. Молодожникова;* Под ред. Р. Н. Молодожниковой.— М.: Изд-во МАИ, 1992.— 20 л.

3. СБОРНИК ЗАДАЧ ПО МАТЕМАТИКЕ ДЛЯ ПОСТУПАЮЩИХ В ВУЗЫ: Учеб. пособие / *А. С. Бортаковский, В. М. Закалюкин, В. Н. Серегин, А. М. Скуридин;* Под ред. Р. Н. Молодожниковой.— М.: Изд-во МАИ, 1992.— 25 л.

4. *Поляков Д. Б., Круглов И. Ю.* ПРОГРАММИРОВАНИЕ В СРЕДЕ ТУРБО ПАСКАЛЬ (версия 5.5): Справ.-метод. пособие.— М.: Изд-во МАИ, 1992.— 45 л.

Заявки направлять по адресу:
125871, Москва, Волоколамское шоссе, 4.
Издательство МАИ.
Справки по телефону: 158-46-52.

Замеченные опечатки

С. 72, 15-я строка снизу. Читать: $M^n \to \{И, Л\}$.

С. 95, 5-я строка снизу. Читать: $a_{i_1} \ldots a_{i_t} \frac{a_j}{q_1} a_{i_{t+1}} \ldots a_{i_f}$

С. 128, 9-я строка сверху. Читать: $b = b_1 \ldots b_{2^r - 1}$