

С. Н. Поздняков
С. В. Рыбин

Дискретная математика

Версия 26.07.2020

Поздняков С. Н., Рыбин С. В.

Дискретная математика.

Учебник «Дискретная математика» предназначен для студентов технических университетов, обучающихся по направлению «информатика». В учебнике подробно освещены разделы теории чисел, связанных с криптографией, что делает учебник полезным для студентов специальностей, связанных с защите информации.

В разделе «Арифметика многочленов» наряду с классическими вопросами теории многочленов рассматриваются алгоритмы, важные для компьютерной математики, например, разложение многочлена на свободные от квадратов множители, основы теории базисов Гребнера.

Раздел «Теория множеств и комбинаторика» объединяет классические комбинаторные идеи и их обобщения с прикладной проблематикой, в том числе, генерированием комбинаторных объектов, кодированием. Подробно обсуждается техника работы с производящими функциями.

В главе «Графы и бинарные отношения», наряду с обсуждением классической теории графов, выделен подраздел, посвященный алгоритмам на графах. Данный подраздел помогает конструктивно взглянуть на основные идеи главы, связывает курс с курсом «информатика». Этот подраздел можно также назвать «прикладной теорией алгоритмов», так как в нем на важных примерах обсуждаются общие принципы доказательства корректности алгоритмов и их эффективности.

Учебник включает материал, обеспечивающий чтение курса «Математической логики и теории алгоритмов» для студентов технических вузов. Это материал третьей главы учебника. В эту главу включен материал по булевым функциям, позволяющий организовать доступные практические занятия на этапе начала изучения курса логики и подготовить студентов к изучению курса. Другой существенной частью главы является обсуждение исчисления высказываний и предикатов с точки зрения автоматизации доказательства теорем. Выбранный конструктивный подход призван мотивировать студентов, ориентированных на получение инженерных специальностей.

Завершает учебник обсуждение различных формализаций понятия алгоритма, формализации понятий языка и грамматики.

Предисловие

Курс дискретной математики, читающийся на первых курсах технических университетов, с первых же лекций открывает интересные перспективы для молодых людей, стремящихся стать специалистами в области информационных технологий.

В чем же особенность этого курса?

□ Во-первых, курс дискретной математики строится на базе известных из курса средней школы математических идей. Идеи эти достаточно разноплановы, чтобы заинтересовать людей с различными интересами и разной математической подготовкой.

□ Во-вторых, обсуждаемые в курсе идеи быстро приводят к интересным приложениям математической теории к практике, появляется возможность экспериментировать с ними уже с первых месяцев обучения в вузе. Эти эксперименты с прикладными алгоритмами могут реализовываться в курсе информатики и программирования, обычно читающемся параллельно с курсом дискретной математики, либо в индивидуальной работе студентов.

□ В-третьих, дискретная математика открывает большие перспективы для самостоятельного углубленного изучения курса, что чрезвычайно важно в подготовке мыслящего ученого или инженера.

В настоящее время курс дискретной математики для специальностей, связанных с информатикой, достаточно устоялся. В него входят такие разделы, как комбинаторика, теория графов, сведения из теории чисел и многочленов, знакомство с общими математическими понятиями, такими как отношения, поля, кольца, группы. В то же время, положение некоторых разделов, например «Булевы функции» неоднозначно, они включается как в курс дискретной математики, так и в курс основ математической логики. Собрав в одном учебнике материалы по дискретному анализу и математической логике, авторы надеются преодолеть это противоречие.

Другой особенностью учебника является большее чем обычно внимание к теории чисел и многочленам. Этот акцент связан с развитием криптографии, существенно опирающейся на результаты теории чисел.

В учебник также включен материал по многочленам нескольких переменных, которые играют большую роль в автоматизации математических вычислений.

Учебник будет полезен студентам технических вузов, специализирующихся в направлении «информатика».

Авторы постарались так скомпоновать материал, чтобы, с одной стороны, дать материал для практических занятий, познакомить студентов с важными идеями на несложных примерах, дать им возможность освоить в совершенстве необходимую технику вычислений, обсуждаемые алгоритмы, а с другой — последовательно и доказательно изложить теоретический материал, который может быть осмыслен на разных уровнях формализма и не обязательно при первом прочтении учебника.

Глава 1

Целочисленные алгоритмы

1.1. Арифметика целых чисел

1.1.1. Деление с остатком

Теорема 1.1 (о делимости с остатком). Для любых $a, b \in \mathbb{Z}$, где $b \neq 0$, существуют, и притом единственны $q, r \in \mathbb{Z}$, такие, что имеет место представление

$$a = bq + r, \quad 0 \leq r < |b|. \quad (1.1)$$

Доказательство. Не умаляя общности, предположим, что $a > b > 0$. Все остальные случаи или сводятся к этому, или являются тривиальными. Далее будем рассматривать числа, кратные b , т. е. числа вида $b, 2b, 3b, \dots, kb, \dots$. Очевидно, что найдется число $q \geq 1$, такое что

$$bq \leq a < (q+1)b. \quad (1.2)$$

Обозначим $r = a - bq$. Из (1.2) следует, что $0 \leq r < b$. Представление (1.1) установлено.

Докажем единственность представления (1.1). Пусть

$$a = bq + r = bq_1 + r_1, \quad 0 \leq r < |b|, \quad 0 \leq r_1 < |b|.$$

Тогда $b(q - q_1) = r_1 - r$. Но $|r_1 - r| < |b|$, а $|b(q - q_1)| \geq |b|$. Противоречие, таким образом, $q = q_1, r = r_1$. ■

Число q в (1.1) называют целой частью дроби $\frac{a}{b}$ и обозначают $\left\lfloor \frac{a}{b} \right\rfloor$ или $a \div b$, а число r — остатком и обозначают $\langle a \rangle_b$ или $a \bmod b$.

Определение 1.1. Будем говорить, что a делится на b (и обозначать¹ как $a : b$), если в представлении (1.1) для a, b остаток r равен нулю.

¹Выражение b делит a часто обозначают, как $b \mid a$.

Историческая справка. Знак $a:b$ для обозначения операции деления числа a на число b впервые ввел Г. Лейбниц в 1684 г.

Сформулируем простейшие свойства делимости, вытекающие непосредственно из определения.

Теорема 1.2. Справедливы следующие утверждения:

- 1) $a:b, b:c$, тогда $a:c$ (транзитивность деления);
- 2) $a_1, \dots, a_k:c$, тогда для любого набора $\lambda_1, \dots, \lambda_k \in \mathbb{Z}$ справедливо $\sum_{i=1}^k a_i \lambda_i : c$;
- 3) $a:b$, тогда $\pm a : \pm b$.

Доказательство. Докажем, например, первое свойство. Из условия теоремы имеем: $a = bq, b = cl$ тогда $a = (cl)q = c(lq)$. ■

Упражнение 1.1. Докажите самостоятельно свойства 2 и 3 теоремы 1.2.

Замечание 1.1. Последнее свойство позволяет в вопросах делимости ограничиться рассмотрением только натуральных чисел.

1.1.2. Наибольший общий делитель, наименьшее общее кратное и их свойства

Определение 1.2. Пусть $a_1, \dots, a_k, c \in \mathbb{N}$. Если $c:a_1, \dots, c:a_k$, то говорят, что c есть *общее кратное* a_1, \dots, a_k . Наименьшее среди всех кратных называют *наименьшим общим кратным* (НОК) и обозначают $M(a_1, \dots, a_k)$ или $\{a_1, \dots, a_k\}$. Заметим, что

$$\max\{a_1, a_2, \dots, a_k\} \leq M(a_1, \dots, a_k) \leq a_1 a_2 \cdots a_k.$$

Очевидно, что любое кратное делится на наименьшее общее кратное. Действительно, пусть m — общее кратное a_1, a_2, \dots, a_k . Разделим m на M с остатком. Тогда согласно (1.1)

$$m = M(a_1, \dots, a_k)q + r, \quad 0 \leq r < M(a_1, \dots, a_k). \quad (1.3)$$

Из (1.3) получаем, что r — также общее кратное a_1, a_2, \dots, a_k , причем меньшее $M(a_1, \dots, a_k)$. Следовательно $r = 0$.

Определение 1.3. Пусть $a_1, \dots, a_k, d \in \mathbb{N}$. Если $a_1:d, \dots, a_k:d$, то говорят, что d есть *общий делитель* a_1, \dots, a_k . Наибольший среди всех делителей называют *наибольшим общим делителем* (НОД) и обозначают $D(a_1, \dots, a_k)$ или (a_1, \dots, a_k) . Заметим, что

$$1 \leq D(a_1, \dots, a_k) \leq \min\{a_1, \dots, a_k\}.$$

Покажем, что если d_1, \dots, d_n — общие делители a_1, \dots, a_k , то

$$D(a_1, \dots, a_k) = M(d_1, \dots, d_n). \quad (1.4)$$

Действительно, любое a_i по определению является общим кратным чисел d_1, \dots, d_n . Следовательно, $a_i \vdots M(d_1, \dots, d_n)$, тогда $M(d_1, \dots, d_n)$ есть общий делитель всех a_i . Но, согласно определению $M(d_1, \dots, d_n) \vdots d_i, i \in \{1, \dots, k\}$, отсюда получаем (1.4).

Замечание 1.2. Равенство (1.4) можно использовать в качестве другого определения наибольшего общего делителя.

Введем одно из важнейших понятий теории делимости.

Определение 1.4. Числа a_1, \dots, a_k называют *взаимно простыми*, если $D(a_1, \dots, a_k) = 1$.

Установим некоторые свойства наибольшего общего делителя и наименьшего общего кратного.

Теорема 1.3. Справедливы следующие утверждения:

- 1) $d = D(a_1, \dots, a_k) \Leftrightarrow D\left(\frac{a_1}{d}, \dots, \frac{a_k}{d}\right) = 1$;
- 2) $d = D(a_1, \dots, a_k)$, тогда $D(a_1b, \dots, a_kb) = db$;
- 3) если c — общий делитель a_1, \dots, a_k , то $D\left(\frac{a_1}{c}, \dots, \frac{a_k}{c}\right) = \frac{d}{c}$;
- 4) $ab = D(a, b)M(a, b)$.

Доказательство. Утверждения 1–3 достаточно просты и предлагаются для самостоятельной работы. Докажем утверждение 4. Имеем ab — кратное a и b , поэтому $ab \vdots M(a, b)$. Тогда

$$d := \frac{ab}{M(a, b)}, \quad \frac{a}{d} = \frac{M(a, b)}{b}, \quad \frac{b}{d} = \frac{M(a, b)}{a}. \quad (1.5)$$

Так как $M(a, b)$ делится на a и b из (1.5) следует, что d — общий делитель a и b . Пусть d_1 — произвольный делитель этих чисел. Тогда

$$\frac{ab}{d_1} = a \left(\frac{b}{d_1} \right) = b \left(\frac{a}{d_1} \right), \quad \text{следовательно } \frac{ab}{d_1} \text{ — общее кратное } a, b, \text{ и}$$

$$\frac{ab}{d_1} : M(a, b) = \frac{d}{d_1} \in \mathbb{Z},$$

получаем для произвольного делителя $d \vdots d_1$, таким образом $d = D(a, b)$. ■

Упражнение 1.2. Докажите самостоятельно утверждения 1–3 теоремы 1.3.

Вопрос о нахождении $D(a, b)$ будет решен далее (см. 2.1). Предположим, что имеется эффективный алгоритм его вычисления. Поставим вопрос о вычислении $D(a_1, \dots, a_k)$. Ответ на этот вопрос дает следующая теорема.

Теорема 1.4. Справедливо равенство

$$D(a_1, a_2, a_3) = D(D(a_1, a_2), a_3).$$

Доказательство. Введем обозначения: $D(a_1, a_2) = e, D(e, a_3) = d$. Тогда в силу транзитивности делимости (теорема 1.2) имеем $a_1, a_2 : d$, но и $a_3 : d$, следовательно, d — общий делитель a_1, a_2, a_3 . Пусть d_1 — произвольный общий делитель a_1, a_2, a_3 . Тогда $e : d_1$, следовательно, d_1 — общий делитель e, a_3 . Тогда $d : d_1$ и, следовательно, $d = D(a_1, a_2, a_3)$. ■

Для наименьшего общего кратного можно получить результат, аналогичный теореме 1.4.

Теорема 1.5. Справедливо равенство

$$M(a_1, a_2, a_3) = M(M(a_1, a_2), a_3).$$

Установим еще несколько свойств взаимно простых чисел.

Теорема 1.6. Если $ab : c$ и $D(a, c) = 1$, тогда $b : c$.

Доказательство. $ab : c$ и $ab : a$, следовательно ab — кратное чисел a и c . Тогда $ab : M(a, c)$, но согласно теореме 1.3 $M(a, c) = ac$. Следовательно $ab : ac$. Отсюда получаем требуемое. ■

Следствие 1.1. Пусть a имеет набор взаимно простых делителей:

$$\left\{ m_1, m_2, \dots, m_k \mid a : m_i, i \in D(m_i, m_j) = 1, i \neq j \right\}.$$

Тогда $m_1 m_2 \cdots m_k$ тоже делитель числа a .

Доказательство. Докажем утверждение для $k = 2$. Имеем $a : m_1$, следовательно $a = cm_1$. Тогда $cm_1 : m_2$ и $D(m_1, m_2) = 1$, по теореме 1.6 получаем $c : m_2$. Таким образом, $a : m_1 m_2$. Проведенные рассуждения легко продолжить для любого k . ■

Упражнение 1.3. Докажите утверждение: если $D(a, c) = 1$, то $D(ab, c) = D(b, c)$.

1.1.3. Простые числа

Определение 1.5. Натуральное число $p > 1$ называют *простым*, если оно делится только на $\pm p$ и на ± 1 .

Теорема 1.7. Пусть p — простое число. Справедливо утверждение: $ab : p$ тогда и только тогда, когда $a : p$ или $b : p$.

Доказательство. Пусть $ab : p$. Предположим, что a не делится на p . Тогда $D(a, p) = 1$. По теореме 1.6 получаем требуемое. В обратную сторону утверждение теоремы очевидно. ■

Лемма 1.1. Любое $a \in \mathbb{Z}, a \neq 1$ имеет по крайней мере один простой делитель.

Доказательство. Пусть d_1, d_2, \dots, d_n — все положительные делители числа a , кроме 1. Положим $p = \min\{d_1, d_2, \dots, d_n\}$. Если бы p было составным, то его делитель (меньший, чем само p) был бы делителем a . Противоречие с определением p . ■

Теорема 1.8 (основная теорема теории делимости). Любое число $a \in \mathbb{Z}$ раскладывается и только одним способом на простые сомножители. Соединив одинаковые множители в степени, получаем **каноническое разложение**

$$a = p^\alpha q^\beta r^\gamma \cdots, \text{ где } p, q, r \text{ — простые числа; } \alpha, \beta, \gamma \geq 1. \quad (1.6)$$

Доказательство. По лемме 1.1 любое число a имеет простой делитель p . Представим его в виде pa_1 . Если a_1 — составное, то воспользуемся леммой 1.1 для a_1 . Заметим, что $a_1 < a$, поэтому на некотором шаге получим a_k простое. Таким образом, имеем представление

$$a = p_1 p_2 p_3 \cdots p_k, \text{ где } p_1, p_2, p_3 \cdots p_k \text{ — простые.}$$

Объединив одинаковые множители в степени, приходим к (1.6). Докажем единственность представления (1.6). Пусть имеется два различных представления

$$a = p_1 p_2 p_3 \cdots p_k = q_1 q_2 q_3 \cdots q_n. \quad (1.7)$$

Тогда $p_1 p_2 p_3 \cdots p_k : q_1$. По теореме 1.7 один из сомножителей слева делится на q_1 . Пусть для определенности, $p_1 : q_1$. Но p_1 и q_1 — простые числа. Поэтому $p_1 = q_1$. Тогда (1.7) можно переписать в виде

$$p_2 p_3 \cdots p_k = q_2 q_3 \cdots q_n.$$

Повторив эту процедуру, приходим к единственности представления (1.6). ■

Следствие 1.2. Пусть a и b представлены в каноническом виде (1.6):

$$a = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_n^{\alpha_n}, \quad b = p_1^{\beta_1} p_2^{\beta_2} \cdots p_n^{\beta_n}.$$

Тогда

$$D(a, b) = p_1^{\gamma_1} p_2^{\gamma_2} \cdots p_n^{\gamma_n}, \quad M(a, b) = p_1^{\delta_1} p_2^{\delta_2} \cdots p_n^{\delta_n},$$

где $\gamma_i = \min\{\alpha_i, \beta_i\}$, а $\delta_i = \max\{\alpha_i, \beta_i\}$.

Историческая справка. В 1876 г. французский математик Ф. Люка доказал, что число $2^{127} - 1$ является простым, и 75 лет оно оставалось наибольшим из известных простых чисел, что не покажется удивительным, если на него взглянуть:

$$2^{127} - 1 = 170141183460469231731687303715884105727.$$

В настоящее время составлены таблицы всех простых чисел, не превосходящих 50 миллионов, далее известны только отдельные их представители. Укажем здесь два самых больших известных в настоящее время простых числа: $2^{44497} - 1$ и $2^{86243} - 1$. Последнее число записано пока в книгу рекордов Гиннесса, в нем 25 962 десятичных знака.

Определение 1.6. Число $a \in \mathbb{Z}$ называется *свободным от квадратов*, или *бесквадратным*, если не делится ни на один квадрат целого числа, кроме 1.

Например, 14 – свободное от квадратов, а 12 – нет, так как $12 = 2^2 \times 3$.

Очевидно следующее утверждение.

Лемма 1.2. Число $n \in \mathbb{N}$ свободно от квадратов тогда и только тогда, когда для любого его разложения на множители $n = ab$, $D(a, b) = 1$.

1.1.4. Решето Эратосфена. Разложение числа на простые множители

В связи с полученным представлением (1.6) возникает необходимость построить эффективные алгоритмы для решения следующих задач:

- 1) найти все простые числа в данном интервале;
- 2) для любого числа $a \in \mathbb{Z}$ получить его разложение в виде (1.6).

Для решения первой задачи рассмотрим алгоритм, который называют **решетом Эратосфена**¹. Он позволяет найти все простые числа не превосходящие N .

Алгоритм 1.1 (решето Эратосфена).

Выписываем подряд все целые числа от 2 до N

¹Эратосфен Киренский (276 – 194 гг. до н.э.) — греческий математик, астроном, географ и поэт. Один из интересных фактов жизни Эратосфена — вычисление радиуса Земли.

```

 $p \leftarrow 2$ 
while  $p < N$  do
    Вычеркиваем все числа, от  $2p$  до  $N$ , делящиеся на  $p$ 
     $p \leftarrow$  первое не вычеркнутое число, большее  $p$ 
end while

```

Замечание 1.3. Алгоритм 1.1 можно улучшить следующим образом. Обрабатывать можно только нечетные числа, начиная с $p = 3$. Числа можно вычеркивать, начиная сразу с числа p^2 , потому что все составные числа меньше его уже будут вычеркнуты к этому времени. И, соответственно, останавливать алгоритм можно, когда $p^2 \geq N$. Это вытекает из теоремы 1.9.

Теорема 1.9. Всякое составное число a имеет делитель $b \leq \sqrt{a}$.

Доказательство. Если $a = bc$, то из пары чисел b, c одно больше \sqrt{a} , а другое меньше, за исключением случая, когда a — точный квадрат, тогда $b = c = \sqrt{a}$. ■

Пример 1.1 (решето Эратосфена). Пусть $N = 50$. Тогда после окончания работы алгоритма (при $p = 7$) получаем следующую картину (на месте вычеркнутых чисел стоит знак *):

3 5 7 * 11 13 * 17 19 * 23 * * 29 31 * * 37 * 41 43 * 47 * .

Возникает вопрос: сколько же простых чисел? Ответ был получен еще Евклидом:

Теорема 1.10 (теорема Евклида). Множество простых чисел бесконечно.

Доказательство. Допустим конечность множества простых чисел: $\{p_1, p_2, p_3, \dots, p_k\}$. Положим $r = p_1 p_2 p_3 \cdots p_k + 1$. Очевидно, что число r не делится ни на одно из p_i . Таким образом, либо r — простое, либо имеет простой делитель, больший любого из p_i . Полученное противоречие и доказывает теорему. ■

Мы показали, что число простых чисел бесконечно. Пусть $\pi(n)$ — функция, возвращающая число простых чисел, меньших n . Точно задать эту функцию не удается, но для нее есть хорошая оценка. Справедлива следующая теорема:

Теорема 1.11 (теорема Чебышева).

$$\lim_{x \rightarrow \infty} \frac{\pi(n)}{n} = \frac{1}{\ln(n)} \tag{1.8}$$

Предельное соотношение (1.8) означает, что у случайно выбранного числа от 1 до n шанс оказаться простым примерно равен $1/\ln(n)$. Таким

Таблица 1.1

n	10^5	10^6	10^9	10^{12}	10^{15}
$\ln(n)$	11.51	13.81	20,72	27.63	34.53
$1/\ln(n)$	0.08686	0.07238	0.048254	0.036191	0.02895

образом, если взять порядка $\ln(n)$ случайных чисел от 1 до n , то при больших n высока вероятность, что хотя бы одно из этих чисел будет простым. Значения $1/\ln(n)$ для различных n приведены в таблице 1.1.

Функция $\pi(n)$ асимптотически сверху приближается к своему пределу, так что оценка (1.8) дает слегка заниженные значения.

Перейдем к решению второй поставленной задачи — разложению числа a на простые множители в виде (1.6). Отыскание простых множителей натурального числа называют для краткости «факторизацией» («фактор», значит — множитель, составная часть; в этом последнем смысле слово обычно и употребляется).

Начнем с простейшего алгоритма, который называют **методом пробных делителей**.

Введем обозначения:

$\{2 = p_0 < \dots < p_n \leq \sqrt{a}\}$ — последовательность *пробных делителей* — простых чисел.

$\{d_0, \dots, d_s \leq \sqrt{a}\}$ — последовательность найденных делителей числа a .

Алгоритм 1.2 (метод пробных делителей).

```

 $k \leftarrow 0, i \leftarrow 0$ 
while  $a > 1$  do
     $q \leftarrow a \div p_k$                                  $\triangleright$  Деление с остатком:  $a = qp_k + r$ 
     $r \leftarrow a \bmod p_k$ 
    if  $r = 0$  then                                 $\triangleright p_k$  — делитель  $a$ , повторяем деление на  $p_k$ 
         $d_i \leftarrow p_k$ 
         $i \leftarrow i + 1$ 
         $a \leftarrow q$ 
    else if  $q > p_k$  then                       $\triangleright$  Переходим к следующему пробному делителю
         $k \leftarrow k + 1$ 
    else
         $d_i \leftarrow a$                                  $\triangleright a$  — простое число, завершаем работу
        return
    end if
end while

```

Пример 1.2 (метод пробных делителей). Рассмотрим работу алгоритма 1.2 на примере.

Пусть $a = 6930$. Возьмем последовательность пробных делителей

$$\{2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, \dots\}$$

- 1) $a = 2 \cdot 3465, d_0 = 2, i = 1, a = 3465;$
- 2) $a = 2 \cdot 1732 + 1, k = 1;$
- 3) $a = 3 \cdot 1155, d_1 = 3, i = 2, a = 1155;$
- 4) $a = 3 \cdot 385, d_2 = 3, i = 3, a = 385;$
- 5) $a = 3 \cdot 128 + 1, k = 2;$
- 6) $a = 5 \cdot 77, d_3 = 5, i = 4, a = 77;$
- 7) $a = 5 \cdot 15 + 2, k = 3;$
- 8) $a = 7 \cdot 11, d_4 = 7, i = 5, a = 11;$
- 9) $a = 7 \cdot 1 + 4, d_5 = 11.$

Имеем: $a = 6930 = 2 \cdot 3 \cdot 3 \cdot 5 \cdot 7 \cdot 11 = 2 \cdot 3^2 \cdot 5 \cdot 7 \cdot 11.$

Легко видеть, что описанный алгоритм 1.2 эффективно работает на небольших числах. При их увеличении быстро растет число «холостых» делений.

Рассмотрим алгоритм, предложенный французским математиком Пьером Ферма, который, *используя только операции умножения и сложения (без делений)*, позволяет представить любое число в виде произведения двух (не обязательно простых) сомножителей.

Историческая справка. Пьер Ферма (1601—1665 гг.), крупный юрист, видный общественный деятель своей родины — города Тулузы, — занимался математикой в часы досуга. О жизни его известно мало, книг он не печатал. Оставшиеся после него рукописи были изданы его сыном уже после смерти отца. Ферма состоял в переписке почти со всеми выдающимися математиками той эпохи; такой крупный учёный, как Паскаль, считал его лучшим математиком своего времени. Одновременно с Декартом Ферма заложил основы аналитической геометрии, вместе с Паскалем — основы теории вероятностей. Но лучшие его открытия принадлежат теории чисел.

Не умаляя общности, можно считать, что исходное число a является нечетным. Выделить степени двойки достаточно легко (сдвигами вправо двоичного представления числа). Будем искать представление a в виде

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

Введем обозначения: $R(x, y) = (x - y)(x + y) - a$. Требуется, поочередно увеличивая x и y на 1, добиться равенства $R(x, y) = 0$. Заметим, что

$$\begin{aligned} R(x + 1, y) &= (x - y + 1)(x + y + 1) - a = R(x, y) + 2x + 1, \\ R(x, y + 1) &= (x - y - 1)(x + y + 1) - a = R(x, y) - (2y + 1). \end{aligned}$$

Чтобы лишний раз не умножать на 2, введем обозначения: $R_x = 2x + 1$, $R_y = 2y + 1$. Тогда последние равенства примут следующий вид:

$$\begin{cases} R(x + 1, y) &= R(x, y) + R_x, \\ R(x, y + 1) &= R(x, y) - R_y. \end{cases} \quad (1.9)$$

Заметим, что при увеличении x или y на единицу R_x или R_y увеличиваются соответственно на двойку.

Будем считать, что известна приблизительная оценка целой части \sqrt{a} , обозначим ее через h :

$$h^2 \leq a < (h+1)^2 \quad (1.10)$$

Далее будет рассмотрен метод построения числа h , удовлетворяющего (1.10) — см. алгоритм 1.4.

Алгоритм 1.3 (метод Ферма).

```

 $R_x \leftarrow 2h + 1$                                  $\triangleright x = h$ 
 $R_y \leftarrow 1$                                       $\triangleright y = 0$ 
 $R(x, y) \leftarrow h^2 - a$ 
while  $R(x, y) \neq 0$  do
    if  $R(x, y) > 0$  then
         $R(x, y) \leftarrow R(x, y) - R_y$             $\triangleright$  Переходим к  $R(x+1, y)$  согласно 1.9
         $R_y \leftarrow R_y + 2$ 
    else
         $R(x, y) \leftarrow R(x, y) + R_x$             $\triangleright$  Переходим к  $R(x, y+1)$  согласно 1.9
         $R_x \leftarrow R_x + 2$ 
    end if
end while
 $a = \frac{R_x - R_y}{2} \left( \frac{R_x + R_y}{2} - 1 \right)$ 

```

Корректность работы описанного алгоритма устанавливает следующая теорема.

Теорема 1.12. Алгоритм 1.3 всегда завершает работу.

Доказательство. При доказательстве рассмотрим отдельно два случая составного и простого числа a на входе. Не умаляя общности считаем, что a — нечетное число.

Пусть a составное число, т. е. $a = bc$, $b \leq c$. При этом b и c нечетные числа. Тогда

$$a = bc = (x - y)(x + y) = x^2 - y^2$$

Поскольку $x - y \leq x + y$, положим $b = x - y$ и $c = x + y$. Получим

$$x = \frac{c+b}{2}, \quad y = \frac{c-b}{2}. \quad (1.11)$$

Так как b и c нечетные числа, x и y — целые. При $b = c$ число a есть полный квадрат и алгоритм Ферма находит делитель на шаге 1. Будем предполагать, что a не является полным квадратом. Тогда,

$$1 < b < c < a.$$

Покажем, что

$$\lfloor \sqrt{a} \rfloor < \frac{c+b}{2} < \frac{a+1}{2}. \quad (1.12)$$

Рассмотрим правое неравенство (1.12). Заменяя a на bc и вычитая из обеих частей $c+1$, мы приходим к неравенству

$$b-1 < bc-c.$$

Однако $b > 1$, поэтому обе части последнего неравенства можно разделить на $b-1$. В результате получаем равносильное неравенство $c > 1$. Таким образом, правая часть неравенства (1.12) доказана.

Установим теперь левое неравенство (1.12). Отметим, что поскольку,

$$\lfloor \sqrt{a} \rfloor \leq \sqrt{a},$$

достаточно доказать, что

$$\sqrt{a} \leq \frac{b+c}{2}.$$

Очевидно, что последнее неравенство выполняется тогда и только тогда, когда

$$a \leq \frac{(b+c)^2}{4}.$$

Запишем очевидное тождество

$$\left(\frac{c+b}{2} \right)^2 - \left(\frac{c-b}{2} \right)^2 = bc = a. \quad (1.13)$$

Отсюда

$$\left(\frac{c+b}{2} \right)^2 - a = \left(\frac{c-b}{2} \right)^2 \geq 0.$$

Обе части неравенства (1.12) установлены.

Вернемся к алгоритму. Напомним, что стартовые значения переменных x и y равны соответственно $\lfloor \sqrt{a} \rfloor$ и 1, одна из них увеличивается на 1 на каждом шаге алгоритма. Из неравенства (1.12) следует, что если число a составное, то x дойдет до значения $(c+b)/2$ раньше, чем до $(a+1)/2$. При этом значения x и y алгоритма связаны равенством (1.11). Поэтому если a составное, то алгоритм всегда останавливается при некотором $x < (a+1)/2$, вычислив два делителя этого числа.

Замечание 1.4. Заметим, что данное составное число a можно представить в виде $a = bc$, где $1 < b < c < a$, возможно, несколькими различными способами. Какое

из разложений находит алгоритм Ферма? Поиск начинается при $x = \lfloor \sqrt{a} \rfloor$, и x увеличивается по ходу алгоритма. Значит, найденные алгоритмом делители b и c таковы, что разность

$$\frac{b+c}{2} - \lfloor \sqrt{a} \rfloor$$

наименьшая из возможных.

Пример 1.3. Возьмем $a = 2520$. Тогда по алгоритму 1.4 находим $h = \lfloor \sqrt{a} \rfloor = 50$. Протокол¹ работы алгоритма представлен в табл. 1.2.

Таблица 1.2

Итерация	$R(x, y)$	R_x	R_y	Итерация	$R(x, y)$	R_x	R_y
0	-20	101	1	6	56	103	11
1	81	103	1	7	45	103	13
2	80	103	3	8	32	103	15
3	77	103	5	9	17	103	17
4	72	103	7	10	0	103	19
5	65	103	9				

Имеем: $a = 2520 = 42 \times 60$.

Пример 1.4. Пусть $a = 221$, тогда $h = \lfloor \sqrt{a} \rfloor = 14$. Протокол работы алгоритма представлен в табл. 1.3.

Таблица 1.3

Итерация	$R(x, y)$	R_x	R_y	Итерация	$R(x, y)$	R_x	R_y
0	-25	29	1	2	3	31	3
1	4	31	3	3	0	31	5

Имеем: $a = 221 = 17 \times 13$.

Комбинируя *методы Ферма и пробных делителей* можно построить эффективный алгоритм для разложения достаточно большого числа a на простые множители в виде (1.6).

Вернемся к вопросу эффективного построения целочисленного приближения квадратного корня, удовлетворяющего неравенству (1.10). Для ее решения можно воспользоваться целочисленной² версией алгоритма на основе *итерационной формулы Герона*³.

¹Поля, изменяющиеся на данной итерации, выделены серым цветом

²Все операции в алгоритме проводятся только в целочисленном формате

³Герон Александрийский — греческий математик и механик. Предположительно,

Алгоритм 1.4 (вычисление квадратного корня в \mathbb{Z}).

```

 $h \leftarrow a$ 
while (1) do
     $\beta \leftarrow \left( h + \frac{a}{h} \right) / 2$ 
    if  $\beta < h$  then
         $h \leftarrow \beta$ 
    else
        return  $h$ 
    end if
end while

```

$\triangleright h = \lfloor \sqrt{a} \rfloor$

Замечание 1.5. При программной реализации алгоритма 1.4, операция деления на двойку реализуется как сдвиг числа на один бит.

Проведем обоснование корректности работы алгоритма 1.4. Справедлива следующая теорема.

Теорема 1.13. Число α , построенное по алгоритму 1.4 удовлетворяет неравенству (1.10).

Доказательство. Заметим, что для любого $h > 0$ справедливо:

$$h + \frac{a}{h} \geq 2\sqrt{a}. \quad (1.14)$$

Действительно, из очевидного неравенства $(h^2 - a)^2 \geq 0$ следует, что

$$(h^2 + a)^2 \geq 4h^2a, \quad \text{или} \quad h^2 + a \geq 2h\sqrt{a}.$$

Последнее неравенство равносильно (1.14).

Таким образом, из (1.14) следует, что последовательность вычисляемых в алгоритме 1.4 значений h убывает и при этом ограничена снизу: $h \geq \sqrt{a}$.

Покажем, что алгоритм 1.4 остановится только при выполнении условия $h = \lfloor \sqrt{a} \rfloor$.

Предположим противное, $h > \lfloor \sqrt{a} \rfloor$ тогда $h^2 > a$. Согласно условию шага 3 алгоритма 1.4 выполнено неравенство $\beta \geq h$. Далее имеем

$$\beta - h = \left(h + \frac{a}{h} \right) / 2 - h = \left(\frac{a}{h} - h \right) / 2 = \frac{a - h^2}{2h} < 0.$$

Полученное противоречие завершает доказательство. ■

Пример 1.5. Рассмотрим работу алгоритма 1.4 для нахождения начального значения $\lfloor \sqrt{a} \rfloor$ из примера 1.3. Протокол работы алгоритма представлен в таблице 1.4.

жил во второй половине первого века н. э. Подробности жизни Герона неизвестны. Его не без оснований считают одним из величайших инженеров за всю историю человечества. Например он первым изобрел паровую турбину, первым начал создавать программируемые устройства.

Таблица 1.4

	h	β		h	β
1	2520	$(2520 + 1)/2 = 1260$	5	162	$(162 + 15)/2 = 88$
2	1260	$(1260 + 2)/2 = 631$	6	88	$(88 + 28)/2 = 58$
3	631	$(631 + 3)/2 = 317$	7	58	$(58 + 43)/2 = 50$
4	317	$(317 + 7)/2 = 162$	8	50	$(50 + 50)/2 = 50$

Протокол работы алгоритма 1.4 для примера 1.4 показан в таблице 1.5.

Таблица 1.5

	h	β		h	β
1	221	$(221 + 1)/2 = 111$	5	18	$(18 + 12)/2 = 15$
2	1260	$(1260 + 2)/2 = 631$	6	15	$(15 + 14)/2 = 14$
3	56	$(56 + 3)/2 = 29$	7	14	$(14 + 15)/2 = 14$
4	29	$(29 + 7)/2 = 18$			

1.1.5. Позиционная запись натуральных чисел

Историческая справка. Изобретение позиционной нумерации, приписывается шумерам и вавилонянам (шестидесятеричная система счисления), затем такая нумерация продолжила свое развитие в Индии. Здесь приблизительно в V веке н. э., возникла десятичная система счисления. Далее, начинается ее быстрое распространение из Индии на Запад и Восток. В IX веке появляются рукописи на арабском языке, в которых излагается эта система счисления. Однако только к XVI веку новая нумерация получила широкое распространение в науке и в житейском обиходе. В России она начинает распространяться при Петре I и в начале XVIII века вытесняет славянскую алфавитную систему счисления¹, которая сохранилась до наших дней в богослужебных книгах.

Определение 1.7. Упорядоченный набор неотрицательных целых чисел $(a_n a_{n-1} \dots a_1 a_0)_p$ называют p -ичной записью натурального числа c (представлением числа c в p -ичной системе счисления или просто p -ичным числом), если

$$c = p^n a_n + p^{n-1} a_{n-1} + \dots + p a_1 + a_0, \quad (1.15)$$

где p — натуральное число, большее 1, $0 \leq a_k < p$ и $a_n \neq 0$.

¹В алфавитных системах для записи чисел использовался буквенный алфавит. В славянской системе роль цифр играли не все буквы, а только те, которые имеются в греческом алфавите. Над буквой, обозначающей цифру, ставился специальный знак — «титло».

Замечание 1.6. Числа a_k в p -ичной записи называют цифрами и обычно обозначают отдельными символами, например $10 = A$, $11 = B$ и т. д.

Теорема 1.14. Каждое натуральное число имеет единственную p -ичную запись.

Доказательство. Пусть c имеет две различные p -ичные записи: $a_n a_{n-1} \dots a_1 a_0$ и $b_m b_{m-1} \dots b_1 b_0$. Тогда

$$\begin{aligned} c &= p^n a_n + \dots + p a_1 + a_0 = p(p^{n-1} a_n + \dots + a_1) + a_0 = p c_1 + a_0 = \\ &= p^m b_m + \dots + p b_1 + b_0 = p(p^{m-1} b_m + \dots + b_1) + b_0 = p c_2 + b_0. \end{aligned}$$

Так как частное и остаток при делении на p определяются однозначно,

$$a_0 = b_0 \text{ и } c_1 = p^{n-1} a_n + \dots + a_1 = p^{m-1} b_m + \dots + b_1 = c_2.$$

Применяя аналогичные рассуждения к c_1 и c_2 , получим $a_1 = b_1$ и т. д. ■

Замечание 1.7. Существуют и другие способы позиционной записи натуральных чисел, т. е. представление их упорядоченными наборами цифр.

Пример 1.6 (факториальная запись).

$$c = (a_n a_{n-1} \dots a_1)! \Leftrightarrow c = a_n n! + a_{n-1} (n-1)! + \dots + a_1 \cdot 1!,$$

где $0 \leq a_k \leq k$, $a_n \neq 0$.

Упражнение 1.4. Докажите единственность факториальной записи натуральных чисел.

Замечание 1.8. Алгоритм построения факториальной записи числа будет получен далее (см. подразд. 1.6.8, алгоритм 1.35).

1.1.6. Алгоритмы арифметических действий с p -ичными записями натуральных чисел

Алгоритмы сложения, вычитания, умножения «столбиком» и деления «уголком» для p -ичных записей чисел совпадают с известными алгоритмами для десятичных записей, если заменить таблицы сложения (вычитания) и умножения (деления). В этом случае говорят о выполнении операций в p -ичной арифметике.

Для рассматриваемых далее алгоритмов введем следующие обозначения: $a = (a_n \dots a_0)_p$, $b = (b_m \dots b_0)_p$ — записи исходных чисел, $c = (c_l \dots c_0)_p$ есть результат работы алгоритма.

Алгоритм 1.5 (сложение p -ичных чисел).

```

if  $n < m$  then
     $a \longleftrightarrow b$ ;  $n \longleftrightarrow m$                                  $\triangleright$  длина числа  $a \geqslant$  длины числа  $b$ 
end if
 $s \leftarrow 0$                                                $\triangleright s$  — величина переноса в старший разряд
for  $i \leftarrow 0, m$  do
     $c_i \leftarrow (a_i + b_i + s) \bmod p$                            $\triangleright$  вычисление очередной с конца цифры результата
     $s \leftarrow (a_i + b_i + s) \div p$                              $\triangleright$  вычисление величины межразрядного переноса
end for
     $\triangleright$  продолжение сложения числа после «прохождения» старшего разряда числа меньшей
длины
for  $i \leftarrow m + 1, n$  do
     $c_i \leftarrow (a_i + s) \bmod p$ 
     $s \leftarrow (a_i + s) \div p$ 
end for
     $\triangleright$  формирование старшего разряда результата, если длина результата больше длины
слагаемых
if  $s > 0$  then
     $c_{n+1} \leftarrow s$ 
end if

```

Рассмотрим алгоритм умножения p -ичного числа на p -ичную цифру α , где $1 \leqslant \alpha < p$.

Алгоритм 1.6 (умножение p -ичного числа на цифру).

```

 $s \leftarrow 0$ 
for  $i \leftarrow 0, n$  do
     $c_i \leftarrow (a_i\alpha + s) \bmod p$ 
     $s \leftarrow (a_i\alpha + s) \div p$ 
end for
if  $s > 0$  then
     $m \leftarrow n + 1$ 
     $c_m \leftarrow s$ 
else
     $m \leftarrow n$ 
end if

```

Рассмотрим алгоритм умножения p -ичного числа на число p^k (сдвиг числа на k разрядов влево). Заметим, что число разрядов сдвинутого числа увеличится на k

Алгоритм 1.7 (сдвиг p -ичного числа на k разрядов влево).

```

for  $i \leftarrow n, 0$  do
     $c_{i+k} \leftarrow a_k$ 
end for
for  $i \leftarrow (k - 1), 0$  do
     $c_i \leftarrow 0$ 
end for

```

Перейдем к перемножению p -ичных чисел.

Алгоритм 1.8 (перемножение p -ичных чисел).

```

 $c \leftarrow 0$ 
for  $i \leftarrow 0, m$  do
     $c \leftarrow c + (ab_i)p^i$ 
     $\triangleright$  поразрядное умножение со сдвигом выполняется по алгоритмам 1.6 и 1.7
end for

```

Задание 1.1. Упростите алгоритмы арифметических операций для двоичных чисел.

Задание 1.2. Сформулируйте алгоритмы сложения и умножения натуральных чисел в факториальной записи.

Задание 1.3. Напишите алгоритмы вычитания и деления нацело, основываясь на таблицах вычитания и деления p -ичных чисел.

Пример 1.7. Используя рассмотренные алгоритмы 1.5 – 1.8 построим таблицы сложения (табл. 1.6,*a*) и умножения (табл. 1.6,*b*) для 5-ичных чисел.

Таблица 1.6

	<i>a</i>					<i>b</i>					
$(+)_5$	0	1	2	3	4	$(\times)_5$	0	1	2	3	4
0	0	1	2	3	4	0	0	0	0	0	0
1	1	2	3	4	10	1	0	1	2	3	4
2	2	3	4	10	11	2	0	2	4	11	13
3	3	4	10	11	12	3	0	3	11	14	22
4	4	10	11	12	13	4	0	4	13	22	31

1.1.7. Алгоритмы перевода p -ичной записи натурального числа в q -ичную

Первый из алгоритмов перевода использует p -ичную арифметику, второй — q -ичную. В основе первого алгоритма лежит та же идея, что и в доказательстве единственности p -ичной записи натурального числа.

Алгоритм 1.9 (перевод из p -ичной записи в q -ичную в p -ичной арифметике).

```

 $i \leftarrow 0$ 
while  $a \neq 0$  do
     $b_i \leftarrow a \bmod q$   $\triangleright$  деление  $a$  на  $q$  выполняется в  $p$ -ичной арифметике
     $a \leftarrow a \div q$ 

```

```

 $i \leftarrow i + 1$ 
end while
 $m \leftarrow i - 1$ 

```

Второй алгоритм основан на так называемой схеме Горнера [см. (1.75)]:

$$\begin{aligned}
a &= (a_n a_{n-1} \cdots a_0)_p = a_n p^n + a_{n-1} p^{n-1} + \cdots + a_1 p + a_0 = \\
&= ((\dots((a_n p + a_{n-1}) p + a_{n-2}) p + \cdots + a_1) p + a_0).
\end{aligned} \tag{1.16}$$

Алгоритм 1.10 (перевод из p -ичной записи в q -ичную в q -ичной арифметике).

```

 $b \leftarrow 0$ 
for  $i \leftarrow n, 0$  do
     $b \leftarrow bp + a_i$  ▷ действия выполняются в  $q$ -ичной арифметике
end for

```

Замечание 1.9. При $p > q$ применение последнего алгоритма подразумевает предварительный перевод всех цифр в q -ичную запись. Аналогично, при $p < q$ алгоритм 1.9 подразумевает дополнительные действия по переводу цифр порождаемой записи из p -ичной записи в q -ичную.

1.1.8. Алгоритм эффективного возведения числа в натуральную степень

Схема Горнера (1.16) может быть применена для построения эффективного алгоритма возведения числа a в натуральную степень m . Рассмотрим двоичную запись числа m :

$$\begin{aligned}
m &= (b_n \cdots b_0)_2 = 2^n b_n + \cdots + 2b_1 + b_0 = \\
&= ((\dots(b_n 2 + b_{n-1}) 2 + b_{n-2}) 2 + \dots) 2 + b_1) 2 + b_0,
\end{aligned} \tag{1.17}$$

где $b_i \in \{0, 1\}$, $b_n \neq 0$.

Тогда

$$\begin{aligned}
a^m &= a^{2m_1+b_0} = (a^{m_1})^2 a^{b_0} = \begin{cases} (a^{m_1})^2 & \text{при } b_0 = 0, \\ (a^{m_1})^2 a & \text{при } b_0 = 1, \end{cases} \text{ где} \\
m_1 &= (b_n \cdots b_1)_2 = 2^n b_n + \cdots + 2b_2 + b_1.
\end{aligned}$$

Алгоритм 1.11 (возведения числа в натуральную степень).

```

 $c \leftarrow 1$  ▷ Результат  $c = a^m$ 
for  $i \leftarrow n, 0$  do
     $c \leftarrow c \cdot c$ 
    if  $b_i > 0$  then
         $c \leftarrow c \cdot a$ 
    end if

```

end for

Замечание 1.10. Алгоритм 1.11 легко модифицируется для возведения a в степень m в кольце \mathbb{Z}_k . Для этого достаточно операции умножения осуществлять в кольце вычетов по модулю k (см. теорему 1.28).

Если двоичное представление степени (1.17) заранее неизвестно, алгоритм 1.11 можно модифицировать, встроив в него побитовую обработку числа m .

Алгоритм 1.12 (модификация алгоритма 1.11).

```

 $c \leftarrow 1$                                       $\triangleright$  Результат  $c = a^m$ 
 $x \leftarrow a$ 
 $y \leftarrow m$ 
while  $y \neq 0$  do
    while  $y \div 2 = 0$  do                          $\triangleright y$  — четное число
         $x \leftarrow x \cdot x$ 
         $y \leftarrow y/2$ 
    end while
     $c \leftarrow c \cdot x$ 
     $y \leftarrow y - 1$ 
end while

```

Пример 1.8. Вычислим $27^{125}(391)$. Согласно алгоритму 1.11, имеем (см. табл. 1.7).

$$m = 125 = (1111101)_2, \quad a = 27, \quad c = 1, \quad i = 6, 5, \dots, 0.$$

Таблица 1.7

Значение i	Значение c
6	$c = c^2 \cdot 27(391) = 27$
5	$c = c^2 \cdot 27 = 27^3(391) = 19\,683(391) = 133$
4	$c = c^2 \cdot 27 = 133^2 \cdot 27(391) = 47\,7603(391) = 192$
3	$c = c^2 \cdot 27 = 192^2 \cdot 27(391) = 995\,328(391) = 233$
2	$c = c^2 \cdot 27 = 233^2 \cdot 27(391) = 1\,465\,803(391) = 335$
1	$c = c^2 = 335^2(391) = 112\,225(391) = 8$
0	$c = c^2 \cdot 27 = 8^2 \cdot 27(391) = 1\,728(391) = 164$

Таким образом, $27^{125}(391) = 164$.

Задание 1.4. Сколько операций умножения (возведение в квадрат — это тоже операция умножения) выполнит алгоритм при вычислении: а) a^{2^k} ; б) a^{2^k-1} .

Задание 1.5. Докажите, что количество умножений в алгоритме возведения числа в степень m не превышает $2 \lceil \log_2 m \rceil$.

Задание 1.6. Найдите формулу, выражающую число умножений в алгоритме возведения a в степень m через l — число цифр в двоичной записи m и r — число нулей в двоичной записи m .

1.2. Алгоритм Евклида и цепные дроби

1.2.1. Классический алгоритм Евклида

Историческая справка. О жизни Евклида не имеется никаких достоверных сведений. Предположительно считается, что он жил во времена Птолемея I (IV-III вв. до н.э.). Наиболее знаменитое и выдающееся произведение Евклида — тринадцать книг его «Начал». Это первые математические труды, которые полностью дошли до нашего времени со времен древних греков. В истории Западного мира «Начала», после Библии, — наибольшее число раз изданная и более всего изучавшаяся книга. Большая часть школьной геометрии заимствована из первых шести книг «Начал».

Вернемся к ранее поставленному вопросу об эффективном алгоритме вычисления $D(a, b)$. Будем считать, что $a > b > 0$. Идея алгоритма Евклида очень проста:

$$\text{Если } a = bq + r, \text{ то } D(a, b) = D(a - bq, b) = D(b, r). \quad (1.18)$$

Таким образом, для нахождения $D(a, b)$ получаем последовательность действий, называемую *классическим алгоритмом Евклида*:

$$\left\{ \begin{array}{l} a = bq_0 + r_0, \\ b = r_0q_1 + r_1, \\ r_0 = r_1q_2 + r_2, \\ \dots \\ r_{k-2} = r_{k-1}q_k + r_k, \\ r_{k-1} = r_kq_{k+1}. \end{array} \right. \quad (1.19)$$

Замечание 1.11. Очевидно, что при некотором k получим $r_{k+1} = 0$, так как

$$r_0 > r_1 > r_2 > \dots > r_k > r_{k+1} = 0.$$

При этом, в силу (1.18):

$$D(a, b) = D(b, r_0) = D(r_0, r_1) = \dots = D(r_{k-2}, r_{k-1}) = D(r_{k-1}, r_k) = r_k.$$

Схему (1.19) легко записать в виде алгоритма.

Алгоритм 1.13 (классический алгоритм Евклида).

```

while  $b \neq 0$  do                                 $\triangleright$  Если  $b = 0$ , то  $D(a, b) = a$ 
     $a \leftarrow a \bmod b$                        $\triangleright a$  – остаток от деления  $a$  на  $b$ 
    swap( $a, b$ )                                $\triangleright$  меняем местами  $a$  и  $b$ 
end while

```

Пример 1.9 (классический алгоритм Евклида). Найдем наибольший общий делитель 4704 и 2808. Применим алгоритм 1.13. Протокол работы алгоритма запишем в виде таблицы (табл. 1.8).

Таблица 1.8

Итерация	a	b	q	r
1	4704	2808	1	1896
2	2808	1896	1	912
3	1896	912	2	72
4	912	72	12	48
5	72	48	1	24
6	48	24	2	0
7	24	0		

Таким образом $D(4704, 2808) = 24$.

Историческая справка. Древнегреческие математики называли этот алгоритм – „взаимное вычитание“. Этот алгоритм не был открыт Евклидом, так как упоминание о нем имеется уже у Аристотеля. Самим Евклидом алгоритм был описан в геометрическом виде и реализуется с помощью циркуля и линейки.

Пусть даны два отрезка длины a и b . Вычтем из большего отрезка меньший и заменим больший отрезок полученной разностью. Повторяя эту операцию, пока отрезки не станут равны. Если это произойдёт, то исходные отрезки соизмеримы, и последний полученный отрезок есть их наибольшая общая мера. Если общей меры нет, то процесс бесконечен.

1.2.2. Бинарный алгоритм Евклида

Классический алгоритм Евклида не всегда является наилучшим способом для нахождения наибольшего общего делителя. Сравнительно недавно¹ для решения этой задачи был предложен совсем иной алгоритм, не требующий операции деления. Основанный исключительно на операциях вычитания, он проверяет, является ли число четным или нет, и сдвигает

¹Обсуждаемый алгоритм был известен еще в Китае I-го века, но опубликован лишь в 1967 году израильским физиком и программистом Джозефом Стайном. Ввиду этого встречается альтернативное название метода – *алгоритм Стайна*.

вправо двоичное представление четного числа (деление пополам). Бинарный алгоритм нахождения $D(a, b)$ основан на нескольких простых свойствах натуральных чисел.

Лемма 1.3. Пусть $a, b \in N$. Тогда справедливы следующие соотношения.

- 1) a и b — четны, тогда $D(a, b) = 2D(a/2, b/2)$;
- 2) a четно и b — нечетно, тогда $D(a, b) = D(a/2, b)$;
- 3) $D(a, b) = D(a - b, b)$
- 4) Если a и b — нечетны, то $a - b$ четно;
- 5) $|a - b| < \max\{a, b\}$.

Замечание 1.12. Свойство 3 использовалось в схеме (1.19)[см. (1.18)].

Алгоритм 1.14 (бинарный алгоритм Евклида).

```

 $k \leftarrow 0$                                 ▷ счетчик для степени двоек, выделяемых по свойству 1 леммы 1.3
while  $a$  и  $b$  — четны do                  ▷ «выдавливаем» из  $a, b$  общую степень двойки
     $k \leftarrow k + 1$ ;  $a \leftarrow a/2$ ;  $b \leftarrow b/2$ 
end while
while  $a \neq b$  do                          ▷ если  $a = b$ , то  $D(a, b) = a \cdot 2^k$ 
    while  $a$  — четно do
         $a \leftarrow a/2$                             ▷ свойство 2 леммы 1.3
    end while
    while  $b$  — четно do
         $b \leftarrow b/2$                             ▷ свойство 2 леммы 1.3
    end while
     $\max(a, b) \leftarrow |a - b|/2$             ▷ свойства 3,5 леммы 1.3
end while
 $D(a, b) = a \cdot 2^k$ 

```

Пример 1.10 (бинарный алгоритм Евклида). Рассмотрим работу алгоритма с данными из примера 1.9. Протокол его работы представлен в табл. 1.9.

Таблица 1.9

k	a	b
$0 \rightarrow 1 \rightarrow 2 \rightarrow 3$	$4704 \rightarrow 2352 \rightarrow 1176 \rightarrow 588$	$2808 \rightarrow 1404 \rightarrow 702 \rightarrow 351$
3	$588 \rightarrow 294 \rightarrow 147$	351
3	147	$204 \rightarrow 102 \rightarrow 51$
3	$96 \rightarrow 48 \rightarrow 24 \rightarrow 12 \rightarrow 6 \rightarrow 3$	51
3	3	$48 \rightarrow 24 \rightarrow 12 \rightarrow 6 \rightarrow 3$

Имеем: $D(4704, 2808) = 2^3 \times 3 = 24$.

Замечание 1.13. Для решения задачи бинарным алгоритмом потребовалось 3 вычитания и 16 делений на 2 (сдвигов мантиссы на 1 разряд вправо). Для решения этой же задачи классическим алгоритмом Евклида требуется 6 делений с остатком .

При использовании классического алгоритма алгоритма Евклида получается дополнительная информация (последовательность неполных частных q_0, q_1, \dots), которая будет необходима для решения других задач.

1.2.3. Линейное представление наибольшего общего делителя

Действуя в схеме (1.19) *обратным ходом* снизу вверх получаем представление $D(a, b)$ через a, b :

$$\begin{aligned} r_k &= r_{k-2} - r_{k-1}q_k = r_{k-2} - (r_{k-3} - r_{k-2}q_{k-1})q_k = \\ &= -r_{k-3}q_k + r_{k-2}(1 + q_kq_{k-1}) = \\ &= -r_{k-3}q_k + (r_{k-4} - r_{k-3}q_{k-2})(1 + q_kq_{k-1}) = \\ &= r_{k-4}(1 + q_kq_{k-1}) + r_{k-3}(-q_k - q_{k-2}(1 + q_kq_{k-1})) = \dots = ax + by. \end{aligned}$$

Таким образом, справедливо следующее утверждение.

Утверждение 1.1. Для любых $a, b \in \mathbb{Z}$ уравнение $ax + by = D(a, b)$ всегда имеет решение.

Нетрудно получить рекуррентные формулы для вычисления x, y . Введем последовательность $\{y_i\}$ следующим образом:

$$y_0 = 0, \quad y_1 = 1, \quad y_{i+1} = y_{i-1} - q_{k+1-i}y_i, \quad i = 1, 2, \dots, k+1. \quad (1.20)$$

Тогда $D(a, b) = ay_{k+1} + by_{k+2}$.

Пример 1.11. Найдем линейное представление для наибольшего общего делителя из примера 1.9 (табл. 1.10).

Таблица 1.10

i	0	1	2	3	4	5	6
y_i	0	1	-1	13	-27	40	-67
q_{6-i}			1	12	2	1	1

Таким образом, $4704 \times 40 - 2808 \times 67 = 24$.

1.2.4. Обобщенный алгоритм Евклида

Из результатов предыдущего подраздела, следует, что справедливо представление

$$d = D(a, b) = ax + by. \quad (1.21)$$

Представление наибольшего общего делителя в виде (1.21) называется *соотношением Безу*¹, а числа x и y – *коэффициентами Безу*.

Равенство (1.21) можно записать в векторной форме, т. е. наибольший общий делитель можно представить вектором (упорядоченным) набором коэффициентов разложения по исходным числам $d = (x, y)$. Легко видеть, что сами числа представляются такими векторами очень просто:

$$\begin{aligned} a &= (1, 0), \text{ так как } a = 1 \times a + 0 \times b, \\ b &= (0, 1), \text{ так как } b = 0 \times a + 1 \times b. \end{aligned}$$

Для нахождения вектора (x, y) применим схему (1.19) классического алгоритма Евклида, но параллельно будем выполнять соответствующие действия с векторами разложений. Напомним, что арифметические действия с упорядоченными наборами чисел, то же что с векторами в координатах, выполняются покомпонентно. Таким образом, одновременно решаются две задачи: вычисление наибольшего общего делителя и нахождение его линейного представления.

Обозначим вектора разложения для чисел a и b как (x_a, y_a) и (x_b, y_b) . Покажем работу алгоритма на примере.

Пример 1.12. Найдем $D(64, 81)$ и его линейное представление. Имеем, $a = 64$, $b = 81$. Тогда $(x_a, y_a) = (1, 0)$, а $(x_b, y_b) = (0, 1)$. Работа алгоритма показана в табл. 1.11.

Таблица 1.11

a	(x_a, y_a)	b	(x_b, y_b)	Комментарий
64	(1, 0)	81	(0, 1)	
64	(1, 0)	17	(-1, 1)	$b = b - a$, $(x_b, y_b) = (x_b, y_b) - (x_a, y_a)$
13	(4, -3)	17	(-1, 1)	$a = a - 3b$, $(x_a, y_a) = (x_a, y_a) - 3(x_b, y_b)$
13	(4, -3)	4	(-5, 4)	$b = b - a$, $(x_b, y_b) = (x_b, y_b) - (x_a, y_a)$
1	(19, -15)	4	(-5, 4)	$a = a - 3b$, $(x_a, y_a) = (x_a, y_a) - 3(x_b, y_b)$

Так как в алгоритме Евклида получена единица (следующий шаг последний — второе число будет равно 0), приходим к искомому разложению, где $x = 19$, $y = -15$:

$$64 \times 19 + 81 \times (-15) = 1 = D(64, 81).$$

Сформулируем алгоритм.

¹ Впервые данный факт опубликовал в 1624 году французский математик Клод Гаспар де Мезириак для случая взаимно простых чисел a и b . Этьен Безу в конце XVIII века обобщил результат, распространив его и на кольцо многочленов.

Алгоритм 1.15 (обобщенный алгоритм Евклида).

```
( $x_a, y_a$ )  $\leftarrow (1, 0)$ , ( $x_b, y_b$ )  $\leftarrow (0, 1)$ 
while  $b \neq 0$  do
     $t \leftarrow a$ ,  $T \leftarrow (x_a, y_a)$ 
     $a \leftarrow b$ , ( $x_a, y_a$ )  $\leftarrow (x_b, y_b)$ 
     $q \leftarrow t \div b$ 
     $b \leftarrow t - qb$ , ( $x_b, y_b$ )  $\leftarrow T - q(x_b, y_b)$ 
end while
```

Наибольший общий делитель будет находиться в a , а коэффициенты разложения в (x_a, y_a) .

Замечание 1.14. Если $a < b$, то после первого шага алгоритма a и b поменяются местами. Если оба числа равны нулю, алгоритм не применим. Переменные t и T используются для временного хранения a и вектора разложения (x_a, y_a) соответственно.

Пример 1.13. Вернемся к примеру 1.12. Протокол работы расширенного алгоритма Евклида удобно записать в виде, показанном в табл. 1.12.

Таблица 1.12

Остатки	64	81	64	17	13	4	1	0
Частные q			0	1	3	1	3	4
x	1	0	1	-1	4	-5	19	-81
y	0	1	0	1	-3	4	-15	64

Пример получения новых значений первых компонент наборов (x_a, y_a) и (x_b, y_b) показан в третьей строке таблицы (клетки выделены серым цветом): из числа, расположенного в первой клетке вычитают число во второй клетке, помноженное на число, стоящее справа от него во второй строке (число в рамке), результат записывают в третью клетку. Аналогично выполняют операции для нахождения вторых компонент (четвертая строка, клетки выделены серым цветом). Легко проследить аналогию с рекуррентными формулами (1.20) предыдущего подраздела.

Заметим, что для каждого столбика выполняется равенство

$$ax + by = \text{остаток}.$$

Пример 1.14. Запишем протокол работы расширенного алгоритма Евклида для примера 1.9 (см. табл. 1.13).

Таким образом

$$D(4704, 2808) = 24 = 4704 \times 40 + 2808 \times (-67).$$

Таблица 1.13

	4704	2808	1896	912	72	48	24	0
q			1	1	2	12	1	2
x	1	0	1	-1	3	-37	40	117
y	0	1	-1	2	-5	62	-67	196

1.2.5. Разложение числа в цепную дробь

Запишем схему (1.19) в виде

$$\left\{ \begin{array}{l} \frac{a}{b} = q_0 + \frac{r_0}{b}, \\ \frac{b}{r_0} = q_1 + \frac{r_1}{r_0}, \\ \dots \dots \dots \\ \frac{r_{k-2}}{r_{k-1}} = q_k + \frac{r_k}{r_{k-1}}, \\ \frac{r_{k-1}}{r_k} = q_{k+1}. \end{array} \right. \quad (1.22)$$

Из (1.22) получаем

$$\frac{a}{b} = q_0 + \frac{1}{\frac{b}{r_0}} = q_0 + \frac{1}{q_1 + \frac{1}{\frac{r_0}{r_1}}} = \dots = q_0 + \frac{1}{q_1 + \frac{1}{q_2 + \frac{1}{\dots + \frac{1}{q_{k+1}}}}}. \quad (1.23)$$

Определение 1.8. Выражение (1.23) называют *цепной, или непрерывной, дробью*. Дроби, возникающие при разложении вида (1.23), называют *подходящими дробями*. Последняя подходящая дробь равна a/b . Числа $q_0, q_1, q_2, \dots, q_{k+1}$ называют *звенями дроби*. Более коротко (1.23) записывают в виде

$$\frac{a}{b} = [q_0; q_1, q_2, \dots, q_{k+1}]. \quad (1.24)$$

Замечание 1.15. В представлении (1.24) q_0 есть целая часть дроби a/b . Легко показать следующее:

- 1) если $0 < a < b$, т. е. a/b — правильная положительная дробь, то представление (1.24) для нее имеет вид $[0; q_1, q_2, \dots, q_{k+1}]$;
- 2) если дробь $a/b < 0$, то $a/b = [q_0; q_1, q_2, \dots, q_{k+1}]$, где $q_0 \leq -1$ есть целая часть дроби a/b ;

3) любое $a \in \mathbb{Z}$ является цепной дробью с одним звеном $[a]$.

Пример 1.15. Пусть $a = -48$, $b = 109$. Тогда

$$\frac{a}{b} = -\frac{48}{109} = -1 + \frac{61}{109} = -1 + \frac{1}{\frac{109}{61}}.$$

Для $a = 109 > b = 61 > 0$ применим алгоритм 1.13 (см. табл. 1.14).

Таблица 1.14

Итерация	a	b	q	r	Итерация	a	b	q	r
1	109	61	1	48	5	9	4	2	1
2	61	48	1	13	6	4	1	4	0
3	48	13	3	9	7	1	0		
4	13	9	1	4					

Тогда окончательно

$$-\frac{48}{109} = [-1; 1, 1, 3, 1, 2, 4].$$

Таким образом, справедлива следующая теорема.

Теорема 1.15. Для любых $a, b \in \mathbb{Z}, b \neq 0$ дробь a/b можно одним и только одним способом представить в виде (1.24), где все звенья q_i начиная с q_1 положительны и последнее звено $q_{k+1} > 1$.

Замечание 1.16. Условие $q_{k+1} > 1$ в теореме обеспечивает единственность разложения (1.24), так как

$$[q_0; q_1, \dots, q_{k+1}] = [q_0; q_1, \dots, q_{k+1} - 1, 1].$$

Действительно, последнее уравнение в (1.19) $r_{k-1} = r_k q_{k+1}$ можно записать в виде

$$\begin{cases} r_{k-1} = r_k(q_{k+1} - 1) + r_k, \\ r_k = r_k \cdot 1. \end{cases}$$

Пример 1.16. Рассмотрим дробь из примера 1.9. Используя результаты примера 1.9 получаем представление (1.24):

$$\frac{4704}{2808} = [1; 1, 2, 12, 1, 2].$$

Заметим, что если полученную цепную дробь снова привести к обычной, то получится *несократимая* дробь: $\frac{4704}{2808} = \frac{196}{117}$.

Историческая справка. Цепные дроби были введены в 1572 г. итальянским математиком Бомбелли. Современное обозначение непрерывных дробей впервые встречается у итальянского математика Катальди в 1613 г. Первый изложил теорию цепных дробей Л. Эйлер¹.

1.2.6. Свойства подходящих дробей и их вычисление

Введем обозначения для подходящих дробей, определенных в предыдущем подразделе при разложении (1.23):

$$\delta_0 = q_0, \quad \delta_1 = q_0 + \frac{1}{q_1}, \quad \delta_2 = q_0 + \frac{1}{q_1 + \frac{1}{q_2}} \quad \dots . \quad (1.25)$$

Обозначим числитель и знаменатель δ_k через P_k и Q_k соответственно. Получим рекуррентные формулы для определения последовательностей числителей $\{P_k\}$ и знаменателей $\{Q_k\}$. Положив $P_{-1} = 1, Q_{-1} = 0$, имеем

$$\begin{aligned} \delta_0 &= \frac{q_0}{1} = \frac{P_0}{Q_0} \text{ (полагаем } P_0 = q_0, Q_0 = 1), \\ \delta_1 &= \frac{q_0 + \frac{1}{q_1}}{1} = \frac{q_0 q_1 + 1}{q_1 + 0} = \frac{q_1 P_0 + P_{-1}}{q_1 Q_0 + Q_{-1}} = \frac{P_1}{Q_1}, \quad \delta_2 = \frac{q_2 P_1 + P_0}{q_2 Q_1 + Q_0} = \frac{P_2}{Q_2}, \dots , \end{aligned}$$

При анализе числителей и знаменателей последовательности $\delta_0, \delta_1, \dots$, становятся очевидными рекуррентные соотношения для P_k и Q_k :

$$\begin{cases} P_s = q_s P_{s-1} + P_{s-2}, \\ Q_s = q_s Q_{s-1} + Q_{s-2}. \end{cases} \quad (1.26)$$

Пример 1.17. Используя результаты примера 1.9, получим последовательность подходящих дробей (табл. 1.15).

Таблица 1.15

s	-1	0	1	2	3	4	5
q_s		1	1	2	12	1	2
P_s	1	1	2	5	62	67	196
Q_s	0	1	1	3	37	40	117

¹См. подраздел 1.3.3.

Таким образом,

$$\delta_0 = 1, \quad \delta_1 = 2, \quad \delta_2 = \frac{5}{3}, \quad \delta_3 = \frac{62}{37}, \quad \delta_4 = \frac{67}{40}, \quad \delta_5 = \frac{196}{11} = \frac{4704}{2808}.$$

Установим важное свойство подходящих дробей. Рассмотрим разность двух соседних подходящих дробей. Имеем

$$\begin{aligned} \delta_s - \delta_{s-1} &= \frac{P_s}{Q_s} - \frac{P_{s-1}}{Q_{s-1}} = \frac{P_s Q_{s-1} - P_{s-1} Q_s}{Q_s Q_{s-1}} = \frac{h_s}{Q_s Q_{s-1}}, \text{ где} \\ h_s &= P_s Q_{s-1} - P_{s-1} Q_s = (q_s P_{s-1} + P_{s-2}) Q_{s-1} - (q_s Q_{s-1} + Q_{s-2}) P_{s-1} = \\ &= -(P_{s-1} Q_{s-2} - P_{s-2} Q_{s-1}) = -h_{s-1}. \end{aligned}$$

Таким образом, $h_s = -h_{s-1}$, но $h_0 = P_0 Q_{-1} - P_{-1} Q_0 = -1$, следовательно, $h_s = (-1)^{s+1}$ и доказана следующая лемма.

Лемма 1.4. Справедливо равенство

$$\frac{P_s}{Q_s} - \frac{P_{s-1}}{Q_{s-1}} = \frac{(-1)^{s+1}}{Q_s Q_{s-1}}. \quad (1.27)$$

Следствие 1.3. Подходящие дроби несократимы.

Доказательство. Так как $P_s Q_{s-1} - P_{s-1} Q_s = (-1)^{s+1}$, то любой общий делитель P_s, Q_s равен 1. Таким образом, $D(P_s, Q_s) = 1$. ■

Следствие 1.4.

$$\lim_{s \rightarrow \infty} \left| \frac{P_s}{Q_s} - \frac{P_{s-1}}{Q_{s-1}} \right| = 0. \quad (1.28)$$

Доказательство. Согласно теореме 1.15, все звенья цепной дроби q_i положительны (кроме, быть может, q_0). Тогда из (1.26) следует, что все Q_s (при $s \geq 0$) также положительны и последовательность натуральных чисел Q_s возрастает. Из (1.27) получаем требуемое. ■

1.2.7. Цепные дроби и анализ алгоритма Евклида

Введенные понятия цепной дроби и подходящих дробей оказываются очень полезными для анализа работы алгоритма Евклида. Дадим необходимые обозначения. Рассмотрим трехдиагональный определитель:

$$\begin{vmatrix} q_0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & q_1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & q_2 & 1 & \cdots & 0 & 0 \\ \dots & & & & & & \\ 0 & 0 & 0 & 0 & \cdots & q_{n-2} & 1 \\ 0 & 0 & 0 & 0 & \cdots & -1 & q_{n-1} \end{vmatrix} = K_n(q_0, q_1, \dots, q_{n-1}). \quad (1.29)$$

Определение 1.9. Определитель (1.29) называют *континуантой* n -го порядка или индекса.

Разложим континуанту n -го порядка по последнему столбцу:

$$K_n(q_0, q_1, \dots, q_{n-1}) = q_{n-1} K_{n-1}(q_0, q_1, \dots, q_{n-2}) + K_{n-2}(q_0, q_1, \dots, q_{n-3}). \quad (1.30)$$

Соотношение (1.30) очень напоминает рекуррентные соотношения (1.26) для числителей и знаменателей подходящих дробей. Это не случайно и две следующие леммы подтверждают предположение о связи континуант и цепных дробей.

Лемма 1.5. Континуанта $K_n(q_0, q_1, \dots, q_{n-1})$ равна сумме всевозможных произведений элементов q_0, q_1, \dots, q_{n-1} , одно из которых содержит все эти элементы, а другие получаются из него выбрасыванием одной или нескольких пар сомножителей с соседними номерами (если выброшены все сомножители, то считаем, что осталась 1).

Доказательство. Индукция по n . База индукции:

$$K_1(q_0) = q_0, \quad K_2(q_0, q_1) = \begin{vmatrix} q_0 & 1 \\ -1 & q_1 \end{vmatrix} = q_0 q_1 + 1,$$

и утверждение леммы справедливо для континуант первого и второго порядков.

Шаг индукции. Пусть утверждение леммы справедливо для континуант $(n-2)$ -го и $(n-1)$ -го порядков. Применив разложение (1.30), получим требуемое. ■

Пример 1.18.

$$\begin{aligned} K_6(q_0, q_1, q_2, q_3, q_4, q_5) = & q_0 q_1 q_2 q_3 q_4 q_5 + q_2 q_3 q_4 q_5 + q_0 q_3 q_4 q_5 + q_0 q_1 q_4 q_5 + \\ & + q_0 q_1 q_2 q_5 + q_0 q_1 q_2 q_3 + q_4 q_5 + q_2 q_5 + q_0 q_5 + \\ & + q_2 q_3 + q_0 q_3 + q_0 q_1 + 1. \end{aligned}$$

Замечание 1.17. Количество слагаемых в континуанте n -го порядка¹ есть сумма числа слагаемых в континуантах $(n-1)$ -го и $(n-2)$ -го порядков, т. е. континуанта $(q_0 q_1 \dots q_{n-1})$ содержит ϕ_{n+1} слагаемых, где ϕ_{n+1} – $(n+1)$ -е число Фибоначчи (см. раздел 1.2.12).

Явная связь континуант и цепных дробей впервые была установлена Эйлером.

Лемма 1.6 (лемма Эйлера). Справедливо тождество

$$[q_0; q_1, \dots, q_{n-1}] = \frac{K_n(q_0, q_1, \dots, q_{n-1})}{K_{n-1}(q_1, q_2, \dots, q_{n-1})}. \quad (1.31)$$

¹Считаем, что континуанта нулевого порядка K_0 содержит 0 слагаемых.

Доказательство. Индукция по n . База индукции:

$$[q_0; q_1] = q_0 + \frac{1}{q_1} = \frac{q_0 q_1 + 1}{q_1} = \frac{K_2(q_0, q_1)}{K_1(q_1)}.$$

Шаг индукции. Пусть тождество (1.31) верно для дробей с $n-1$ звеном включительно. Представим n -звенную дробь $(q_0, q_1, \dots, q_{n-1})$ дробью с $n-1$ звеном, где последнее звено имеет вид $q_{n-2} + \frac{1}{q_{n-1}}$. Применив к этой дроби индукционное предположение, с учетом разложения (1.30), имеем

$$\begin{aligned} [q_0; q_1, \dots, q_{n-1}] &= \left[q_0; q_1, \dots, q_{n-2} + \frac{1}{q_{n-1}} \right] = \\ &= \frac{K_{n-1} \left(q_0, q_1, \dots, \left(q_{n-2} + \frac{1}{q_{n-1}} \right) \right)}{K_{n-2} \left(q_1, q_2, \dots, \left(q_{n-2} + \frac{1}{q_{n-1}} \right) \right)} = \\ &= \frac{\left(q_{n-2} + \frac{1}{q_{n-1}} \right) K_{n-2}(q_0, q_1, \dots, q_{n-3}) + K_{n-3}(q_0, q_1, \dots, q_{n-4})}{\left(q_{n-2} + \frac{1}{q_{n-1}} \right) K_{n-3}(q_1, q_2, \dots, q_{n-3}) + K_{n-4}(q_1, q_2, \dots, q_{n-4})} = \\ &= \frac{K_{n-1}(q_0, q_1, \dots, q_{n-2}) + \frac{K_{n-2}(q_0, q_1, \dots, q_{n-3})}{q_{n-1}}}{K_{n-2}(q_1, q_2, \dots, q_{n-2}) + \frac{K_{n-3}(q_1, q_2, \dots, q_{n-3})}{q_{n-1}}} = \frac{K_n(q_0, q_1, \dots, q_{n-1})}{K_{n-1}(q_1, q_2, \dots, q_{n-1})}. \end{aligned}$$

■

Замечание 1.18. Из (1.31) становится очевидно, что

$$\frac{K_n(q_0 q_1 \dots q_{n-1})}{K_{n-1}(q_1 q_2 \dots q_{n-1})} = \frac{P_{n-1}}{Q_{n-1}} = \delta_{n-1}.$$

Таким образом, введенное понятие континуанты это другой взгляд на подходящие дроби, дающий возможность связать их с последовательностью Фибоначчи (см. лемму 1.5).

Перейдем к анализу алгоритма Евклида. Нас будет интересовать наихудший случай — когда алгоритм Евклида работает особенно долго. Сформулируем вопрос точнее: для каких двух наименьших чисел надо применить алгоритм Евклида, чтобы он работал в точности заданное число шагов? Ответ на этот вопрос дает следующая теорема.

Теорема 1.16 (теорема Ламе¹, 1845 г.). Пусть n — произвольное натуральное число, и $a > b > 0$ такие, что алгоритму Евклида для обработки a и b необходимо выполнить точно n шагов (делений с остатком), причем a — наименьшее натуральное число с таким свойством. Тогда

$$a = \phi_{n+2}, \quad b = \phi_{n+1},$$

где ϕ_k — k -е число Фибоначчи.

Доказательство. Разложим a/b в цепную дробь. Согласно лемме 1.6 получаем,

$$\frac{a}{b} = [q_0; q_1, \dots, q_{n-1}] = \frac{K_n(q_0, q_1, \dots, q_{n-1})}{K_{n-1}(q_1, q_2, \dots, q_{n-1})},$$

где q_0, q_1, \dots, q_{n-1} — неполные частные из алгоритма Евклида. По условию теоремы, их ровно n . Принимая во внимание несократимость подходящих дробей (см. следствие 1.3 к лемме 1.4 и замечание 1.18) становится очевидно, что континуанты $(q_0 q_1 \dots q_{n-1})$ и $(q_1 q_2 \dots q_{n-1})$ взаимно просты. Пусть $D(a, b) = d$. Тогда

$$\begin{cases} a = K_n(q_0, q_1, \dots, q_{n-1}) d, \\ b = K_{n-1}(q_1, q_2, \dots, q_{n-1}) d. \end{cases} \quad (1.32)$$

В силу единственности разложения в цепную дробь (см. теорему 1.15), в случае $a > b > 0$ справедливы неравенства $q_0, q_1, \dots, q_{n-2} \geq 1$, $q_{n-1} \geq 2$. Очевидно, что $d \geq 1$. По лемме 1.5, континуанта есть многочлен с неотрицательными коэффициентами от переменных q_0, q_1, \dots . Его минимальное значение очевидно достигается при $q_0 = q_1 = \dots = q_{n-2} = 1, q_{n-1} = 2$. Положив $d = 1$ и подставив эти значения q_i в (1.32), получим требуемое. ■

Следствие 1.5. Если натуральные числа a и b не превосходят N , то число шагов (операций деления с остатком), необходимых алгоритму Евклида для обработки a и b не превышает¹

$$\left\lceil \log_\alpha \left(\sqrt{5} N \right) \right\rceil - 2, \quad \text{где } \alpha = \frac{1 + \sqrt{5}}{2}.$$

Коэффициент $\alpha = 1.6180337\dots$ известен в геометрии как «золотое сечение».

Замечание 1.19. Легко подсчитать, что

$$\log_\alpha \left(\sqrt{5} N \right) \cong 4.785 \cdot \lg N + 1.672,$$

¹Г. Ламе (1795–1870) — французский математик, физик и инженер

¹Доказательство следствия 1.5 очевидно из примера 1.133, (см. подразд. 1.7.1)

поэтому, например, для вычисления НОД пары чисел, меньших миллиона, алгоритму Евклида потребуется не более, чем

$$4.785 \cdot 6 + 1.672 - 2 = 31 - 2 = 29$$

шагов.

Из теоремы Ламе следует, что наихудший случай для алгоритма Евклида — два последовательных числа Фибоначчи. Рекуррентное соотношение для чисел Фибоначчи (1.46) по сути является строкой алгоритма Евклида при делении ϕ_{k+2} на ϕ_{k+1} с остатком.

В заключение подраздела, используя введенную последовательность Фибоначчи (1.46), получим еще одно свойство подходящих дробей.

Теорема 1.17. Знаменатели подходящих дробей растут не медленнее последовательности Фибоначчи:

$$Q_s \geq \phi_s, \text{ при } s \geq 1,$$

где ϕ_s — s -е число Фибоначчи.

Доказательство. Наиболее медленный рост знаменателей подходящих дробей будет наблюдаться при

$$Q_s = Q_{s-1} + Q_{s-2}, \text{ т. е. при } q_1 = q_2 = \dots = q_s = 1.$$

Последнее рекуррентное соотношение вместе с начальными условиями $Q_0 = 0, Q_1 = 1$ задает последовательность Фибоначчи. ■

1.2.8. Бесконечная цепная дробь и ее вычисление

Попробуем представить в виде цепной дроби произвольное вещественное число. Заметим, что для рациональной дроби этот процесс совпадает с (1.19). Пусть $\alpha \in \mathbb{R}$. Имеем

$$\begin{aligned} \alpha &= q_0 + \eta_1 = q_0 + \frac{1}{\alpha_1}, \quad \alpha_1 > 1, \quad q_0 = [\alpha], \\ \alpha_1 &= q_1 + \eta_2 = q_1 + \frac{1}{\alpha_2}, \quad \alpha_2 > 1, \quad q_1 = [\alpha_1], \\ &\dots \\ \alpha_{s-2} &= q_{s-2} + \eta_{s-1} = q_{s-2} + \frac{1}{\alpha_{s-1}}, \quad \alpha_{s-1} > 1, \quad q_{s-2} = [\alpha_{s-2}], \quad (1.33) \\ \alpha_{s-1} &= q_{s-1} + \eta_s. \end{aligned}$$

Из (1.33) получаем следующее разложение α в непрерывную дробь:

$$\alpha = q_0 + \cfrac{1}{q_1 + \cfrac{1}{q_2 + \cfrac{1}{\dots + \cfrac{1}{q_{s-1} + \eta_s}}}}. \quad (1.34)$$

Если $\alpha \in \mathbb{Q}$, то на неком шаге будет получено $\eta_s = 0$ и процесс (1.33) завершится разложением α в конечную цепную дробь согласно (1.19). Если же α не является рациональной дробью, то процесс (1.33) можно продолжать бесконечно. Зададимся вопросами:

1. Что понимать под такой бесконечной дробью?
2. Как ее приближенно вычислять?

Прежде, чем ответить на поставленные вопросы, рассмотрим представление (1.33) на конкретном примере.

Пример 1.19. Пусть $\alpha = \sqrt{28}$. Имеем

$$\begin{aligned}\alpha &= 5 + \frac{1}{\alpha_1}, \\ \alpha_1 &= \frac{1}{\sqrt{28} - 5} = \frac{\sqrt{28} + 5}{3} = 3 + \frac{1}{\alpha_2}, \\ \alpha_2 &= \frac{3}{\sqrt{28} - 4} = \frac{\sqrt{28} + 4}{4} = 2 + \frac{1}{\alpha_3}, \\ \alpha_3 &= \frac{4}{\sqrt{28} - 4} = \frac{\sqrt{28} + 4}{3} = 3 + \frac{1}{\alpha_4}, \\ \alpha_4 &= \frac{3}{\sqrt{28} - 5} = \sqrt{28} + 5 = 10 + \frac{1}{\alpha_5} \\ \alpha_5 &= \frac{1}{\sqrt{28} - 5} = \frac{\sqrt{28} + 5}{3} = 3 + \frac{1}{\alpha_6}.\end{aligned}$$

Заметим, что $\alpha_5 = \alpha_1$ и происходит возврат к первому уравнению. Процесс «зацикливается». Таким образом, получаем представление $\sqrt{28}$ в виде бесконечной периодической последовательности. Строгое обоснование этому представлению будет дано далее.

Установим некоторые свойства разложения (1.33).

Теорема 1.18. Точное значение числа $\alpha \in \mathbb{R}$ всегда находится между соседними подходящими дробями, причем оно ближе к последующей, чем к предыдущей дроби. При этом подходящие дроби с четными номерами находятся слева, а с нечетными — справа от α .

Доказательство. Считая в разложении (1.34) $\alpha_s = \frac{1}{\eta_s}$ — последним знаменателем (аналогом q_s) запишем это разложение в виде

$$\alpha = [q_0; q_1, \dots, q_{s-1}, \alpha_s]. \quad (1.35)$$

Тогда α можно считать s -й подходящей дробью. Применив (1.26), получим

$$\alpha = \frac{P_{s-1}\alpha_s + P_{s-2}}{Q_{s-1}\alpha_s + Q_{s-2}}. \quad (1.36)$$

Из (1.36) имеем

$$\begin{aligned} \alpha Q_{s-1}\alpha_s + \alpha Q_{s-2} &= P_{s-1}\alpha_s + P_{s-2}, \\ \alpha_s(\alpha Q_{s-1} - P_{s-1}) &= P_{s-2} - \alpha Q_{s-2}, \\ \alpha_s Q_{s-1} \left(\alpha - \frac{P_{s-1}}{Q_{s-1}} \right) &= Q_{s-2} \left(\frac{P_{s-2}}{Q_{s-2}} - \alpha \right), \end{aligned} \quad (1.37)$$

но $\alpha_s > 1$, $Q_{s-1} > Q_{s-2} > 0$, т. е. $\alpha_s Q_{s-1} > Q_{s-2}$, тогда из (1.37)

$$\left| \alpha - \frac{P_{s-1}}{Q_{s-1}} \right| < \left| \frac{P_{s-2}}{Q_{s-2}} - \alpha \right|.$$

При этом знаки у $\alpha - \delta_{s-1}$ и $\delta_{s-2} - \alpha$ одинаковы. Замечание, что

$$\delta_0 = q_0 < \alpha,$$

т. е. лежит слева от α , завершает доказательство. ■

Учитывая теорему 1.18 и свойство (1.27), получаем следующее следствие (рис. 1.1).

Следствие 1.6.

$$|\alpha - \delta_s| \leq |\delta_{s+1} - \delta_s| = \frac{1}{Q_{s+1}Q_s}. \quad (1.38)$$

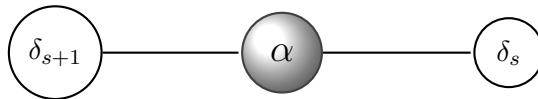


Рис. 1.1

Замечание 1.20. Равенство в (1.38) достигается только когда α — рациональная дробь, а δ_{s+1} — последняя подходящая дробь.

Перейдем к формальному определению бесконечной цепной дроби. Разобьем последовательность подходящих дробей на две подпоследовательности с четными и нечетными номерами. Как уже отмечалось, последовательность знаменателей $\{Q_k\}$ возрастает. Применив (1.27) и теорему 1.18, имеем:

$$\begin{aligned} \frac{P_{2m-1}}{Q_{2m-1}} - \frac{P_{2m}}{Q_{2m}} &= \frac{1}{Q_{2m-1}Q_{2m}} > \frac{P_{2m+1}}{Q_{2m+1}} - \frac{P_{2m}}{Q_{2m}} = \frac{1}{Q_{2m+1}Q_{2m}} > \\ &> \frac{P_{2m+1}}{Q_{2m+1}} - \frac{P_{2m+2}}{Q_{2m+2}} = \frac{1}{Q_{2m+1}Q_{2m+2}} > 0. \end{aligned}$$

Из последней цепочки равенств получаем:

$$\frac{P_{2m-1}}{Q_{2m-1}} > \frac{P_{2m+1}}{Q_{2m+1}}, \quad \frac{P_{2m}}{Q_{2m}} < \frac{P_{2m+2}}{Q_{2m+2}}.$$

Таким образом, подпоследовательность $\left\{ \frac{P_{2m+1}}{Q_{2m+1}} \right\}$ убывает и по (1.27) ограничена снизу (например, δ_2), следовательно, имеет предел. Аналогично подпоследовательность $\left\{ \frac{P_{2m}}{Q_{2m}} \right\}$ возрастает и ограничена сверху (например, δ_1), следовательно, также имеет предел. В силу (1.28) эти пределы совпадают.

Определение 1.10. Бесконечной цепной дробью будем называть предел последовательности подходящих дробей.

Перейдем к вопросу вычисления бесконечной цепной дроби.

Утверждение 1.2. Бесконечную цепную дробь можно вычислить с любой степенью точности ε .

Доказательство. Используя неравенство (1.38) и возрастание последовательности $\{Q_k\}$, получим оценку

$$|\alpha - \delta_m| \leq \frac{1}{Q_{m+1}Q_m} < \frac{1}{Q_m^2} < \varepsilon. \quad (1.39)$$

Таким образом для достижения нужной точности вычисляем подходящие дроби $\{\delta_m\}$ до тех пор, пока $Q_m^2 \leq \frac{1}{\varepsilon}$. ■

Замечание 1.21. Оценка $1/Q_m^2$ в (1.39) является менее точной, чем оценка $1/(Q_{m+1}Q_m)$. Но для последней нужно проводить вычисления знаменателей подходящих дробей на один шаг вперед.

Пример 1.20. Вернемся к вычислению $\sqrt{28}$. В примере 1.19 была получена бесконечная цепная дробь с периодом $(3, 2, 3, 10)$.

Определение 1.11. Бесконечную цепную дробь, в которой определенная последовательность неполных частных, начиная с некоторого места, периодически повторяется, называют *периодической цепной дробью*. Период бесконечной цепной дроби будем записывать в круглых скобках следующим образом:

$$\sqrt{28} = [5; 3, 2, 3, 10, 3, 2, 3, 10, \dots, 3, 2, 3, 10, \dots] = [5; (3, 2, 3, 10)].$$

Таблица 1.16

s	-1	0	1	2	3	4	5	6
q_s		5	3	2	3	10	3	2
P_s	1	5	16	37	127	1 307	4 048	9 403
Q_s	0	1	3	7	24	247	765	1 777

Вычислим $\sqrt{28}$ с точностью до $\varepsilon = 10^{-4}$. Используя (1.26), получаем последовательность подходящих дробей (табл. 1.16).

Во время вычисления производим контроль точности по знаменателю.

1) $\frac{P_1}{Q_1} = \frac{16}{3} = 5,3(3)$, точность $\frac{1}{3^2} = 0,1(1)$ — гарантирована одна значащая цифра.

2) $\frac{P_2}{Q_2} = \frac{37}{7} = 5,2857\dots$, точность $\frac{1}{7^2} = 0,02048\dots$ — две значащие цифры гарантированы.

3) $\frac{P_3}{Q_3} = \frac{127}{24} = 5,2916(6)$, точность $\frac{1}{24^2} = 0,0017\dots$ — три значащие цифры гарантированы. (В данном случае реально получаем четыре, так как оценка (1.39) достаточно грубая).

4) $\frac{P_4}{Q_4} = \frac{1307}{247} = 5,291500\dots$, точность $\frac{1}{247^2} = 0,000016\dots$ — получаем пять, а реально шесть значащих цифр.

Достигнута нужная оценка. Останавливаем процесс вычислений.

$$\sqrt{28} \approx 5,291500.$$

Установим важное свойство бесконечной цепной дроби. Дадим необходимые определения.

Определение 1.12. Иррациональное число, являющееся корнем некоторого квадратного уравнения с целыми коэффициентами, называют *квадратичной иррациональностью*.

Пример 1.21. Квадратичными иррациональностями являются:

$$\frac{5 + \sqrt{17}}{4}, \frac{7 + \sqrt{29}}{10}, 3 + \sqrt{7}.$$

Выражения

$$\sqrt[3]{7}, \sqrt[5]{3} + 13, \frac{\sqrt[7]{7} + 13}{8}, \pi$$

квадратичными иррациональностями не являются.

Теорема 1.19 (теорема Лагранжа). Квадратичные иррациональности и только они представимы в виде бесконечной периодической цепной дроби.

Доказательство. Пусть $\alpha = (q_1, q_2, \dots, q_n, \dots)$ — бесконечная периодическая цепная дробь. Воспользуемся доказательством теоремы 1.18. Запишем для α представление (1.35).

Очевидно, что в силу периодичности цепной дроби величины α_s , удовлетворяют соотношению: $\alpha_{s+h} = \alpha_s$, где $s > 1$, а h — период дроби. Используя (1.36) имеем

$$\alpha = \frac{P_{s-1}\alpha_s + P_{s-2}}{Q_{s-1}\alpha_s + Q_{s-2}} = \frac{P_{s+h-1}\alpha_{s+h} + P_{s+h-2}}{Q_{s+h-1}\alpha_{s+h} + Q_{s+h-2}} = \frac{P_{s+h-1}\alpha_s + P_{s+h-2}}{Q_{s+h-1}\alpha_s + Q_{s+h-2}}. \quad (1.40)$$

Из (1.40) получаем

$$\frac{P_{s-1}\alpha_s + P_{s-2}}{Q_{s-1}\alpha_s + Q_{s-2}} = \frac{P_{s+h-1}\alpha_s + P_{s+h-2}}{Q_{s+h-1}\alpha_s + Q_{s+h-2}}$$

квадратное уравнение с целыми коэффициентами для нахождения α_s . Значит, α_s — квадратичная иррациональность, тогда в силу (1.36) α — тоже квадратичная иррациональность.

Докажем достаточность. Пусть α удовлетворяет квадратному уравнению с целыми коэффициентами

$$A\alpha^2 + B\alpha + C = 0. \quad (1.41)$$

Разложим α в цепную дробь и подставим в уравнение (1.41) вместо α его выражение в (1.36). После преобразований получается квадратное уравнение

$$A_s\alpha_s^2 + B_s\alpha_s + C_s = 0, \quad (1.42)$$

где

$$\begin{cases} A_s = AP_{s-1}^2 + BP_{s-1}Q_{s-1} + CQ_{s-1}^2, \\ B_s = 2AP_{s-1}P_{s-2} + B(P_{s-1}Q_{s-2} + P_{s-2}Q_{s-1}) + 2CQ_{s-1}Q_{s-2}, \\ C_s = AP_{s-2}^2 + BP_{s-2}Q_{s-2} + CQ_{s-2}^2. \end{cases}$$

Очевидно, что $A_s, B_s, C_s \in \mathbb{Z}$ и $C_s = A_{s-1}$. Покажем, что дискrimинанты квадратных уравнений (1.41) и (1.42) совпадают при всех s . При доказательстве леммы 1.4 было установлено равенство

$$P_sQ_{s-1} - P_{s-1}Q_s = (-1)^{s+1}.$$

Тогда

$$\begin{aligned} B_s^2 - 4A_sC_s &= (B^2 - 4AC)(P_{s-1}Q_{s-2} + P_{s-2}Q_{s-1})^2 = (B^2 - 4AC)(-1)^{2s} = \\ &= B^2 - 4AC. \end{aligned}$$

Согласно (1.39),

$$\left| \alpha - \frac{P_{s-1}}{Q_{s-1}} \right| < \frac{1}{Q_{s-1}^2}.$$

Тогда найдется число γ_{s-1} , такое, что

$$P_{s-1} = \alpha Q_{s-1} + \frac{\gamma_{s-1}}{Q_{s-1}}, \quad \text{где } |\gamma_{s-1}| < 1.$$

Вычислим коэффициент A_s в квадратном уравнении (1.42):

$$\begin{aligned} A_s &= AP_{s-1}^2 + BP_{s-1}Q_{s-1} + CQ_{s-1}^2 = \\ &= A \left(\alpha Q_{s-1} + \frac{\gamma_{s-1}}{Q_{s-1}} \right)^2 + B \left(\alpha Q_{s-1} + \frac{\gamma_{s-1}}{Q_{s-1}} \right) Q_{s-1} + CQ_{s-1}^2 = \\ &= (A\alpha^2 + B\alpha + C) Q_{s-1}^2 + 2A\alpha\gamma_{s-1} + A \frac{\gamma_{s-1}^2}{Q_{s-1}^2} + B\gamma_{s-1} = \\ &= 2A\alpha\gamma_{s-1} + A \frac{\gamma_{s-1}^2}{Q_{s-1}^2} + B\gamma_{s-1}, \end{aligned}$$

так как α — корень уравнения (1.41). Тогда для любого натурального s ,

$$|A_s| = \left| 2A\alpha\gamma_{s-1} + A \frac{\gamma_{s-1}^2}{Q_{s-1}^2} + B\gamma_{s-1} \right| < 2|A\alpha| + |A| + |B|.$$

Как было замечено ранее $C_s = A_{s-1}$, поэтому для C_s справедлива та же оценка. Таким образом, целые коэффициенты A_s и C_s уравнения (1.42) ограничены по абсолютному значению и, следовательно, при изменении s могут принимать лишь конечное число различных значений. Так как дискриминанты уравнений (1.41) и (1.42) совпадают, то и коэффициент B_s может принимать лишь конечное число различных значений. Поэтому, при изменении s от 1 до ∞ , может встретиться только конечное число различных уравнений вида (1.42), т. е. лишь конечное число различных чисел α_s . Это значит, что некоторые два числа α_s и α_{s+h} с разными номерами обязательно совпадают, что и указывает на периодичность цепной дроби. ■

Замечание 1.22. Отметим без доказательства следующие свойства разложений квадратических иррациональностей:

- 1) при разложении корня квадратного из целого положительного числа, не являющегося полным квадратом, период всегда начинается со второго звена;
- 2) периодическая цепная дробь, у которой период начинается с первого звена получается тогда и только тогда, когда квадратическая иррациональность больше 1, а сопряженная иррациональность лежит в интервале $(-1, 0)$ (это свойство было доказано Э. Галуа в 1828 г. Он доказал также, что в этом случае сопряженная квадратическая иррациональность имеет те же элементы, но расположенные в обратном порядке).

Пример 1.22. Как известно, число e — трансцендентное и не является квадратичной иррациональностью. Тем не менее, интересно разложение числа e в бесконечную (непериодическую) цепную дробь.

$$e = [2; 1, 2, 1, 1, 4, 1, 1, 6, 1, 1, 8, 1, 1, 10, 1, \dots].$$

1.2.9. Наилучшие приближения

Определение 1.13. Дробь $\frac{a}{b}$ называют *наилучшим приближением* к числу α , если не существует дроби $\frac{x}{y}$, у которой

$$0 < y \leq b \text{ и } \left| \alpha - \frac{x}{y} \right| < \left| \alpha - \frac{a}{b} \right|.$$

Замечание 1.23. Знаменатели дробей, не умалляя общности, считаем положительными.

Замечание 1.24. Наилучшее приближение не обязательно является единственным. У иррационального числа наилучших приближений бесконечно много.

Пример 1.23. Пусть $\alpha = \frac{6}{17}$, тогда $\frac{1}{2}, \frac{1}{3}, \frac{4}{11}, \frac{5}{14}$ являются наилучшими приближениями, причем других нет.

Теорема 1.20. Рассмотрим промежуток $\left(\frac{a}{b}; \frac{c}{d} \right)$, обладающий свойством $bc - ad = 1$, т. е. $\frac{c}{d} - \frac{a}{b} = \frac{1}{bd}$. Тогда знаменатели всех рациональных чисел, лежащих внутри промежутка, больше знаменателей его концов, т. е. если $\frac{x}{y} \in \left(\frac{a}{b}; \frac{c}{d} \right)$, то $y > b$ и $y > d$.

Доказательство. Рассмотрим разность

$$\frac{x}{y} - \frac{a}{b} > 0 : \frac{x}{y} - \frac{a}{b} = \frac{xb - ay}{by} \geq \frac{1}{by},$$

так как $xb - ay$ — большее нуля целое. С другой стороны,

$$\frac{x}{y} - \frac{a}{b} < \frac{c}{d} - \frac{a}{b} = \frac{bc - ad}{bd} = \frac{1}{bd}.$$

Откуда $\frac{1}{bd} > \frac{1}{by}$, следовательно, $y > d$. Аналогично доказывается, что $y > b$. ■

Следствие 1.7. Пусть в условиях теоремы 1.20 $\alpha \in \left(\frac{a}{b} ; \frac{c}{d} \right)$, тогда та из двух границ $\frac{a}{b}$ и $\frac{c}{d}$, которая лежит ближе к числу α , является наилучшим приближением к нему.

Следствие 1.8. Все подходящие дроби для разложения числа α в цепную (непрерывную) дробь, начиная со второй, являются наилучшими приближениями к α .

Доказательство. Рассмотрим последовательность подходящих дробей к α :

$$\frac{P_0}{Q_0}, \frac{P_1}{Q_1}, \dots, \frac{P_n}{Q_n}, \frac{P_{n+1}}{Q_{n+1}}, \dots$$

Так как $P_{n+1}Q_n - P_nQ_{n+1} = (-1)^n$, то любые две подходящие дроби, взятые в нужном порядке, обладают свойствами концов промежутка, описанного в теореме. Согласно свойствам подходящих дробей, α всегда находится между последовательными подходящими дробями. Поэтому ближайшая к α граница является наилучшим приближением. Согласно свойствам подходящих дробей, это всегда дробь с большим номером. ■

Упражнение 1.5. Найдите три наилучших приближения к π .

1.2.10. Некоторые применения цепных дробей

Зубчатые колеса Рассмотрим задачу, аналогичную той, с которой встретился голландский математик Х. Гюйгенс (1629–1695) при построении модели солнечной системы с помощью набора зубчатых колес и которая привела его к открытию ряда важных свойств цепных дробей. Пусть требуется соединить два вала звездочками колесами так, чтобы отношение их угловых скоростей равнялось данному числу α . Известно, что угловые скорости колес обратно пропорциональны количеству зубцов, значит обратное отношение чисел зубцов равно α . Но — количества зубцов целые числа и не очень большие, тогда как α может быть и иррациональным. Следовательно, задачу можно решить только приближенно, взяв для α приближенное значение в форме простой дроби с не очень большим знаменателем. Из предыдущих рассуждений вытекает, что наиболее выгодно взять одну из подходящих дробей цепной дроби, в которую раскладывается число α .

Календарный стиль Известно, что в году 365, 24220... так называемых «средних» суток или 365 суток 5 часов 48 минут и 46 секунд. Конечно, такое сложное отношение года к суткам в практической жизни совершенно неудобно. Нужно заменить его более простым, хотя и менее точным.

Разложив $365,24220\dots$ в цепную дробь, получим:

$$365,24220\dots = 365 + \cfrac{1}{4 + \cfrac{1}{7 + \cfrac{1}{1 + \cfrac{1}{3 + \ddots}}}} = (365, 4, 7, 1, 3, 5, 20, 6, 12\dots).$$

Первые подходящие дроби здесь будут:

$$365, \quad 365\frac{1}{4}, \quad 365\frac{7}{29}, \quad 365\frac{8}{33}, \quad 365\frac{31}{128}.$$

Историческая справка. Приближение $365\frac{1}{4}$ было известно еще древним египтянам, ассирийцам, китайцам, хотя они не имели регулярных високосных годов. 7 марта 238 г. до н. э. вышел Канопский декрет Птолемея Эвергета, которым предписывалось, чтобы каждый четвертый год имел не 365, а 366 сут. Но через 40 лет этот декрет был забыт, и только в 47 году до н. э. Юлий Цезарь, при участии Сосигена, возобновил его, вставив в каждом четвертом году лишний день в феврале. Этот день называли *bissextilis*, откуда происходит и русское название «високосный». Это — старый, или Юлианский, календарь.

Новый, или Григорианский, стиль дает приближение $365\frac{97}{400}$ или $365,2425$ суток т. е. 365 суток 5 часов 49 минут и 12 секунд. Это значение больше истинного лишь на 26 с. Такая дробь дает приближение немного хуже, чем вторая подходящая дробь $365\frac{7}{29}$. Этот стиль отличается от Юлианского тем, что в нем каждый сотый год — не високосный, кроме тех, число сотен которых делится на 4. Так, 1700-й, 1800-й, 1900-й годы не високосные, тогда как 1600-й, 2000-й годы високосные, т. е. 400 лет имеют 97 лишних суток, а не 100 суток, как в Юлианском стиле.

Уже в XV в. было замечено отставание Юлианского стиля (тогда на 10 суток) и поступили предложения реформировать календарь, что осуществилось только в конце XVI в. в католических странах буллою папы Григория XIII от 1 марта 1582 г. 10 суток (5–14 октября) были вычеркнуты, т. е. вместо пятого сразу наступило 15 октября 1582 г.

Интересная система календаря была предложена в Персии в 1079 г. персидским астрономом и математиком (и поэтом) Омаром Альхайями. Он ввел цикл из 33 лет, в котором семь раз високосным годом считался четвертый, а восьмой раз — пятым, таким образом, продолжительность года $365\frac{8}{33}$ сут — это четвертая подходящая дробь. Погрешность календаря составляет всего 19 с в год.

В 1864 г. русский астроном И. Медлер предложил с XX в. ввести в России следующую поправку к юлианскому календарю: через каждые 128 лет пропускать один високосный год из 32, которые выпадают на этот период. Этот календарь самый точный из всех перечисленных. Средняя продолжительность года по этому календарю составляет $365\frac{31}{128}$ суток, или 365 суток 5 часов 48 минут и 45 секунд. При этом погрешность сокращается всего до 1 секунды. Однако календарь И. Медлера не был принят, видимо, из-за того, что период 128 лет не является круглым числом.

1.2.11. Диофантовы уравнения

Определение 1.14. Пусть $a, b, c, x, y \in \mathbb{Z}$. Уравнение вида

$$ax + by = c, \quad (1.43)$$

где x, y — неизвестные, называют *диофантовым*.

Историческая справка. О подробностях жизни Диофанта Александрийского практически ничего не известно. Все, что знают историки, почерпнуто из надписи на его гробнице. Основной труд Диофанта — «Арифметика». Уцелели только шесть книг оригинала, общее их число — точно неизвестно (предположительно — тринадцать). Его труд — один из наиболее увлекательных трактатов, сохранившихся от греко-римской древности. В нем впервые встречается систематическое использование алгебраических символов, есть особые знаки для обозначения неизвестного, минуса, обратной величины, возведения в степень. Среди уравнений, решаемых Диофантом, присутствуют такие, как

$$x^2 - 26y^2 = 1 \text{ и } x^2 - 30y^2 = 1,$$

теперь известные как частные случаи «уравнения Пелля», причем Диофант интересуется их решениями именно в целых числах.

«Арифметика» Диофанта неожиданно оказалась еще и огромное косвенное влияние на развитие математической науки последних трех столетий. Дело в том, что фр. математик П. Ферма (1601 – 1665), изучая «Арифметику», сделал на полях этой книги знаменитую пометку: «Я нашел воистину удивительное доказательство того, что уравнение

$$x^n + y^n = z^n \text{ при } n > 2,$$

не имеет решений в целых числах, однако поля этой книги слишком малы, чтобы здесь его уместить». Это утверждение получило название «Великой теоремы Ферма»

Установим критерий разрешимости диофантового уравнения.

Теорема 1.21. Пусть $d = D(a, b)$. Уравнение (1.43) разрешимо тогда и только тогда, когда $c \mid d$. При этом все множество решений описывается уравнениями:

$$x = x_0 - \frac{b}{d}t, \quad y = y_0 + \frac{a}{d}t, \quad (1.44)$$

где x_0, y_0 — любое частное решение (1.43); t — произвольное целое число.

Доказательство. Так как $d = D(a, b)$, то левая часть уравнения (1.43) делится на d . Следовательно, для разрешимости необходимо и достаточно, чтобы и правая часть делилась на d . Пусть $c = de$. Согласно утверждению 1.1, уравнение $ax + by = d$ всегда имеет решение. Обозначим его через x_1, y_1 . Тогда $x_0 = x_1e, y_0 = y_1e$ — решение исходного уравнения. Найдем общий вид решения. Пусть x, y — произвольное решение (1.43). Покажем, что оно представимо в виде (1.44). Имеем

$$a(x - x_0) + b(y - y_0) = 0, \quad \text{тогда} \quad \frac{a}{d}(x - x_0) = -\frac{b}{d}(y - y_0).$$

Согласно теореме 1.3, $D\left(\frac{a}{d}, \frac{b}{d}\right) = 1$. Тогда, согласно теореме 1.6, $(y - y_0) : \frac{a}{d}$, следовательно, $y = y_0 + \frac{a}{d} t$. Подставив найденое значение y в уравнение, получаем $x = x_0 - \frac{b}{d} t$. Таким образом, если x, y — решение (1.43), то оно имеет вид (1.44). Обратное утверждение проверяется непосредственной подстановкой. ■

Следствие 1.9. Справедливы следующие утверждения:

- 1) если $D(a, b) = 1$, то уравнение (1.43) всегда разрешимо;
- 2) решения уравнений $ax + by = c$ и $|a|x + |b|y = c$ могут отличаться только знаками.

Доказательство. Первое утверждение очевидно. Пусть x, y — решения уравнения $ax + by = c$, а x_1, y_1 — соответственно уравнения $|a|x + |b|y = c$. Тогда $x = x_1 \text{sign}(a)$, $y = y_1 \text{sign}(b)$. ■

Опишем алгоритм решения уравнения (1.43).

Алгоритм 1.16 (решение диофанта уравнения).

Находим линейное представление $D(|a|, |b|)$, используя обобщенный алгоритм Евклида.

$$d = D(|a|, |b|) = |a|x_1 + |b|y_1.$$

```

if  $c = d\alpha$  then ▷  $c$  делится на  $d$ 
     $x_0 \leftarrow \alpha x_1 \text{sign}(a)$  ▷  $(x_0, y_0)$  — частное решение (1.43)
     $y_0 \leftarrow \alpha y_1 \text{sign}(b)$ 
    Подставив  $(x_0, y_0)$  в (1.44), получаем общее решение исходного уравнения.
else
    уравнение (1.43) не имеет решений
end if

```

Замечание 1.25. Заметим, что коэффициенты последнего столбца таблицы протокола расширенного алгоритма Евклида равны соответственно $-b/d$ и a/d . Это следует из представления:

$$0 = a \left(-\frac{b}{d} \right) + b \left(\frac{a}{d} \right).$$

Пример 1.24. Решим уравнение $126x - 102y = 18$.

Применив расширенный алгоритм Евклида для чисел 126 и 102, запишем протокол его работы (табл. 1.17).

Таким образом,

$$D(126, 102) = 6 = 126(-4) + 102(5) \text{ тогда } 18 = 126(-12) - 102(-15).$$

Таблица 1.17

	126	102	24	6	0
q			1	4	4
x	1	0	1	-4	17
y	0	1	-1	5	-21

Уравнение разрешимо, так как $18 : 6$. Тогда $x_0 = -12$, $y_0 = -15$. Согласно замечанию 1.25 к алгоритму 1.16 имеем¹ $\frac{b}{d} = -17$, $\frac{a}{d} = 21$. Запишем общее решение:

$$x = -12 + 17t, \quad y = -15 + 21t, \quad t \in \mathbb{Z}.$$

Пример 1.25. Решим уравнение $253x - 449y = 1$.

Применив расширенный алгоритм Евклида для чисел 253 и 449, запишем протокол его работы (табл. 1.18).

Таблица 1.18

	253	449	253	196	57	25	7	4	3	1	0
q			0	1	1	3	2	3	1	1	3
x	1	0	1	-1	2	-7	16	-55	71	-126	449
y	0	1	0	1	-1	4	-9	31	-40	71	-253

Таким образом

$$D(253, 449) = 1 = 253(-126) + 449(71) = 253(-126) - 449(-71).$$

Уравнение разрешимо, так как $1 : 1$. Тогда $x_0 = -126$, $y_0 = -71$. Запишем общее решение:

$$x = -126 + 449t, \quad y = -71 + 253t, \quad t \in \mathbb{Z}.$$

Пример 1.26. Решим уравнение $2071x - 2318y = -95$.

Применив расширенный алгоритм Евклида для чисел 2071 и 2318, запишем протокол его работы (табл. 1.19).

Таким образом, представление для наибольшего общего делителя:

$$D(2071, 2318) = 19 = 2071(-47) + 2318(42).$$

Для исходного уравнения получаем:

$$2071(235) - 2318(210) = -95.$$

¹С учетом знака b .

Таблица 1.19

	2071	2318	2071	247	95	57	38	19	0
q			0	1	8	2	1	1	2
x	1	0	1	-1	9	-19	28	-47	122
y	0	1	0	1	-8	17	-25	42	-109

Уравнение разрешимо, так как $95 : 19$. Тогда $x_0 = 235$, $y_0 = 210$. Запишем общее решение:

$$x = 235 + 122t, \quad y = 210 + 109t, \quad t \in \mathbb{Z}.$$

Теорему 1.21 можно обобщить на случай нескольких переменных.

Теорема 1.22. Уравнение $a_1x_1 + a_2x_2 + \cdots + a_nx_n = c$ разрешимо тогда и только тогда, когда $c : D(a_1, a_2, \dots, a_n)$.

Доказательство. Пусть $d = D(a_1, a_2, \dots, a_n)$. Так как $a_1, \dots, a_n : d$, то и $a_1x_1 + a_2x_2 + \cdots + a_nx_n : d$. Следовательно, если c не делится на d , то решений нет. Пусть $c = de$. Докажем по индукции, что уравнение

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = D(a_1, a_2, \dots, a_n) \quad (1.45)$$

всегда разрешимо. База индукции (при $n = 2$) вытекает из теоремы 1.21. Предположим разрешимость уравнения для произвольного n и докажем этот факт для $n + 1$. Рассмотрим уравнение

$$a_1u_1 + a_2u_2 + \cdots + a_nu_n + a_{n+1}u_{n+1} = D(a_1, a_2, \dots, a_n, a_{n+1}),$$

при этом, согласно теореме 1.4,

$$D(a_1, a_2, \dots, a_n, a_{n+1}) = D(D(a_1, a_2, \dots, a_n), a_{n+1}) = D(d, a_{n+1}).$$

Обозначим $\delta = D(d, a_{n+1})$. Тогда по теореме 1.21 уравнение

$$dy + a_{n+1}u_{n+1} = \delta$$

разрешимо. Подставив вместо $d \rightarrow a_1x_1 + a_2x_2 + \cdots + a_nx_n$, получаем решение уравнения

$$a_1(x_1y) + a_2(x_2y) + \cdots + a_n(x_ny) + a_{n+1}u_{n+1} = \delta.$$

Индукционный переход доказан.

Таким образом, уравнение (1.45) разрешимо для всех n . Учитывая, что $c = de$, получаем решение исходного уравнения

$$a_1(x_1e) + a_2(x_2e) + \cdots + a_n(x_ne) = c. \quad \blacksquare$$

1.2.12. Числа Фибоначчи

Историческая справка. Фибоначчи — «Сын Боначчо» или Леонардо Пизанский (1180 - 1240) — известный средневековый математик, философ, купец и т. д. Путешествовал и торговал в странах Востока. По возвращении в Европу он записал собранные сведения, добавил много собственных исследований и издал книги «Практика геометрии» и «Книга абака». Последовательность Фибоначчи возникает у самого Леонардо при решении следующей задачи: сколько пар кроликов может произойти от одной пары в течение года, если а) каждая пара ежемесячно порождает новую пару, которая со второго месяца способна к размножению, и б) кролики не погибают.

Последовательность Фибоначчи появляется не только при изучении цепных дробей, но и во многих других разделах математики, физики, биологии, искусствоведения. Добавим, что, «Книга абака» была одним из решающих источников проникновения в Западную Европу десятичной системы счисления и арабских цифр.

Название «числа Фибоначчи» стало общеупотребительным только в 19 веке, благодаря математику Люка. Однако, индийские математики упоминали числа этой последовательности еще в 12 веке. Для них это было связано с задачей о количестве ритмических рисунков, образующихся в результате чередования долгих и кратких слогов в стихах или сильных и слабых долей в музыке. Число таких рисунков, имеющих в целом n долей, равно F_n .

Числа Фибоначчи появляются и в работе Кеплера 1611 года, который размышлял о числах, встречающихся в природе (работа «О шестиугольных снежинках»).

Интересен пример растения — тысячелистника, у которого число стеблей (а значит и цветков) всегда есть число Фибоначчи. Причина этого проста: будучи изначально с единственным стеблем, этот стебель затем делится на два, затем от главного стебля отвертывается ещё один, затем первые два стебля снова разветвляются, затем все стебли, кроме двух последних, разветвляются, и так далее. Таким образом, каждый стебель после своего появления «пропускает» одно разветвление, а затем начинает делиться на каждом уровне разветвлений, что и даёт в результате числа Фибоначчи. Вообще говоря, у многих цветов (например, лилий) число лепестков является тем или иным числом Фибоначчи.

Перейдем к точным определениям.

Определение 1.15. Рекуррентное соотношение для чисел Фибоначчи имеет вид

$$\phi_{k+2} = \phi_{k+1} + \phi_k, \quad k \geq 1, \quad \phi_1 = \phi_2 = 1. \quad (1.46)$$

Фибоначчи в 1228 г. привел первые 14 чисел этой последовательности

$$1, 1, 2, 3, 5, 8, 13, 21, 34, \dots, 39\,088\,169, \dots$$

Замечание 1.26. Максимальным числом Фибоначчи, которое помещается в тип *int*, является $\phi_{46} = 1836311903$. В 64-битовый целочисленный тип *long long* помещается максимум — ϕ_{92} . Если в задаче требуется находить значения ϕ_n для $n > 92$, то следует воспользоваться длинной арифметикой

Рассмотрим некоторые свойства чисел Фибоначчи.

Лемма 1.7 (соотношение Кассини¹). Для последовательности чисел Фибоначчи справедливо тождество:

$$\phi_{n-1}\phi_{n+1} - \phi_n^2 = (-1)^n$$

Доказательство. Введем обозначение: $\Delta_n = \phi_{n-1}\phi_{n+1} - \phi_n^2$. Далее, согласно (1.46) имеем:

$$\begin{aligned} \Delta_{n+1} &= \phi_n\phi_{n+2} - \phi_{n+1}^2 = (\phi_{n+1} - \phi_n)\phi_n - (\phi_n + \phi_{n-1})^2 = \\ &= \phi_{n+1}\phi_n + \phi_n^2 - \phi_n^2 - 2\phi_n\phi_{n-1} - \phi_{n-1}^2 = \\ &= \phi_{n+1}\phi_n - 2\phi_n\phi_{n-1} - \phi_{n-1}^2 = (\phi_n + \phi_{n-1})\phi_n - 2\phi_n\phi_{n-1} - \phi_{n-1}^2 = \\ &= \phi_n^2 + \phi_{n-1}\phi_n - 2\phi_n\phi_{n-1} - \phi_{n-1}^2 = \phi_n^2 - \phi_n\phi_{n-1} - \phi_{n-1}^2 = \\ &= \phi_n^2 - (\phi_n + \phi_{n-1})\phi_{n-1} = \phi_n^2 - \phi_{n+1}\phi_{n-1} = -\Delta_n. \end{aligned}$$

Таким образом, $\Delta_{n+1} = -\Delta_n$. Учитывая, что $\Delta_2 = 1$, получаем требуемое. ■

Теорема 1.23 (теорема Цекендорфа). Любое натуральное число N можно представить единственным образом в виде суммы чисел Фибоначчи:

$$N = \phi_{k_1} + \phi_{k_2} + \dots + \phi_{k_r}, \quad k_1 > k_2 + 1, k_2 > k_3 + 1, \dots, k_r \geq 1, \quad (1.47)$$

т. е. в записи (1.47) нельзя использовать два соседних числа Фибоначчи.

Доказательство. Воспользуемся индукцией по номерам чисел Фибоначчи n . Из равенства (1.46) легко заметить, что любое натуральное число N попадает в промежуток между двумя соседними числами Фибоначчи, т. е. для некоторого $n \geq 2$ верно неравенство:

$$\phi_n \leq N < \phi_{n+1}.$$

Таким образом, $N = \phi_n + N_1$, где $N_1 = N - \phi_n < \phi_{n-1}$, так что разложение числа N_1 уже не будет содержать слагаемого ϕ_{n-1} . ■

¹Кассини — это знаменитая династия французских астрономов. Наиболее известным из них считается основатель этой династии Джованни Доменико Кассини (1625–1712). Именем Джованни Кассини названы многие астрономические объекты: «Кратер Кассини» на Луне, «Кратер Кассини» на Марсе, «Щель Кассини» — промежуток в кольцах Сатурна, «Законы Кассини» — три открытые Кассини законы движения Луны.

Из равенства 1.47 очевидно следует, что любое число можно однозначно записать в *фибоначчиевой системе счисления*, например:

$$\begin{aligned} 9 &= 8 + 1 = \phi_6 + \phi_1 = (100001)_\phi \\ 6 &= 5 + 1 = \phi_5 + \phi_1 = (10001)_\phi \\ 19 &= 13 + 5 + 1 = \phi_7 + \phi_5 + \phi_1 = (1010001)_\phi, \end{aligned}$$

причём ни в каком числе не могут идти две единицы подряд.

Перевод числа N в фибоначчиеву систему счисления осуществляется простым алгоритмом: перебираем числа Фибоначчи от больших к меньшим и, если некоторое $\phi_k \leq N$, то включаем ϕ_k в запись числа N , полагаем $N = N - \phi_k$ и продолжаем поиск.

Формула для общего члена последовательности чисел Фибоначчи (1.120) будет получена далее в разделе 1.7.1 с помощью аппарата производящих функций.

Числа Фибоначчи тесно связаны с *числом Фидия*² — Φ , характеризующим так называемое «золотое сечение» или «божественную пропорцию».

Определение 1.16. Рассмотрим деление единичного отрезка на две части. Большую часть отрезка обозначим через α . Если большая часть так относится к меньшей, как весь отрезок к большей его части, то говорят, что имеет место «золотое сечение» отрезка (рис. 1.2). В этом случае данное отношение называют числом Фидия.

$$\Phi = \frac{\alpha}{1 - \alpha} = \frac{1}{\alpha}. \quad (1.48)$$

Замечание 1.27. Легко показать, что в определении 1.16 можно рассматривать отрезок любой длины.

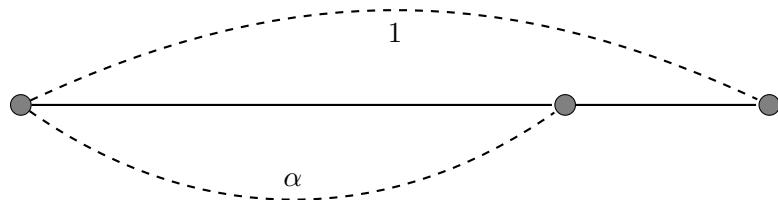


Рис. 1.2

Из (1.48) просто вычислить число Фидия. Имеем:

$$\alpha^2 = 1 - \alpha, \quad \Phi = \frac{1 + \sqrt{5}}{2}$$

²Названо так в честь древнегреческого скульптора и архитектора, для работ которого характерно применение золотого сечения.

Установим связь чисел Фибоначчи с числом Фидия.

Теорема 1.24. Справедливо следующее соотношение:

$$\phi_n = \frac{\Phi^n - (1 - \Phi)^n}{\sqrt{5}} \quad (1.49)$$

Доказательство. Применим метод математической индукции. Нетрудно проверить, что соотношение (1.49) выполняется для ϕ_1 и ϕ_2 . База индукции установлена. Предположим, что равенство (1.49) выполняется для всех $\phi_k, k < n$ и покажем его справедливость для ϕ_n . Заметим, что из (1.48) следуют очевидные равенства:

$$1 + \Phi = \Phi^2, \quad 2 - \Phi = (1 - \Phi)^2. \quad (1.50)$$

Далее, с учетом (1.50) получаем:

$$\begin{aligned} \phi_n &= \phi_{n-1} + \phi_{n-2} = \frac{\Phi^{n-1} - (1 - \Phi)^{n-1}}{\sqrt{5}} + \frac{\Phi^{n-2} - (1 - \Phi)^{n-2}}{\sqrt{5}} = \\ &= \frac{\Phi^{n-2}(1 + \Phi) - (1 - \Phi)^{n-2}(2 - \Phi)}{\sqrt{5}} = \frac{\Phi^n - (1 - \Phi)^n}{\sqrt{5}} \end{aligned}$$

■

В следующем разделе будет получено интересное свойство периодичности последовательности чисел Фибоначчи, связанное с модульной арифметикой (теорема 1.33).

1.3. Арифметика остатков

1.3.1. Алгебраические структуры

Для последующих подразделов нам понадобятся некоторые понятия алгебры. Дадим необходимые определения. В дальнейшем запись (Ω, \circ) означает, что на непустом множестве Ω задана операция \circ .

Определение 1.17. Множество (Ω, \circ) называют *группой*, если выполнены следующие условия:

1) операция \circ ассоциативна, т. е. для любых $a, b, c \in \Omega$ выполнено

$$a \circ (b \circ c) = (a \circ b) \circ c;$$

2) множество Ω обладает нейтральным элементом относительно операции \circ , т. е. существует элемент $e \in \Omega$, такой, что для любого элемента $a \in \Omega$, выполнено

$$a \circ e = e \circ a = a;$$

3) для любого элемента $a \in \Omega$, существует противоположный элемент $\tilde{a} \in \Omega$, такой, что

$$a \circ \tilde{a} = \tilde{a} \circ a = e.$$

Если для любых $a, b \in \Omega$ справедливо

$$a \circ b = b \circ a,$$

то такую группу называют *коммутативной* или *абелевой*. В коммутативных группах операцию \circ обычно называют операцией сложения и обозначают $+$, нейтральный элемент e называют нулем группы и обозначают 0 , хотя этот элемент, вообще говоря, числом не является. Противоположный элемент \tilde{a} обозначают $-a$.

Пример 1.27. Примеры абелевых групп:

- а) аддитивные группы (группы по сложению) целых чисел \mathbb{Z} , рациональных чисел \mathbb{Q} и вещественных чисел \mathbb{R} ;
- б) мультипликативные группы (группы по умножению) положительных рациональных чисел \mathbb{Q}_+ и положительных вещественных чисел \mathbb{R}_+ .

Определение 1.18. Непустое множество Ω , на котором заданы операции сложения и умножения, называют *кольцом*, если выполнены следующие два условия:

- а) $(\Omega, +)$ — абелева группа;
- б) умножение дистрибутивно относительно сложения, т. е. для любых элементов $x, y, z \in \Omega$ выполнены равенства:

$$(x + y)z = xz + yz; \quad x(y + z) = xy + xz.$$

Кольцо называют *коммутативным*, если операция умножения в нем коммутативна; кольцо называют *ассоциативным*, если операция умножения в нем ассоциативна. Кольцо называют *кольцом с единицей*, если оно обладает нейтральным элементом относительно умножения.

Пусть Ω — ассоциативное кольцо с единицей (обозначим ее как e). Элемент $a \in \Omega$ называют *обратимым*, если существует элемент $b \in \Omega$, такой, что

$$ab = ba = e.$$

Легко проверить, что элемент b , о котором идет речь, находится однозначно, поэтому его обозначают a^{-1} и называют элементом, обратным к a .

Если для ненулевых элементов $a, b \in \Omega$ справедливо $ab = 0$, то такие элементы называют *делителями нуля*. Очевидно, что делители нуля необратимы.

Пример 1.28. Примеры колец:

- а) кольцо целых чисел \mathbb{Z} , кольцо рациональных чисел \mathbb{Q} , кольцо вещественных чисел \mathbb{R} ;
- б) кольцо $\mathbb{Z}[i]$ целых гауссовых чисел вида $a + bi$, где $a, b \in \mathbb{Z}$;
- в) кольцо $\mathbb{Z}[\sqrt{2}]$ вещественных чисел вида $a + b\sqrt{2}$ с целыми a, b .

Важнейшим типом колец являются поля.

Определение 1.19. Ассоциативное коммутативное кольцо с единицей называют *полем*, если в нем всякий ненулевой элемент обратим.

Пример 1.29. Примеры полей:

- а) числовые поля \mathbb{Q}, \mathbb{R} ;
- б) поле $\mathbb{Q}[i]$ рациональных чисел вида $a + bi$, где $a, b \in \mathbb{Q}$;
- в) поле $\mathbb{Q}[\sqrt{2}]$ вещественных чисел вида $a + b\sqrt{2}$ с рациональными a, b .

Другие примеры колец и полей будут приведены в следующем подразделе.

1.3.2. Арифметика и свойства сравнений

Определение 1.20. Пусть $a, b, m \in \mathbb{Z}, m > 0$. Говорят, что a сравнимо с b по модулю m ($a \equiv b \pmod{m}$), или сокращенно $a \equiv b(m)$, если $(a - b) : m$.

Историческая справка. Знак \equiv для обозначения операции сравнения впервые ввел К. Гаусс в 1801 г.

Теорема 1.25. Для отношения сравнения справедливо:

- 1) $a \equiv b(m)$, тогда $b \equiv a(m)$ (симметричность);
- 2) $a \equiv b(m)$, $b \equiv c(m)$, тогда $a \equiv c(m)$ (транзитивность);
- 3) $a \equiv a(m)$ (рефлексивность).

Замечание 1.28. Известные бинарные отношения $(>, <, \leqslant, \geqslant, :)$ такими свойствами не обладают.

Определение 1.21. Классом вычетов по модулю m называют множество чисел с одинаковым остатком при делении на m .

Теорема 1.26. Все числа из \mathbb{Z} распределяются относительно данного модуля на m непересекающихся классов вычетов со следующими свойствами:

- 1) все числа одного класса сравнимы друг с другом по модулю m ;

2) числа из разных классов друг с другом не сравнимы.

Доказательство. Пусть $a \in \mathbb{Z}$. Тогда по теореме 1.1 $a = mq + r$, $0 \leq r < m$. Очевидно, что классов вычетов ровно m и все числа из одного класса сравнимы друг с другом. Возьмем числа из разных классов: $a \equiv r(m), b \equiv r_1(m), r \neq r_1$. Тогда $a \not\equiv b \pmod{m}$. Действительно, предположим противное. Пусть $(a - b) \mid m$, но $a = qm + r, b = q_1m + r_1$, тогда $(r - r_1) \mid m$, следовательно, $r = r_1$. ■

Определение 1.22. Если взять по одному представителю из каждого класса вычетов, то эти m чисел образуют *полную систему вычетов* по модулю m .

Пример 1.30. Простейшие полные системы вычетов:

- 1) $\{0, 1, 2, \dots, m - 1\}$ — наименьшие положительные вычеты;
- 2) $\{0, -1, -2, \dots, -(m - 1)\}$ — наименьшие отрицательные вычеты;
- 3) для произвольного $a \in \mathbb{Z}$ $\{a, a + 1, a + 2, \dots, a + (m - 1)\}$;
- 4) если $D(a, m) = 1$, то $\{0, a, 2a, \dots, (m - 1)a\}$.

Теорема 1.27. Справедливы следующие свойства сравнений:

- 1) $a \equiv b(m), m \mid k$, тогда $a \equiv b(k)$;
- 2) $a \equiv b(k_1), \dots, a \equiv b(k_n)$, тогда $a \equiv b(M(k_1, \dots, k_n))$;
- 3) $a \equiv b(m)$, тогда $ac \equiv bc \pmod{mc}$.
- 4) $a \equiv b(m_1), \dots, a \equiv b(m_k)$, $D(m_i, m_j) = 1$, $i \neq j$, тогда

$$a \equiv b \pmod{m_1 m_2 \cdots m_k}.$$

Доказательство. Свойство 1 очевидно из определения. Свойство 2 следует из того, что $(a - b)$ — общее кратное чисел k_1, k_2, \dots, k_n , а любое кратное делится на наименьшее (см. подраздел 1.1.2). Для доказательства свойства 3 заметим, что

$$\frac{ac - bc}{mc} = \frac{a - b}{m}.$$

Свойство 4 следует из определения операции сравнения по модулю и следствия к теореме 1.6. ■

Следствие 1.10. $a \equiv b(m)$, тогда $ac \equiv bc \pmod{m}$ для любого $c \in \mathbb{Z}$.

Теорема 1.28 (арифметика сравнений). Сравнения по одному модулю можно почленно складывать, вычитать и перемножать.

Доказательство. Пусть $a \equiv b(m)$, $a_1 \equiv b_1(m)$. Тогда

$$((a - b) + (a_1 - b_1)) = ((a + a_1) - (b + b_1)) \vdots m,$$

следовательно, $a + a_1 \equiv b + b_1(m)$. С другой стороны, так как $a \equiv b(m)$, по следствию 1.10 $-a \equiv -b(m)$. Повторив предыдущие рассуждения, получим $a - a_1 \equiv b - b_1(m)$. Используя следствие 1.10 для сравнения $a \equiv b(m)$, получаем $aa_1 \equiv ba_1(m)$, применяя это же следствие для второго сравнения $a_1 \equiv b_1(m)$, имеем $ba_1 \equiv bb_1(m)$. Тогда по транзитивности сравнений (см. теорему 1.25) $aa_1 \equiv bb_1(m)$. ■

Следствие 1.11. Утверждение теоремы 1.28 можно обобщить на любое конечное число сравнений. В частности, если $a \equiv b(m)$, то

$$a^n \equiv b^n \pmod{m} \quad \text{для любого } n \in \mathbb{N}.$$

Теорема 1.28 определяет действия над классами вычетов по данному модулю m . Если $a \in A, b \in B$, тогда $A + B$ — класс вычетов, к которому принадлежит $a + b$, а AB — это класс, к которому принадлежит ab .

Таким образом, справедлива следующая теорема.

Теорема 1.29. Множество классов вычетов (обозначают \mathbb{Z}_m , также используют обозначения $\mathbb{Z}/(m)$ и $\mathbb{Z}/m\mathbb{Z}$) является ассоциативным коммутативным кольцом с единицей.

Доказательство предлагается провести самостоятельно с использованием определения 1.18.

Пример 1.31. Построим несколько простейших признаков делимости, как иллюстрацию арифметики сравнений. Пусть число m представлено в десятичной системе счисления. Согласно (1.15), для $p = 10$ имеем

$$m = 10^n a_n + 10^{n-1} a_{n-1} + \cdots + 10a_1 + a_0, \quad (1.51)$$

1. Заметим, что $10 \equiv 1 \pmod{9}$. Тогда $10^k \equiv 1 \pmod{9}$, $k \in \mathbb{N}$. Из (1.51) получаем признак делимости на 9:

$$m \equiv a_n + a_{n-1} + \cdots + a_1 + a_0 \pmod{9}$$

2. Очевидно, что $10 \equiv -1 \pmod{11}$. Тогда $10^k \equiv (-1)^k \pmod{11}$ для $k \in \mathbb{N}$. Получаем признак делимости на 11:

$$m \equiv a_0 - a_1 + a_2 + \cdots + (-1)^n a_n \pmod{11}.$$

Теорема 1.30. Если модуль m — составное число, то \mathbb{Z}_m не является полем.

Доказательство. Пусть

$$m = p_1 p_2, \quad 1 < p_1, p_2 < m. \quad (1.52)$$

Будем считать, что $P_1, P_2 \in \mathbb{Z}_m$ — классы вычетов, которым принадлежат p_1, p_2 . Тогда, в силу (1.52)

$$P_1 P_2 = 0, \quad P_1, P_2 \neq 0,$$

и элементы P_1, P_2 — необратимы. Следовательно, \mathbb{Z}_m — не поле. ■

Далее будет показано (см. следствие 1.17 к теореме 1.39), что в случае простого модуля \mathbb{Z}_m — есть поле.

Пример 1.32. Рассмотрим таблицы арифметических действий для \mathbb{Z}_6 (табл. 1.20 *a* и *b*).

В \mathbb{Z}_6 имеются делители нуля. Действительно: $2 \cdot 3 = 0, 3 \cdot 4 = 0$. Для элементов $\{2, 3, 4\}$ нет обратных.

Таблица 1.20

a

b

+	0	1	2	3	4	5	×	0	1	2	3	4	5
0	0	1	2	3	4	5	0	0	0	0	0	0	0
1	1	2	3	4	5	0	1	0	1	2	3	4	5
2	2	3	4	5	0	1	2	0	2	4	0	2	4
3	3	4	5	0	1	2	3	0	3	0	3	0	3
4	4	5	0	1	2	3	4	0	4	2	0	4	2
5	5	0	1	2	3	4	5	0	5	4	3	2	1

Замечание 1.29. Как легко заметить из последнего примера, таблица по сложению (табл. 1.20 *a*) в \mathbb{Z}_6 является левоциркулярной матрицей. Этот факт нетрудно доказать для произвольного модуля.

Перейдем к делению сравнений.

Теорема 1.31. Пусть $ac \equiv bc \pmod{m}$, $D(c, m) = d$, тогда $a \equiv b \left(\frac{m}{d} \right)$.

Доказательство. Пусть $c = c_1 d$, $m = m_1 d$, тогда $ac - bc = lm$, таким образом,

$$l = \frac{ac - bc}{m} = \frac{(a - b)c_1}{m_1},$$

следовательно, $(a - b)c_1 \mid m_1$, но $(c_1, m_1) = 1$, тем самым $(a - b) \mid m_1$ или, что то же самое, $a \equiv b \left(\frac{m}{d} \right)$. ■

Следствие 1.12. Рассмотрим два важных предельных случая теоремы 1.31:

1. если $m \mid c$, тогда из $ac \equiv bc(m)$ следует $a \equiv b \left(\frac{m}{c} \right)$;
2. если $D(m, c) = 1$, тогда из $ac \equiv bc(m)$ следует $a \equiv b(m)$.

Результаты, полученные в этом подразделе, обобщает следующая теорема.

Теорема 1.32. Пусть $F(a, b, \dots)$ — произвольная целая рациональная функция от $a, b, \dots \in \mathbb{Z}$, т. е.

$$F(a, b, \dots) = \sum_k C_k a^{\alpha_k} b^{\beta_k} \dots, \text{ где } C_k, \alpha_k, \beta_k, \dots \in \mathbb{Z}, \alpha_k, \beta_k, \dots > 0.$$

Если $a \equiv a_1(m), b \equiv b_1(m), \dots$, то $F(a, b, \dots) \equiv F(a_1, b_1, \dots)(m)$.

В качестве простейшего приложения теоремы докажем следующее утверждение.

Утверждение 1.3. Квадрат всякого нечетного числа сравним с единицей по модулю 8.

Доказательство. Действительно, любое нечетное число можно представить в виде $4k \pm 1$. Тогда $(4k \pm 1)^2 = 16k^2 \pm 8k + 1 \equiv 1 \pmod{8}$. ■

С модульной арифметикой связано интересное свойство последовательности чисел Фибоначчи (раздел 1.2.12).

Теорема 1.33. Последовательность чисел Фибоначчи $\{\phi_i\}$, взятых по модулю m , является периодичной, причем период начинается с $\phi_1 = 1$.

Доказательство. Рассмотрим $m^2 + 1$ пар чисел Фибоначчи, взятых по модулю m :

$$(\phi_1, \phi_2), (\phi_2, \phi_3), \dots, (\phi_{m^2+1}, \phi_{m^2+2})$$

По модулю m может быть только m^2 различных пар, поэтому среди этой последовательности найдется как минимум две одинаковые пары. Это означает, что последовательность периодична.

Возьмем среди всех таких одинаковых пар две пары с наименьшими номерами. Обозначим их как (ϕ_n, ϕ_{n+1}) и (ϕ_k, ϕ_{k+1}) . Если $n > 1$, тогда предыдущие пары (ϕ_{n-1}, ϕ_n) и (ϕ_{k-1}, ϕ_k) по свойству чисел Фибоначчи, тоже будут равны друг другу. Однако это противоречит выбору совпадающих пар с наименьшими номерами, таким образом, $n = 1$. ■

1.3.3. Функция Эйлера и ее свойства

Историческая справка. Леонард Эйлер (1707–1783) — самый плодовитый математик XVIII в., если только не всех времен. Опубликовано более двухсот томов его научных трудов, но это еще далеко не полное собрание сочинений.

Л. Эйлер закончил университет в швейцарском городе Базеле, где изучал математику под руководством Иоганна Бернулли, а когда в 1725 г. сыновья Иоганна Николай и Даниил уехали в Петербург, Леонардо последовал за ними в недавно учрежденную по замыслу Петра I Российскую (Петербургскую) Академию Наук. Эйлер жил в России до 1741 г., затем работал в Берлинской академии под особым покровительством Фридриха II, а с 1766 г. до конца жизни снова в России. Очевидно, что Эйлера с полным правом можно считать российским ученым, ибо основные годы его творчества прошли в Петербурге и он являлся академиком именно Петербургской Академии наук под особым покровительством Екатерины II.

В 1766 г. ученый ослеп, но продолжал работать. Пользуясь своей феноменальной памятью, он диктовал статьи и книги, общее число которых достигло 886. Его работы посвящены анализу, алгебре, дискретной математике (теории графов), вариационному исчислению, функциям комплексного переменного, астрономии, гидравлике, теоретической механике, кораблестроению, артиллерии, теории музыки и т. д. Колossalная продуктивность Эйлера в разных областях математики и в других областях науки была и остается поводом для изумления. Обозначения Эйлера почти современны, точнее современная математическая символика почти Эйлерова. Можно составить огромный список известных и важных математических открытий, приоритет в которых принадлежит великому ученому. Многие его идеи ждут разработки до настоящего времени.

Определение 1.23. Функция Эйлера $\varphi(m)$ ставит в соответствие каждому натуральному m количество чисел, меньших m и взаимно простых с m . Будем полагать $\varphi(1) = 1$.

Вычислять функцию Эйлера согласно определению затруднительно. Выведем общую формулу для произвольного аргумента.

Лемма 1.8. Пусть $m = p^n$, где p — простое число. Тогда

$$\varphi(m) = p^n \left(1 - \frac{1}{p}\right).$$

Доказательство. Из интервала $(0, p^n)$ на p делятся только числа, кратные p . Это $p, 2p, 3p, \dots, p^{n-1}p$. Их всего $p^{n-1} - 1$. Остальные числа из данного интервала взаимно просты с p . Поэтому

$$\varphi(m) = \varphi(p^n) = p^n - (p^{n-1} - 1) = p^n \left(1 - \frac{1}{p}\right). \quad \blacksquare$$

Докажем вспомогательную теорему.

Теорема 1.34. Пусть x пробегает полную систему вычетов по модулю a , а y — соответственно по модулю b . При этом $D(a, b) = 1$. Тогда:

- 1) $z = ay + bx$ пробегает полную систему вычетов по модулю ab .

2) $D(z, ab) = 1$ тогда и только тогда, когда $D(x, a) = 1$, $D(y, b) = 1$.

Доказательство. x принимает a значений, y принимает b значений, следовательно, z принимает ab значений. Покажем, что никакие два значения z не сравнимы друг с другом по модулю ab . Пусть

$$z_1 = ay_1 + bx_1, \quad z_2 = ay_2 + bx_2, \quad ay_1 + bx_1 \equiv ay_2 + bx_2 \pmod{ab}$$

тогда, согласно теореме 1.27,

$$ay_1 + bx_1 \equiv ay_2 + bx_2 \pmod{a}, \quad ay_1 + bx_1 \equiv ay_2 + bx_2 \pmod{b}.$$

По следствию 1.12(п.2) (так как $D(a, b) = 1$) имеем

$$x_1 \equiv x_2 \pmod{a}, \quad y_1 \equiv y_2 \pmod{b}.$$

Но x_1, x_2 — из полной системы вычетов по модулю a , а y_1, y_2 — из полной системы вычетов по модулю b . Тогда $x_1 = x_2, y_1 = y_2, z_1 = z_2$. Таким образом, первая часть теоремы доказана.

Пусть $D(z, ab) = 1$, тогда очевидно, что

$$D(ay + bx, a) = 1, \quad D(ay + bx, b) = 1.$$

Следовательно, $z - ay = bx$ взаимно просто с a , а $z - bx = ay$ взаимно просто с b . Так как $D(a, b) = 1$, то $D(a, x) = 1, D(b, y) = 1$. Повторив последние рассуждения в обратном порядке, получаем достаточность п.2. ■

Следствие 1.13.

$$D(a, b) = 1, \text{ тогда } \varphi(ab) = \varphi(a)\varphi(b). \quad (1.53)$$

Доказательство. В полной системе вычетов по модулю a существует $\varphi(a)$ значений x , таких, что $D(a, x) = 1$. Аналогично для модуля b получаем $\varphi(b)$ значений y , таких, что $D(b, y) = 1$. Следовательно, всего имеется $\varphi(a)\varphi(b)$ значений z , взаимно простых с ab . Но значения z образуют полную систему вычетов по модулю ab , и чисел, взаимно простых с ab , в ней $\varphi(ab)$. ■

Пример 1.33. Для иллюстрации доказательства следствия 1.13 составим таблицу (табл. 1.21) величин (x, y) при $a = 5, b = 8$. Возможные значения для x — числа $0, 1, 2, 3, 4$, возможные значения для y — числа $0, 1, 2, 3, 4, 5, 6, 7$. Из них для x имеется четыре значения взаимно простых с a (так как $\varphi(5) = 4$). Соответственно для y также есть четыре значения взаимно простых с b ($\varphi(8) = 4$). Эти значения выделены в таблице «серым» цветом, как и соответствующие им значения $z = ay + bx$.

Выделенные значения z дают 16 чисел, меньших 40 и взаимно простых с ним. Таким образом:

$$\varphi(40) = \varphi(5)\varphi(8) = 4 \cdot 4 = 16.$$

Таблица 1.21

x	y							
	0	1	2	3	4	5	6	7
0	0	5	10	15	20	25	30	35
1	8	13	18	23	28	33	38	3
2	16	21	26	31	36	1	6	11
3	24	29	34	39	4	9	14	19
4	32	37	2	7	12	17	22	27

Замечание 1.30. Равенство (1.53) легко обобщить на случай нескольких попарно взаимно простых сомножителей.

Следствие 1.14. Пусть $m \in \mathbb{Z}$. Согласно (1.6), m представимо в виде

$m = p^\alpha q^\beta \cdots$, где p, q, \dots — простые числа, $\alpha, \beta, \dots \geq 1$. Тогда

$$\varphi(m) = p^{\alpha-1}(p-1)q^{\beta-1}(q-1)\cdots = m \left(1 - \frac{1}{p}\right) \left(1 - \frac{1}{q}\right) \cdots. \quad (1.54)$$

Доказательство. Так как p^α, q^β, \dots попарно взаимно просты, то, применяя последовательно следствие 1.13 и лемму 1.8, получаем требуемое. ■

Пример 1.34.

$$\varphi(49) = \varphi(7^2) = 7^2 - 7 = 42,$$

$$\varphi(30) = \varphi(2 \cdot 3 \cdot 5) = \varphi(2)\varphi(3)\varphi(5) = (2-1)(3-1)(5-1) = 8,$$

$$\varphi(60) = 60 \left(1 - \frac{1}{2}\right) \left(1 - \frac{1}{3}\right) \left(1 - \frac{1}{5}\right) = 16.$$

Рассмотрим важное применение формулы (1.54).

Теорема 1.35 (формула Гаусса). Пусть d пробегает все делители числа m . Тогда¹ $m = \sum_{d|m} \varphi(d)$.

Доказательство. Пусть m представимо в виде (1.6). Тогда для любого делителя d справедливо представление

$$d = p^\chi q^\lambda \cdots, \text{ где } 0 \leq \chi \leq \alpha, 0 \leq \lambda \leq \beta, \dots$$

¹Индекс $d|m$ у знака \sum означает суммирование по всем делителям числа m .

Рассмотрим произведение

$$[1 + \varphi(p) + \varphi(p^2) + \cdots + \varphi(p^\alpha)][1 + \varphi(q) + \varphi(q^2) + \cdots + \varphi(q^\beta)] \cdots .$$

Перемножив все скобки, имеем

$$\sum_{\chi, \lambda \dots} \varphi(p^\chi) \varphi(q^\lambda) \cdots = \sum_{\chi, \lambda \dots} \varphi(p^\chi q^\lambda \cdots) = \sum_d \varphi(d) .$$

С другой стороны, согласно лемме 1.8, каждую скобку можно представить в виде

$$1 + \varphi(p) + \varphi(p^2) + \cdots + \varphi(p^\alpha) = 1 + (p - 1) + (p^2 - p) + \cdots + (p^\alpha - p^{\alpha-1}) = p^\alpha .$$

Перемножив, получим $p^\alpha q^\beta \cdots = m$. ■

Пример 1.35.

$$\begin{aligned} \sum_{d|30} \varphi(d) &= \varphi(1) + \varphi(2) + \varphi(3) + \varphi(5) + \varphi(6) + \varphi(10) + \varphi(15) + \varphi(30) = \\ &= 1 + 1 + 2 + 4 + 2 + 4 + 8 + 8 = 30. \end{aligned}$$

Функция Эйлера обладает еще одним интересным свойством.

Теорема 1.36. Если $n \in \mathbb{N}$ и $n > 2$, то $\varphi(n)$ — четное число.

Доказательство. Так как $n > 2$, то возможны два случая.

$$n = \begin{cases} 2^k n_1, & k > 1, \\ p^s n_2, & s \geq 1, p — \text{простое число и } p \geq 3. \end{cases}$$

В обоих случаях четность $\varphi(n)$ очевидным образом следует из (1.54). ■

Для функции Эйлера существуют задачи не разрешенные и по сей день.

Гипотеза (Кармайкл, 1907). Если посмотреть на последовательность значений функции Эйлера: $\{1, 1, 2, 2, 4, 2, 6, 4, 6, 4, \dots\}$, можно заметить, что она содержит много повторяющихся чисел.

Гипотеза Кармайкла состоит в том, что не существует такого значения n , которое функция Эйлера принимала бы только один раз.

Введем важное определение.

Определение 1.24. Функцию $f : \mathbb{R} \rightarrow \mathbb{R}$ называют *мультипликативной* если:

- 1) функция f определена всюду на \mathbb{N} и существует $a \in \mathbb{N}$, такой, что $f(a) \neq 0$;

2) для любых взаимно простых натуральных чисел a_1 и a_2 выполняется $f(a_1a_2) = f(a_1)f(a_2)$.

Очевидно, что в силу следствия 1.13 к теореме 1.34 функция Эйлера является мультипликативной функцией, но далеко не единственной. Рассмотрим еще один важный пример мультипликативной функции.

Определение 1.25. Функция Мёбиуса $\mu(m)$ определяется следующим образом.

1. Полагаем $\mu(1) = 1$.
2. Пусть число $m \in \mathbb{N}$ представлено в каноническом виде (1.6)

$$m = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_k^{\alpha_k}, \quad \alpha_i \geq 1, \quad i \in 1 : k. \quad (1.55)$$

Тогда

$$\mu(m) = \begin{cases} 0, & \text{если } \max_{i \in 1:k} \{\alpha_i\} > 1, \\ (-1)^k & \text{в противном случае.} \end{cases}$$

Таким образом, если число m делится на квадрат натурального числа, отличный от единицы, то $\mu(m) = 0$.

Пример 1.36.

$$\mu(30) = \mu(2 \cdot 3 \cdot 5) = (-1)^3 = -1, \quad \mu(60) = \mu(2^2 \cdot 3 \cdot 5) = 0.$$

Упражнение 1.6. Покажите, что функция Мёбиуса — мультипликативна.

Для функции Мёбиуса можно получить результат, аналогичный теореме 1.35.

Теорема 1.37. Справедливо тождество

$$\sum_{d|m} \mu(d) = \begin{cases} 0, & \text{если } m \geq 2, \\ 1 & \text{в противном случае.} \end{cases}$$

Доказательство. Если $m = 1$, то $\mu(m) = 1$ по определению. Пусть $m \geq 2$ представлено в виде (1.55). Тогда

$$\begin{aligned} \sum_{d|m} \mu(d) &= \mu(1) + [\mu(p_1) + \cdots + \mu(p_k)] + [\mu(p_1p_2) + \cdots + \mu(p_{k-1}p_k)] + \\ &\quad + [\mu(p_1p_2p_3) + \cdots + \mu(p_{k-2}p_{k-1}p_k)] + \cdots + \mu(p_1p_2 \cdots p_k) = \\ &= 1 - k + C_k^2 + \cdots + (-1)^s C_k^s + \cdots + (-1)^k = \sum_{s=0}^k (-1)^s C_k^s = 0. \end{aligned} \quad (1.56)$$



Замечание 1.31. Подробно биномиальные коэффициенты C_k^s будут рассмотрены в подразделе 1.6.3. Последнее равенство в (1.56) будет доказано в следствии 1.22 к теореме 1.62.

Установим еще одно свойство функции Эйлера.

Теорема 1.38 (теорема Эйлера–Ферма). Пусть $D(a, m) = 1$, тогда $a^{\varphi(m)} \equiv 1(m)$.

Доказательство. Пусть $D(a, m) = 1$, $\varphi(m) = k$. Тогда классов вычетов, взаимно простых с m , будет k . Пусть $\{a_1, a_2, \dots, a_k\}$ — представители этих классов. Тогда $\{aa_1, aa_2, \dots, aa_k\}$ также взаимно прости с m . Предположим, что $aa_i \equiv aa_j(m)$, тогда $a_i \equiv a_j(m)$. Следовательно, числа aa_i — также представители всех классов вычетов, взаимно прости с m . Тогда каждое aa_i сравнимо с одним и только одним a_j , т. е.

$$aa_1 \equiv a_\alpha \pmod{m}, \quad aa_2 \equiv a_\beta \pmod{m}, \dots, \quad aa_\mu \equiv a_\theta \pmod{m}, \text{ где } 1 \leq \alpha, \beta, \theta \leq k.$$

Перемножив последние k сравнений, имеем $a_1 \cdots a_k a^k \equiv a_\alpha \cdots a_\theta(m)$. Заметим, что $a_\alpha, \dots, a_\theta$ — это те же числа, что и a_1, \dots, a_k , только в другом порядке (см. пример 1.37). После сокращения (в силу $D(a_i, m) = 1$) получаем нужное сравнение. ■

Пример 1.37. Проиллюстрируем доказательство теоремы 1.38. Пусть, $a = 4$, $m = 7$. Тогда

$$1*4 \equiv 4 \pmod{7}, \quad 2*4 \equiv 1 \pmod{7}, \quad 3*4 \equiv 5 \pmod{7}, \quad 4*4 \equiv 2 \pmod{7}, \quad 5*4 \equiv 6 \pmod{7}, \quad 6*4 \equiv 3 \pmod{7}$$

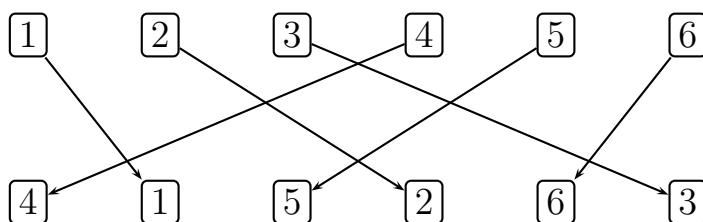


Рис. 1.3

Следствие 1.15 («малая» теорема Ферма). Если p — простое, то $a^p \equiv a(p)$.

Замечание 1.32. Этот частный случай установил П. Ферма в XVII в., а полностью теорема 1.38 была доказана Л. Эйлером.

Следствие 1.16. Для простого p справедливо

$$(a + b)^p \equiv a^p + b^p \pmod{p}$$

Доказательство. Достаточно заметить, что

$$(a + b)^p \equiv a + b \pmod{p}, \quad a^p \equiv a \pmod{p} \text{ и } b^p \equiv b \pmod{p}.$$

■

Пример 1.38. Найдем остаток от деления $25^{11^{35}}$ на 34.

Вычислим $\varphi(34) = \varphi(2 \cdot 17) = \varphi(2) \cdot \varphi(17) = 1 \cdot 16 = 16$. Так как $D(25, 34) = 1$, то, согласно теореме 1.38 (Эйлера–Ферма),

$$25^{16} \equiv 1 \pmod{34}$$

Далее имеем $11^{35} = 16k + b$, где b — неизвестный остаток. Если число b известно, то в \mathbb{Z}_{34} справедливо

$$25^{11^{35}} = 25^{16k+b} = 25^{16k} \cdot 25^b = (25^{16})^k \cdot 25^b = 25^b.$$

Найдем теперь остаток b . Так как

$$D(11, 16) = 1, \text{ и } \varphi(16) = \varphi(2^4) = 2^4 - 2^3 = 8,$$

то, согласно теореме 1.38,

$$11^8 \equiv 1 \pmod{16}$$

Тогда в \mathbb{Z}_{16} справедливо

$$b = 11^{35} = 11^{32} \cdot 11^3 = (11^8)^4 \cdot 11^3 = 1 \cdot 11^3 = 1331 = 3.$$

Решаем теперь исходную задачу. Имеем в \mathbb{Z}_{34} :

$$25^{11^{35}} = 25^3 (34) = 15\,625 = 19 (34).$$

1.3.4. Линейные сравнения

Определение 1.26. *Линейным сравнением¹* называют уравнение вида:

$$ax \equiv b \pmod{m}, \text{ где } a, b, x, m \in \mathbb{Z}. \quad (1.57)$$

В качестве решений уравнения (1.57) рассмотрим классы вычетов по модулю m , тогда сравнение может иметь не более m решений. Установим критерий разрешимости сравнения (1.57).

Теорема 1.39. Справедливо следующее:

1) пусть $d = D(a, m)$. Сравнение (1.57) разрешимо тогда и только тогда, когда $d | b$;

¹По аналогии с линейными уравнениями в элементарной алгебре.

2) при этом существует d классов решений

$$\left\{ x_0, x_0 + \frac{m}{d}, x_0 + 2\frac{m}{d}, \dots, x_0 + (d-1)\frac{m}{d} \right\}. \quad (1.58)$$

Доказательство. Очевидно, что разрешимость сравнения (1.57) эквивалентна разрешимости диофантового уравнения $ax - my = b$. Применив теорему 1.21 устанавливаем п. 1. Далее, все решения соответствующего диофантового уравнения (а нас интересует только неизвестное x) описываются согласно (1.44) следующим образом: $x = x_0 - \frac{m}{d} t$. Различных по модулю m среди них ровно d , т. е. п. 2 установлен. ■

Следствие 1.17. Если p является простым числом, то сравнение $ax \equiv b \pmod{p}$ всегда имеет единственное решение (кроме случая $p | a$).

Таким образом (см. теорему 1.30), в случае простого модуля для любого ненулевого $a \in \mathbb{Z}_p$ существует обратный элемент a^{-1} , т. е. \mathbb{Z}_p является полем. Такие конечные поля называют *полями Галуа* и обозначают $GF(p)$

На основании теоремы 1.39 опишем алгоритм решения линейного сравнения.

Алгоритм 1.17 (решение линейного сравнения).

```

Алгоритмом 1.16 решаем диофантово уравнение  $ax - my = b$ 
 $d \leftarrow D(a, m)$ 
 $\text{if } d \mid b \text{ then}$ 
     $x_0 \leftarrow$  любое решение диофантова уравнения
     $\triangleright$  Согласно теореме 1.39 получаем все решения сравнения (1.57) в виде (1.58)
     $\text{for } i \leftarrow 1, d \text{ do}$ 
         $x_0 \leftarrow x_0 + m/d$ 
     $\text{end for}$ 
 $\text{else}$ 
    сравнение (1.57) не имеет решений.
 $\text{end if}$ 
```

Пример 1.39. Решить сравнение $378x \equiv 90 \pmod{120}$.

Следуя теореме 1.39, решим диофантово уравнение

$$378x - 120y = 90. \quad (1.59)$$

Воспользуемся алгоритмом 1.16. Запишем протокол работы расширенного алгоритма Евклида для чисел 378 и 120 (табл. 1.22).

Таким образом

$$D(378, 120) = 6 = 378 \cdot 7 + 120 \cdot (-22) \Rightarrow 90 = 378 \cdot 105 - 120 \cdot 330.$$

Уравнение разрешимо, так как $90 \div 6$, при этом $x_0 = 105$. Согласно теореме 1.39, существует $d = 6$ классов решений по модулю 120, получающихся из x_0 сдвигом на $m/d = 120/6 = 20$:

$$105, 125, 145, 165, 185, 205.$$

Таблица 1.22

	378	120	18	12	6	0
q			3	6	1	2
x	1	0	1	-6	7	-20
y	0	1	-3	19	-22	63

Замечание 1.33. В качестве x_0 в примере 1.39 вместо 105 можно взять любое решение диофантова уравнения (1.59):

$$x = x_0 + \frac{b}{d}t = 105 + 20t, \quad t \in \mathbb{Z},$$

например, -15. Добавляя последовательно 20, получаем те же шесть классов решений по модулю 120:

$$\begin{aligned} -15 &\equiv 105(120), \quad 5 \equiv 125(120), \quad 25 \equiv 145(120), \\ 45 &\equiv 165(120), \quad 65 \equiv 185(120), \quad 85 \equiv 205(120). \end{aligned}$$

Пример 1.40. Вычислить $\frac{46}{65}$ по модулю 71.

Найдем $x = 65^{-1} \pmod{71}$. Перепишем последнее равенство в виде $65x = 1 \pmod{71}$. Решим это сравнение. Воспользуемся алгоритмом 1.16. Запишем протокол работы расширенного алгоритма Евклида для чисел 65 и 71 (табл. 1.23).

Таблица 1.23

	65	71	65	6	5	1	0
q			0	1	10	1	5
x	1	0	1	-1	11	-12	71
y	0	1	0	1	-10	11	-65

Таким образом,

$$D(71, 65) = 1 = 65 \cdot (-12) + 71 \cdot 11 = 65 \cdot (-12) - 71 \cdot (-11).$$

Отсюда $x_1 = -12 \pmod{71} = 59 = 65^{-1}$. Вычисляем

$$46 \cdot 65^{-1} \pmod{71} = 46 \cdot 59 \pmod{71} = 2714 \pmod{71} = 16 \pmod{71}.$$

Рассмотрим еще один способ решения сравнения (1.57), принадлежащий Л. Эйлеру.

Будем считать, что числа a и m взаимно прости. К этой ситуации всегда можно привести сравнение, разделив a, b и m на $D(a, m)$. Согласно теореме 1.38

$$a^{\varphi(m)} \equiv 1 \pmod{m}, \quad a^{\varphi(m)}b \equiv b \pmod{m}, \quad a(a^{\varphi(m)-1}b) \equiv b \pmod{m}.$$

Тогда

$$x^* = (a^{\varphi(m)-1}b) \pmod{m} \quad (1.60)$$

является решением (1.57).

Замечание 1.34. Недостаток такого подхода в том, что число a нужно возводить в степень $\varphi(m)$. Вычисления упрощаются, если возводить в степень в кольце \mathbb{Z}_m (см. замечание 1.10 к алгоритму 1.11).

Пример 1.41. Решить сравнение

$$11x \equiv 15 \pmod{24}.$$

Здесь $D(11, 24) = 1$, а $\varphi(24) = 8$. Согласно (1.60), нужно вычислить 11^7 .

Имеем (в Z_{34}):

$$11^2 = 1 \implies 11^6 = 1 \implies 11^7 = 11 \implies \text{в } Z_{34} \quad x^* = 11 \cdot 15 = -3.$$

В случае простого модуля решение (а оно в этом случае единственное) сравнения (1.57) можно получить в явном виде. Справедлива теорема.

Теорема 1.40. Пусть p — простое число, $0 < a < p$. Тогда сравнение $ax \equiv b \pmod{p}$ имеет решение:

$$x \equiv b(-1)^{a-1} \frac{C_p^a}{p} \pmod{p}. \quad (1.61)$$

Доказательство. Запишем набор из a тривиальных сравнений:

$$p-1 \equiv -1 \pmod{p}, \quad p-2 \equiv -2 \pmod{p}, \quad \dots, \quad p-(a-1) \equiv -(a-1) \pmod{p}, \quad b \equiv b \pmod{p}.$$

Перемножив эти a сравнений, получим

$$b(-1)^{a-1}(p-1)(p-2)\cdots(p-(a-1)) \equiv b \cdot 1 \cdot 2 \cdots (a-1) \pmod{p}. \quad (1.62)$$

Домножим и разделим левую часть сравнения (1.62) на $a!$. Получим

$$b(-1)^{a-1}a! \frac{(p-1)(p-2)\cdots(p-(a-1))}{a!} \equiv b \cdot 1 \cdot 2 \cdots (a-1) \pmod{p}. \quad (1.63)$$

Так как по условию теоремы $0 < a < p$ и p — простое число, то

$$D(1 \cdot 2 \cdots (a-1), p) = 1.$$

Разделим сравнение (1.63) почленно на $(a-1)!$. Согласно следствию 1.12 к теореме 1.31, получаем требуемое. ■

Пример 1.42. Решим сравнение $8x \equiv 2 \pmod{13}$. Вычисляем решение согласно (1.61):

$$x \equiv 2(-1)^7 \frac{C_{13}^8}{13} (13) \equiv -\frac{2}{13} C_{13}^8 (13) \equiv -198 (13) \equiv 10 (13).$$

С помощью линейных сравнений установим критерий простоты числа.

Теорема 1.41 (теорема Вильсона¹). Число $p \in \mathbb{Z}$ является простым тогда и только тогда, когда $(p-1)! \equiv -1 \pmod{p}$.

Доказательство. Пусть p — простое. Не умаляя общности, считаем, что $p > 3$. Тогда для любого $a \in \mathbb{Z}$, не делящегося на p , сравнение

$$ax \equiv 1 \pmod{p} \quad (1.64)$$

имеет единственный класс решений $x \equiv b(p)$, где b — некоторый вычет из $\{1, 2, \dots, p-1\}$. Заметим, что $b \neq 0$, в противном случае, если $x \equiv 0(p)$, то $p | x$. Рассмотрим все пары $\{a, b\}$, взаимно простые с p , из той же системы вычетов, что и решения (1.64), такие, что $ab \equiv 1(p)$. Пусть $a = b$. Имеем $a^2 \equiv 1(p)$, тогда $(a+1)(a-1) : p$, следовательно, $a \equiv \pm 1(p)$, т. е. $a = 1$ или $a = -1$. Тогда, для пар $\{a, b\}$, таких, что $2 \leq a, b \leq p-2$, справедливо $a \neq b$. Таких пар всего $(p-3)/2$.

Перемножив $(p-3)/2$ сравнения вида $ab \equiv 1(p)$, получим

$$2 \cdot 3 \cdots (p-2) \equiv 1 \pmod{p}.$$

Домножив на тривиальное сравнение $p-1 \equiv -1(p)$ получим критерий Вильсона.

Пусть теперь p — составное число. Тогда $D((p-1)!, p) > 1$, следовательно, $(p-1)! + 1$ не может делиться на p . ■

1.3.5. Китайская теорема об остатках

Перейдем к системам сравнений первой степени. Рассмотрим важный частный случай, когда модули сравнений системы попарно взаимно просты.

Теорема 1.42 (китайская теорема об остатках). Система сравнений

$$x \equiv c_1 (m_1), \quad x \equiv c_2 (m_2), \dots, \quad x \equiv c_k (m_k), \quad (1.65)$$

где $D(m_i, m_j) = 1$ для $i \neq j$, имеет единственное решение (по модулю $m_1 m_2 \cdots m_k$).

¹А. Вильсон (1714–1786) — шотландский астроном и математик, профессор астрономии в Глазго.

Доказательство. Введем обозначения:

$$M = \prod_{i=1}^k m_i, \quad M_l = \frac{M}{m_l}, \quad l \in 1 : k.$$

Так как $D(M_l, m_l) = 1$ для всех $l \in 1 \dots k$, то согласно теореме 1.39, каждое сравнение $M_l x \equiv 1 (m_l)$ имеет единственное решение (x_l) . Тогда легко построить решение системы (1.65):

$$x^* \equiv \sum_{l=1}^k M_l x_l c_l \pmod{M}. \quad (1.66)$$

Подставив x^* в каждое сравнение (1.65), убеждаемся, что (1.66) — решение системы. Подставим, для примера, x^* в первое уравнение (1.65). Тогда $M_2 x_2 c_2 + M_3 x_3 c_3 + \dots + M_k x_k c_k \equiv 0 \pmod{m_1}$, но $M_1 x_1 \equiv 1 \pmod{m_1}$, следовательно, $M_1 x_1 c_1 \equiv c_1 \pmod{m_1}$ и $x^* \equiv c_1 \pmod{m_1}$.

Покажем, что решение x^* единственно по модулю M . Предположим, что x — другое решение системы (1.65). Тогда выполнены сравнения

$$x^* - x \equiv 0 \pmod{m_1}, \dots, x^* - x \equiv 0 \pmod{m_k},$$

откуда

$$(x^* - x) \vdots m_1, (x^* - x) \vdots m_2 \dots, (x^* - x) \vdots m_k.$$

По следствию к теореме 1.6, получаем, что

$$(x^* - x) \vdots \prod_{i=1}^k m_i, \quad x^* \equiv x \pmod{M}.$$

Последнее сравнение и завершает доказательство. ■

Определение 1.27. Набор c_1, c_2, \dots, c_k называют *китайским кодом* числа x^* .

Историческая справка. Теорема 1.42 была впервые описана в трактате китайского математика Сунь Ізы «Математическое руководство», предположительно в третьем веке н. э. О самом авторе не известно ничего, кроме того, что он является автором этой книги; годы его жизни устанавливались историками на основе анализа текста трактата.

Пример 1.43. Найдем наименьшее натуральное число x , удовлетворяющее условиям:

$$x \equiv 2 \pmod{29}, \quad x \equiv 12 \pmod{20}, \quad x \equiv 20 \pmod{23}, \quad x \equiv 4 \pmod{13}.$$

Согласно обозначениям теоремы 1.65, имеем:

$$\begin{aligned}M &= m_1 m_2 m_3 m_4 = 29 \cdot 20 \cdot 23 \cdot 13 = 173\,420, \\M_1 &= m_2 m_3 m_4 = 20 \cdot 23 \cdot 13 = 5\,980, \\M_2 &= m_1 m_3 m_4 = 29 \cdot 23 \cdot 13 = 8\,671, \\M_3 &= m_1 m_2 m_4 = 29 \cdot 20 \cdot 13 = 7\,540, \\M_4 &= m_1 m_2 m_3 = 29 \cdot 20 \cdot 23 = 13\,340.\end{aligned}$$

Решим сравнение

$$5\,980x_1 \equiv 1 \pmod{29}.$$

Воспользуемся алгоритмом 1.16. Запишем протокол работы расширенного алгоритма Евклида для чисел 5980 и 29 (табл. 1.24).

Таблица 1.24

	5 980	29	6	5	1	0
<i>q</i>			206	4	1	5
<i>x</i>	1	0	1	-4	5	-29
<i>y</i>	0	1	-206	825	-1 031	5 980

Таким образом

$$D(5\,980, 29) = 1 = 5\,980 \cdot 5 + 29 \cdot (-1\,031).$$

Откуда $x_1 = 5$, $M_1 x_1 = 29\,900$.

Решим сравнение

$$8\,671x_2 \equiv 1 \pmod{20}.$$

Применив алгоритм 1.16, запишем протокол работы расширенного алгоритма Евклида для чисел 8 671 и 20 (табл. 1.25).

Таким образом

$$D(8\,671, 20) = 1 = 8\,671 \cdot (-9) + 20 \cdot 3\,920.$$

Откуда $x_2 = -9 = 11(20)$, $M_2 x_2 = 95\,381$.

Решим сравнение

$$7\,540x_3 \equiv 1 \pmod{23}.$$

Использовав алгоритм 1.16, запишем протокол работы расширенного алгоритма Евклида для чисел 7 540 и 23 (табл. 1.26).

Таким образом

$$D(7\,540, 23) = 1 = 7\,540 \cdot (-6) + 23 \cdot 1\,967.$$

Откуда $x_3 = -6 = 17(23)$, $M_3 x_3 = 128\,180$.

Таблица 1.25

	8671	20	11	9	2	1	0
q			433	1	1	4	2
x	1	0	1	-1	2	-9	20
y	0	1	-433	434	-867	3902	-8671

Таблица 1.26

	7 540	23	19	1	4	1	0
q			327	1	4	1	3
x	1	0	1	-1	5	-6	23
y	0	1	-327	328	-1 639	1 967	-7 540

Решим сравнение

$$13\ 340x_4 \equiv 1 \pmod{13}.$$

Применив алгоритм 1.16, запишем протокол работы расширенного алгоритма Евклида для чисел 13 340 и 13 (табл. 1.27).

Таблица 1.27

	13 340	13	2	1	0
q			1 026	6	2
x	1	0	1	-6	13
y	0	1	-1 026	6 157	-13 340

Таким образом,

$$D(13 340, 13) = 1 = 13 340 \cdot (-6) + 13 \cdot 6 157.$$

Откуда $x_4 = -6 = 7(13)$, $M_4x_4 = 93 380$.

Находим решение x^* .

$$\begin{aligned} x^* &= (M_1x_1c_1 + M_2x_2c_2 + M_3x_3c_3 + M_4x_4c_4) \pmod{M} \\ &= (29 900 \cdot 2 + 95 381 \cdot 12 + 128 180 \cdot 20 + 93 380 \cdot 4) \pmod{173 420} = \\ &= 152 832. \end{aligned}$$

Замечание 1.35. Основную трудоемкость представляет вычисление значений $M_1x_1, M_2x_2, \dots, M_kx_k$. Однако для фиксированной системы модулей m_1, m_2, \dots, m_k они вычисляются только один раз и не зависят от китайского кода самого числа.

1.3.6. Система остаточных классов

Рассмотренная выше китайская теорема об остатках (теорема 1.42) является основой одной из «непозиционных систем счисления» — так называемой «системой остаточных классов»¹ или «модулярной арифметике».

Современные исследования по параллельным вычислениям показали, что в рамках обычной позиционной системы счисления значительного ускорения выполнения арифметических операций добиться невозможно. Это объясняется тем, что в таких системах значение разряда любого числа (кроме младшего) зависит не только от значения одноименных операндов, но и от всех предшествующих разрядов, т. е. позиционные системы обладают *последовательной структурой*.

Сегодня, все большее внимание привлекают системы счисления, обладающие способностями к параллельной обработке информации. Такими возможностями обладает система остаточных классов, которая позволяет реализовать идею распараллеливания операций на уровне выполнения элементарных арифметических действий.

В системе остаточных классов согласно теореме 1.42 каждое число $0 \leq N \leq M - 1$ однозначно представляется вектором своего китайского кода (c_1, \dots, c_k) — см. определение 1.27. Набор модулей (p_1, \dots, p_k) называют *базисом* системы. Представление самого числа в системе обычно записывают следующим образом:

$$N = (c_1 | \cdots | c_k)_{(p_1 | \cdots | p_k)}.$$

Выполнение арифметических операций в RNS производится поразрядно, *независимо* по каждому из модулей, что позволяет распараллелить алгоритмы при выполнении арифметических операций. Это избавляет от необходимости «занимать» или «переносить» единицу старшего разряда.

$$X = (x_1 | \cdots | x_k)_{(p_1 | \cdots | p_k)}, \quad Y = (y_1 | \cdots | y_k)_{(p_1 | \cdots | p_k)}.$$

Тогда

$$X + Y = ((x_1 + y_1) \bmod p_1 | \cdots | (x_k + y_k) \bmod p_k)_{(p_1 | \cdots | p_k)}, \quad (1.67)$$

$$XY = (x_1 y_1 \bmod p_1 | \cdots | x_k y_k \bmod p_k)_{(p_1 | \cdots | p_k)}$$

Систему используемых модулей (p_1, \dots, p_k) подбирают под конкретную задачу. Например, для представления 32-х битных чисел достаточно взять систему модулей: $(7, 11, 13, 17, 19, 23, 29, 31)$ — все модули простые, следовательно попарно взаимнопросты и

$$M = \prod_{i=1}^k p_i = 6685349671 > 4294967296 = 2^{32}.$$

¹Оригинальное английское название — Residue Number System (RNS)

Каждый из модулей не превышает 5 бит, таким образом поразрядные операции сложения и умножения будут производиться над 5-битными числами.

Пример 1.44. Рассмотрим арифметические действия с числами 13 и 11 в системе остаточных классов с базисом $(3, 7, 8)$. Тогда, согласно (1.67), имеем:

$$\begin{aligned} 13 &= (1|6|5)_{(3|7|8)}, \quad 11 = (2|4|3)_{(3|7|8)}, \\ 13 + 11 &= (0|3|0)_{(3|7|8)}, \quad 13 \cdot 11 = (2|3|7)_{(3|7|8)}. \end{aligned}$$

В настоящее время система остаточных классов активно применяется в цифровой обработке сигналов, криптографии, обработки изображений и. т. д.

Из недостатков RNS можно отметить следующие:

- a) ограниченный диапазон представления чисел;
- b) сложные алгоритмы деления;
- c) трудности в обнаружении переполнения.

Для более детального изучения данной темы можно рекомендовать монографию [59] (на английском языке). В ней — обширная информация о применении RNS с большим количеством подробных примеров.

1.3.7. Применение теоремы Эйлера в криптографии. Система шифрования RSA

Развитие телекоммуникаций привело к повышению роли алгоритмов шифровки информации и развитию нового направления в криптографии — науке о способах преобразования (шифрования) информации с целью ее защиты от незаконных пользователей. Центральным понятием этого направления является понятие односторонней функции.

Определение 1.28. Функцию f называют *односторонней*, если существует эффективный (полиномиальный) алгоритм вычисления ее значений $f(x)$, но для решения уравнения $f(x) = a$ (обращения функции f) эффективного алгоритма нет.

Несмотря на то что построенные на понятии односторонней функции криптографические системы уже работают и понятия электронной подписи и электронных денег вошли в экономическую практику и закреплены законодательно, математического доказательства того, что эти функции действительно односторонние, пока не получено. Поэтому неизвестно, существуют ли односторонние функции.

Электронная подпись и электронные деньги основаны на так называемой функции с секретом (или ловушкой).

Определение 1.29. Функцией с секретом K называют функцию f_K , зависимую от параметра K и обладающую тремя свойствами:

- 1) при любом K существует эффективный (полиномиальный) алгоритм вычисления f_K ;
- 2) при неизвестном K не существует эффективного алгоритма обращения f_K ;
- 3) при известном K существует эффективный алгоритм обращения f_K .

Наиболее известной и популярной односторонней функцией является функция $f(x) = x^e \bmod m$, лежащая в основе шифра RSA(Rivest, Shamir, Adleman). Рассмотрим эту систему шифрования более подробно.

Историческая справка. Основы криптографии с открытыми ключами были выдвинуты Уитфилдом Диффи и Мартином Хеллманом. Идея состояла в том, что ключи можно использовать парами — ключ шифрования и ключ дешифрирования, и что невозможно получить один ключ из другого. Диффи и Хеллман впервые представили этот подход в 1976 г¹., а спустя год появился первый алгоритм шифрования — RSA, основанный на этом подходе.

Еще до выхода из печати статья авторов системы RSA была послана известному популяризатору математики М. Гарднеру, который в 1977 г. в журнале *Scientific American* опубликовал статью², посвященную этой системе шифрования. В русском переводе название статьи Гарднера звучит так: «Новый вид шифра, на расшифровку которого потребуются миллионы лет». Именно эта статья Гарднера сыграла важнейшую роль в распространении информации об RSA, привлекла к криптографии внимание широких кругов неспециалистов и фактически способствовала бурному прогрессу этой области, произошедшему в последующие 20 лет.

Будем считать, что шифруемые сообщения представлены (закодированы) длинными целыми числами.

Пусть m и e — натуральные числа. Функция f , осуществляющая шифрование, определяется следующим образом:

$$f(x) = x^e \bmod m.$$

Для дешифрования сообщения нужно решить сравнение

$$x^e \equiv a \pmod{m}.$$

Для возведения x в степень e (в кольце \mathbb{Z}_m) есть эффективный алгоритм (см. алгоритм 1.11). Для решения же обратной задачи — нахождения x по e и a — такого алгоритма предположительно нет. Поэтому f — односторонняя функция.

¹ W. Diffie and M. E. Hellman. New Directions in Cryptography // IEEE Transactions on Information Theory, vol. IT-22, Nov. 1976, PP. 644—654.

² Gardner M. A new kind of cipher that would take millions of years to break // Scientific American. — 1977. — P. 120–124.

Оказывается, при некоторых условиях, налагаемых на m и e , функцию f можно использовать как функцию с секретом. Найдем эти условия.

Пусть число e взаимно просто с $\varphi(m)$, где φ — функция Эйлера (см. определение 1.23), а d — решение сравнения $de \equiv 1 \pmod{\varphi(m)}$. Тогда

$$de = 1 + k\varphi(m), \quad k \in \mathbb{Z}, \quad 1 \leq d < \varphi(m).$$

Так как $D(e, \varphi(m)) = 1$, то, согласно теореме 1.39, существует единственный класс решений $d \pmod{\varphi(m)}$.

Авторы шифра RSA предложили выбирать число m в виде произведения двух простых множителей p и q , примерно одинаковых по значению. Тогда

$$\varphi(m) = \varphi(pq) = \varphi(p) \cdot \varphi(q) = (p-1)(q-1)$$

и условия взаимной простоты e с $\varphi(m)$ сводятся к условиям взаимной простоты e с $p-1$ и с $q-1$.

Покажем, что число d может служить секретом.

Теорема 1.43. Для любого $x \in \mathbb{Z}_m$, $x^{ed} \equiv x \pmod{m}$.

Доказательство. В силу выбора чисел d, e и m

$$de = 1 + k \cdot \varphi(m) = 1 + k(p-1)(q-1)$$

при некотором k . Если $x \not\equiv 0 \pmod{p}$, то по следствию 1.15 имеем:

$$x^{ed} \equiv x (x^{p-1})^{k(q-1)} \equiv x \cdot 1^{k(q-1)} \equiv x \pmod{p}.$$

Если $x \equiv 0 \pmod{p}$, то сравнение $x^{ed} \equiv x \pmod{p}$, очевидно, выполняется. Таким образом,

$$\forall x \in \mathbb{Z}_m, \quad x^{ed} \equiv x \pmod{p}.$$

Аналогично получаем

$$\forall x \in \mathbb{Z}_m, \quad x^{ed} \equiv x \pmod{q}.$$

Применяя пункт 4 теоремы 1.27, получаем требуемое. ■

Итак, лицо, заинтересованное в организации шифрованной переписки с помощью шифра RSA, выбирает два достаточно больших простых числа p и q и перемножает их: $m = pq$. Затем выбирается число e , взаимно простое с $p-1$ и с $q-1$; вычисляется $\varphi(m)$ и d . Числа m и e публикуются, а число d держится в секрете. Теперь любой может отправить шифрованное сообщение лицу, опубликовавшему m и e , расшифровать которое сможет только последний.

Замечание 1.36. Для вычисления функции $f(x) = x^e \bmod m$ достаточно знать лишь числа e и m . Именно они составляют открытый ключ для шифрования. А вот для вычисления обратной функции требуется знать число d , оно и является «секретом».

Казалось бы, ничего не стоит, зная число m , разложить его на простые сомножители, вычислить затем значение $\varphi(m)$ и определить нужное число d . Все шаги этого вычисления могут быть реализованы достаточно быстро, за исключением первого. Именно разложение числа на простые множители и составляет наиболее трудоемкую часть вычислений. В теории чисел несмотря на многолетнюю ее историю и интенсивные поиски в течение последних 20 лет, эффективный алгоритм разложения натуральных чисел на множители так и не найден. Конечно, можно, перебирая все простые числа до \sqrt{m} , и, деля на них m , найти требуемое разложение. Но, учитывая, что количество простых чисел в этом промежутке, асимптотически равно

$$\frac{2\sqrt{m}}{\ln m},$$

находим, что при m , записываемом 100 десятичными цифрами, найдется не менее $4 \cdot 10^{42}$ простых чисел, на которые придется делить m при разложении его на множители. Очень грубые прикидки показывают, что компьютеру, выполняющему миллион делений в секунду, для разложения числа $m > 10^{99}$ таким способом на простые сомножители потребуется не менее, чем 10^{35} лет. Известны и более эффективные способы разложения целых чисел на множители, чем перебор простых делителей, но и они работают очень медленно.

Для иллюстрации своего метода авторы RSA зашифровали фразу «The magic words are squeeamish ossifrage», предварительно заменив ее символы наборами из двух десятичных цифр: 00 — пробел, 01 — а, 02 — б, …, 26 — з.

При этом m было равно числу ..., которое получилось перемножением простых чисел, записываемых соответственно 64 и 65 десятичными знаками.

Число e взяли равным 9007.

Это сообщение ждало своей расшифровки 17 лет и ожидание завершилось в 1994 г., когда D. Atkins, M. Graff, A. K. Lenstra и P. C. Leyland сообщили о дешифровке фразы.

Для дешифровки сообщения пришлось искать разложение на множители 129-значного десятичного числа. Это было непросто, поскольку для разложения числа на множители до сих пор не найдено эффективного алгоритма. Все известные алгоритмы, в том числе и использованный в решении метод квадратичного решета, основаны на методе полного перебора.

Интересно, что нахождение этого разложения стало, по-видимому, первым серьезным сетевым проектом. В работе, возглавляемой четырьмя авторами проекта и продолжавшейся 220 дней, участвовало 660 человек и 1 600 компьютеров, объединенных сетью Интернет.

При практическом применении шифров с открытым ключем в настоящее время используются простые числа до 400 и более цифр в десятичной записи. В результате вскрыть этот шифр оказывается невозможно (в обозримое время).

Рассмотрим работу алгоритма RSA на примере.

Пример 1.45.

Генерация ключей. Выбираем пару простых чисел $p = 3$, $q = 11$. Тогда

$$m = pq = 3 \cdot 11 = 33; \quad (p - 1)(q - 1) = 2 \cdot 10 = 20.$$

Выбираем число e , взаимно простое с $\varphi(m) = 20$, например $e = 7$. Тогда открытая часть ключа $(m, e) = (33, 7)$. Вычислим закрытую часть ключа d . Имеем

$$7d \equiv 1 \pmod{20}, \text{ тогда } d = 3 \ (7 \cdot 3 = 21 \pmod{20} = 1).$$

Шифрование сообщения. Зашифруем слово LIFE. Шифруем сообщение S .

$$S = [S_1, S_2, S_3, S_4], \text{ где } S_1 = L = 12, S_2 = I = 9, S_3 = F = 6, S_4 = E = 5.$$

Тогда кодированное сообщение имеет вид $C = [C_1, C_2, C_3, C_4]$, где

$$\begin{aligned} C_1 &= S_1^7 \pmod{33} = 12^7 \pmod{33} = 12, \\ C_2 &= S_2^7 \pmod{33} = 9^7 \pmod{33} = 15, \\ C_3 &= S_3^7 \pmod{33} = 6^7 \pmod{33} = 30, \\ C_4 &= S_4^7 \pmod{33} = 5^7 \pmod{33} = 14. \end{aligned}$$

Декодирование сообщения. Декодируем принятое сообщение. Имеем

$$\begin{aligned} D_1 &= C_1^3 \pmod{33} = 12^3 \pmod{33} = 1728 \pmod{33} = 12, \\ D_2 &= C_2^3 \pmod{33} = 15^3 \pmod{33} = 3375 \pmod{33} = 9, \\ D_3 &= C_3^3 \pmod{33} = 30^3 \pmod{33} = 27\,000 \pmod{33} = 6, \\ D_4 &= C_4^3 \pmod{33} = 14^3 \pmod{33} = 2\,744 \pmod{33} = 5. \end{aligned}$$

Рассмотри еще один пример работы алгоритма RSA.

Пример 1.46. Русское слово из четырех букв закодированного с помощью алгоритма RSA с открытым ключом $e = 19, m = 46$. Зашифрованное сообщение имеет вид: $(5, 12, 16, 26)$. Буквам русского алфавита соответствуют числа в диапазоне от 2 до 32 (исключены буквы „Ё“ и „Ъ“).

Нужно подобрать закрытую часть ключа и дешифровать исходное слово.

Определяем закрытую часть ключа d :

$$\varphi(m) = \varphi(2 \cdot 23) = 22, 19d \equiv 1 \pmod{22}, d = 7 = (1, 1, 1)_2.$$

Используя алгоритм быстрого возведения в степень по модулю (алгоритм 1.11), получаем

$$\begin{aligned} 1) c &= 1, 2) c = 1^2 \cdot 5 = 5, 3) c = 5^2 \cdot 5 = 125(46) = 33, \\ 4) c &= 33^2 \cdot 5 = 5445(46) = 17. \end{aligned}$$

Первая буква сообщения П (17). Далее, аналогично,

- 1) $c = 1$,
- 2) $c = 1^2 \cdot 12 = 12$,
- 3) $c = 12^2 \cdot 12 = 1728(46) = 26$,
- 4) $c = 26^2 \cdot 12 = 8112(46) = 16$.

Вторая буква сообщения О (16). Далее:

- 1) $c = 1$,
- 2) $c = 1^2 \cdot 16 = 16$,
- 3) $c = 16^2 \cdot 16 = 4096(46) = 2$,
- 4) $c = 2^2 \cdot 16 = 64(46) = 18$.

Третья буква сообщения Р (18). Определяем последнюю букву слова:

- 1) $c = 1$,
- 2) $c = 1^2 \cdot 26 = 12$,
- 3) $c = 26^2 \cdot 26 = 17576(46) = 4$,
- 4) $c = 4^2 \cdot 26 = 416(46) = 2$.

Четвертая буква сообщения А (2). Слово „ПОРА“.

1.3.8. Простейшие атаки на систему RSA

Появление новой системы шифрования RSA привело к многочисленным попыткам ее взлома, она проверяется «на прочность» многими исследователями с момента первой публикации. Хотя четверть века исследований привели к большому числу крайне интересных атак, ни одна из них не представляет большой угрозы. Они скорее иллюстрируют опасность неправильного использования RSA. Рассмотрим несколько простейших вариантов атак на RSA, которые позволяют взломать ее, не используя прямого разложения модуля m на произведение двух простых чисел.

□ Схема с нотариусом

Начальные условия. Известны открытая часть ключа нотариуса (m, e) .

Задача. Получить подпись нотариуса на сообщении S .

Решение. Противник выбирает затемняющий множитель $r \in \mathbb{Z}_m$ и $r \neq 0$, затем вычисляет

$$\widehat{S} = r^e S \pmod{m} \quad (1.68)$$

и отправляет это сообщение на подпись нотариусу. Если нотариус подписывает это сообщение:

$$\widehat{C} = \widehat{S}^d \pmod{m} = r^{ed} S^d \pmod{m} = r S^d \pmod{m},$$

то противник, легко сняв затемняющий множитель r , получает подпись сообщения $S^d \pmod{m}$.

Замечание 1.37. Из-за использования затемняющего множителя этот метод иногда называют *ослеплением* (*blinding*). В данном контексте ослепление является способом атаки на RSA, однако, как будет показано ниже, это полезное свойство для предоставления анонимности электронных платежей: «электронные деньги» позволяют покупать товар, не указывая конкретно покупателя.

□ Метод бесключевого чтения RSA, или циклическая атака

Рассмотрим еще один пример алгоритма такого рода — метод бесключевого чтения RSA. Этот алгоритм известен еще под названием *циклическая атака*.

Начальные условия. Противнику стали известны открытая часть ключа (m, e) и кодированное сообщение C .

Задача. Восстановить исходное сообщение S .

Решение. Противник подбирает число j , для которого выполняется следующее соотношение:

$$C^{e^j} \pmod{m} = C.$$

Таким образом, противник просто проводит j раз шифрование на открытом ключе перехваченного кодированного сообщения C . Это выглядит следующим образом:

$$(C^e)^e \dots)^e \pmod{m} = C^{e^j} \pmod{m}.$$

Найдя такое j , противник вычисляет $C^{e^{j-1}} \pmod{m}$ (т. е. $j - 1$ раз повторяет операцию шифрования) — это значение и есть исходное сообщение S . Это следует из того, что

$$C^{e^j} \pmod{m} = (C^{e^{j-1}})^e \pmod{m} = C = S^e \pmod{m}.$$

Пример 1.47. Рассмотрим следующие параметры шифра RSA¹:

$$p = 983, q = 563, m = pq = 553\,429.$$

Выберем для открытой части ключа $e = 49$, а сообщение $S = 123\,456$. Тогда кодированное (перехваченное) сообщение C имеет вид:

$$C = S^{49} \pmod{m} = 1\,603.$$

Проводим циклическую атаку (табл. 1.28).

Откуда получаем

$$S = C^{49^8} \pmod{m} = 123\,456.$$

¹ G.J. Simmons and M.J. Norris. Preliminary comment on the M.I.T. public key cryptosystem // Cryptologia, 1977.

Таблица 1.28

j	$C^{e^j} \pmod{m}$
...
7	$C^{49^7} \pmod{m} = 85\,978$
8	$C^{49^8} \pmod{m} = 123\,456$
9	$C^{49^9} \pmod{m} = 1\,603 = C$

1.3.9. Другие примеры использования RSA

□ Электронная подпись

Электронная подпись существенно отличается от привычной. Ее смысл — так зашифровать передаваемое сообщение, чтобы расшифровать его можно было только с помощью открытой части кода, принадлежащей лицу, «поставившему подпись». Таким образом, проверяется принадлежность письма предполагаемому автору.

Как следует из теоремы 1.43, операции зашифровки и расшифровки различаются только показателем степени и, следовательно, коммутируют:

$$x = (x^e)^d \pmod{m} = x^{ed} \pmod{m} = x^{de} \pmod{m} = (x^d)^e \pmod{m} = x.$$

На этом свойстве и основан прием, называемый *электронной подписью*.

1. Отправитель кодирует сообщение x своим закрытым ключом d : $y = x^d \pmod{m}$ и посыпает адресату пару $\{x, y\}$, т. е. подписанное сообщение.

2. Адресат декодирует подпись сообщения с помощью открытой части кода e и m , т. е. вычисляет $\bar{x} = y^e \pmod{m}$. Если $\bar{x} = x$, то сообщение правильное, если же $\bar{x} \neq x$, то сообщение было искажено при передаче или фальсифицировано.

□ Электронные деньги

Всюду ниже под электронными деньгами будем понимать электронные платежные средства, обеспечивающие *неотслеживаемость*. Понятие *неотслеживаемости*, по-видимому, не может быть formalизовано и будет пояснено на конкретном примере протокола.

Для работы с электронными деньгами разрабатываются специальные криптографические протоколы, которые называют *протоколами электронных платежей*. В таком протоколе задействованы три участника, ко-

торых будем называть: *банк, покупатель и продавец*. Покупатель и продавец, каждый, имеют счет в банке, и покупатель желает заплатить продавцу за товар или услугу.

В платежной системе используются три основные транзакции:

- а) снятие со счета;
- б) платеж;
- в) депозит.

В транзакции снятия со счета покупатель получает подписанную банком электронную банкноту на затребованную сумму. При этом счет покупателя уменьшается на эту сумму.

В транзакции платежа покупатель передает банкноту продавцу и указывает сумму платежа. Продавец, в свою очередь, передает эту информацию банку, который проверяет подлинность банкноты. Если банкнота подлинная, банк проверяет, не была ли она потрачена ранее. Если нет, то банк заносит банкноту в специальный регистр, зачисляет требуемую сумму на счет продавца, уведомляет продавца об этом, и, если достоинство банкноты выше, чем сумма платежа, возвращает покупателю «сдачу» (через продавца).

С помощью транзакции депозита, покупатель может положить «сдачу» на свой счет в банке.

Замечание 1.38. Транзакция депозита в данном примере не рассматривается. Подробнее с этим можно ознакомиться в [9].

Безопасность банка основывается на невозможности подделать банковскую подпись для создания фальшивой банкноты, или, более общим образом, на невозможности, получив набор подлинных электронных банкнот, подделать подпись еще хотя бы для одной банкноты. Для неотслеживаемости покупателя необходимо, чтобы банк, получив банкноту в транзакции платежа, не мог установить, кому она была выдана. То же относится и к «сдаче». Это, казалось бы, парадоксальное требование удовлетворяется с помощью схемы так называемой затемненной (слепой) подписи: в транзакции снятия со счета банк подписывает не банкноту, а некоторую «абракадабру», из которой покупатель восстанавливает подписанную банкноту. Таким образом, неотслеживаемость гарантируется тем, что банк просто не знает, что именно он подписал.

Рассмотрим простейший вариант платежной системы, в которой используется затемненная подпись, соответствующая схеме подписи RSA.

Банк выбирает открытую (m, e) и закрытую d части ключа и публикует открытую часть, а также некоторую одностороннюю функцию $f : \mathbb{Z}_m \rightarrow \mathbb{Z}_m$, назначение которой станет ясно из дальнейшего. Ключ

$\{(m, e), d\}$ используется банком исключительно для создания электронных банкнот, т. е. устанавливается соглашение о том, что электронной подписи, сгенерированной на этом ключе, соответствует электронная банкнота достоинством, скажем, в 1 сантику.

Транзакция снятия со счета. Покупатель выбирает случайное число $n \in \mathbb{Z}_m$ и вычисляет $f(n)$. Ему нужно получить подпись банка на этой банкноте, т. е. значение $f^d(n) \bmod m$. Но просто послать значение $f(n)$ банку покупатель не может, поскольку для снятия денег со счета он должен идентифицировать себя. Поэтому, если банк получает $f(n)$, он в дальнейшем всегда узнает данную банкноту и неотслеживаемость будет потеряна.

Решение проблемы состоит в использовании затемненной подписи («схема с нотариусом» — (1.68)): покупатель выбирает случайное число $r \in \mathbb{Z}_m, r \neq 0$, затем вычисляет $f(n)r^e \bmod m$ и посыпает это значение банку. Как уже отмечалось ранее, множитель r^e часто называют *затемняющим множителем*. Банк подписывает это число, т. е. вычисляет значение

$$(f(n)r^e)^d \equiv f(n)^d r \bmod m$$

и возвращает его покупателю. Покупатель делением на r «снимает» затемняющий множитель и получает подписанную банкноту $(n, f^d(n) \bmod m)$.

Транзакция платежа. Покупатель передает продавцу электронную банкноту $(n, f^d(n) \bmod m)$. В принципе, продавец может проверить подлинность любой банкноты (n, s) самостоятельно. Для этого достаточно вычислить $f(n)$ и проверить, что при этом $f(n) \equiv s^e \bmod m$. Но дело в том, что электронные банкноты, как и любую другую информацию, представленную в электронной форме, легко копировать. Поэтому нечестный покупатель может заплатить одной и той же электронной банкнотой многократно. Для предотвращения подобного злоупотребления продавец передает банкноту на проверку банку. Банк проверяет по специальному регистру, не была ли эта банкнота потрачена ранее, и если нет, то зачисляет 1 сантику на счет продавца и уведомляет его об этом.

Безопасность банка в этой системе электронных платежей основывается на вере в стойкость схемы электронной подписи RSA. Применение функции f в этой конструкции необходимо ввиду известного свойства мультипликативности схемы RSA: если s_1 и s_2 — подписи для m_1 и m_2 соответственно, то

$$s_1 s_2 \equiv m_1^d m_2^d \bmod m$$

есть подпись для $m_1 m_2$. Поэтому, если бы в системе электронных платежей использовались банкноты вида $(n, n^d \bmod m)$, то из двух подлинных банкнот всегда можно было бы изготовить третью. Неотслеживаемость клиентов в данной системе абсолютна. Все, что остается у банка от транзакции

снятия со счета, $f^d(n)r \bmod m$ — это значение, которое благодаря затемняющему множителю r представляет собой просто случайное число из \mathbb{Z}_m . Поэтому у банка нет информации о том, какую именно банкноту он выдал данному клиенту.

В рассмотренном случае банк выдает банкноты только достоинством в 1 сантим и все платежи должны быть кратны этой величине. Оказывается, можно реализовать и более гибкую систему. С подробностями можно ознакомиться в уже указанной монографии [9].

1.4. Арифметика многочленов

1.4.1. Основные операции и свойства

Определение 1.30. Алгебраическое выражение с переменной x называют *многочленом над полем K* , если оно может быть представлено в виде

$$P_n(x) = p_n x^n + p_{n-1} x^{n-1} + \cdots + p_1 x + p_0, \text{ где } p_i \in K. \quad (1.69)$$

При этом используются следующие обозначения:

- $K[x]$ — множество многочленов от переменной x над полем K ,
- $l(P_n(x))$ — старший коэффициент многочлена, $l(P_n(x)) = p_n \neq 0$,
- $\deg P_n(x)$ — степень многочлена, $\deg P_n(x) = n$.

Если в формуле (1.69) все p_i равны нулю, то такой многочлен называют *нулевым*. Степень нулевого многочлена не определена (иногда ее удобно полагать равной нулю, см. теорему 1.45).

Арифметические операции с многочленами в некотором смысле проще операций с натуральными числами, так как операции межразрядного переноса для многочленов отсутствуют.

Многочлен $P_n(x)$ степени n определяется упорядоченным набором своих $n+1$ коэффициентов: $\{p_n, p_{n-1}, \dots, p_0\}$.

Замечание 1.39.

1. Между многочленами и позиционными записями натуральных чисел есть очевидная аналогия, которая распространяется и на алгоритмы выполнения арифметических операций над многочленами.

2. В дальнейшем, когда это не вызывает разночтений, вместо $P_n(x)$ будем писать P_n или просто P .

Сложение и умножение многочлена на число осуществляется как сложение и вычитание векторов.

Для рассматриваемых в этом разделе алгоритмов введем следующие обозначения:

$$P_n(x) = p_n x^n + \cdots + p_0, \quad Q_m(x) = q_m x^m + \cdots + q_0, \quad n \geq m$$

исходные многочлены,

$$S_n(x) = s_k x^k + \cdots + s_0$$

результат работы алгоритма.

Алгоритм 1.18 (сложение двух многочленов).

```

for  $i \leftarrow 0, m$  do
     $s_i \leftarrow p_i + q_i$ 
end for
for  $i \leftarrow m + 1, n$  do
     $s_i \leftarrow p_i$ 
end for

```

Произведение многочленов P_n степени n и Q_m степени m вычисляется по формуле

$$S_{n+m} = P_n Q_m = \sum_{k=0}^{n+m} \left(\sum_{i+j=k} p_i q_j \right) x^k. \quad (1.70)$$

Замечание 1.40. Для простоты вычислений выражение в правой части (1.70) удобно переписать в виде:

$$\sum_{k=0}^m \sum_{j=0}^k p_j q_{k-j} x^k + \sum_{k=m+1}^n \sum_{j=k-m}^k p_j q_{k-j} x^k + \sum_{k=n+1}^{n+m} \sum_{j=k-m}^n p_j q_{k-j} x^k. \quad (1.71)$$

Из (1.71) очевидно, что

$$l(PQ) = l(P)l(Q), \quad \deg PQ = \deg P + \deg Q.$$

Запишем алгоритм умножения многочленов на основе равенства (1.71).

Алгоритм 1.19 (умножение многочленов).

```

for  $k \leftarrow 0, n + m$  do
     $s_k \leftarrow 0$ 
end for
for  $k \leftarrow 0, m$  do
    for  $j \leftarrow 0, k$  do
         $s_k \leftarrow s_k + p_j q_{k-j}$ 
    end for
end for
for  $k \leftarrow m + 1, n$  do
    for  $j \leftarrow k - m, k$  do
         $s_k \leftarrow s_k + p_j q_{k-j}$ 
    end for

```

```

    end for
end for
for  $k \leftarrow n+1, n+m$  do
    for  $j \leftarrow k-m, n$  do
         $s_k \leftarrow s_k + p_j q_{k-j}$ 
    end for
end for

```

При таком определении сложения и умножения многочленов очевидна следующая теорема.

Теорема 1.44. Множество многочленов $K[x]$ есть ассоциативное коммутативное кольцо с единицей.

Упражнение 1.7. Покажите, что кольцо многочленов $\mathbb{R}[x]$ не является полем.

Для многочленов в $K[x]$ справедливо представление, аналогичное (1.1) для чисел из \mathbb{Z} .

Теорема 1.45. Пусть $A(x), B(x) \in K[x]$, тогда если $B(x) \neq \text{const}$, то существуют и единственны многочлены $Q(x)$ и $R(x)$, такие, что справедливо представление

$$A(x) = B(x)Q(x) + R(x), \quad 0 \leq \deg R < \deg B. \quad (1.72)$$

Здесь полагаем, что степень нулевого многочлена равна 0. Как и ранее для чисел, многочлен $Q(x)$ называют частным от деления $A(x)$ на $B(x)$, а $R(x)$ соответственно остатком, $R(x) = \langle A(x) \rangle_{B(x)}$.

Доказательство. Существование многочленов Q и R в представлении (1.72) можно обосновать конструктивно. Построим для этого алгоритм, аналогичный делению чисел уголком.

Алгоритм 1.20 (первый алгоритм деления многочленов).

```

 $R \leftarrow A$ 
 $Q \leftarrow 0$ 
     $\triangleright$  Вычисляем обратное значение старшего коэффициента многочлена делителя  $l(B)^{-1}$ 
if  $l(B) \neq 1$  then
     $\alpha \leftarrow 1/l(B)$ 
else
     $\alpha \leftarrow 1$ 
end if
while  $\deg R \geq \deg B$  do
     $T \leftarrow l(R) \cdot \alpha \cdot x^{\deg R - \deg B}$ 
     $Q \leftarrow Q + T$ 
     $R \leftarrow R - T \cdot B$ 
end while

```

Замечание 1.41. Очевидно, что после каждого шага цикла степень R уменьшается, но равенство $A = BQ + R$ сохраняется. Когда степень R станет меньше степени B , алгоритм 1.20 свою работу заканчивает.

Единственность представления (1.72) докажем от противного. Предположим, что $A = BQ_1 + R_1 = BQ_2 + R_2$. Тогда, после вычитания равенств, получим $B(Q_1 - Q_2) = R_2 - R_1$. Степень левой части не меньше $\deg B$, если только Q_1 не совпадает с Q_2 ; степень же правой части меньше $\deg B$, так как $\deg R_1 < \deg B$ и $\deg R_2 < \deg B$. Итак, равенство будет справедливо только при $Q_1 = Q_2$ и, следовательно, $R_1 = R_2$. ■

Представление (1.72) для многочленов над конечными полями играет важную роль в криптографии и будет использоваться в дальнейшем в разделе 1.6.5. Рассмотрим несколько примеров деления многочленов с остатком в \mathbb{Z}_m для различных простых m .

Пример 1.48. Используя алгоритм 1.20, найдем представление (1.72) для многочленов $A(x) = x^5 + 2x^3 + x^2 + 2$ и $B(x) = 2x^3 + 2x^2 + x + 2$ над полем \mathbb{Z}_3 .

Для вычислений понадобятся таблицы арифметических действий для \mathbb{Z}_3 (табл. 1.29 *а* и *б*).

Таблица 1.29

а *б*

+	0	1	2	×	0	1	2
0	0	1	2	0	0	0	0
1	1	2	0	1	0	1	2
2	2	0	1	2	0	2	1

Замечание 1.42. Для сложения и вычитания в кольце вычетов \mathbb{Z}_m можно обойтись и без таблицы сложения, просто выполнив действия в \mathbb{Z} и взяв результат из наименьшей положительной системы вычетов (см. пример 1.30). Например, для \mathbb{Z}_3

$$1 - 2 = -1 = 2 \pmod{3}$$

Шаг 1. $R = A$, $Q = 0$. Вычислим предварительно $1/l(B) = 2^{-1} = 2$. Тогда

$$l(R)/l(B) = 2, T = 2x^{\deg R - \deg B} = 2x^2, Q = 2x^2, R = 2x^4 + 2.$$

Вычисление $R = R - TB$ покажем подробно:

$$\begin{array}{r} x^5 + 2x^3 + x^2 \\ x^5 + x^4 + 2x^3 + x^2 \\ \hline 2x^4 \end{array} \quad \begin{array}{r} + 2 \\ \hline \left| \begin{array}{r} 2x^3 + 2x^2 + x + 2 \\ 2x^2 \\ \hline \end{array} \right. \\ + 2 \end{array}$$

Шаг 2. $R = 2x^4 + 2$, $Q = 2x^2$, $\deg R - \deg B = 1$.

$$l(R)/l(B) = 2 \cdot 2 = 1, T = x^{\deg R - \deg B} = x, Q = 2x^2 + x, R = x^3 + 2x^2 + x + 2.$$

Вычисляем $R = R - TB$.

$$\begin{array}{r} 2x^4 \\ 2x^4 + 2x^3 + x^2 + 2x \\ \hline x^3 + 2x^2 + x + 2 \end{array} \quad \begin{array}{c} + 2 \\ \hline 2x^3 + 2x^2 + x + 2 \\ x \end{array}$$

Шаг 3. $R = x^3 + 2x^2 + x + 2$, $Q = 2x^2 + x + 2$, $\deg R - \deg B = 0$.

$$l(R)/l(B) = 1 \cdot 2 = 2, T = 2x^{\deg R - \deg B} = 2, Q = 2x^2 + x + 2, R = x^2 + 2x + 1.$$

Вычисляем $R = R - TB$.

$$\begin{array}{r} x^3 + 2x^2 + x + 2 \\ x^3 + x^2 + 2x + 1 \\ \hline x^2 + 2x + 1 \end{array} \quad \begin{array}{c} | 2x^3 + 2x^2 + x + 2 \\ 2 \end{array}$$

Окончательно получаем представление

$$x^5 + 2x^3 + x^2 + 2 = (2x^3 + 2x^2 + x + 2)(2x^2 + x + 2) + x^2 + 2x + 1.$$

Пример 1.49. Найдем представление (1.72) для многочленов $x^5 + x^4 + 1$ и $x^4 + x^2 + 1$ над полем \mathbb{Z}_2

Все действия будем выполнять в \mathbb{Z}_2 . Поэтому, нам потребуются таблицы арифметических действий (табл. 1.30 *a* и *б*).

Таблица 1.30

а *б*

+	0	1
0	0	1
1	1	0

×	0	1
0	0	0
1	0	1

Заметим, что из табл. 1.30 *a* следует, что $1 - 1 = 1$. Вычисляем предварительно $1/l(B) = 1^{-1} = 1$. Далее имеем.

Шаг 1. $R = x^5 + x^4 + 1$, $Q = 0$, $\deg R - \deg B = 1$.

$$l(R)/l(B) = 1 \cdot 1 = 1, T = x^{\deg R - \deg B} = x, Q = x.$$

Вычисляем $R = R - TB$

$$\begin{array}{r} x^5 + x^4 \\ x^5 + x^3 + x \\ \hline x^4 + x^3 + x + 1 \end{array} \quad \begin{array}{c} + 1 \\ \hline x^4 + x^2 + 1 \\ x \end{array}$$

Шаг 2. $R = x^4 + x^3 + x + 1$, $Q = x$, $\deg R - \deg B = 0$.

$$l(R)/l(B) = 1 \cdot 1 = 1, T = x^{\deg R - \deg B} = 1, Q = x + 1.$$

Вычисляем $R = R - TB$

$$\begin{array}{r} x^4 + x^3 + x + 1 \\ x^4 + x^2 + 1 \\ \hline x^3 + x^2 + x \end{array}$$

Окончательно получаем представление

$$x^5 + x^4 + 1 = (x^4 + x^2 + 1)(x + 1) + x^3 + x^2 + x.$$

Пример 1.50. Найдем представление (1.72) для многочленов $4x^5 + 4x^4 + x^3 + 4x^2 + x + 1$ и $4x^3 + 2x^2 + 2x + 2$ над полем \mathbb{Z}_5 .

Используем таблицы арифметических действий (табл. 1.31 *a* и *b*).

Таблица 1.31

	<i>a</i>						<i>b</i>					
+	0	1	2	3	4		×	0	1	2	3	4
0	0	1	2	3	4		0	0	0	0	0	0
1	1	2	3	4	0		1	0	1	2	3	4
2	2	3	4	0	1		2	0	2	4	1	3
3	3	4	0	1	2		3	0	3	1	4	2
4	4	0	1	2	3		4	0	4	3	2	1

Предварительно вычисляем $1/l(B) = 4^{-1} = 4$. Далее имеем.

Шаг 1. $R = 4x^5 + 4x^4 + x^3 + 4x^2 + x + 1$, $Q = 0$, $\deg R - \deg B = 2$.

$$l(R)/l(B) = 4 \cdot 4 = 1, T = x^{\deg R - \deg B} = x^2, Q = x^2.$$

Вычисляем $R = R - TB$

$$\begin{array}{r} 4x^5 + 4x^4 + x^3 + 4x^2 + x + 1 \\ 4x^5 + 2x^4 + 2x^3 + 2x^2 \\ \hline 2x^4 + 4x^3 + 2x^2 + x + 1 \end{array}$$

Шаг 2. $R = 2x^4 + 4x^3 + 2x^2 + x + 1$, $Q = x^2$, $\deg R - \deg B = 1$.

$$l(R)/l(B) = 2 \cdot 4 = 3, T = x^{\deg R - \deg B} = 3x, Q = x^2 + 3x.$$

Вычисляем $R = R - TB$

$$\begin{array}{r} 2x^4 + 4x^3 + 2x^2 + x + 1 \\ 2x^4 + x^3 + x^2 + x \\ \hline 3x^3 + x^2 + 1 \end{array}$$

Шаг 3. $R = 3x^3 + x^2 + 1$, $Q = x^2 + 3x$, $\deg R - \deg B = 0$.

$$l(R)/l(B) = 3 \cdot 4 = 2, T = x^{\deg R - \deg B} = 2, Q = x^2 + 3x + 2.$$

Вычисляем $R = R - TB$

$$\begin{array}{r} 3x^3 + x^2 + 1 \\ 3x^3 + 4x^2 + 4x + 4 \\ \hline 2x^2 + x + 2 \end{array} \left| \begin{array}{r} 4x^3 + 2x^2 + 2x + 2 \\ 3x \\ \hline \end{array} \right.$$

Окончательно получаем представление

$$4x^5 + 4x^4 + x^3 + 4x^2 + x + 1 = (4x^3 + 2x^2 + 2x + 2)(x^2 + 3x + 2) + 2x^2 + x + 2.$$

Запишем алгоритм деления многочленов в другом виде.

Для многочленов $A(x) = a_n x^n + \dots + a_0$ и $B(x) = b_m x^m + \dots + b_0$ требуется получить представление:

$$A(x) = B(x)(q_{n-m} x^{n-m} + \dots + q_0) + r_{m-1} x^{m-1} + \dots + r_0. \quad (1.73)$$

Замечание 1.43. В отличие от алгоритма 1.20 коэффициенты многочлена остатка r_i будем хранить в коэффициентах многочлена делимого a_i .

Алгоритм 1.21 (второй алгоритм деления многочленов).

```

 $k \leftarrow \deg A - \deg B = n - m$ 
     $\triangleright$  Вычисляем обратное значение старшего коэффициента многочлена делителя  $l(B)^{-1}$ 
 $\text{if } b_m \neq 1 \text{ then}$ 
     $\alpha \leftarrow 1/b_m$ 
 $\text{else}$ 
     $\alpha \leftarrow 1$ 
 $\text{end if}$ 
 $\text{while } k \geq 0 \text{ do}$ 
     $q_k \leftarrow \alpha \cdot a_{m+k}$ 
     $\text{for } j \leftarrow m + k - 1, k \text{ do}$ 
         $a_j \leftarrow a_j - q_k \cdot b_{j-k}$ 
     $\text{end for}$ 
 $\text{end while}$ 
     $\triangleright$  пересчитываем коэффициенты  $a_j$ 

```

Пример 1.51. Используя алгоритм 1.21, разделим с остатком над полем \mathbb{R} многочлен $x^4 + x^3 - 3x^2 - 4x - 1$ на $x^3 + x^2 - x - 1$. Протокол работы алгоритма приведен в табл. 1.32. Имеем $n = 4$, $m = 3$. Тогда

Таким образом получаем представление:

$$x^4 + x^3 - 3x^2 - 4x - 1 = (x^3 + x^2 - x - 1)(x + 0) + (-2x^2 - 3x - 1)$$

Таблица 1.32

k	q_k	Формула пересчета	a_4	a_3	a_2	a_1	a_0
			1	1	-3	-4	-1
1	$q_1 = a_4/b_3 = 1$	$a_j = a_j - b_{j-1}, j = 3, 2, 1$		0	-2	-3	
0	$q_0 = a_3/b_3 = 0$	$a_j = a_j, j = 2, 1, 0$			-2	-3	-1

1.4.2. Схема Горнера

Рассмотрим результат деления многочлена $P(x) \in K[x]$ на многочлен $Q(x) = x - \alpha \in K[x]$:

$$P(x) = (x - \alpha)Q(x) + R(x), \text{ где } \deg R(x) < \deg(x - \alpha) = 1,$$

т. е. $\deg R(x) = 0$ и $R(x) = \text{const} = r$.

Замечание 1.44. $\deg Q(x) = \deg P(x) - 1$.

Теорема 1.46 (теорема Безу). Остаток от деления многочлена $P(x)$ на $x - \alpha$ равен значению многочлена $P(\alpha)$.

Доказательство. $P(x) = (x - \alpha)Q(x) + r$, следовательно, $P(\alpha) = r$. ■

Следствие 1.18. Многочлен $x - \alpha$ является делителем многочлена $P(x)$ тогда и только тогда, когда $P(\alpha) = 0$.

Замечание 1.45. Теорема Безу связывает два разных взгляда на многочлены — алгебраический (деление многочленов, остаток) и аналитический (вычисление значений многочлена как функции r).

Согласно теореме Безу в основу вычисления многочлена может быть положен алгоритм деления многочлена на двучлен. Этот алгоритм называют *схемой Горнера*. Рассмотрим его подробнее, как частный случай алгоритма 1.21, когда $Q(x) = x - \alpha$.

Имеем $t = 1$, $b_1 = 1$, $b_0 = -\alpha$. Тогда получаем следующую схему

$$q_k = a_{k+1}, \quad a_k = a_k - a_{k+1}b_0 = a_k + a_{k+1}\alpha, \quad k = n-1, \dots, 0. \quad (1.74)$$

При этом значение остатка находится в a_0 .

Схема (1.74) иногда может быть неудобна, так как в ней изменяются коэффициенты исходного многочлена $A_n(x)$. Модифицируем схему (1.74), так, чтобы их сохранить. В представлении (1.73) коэффициенты частного будем хранить в r_{n-1}, \dots, r_0 , а остатка в r_{-1} . Тогда схема (1.74) преобразуется к виду

$$r_{n-1} = a_n, \quad r_{k-1} = a_k - a_{k+1}b_0 = a_k - r_k b_0 = a_k + r_k \alpha, \quad k = n-1, \dots, 0. \quad (1.75)$$

Обычно схемой Горнера называют именно схему (1.75).

Замечание 1.46. Схема Горнера эквивалентна представлению многочлена $A(x)$ в виде

$$A(x) = a_n x^n + \cdots + a_0 = (\cdots ((a_n x + a_{n-1}) x + a_{n-2}) x + \cdots + a_1) x + a_0$$

Замечание 1.47. При ручных вычислениях промежуточные результаты схемы Горнера удобно помещать в следующей таблице:

a_n	a_{n-1}	a_{n-2}	\cdots	a_1	a_0
r_{n-1}	r_{n-2}	r_{n-3}	\cdots	r_0	$r_{-1} = s_0 = p(\alpha)$

Каждое следующее вычисление использует только значение α и числа, помещенные в соседних (левой и верхней) клетках таблицы:

\cdots	a_k	\cdots
r_k	$r_{k-1} = a_k + \alpha r_k$	\cdots

Следствие 1.19. Схему Горнера можно применять многократно: к частному $Q(x)$ от деления $A(x)$ на $x - \alpha$, затем к получившемуся частному $Q_1(x)$ и т. д.

В результате получим коэффициенты разложения многочлена $P(x)$ по степеням $x - \alpha$ (от младших степеней к старшим):

$$\begin{aligned} A(x) &= (x - \alpha)Q(x) + s_0 = (x - \alpha)[(x - \alpha)Q_1(x) + s_1] + s_0 = (x - \alpha)^2 Q_1(x) + \\ &+ s_1(x - \alpha) + s_0 = \cdots = s_n(x - \alpha)^n + s_{n-1}(x - \alpha)^{n-1} + \cdots + s_1(x - \alpha) + s_0. \end{aligned}$$

Замечание 1.48. При выполнении вычислений вручную протокол вычислений удобно вести в следующей «треугольной» таблице:

a_n	a_{n-1}	a_{n-2}	\cdots	a_1	a_0
r_{n-1}	r_{n-2}	\cdots	\cdots	r_0	$r_{-1} = s_0$
c_{n-2}	\cdots	\cdots	\cdots	$c_{-1} = s_1$	
\cdots	\cdots	\cdots			
\cdots	\cdots				
\cdots					
s_n					

Таблица заполняется слева направо и сверху вниз по правилу, указанному ранее.

Пример 1.52 (схема Горнера). Выполним деление с остатком многочлена $x^4 + 2x^3 - 3x^2 + x - 5$ на двучлен $x - 2$.

Согласно схеме Горнера имеем (табл. 1.33):

$$x^4 + 2x^3 - 3x^2 + x - 5 = (x^3 + 4x^2 + 5x + 11)(x - 2) + 17.$$

Используя обобщенную схему Горнера, найдем разложение многочлена $x^4 + 2x^3 - 3x^2 + x - 5$ по степеням $x - 2$. Последовательно применяя схему Горнера, получаем (табл. 1.34):

$$x^4 + 2x^3 - 3x^2 + x - 5 = (x - 2)^4 + 10(x - 2)^3 + 33(x - 2)^2 + 45(x - 2) + 17.$$

Упражнение 1.8. Напишите алгоритм вычисления коэффициентов многочлена $S(x) : s_0, s_1, \dots, s_n$, полученного из многочлена $A(x)$ с коэффициентами a_n, a_{n-1}, \dots, a_0 разложением по степеням $x - \alpha$.

Таблица 1.33

a_4	a_3	a_2	a_1	a_0
1	2	-3	1	-5
1	4	5	11	17
r_3	r_2	r_1	r_0	s_0

Таблица 1.34

1	2	-3	1	-5
1	4	5	11	17
1	6	17	45	
1	8	33		
1	10			
1				

1.4.3. Алгоритм Евклида для многочленов. Линейное представление НОД

Алгоритм Евклида для многочленов не отличается от алгоритма деления чисел при соответствующей интерпретации входящих в него переменных и операций.

Сравним определения деления нацело, лежащие в основе алгоритма Евклида, для чисел и многочленов.

$a = b q + r$	$A(x) = B(x) Q(x) + R(x)$
$0 \leq r < b $	$0 \leq \deg R(x) < \deg B(x)$
q — частное, r — остаток	$Q(x)$ — частное, $R(x)$ — остаток

Видно, что определения отличаются лишь характеристиками (абсолютным значением числа, степенью многочлена), используемыми для нахождения остатка и частного.

Несмотря на то что частное и остаток при делении как целых чисел, так и многочленов, определяются однозначно, выбор наибольшего общего делителя многочленов неоднозначен. Это связано с тем, что НОД определяется как общий делитель наибольшей степени, а таких многочленов может быть бесконечно много. Например, если $G(x) = D(A(x), B(x))$ для $A, B \in \mathbb{R}[x]$, то $\alpha G(x)$ при $\alpha \neq 0$ также будет наибольшим общим делителем этих многочленов. Можно было бы среди всех наибольших общих делителей выбрать многочлен с единичным старшим коэффициентом, тогда НОД будет определен единственным образом. Однако для вычислений это не всегда удобно. Например, можно рассмотреть модификацию классического алгоритма Евклида с домножением промежуточных результатов на ненулевую константу, что позволит, находя НОД целочисленных многочленов, избежать дробных коэффициентов.

Пример 1.53. Найдем $D(2x^2 + x - 1, 3x^2 + 2x - 1)$.

$$1) D(2x^2 + x - 1, 3x^2 + 2x - 1) = D(6x^2 + 3x + 3, 3x^2 + 2x - 1).$$

$$2) D(6x^2 + 3x - 3, 3x^2 + 2x - 1) = D(-x - 1, 3x^2 + 2x - 1).$$

$$3) D(-x - 1, 3x^2 + 2x - 1) = D(-x - 1, 0).$$

$$4) D(-x - 1, 0) = D(x + 1, 0) = x + 1.$$

Определение 1.31. Многочлены P и G называют *взаимно простыми*, если они не имеют отличных от константы общих делителей.

Для нахождения линейного представления наибольшего общего делителя многочленов можно использовать аналог расширенного алгоритма Евклида для чисел (алгоритм 1.15). Рассмотрим примеры.

Пример 1.54. С помощью расширенного алгоритма Евклида найдем $D(x^4 + x^3 - 3x^2 - 4x - 1, x^3 + x^2 - x - 1)$ и его линейное представление над полем \mathbb{R} . Протокол работы алгоритма запишем в виде таблицы (табл. 1.35):

Таблица 1.35

	$x^4 + x^3 - 3x^2 - 4x - 1$	$x^3 + x^2 - x - 1$	$-2x^2 - 3x - 1$
$q(x)$			x
$u(x)$	1	0	1
$v(x)$	0	1	$-x$
	$-3/4(x + 1)$	0	
$q(x)$	$1/4(-2x + 1)$	$4/3(2x + 1)$	
$u(x)$	$-1/4(-2x + 1)$	$-4/3(x^2 - 1)$	
$v(x)$	$1/4(-2x^2 + x + 4)$	$4/3(x^3 - 3x - 1)$	

Следовательно,

$$D(x^4 + x^3 - 3x^2 - 4x - 1, x^3 + x^2 - x - 1) = -\frac{3}{4}(x + 1).$$

Произведя нормировку (умножив на $-4/3$), получаем

$$D(x^4 + x^3 - 3x^2 - 4x - 1, x^3 + x^2 - x - 1) = x + 1.$$

Таким образом:

$$(x^4 + x^3 - 3x^2 - 4x - 1) \left(\frac{x}{2} - \frac{1}{4} \right) + (x^3 + x^2 - x - 1) \left(-\frac{x^2}{2} + \frac{x}{4} + 1 \right) = -\frac{3}{4}(x + 1).$$

Произведя нормировку (умножив на $-4/3$), получаем

$$(x^4 + x^3 - 3x^2 - 4x - 1) \left(\frac{-2x + 1}{3} \right) + (x^3 + x^2 - x - 1) \left(\frac{2x^2 - x - 4}{3} \right) = x + 1.$$

Пример 1.55. Найти $D(x^5+x^4+1, x^4+x^2+1)$ и его линейное представление над полем \mathbb{Z}_2 . Запишем протокол работы расширенного алгоритма Евклида (табл 1.36). Тогда:

Таблица 1.36

	$x^5 + x^4 + 1$	$x^4 + x^2 + 1$	$x^3 + x^2 + x$	$x^2 + x + 1$	0
$q(x)$			$x + 1$	$x + 1$	x
$u(x)$	1	0	1	$x + 1$	$x^2 + x + 1$
$v(x)$	0	1	$x + 1$	x^2	$x^3 + x + 1$

$$(x^5 + x^4 + 1)(x + 1) + (x^4 + x^2 + 1)(x^2) = x^2 + x + 1.$$

1.4.4. Интерполяционная формула Лагранжа

Аналогично алгоритму Евклида, на многочлены переносится и китайская теорема об остатках.

Теорема 1.47. Рассмотрим систему сравнений

$$P(x) \equiv R_1(x) \pmod{M_1(x)}, \dots, P(x) \equiv R_n(x) \pmod{M_n(x)}.$$

Если многочлены $M_1(x), \dots, M_n(x)$ попарно взаимно просты, то

$$P(x) \equiv \sum_{i=1}^n C_i(x) D_i(x) R_i(x) \pmod{M(x)}, \text{ где}$$

$$M(x) = \prod_{i=1}^n M_i(x), \quad C_i(x) = \frac{M(x)}{M_i(x)}, \quad D_i(x) C_i(x) \equiv 1 \pmod{M_i(x)}.$$

Замечание 1.49. Доказательство теоремы идентично доказательству аналогичной теоремы для чисел (см. теорему 1.42).

Решение сравнения $D_i(x)C_i(x) \equiv 1 \pmod{M_i(x)}$ осуществляется применением расширенного алгоритма Евклида к взаимно простым многочленам $C_i(x), M_i(x)$:

$$C_i(x)U(x) + M_i(x)V(x) = \text{const} \neq 0, \text{ тогда } \frac{C_i(x)U(x)}{\text{const}} \equiv 1 \pmod{M_i(x)}, \text{ откуда}$$

$$D_i(x) \equiv \frac{U(x)}{\text{const}} \pmod{M_i(x)}.$$

Теорема 1.48 (интерполяционная формула Лагранжа). Пусть $\deg P = n$ и известны значения многочлена $P(x)$ в $n+1$ различных точках:

$$P(x_0) = y_0, \quad P(x_1) = y_1, \dots, \quad P(x_n) = y_n,$$

тогда

$$P(x) = \sum_{i=0}^n \frac{(x - x_0) \cdots (x - x_{i-1})(x - x_{i+1}) \cdots (x - x_n)}{(x_i - x_0) \cdots (x_i - x_{i-1})(x_i - x_{i+1}) \cdots (x_i - x_n)} y_i. \quad (1.76)$$

Доказательство. Интерпретируем задачу в терминах арифметики многочленов. По теореме Безу $P(x_i) = y_i$ тогда и только тогда, когда $P(x) \equiv y_i \pmod{(x - x_i)}$. Многочлены $M_i(x) = x - x_i$ попарно взаимно просты, и $P(x)$ можно найти применением китайской теоремы об остатках:

$$R_i(x) = y_i; \quad M(x) = \prod_{k=0}^n (x - x_k); \quad C_i(x) = \frac{M(x)}{M_i(x)} = \prod_{k \neq i} (x - x_k),$$

тогда $D_i(x)$ удовлетворяет уравнению

$$C_i(x)D_i(x) + M_i(x)B(x) = 1, \quad \deg D_i < \deg M_i = \deg(x - x_i) = 1, \quad (1.77)$$

откуда $\deg D_i = 0$, т. е. $D_i(x) = \text{const}$. Подставив в (1.77) $x = x_i$, получим

$$C_i(x_i)D_i + M_i(x_i)B(x_i) = 1,$$

но $M_i(x_i) = x_i - x_i = 0$, следовательно, $D_i = \frac{1}{C_i(x_i)}$. Отсюда по китайской теореме об остатках

$$P(x) \equiv \sum_{i=0}^n C_i(x)D_i(x)R_i(x) = \sum_{i=0}^n \frac{C_i(x)}{C_i(x_i)} y_i = \sum_{i=0}^n y_i \prod_{i \neq k} \left(\frac{x - x_i}{x_i - x_k} \right). \quad \blacksquare$$

Пример 1.56. По формуле Лагранжа (1.76) найти интерполяционный многочлен $P(x)$, удовлетворяющий условиям:

$$P(-2) = 23, \quad P(1) = -1, \quad P(3) = 13.$$

Имеем

$$\begin{aligned} P(x) &= 23 \frac{(x - 1)(x - 3)}{(-2 - 1)(-2 - 3)} - \frac{(x + 2)(x - 3)}{(1 + 2)(1 - 3)} + 13 \frac{(x + 2)(x - 1)}{(3 + 2)(3 - 1)} = \\ &= \frac{23}{15}(x^2 - 4x + 3) + \frac{1}{6}(x^2 - x - 6) + \frac{13}{10}(x^2 + x - 2) = \\ &= 3x^2 - 5x + 1. \end{aligned}$$

1.4.5. Разложение многочлена на свободные от квадратов множители

Определение 1.32. Многочлен $P(x)$ из кольца $K[x]$ называют *неприводимым* (*примитивным*) над полем K , если не существует нетривиальных, т. е. отличных от константы, многочленов $P_1(x)$ и $P_2(x)$ из $K[x]$ таких, что $P(x) = P_1(x)P_2(x)$.

Пример 1.57. Многочлены $P(x) = x^2 + 1$, $Q(x) = x^2 + x + 2$ являются неприводимыми над полем \mathbb{R} , многочлены $T(x) = x^3 + 3x^2 + 9x + 6$ и $S(x) = x^4 + 2x^3 - 8x^2 - 2x + 3$ неприводимы над полем \mathbb{Q} (см. теорему 1.51).

Аналогично определению 1.6 для чисел можно ввести соответствующее определение для многочленов.

Определение 1.33. Многочлен $P(x)$ называется *свободным от квадратов*, если не существует многочлена $Q(x)$, $\deg Q > 0$ такого, что $Q^2(x) | P(x)$.

Определение 1.34. Представление многочлена $P(x) \in K[x]$ в виде

$$P(x) = P_1(x)P_2^2(x)P_3^3(x) \cdots P_n^n(x), \quad P_1, P_2, \dots, P_n \in K[x],$$

где $P_i(x)$ — произведение различных свободных от квадратов многочленов, причем P_i взаимно прост с P_j ($j \neq i$), называют разложением многочлена $P(x)$ на *свободные от квадратов множители* P_1, P_2, \dots, P_n .

Алгоритм разложения на свободные от квадратов множители основан на использовании алгоритма Евклида.

В данном подразделе будем рассматривать многочлены над полем вещественных чисел.

Лемма 1.9. Если $P = Q^k M$, где $P, Q, M \in \mathbb{R}[x]$, а Q — неприводим в $\mathbb{R}[x]$ и взаимно прост с M , то производная P' может быть записана в виде $Q^{k-1} N$, где Q и N — взаимно просты.

Доказательство.

$$P' = (Q^k M)' = k Q^{k-1} Q' M + Q^k M' = Q^{k-1} (k M Q' + Q M').$$

Пусть $N = k M Q' + Q M'$.

Докажем, что Q и N взаимно просты. Так как Q неприводим в $\mathbb{R}[x]$, то взаимная простота Q и N равносильна тому, что N не делится на Q . Если бы многочлен N делился на Q , то на Q делился бы и многочлен $M Q'$. Но Q' на Q не делится, так как $\deg Q' < \deg Q$, поэтому на Q должен делиться многочлен M , что противоречит условию. ■

Следствие 1.20. Если $P = Q_1^{k_1} \cdots Q_n^{k_n}$, где Q_1, \dots, Q_n — неприводимы в $\mathbb{R}[x]$, то $P' = Q_1^{k_1-1} \cdots Q_n^{k_n-1} Q$, где Q — взаимно прост с Q_1, \dots, Q_n .

Рассмотрим алгоритм разложения многочлена на свободные от квадратов множители.

Алгоритм 1.22.

```

 $j \leftarrow 1$ 
while  $P \neq \text{const}$  do
     $R \leftarrow D(P, P')$                                  $\triangleright P = Q_1^{k_1} \cdots Q_n^{k_n}$ 
     $S \leftarrow \frac{P}{R}$                                  $\triangleright R = Q_1^{k_1-1} \cdots Q_n^{k_n-1}$ 
     $M \leftarrow D(S, P')$                              $\triangleright S = Q_1 \cdots Q_n$ 
     $P_j \leftarrow \frac{S}{M}$                                  $\triangleright M = Q_{i_1} \cdots Q_{i_s} \mid k_{i_p} > 1$ 
     $j \leftarrow j + 1$                                  $\triangleright P_j = Q_{i_1} \cdots Q_{i_s} \mid k_{i_p} = 1$ 
     $\triangleright$  повторяем алгоритм для многочлена, у которого степени неприводимых множителей уменьшены на 1
     $P \leftarrow R$ 
end while

```

1.4.6. Разложение многочленов на неприводимые множители

Неприводимые многочлены являются аналогом простых чисел для многочленов. Возможность разложения многочлена на множители зависит от поля, над которым рассматривается многочлен, т. е., какому полю принадлежат коэффициенты.

Пример 1.58.

1) Многочлен $P(x) = x^2 - 2$ является неприводимым над \mathbb{Q} — полем рациональных чисел, но приводимым над \mathbb{R} — полем вещественных чисел:

$$x^2 - 2 = (x - \sqrt{2})(x + \sqrt{2}).$$

2) Многочлен $P(x) = x^4 + 1$ также является неприводимым над \mathbb{Q} — полем рациональных чисел, но приводимым над \mathbb{R} — полем вещественных чисел:

$$x^4 + 1 = x^4 + 2x^2 + 1 - 2x^2 = (x^2 + 1)^2 - (\sqrt{2}x)^2 = (x^2 - \sqrt{2}x + 1)(x^2 + \sqrt{2}x + 1).$$

3) Многочлен $P(x) = x^2 + 1$ является неприводимым над \mathbb{R} — полем вещественных чисел, но приводимым над \mathbb{C} — полем комплексных чисел:

$$x^2 + 1 = (x + i)(x - i).$$

4) Многочлен $P(x) = x^2 + 1$ также приводим над полем вычетов по модулю два \mathbb{Z}_2 :

$$x^2 + 1 = (x + 1)(x + 1).$$

Очевидно, многочлены вида $x - a$ неразложимы над любым полем. Это простейшие примитивные многочлены — линейные многочлены или двучлены.

□ Приводимость над \mathbb{Q}

Наибольший интерес в рамках курса дискретной математики представляют многочлены с целыми и рациональными коэффициентами. Разложимость многочлена с рациональными коэффициентами над полем рациональных чисел \mathbb{Q} легко сводится к разложимости многочленов с целыми коэффициентами над \mathbb{Q} вынесением за скобки наибольшего общего знаменателя всех дробных коэффициентов многочлена.

Рассмотрим сначала вопрос о возможности выделения у многочлена линейного множителя.

По теореме Безу (см. теорему 1.46) $P(x) = (x - a)P_1(x)$ тогда и только тогда, когда $P(a) = 0$.

Возьмем многочлен с целыми коэффициентами

$$P(x) = a_n x^n + \cdots + a_1 x + a_0.$$

Найдем условия, при которых рациональное число p/q будет его корнем. Положим, что дробь p/q несократима и подставим ее в многочлен $P(x)$:

$$a_n \left(\frac{p}{q} \right)^n + \cdots + a_1 \frac{p}{q} + a_0 = 0 \iff a_n p^n + \cdots + a_0 q^n = 0.$$

Правая часть (равная нулю) очевидно делится на p . По условию теоремы все члены левой части, кроме последнего, делятся на p . Тогда и последнее слагаемое должно делиться на p , но, так как q на p не делится, то свободный член a_0 делится на p . Аналогично доказываем, что старший член a_n многочлена P делится на q .

Таким образом доказана следующая теорема.

Теорема 1.49. Для существования линейного множителя $(x - p/q)$ у многочлена P с целыми коэффициентами необходимо, чтобы свободный член многочлена делился на p , а коэффициент старшего члена — на q .

Пример 1.59. Многочлен $P(x) = x^5 + 5x^3 + x^2 + 6x + 2$ нельзя разложить над полем \mathbb{Q} так, чтобы один из множителей был линейным. Действительно, по теореме в качестве возможных множителей достаточно проверить четыре: $x - 2, x + 2, x - 1$ и $x + 1$, но ни один из них не подходит. Заметим, что этот многочлен приводим над \mathbb{Q} :

$$P(x) = (x^3 + 3x + 1)(x^2 + 2).$$

Представляет интерес следующая теорема.

Теорема 1.50. Если многочлен с целыми коэффициентами раскладывается на множители над \mathbb{Q} , то он может быть разложен на множители с целыми коэффициентами.

Пример 1.60. Многочлен $6x^2 - x - 2$ раскладывается на множители

$$6x^2 - x - 2 = 6 \left(x + \frac{1}{2} \right) \left(x - \frac{2}{3} \right).$$

Это разложение можно переписать так, что коэффициенты многочленов-множителей будут целыми:

$$6x^2 - x - 2 = (2x + 1)(3x - 2).$$

Задача 1.1. Докажите теорему 1.50 самостоятельно.

Рассмотрим критерий неприводимости многочлена над \mathbb{Q} .

Теорема 1.51 (критерий Эйзенштейна). Если все коэффициенты многочлена с целыми коэффициентами, кроме коэффициента старшего члена, делятся на простое число p , а свободный член не делится на p^2 , то многочлен неприводим над \mathbb{Q} .

Доказательство. От противного. Пусть имеется требуемое разложение, в котором $k, m \geq 1$. По теореме 1.50 коэффициенты b_i и c_j можно считать целыми:

$$a_n x^n + \cdots + a_1 x + a_0 = (b_k x^k + \cdots + b_1 x + b_0)(c_m x^m + \cdots + c_1 x + c_0),$$

тогда по формуле умножения многочленов (1.70) получаем

$$\begin{aligned} a_0 &= b_0 c_0, \\ a_1 &= b_1 c_0 + b_0 c_1, \\ a_2 &= b_2 c_0 + b_1 c_1 + b_0 c_2, \\ &\dots, \\ a_i &= b_i c_0 + \cdots + b_0 c_i, \\ &\dots \end{aligned}$$

Из первого равенства следует, что ровно один из множителей b_0, c_0 делится на p (если бы делились оба, то их произведение делилось бы на p^2). Пусть для определенности это будет b_0 . Тогда из второго равенства получим, что b_1 делится на p и т. д. Дойдя до равенства с левой частью a_k (или a_m , в зависимости от того, что будет больше, k или m , но в любом случае $k, m < n$), получим, что все коэффициенты b_i делятся на p , а значит p можно вынести за скобку. Последнее означает, что все коэффициенты исходного многочлена делятся на p , а это противоречит условию. ■

Пример 1.61. Многочлен $x^5 + 2x^4 + 4x^3 + 6x + 2$ неразложим над \mathbb{Q} по критерию Эйзенштейна для $p = 2$.

□ **Связь неприводимости многочленов с целыми коэффициентами над \mathbb{Z}_p и \mathbb{Q}**

Теорема 1.52. Если многочлен с целыми коэффициентами и со старшим членом, не кратным простому числу p , приводим над полем \mathbb{Q} , то он приводим и над полем \mathbb{Z}_p .

Доказательство. Рассмотрим разложение многочлена над полем \mathbb{Q} . По теореме 1.50 его можно преобразовать в разложение с целыми коэффициентами. Перейдем в полученном равенстве к модульной арифметике, заменив все коэффициенты остатками от деления на p . Справедливость равенства при этом не пострадает. Осталось объяснить, почему степени множителей сохраняются (иначе один из многочленов-множителей мог бы превратиться в константу, и разложения многочлена на нетривиальные множители не получилось бы).

Исчезновение старшего члена множителя означало бы, что его коэффициент делится нацело на p . Но тогда и старший коэффициент исходного многочлена делился бы на p , что противоречит условию. ■

Пример 1.62. Рассмотрим разложение из примера 1.59

$$x^5 + 5x^3 + x^2 + 6x + 2 = (x^3 + 3x + 1)(x^2 + 2).$$

Рассмотрим это равенство последовательно как равенство в \mathbb{Z}_2 , \mathbb{Z}_3 , \mathbb{Z}_5 . Получим равенства, справедливость которых легко проверить умножением:

$$\begin{aligned} \text{в } \mathbb{Z}_2: & x^5 + x^3 + x^2 = (x^3 + x + 1)x^2, \\ \text{в } \mathbb{Z}_3: & x^5 + 2x^3 + x^2 + 2 = (x^3 + 1)(x^2 + 2), \\ \text{в } \mathbb{Z}_5: & x^5 + x^2 + x + 2 = (x^3 + 3x + 1)(x^2 + 2). \end{aligned}$$

Наибольший интерес для применения теоремы 1.52 имеет следующее ее следствие.

Следствие 1.21. Если старший член многочлена с целыми коэффициентами не делится на простое число p , и если многочлен неприводим как многочлен над \mathbb{Z}_p , то он неприводим над \mathbb{Q} .

Доказательство. От противного. Если бы многочлен был приводим над \mathbb{Q} , то по теореме 1.52 он был бы приводим и над \mathbb{Z}_p , что не так. ■

Полученное следствие дает простой и достаточно мощный метод проверки многочлена на неприводимость над \mathbb{Q} . Надо последовательно рассматривать его как многочлен над \mathbb{Z}_p для последовательно возрастающих простых чисел p . Неприводимые многочлены над полем вычетов, степень которых ограничена сверху, можно перечислить явно (перебором и проверкой неделимости на все многочлены меньшей степени). Если таким образом

исходный многочлен будет сведен к неприводимому, то и неприводимость исходного доказана.

Теорема 1.52 дает только достаточные условия. Поэтому, если исходный многочлен оказывается приводимым над \mathbb{Z}_p для достаточно большого числа простых чисел p , то это еще не означает, что он не может оказаться приводимым над \mathbb{Q} .

Пример 1.63. Рассмотрим все неприводимые над \mathbb{Z}_2 нетривиальные многочлены степени не выше двух:

$$\begin{aligned} P_1(x) &= x, \\ P_2(x) &= x + 1 (x - 1, -x - 1, -x + 1 — это тот же самый двучлен), \\ P_3(x) &= x^2 + 1. \end{aligned}$$

Пример 1.64. Является ли многочлен $x^3 + 7x^2 + 4x + 3$ неприводимым над \mathbb{Q} ?

Рассмотрим этот многочлен как многочлен над \mathbb{Z}_2 . Получим многочлен $x^3 + x^2 + 1$, который является неприводимым, так как при делении на x и на $x + 1$ дает 1 (не делится на P_1 и P_2 из примера 1.63). Проверка делимости на $P_3(x) = x^2 + 1$ не требуется, так как в случае делимости был бы получен другой делитель с меньшей степенью. Из решения следует, что и любой многочлен вида

$$(2n + 1)x^3 + (2m + 1)x^2 + 2kx + 2s + 1$$

будет неприводимым над \mathbb{Q} для любых целых неотрицательных n, m, k, s .

1.5. Базисы Гребнера

1.5.1. Задачи на доказательство и многочлены

Для многочленов одной переменной расширенный алгоритм Евклида позволяет решить задачу, можно ли выразить данный многочлен $P(x)$ как линейную комбинацию заданных многочленов $P_i(x)$ над тем же полем, коэффициентами которой являются многочлены над тем же полем:

$$P(x) = Q_1(x)P_1(x) + Q_2(x)P_2(x) + \cdots + Q_n(x)P_n(x),$$

где $P(x), P_1(x), \dots, P_n(x), Q_1(x), \dots, Q_n(x) \in K[x]$.

Можно ли решить аналогичную задачу для многочленов от нескольких переменных?

Если бы такой алгоритм существовал, то некоторые классы задач на доказательство могли бы быть решены вычислением по алгоритму.

Например, в геометрических задачах на доказательство, как условия, так и вывод можно описать равенствами нулю многочленов от нескольких переменных. Поэтому, если бы удалось выразить подобным образом «многочлен вывода» через «многочлены условий», то равенство нулю последних влекло бы равенство нулю первого, т. е., из справедливости условий следовала бы справедливость утверждения теоремы.

Пример 1.65. Доказать, что, если из заданной точки M провести прямую, пересекающую окружность, то произведение расстояний от этой точки до точек пересечения A и B прямой с окружностью будет постоянным: $AM \cdot BM = \text{const}$ (рис. 1.4).

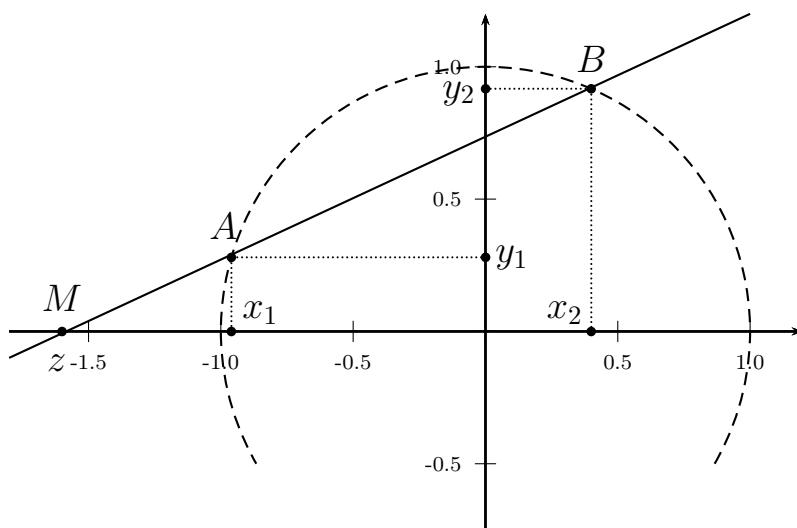


Рис. 1.4

Введем координаты с началом в центре окружности (см. рис. 1.4), тогда условия можно записать в виде равенства нулю трех многочленов. Радиус окружности примем за 1, отметим точку M на оси абсцисс с координатой z . Обозначив координаты точки A как (x_1, y_1) , точки B — (x_2, y_2) , получим условия в виде равенства нулю таких многочленов:

$$P_1(x_1, y_1) = 0 \Leftrightarrow x_1^2 + y_1^2 - 1 = 0 \quad \text{точка } A \text{ лежит на окружности};$$

$$P_2(x_2, y_2) = 0 \Leftrightarrow x_2^2 + y_2^2 - 1 = 0 \quad \text{точка } B \text{ лежит на окружности};$$

$$P_3(x_1, y_1, x_2, y_2, z) = 0 \Leftrightarrow y_1(x_2 - z) = y_2(x_1 - z) \quad \text{точка } M \text{ лежит на прямой } AB.$$

Чтобы записать вывод, надо взять две другие точки A' и B' , записать для них те же условия, а затем приравнять указанные в утверждении произведения. Можно также взять точки A' и B' на диаметре, тогда расстояния будут: $1 - r$ и $1 + r$, если точка M лежит внутри круга радиуса 1, и $r - 1$ и $r + 1$, если точка лежит вне этого круга. Здесь $r = |z|$. Тогда произведение расстояний будет $|1 - r^2|$, а вывод может быть записан так:

$$\sqrt{(z - x_1)^2 + y_1^2} \sqrt{(z - x_2)^2 + y_2^2} = |1 - z^2|$$

или после возвведения в квадрат (что не нарушает эквивалентности, так как обе части неотрицательны):

$$((z - x_1)^2 + y_1^2)((z - x_2)^2 + y_2^2) = (1 - z^2)^2.$$

Перенося все члены в одну часть, получим условие в виде равенства нулю многочлена от нескольких переменных:

$$P(x_1, y_1, x_2, y_2, z) = 0 \Leftrightarrow ((z - x_1)^2 + y_1^2)((z - x_2)^2 + y_2^2) - (1 - z^2)^2 = 0.$$

Если удастся выразить P через P_1, P_2, P_3 как указано ранее, то при любых значениях переменных выполнение условий обеспечивает выполнение утверждения. Тем самым, утверждение будет доказано.

1.5.2. Случай многочленов одной переменной

Проанализируем решение задачи о возможности выражения одного многочлена через другие для случая многочленов одной переменной.

В этом случае, как известно из расширенного алгоритма Евклида для многочленов, многочлен P можно выразить через многочлены P_1, \dots, P_n тогда и только тогда, когда P делится на наибольший общий делитель многочленов P_1, \dots, P_n .

Пример 1.66. Многочлен $P = x^3 + x - 2$ выражается через $P_1 = x^2 - 1$ и $P_2 = (x - 1)^2$, так как делится на многочлен $x - 1$, который является НОД многочленов P_1 и P_2 :

$$x^3 + x - 2 = A(x)(x^2 - 1) + B(x)(x - 1)^2, \text{ где } A(x) = x + 1, B(x) = -1.$$

В случае, если необходимо выразить многочлен P через единственный многочлен P_1 , то задача упрощается — надо поделить P на P_1 «уголком» (см. алгоритм 1.20): если получится остаток ноль, то многочлен P выражается через многочлен P_1 (будем говорить, редуцируется к нулю через P_1).

Алгоритм Евклида позволяет также решить задачу с несколькими многочленами:

- 1) сначала находим $D(P_1, \dots, P_n)$, используя алгоритм деления многочленов с остатком;
- 2) затем делим многочлен P на $D(P_1, \dots, P_n)$: если остаток равен нулю, то многочлен P выражается через многочлены P_1, \dots, P_n , иначе — нет.

Пример 1.67. В примере 1.66 деление P_1 на P_2 дает в остатке многочлен $P_3 = 2x - 2$, на который делится многочлен P . Поэтому, если к исходным многочленам P_1 и P_2 добавить P_3 , то многочлен P , как будем говорить в дальнейшем, может быть «редуцирован к нулю по системе многочленов» P_1, P_2, P_3 . При этом для редукции будет использован только многочлен P_3 , отсутствующий в исходной системе многочленов.

1.5.3. Понятие идеала

Аналогично определению 1.30 дадим формальное определение для многочлена нескольких переменных.

Определение 1.35. Конечную сумму вида

$$P(x_1, x_2, \dots, x_n) = \sum_{J=\{j_1, j_2, \dots, j_n\}} c_J x_1^{j_1} x_2^{j_2} \cdots x_n^{j_n}, \quad c_J \in K$$

называют *многочленом от n переменных над полем K* , а набор из целых неотрицательных чисел $J = \{j_1, j_2, \dots, j_n\}$ называют *мультииндексом*.

Многочлен вида $c x_1^{j_1} x_2^{j_2} \cdots x_n^{j_n}$ называют *одночленом*. Степенью (ненулевого) одночлена называют целое число $|J| = j_1 + j_2 + \cdots + j_n$. Степенью многочлена называют максимальную из степеней его одночленов, тождественный ноль не имеет степени.

Множество многочленов от n переменных над полем K обозначают $K[x_1, x_2, \dots, x_n]$.

Замечание 1.50. Если в определении 1.35 допустить отрицательные степени, то полученное выражение называют *многочленом Лорана*.

Для многочленов от n переменных легко доказать теорему аналогичную теореме 1.44.

Теорема 1.53. Множество многочленов $K[x_1, x_2, \dots, x_n]$ есть ассоциативное коммутативное кольцо с единицей.

Исходная задача может быть интерпретирована по-другому.

Для многочленов $P_1, \dots, P_n \in K[x_1, x_2, \dots, x_n]$ строим множество многочленов вида:

$$I = \{Q_1 P_1 + Q_2 P_2 + \cdots + Q_n P_n \mid Q_1, \dots, Q_n \in K[x_1, x_2, \dots, x_n]\}. \quad (1.78)$$

Требуется определить принадлежность многочлена P этому множеству.

Многочлены P_1, \dots, P_n называют *базисом I* (несмотря на то, что они не обязательно образуют минимальное множество).

Пример 1.68. В примере 1.67 базис образуют как многочлены P_1, P_2 , так и многочлены P_1, P_2, P_3 . Кроме того, базис образует и единственный многочлен P_3 .

Так построенное множество многочленов образует новую математическую структуру — идеал — во множестве всех многочленов $K[x_1, x_2, \dots, x_n]$.

Определение 1.36. Непустое подмножество I кольца M называют *идеалом* этого множества, если разность двух элементов этого подмножества и умножение его элемента на любой элемент множества M , не выводят за пределы I :

- 1) $a \in I, b \in I \Rightarrow a - b \in I,$
- 2) $a \in I, m \in M \Rightarrow am \in I.$

Упражнение 1.9. Докажите, что для любых P_1, \dots, P_n принадлежащих $K[x_1, x_2, \dots, x_n]$, множество I , определенное в (1.78), есть идеал в кольце многочленов $K[x_1, x_2, \dots, x_n]$.

Историческая справка. Понятие идеала (или в первоначальной терминологии идеального числа) было введено в 1847 г. немецким математиком Э. Куммером для одного частного случая числовых полей в попытке доказать великую теорему Ферма.

□ Идеалы в кольце многочленов одной переменной

В случае многочленов одной переменной множество I , определенное в (1.78), совпадает с множеством, которое образуют многочлены, кратные $D(P_1, \dots, P_n)$. Процедура определения принадлежности многочлена одной переменной P идеалу, определяемому многочленами P_1, \dots, P_n , крайне проста, если среди набора есть НОД этих многочленов R . Редукция осуществляется вычитанием из P многочлена R , умноженного на многочлен Q — частное многочленов P и R (предварительно можно сделать редукцию исходного многочлена и по другим многочленам P_1, \dots, P_n). Если же среди многочленов не содержится их НОД, то подобная редукция не даст результата. Поэтому для применения того же приема, базис идеала I придется расширить, добавив НОД многочленов (или, что то же самое, остатки от деления многочленов P_i друг на друга — рано или поздно НОД среди них появится).

Теорема 1.54. Добавление к базису идеала остатков от деления многочленов идеала друг на друга не меняет идеала.

Доказательство. Заметим, что если $R(x) \equiv P_j(x) \pmod{P_i(x)}$, т. е. $R(x)$ есть остаток от деления $P_j(x)$ на $P_i(x)$, то существует многочлен $Q(x)$ — частное, — такой что $R(x) = P_j(x) - Q(x)P_i(x)$. По свойству идеала, умножение многочлена, принадлежащего идеалу, на любой многочлен дает многочлен идеала, и вычитание многочленов идеала не выводит за его пределы. Так что $P_i(x) \in I, P_j(x) \in I \Rightarrow R(x) \in I$. ■

1.5.4. Линейные многочлены многих переменных. Понятие базиса Гребнера

Рассмотрим решение задачи о разложении многочлена для линейных многочленов многих переменных. В этом случае для построения базиса исходной системы и осуществления редукции по нему можно применить метод Гаусса.

Пример 1.69. Можно ли выразить многочлен

$$P(x, y, z, u) = 3y - 2u + 5z - 4 + 9x$$

через многочлены

$$\begin{aligned} P_1(x, y, z, u) &= 2 - u + 2z - y + 5x, & P_2(x, y, z, u) &= -z - u + 2y + 3x - 5, \\ P_3(x, y, z, u) &= 4x - 3y - 2z + 3 - u, & P_4(x, y, z, u) &= z - u + 6x - 6y + 10. \end{aligned}$$

Расположим одночлены каждого многочлена в лексикографическом порядке (далее будет дано формальное определение лексиграфического порядка — определение 1.50):

$$\begin{aligned} P(x, y, z, u) &= 9x + 3y + 5z - 2u - 4, \\ P_1(x, y, z, u) &= 5x - y + 2z - u + 2, & P_2(x, y, z, u) &= 3x + 2y - z - u - 5, \\ P_3(x, y, z, u) &= 4x - 3y - 2z - u + 3, & P_4(x, y, z, u) &= 6x - 6y + z - u + 10. \end{aligned}$$

Если не внести в набор P_1, P_2, P_3, P_4 изменений, то редуцировать по нему многочлен P не удастся, так как после первой редукции, например, через многочлен P_1 , коэффициент при x станет равным нулю, в то время как среди оставшихся будут ненулевые, и дальнейшая редукция невозможна.

Сделаем типичное для метода Гаусса элементарное преобразование (в дальнейшем такое преобразование будем называть построением S -многочлена): построим многочлен

$$S(P_1, P_2) = 3P_1 - 5P_2 = -13y + 11z + 2u + 31,$$

исключив старший член. Добавим полученный S -многочлен к системе как P_5 (предварительно попробовав редуцировать через остальные, но здесь редукции не будет, так как все остальные многочлены содержат член со старшей переменной x).

Далее построим $S(P_1, P_3) = 4P_1 - 5P_3 = 11y + 18z + u - 7$. Редуцируя его через многочлен P_5 (через другие это невозможно), получим многочлен $71z + 7u + 50$ (после домножения на подходящую константу, чтобы иметь дело с целыми коэффициентами). Поскольку более он не редуцируется, обозначим его P_6 и добавим к системе.

Затем построим $S(P_1, P_4) = 6P_1 - 5P_4 = 24y + 7z - u - 38$. Редуцируем его через P_5 и P_6 . После первого редуцирования (и домножения на удобный коэффициент) получим $71z + 7u + 50$, который равен P_6 и поэтому редуцируется через него к нулю.

Теперь построим S -многочлены для P_2 и многочленов, больших его по номеру (все S -многочлены для P_1 уже построены).

Получим: $S(P_2, P_3) = 4P_2 - 3P_3 = 17y + 2z - u - 29$. Редуцируем его через P_5 и получим (после домножения на соответствующий коэффициент) $71z + 7u + 50$, который редуцируется к нулю через P_6 .

Далее: $S(P_2, P_4) = 2P_2 - P_4 = 10y - 3z - u - 20$. Редуцируем его через P_5 и получим опять $71z + 7u + 50$, который редуцируется к нулю через P_6 .

Наконец, $S(P_3, P_4) = 3P_3 - 2P_4 = 3y - 8z - u - 11$. Редуцируем его через P_5 и P_6 и получим ноль:

$$3y - 8z - u - 11 - \left(-\frac{3}{13}\right)(-13y + 11z + 2u + 31). \quad (1.79)$$

Домножив (1.79) на числовой коэффициент 13, позволяющий сделать коэффициенты результата целыми, получим равенство

$$-68z - 7u - 50 - 71z - 7u - 50 - (-1)(71z + 7u + 50) = 0.$$

Других S -многочленов построить нельзя. Построенную таким образом систему $\{P_1, P_2, P_3, P_4, P_5, P_6\}$ называют базисом Гребнера для идеала, определяемого исходной системой $\{P_1, P_2, P_3, P_4\}$. Базис получил специальное название (базис Гребнера) потому, что через него любой многочлен идеала редуцируется к нулю. Этот базис не обязан содержать минимального числа элементов, но при возможности количество элементов следует делать меньше. В данном случае два из трех многочленов P_1, P_2, P_3 можно удалить, так как они выражаются через оставшиеся три многочлена системы (что ясно из их построения).

Оставим базис $\{P_1, P_5, P_6\}$.

Теперь для решения вопроса, можно ли выразить P через исходные многочлены, достаточно проверить редуцируемость P к нулю по базису $\{P_1, P_5, P_6\}$.

После редуцирования P через P_1 получаем $Q_1 = 24y + 7z - u - 38$. После редуцирования Q_1 через P_2 (и домножения на удобный коэффициент) имеем $71z + 7u + 50$, редуцируемый к нулю через P_6 .

Таким образом, исходный вопрос решен положительно: P выражается через P_1, P_2, P_3, P_4 .

Упражнение 1.10. Найдите многочлены Q_1, Q_2, Q_3, Q_4 — коэффициенты разложения P по P_1, P_2, P_3, P_4 .

Замечание 1.51. В полученном алгоритме нетрудно видеть алгоритм Гаусса для определения ранга расширенной системы (если добавление уравнения $P = 0$ не меняет ранга системы, значит $P = 0$ не влияет на множество решений системы и выражается как линейная комбинация других уравнений).

Полученные приемы обобщает на случай системы многочленов многих переменных алгоритм Бухбергера. Алгоритм Бухбергера позволяет получить базис Гребнера идеала, заданного произвольным конечным набором многочленов.

Если базис Гребнера построен, то для решения вопроса о принадлежности многочлена идеалу его достаточно редуцировать по многочленам этой системы.

1.5.5. Алгоритм Бухбергера

Для формулировки алгоритма Бухбергера нам понадобится определить аккуратно понятия:

- 1) главного члена многочлена с несколькими переменными;
- 2) редукции многочлена;
- 3) S -многочлена.

□ Главный член многочлена с несколькими переменными

Одночлены многочлена с несколькими переменными можно упорядочить по разному. Рассмотрим один из способов.

Приведем многочлен к нормальному виду, т. е., раскроем скобки и приведем подобные члены. Теперь расположим все множители одночлена в лексикографическом порядке, считая, например, переменную x_1 главной, x_2 следующей по старшинству и так далее. Наконец, упорядочим все одночлены в лексикографическом порядке. Так, если два одночлена включают переменную x_1 в разных степенях, то старшим будем считать одночлен с большей степенью переменной. Если степени главной переменной одинаковы, то сравниваем следующие по старшинству переменные и т. д.

Определение 1.37. Одночлен $x_1^{k_1}x_2^{k_2}\cdots x_n^{k_n}$ считается старше одночлена $x_1^{m_1}x_2^{m_2}\cdots x_n^{m_n}$, если существует i , при котором $k_i > m_i$, а все предыдущие степени одинаковы: $k_j = m_j$ при $j < i$.

По аналогии с многочленами одной переменной можно обозначить $\deg(x_1^{k_1}x_2^{k_2}\cdots x_n^{k_n}) = (k_1, k_2, \dots, k_n)$. Тогда порядок на множестве одночленов определяется лексикографическим порядком на наборах, задающих степени многочленов. Иногда одночлен $x_1^{k_1}x_2^{k_2}\cdots x_n^{k_n}$ записывают в виде x^k , имея в виду $x = (x_1, x_2, \dots, x_n)$, $k = (k_1, k_2, \dots, k_n)$.

Пример 1.70. Упорядочим одночлены по убыванию их старшинства в многочлене $P(x, y, z) = y^2z + z^4 + xy^3$ (здесь x соответствует старшей переменной x_1 , y — переменной x_2 , z — переменной x_3). Первые два члена не включают x , т. е. соответствующий показатель степени при x в этих одночленах равен нулю, а у последнего одночлена — единице, значит он и есть главный. Аналогично член y^2z старше z^4 . Получаем:

$$P(x, y, z) = xy^3 + y^2z + z^4.$$

Если обозначить старший член многочлена P (как и для многочленов одной переменной) $l(P)$, то в примере 1.70 $l(P) = xy^3$, а результат упорядочения может быть записан так:

$$\deg(xy^3) > \deg(y^2z) > \deg(z^4).$$

Упражнение 1.11. Упорядочьте по старшинству члены следующего многочлена: $P(x_1, x_2, x_3, x_4) = x_1x_3^4x_4^2 + x_3^4x_4^2 + x_1^2x_2^2x_4^2 + x_1^2x_2^2x_3^4$.

□ Редукция многочлена с несколькими переменными

Определение 1.38. Если старший член многочлена P делится на старший член другого многочлена Q и дает в результате одночлен t , то операцию $P - tQ$ называют *редукцией* многочлена P с помощью Q .

Теорема 1.55. Замена редуцируемого базисного многочлена P многочленом P^* , полученным в результате редукции по другим многочленам базиса, не меняет идеала.

Идея доказательства.

При однократной редукции P по многочлену Q получаем $P^* = P - tQ$, где t — некоторый одночлен. Отсюда $P = P^* + tQ$, т. е. исходный многочлен может быть выражен через многочлены нового базиса. Аналогичное утверждение справедливо и при дальнейших редукциях.

Замечание 1.52. Поскольку на делимость многочленов умножение на ненулевую константу не влияет, редукцией можно называть и операцию вида $c(P - tQ)$ с любым c отличным от нуля.

Пример 1.71. Многочлен $P(x, y, z) = xy^3 + y^2z + z^4$ из примера 1.70 редуцируется с помощью многочлена $P_1(x, y, z) = 3y^2 + y + z^5$ к многочлену $y^2z + z^4 - 1/3xy^2 - 1/3xyz^5$, или (после умножения на 3) к многочлену $P_2(x, y, z) = -xy^2 - xyz^5 + 3y^2z + 3z^4$, старший член полученного многочлена меньше, чем был раньше (хотя членов стало больше). Полученный многочлен P_2 можно еще раз редуцировать с помощью P_1 :

$$P_3(x, y, z) = -3xyz^5 + xy + xz^5 + 9y^2z + 9z^4.$$

Теперь его старший член меньше старшего члена многочлена P_1 и дальнейшая редукция невозможна.

Это решение можно записать в виде деления «уголком». Используя обобщение алгоритма 1.20 (см. пример 1.48), получаем:

$$\begin{array}{r}
 xy^3 + y^2z + z^4 \\
 \underline{xy^3 + \frac{1}{3}xy^2 + \frac{1}{3}xyz^5} \\
 -\frac{1}{3}xy^2 - \frac{1}{3}xyz^5 + y^2z + z^4 \\
 \underline{-\frac{1}{3}xy^2 - \frac{1}{9}xy - \frac{1}{9}xz^5} \\
 -\frac{1}{3}xyz^5 + \frac{1}{9}xy + \frac{1}{9}xz^5 + y^2z + z^4
 \end{array}
 \quad \left| \begin{array}{l} 3y^2 + y + z^5 \\ \hline \frac{1}{3}xy - \frac{1}{9}x \end{array} \right.$$

Таким образом видно, что редукция одного многочлена по другому является обобщением алгоритма деления многочленов на многочлены нескольких переменных.

Запишем алгоритм редукции многочлена многих переменных по базису Гребнера.

Нам потребуется вспомогательный алгоритм редукции многочлена P по Q . Результат работы алгоритма 1.23 — редуцированный многочлен сохраняется в P . Напомним, что $l(P)$ и $l(Q)$ — обозначение старших коэффициентов P и Q соответственно.

Алгоритм 1.23 (редукция многочлена P по Q).

```

while  $l(P)$  делится на  $l(Q)$  do
     $P \leftarrow P - \frac{l(P)}{l(Q)} \cdot Q$ 
end while

```

Редукция многочлена по базису Гребнера сводится к редукции многочлена по каждому из многочленов базиса. Упорядочим многочлены базиса по убыванию их степени. Тогда сначала многочлен редуцируется по первому многочлену базиса, затем по второму и т. д. Результат работы алгоритма 1.24 — редуцированный многочлен сохраняется в P .

Алгоритм 1.24 (редукция многочлена P по базису Гребнера P_1, \dots, P_n).

```

while есть многочлен  $P_i$ , по которому редуцируется  $P$  do
     $Q \leftarrow$  первый из многочленов  $P_i$ , у которого  $l(P)$  делится на  $l(P_i)$ 
     $P \leftarrow$  результат редукции  $P$  по  $Q$  (Алгоритм 1.23)
end while

```

Из примера 1.71 видно, что в процессе редукции степени переменных младших членов быстро растут. Может ли быть так, что процесс редукции не остановится? Нет.

Докажем предварительно вспомогательное утверждение.

Лемма 1.10. При редукции степень одночлена уменьшается.

Доказательство. После домножения делителя на частное от деления их старших коэффициентов степени делимого и делителя сравняются, но другие одночлены делителя, как и делимого, будут иметь меньшую степень. После вычитания старшие члены уничтожаются, старшим членом станет один из одночленов меньшей степени, и степень многочлена уменьшится. ■

Для дальнейших рассуждений полезно использовать геометрическую интерпретацию степеней одночленов (момономов). Сопоставим одночлену от n переменных $x_1^{k_1}x_2^{k_2}\cdots x_n^{k_n}$ точку с координатами (k_1, k_2, \dots, k_n) (если переменная в одночлене отсутствует, соответствующий показатель степени берется нулевым). Заметим, что координаты являются неотрицательными целыми числами.

Пример 1.72. На рис. 1.5, а показаны графические изображения степеней всех одночленов многочлена $x^4y + 2x^3y^2 - x^3 + 5x^2y^3 + y^2$.

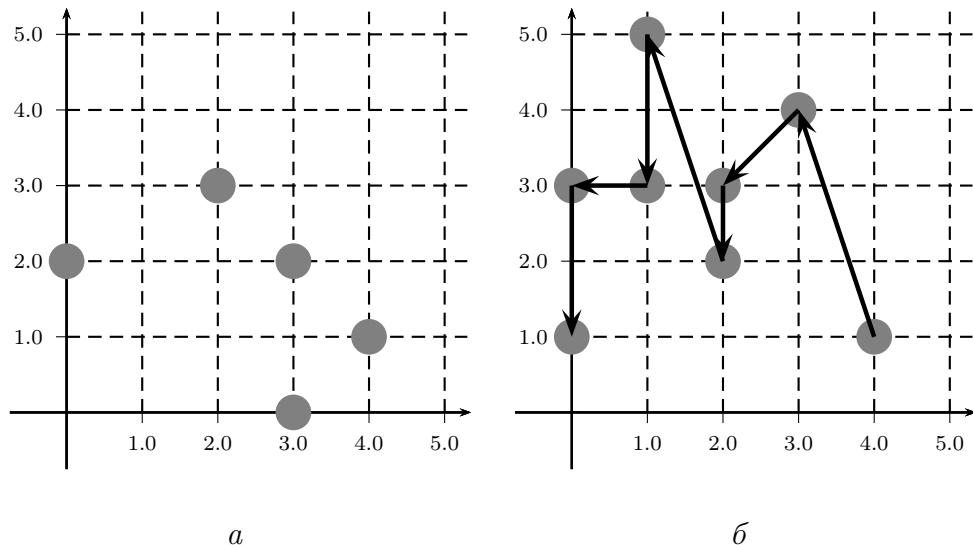


Рис. 1.5

Теорема 1.56. Процесс редукции заканчивается за конечное число шагов.

Доказательство. Докажем теорему для многочленов двух переменных. Для многочленов большего числа переменных доказательство аналогичное. Заметим, что чем меньше степень одночлена, тем левее лежит точка, изображающая его степень. Если же точки лежат на одной вертикали, то меньшие степени изображаются точками, лежащими ниже.

Таким образом, если проследить за движением точки, изображающей степень старшего члена многочлена при его редукции, то получится траектория, состоящая из перемещения точки влево (и, возможно, одновременно вверх или вниз), либо вниз по вертикали (рис. 1.5,б).

Поскольку координаты точки целые и неотрицательные, траектория может состоять только из конечного числа звеньев, что и означает окончание процесса редукции за конечное число шагов. ■

□ Построение S -многочленов

Определение 1.39. Для двух многочленов P и Q определим S -многочлен следующим образом:

$$S(P, Q) = Pq - Qp, \text{ где:}$$

$l(P) = pd, l(Q) = qd, d = D(l(P), l(Q))$ — нетривиальный НОД старших членов многочленов P и Q .

Если $D(l(P), l(Q)) = \text{const}$, то говорят, что многочлены не имеют *зацепления*.

Теорема 1.57. Добавление к базису идеала S -многочленов, построенных по базисным многочленам, не меняет идеала.

Доказательство. Если многочлен $R = S(P, Q)$, то $R = Pq - Qp$, где p и q — некоторые одночлены. Значит, R принадлежит идеалу, которому принадлежат P и Q . ■

Пример 1.73. Многочлен $Q = x^3y + x^3$ не редуцируется с помощью многочлена $P = x^2y^4 + x^2y^2$, но имеет с ним зацепление. Тогда S -многочлен для них будет $S(Q, P) = y^3Q - xP = x^3y^3 - x^3y^2$.

Заметим, что степени обоих одночленов результата превышают степени исходных многочленов. Таким образом, при использовании S -многочленов для построения базисов Гребнера нужно обосновать конечность процесса добавления этих многочленов к исходному набору.

□ Построение базиса Гребнера

Пусть дана система m многочленов от n переменных P_1, P_2, \dots, P_m . Она определяет идеал

$$I = \{Q_1P_1 + Q_2P_2 + \dots + Q_mP_m \mid Q_1, \dots, Q_m \in K[x_1, x_2, \dots, x_m]\}.$$

Требуется построить базис Гребнера этого идеала.

Определение 1.40. *Базисом Гребнера* называют такой базис идеала, для которого принадлежность многочлена идеалу определяется редукцией по многочленам базиса.

Алгоритм 1.25 (алгоритм Бухбергера).

```

while существуют многочлены, имеющие зацепление do
    > проводим поиск только среди тех пар, которые еще не рассматривались
    Для найденного зацепления строим  $S$ -многочлен
    Редуцируем его по всем многочленам системы
        > подойдет любая последовательность редукций
    if многочлен редуцировался к ненулевому многочлену then
        добавляем его к системе
    end if
end while

```

Для дальнейших рассуждений потребуется вспомогательная лемма.

Лемма 1.11. Любой мономиальный идеал имеет конечный базис.

Доказательство. Проиллюстрируем доказательство на многочленах двух переменных. Для многочленов многих переменных доказательство проводится аналогично.

Сначала заметим, что каждый одночлен (моном) порождает все одночлены, лежащие правее (умножение на первую переменную) и/или выше (умножение на вторую переменную) (рис. 1.6,*a*). Все эти одночлены делятся на данный моном.

Геометрически задача означает, что имеется некоторая область, лежащая в первом квадранте и являющаяся объединением «углов» (рис. 1.6,*b*). Требуется найти конечное покрытие этой области такими «углами».

Попробуем, наоборот, построить область из бесконечного числа «углов». После того, как поставлен первый «угол», следующий нужно ставить или левее или ниже (иначе область не изменится) (рис. 1.6,*b*).

Но, как и в случае описанной выше траектории «левее или ниже», число шагов, меняющих область, конечно, поэтому любая область подобного вида имеет конечное покрытие «углами». ■

Докажем теперь основное утверждение.

Теорема 1.58. Предложенный алгоритм 1.25 строит базис Гребнера для любого конечного набора многочленов от нескольких переменных.

Доказательство. Требуется доказать, что:

1) алгоритм заканчивает свою работу за конечное число шагов (как видно из примера 1.73 степень S -многочлена может превышать степени исходных многочленов, поэтому это утверждение неочевидно);

2) полученный базис является базисом Гребнера.

Доказательство первого пункта будем вести от противного, т. е., положим, что применение операций построения S -многочленов и редукции привело к бесконечному множеству многочленов.

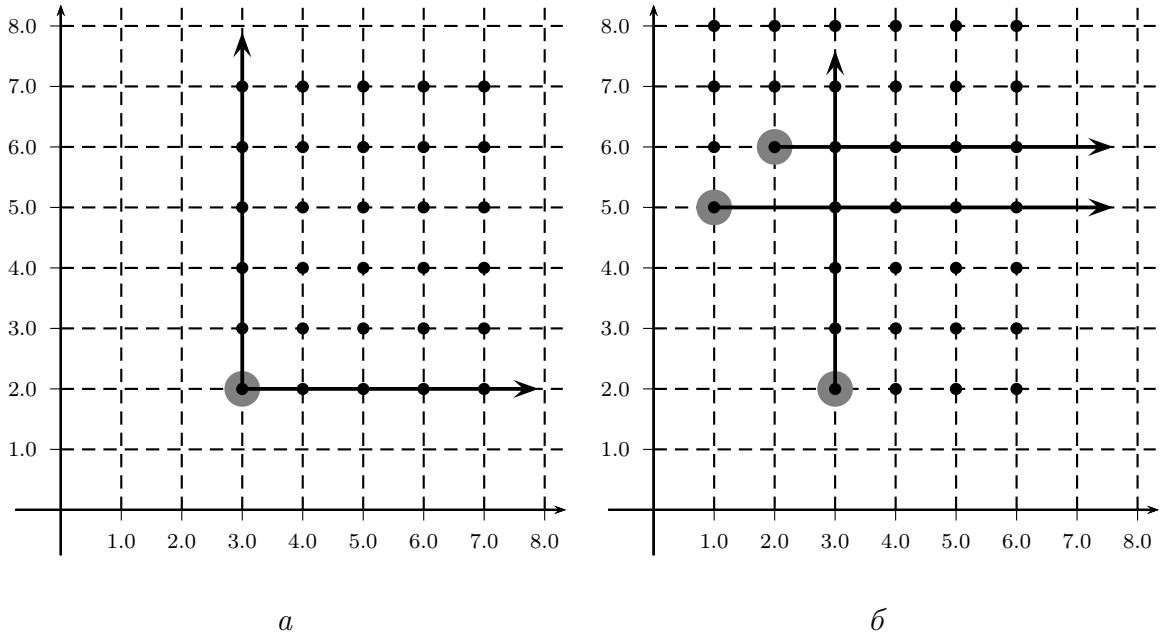


Рис. 1.6

Проанализируем множество их старших членов (момомов) и докажем, что если рассмотреть идеал, построенный на них (мономиальный идеал), то в нем можно будет построить конечный базис.

Установим, что процесс построения S -многочленов с редукцией по уже имеющимся многочленам не может продолжаться бесконечно. От противного. Пусть описанные процедуры порождают бесконечное множество нередуцируемых друг через друга многочленов. Рассмотрим конечный базис множества их старших членов — мономов. Возьмем многочлены, старший член которых не вошел в этот базис. По лемме их старшие коэффициенты делятся на старшие члены некоторых многочленов, вошедших в базис, а значит соответствующие им многочлены должны редуцироваться по выделенным таким образом многочленам, что противоречит нередуцируемости построенных многочленов.

Докажем редуцируемость любого многочлена идеала по построенному базису. От противного. Пусть имеется многочлен $P = \sum H_i P_i$, который нельзя редуцировать по построенному базису. Старший член такого многочлена не должен делиться ни на один из старших членов, построенных по алгоритму Бухбергера многочленов P_i . Это означает, что в представленной сумме старшие члены уничтожились, и получившийся после приведения одночленов старший член не делится ни на один из старших членов P_i . Будем считать, что из всех представлений выбрано то, для которого степень самого старшего члена правой части наименьшая из возможных:

$$P = \sum H_i P_i. \quad (1.80)$$

Покажем, что эту сумму можно преобразовать так, что степень старшего члена уменьшится, тем самым прида к противоречию, доказывающе-

му утверждение теоремы. Разобьем сумму на две: первая будет содержать только члены старшей степени, вторая — члены меньшей степени (поэтому в дальнейшем ее можно не рассматривать).

Первую сумму можно разбить еще на две суммы, выделив у многочленов H_i старший член. Тогда первая из этих сумм

$$P = \sum l(H_i)P_i, \quad (1.81)$$

а вторая будет содержать члены меньшей степени и поэтому ее можно, как и в предыдущем случае, не рассматривать.

Оказывается, вследствие сокращения старших членов сумму (1.81) можно переписать в виде

$$\sum l_{ij}S(l(H_i)P_i, l(H_j)P_j), \quad (1.82)$$

где $l(H_i)$ — старший член многочлена H_i (моном), l_{ij} — числовые коэффициенты (см. задачу 1.2). Заметим, что при построении S -многочлена для многочленов P и Q одинаковой степени в выражении $S(P, Q) = Pq - Qp$, p и q — числа, поэтому степень S -многочлена будет меньше степеней исходных многочленов. Значит у полученной суммы (1.82) степень старшего члена меньше, чем у суммы (1.81).

Выражение $S(l(H_i)P_i, l(H_j)P_j)$ можно записать так: $x^m S(P_i, P_j)$, где $x = (x_1, x_2, \dots, x_n)$, $k = (k_1, k_2, \dots, k_n)$ (см. задачу 1.3). Поскольку все нередуцируемые S -многочлены уже включены в базис, многочлены $S(P_i, P_j)$ редуцируются к нулю, а значит, их можно представить в виде суммы $\sum G_s P_s$, степень старшего члена которой совпадает со степенью $S(P_i, P_j)$ (см. задачу 1.4).

Подставив полученные выражения в сумму (1.82), получим разложение, аналогичное разложению (1.80), но со старшим членом меньшей степени. Добавив две другие суммы, снова получим сумму того же вида (1.80) и снова со старшим членом меньшей степени, что противоречит предположению о выборе разложения. ■

Задача 1.2. Обозначим через x^k одночлен вида $x_1^{k_1}x_2^{k_2}\cdots x_n^{k_n}$. Пусть $f_i = a_i x^k + \dots$

Выразите коэффициенты l_{ij} через b_i так, чтобы выполнялось равенство $\sum b_i f_i = \sum l_{ij} S(f_i, f_j)$, если известно, что старшие коэффициенты в сумме уничтожаются, а b_i и l_{ij} — числа.

Задача 1.3. Докажите, что $S(M_i P_i, M_j P_j)$ делится на $S(P_i, P_j)$, где M_i — мономы.

Задача 1.4. Докажите, что, если многочлен P редуцируется к нулю по многочленам P_i , его можно представить в виде суммы $\sum G_i P_i$, степень которой совпадает со степенью P .

Замечание 1.53. Задачи 1.2 и 1.3 — повышенной сложности.

В процессе построения базисов Гребнера, как отмечалось ранее, могут появиться многочлены больших степеней, что затрудняет решение задач вручную. Применение пакетов символьных вычислений позволяет выполнять символьные вычисления с многочленами больших степеней.

Вот как выглядит решение задачи из начала раздела (см. пример 1.65) с помощью пакета Maple (в текст внесены комментарии к решению).

Пример 1.74 (построение базиса Гребнера).

```
>P:=((z-x1)^2+y1^2)*((z-x2)^2+y2^2)-(1-z^2)^2;
```

$$P := ((z - x_1)^2 + y_1^2)((z - x_2)^2 + y_2^2) - (1 - z^2)^2;$$

Так записывается многочлен, принадлежность которого к идеалу нужно установить.

```
>with(Groebner):G:=[x1^2+y1^2-1,x2^2+y2^2-1,y1*(x2 - z)-
y2*(x1 - z),(y1-y2)*w-1]:
```

Так описывается команда подключения операций, используемых при построении базисов Гребнера, и указывается набор многочленов, определяющих идеал. Обратите внимание, что кроме трех многочленов, описывающих принадлежность точек окружностям и одной прямой, появилось четвертое условие $(y_1 - y_2)w - 1$, которое указывает на то, что ординаты точек y_1 и y_2 не могут быть одинаковыми, иначе прямая, проведенная через них, не пересечет ось x (для задания этого условия в виде равенства нулю некоторого многочлена пришлось ввести новую переменную w). Подобные условия в геометрических задачах мы часто не замечаем, принимая их как очевидные. Однако без этого условия решения мы не получим, что легко проверить.

```
>GB:=gbasis(G,plex(x1,x2,y1,y2,z,w));
```

$$\begin{aligned} GB := & [8z^2y_2^3w + 4z^2y_2^2 + 4wy_2 + 4wy_2z^2 + 4z^2y_2^4w^2 - 4w^2z^2y_2^2 + z^4 - 2z^3 + \\ & + 1, wy_1 - wy_2 - 1, 3y_2 - 2y_1z^2 + y_1 + 4y_1z^2y_2^2 + y_1z^4 - y_2z^2 + \\ & + 4y_2^2w - 2y_2z^2 + 4z^2y_2^4w + 4y_2^3z^2 - 4wz^2y_2^2, -zx_2 + 2wy_2 - z^2 + \\ & + 1 + x_2z^3 + 2z^2y_2^3w + 2z^2y_2^2 - 2wy_2z^2, 2x_2wy_2 + z^2x_2 + x_2 + \\ & + 2y_2^3zw + 2y_2^2z - 2zwy_2 - 2z, 4y_1zy_2^2 - 3y_1z + 2x_2y_2 - 2x_2y_2z^2 + \\ & + 2y_1x_2 + y_2z + y_1z^3 - y_2z^3, x_2^2 + y_2^2 - 1, x_1z^2 - x_1 + z^2x_2 - x_2 - \\ & - 2zy_1y_2, 2y_2x_1 - 2x_2y_2z^2 + 2x_2y_2 + 4y_1zy_2^2 + y_1z^3 - y_1z - y_2z^3 - \\ & - y_2z, x_1x_2 - x_1z - zx_2 + y_1y_2 + 1, x_1^2 + y_1^2 - 1] \end{aligned}$$

Эта процедура вычисляет базис Гребнера, учитывая лексикографический порядок, явно определяемый перечислением переменных. В базисе десять многочленов

```
>reduce(P,GB,plex(x1,x2,y1,y2,z,w));  
0
```

Последняя процедура выполняет редуцирование исходного многочлена по построенному базису. В силу того, что он редуцировался к нулю, можно заключить, что утверждение теоремы истинно и доказано.

Замечание 1.54. При практических вычислениях порядок, в котором идут переменные, может существенно повлиять на трудоемкость вычислений.

Например, если в примере 1.74 переменную z сделать старшей, то базис Гребнера будет существенно проще.

```
> GB:=gbasis(G,plex(z,x1,x2,y1,y2,w));  
GB := [wy1 - wy2 - 1, x2^2 + y2^2 - 1, x1^2 + y1^2 - 1, z + x1wy2 - x2wy2 - x2]  
> reduce(P,GB,plex(z,x1,x2,y1,y2,w));  
0
```

Замечание 1.55. Базис Гребнера можно уменьшить и другими способами. Рассмотрим два приема, которые позволяют упростить базис (самостоятельно докажите корректность таких упрощений).

Прием 1. Если старший член одного из многочленов базиса делится на старший член другого, то первый из многочленов можно удалить из базиса.

Прием 2. Если нестарший член одного многочлена базиса делится на старший член другого, то первый многочлен можно упростить, редуцировав его по второму многочлену (относительно выделенного нестаршего члена), т. е., домножить второй многочлен на некоторый одночлен и результат вычесть из исходного многочлена так, чтобы выделенный одночлен уничтожился.

Пример 1.75. В примерах 1.73 и 1.74 не все исходные многочлены присутствуют в построенных базисах Гребнера. В обоих примерах многочлен, описывающий принадлежность точек прямой, был изменен.

1.6. Теория множеств и комбинаторика

Комбинаторика изучает вопросы о том, сколько различных комбинаций, подчиненных тем или иным условиям, можно составить из заданных объектов и как перечислить эти объекты.

Прежде чем перейти к систематическому изложению, рассмотрим несколько классических задач, отражающих ее базовые идеи.

1.6.1. Вводные задачи

□ Принцип умножения

Принцип умножения особенно ярко проявился в открытии Д. И. Менделеевым периодической таблицы химических элементов. Пытаясь найти скрытую закономерность в свойствах элементов, ученый выделил базовые (возможно скрытые от непосредственного восприятия) признаки, комбинации которых характеризуют свойства элементов.

Более формально принцип умножения связан с перечислением (в виде таблицы) всех комбинаций двух признаков различной природы. Эта идея привела к понятию декартова произведения множеств (признаков).

$$A \times B = \{(a, b) | a \in A, b \in B\}$$

Задача 1.5. Сколько существует автомобильных номеров, составленных из трех произвольных букв русского алфавита и четырех десятичных цифр.

Решение. Автомобильный номер можно интерпретировать как один из элементов декартова произведения трехбуквенных «слов» {ААА; ААБ; ...} и четырехзначных наборов цифр {0000; 0001; ...}. С другой стороны, множество трехбуквенных слов может быть представлено как комбинация буквы и двухбуквенного слова и т. д. В результате можно рассмотреть множество всех потенциально допустимых номеров автомобилей как декартово произведение семи множеств:

$$A \times A \times A \times D \times D \times D \times D = A^3 D^4,$$

где A — алфавит (множество букв русского языка); D — множество десятичных цифр.

Понятно, что написанная формула может рассматриваться как формула для подсчета числа элементов декартова произведения множеств:

$$N = |A|^3 \times |D|^4 = 33^3 \times 10^4 = 359\,370\,000$$

— число всех автомобильных номеров.

□ Принцип сложения

Принцип сложения тоже можно иллюстрировать замечательным открытием - открытием телескопа Г. Галилеем. Г. Галилей, используя комбинаторные рассуждения, «переоткрыл» устройство телескопа, которое являлось в то время коммерческим секретом. Удивительно, но сохранилась запись рассуждений Галилея об этом открытии: «Мои рассуждения были такими. Устройство должно иметь либо одно стекло, либо больше. Оно не

может состоять из единственного стекла, потому что стекло может быть выпуклым, вогнутым или иметь параллельные грани. Но последнее из упомянутых стекол не меняет изображения, вогнутое же уменьшает предметы, а выпуклое увеличивает, но делает их, в то же время, неотчетливыми и искаженными. Перейдя к устройству с двумя стеклами и помня, что стекло с параллельными гранями ничего не меняет, я сделал вывод, что нужный эффект не может быть достигнут комбинированием этого стекла с другими. Таким образом, мне осталось исследовать взаимоположение выпуклого и вогнутого стекол».

В этих рассуждениях Галилей сначала разбивает все возможные комбинации стекол (варианты устройства телескопа) на группы по числу стекол в устройстве. Затем в каждой из групп перебирает комбинации стекол различного вида.

Идея разбить все варианты на группы, каждая из которых состоит из «одинаково устроенных» комбинаций в теории прикладных алгоритмов получила название «разделяй и властвуй» (divide and conquer).

Принцип сложения лежит в основе построения рекуррентных формул.

Задача 1.6. Вывести рекуррентную формулу для подсчета C_n^k — числа k -элементных подмножеств n -элементного множества (читается как «число сочетаний из n по k »).

Решение. Выделим из n элементов один и разобьем все k -элементные подмножества на два класса:

- содержащие выделенный элемент;
- не содержащие его.

Первых будет C_{n-1}^{k-1} , так как один элемент уже выбран, и из оставшихся $n - 1$ элементов надо выбрать еще $k - 1$ элемент.

Вторых будет C_{n-1}^k , так как один элемент запрещается выбирать, и надо выбрать все k элементов из оставшихся $n - 1$. Так как любое подмножество (сочетание) принадлежит либо одному, либо другому классу и классы не пересекаются, получаем рекуррентную формулу:

$$C_n^k = C_{n-1}^{k-1} + C_{n-1}^k.$$

Процесс вычислений по этому рекуррентному соотношению можно представить в виде пирамиды, которую называют *треугольником Паскаля* (рис. 1.7).

В этом треугольнике каждое число по выведенной рекуррентной формуле есть сумма двух, стоящих над ним, а по границам треугольника стоят единицы.

Задача 1.7 (о размене монет). Сколькими способами можно разменять один рубль монетами достоинством 50, 10 и 5 копеек.

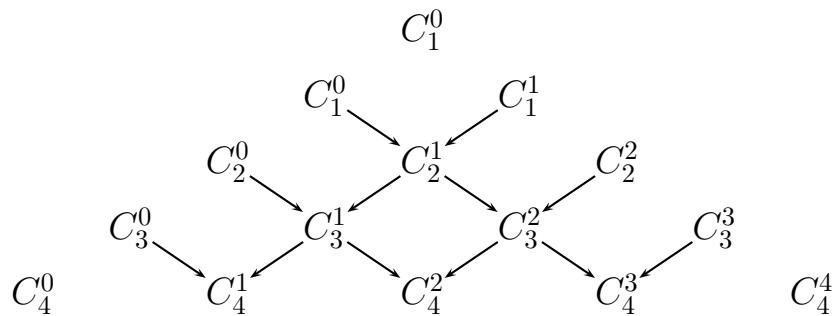


Рис. 1.7

Решение. Обозначим количество способов размена монеты достоинством k на монеты меньшего номинала как D_k . Любой вариант размена можно записать, указав количество монет данного номинала (например, $(1; 3; 4)$ соответствует размену $50 + 10 + 10 + 10 + 5 + 5 + 5 + 5$). При этом первым можно указать количество монет номинала 50, затем 10 и, наконец 5 (будем считать порядок, в котором «выдаются» монеты после размена, несущественным, и разные комбинации в нашей постановке задачи отличаются только составом монет).

Тогда любой размен начинается либо с 50, либо с 10, либо с 5. Соответственно, количество разменов оставшейся суммы для первого вида размена (в общем случае) будет D_{k-50} , для второго — D'_{k-10} , для третьего — D''_{k-5} , где D' обозначает размен без использования монет достоинством 50 копеек, а D'' — без использования монет достоинством 50 и 10 копеек. Тем самым D'' всегда равно 1. Получаем формулы

$$D_k = D_{k-50} + D'_{k-10} + 1, \quad D'_k = D'_{k-10} + 1,$$

по которым рекурсией (сверху вниз) можно получить искомое число разменов. Как легко заметить, $D_0 = 1$ — это соответствует размену $10 = 50 + 50$. Далее имеем:

$$\begin{aligned} D_{100} &= D_{50} + D'_{90} + 1 = D_0 + D'_{40} + 1 + D'_{80} + 1 + 1 = D'_{40} + D'_{80} + 4 = \\ &= D'_{30} + 1 + D'_{70} + 1 + 4 = D'_{30} + D'_{70} + 6 = D'_0 + D'_{40} + 12 = \\ &= D'_{40} + 13 = D'_0 + 17 = 18. \end{aligned}$$

□ Использование взаимно-однозначного соответствия множеств

Рассмотрим примеры, когда вместо прямого подсчета количества элементов данного множества A можно установить взаимно-однозначное соответствие между A и другим множеством B , количество элементов которого уже подсчитано или считается легче.

Задача 1.8. Какое из чисел больше: C_n^k или C_n^{n-k} ?

Решение. По определению, C_n^k есть количество k -элементных подмножеств n -элементного множества. Например, C_5^2 показывает, сколькими способами можно выбрать два предмета из пяти различных.

Если обозначить последнее множество как $A = \{a, b, c, d, e\}$, то все варианты выбора легко перечислить явно: $ab, ac, ad, ae, bc, bd, be, cd, ce, de$. Всего их 10.

Теперь, положим, нас интересует вопрос о том, сколькими способами можно выбрать три предмета из пяти различных. Если обратить внимание на то, что, выбирая два элемента из пяти, мы оставляем три, то между подмножествами выбранных и остающихся элементов возникает естественное взаимно-однозначное соответствие:

$$\begin{aligned} \{a; b\} &- \{c; d; e\}, \quad \{a; c\} - \{b; d; e\}, \quad \{a; d\} - \{b; c; e\}, \quad \{b; c\} - \{a; d; e\}, \\ \{b; d\} &- \{a; c; e\} \text{ и т. д.} \end{aligned}$$

В итоге получаем равенство:

$$C_n^k = C_n^{n-k}.$$

Задача 1.9 (о числе перестановок). Сколькими способами можно переставить между собой цифры числа 12345?

Решение. Рассуждения для решения этой задачи таковы: на первом месте может стоять любая из пяти цифр $\{1, 2, 3, 4, 5\}$, с каждой из этих цифр может комбинироваться одна из четырех, кроме той, которая уже поставлена, что дает 5×4 комбинаций. Далее с любой из этих комбинаций комбинируется любая цифра из трех оставшихся (кроме двух уже использованных) и т. д. По принципу умножения получаем $5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 5! = 120$.

Однако заметим, что принцип умножения в сформулированном выше виде неприменим, так как у нас нет декартова произведения множеств! Когда, например, на первом месте стоит 1, на втором месте могут стоять любые перестановки цифр $\{2, 3, 4, 5\}$, а когда на первом месте стоит 2, то на втором месте стоят уже элементы другого множества — множества перестановок цифр $\{1, 3, 4, 5\}$. Но хотя эти множества и разные, легко понять, что число элементов в них одинаково. Для доказательства достаточно в каждой перестановке цифр первого множества поменять 2 на 1, и получится перестановка цифр второго множества. Аналогично делается и обратный переход. Таким образом, все перестановки четырех элементов имеют одинаковое число элементов, и можно сопоставить их одному фиксированному множеству (из 24 элементов). После этого применение принципа умножения оправдано, и можно записать:

$$P_n = nP_{n-1},$$

где P_n — число перестановок n элементов. По этой рекуррентной формуле легко находим, что

$$P_n = n!.$$

Задача 1.10. Доказать, что количество шестизначных номеров, у которых сумма первых трех цифр равна сумме оставшихся трех («счастливые» номера билетов) равно количеству шестизначных номеров с суммой цифр 27.

Решение. В этой задаче прямой подсчет достаточно труден для нахождения обоих количеств, однако доказательства равенства количества элементов обоих множеств элементарно!

Следующее сопоставление решает задачу:

$$(a; b; c; d; e; f) - (a; b; c; 9 - d; 9 - e; 9 - f).$$

Действительно, то что сопоставление взаимно-однозначно очевидно. С другой стороны, сумма цифр второго билета в показанном сопоставлении равна

$$a + b + c + (9 - d) + (9 - e) + (9 - f) = 27 + [(a + b + c) - (d + e + f)].$$

Поэтому, если первый билет «счастливый», то

$$a + b + c = d + e + f$$

и второй билет имеет сумму цифр 27. Легко показать и обратное.

□ Различные способы рассуждений

Идею взаимно-однозначного соответствия можно видоизменить так, чтобы она не только приводила к факту равнomoщности некоторых множеств, но и давала алгоритм для вычисления количества элементов этих множеств. Для этого можно применить несколько разных способов рассуждений к вычислению количества элементов одного и того же множества, а затем сравнить результаты.

Задача 1.11 (о футбольной команде). Сколькими способами из n претендентов можно выбрать футбольную команду из k человек, а в ней капитана?

Решение. Первый способ рассуждений. Выберем из n человек команду из k человек — это можно сделать C_n^k способами. Далее выберем из этих k человек капитана, что можно сделать k способами. По принципу умножения будем иметь kC_n^k команд, различающихся либо составом, либо капитанами.

Второй способ рассуждений. Выберем из n человек капитана, что можно сделать n способами. Далее из оставшихся человек выберем остальную часть команды, что можно сделать C_{n-1}^{k-1} способами. По принципу умножения получаем другую запись искомого количества nC_{n-1}^{k-1} . Заключаем, что:

$$kC_n^k = nC_{n-1}^{k-1},$$

откуда получаем рекуррентную формулу для вычисления C_n^k :

$$C_n^k = \frac{n}{k} C_{n-1}^{k-1}.$$

Именно благодаря этому число сочетаний из n по k часто обозначают как $\binom{n}{k}$. Например, если $n = 5, k = 3$, получаем

$$C_5^3 = \frac{5}{3} \cdot \frac{4}{2} C_3^1.$$

Но $C_3^1 = 3$, что очевидно из определения C_n^k . Таким образом, $C_5^3 = 10$. В общем случае,

$$C_n^k = \frac{n}{k} C_{n-1}^{k-1} = \frac{n}{k} \frac{n-1}{k-1} C_{n-2}^{k-2} = \dots = \frac{n}{k} \frac{n-1}{k-1} \dots \frac{n-k+1}{1}.$$

□ Принцип использования разных языков описания

Идея взаимно-однозначного соответствия часто проявляется в форме переформулировки исходной задачи, «переводе» с одного языка описания на другой. Например, ранее мы уже столкнулись с языком теории множеств (подсчет элементов декартова произведения) и подсчета слов в некотором алфавите. Это два самых важных, но не единственных языка описания комбинаторных задач. Часто переход от одного описания задачи к другому позволяет применить приемы, известные в терминах другой интерпретации. Таким образом, одним из важных приемов решения задач является установление взаимно-однозначного соответствия между множествами комбинаций в разных по форме задачах.

Задача 1.12. Найти число двоичных последовательностей с k единицами и $n - k$ нулями.

Решение. Интерпретируем k -элементные подмножества n -элементного множества как двоичные последовательности. Для этого упорядочим элементы множества, и каждое подмножество закодируем набором из единиц и нулей, в зависимости от того, входит ли очередной элемент в подмножество или не входит. Например, для множества $\{a; b; c; d; e\}$ подмножество $\{b; c; e\}$ кодируется двоичным набором 01101.

Таким образом, число двоичных последовательностей с k единицами и $n - k$ нулями равно C_n^k . Заметим, что отсюда следует, что число всех подмножеств множества из n элементов равно 2^n , так как каждая цепочка есть элемент n -кратного декартова произведения \mathbf{B}^n бинарных множеств $B = \{0; 1\}$.

Задача 1.13. Сколько существует способов представить число 10 в виде суммы пяти неотрицательных целых слагаемых (с учетом порядка слагаемых, т. е. разложения $4 + 1 + 3 + 0 + 2$ и $2 + 0 + 3 + 4 + 1$ считаются разными).

Решение. Каждое из представлений можно закодировать набором нулей и единиц, где суммарное число единиц равно 10, а нули заменяют плюсы: 11110101110011 кодирует первое из приведенных выше разложений, а 11001110111101 второе. В задаче 1.12 показано, что число таких двоичных наборов равно

$$\frac{14}{4} \cdot \frac{13}{3} \cdot \frac{12}{2} \cdot \frac{11}{1} = 1001.$$

Эту же задачу можно сформулировать по-другому. Например, так. В кондитерской продают пирожные пяти сортов. Надо купить 10 пирожных. Сколькими способами это можно сделать? Каждая покупка кодируется упорядоченным разложением 10 в сумму пяти неотрицательных слагаемых. Так разложение $4 + 1 + 3 + 0 + 2$ кодирует покупку: 4 пирожных первого вида, 1 пирожное — второго и т. д. Таким образом, ответ в задаче такой же — 1001.

В общем случае эту задачу называют нахождением числа сочетаний с повторениями: имеется k видов элементов, каждого вида элементов бесконечное множество. Сколькими способами можно выбрать n элементов?

Ответ в задаче C_{n+k-1}^{k-1} .

Часто процесс комбинирования можно свести к формальному оперированию алгебраическими объектами.

□ Принцип формального оперирования алгебраическими объектами вместо комбинаторных рассуждений

Задача 1.14. Рассмотрим следующее выражение, содержащее n множителей:

$$(x+y)(x+y)(x+y) \cdots (x+y).$$

Требуется раскрыть скобки и упростить выражение.

Решение. Посчитаем количество одночленов, содержащих k букв x (остальные $n - k$ букв будут y).

Каждый такой одночлен может быть закодирован двоичным набором из n цифр: 1 — если мы выбираем x из очередной скобки и 0, если выбираем y . Таким образом, каждый одночлен, содержащий k букв x и $(n-k)$ букв y , т. е. одночлен $x^k y^{n-k}$, будет встречаться C_n^k раз. Итак, получаем формулу:

$$(x+y)^n = x^n + C_n^{n-1} x^{n-1} y + \cdots + C_n^k x^k y^{n-k} + \cdots + y^n.$$

Эту формулу называют *биномом Ньютона*. С точки зрения комбинаторики важность этой формулы в том, что в ней комбинаторные, по сути,

объекты — числа сочетаний — становятся коэффициентами многочлена, с которым можно выполнять алгебраические операции, не задумываясь о комбинаторном смысле действий. Например, подставив вместо x и y единицы, получим тождество:

$$(1+1)^n = 1^n + C_n^{n-1}1^{n-1}1 + \cdots + C_n^k 1^k 1^{n-k} + \cdots + 1^n = \sum_{k=0}^n C_n^k,$$

комбинаторный смысл которого в том, что все подмножества (которых 2^n), можно разбить на классы, в каждом из которых содержатся подмножества с одинаковым числом элементов (число подмножеств с количеством элементов k равно C_n^k). При подстановке $x = 1, y = -1$ получим равенство

$$\sum_k C_n^{2k-1} = \sum_k C_n^{2k},$$

для которого провести комбинаторное рассуждение уже не так просто.

Еще менее очевидный результат получается при подстановке 1 вместо x и y в формулу, полученную дифференцированием бинома Ньютона по x :

$$\begin{aligned} n(x+y)^{n-1} &= nx^{n-1} + (n-1)C_n^{n-1}x^{n-2}y + \cdots + kC_n^kx^{k-1}y^{n-k} + \cdots + 0, \\ n2^{n-1} &= n1^{n-1} + (n-1)C_n^{n-1}1^{n-2}1 + \cdots + kC_n^k1^{k-1}1^{n-k} + \cdots + 0 = \\ &= \sum_{k=0}^n kC_n^k = \sum_{k=1}^n kC_n^k. \end{aligned}$$

Такой переход от оперирования комбинаторными объектами к действиям с функциями называют методом *производящих функций*. Подробно производящие функции рассмотрим в следующем подразделе.

□ Принцип Дирихле

Самая распространенная формулировка принципа Дирихле¹ звучит следующим образом:

если в n клетках сидят больше n зайцев, то найдется клетка, в которой сидят по крайней мере два зайца.

В роли «зайцев» могут выступать различные предметы и математические объекты. Принцип Дирихле можно сформулировать и на языке множеств и отображений:

при любом отображении множества A , состоящего из $n+1$ элемента, в множество B из n элементов, найдутся два элемента множества A , имеющие один и тот же образ.

¹Дирихле Петер Густав Лежен (1805–1859) — выдающийся немецкий математик. Ученик Ж. Фурье. После смерти К. Гаусса возглавил его кафедру в Геттингенском университете. Сделал ряд крупных открытий в теории чисел и математическом анализе.

Несмотря на совершенную очевидность этого принципа, его применение является весьма эффективным методом решения задач, дающим во многих случаях наиболее простое и изящное решение.

Задача 1.15. Имеются n целых чисел a_1, \dots, a_n . Доказать, что среди них найдутся одно или несколько чисел, сумма которых делится на n .

Рассмотрим n сумм:

$$\sum_{i=1}^k a_i, \quad 1 \leq k \leq n$$

Если одна из этих сумм делится на n — задача решена. Если же ни одна из этих сумм не делится на n , тогда по принципу Дирихле существуют суммы принадлежащие одному классу вычетов модулю n :

$$\sum_{i=1}^s a_i \equiv \sum_{j=1}^k a_j \pmod{n}.$$

Утверждение задачи следует из определения сравнения по модулю.

Задача 1.16. На клетчатой бумаге отметили пять точек, расположенных в узлах клеток. Доказать, что хотя бы один из отрезков, соединяющих эти точки, проходит через узел клетки.

Введем на клетчатой бумаге систему координат с началом координат в одном из узлов, осями, направленными вдоль линий сетки, и единичным отрезком, равным стороне клетки. Тогда все отмеченные точки будут иметь целочисленные координаты. Покажем, что найдутся две точки из пяти, у которых одна и та же четность координат (x, y) .

Возможно четыре варианта «четности» координат: (четная, четная), (четная, нечетная), (нечетная, четная), (нечетная, нечетная). Следовательно из пяти точек две будут иметь один вариант «четности». Обозначим их (x_1, y_1) и (x_2, y_2) .

Тогда середина отрезка, соединяющего эти две точки, имеет координаты

$$\left(\frac{x_1 + x_2}{2}, \frac{y_1 + y_2}{2} \right),$$

которые являются целыми числами в силу одинаковой четности x_1, x_2 и y_1, y_2 . Таким образом, середина этого отрезка лежит в узле сетки, т. е. данный отрезок является искомым.

Перейдем к более формальному описанию базовых понятий комбинаторики.

1.6.2. Размещения и сочетания

Определение 1.41. Способ расположения в определенном порядке некоторого числа элементов из заданного конечного множества S называют *размещением*. Иначе говоря — способ выбора этих элементов из S , когда *существенна* последовательность выбора. Если последовательность выбора *несущественна*, то такой способ называют *сочетанием*. При этом может допускаться или нет повторение элементов.

Пример 1.76. Пусть $S = \{a, b, c\}$. Будем выбирать из множества S по два элемента. Тогда:

- 1) размещения без повторений: $\{ab, ac, ba, bc, ca, cb\}$;
- 2) сочетания без повторений: $\{ab, ac, bc\}$;
- 3) размещения с повторениями: $\{aa, ab, ac, ba, bb, bc, ca, cb, cc\}$;
- 4) сочетания с повторениями: $\{aa, ab, ac, bb, bc, cc\}$.

Введем следующие обозначения:

- число размещений без повторений из n элементов по k обозначим A_n^k ;
- число сочетаний без повторений обозначим C_n^k или $\binom{n}{k}$.

Определение 1.42. Размещение без повторений из n элементов по n называют *перестановкой* n элементов и обозначают P_n .

По определению положим $C_n^0 = 1$, $A_n^0 = 1$. Справедлива следующая теорема.

Теорема 1.59.

$$A_n^r = n(n - 1) \cdots (n - r + 1) = \frac{n!}{(n - r)!}, \quad C_n^r = \frac{A_n^r}{r!} = \frac{n!}{(n - r)!r!}.$$

Доказательство. Необходимо заполнить r позиций, выбирая из n -элементного множества без повторений. Первого кандидата можно выбрать n способами. Для каждого из этих способов второго кандидата можно выбрать $n - 1$ способом. Для каждого из $n(n - 1)$ способов третьего кандидата можно выбрать $n - 2$ способами и т. д. Следовательно всего существует $n(n - 1)(n - 2) \cdots (n - r + 1)$ способ выбора.

Замечание 1.56. Заметим в частности, что $P_r = r!$.

Далее, если при выборе не учитывать порядок элементов, то комбинация, состоящая из одних и тех же элементов учитывается в A_n^r ровно $P_r = r!$ раз. Отсюда получаем требуемое. ■

Замечание 1.57. Понятие числа сочетаний появилось в задаче 1.6. Формула для числа перестановок P_n была получена в задаче 1.9. Результаты теоремы 1.59 обсуждаются в задаче 1.11.

Вычислим теперь число размещений и сочетаний с повторениями. Число размещений с повторениями из n элементов по r будем обозначать U_n^r . Справедлива следующая теорема.

Теорема 1.60. $U_n^r = n^r$.

Доказательство. Необходимо заполнить r позиций, выбирая из n -элементного множества с повторениями. Первого кандидата можно выбрать n способами. Для каждого из этих способов второго кандидата можно выбрать также n способом. Для каждого из этих n^2 способов третьего кандидата можно выбрать n способами и т. д. Следовательно, всего существует n^r способ выбора. ■

Пример 1.77. Посчитаем, сколько чисел можно представить m двоичными разрядами. Число A представимо в виде:

$$A = \boxed{a_0 \mid a_1 \mid \cdots \mid a_{m-1}},$$

где $a_i \in \{0, 1\}$. Имеем: $S = \{0, 1\}$, $n = 2$, $r = m$. Тогда всего существует 2^m чисел (от 0 до $2^m - 1$).

Найдем теперь число сочетаний с повторениями. Число сочетаний с повторениями из n элементов по r будем обозначать V_n^r .

Теорема 1.61. $V_n^r = C_{n+r-1}^{n-1} = C_{n+r-1}^r$.

Доказательство. Пусть $S = \{a_1, a_2, \dots, a_n\}$. Рассмотрим произвольное сочетание. Пусть в нем:

$$\left\{ \begin{array}{l} \text{элемент } a_1 \text{ встречается } x_1 \text{ раз,} \\ \text{элемент } a_2 \text{ встречается } x_2 \text{ раз,} \\ \dots \dots \dots \dots \dots \dots, \\ \text{элемент } a_n \text{ встречается } x_n \text{ раз.} \end{array} \right.$$

Очевидно, что

$$x_1 + x_2 + \cdots + x_n = r, \quad x_i \geq 0, \quad i = 1, 2, \dots, n.$$

Проведем некоторое преобразование данного сочетания. Добавим по одному элементу a_1, a_2, \dots, a_n . Тогда в новом сочетании:

$$\left\{ \begin{array}{l} \text{элемент } a_1 \text{ встречается } y_1 = x_1 + 1 \text{ раз,} \\ \text{элемент } a_2 \text{ встречается } y_2 = x_2 + 1 \text{ раз,} \\ \dots \dots \dots \dots \dots \dots, \\ \text{элемент } a_n \text{ встречается } y_n = x_n + 1 \text{ раз.} \end{array} \right.$$

При этом

$$y_1 + y_2 + \cdots + y_n = n + r, \quad y_i > 0, \quad i = 1, 2, \dots, n.$$

Объединим одинаковые элементы сочетания в группы (нам не важен порядок элементов), а между группами поставим разделители. Заметим, что в силу проведенного преобразования между любыми двумя разделителями присутствует хотя бы один элемент ($y_i > 0, i = 1, 2, \dots, n$).

$$\underbrace{a_1, a_1, \dots, a_1}_{y_1} \mid \underbrace{a_2, a_2, \dots, a_2}_{y_2} \mid \dots \dots \mid \underbrace{a_n, a_n, \dots, a_n}_{y_n}. \quad (1.83)$$

Разделителей между группами в (1.83) всего $n - 1$. Мест для их размещения всего $n + r - 1$ (после последнего элемента нет разделителя). Очевидно, что есть взаимно-однозначное соответствие между произвольным исходным сочетанием и взаиморасположением разделителей в (1.83). Отсюда получаем, что число сочетаний с повторениями из n элементов по r равно числу способов выбора $n - 1$ места для разделителей из $n + r - 1$ возможных, т. е. C_{n+r-1}^{n-1} . ■

Замечание 1.58. Частный случай теоремы 1.61 рассмотрен в задаче 1.12.

В полученных формулах для сочетаний и размещений присутствует факториал. Поэтому, применяя эти формулы для решения практических задач нужно помнить, что факториал — это очень быстрорастущая функция (факториал растет быстрее экспоненты). Согласно формуле Стирлинга,

$$\lim_{n \rightarrow \infty} \frac{n!}{n^n e^{-n} \sqrt{2\pi n}} = 1 \quad \text{или} \quad n! \simeq n^n e^{-n} \sqrt{2\pi n} \quad \text{при } n \rightarrow \infty.$$

Приведем и более специальные формулы

$$n^n e^{-n} \sqrt{2\pi n} < n! < n^n \sqrt{2\pi n} \exp\left(-n + \frac{1}{12n}\right),$$

$$n! \simeq n^n \sqrt{2\pi n} \exp\left(-n + \frac{1}{12n} - \frac{1}{360n^2} + \dots\right) \quad \text{при } n \rightarrow \infty.$$

1.6.3. Биномиальные коэффициенты

Оказывается, что числа сочетаний C_n^k , полученные в предыдущем подразделе, обладают целым рядом свойств, которые оказываются очень полезными при решении ряда комбинаторных задач. Числа сочетаний C_n^k называют также *биномиальными коэффициентами*. Смысл этого названия устанавливается следующей теоремой, известной также как формула *бина-ма Ньютона*.

Теорема 1.62.

$$(x+y)^n = \sum_{k=0}^n C_n^k x^k y^{n-k}. \quad (1.84)$$

Доказательство. Коэффициент при $x^k y^{n-k}$ равен числу способов которыми при раскрытии скобок в $(x+y)^n$ из n скобок можно выбрать k скобок с x , а в остальных $n - k$ взять y . ■

Следствие 1.22. Справедливы следующие свойства C_n^k :

$$1) \sum_{k=0}^n C_n^k = 2^n;$$

$$2) \sum_{k=0}^n (-1)^k C_n^k = 0;$$

3) каждое n -элементное множество имеет ровно 2^n подмножеств (включая пустое подмножество и его самого).

Доказательство. Для доказательства п. 1 и 2 в (1.84) нужно положить соответственно $x = y = 1$ и $x = -1, y = 1$. П. 3 следует непосредственно из определения сочетания без повторений. ■

Замечание 1.59. Формула бинома Ньютона и результаты следствия 1.22 к теореме 1.62 обсуждались при изучении алгебраических объектов комбинаторными способами (см. задачу 1.14).

Теорема 1.63. Справедливы следующие равенства

$$1) C_n^k = C_n^{n-k},$$

$$2) C_n^k = C_{n-1}^k + C_{n-1}^{k-1}.$$

Доказательство. Каждой выборке k элементов из n однозначно соответствует выборка оставшихся $n - k$ элементов из n . Это доказывает п. 1. Для доказательства п. 2 рассмотрим n -элементное множество $S = \{s_1, s_2, \dots, s_n\}$ и его произвольный элемент s_i . Все выборки k элементов из n распадаются на две непересекающиеся группы.

1. Выборки, содержащие элемент s_i . Таких выборок всего C_{n-1}^{k-1} , так как выбираем из $n - 1$ кандидата (элемент s_i уже выбран) на оставшееся $k - 1$ место (одно место уже занято элементом s_i).

2. Выборки, не содержащие элемент s_i . Таких выборок всего C_{n-1}^k , так как выбираем из $n - 1$ кандидата (элемент s_i не может принадлежать выборке) на k мест.

Сложив оба числа, получаем требуемое. ■

Следствие 1.23 (треугольник Паскаля). Треугольник Паскаля построен по следующему правилу: его строку с номером i образуют коэффициенты разложения бинома $(x+y)^i$, т. е. числа $C_i^0, C_i^1, \dots, C_i^i$. Согласно теореме 1.63 треугольник имеет вид, представленный на рис. 1.8.

			1					
			1	1				
			1	2	1			
			1	3	3	1		
			1	4	6	4	6	
			1	5	10	10	5	1
			1	6	15	20	15	6
			1	7	21	35	35	21
		

Рис. 1.8

Замечание 1.60. Результаты теоремы 1.63 обсуждаются в задачах 1.6 и 1.8.

Теорема 1.64 (тождество Коши).

$$C_{n+m}^k = \sum_{s=0}^k C_m^s C_n^{k-s}.$$

Доказательство. Предположим, что имеется m белых и n красных шаров. Из них нужно выбрать k любых шаров. Это можно сделать C_{n+m}^k способами. Рассмотрим произвольную выборку, в которой s белых шаров ($s \geq 0$). Способов получить такую выборку C_m^s , и для каждого этого способа оставшиеся $k - s$ красных шаров можно выбрать C_n^{k-s} способами. Поэтому всего способов выбрать k шаров из $m + n$ при фиксированном числе белых шаров s равно $C_m^s C_n^{k-s}$. Остается просуммировать по всем возможным значениям $s = 0, 1, \dots, k$. ■

Следствие 1.24.

$$C_{2n}^n = \sum_{s=0}^n (C_n^s)^2.$$

Доказательство. Достаточно применить теорему 1.64 при $m = k$ и $n = k$. ■

Приведем еще несколько свойств биномиальных коэффициентов для самостоятельного доказательства.

Теорема 1.65. Справедливы соотношения:

- 1) $C_n^k C_k^m = C_n^m C_{n-m}^{k-m}$;
- 2) $\sum_{n=0}^m n C_m^n = m 2^{m-1}$.

1.6.4. Кодирование с исправлением ошибок. Граница Хемминга

Определение 1.43. Неотрицательную вещественную функцию

$$d(a, b) : M \times M \rightarrow \mathbb{R}_+$$

называют *функцией расстояния* (или *метрикой*) на множестве M , если выполнены следующие условия (аксиомы метрики):

- 1) $d(a, b) \geq 0$, $d(a, b) = 0$ тогда и только тогда, когда $a = b$;
- 2) $d(a, b) = d(b, a)$;
- 3) $d(a, b) + d(b, c) \geq d(a, c)$.

Определение 1.44. Пусть на каждый кодовый символ a выделено m двоичных разрядов,

$$a = (a_1, a_2, \dots, a_m), \quad a_i = 0, 1, \quad i \in 1 : m.$$

Определим бинарную операцию на множестве кодовых символов:

$$a \oplus b = c = (c_1, c_2, \dots, c_m), \quad \text{где } c_i = a_i + b_i \bmod 2.$$

Введем понятие расстояния между кодовыми символами: $d(a, b)$ — число позиций $i : a_i \neq b_i$. Функцию d называют *кодовым расстоянием Хемминга*.

Пример 1.78. $a = (0, 1, 1, 1)$, $b = (1, 0, 1, 0)$, тогда $a \oplus b = (1, 1, 0, 1)$, $d(a, b) = 3$.

Очевидно следующее утверждение.

Лемма 1.12. Введенная функция $d(a, b)$ является функцией расстояния.

Доказательство. Условия 1, 2 очевидны. Установим условие 3. Если $a_i = c_i$, то $d(a, c)$ не меняется, а выражение в левой части неравенства может возрасти. Если же $a_i \neq c_i$, то либо $a_i \neq b_i$, либо $b_i \neq c_i$. ■

Предположим, что при приеме кодируемых символов возможно появление ошибок не более, чем в $r \leq m$ разрядах. Поставим задачу исправления ошибок в разрядах за счет уменьшения числа кодовых символов. Пусть $A = \{a^{(1)}, a^{(2)}, \dots, a^{(p)}\}$ — множество выбранных кодовых символов, $p \leq 2^m$. Справедлива следующая теорема.

Теорема 1.66. Если $d(a^{(i)}, a^{(j)}) \geq 2r + 1$, то можно исправлять ошибки, возникающие не более, чем в r разрядах.

Доказательство. Пусть при приеме символа $a^{(i)}$ не более чем в r разрядах возникли ошибки. Тогда принятый символ b попадает в окрестность символа $a^{(i)}$ радиуса r , т. е. справедливо $d(a^{(i)}, b) \leq r$. Обозначим соответствующую окрестность $A_i = \{a | d(a^{(i)}, a) \leq r\}$. Тогда $A_i \cap A_j = \emptyset$. Действительно, пусть $w \in A_i \cap A_j$, тогда

$$d(a^{(i)}, w) \leq r, \quad d(a^{(j)}, w) \leq r, \quad d(a^{(i)}, a^{(j)}) \leq d(a^{(i)}, w) + d(a^{(j)}, w) \leq 2r.$$

Полученное противоречие и доказывает утверждение. ■

Определение 1.45. Множество A_i назовем *областью декодирования* символа $a^{(i)}$. Любой принятый символ b , попадающий в A_i , отождествляем с $a^{(i)}$.

Сколько элементов в множестве A_i ? Это число способов выбора некоторых k ($0 \leq k \leq r$) компонент кодового символа $a^{(i)}$, т. е. число символов, отличающихся от символа $a^{(i)}$ в $0, 1, 2, \dots, r$ разрядах. Тогда

$$|A_i| = C_m^0 + C_m^1 + \cdots + C_m^r = \sum_{k=0}^r C_m^k.$$

Теперь легко оценить число элементов в множестве выбранных символов $A = \{a^{(1)}, a^{(2)}, \dots, a^{(p)}\}$. Так как всего 2^m кодируемых символов, то для p получаем оценку (*граница Хемминга*):

$$p \sum_{k=0}^r C_m^k \leq 2^m, \quad p \leq \frac{2^m}{\sum_{k=0}^r C_m^k}. \quad (1.85)$$

Пример 1.79. Пусть $m = 3$, $r = 1$. Тогда, согласно (1.85),

$$p \leq \frac{2^3}{(C_3^0 + C_3^1)} = \frac{8}{1+3} = 2.$$

В качестве двух кодовых символов, удовлетворяющих условию теоремы 1.66, можно взять следующие:

$$a^{(0)} = (0, 0, 0), \quad a^{(1)} = (1, 1, 1),$$

при этом

$$\begin{aligned} A_0 &= \{(0, 0, 1), (0, 1, 0), (1, 0, 0), (0, 0, 0)\}, \\ A_1 &= \{(1, 1, 0), (1, 0, 1), (0, 1, 1), (1, 1, 1)\}. \end{aligned}$$

На рис. 1.9 элементы области декодирования A_0 выделены черным цветом, а элементы области декодирования A_1 — белым.

Замечание 1.61. В общем случае вопрос о способе выбора элементов $a^{(i)}$, удовлетворяющих условию теоремы 1.66, является достаточно сложным и остается вне рамок данной книги. Подробно этот вопрос изучен в [5].

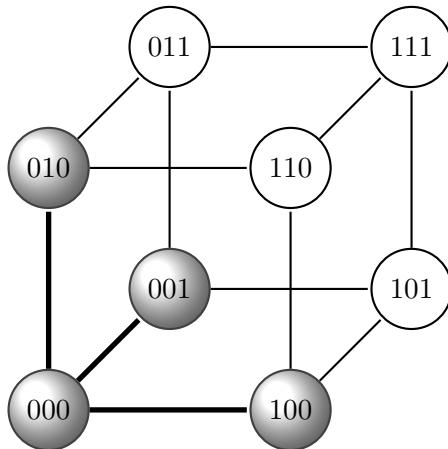


Рис. 1.9

1.6.5. Полиномиальное кодирование

Будем решать задачу предыдущего подраздела (см. пример 1.79), не уменьшая количество кодовых слов, согласно границе Хемминга, а увеличивая разрядность кода.

Определение 1.46. Код называют *линейным*, если для любых двух кодовых слов a и b их сумма $a \oplus b$, введенная в определении 1.44, также является кодовым словом и кодовым словом является λa — произведение кодового слова a на элемент поля $\lambda \in \{0, 1\}$.

Определение 1.47. Вес Хемминга $w(a)$ кодового слова a есть число его ненулевых компонент. Минимальным кодовым расстоянием кода называют величину d^*

$$d^* = \min_{a \neq b} d(a, b),$$

где d — расстояние Хемминга (см. определение 1.44) и минимум берется по всем кодовым словам.

Из определений 1.46 и 1.47 очевидно вытекает следующее утверждение.

Теорема 1.67. Для линейного кода минимальное кодовое расстояние d^* определяется из равенства

$$d^* = \min_{c \neq 0} w(c),$$

где минимум берется по всем кодовым словам, кроме нулевого.

Пример 1.80. Код из примера 1.79 является линейным и $d^* = 3$.

Определение 1.48. Совершенный или плотно-упакованный код есть код, для которого сферы (окрестности) одинакового радиуса вокруг кодовых слов не пересекаясь покрывают пространство всех двоичных векторов

длины m . Совершенный код удовлетворяет границе Хемминга (1.85) с равенством.

Пример 1.81. Заметим, что код из примера 1.79 — совершенный. Однако на практике крайне редко удается построение совершенных кодов.

Определение 1.49. Линейный код называют *циклическим*, если для любого кодового слова $\{a_1, a_2, \dots, a_m\}$ циклическая перестановка символов $\{a_m, a_1, \dots, a_{m-1}\}$ также дает кодовое слово.

Пример 1.82. Код из примера 1.79 — циклический.

Для построения линейных циклических кодов потребуется перейти от векторного описания кодов к полиномиальному. Последовательность символов основного алфавита $\{0, 1\}$, составляющих сообщения и кодовые слова будем интерпретировать как коэффициенты полиномов. Например, считая, что коэффициенты записаны в *порядке возрастания степени*, кодовое слово $(1, 0, 1, 0)$ запишем в виде многочлена $1 + x^2$. Кодирование сообщения в более «длинное» кодовое слово будет проводиться умножением этого многочлена на другой, что даст в результате многочлен более высокой степени.

Замечательным свойством полиномиального представления кодов является возможность осуществить циклический сдвиг на одну позицию вправо простым умножением кодового многочлена степени $m - 1$ на многочлен x и нахождения остатка от деления на $x^m + 1$.

Пример 1.83. Осуществить единичный правый циклический сдвиг кодового слова $(1, 0, 1, 1)$, используя полиномиальную интерпретацию.

Вектору $(1, 0, 1, 1)$ соответствует полином $1 + x^2 + x^3$. Умножим его на x , что дает $x + x^3 + x^4$. Далее, найдем остаток от деления на $x^4 + 1$, т. е. $x + x^3 + x^4 \pmod{x^4 + 1}$. Используя алгоритм 1.21 или алгоритм 1.20, получаем

$$x + x^3 + x^4 = (x^4 + 1) \cdot 1 + (1 + x + x^3).$$

Многочлен остатка $1 + x + x^3$ соответствует вектору $(1, 1, 0, 1)$, который получается из $(1, 0, 1, 1)$ правым циклическим сдвигом на одну позицию.

Процедура кодирования в полиномиальной интерпретации сводится к умножению многочлена-сообщения на подходящий многочлен, называемый *порождающим многочленом* данного кода. Очевидно, что не любой многочлен может служить порождающим.

Пример 1.84. Многочлен $G(x) = x$ не является порождающим. Действительно, все кодовые слова, полученные при умножении на $G(x)$ имеют вид $(0, a_2, a_3, a_4)$ и не удовлетворяют требованию цикличности.

Справедлива следующая теорема¹.

¹Доказательство можно найти в [5].

Теорема 1.68. Многочлен $G(x)$ является порождающим многочленом линейного циклического кода длины n тогда и только тогда, когда $g(x)$ делит $1 + x^n$.

Следствие 1.25. Если $G(x)$ — порождающий многочлен, то его свободный член $g_0 \neq 0$.

Доказательство. Если $g_0 = 0$, то многочлен x — делитель $G(x)$ и, следовательно, делитель $1 + x^n$. Противоречие. ■

Из теоремы 1.68 следует, что для получения порождающего многочлена $G(x)$ необходимо разложить на множители $x^n + 1$ и выделить многочлен такой степени, которая соответствует длине кодового слова. Длина кодового слова n и длина исходного сообщения m связаны соотношением

$$n = m + \deg(G(x)).$$

Пусть $C(x)$ обозначает переданное кодовое слово. Это значит, что символами переданного слова были коэффициенты многочлена $C(x)$. Пусть многочлен $V(x)$ обозначает принятое кодовое слово и пусть

$$E(x) = V(x) - C(x). \quad (1.86)$$

Многочлен $E(x)$ в (1.86) называют *многочленом ошибок*. Ненулевые коэффициенты этого многочлена стоят на тех позициях, где в канале произошли ошибки. Определим также понятие *синдромного многочлена* $S(x)$ как остаток от деления $V(x)$ на $G(x)$

$$S(x) = \langle V(x) \rangle_{G(x)} = \langle E(x) \rangle_{G(x)}. \quad (1.87)$$

Очевидно, что если кодовое слово было передано без помех, то синдром равен нулю. При некоторых условиях на код можно установить взаимно-однозначное соответствие между неизвестным многочленом ошибки и вычисляемым на приеме синдромным многочленом. Имеет место теорема.

Теорема 1.69. Пусть d^* — минимальное кодовое расстояние линейного циклического кода. Тогда каждому многочлену ошибок с весом меньше $d^*/2$ соответствует единственный синдромный многочлен.

Доказательство. Предположим, что двум многочленам ошибок $E_1(x)$ и $E_2(x)$ с весами меньшими $d^*/2$ соответствует один и тот же синдромный многочлен $S(x)$. Тогда, согласно (1.86) и (1.87), и учитывая, что $C(x)$ по определению делится на $G(x)$, получаем представления

$$E_1(x) = Q_1(x)G(x) + S(x), \quad E_2(x) = Q_2(x)G(x) + S(x).$$

Тогда

$$E_1(x) - E_2(x) = (Q_1(x) - Q_2(x))G(x). \quad (1.88)$$

По предположению, вес каждого из многочленов E_1 и E_2 меньше $d^*/2$, поэтому вес их разности меньше d^* . С другой стороны, в правой части (1.88) стоит по определению некоторое кодовое слово, следовательно, если $Q_1 \neq Q_2$, то вес правой части не менее d^* . Полученное противоречие показывает, что Q_1 и Q_2 равны и, следовательно, равны E_1 и E_2 . ■

Таким образом, задача исправления ошибок сводится к вычислению многочлена ошибки $E(x)$ с наименьшим весом, удовлетворяющего условию

$$S(x) = \langle E(x) \rangle_{G(x)}.$$

При небольших длинах кода, согласно теореме 1.69 для каждого многочлена ошибки вычисляется и табулируется соответствующий синдромный многочлен. Такая таблица носит название *таблицы синдромов*. Вычисляя синдром $S(x)$ по принятому кодовому слову $V(x)$, декодер находит в таблице соответствующий многочлен $E(x)$. При больших длинах кодов такой подход неприменим. Далее, в примере 1.86 будет рассмотрен более эффективный алгоритм декодирования.

Найдем все порождающие многочлены для циклического кода длины n . Согласно теореме 1.68, нужно разложить многочлен $1 + x^n$ на простые множители:

$$1 + x^n = F_1(x) F_2(x) \cdots F_l(x), \quad (1.89)$$

где l — число простых сомножителей. Произведение произвольного подмножества этих множителей даст порождающий многочлен $G(x)$. Если все простые множители в разложении (1.89) различны, то всего имеется $2^l - 2$ различных нетривиальных циклических кодов длины n (исключаются тривиальные случаи $G(x) = 1 + x^n$ и $G(x) = 1$). Вопрос о том, какие из них дают коды с большим минимальным расстоянием d^* решается достаточно сложно и остается вне рамок данной книги¹.

Пример 1.85. Найти порождающий многочлен линейного циклического кода длины $n = 15$, который осуществляет кодирование сообщений длины $m = 7$. Затем кодировать сообщение $(0, 1, 1, 0, 1, 1, 0)$.

Для построения требуемого порождающего многочлена нужно найти делитель $x^{15} + 1$ степени $15 - 7 = 8$. Имеем

$$x^{15} + 1 = (1 + x)(1 + x + x^2)(1 + x + x^2 + x^3 + x^4)(1 + x + x^4)(1 + x^3 + x^4),$$

поэтому можно взять

$$G(x) = (1 + x + x^2 + x^3 + x^4)(1 + x + x^4) = 1 + x^4 + x^6 + x^7 + x^8.$$

¹Подобный анализ подробно проведен в монографии [5].

Так как многочлен $G(x)$ сам является кодовым словом ($G(x) = G(x) \cdot 1$) и $w(G(x)) = 5$, то по теореме 1.67 $d^* \leq 5$. На самом деле в данном случае имеет место равенство. Таким образом, согласно теореме 1.69 данный код исправляет две ошибки.

В полиномиальной интерпретации вектору $(0, 1, 1, 0, 1, 1, 0)$ соответствует многочлен $x + x^2 + x^4 + x^5$. Умножим его на порождающий многочлен $G(x)$. Имеем

$$(x + x^2 + x^4 + x^5)(1 + x^4 + x^6 + x^7 + x^8) = x + x^2 + x^4 + x^6 + x^7 + x^8 + x^9 + x^{13}.$$

Соответствующее кодовое слово $(0, 1, 1, 0, 1, 0, 1, 1, 1, 0, 0, 0, 1, 0)$.

Рассмотрим более эффективный алгоритм декодирования. В нем выполняется циклический сдвиг принятого кодового слова пока вес остатка от деления не будет равен известному синдрому. Подробное обоснование можно найти также в [5].

Входные данные: многочлен $V(x)$ — принятое слово, $G(x)$ — порождающий многочлен кода, максимальное число исправляемых кодом ошибок равно t .

На выходе получаем исправленное слово $V(x)$ или отказ в декодировании.

Алгоритм 1.26 (полиномиальное декодирование).

```

 $S(x) \leftarrow \langle V(x) \rangle_{G(x)}$                                 ▷ находим синдромный многочлен
for  $i \leftarrow 0, n - 1$  do
     $S_i(x) \leftarrow \langle x^i S(x) \rangle_{G(x)}$ 
    if  $w(S_i) \leq t$  then
         $E(x) \leftarrow \langle x^{n-i} S_i(x) \rangle_{1+x^n}$                 ▷ Вычисляем полином ошибки
        Производим исправления в  $V(x)$  с учетом  $E(x)$ 
        Декодирование успешно завершено
    end if
end for

```

В принятом слове более t ошибок, декодирование **невозможно**

Пример 1.86. Декодируем слово $(0, 1, 1, 0, 1, 0, 1, 1, 1, 0, 1, 0, 0, 1, 0)$, которое было отправлено после кодирования кодом из примера 1.85. Соответствующий многочлен есть $x + x^2 + x^4 + x^6 + x^7 + x^8 + x^{10} + x^{13}$.

Согласно алгоритму 1.26 находим многочлен-синдром

$$S(x) = \langle x + x^2 + x^4 + x^6 + x^7 + x^8 + x^{10} + x^{13} \rangle_{1+x^4+x^6+x^7+x^8} = 1 + x^2 + x^4 + x^7.$$

Для кодового слова синдром, как известно, равен 0. В данном случае это не так, посланное слово было искажено помехой. В соответствии с описанной процедурой декодирования будем вычислять

$$S_i(x) = \langle x^i (1 + x^2 + x^4 + x^7) \rangle_{G(x)}$$

для последовательных возрастающих значений $i = 0, 1, \dots, n - 1$ пока не найдем многочлен степени меньшей или равной двум (число ошибок $t = 2$). Имеем

$$\begin{aligned} S_1 &= \langle xs(x) \rangle_{G(x)} = \langle x + x^3 + x^5 + x^8 \rangle_{G(x)} = 1 + x + x^3 + x^4 + x^5 + x^6 + x^7, \\ S_2 &= \langle x^2 s(x) \rangle_{G(x)} = \langle x^2 + x^4 + x^6 + x^9 \rangle_{G(x)} = 1 + x + x^2 + x^5, \\ S_3 &= \langle x^3 s(x) \rangle_{G(x)} = \langle x^3 + x^5 + x^7 + x^{10} \rangle_{G(x)} = x + x^2 + x^3 + x^6, \\ S_4 &= \langle x^4 s(x) \rangle_{G(x)} = \langle x^4 + x^6 + x^8 + x^{11} \rangle_{G(x)} = x^2 + x^3 + x^4 + x^7, \\ S_5 &= \langle x^5 s(x) \rangle_{G(x)} = \langle x^5 + x^7 + x^9 + x^{12} \rangle_{G(x)} = 1 + x^3 + x^5 + x^6 + x^7, \\ S_6 &= \langle x^6 s(x) \rangle_{G(x)} = \langle x^6 + x^8 + x^{10} + x^{13} \rangle_{G(x)} = 1 + x. \end{aligned}$$

Многочлен $x + 1$ имеет вес 2, поэтому для нахождения многочлена ошибок вычисляем

$$E(x) = \langle x^{15-6} s_6(x) \rangle_{1+x^n} = \langle x^9(x+1) \rangle_{1+x^{15}} = x^{10} + x^9.$$

Итак, если отправленное кодовое слово имеет не более двух ошибок, то оно было таким

$$(x+x^2+x^4+x^6+x^7+x^8+x^{10}+x^{13})+(x^{10}+x^9)=x+x^2+x^4+x^6+x^7+x^8+x^9+x^{13}.$$

Этот многочлен соответствует слову $(0, 1, 1, 0, 1, 0, 1, 1, 1, 1, 0, 0, 0, 1, 0)$. Чтобы восстановить само сообщение, надо разделить кодовое слово на порождающий многочлен $G(x)$ и получить $x + x^2 + x^4 + x^5$, значит отправленное сообщение было $(0, 1, 1, 0, 1, 1, 0)$.

Справедлива теорема.

Теорема 1.70. Если порождающий многочлен $G(x)$ делится на $1 + x$, то любое кодовое слово имеет четное число ненулевых коэффициентов.

Доказательство. Любое кодовое слово $C(x)$ получается умножением исходного слова $A(x)$ на $G(x)$. Тогда

$$C(x) = c_0 + c_1 x + \cdots + c_{n-1} x^{n-1} = A(x)G(x) = (1 + x)H(x). \quad (1.90)$$

Подставив в (1.90) $x = 1$, получим $c_0 + c_1 + \cdots + c_{n-1} = (1 + 1)H(1) = 0$. ■

Это есть *код с проверкой на четность*, обнаруживающий любое нечетное число ошибок передачи. Все кодовые слова — «четные».

1.6.6. Код Шеннона–Фэно и алгоритм Хаффмена

В предыдущих подразделах на каждый кодовый символ отводилось одинаковое число двоичных разрядов. К. Шеннон и Фэно предложили конструкцию *кода переменной длины*. При таком коде у каждого символа своя длина кодовой последовательности, поэтому необходимо побеспокоиться о том, как определять конец кода отдельного символа. Предлагается такое ограничение на код: *никакая кодовая последовательность не является началом другой кодовой последовательности*. Это свойство называют *свойством префикса*, а код, обладающий таким свойством, — *беспрефиксным кодом*.

Пусть $\{p_1, p_2, \dots, p_n\}$ — частоты появления (использования) кодируемых символов, а $\{s_1, s_2, \dots, s_n\}$ — длины их кодовых последовательностей. Будем стремиться уменьшить среднее число двоичных разрядов, приходящихся на один кодовый символ, т. е. величину $\sum_{i=1}^n p_i s_i$.

К. Шеннон и Фэно предложили метод построения кода, близкого к оптимальному: разбить все символы на две группы с приблизительно одинаковой суммарной частотой появления, коды первой группы начать с нуля, а второй — с единицы. Внутри группы делать то же самое, пока в каждой группе не останется только по одному символу.

Элегантный алгоритм для точного решения этой задачи, основанный на нескольких очевидных свойствах оптимальной кодовой последовательности, предложил Д. Хаффмен.

Очевидно следующее утверждение.

Лемма 1.13. Если $p_1 \geq p_2 \geq \dots \geq p_n$, то $s_1 \leq s_2 \leq \dots \leq s_n$.

Какими свойствами должен обладать оптимальный код? Посмотрим на самое длинное кодовое слово. Оно должно быть парным, то есть слово, которое отличается в последнем бите, тоже должно быть кодовым словом. Например, если мы какую-то букву кодируем как $b_0 \dots b_n 0$, а как $b_0 \dots b_n 1$ ничего не кодируем, то почему же мы не используем просто битовую последовательность $b_0 \dots b_n$? Ведь получится тоже беспрефиксный код, и будет короче.

Таким образом справедлива теорема.

Теорема 1.71. В предположениях леммы 1.13 справедливо $s_{n-1} = s_n$.

Опишем теперь собственно алгоритм: два символа кодируем 0 и 1, а если символов больше, то соединяем два самых редких символа в один новый символ, решаем получившуюся задачу, а затем вновь разделяем этот новый символ на два, приписав соответственно 0 и 1 к его кодовой последовательности.

Пример 1.87. Пусть имеется семь кодовых символов a, b, c, d, e, f, g с частотами появления соответственно

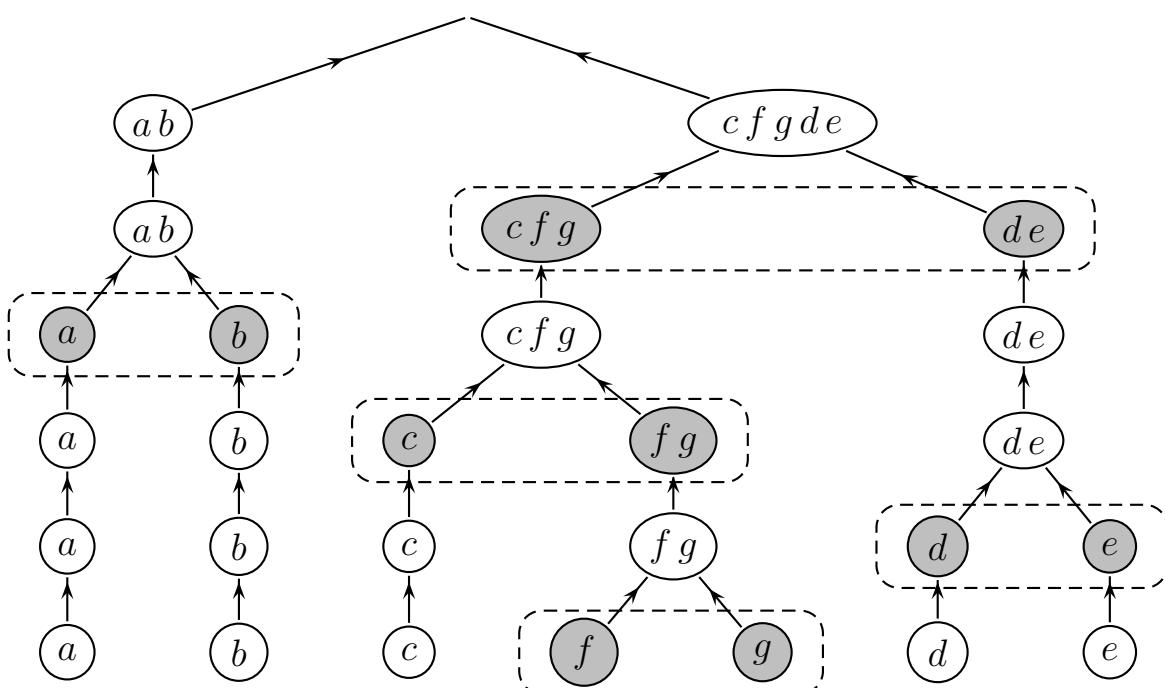
0.22(*a*), 0.20(*b*), 0.15(*c*), 0.13(*d*), 0.12(*e*), 0.10(*f*), 0.08(*g*).

С помощью алгоритма Хаффмена построить код Шеннона–Фэно и зашифровать сообщение: [beda].

Воспользуемся алгоритмом Хаффмена. Проводя объединение получаем:

- 1) 0.22(a), 0.20(b), 0.18(fg), 0.15(c), 0.13(d), 0.12(e);
 - 2) 0.25(de), 0.22(a), 0.20(b), 0.18(fg), 0.15(c);
 - 3) 0.33(cfg), 0.25(de), 0.22(a), 0.20(b);
 - 4) 0.42(ab), 0.33(cfg), 0.25(de);
 - 5) 0.58($cfgde$), 0.42(ab).

Последовательность объединения схематично проиллюстрирована на рис. 1.10. Алгоритм работает «снизу вверх» (движение по стрелкам). Объединяемые наборы символов отмечены серым цветом и пунктирными линиями.



Puc. 1.10

Расщепляем символы и приписываем 0 и 1 справа к кодовым последовательностям:

- 1) $0(ab), \ 1(cfgde);$
 - 2) $0(ab), \ 11(cfg), \ 10(de);$

- 3) 01(a), 00(b), 11(cfg), 10(de);
- 4) 01(a), 00(b), 110(c), 111(fg), 10(de);
- 5) 01(a), 00(b), 110(c), 111(fg), 101(d), 100(e);
- 6) 01(a), 00(b), 110(c), 1111(f), 1110(g), 101(d), 100(e).

Таким образом кодовая таблица имеет вид (табл. 1.37):

Таблица 1.37

<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>
01	00	110	101	100	1111	1110

Шифруем исходное сообщение: 0010010101. Согласно кодовой табл. 1.37 легко произвести дешифровку.

Замечание 1.62. Рассмотренный алгоритм является примером *жадного алгоритма*. Жадный алгоритм (англ. *Greedy algorithm*) — это алгоритм, заключающийся в принятии локально оптимальных решений на каждом этапе, допуская, что конечное решение также окажется оптимальным.

1.6.7. Лексикографический порядок. Генерирование k -элементных подмножеств

Рассмотрим задачу генерации всех k -элементных подмножеств данного n -элементного множества $X = \{x_1, \dots, x_n\}$. Очевидно, что всего их C_n^k . Не умаляя общности, будем считать, что $X = \{1, \dots, n\}$. Определим упорядоченность k -элементных подмножеств на основе упорядоченности элементов.

Определение 1.50. Рассмотрим два k -элементных подмножества $A = \{x_1, \dots, x_k\}$ и $B = \{y_1, \dots, y_k\}$. Говорят, что

1) подмножество A лексикографически предшествует подмножеству B , если найдется индекс $m \leq k$ такой, что

$$x_i = y_i, \quad x_m < y_m, \quad 1 \leq i < m; \quad (1.91)$$

2) подмножество A антилексикографически предшествует подмножеству B , если существует индекс $m \leq k$ такой, что

$$x_i = y_i, \quad x_m > y_m, \quad m < i \leq k.$$

Если в (1.91) выполнено $x_m > y_m$, то такой порядок называют *обратным лексикографическим*.

Замечание 1.63. В общем случае будем считать, что для каждого индекса $1 \leq i \leq k$ определено сравнение элементов x_i и y_i . Это позволяет лексикографически сравнивать последовательности, состоящие из элементов разных типов.

Пример 1.88. Если в качестве элементов X взять буквы алфавита, то слова длины k появляются в словаре в лексикографическом порядке.

В качестве примера построим алгоритм генерации k -элементных подмножеств X в лексикографическом порядке. Проведем некоторые предварительные рассуждения. Очевидно, что первое подмножество — это $\{1, 2, 3, \dots, k\}$, а за подмножеством $\{a_1, a_2, \dots, a_k\}$ следует подмножество

$$\{b_1, b_2, \dots, b_k\} = \{a_1, a_2, \dots, a_{p-1}, a_p + 1, a_p + 2, \dots, a_p + k - p + 1\},$$

где p — максимальный индекс, для которого $b_k = a_p + k - p + 1 \leq n$.

Очередное подмножество получается из предыдущего заменой последних элементов на идущие подряд целые числа так, чтобы последний элемент не превосходил n , а первый изменяемый элемент был на 1 больше, чем соответствующий элемент предыдущего подмножества. Таким образом, индекс p , начиная с которого следует проводить изменения, определяется по значению a_k . Если $a_k < n$, то следует изменить только a_k , и при этом $p = k$. Если же $a_k = n$, то нужно уменьшать индекс $p = p - 1$, увеличивая тем самым длину изменяемой части подмножества.

Запишем теперь сам алгоритм.

Алгоритм 1.27 (генерация k -элементных подмножеств).

```

for  $i \leftarrow 1, k$  do ▷ первое подмножество
     $a_i \leftarrow i$ 
end for
 $p \leftarrow k$ 
while  $p \geq 1$  do
    if  $a_k = n$  then
         $p \leftarrow p - 1$ 
    else
         $p \leftarrow k$ 
    end if
    if  $p \geq 1$  then ▷  $p$  могло стать меньше 1 на предыдущем шаге
        for  $i \leftarrow p, k$  do
             $a_i \leftarrow a_p + i - p + 1$ 
        end for
    end if
end while

```

Пример 1.89. Пусть $X = \{1, 2, 3, 4, 5, 6\}$, $k = 4$. Протокол работы алгоритма 1.27 дан в табл. 1.38.

Таблица 1.38

Шаг	p	Текущее разбиение	Шаг	p	Текущее разбиение
1	4	{1, 2, 3, 4}	9	3	{1, 3, 5, 6}
2	4	{1, 2, 3, 5}	10	2	{1, 4, 5, 6}
3	4	{1, 2, 3, 6}	11	1	{2, 3, 4, 5}
4	3	{1, 2, 4, 5}	12	4	{2, 3, 4, 6}
5	4	{1, 2, 4, 6}	13	3	{2, 3, 5, 6}
6	3	{1, 2, 5, 6}	14	2	{2, 4, 5, 6}
7	2	{1, 3, 4, 5}	15	1	{3, 4, 5, 6}
8	4	{1, 3, 4, 6}		0	Генерация закончена

1.6.8. Перечислительная комбинаторика

Рассмотрим комбинаторные задачи следующего вида: нужно пронумеровать все комбинации с заданным свойством и найти алгоритм, который по заданному номеру будет строить соответствующую комбинацию и алгоритм, который будет решать обратную задачу — по заданной комбинации вычислять ее номер. Эту же задачу можно назвать задачей перечисления всех комбинаций. В предыдущем подразделе был приведен частный случай решения такой задачи (алгоритм 1.27). Разберем общие приемы, позволяющие перечислять подмножества данного множества, элементы декартова произведения, перестановки, сочетания.

□ Перечисление подмножеств

Пусть задано множество, содержащее n элементов: $M = \{m_1, \dots, m_n\}$.

Требуется найти алгоритм, генерирующий все подмножества данного множества.

Для решения задачи перечисления подмножеств используем часто применяемый в комбинаторике прием — конструирование взаимно-однозначного соответствия для того, чтобы свести одну комбинаторную задачу к другой.

Рассмотрим двоичные наборы, содержащие n цифр. Каждый такой набор можно интерпретировать как шифр некоторого подмножества: Единица на k -ом месте означает, что элемент m_k включается в соответствующее подмножество, а ноль — что элемент не включается.

Пример 1.90. У множества из трех элементов $M = \{a, b, c\}$ имеется восемь подмножеств, два из них — пустое подмножество и само

множество M , и еще шесть так называемых собственных подмножеств $\{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}$.

Указанные собственные подмножества множества из трех элементов шифруются так: 100, 010, 001, 110, 101, 011. Пустое подмножество имеет шифр 000, а само множество шифр 111. Если эти наборы интерпретировать как двоичные записи чисел, то получим соответственно 4, 2, 1, 6, 5, 3 и 0, 7, т. е. все целые числа от 0 до 7. Это дает идею решения всех поставленных задач.

Число подмножеств множества из n элементов равно 2^n . Генерация подмножеств делается так: надо, начиная с $i = 0$, переводить i в двоичную запись, которая определяет очередное подмножество, а затем увеличивать i на единицу вплоть до $2^n - 1$.

Алгоритм 1.28 (генерация всех подмножеств n -элементного множества).

```

for  $i \leftarrow 1, 2^n - 1$  do
    перевести  $i$  в двоичную запись
    дополнить полученную запись спереди нулями так, чтобы число цифр стало равным  $n$ 
    построить по двоичному набору подмножество
end for

```

Замечание 1.64. При перечислении можно оперировать двоичной арифметикой, тогда перевод номера подмножества в двоичную запись станет излишним.

Нахождение десятичного номера подмножества осуществляется алгоритмом перевода числа из двоичной системы в десятичную, а нахождение двоичного набора, описывающего данное подмножество, по его десятичному номеру — переводом из десятичной в двоичную.

Задача перечисления подмножеств станет более интересной, если наложить дополнительное условие на их порядок. Например, естественным ограничением будет такое: два последовательных подмножества должны отличаться ровно одним элементом.

Пример 1.91. На столе лежат четыре разноцветных шара. За один ход можно взять или положить ровно один шар. Как за наименьшее число ходов перебрать все возможные подмножества этих шаров.

Понятно, что если найти алгоритм, позволяющий после каждого взятия или добавления шара получать новое подмножество, то такая стратегия даст наименьшее число ходов. Поскольку всего подмножеств (включая пустое) $2^4 = 16$, то достаточно будет 15 ходов (если учесть, что уже до первого хода одно из подмножеств — само множество — уже представлено).

Приведем протокол решения этой задачи (табл. 1.39), сравнивая двоичные наборы, определяющие текущие подмножества, с теми, которые дает предыдущий алгоритм (поскольку мы начали перебор не с пустого подмножества, а, наоборот, с самого множества, будем обозначать нулем оставленный шар, а единицей — взятый).

Таблица 1.39

Номера шаров, лежащих на столе	Комментарий к очередному действию	Двоичный код подмножества шаров B	Двоичный номер шага A	Комментарий к правилу изменения B по изменению A
1234	Исходная комбинация	0000	0000	
123	Сняли 4-й шар	0001	0001	В 4-м разряде 0 заменили на 1
12	Сняли 3-й шар	0011	0010	В 3-м разряде 0 заменили на 1
124	Добавили 4-й шар	0010	0011	В 4-м разряде 0 заменили на 1
14	Сняли 2-й шар	0110	0100	Во 2-м разряде 0 заменили на 1
1	Сняли 4-й шар	0111	0101	В 4-м разряде 0 заменили на 1
13	Добавили 3-й шар	0101	0110	В 3-м разряде 0 заменили на 1
134	Добавили 4-й шар	0100	0111	В 4-м разряде 0 заменили на 1
34	Сняли 1-й шар	1100	1000	В 1-м разряде 0 заменили на 1
3	Сняли 4-й шар	1101	1001	В 4-м разряде 0 заменили на 1
	Сняли 3-й шар	1111	1010	В 3-м разряде 0 заменили на 1
4	Добавили 4-й шар	1100	1011	В 4-м разряде 0 заменили на 1
24	Добавили 2-й шар	1010	1100	Во 2-м разряде 0 заменили на 1
2	Сняли 4-й шар	1011	1101	В 4-м разряде 0 заменили на 1
23	Добавили 3-й шар	1001	1110	В 3-м разряде 0 заменили на 1
234	Добавили 4-й шар	1000	1111	В 4-м разряде 0 заменили на 1

Можно заметить следующую закономерность, когда при обычном перечислении двоичными числами в k -м разряде 0 сменяет 1, то с k -м шаром совершают операцию перекладывания: если он лежал на столе — снимают, если его не было на столе — добавляют.

Пусть массив A_i представляет двоичные номера при обычном переборе, а B_i — при переборе с заданными выше ограничениями. Пусть в начальный момент оба массива заполнены нулями. Тогда, двигаясь сначала

справа налево по разрядам двоичного набора A_i , надо найти первый нулевой разряд, запомнить его номер k и записать в него 1 : $A_k := 1$, а все разряды правее k -го заполнить нулями.

Затем надо поменять k -й разряд в массиве B_i и вывести очередное подмножество. Все это надо повторять в цикле, пока в массиве A_i не будут стоять одни единицы. Таким образом, получаем следующий алгоритм.

Алгоритм 1.29 (генерация всех подмножеств n -элементного множества).

```

for  $i \leftarrow 1, n$  do
     $A_i \leftarrow 0$ 
     $B_i \leftarrow 0$ 
end for
repeat
     $k \leftarrow n$ 
    while  $A_k \neq 0$  do                                 $\triangleright$  ищем первый нулевой разряд
         $k \leftarrow k - 1$ 
    end while
    if  $k > 0$  then
         $A_k \leftarrow 1$ 
        for  $i \leftarrow k + 1, n$  do                   $\triangleright$  все разряды правее  $k$ -го заполняем нулями
             $A_i \leftarrow 0$ 
        end for
         $B_k \leftarrow B_k + 1 \pmod{2}$ 
        return  $B$                                  $\triangleright$  вывод очередного подмножества  $B$ 
    end if
until  $k = 0$ 

```

Для того чтобы обосновать корректность алгоритма 1.29, посмотрим на него с «рекурсивной» точки зрения. Действительно, анализируя пример, нетрудно заметить, что первые восемь подмножеств из шестнадцати содержат 1-й шар, оставшиеся восемь — нет. Так же каждую из этих восьмерок можно разделить на четверки по признаку, содержат ли они 2-й шар или нет (заметим, что, если в первой восьмерке этот шар содержит первые четыре подмножества, то во второй — последние четыре). Затем в каждой четверке можно сравнить двойки, которые будут отличаться наличием 3-го шара. И, наконец, каждая двойка разбивается на отдельные подмножества, которые отличаются наличием 4-го шара.

Получается, что любые два подмножества отличаются друг от друга. То, что два соседних (относительно порядка перебора) подмножества отличаются ровно одним элементом, объясняется так. Если два подмножества принадлежат одной «двойке», то они отличаются только 4-м шаром. Если они принадлежат одной четверке так, что одно является вторым для первой двойки, а второе — первым для второй, то они отличаются только 3-м шаром, так как 4-й шар будет в обоих подмножествах и т. д. То, что нумера разрядов, которые соответствуют перемещаемым шарам, получаются из алгоритма добавления 1 к двоичному числу, следует из того, что первые

восемь двоичных чисел имеют 0 в старшем разряде, а следующие восемь — 1. Аналогичное утверждение можно сделать про четверки и т. д.

Замечание 1.65. Более подробно вопросы, связанные с перечислением подмножеств в примере 1.91 и алгоритмом 1.29 будут рассмотрены в разделе 1.6.9.

Задача 1.17. Решает ли следующий алгоритм поставленную задачу о переборе всех подмножеств. Ответ обосновать.

Алгоритм 1.30 (генерация всех подмножеств n -элементного множества).

```

while число сгенерированных подмножеств меньше  $2^n$  do
    for  $i \leftarrow n, 1$  do
        if перемещение  $i$ -го шара не приводит к уже встречавшейся комбинации then
            переместить шар (снять или добавить)
        end if
    end for
end while

```

Пример 1.92. Для приведенного выше примера 1.91 последовательность подмножеств, генерируемая алгоритмом 1.30, получается такой:

1234, 123, 12, 1, \emptyset , 4, 34, 234, 23, 2, 24, 124, 14, 134, 13, 3.

□ Перечисление элементов декартова произведения множеств

Пусть нам нужно перечислить элементы множества $M = M_1 \times \dots \times M_n$, где

$$M_1 = \{a_1, a_2, \dots, a_{m_1}\}, M_2 = \{b_1, b_2, \dots, b_{m_2}\}, \dots, M_n = \{x_1, x_2, \dots, x_{m_n}\}.$$

Заметим, что в комбинаторике проще работать с номерами элементов, забывая о том какую природу имеют эти элементы (буквы, множества, фигуры и пр.). Так и будем делать в дальнейшем. Тогда

$$M = \{(i_1, i_2, \dots, i_n) \mid 0 \leq i_1 < m_1, 0 \leq i_2 < m_2, \dots, 0 \leq i_n < m_n\}. \quad (1.92)$$

Пример 1.93. Элемент декартова произведения, описываемый набором $(2, 0, \dots, 3)$ для обозначений (1.92), будет (a_3, b_1, \dots, x_4) .

В соответствии с определением наборы декартова произведения можно генерировать с помощью n вложенных циклов.

Алгоритм 1.31 (перечисление декартова произведения).

```

for  $i_1 \leftarrow 0, m_1 - 1$  do
    for  $i_2 \leftarrow 0, m_2 - 1$  do
        .....
        for  $i_n \leftarrow 0, m_n - 1$  do
            return набор  $(i_1, i_2, \dots, i_n)$ 
        end for

```

```
.....  
    end for  
end for
```

По этому алгоритму легко определить, каким по счету будет сгенерирован набор (i_1, i_2, \dots, i_n) .

$$N(i_1, i_2, \dots, i_n) = m_2 m_3 \cdots m_n i_1 + m_3 m_4 \cdots m_n i_2 + \cdots + m_n i_{n-1} + i_n.$$

Замечание 1.66. Номера $N(i_1, i_2, \dots, i_n)$ также будут начинаться с 0. Для привычной нумерации с 1 к полученному выражению надо добавить 1. Полученное выражение можно рассматривать как представление числа N в системе с переменным основанием. Перевод числа в такую систему — построение набора по его номеру — можно осуществить последовательными делениями с остатком.

Представим N следующим образом:

$$N(i_1, i_2, \dots, i_n) = (\cdots (i_1 m_2 + i_2) m_3 + i_3) m_4 + \cdots + i_{n-1}) m_n + i_n.$$

Тогда $i_n = N \pmod{m_n}$, а выражение в скобках перед m_n равно $N \div m_n$. Алгоритм вычисления всех элементов набора будет такой.

Алгоритм 1.32 (вычисление элементов набора (i_1, i_2, \dots, i_n)).

```
for k ← n, 1 do  
     $i_k \leftarrow N \pmod{m_k}$   
     $N \leftarrow N \div m_k$   
end for
```

Замечание 1.67. В алгоритме 1.32 предполагается знание длины набора n . Если перенумеровать элементы набора от последнего к первому, то значение n не понадобится. В следующем алгоритме этот прием использован.

На основе этого алгоритма алгоритм перечисления элементов декартова произведения можно записать, используя всего два вложенных цикла:

Алгоритм 1.33 (перечисление декартова произведения).

```
return (0, ..., 0)  
for N ← 1,  $m_1 m_2 \cdots m_n - 1$  do  
    k ← 1  
    R ← N  
    while R > 0 do  
         $i_k \leftarrow R \pmod{m_k}$   
        R ← R  $\div m_k$   
        k ← k + 1  
    end while  
    return ( $i_n, \dots, i_1$ )  
end for
```

□ Перечисление перестановок

Рассмотрим задачу нумерации перестановок и составим алгоритм, который по перестановке определяет ее номер, а также алгоритм, который по номеру определяет соответствующую перестановку.

Расположим наборы в лексикографическом порядке и пронумеруем их. Нумерацию будем начинать с 0.

Пример 1.94. Для $n = 4$ имеем $4! = 24$ перестановки (табл. 1.40).

Таблица 1.40

0	1234	6	2134	12	3124	18	4123
1	1243	7	2143	13	3142	19	4132
2	1324	8	2314	14	3214	20	4213
3	1342	9	2341	15	3241	21	4231
4	1423	10	2413	16	3412	22	4312
5	1432	11	2431	17	3421	23	4321

Построим сначала алгоритм перехода от одной перестановки к следующей за ней в лексикографическом порядке. Сравнив, например, переход от 5-й к 6-й перестановке в предыдущем примере, можно сформулировать такой алгоритм.

Алгоритм 1.34.

Находим упорядоченный по убыванию «остаток» перестановки наибольшей длины

if Если «остаток» не совпадает со всей перестановкой **then**

Элемент, предшествующий «остатку», поменяем с наименьшим элементом «остатка», *большим его*

Элементы «остатка» упорядочиваем в порядке возрастания

end if

Замечание 1.68. На шаге 2 алгоритма 1.34 в «остатке» всегда найдется элемент для обмена (больший, чем элемент, предшествующий «остатку») В противном случае, элемент, предшествующий «остатку», сам был членом «остатка».

Пример 1.95. Какая перестановка идет следующей за перестановкой 61287543?

Искомый «остаток» — 87543, его наименьший элемент — 3. Меняем его с элементом, предшествующим «остатку» — 2, и упорядочиваем «остаток» по возрастанию. Получаем 61324578.

Задача 1.18. Предложите алгоритм перехода к следующей перестановке для обратного лексикографического порядка.

Из приведенной таблицы для перечисления перестановок из четырех элементов в лексикографическом порядке (см. табл. 1.40) можно заметить и другие закономерности, например, что перестановки в разных столбиках (24 перестановки были разбиты на четыре столбика по 6 элементов) начинаются с разных цифр, а в одном — с одной.

Это напоминает систему счисления. Действительно, посчитаем, какой по счету идет перестановка (i_1, i_2, \dots, i_n) в лексикографическом порядке.

До этой перестановки стоят $(i_1 - 1)(n - 1)!$ перестановок, у которых первый элемент меньше (а остальные переставляются всеми $(n - 1)!$ возможными способами). Далее, перед этой перестановкой стоят все те перестановки, у которых первый элемент такой же, а второй меньше. Таких перестановок будет $(i_2 - 1)(n - 2)!$ либо $(i_2 - 2)(n - 2)!$, если первая цифра меньше второй. В общем случае полученный результат можно записать в форме следующего утверждения.

Лемма 1.14. Номер N перестановки (i_1, i_2, \dots, i_n) вычисляется по формуле:

$$N = (i_1 - 1)(n - 1)! + (i_2 - 1 - k_2)(n - 2)! + \dots + (i_{n-1} - 1 - k_{n-1})1! + i_n - 1 - k_n, \quad (1.93)$$

где k_j — число элементов, меньших i_j и стоящих левее его.

Пример 1.96. Какой по счету идет перестановка 3412?

$$N = 2 \times 3! + (3 - 1) \times 2! + 0 \times 1! + (1 - 1) = 12 + 4 + 0 + 0 = 16.$$

Формулу (1.93) можно записать изящнее, используя факториальную систему счисления.

Теорема 1.72. При лексикографическом упорядочении перестановок, номер перестановки (i_1, i_2, \dots, i_n) вычисляется по формуле

$$N = a_1(n - 1)! + a_2(n - 2)! + \dots + a_{n-1}1!, \quad (1.94)$$

где a_j — количество элементов, меньших i_j и стоящих правее его, причем $0 \leq a_j \leq n - j$.

Доказательство. В формуле (1.93) обозначим коэффициент перед факториалом $i_j - 1 - k_j$ через a_j . Его можно интерпретировать как количество элементов, меньших i_j и стоящих правее его. Действительно, всего существует $i_j - 1$ натуральное число, меньшее i_j , все они входят в перестановку, а k_j из них стоят левее. Значит оставшиеся $i_j - 1 - k_j$ элементов стоят правее. Неотрицательность a_j следует из определения. Так же объясняется и неравенство $a_j \leq n - j$: количество элементов меньших i_j и стоящих правее его не может быть больше $n - j$, так как равенство достигается, когда все элементы стоящие правее i_j , меньше его. Это и доказывает теорему. ■

Замечание 1.69. В более привычных для позиционных систем обозначениях можно записать:

$$N = b_{n-1}(n-1)! + b_{n-2}(n-2)! + \cdots + b_1 1! = (b_{n-1} b_{n-2} \dots b_1)_!,$$

где $b_j = a_{n-j}$.

Следствие 1.26. Если перестановки пронумерованы в лексикографическом порядке, то для отыскания перестановки по ее номеру надо представить номер в факториальной системе счисления, а затем восстановить перестановку, учитывая, что j -я цифра показывает число элементов, меньших j -го элемента и стоящих правее его.

Пример 1.97. Найти перестановку, определяемую номером в факториальной системе счисления $(2\ 3\ 0\ 2\ 1\ 1)_!$.

Первый элемент набора 2, значит, в перестановке правее него стоят два элемента, меньших его. Тогда искомый элемент — 3.

На втором месте стоит 3, значит, в перестановке правее него стоит три элемента, меньших его. Тогда искомый элемент — 5 (а не 4, так как 3 стоит на первом месте).

Весь ход решения отражен в табл. 1.41.

Таблица 1.41

Факториальное представление номера перестановки	Выбор элемента перестановки	Элементы перестановки
2 3 0 2 1 1	7 6 5 4 3 2 1	3
2 3 0 2 1 1	7 6 5 4 2 1	5
2 3 0 2 1 1	7 6 4 2 1	1
2 3 0 2 1 1	7 6 4 2	6
2 3 0 2 1 1	7 4 2	4
2 3 0 2 1 1	7 2	7
2 3 0 2 1 1 \emptyset	2	2

Ответ: 3516472.

Рассмотрим решение обратной задачи для примера 1.97 — по перестановке определить ее номер.

Пример 1.98. Данна перестановка 3516472. Определить ее номер N в факториальной системе счисления.

Согласно теореме 1.72 строим факториальную запись перестановки:

$$(2\ 3\ 0\ 2\ 1\ 1)_!.$$

Применяя формулу (1.94) и схему Горнера для факториальной записи, получаем

$$\begin{aligned} N &= 1 \cdot 1! + 1 \cdot 2! + 2 \cdot 3! + 0 \cdot 4! + 3 \cdot 5! + 2 \cdot 6! = \\ &= (((6 \cdot 2 + 3)5 + 0)4 + 2)3 + 1)2 + 1 = 1815. \end{aligned}$$

Рассмотрим еще один пример.

Пример 1.99. Найдем перестановку по ее номеру 3973. Получим из номера факториальную запись перестановки. Производя последовательно деления на $2, 3, \dots, n$ имеем

$$\begin{aligned} 3973 &= 1986 \cdot 2 + 1, \\ 1986 &= 662 \cdot 3 + 0, \\ 662 &= 165 \cdot 4 + 2, \\ 165 &= 33 \cdot 5 + 0, \\ 33 &= 5 \cdot 6 + 3, \\ 5 &= 0 \cdot 7 + 5. \end{aligned}$$

Таким образом, факториальная запись искомой перестановки 5 3 0 2 0 1. Отсюда получаем перестановку 6415273.

Задача 1.19. Перестановки n элементов упорядочены в лексикографическом порядке. Постройте алгоритмы:

- 1) нахождения номера перестановки,
- 2) нахождение перестановки по ее номеру.

Приведем решение для п. 2.

Кратко алгоритм генерации перестановок можно записать так.

Алгоритм 1.35 (генерация перестановок).

```
for  $i \leftarrow 1, n! - 1$  do
    строим факториальную запись числа  $i$  — procedure RECORD
    строим перестановку по факториальной записи — procedure PERMUTATION
end for
```

Используемые процедуры будут такими:

- a) построить факториальную запись числа $(i; n; a)$, где i — номер перестановки; n — число элементов перестановки; a — массив цифр факториальной записи с индексами от 1 до $n - 1$, начиная с младших разрядов.

```
procedure RECORD( $i, n, a$ )
     $t \leftarrow i$ 
```

```

for  $k \leftarrow 2, n$  do
     $a_{k-1} \leftarrow t \bmod k$ 
     $t \leftarrow t \div k$ 
end for
end procedure

```

б) построить перестановку по факториальной записи.

```

procedure PERMUTATION( $n, a$ )
    for  $k \leftarrow 1, n$  do
         $b_k \leftarrow k$ 
    end for
       ▷  $b_k$  — состав элементов перестановки; исходно в него входят все элементы в
        возрастающем порядке
    for  $k \leftarrow 1, n - 1$  do
       ▷ построение очередного элемента искомой перестановки  $c_k$ ; учитывается, что
        нумерация элементов факториальной записи велась с конца
         $c_k \leftarrow b_{a_{n-k}+1}$ 
    for  $j \leftarrow a_{n-k} + 1, n - k$  do
       ▷ удаление из состава элементов перестановки того, который был использован
        на текущем шаге — сдвиг элементов влево для заполнения освободившегося места
         $b_j \leftarrow b_{j+1}$ 
    end for
    end for
end procedure

```

Задача 1.20. Решите задачу 1.19 для обратного лексикографического порядка.

Для перечисления перестановок также можно поставить задачу на перечисление с ограничениями: две последующие перестановки должны получаться друг из друга перестановкой двух соседних элементов.

Рассмотрим идею такого алгоритма на примере перечисления перестановок четырех элементов: [1 2 3 4]:

- 1) возьмем последний элемент перестановки и будем попарными обменами «двигать» его влево: [1 2 4 3], [1 4 2 3], [4 1 2 3];
- 2) после того, как элемент 4 достиг левой границы набора, сдвинем элемент 3, стоящий на последнем месте, влево, обменяв с предыдущим: [4 1 3 2];
- 3) теперь снова начнем «движение» элемента 4, но в обратную сторону: [1 4 3 2], [1 3 4 2], [1 3 2 4];
- 4) теперь сдвинем элемент 3 еще на 1 влево и, получив перестановку [3 1 2 4], вновь повторим «движение» элемента 4: [3 1 4 2], [3 4 1 2], [4 3 1 2];
- 5) теперь (так как элемент 3 достиг своего левого положения) начнем «движение» элемента 2 влево, поменяв с предыдущим [4 3 2 1];

6) после чего опять начнем «движение» элемента 4: [3 4 2 1], [3 2 4 1], [3 2 1 4];

7) затем опять продвинем элемент 3 вправо, получив [2 3 1 4], а затем снова начнем «движение» элемента 4: [2 3 4 1], [2 4 3 1], [4 2 3 1];

8) затем опять продвинем элемент 3 вправо, получив [4 2 1 3], а затем снова начнем «движение» элемента 4: [2 4 1 3], [2 1 4 3], [2 1 3 4].

Получены все перестановки четырех элементов.

1.6.9. Бинарный код Грэя

Рассмотрим в общем виде идею кодирования, частный случай которой, был рассмотрен при перечислении перестановок в примере 1.91.

Определение 1.51. Кодом Грэя называется такая система нумерования неотрицательных чисел, когда коды двух соседних чисел отличаются ровно в одном бите. Код числа n обозначается $G(n)$. Этот код был изобретен Фрэнком Грэем (Frank Gray) в 1953 году.

Замечание 1.70. В современной технике, коды Грэя активно используются для минимизации ошибок при преобразовании аналоговых сигналов в цифровые (например, в датчиках). Отметим, что коды Грэя и были открыты в связи с этим применением.

Пример 1.100. Для чисел длины 3 бита имеем такую последовательность кодов Грэя:

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

Например, $G(4) = 6$.

Рассмотрим теперь обобщение алгоритма 1.29.

Нахождение кода Грэя числа N Рассмотрим биты числа N и биты числа $G(N)$. Заметим, что i -ый бит $G(N)$ равен единице только в том случае, когда i -ый бит N равен единице, а $i + 1$ -ый бит равен нулю, или наоборот (i -ый бит равен нулю, а $i + 1$ -ый равен единице). Таким образом, имеем:

$$G(N) = N \oplus \frac{N}{2}$$

Рассмотрим обратную задачу: по $G(N)$ восстановить число N .

Нахождение обратного кода Грэя Идем от старших битов к младшим (нумерация от 1 до k). Получаем такие соотношения между битами

N_i числа N и битами G_i числа $G(N)$:

$$\begin{cases} N_k &= G_k, \\ N_{k-1} &= G_{k-1} \oplus N_k = G_k \oplus G_{k-1}, \\ N_{k-2} &= G_{k-2} \oplus N_{k-1} = G_k \oplus G_{k-1} \oplus G_{k-2}, \\ N_{k-3} &= G_{k-3} \oplus N_{k-2} = G_k \oplus G_{k-1} \oplus G_{k-2} \oplus G_{k-3}, \\ \dots &= \dots \dots \dots \end{cases}$$

□ Код Грея и задача о Ханойских башнях

Коды Грея применяются в решении задачи о Ханойских башнях.

Постановка задачи. Даны три стержня, на один из которых нанизаны n колец¹, причем кольца отличаются размером и лежат меньшее на большем. Задача состоит в том, чтобы перенести пирамиду из n колец за наименьшее число ходов на другой стержень. За один раз разрешается переносить только одно кольцо, причём нельзя класть большее кольцо на меньшее.

Решение. Для представления перемещений n колец будем использовать n -разрядный код Грея.

Будем двигаться по кодам Грея по возрастанию:

$$G(0) = (\underbrace{0, \dots, 0}_n), \dots, G(2^n - 1) = (\underbrace{1, \dots, 1}_n).$$

Каждому биту i текущего кода Грея поставим в соответствие диск i (младшему биту соответствует наименьший по размеру диск, а старшему биту — наибольший).

Применим тот же подход, что и в задаче 1.91 о перечислении подмножеств шаров на столе. Так как на каждом шаге изменяется ровно один бит, будем считать, что изменение бита i соответствует перемещению диска i . Заметим, что для всех дисков, кроме наименьшего, на каждом шаге имеется ровно один вариант хода (за исключением стартовой и финальной позиций). Для наименьшего диска всегда имеется два варианта хода, однако имеется стратегия выбора хода, всегда приводящая к ответу.

Обозначим f — стартовый стержень, t — финальный стержень, r — оставшийся стержень. Тогда последовательность перемещений наименьшего диска имеет вид

$$\begin{cases} f \rightarrow t \rightarrow r \rightarrow f \rightarrow t \rightarrow r \rightarrow \dots & n \text{ — нечетно,} \\ f \rightarrow r \rightarrow t \rightarrow f \rightarrow r \rightarrow t \rightarrow \dots & n \text{ — четно.} \end{cases}$$

¹Первоначально, игра была предложена французским математиком Эдуардом Люка, в 1883 году для трех колец.

Очевидно, что для решения задачи потребуется $2^n - 1$ перемещений колец.

Историческая справка. Одно из любопытных применений кода Грэя связано с театром. Это код *Беккета-Грея*, он получил свое название в честь ирландского писателя Сэмюэла Беккета.

Одна из его пьес была написана для четырех акторов и состояла из 16 актов. В конце каждого акта один из четырех актеров выходил на сцену или же, уходил с нее. Пьеса начиналась на пустой сцене, и Беккет хотел, чтобы каждое подмножество актеров появлялось ровно один раз. Очевидно, что множество актеров на сцене может быть представлено в виде двоичного кода Грэя для $n = 4$.

Драматург, однако, добавил дополнительное условие в сценарий: чтобы со сцены уходил всегда тот из актеров, кто находился на ней дольше остальных. Беккет не смог найти код Беккета-Грея для своей пьесы. Позднее было доказано, что такие коды существуют для $n = 2, 5, 6, 7, 8$ и не существуют для $n = 3, 4$.

1.6.10. Числа Стирлинга первого и второго родов

Определение 1.52. Числом *Стирлинга¹ второго рода* (обозначают $S_{n,k}$) называют число разбиений n -элементного множества на k непересекающихся непустых подмножеств.

Очевидно, что $S_{n,k} = 0$ для $k > n$. Положим $S_{0,0} = 1$.

Лемма 1.15. Для $S_{n,k}$ справедливы следующие свойства:

- 1) $S_{n,0} = 0$, $n > 0$;
- 2) $S_{n,n} = 1$, $n \geq 0$;
- 3) $S_{n,k} = S_{n-1,k-1} + kS_{n-1,k}$, $0 < k < n$.

Доказательство. Свойства 1 и 2 очевидны. Установим свойство 3. Рассмотрим множество всех разбиений n -элементного множества $\{1, 2, \dots, n\}$ на k непересекающихся подмножеств. Эти разбиения распадаются на две группы:

- разбиения, которые содержат $\{n\}$ как одноэлементное множество;
- разбиения, где n содержится в составе по крайней мере двухэлементных множеств.

Для первой группы существует $S_{n-1,k-1}$ вариантов разбиений, т. е. разбиений множества $\{1, \dots, n-1\}$ на $k-1$ непересекающихся подмножеств. Для второй группы каждому разбиению множества $\{1, \dots, n-1\}$ на k непересекающихся подмножеств ($S_{n-1,k}$ вариантов) соответствует k разбиений, получающихся поочередным добавлением элемента n к каждому подмножеству. Следовательно, в этой группе $kS_{n-1,k}$ вариантов разбиений. ■

¹Джеймс Стирлинг (1692–1770), — шотландский математик, член Лондонского королевского общества.

Эти свойства позволяют легко вычислять числа Стирлинга для произвольных n и k . В табл. 1.42 даны числа Стирлинга второго рода.

Таблица 1.42

n	k					
	1	2	3	4	5	6
1	1	0	0	0	0	0
2	1	1	0	0	0	0
3	1	3	1	0	0	0
4	1	7	6	1	0	0
5	1	15	25	10	1	0
6	1	31	90	65	15	1

Существует другая рекуррентная формула для вычисления чисел Стирлинга второго рода.

Теорема 1.73. Для $k \geq 2$ справедливо соотношение:

$$S_{n,k} = \sum_{i=k-1}^{n-1} C_{n-1}^i S_{i,k-1}. \quad (1.95)$$

Доказательство. Рассмотрим n -элементное множество $\{1, \dots, n\}$ и зафиксируем в нем последний элемент n . Из оставшихся элементов для каждого $j \in 0 : n - k$ выделим j (это можно сделать C_{n-1}^j способами) и добавим к ним зафиксированный последний элемент n . Рассматриваем выбранные $j + 1$ элементы, как одно подмножество. Для оставшегося $(n - j - 1)$ -элементного множества существует $S_{n-j-1, k-1}$ разбиений. Таким образом,

$$S_{n,k} = \sum_{j=0}^{n-1} C_{n-1}^j S_{n-j-1, k-1}. \quad (1.96)$$

Сделав в (1.96) замену переменной $i = n - j - 1$ и учитывая свойство биномиальных коэффициентов ($C_{n-1}^j = C_{n-1}^{n-1-j}$), установленное в теореме 1.63, получаем требуемое. ■

Определение 1.53. Числом *Белла* (обозначают B_n) называют число всех разбиений n -элементного множества на $1, \dots, n$ непересекающихся подмножеств, т. е. $B_n = \sum_{k=0}^n S_{n,k}$.

Положим $B_0 = 1$. Докажем следующее рекуррентное соотношение.

Теорема 1.74. Справедливо равенство

$$B_{n+1} = \sum_{i=0}^n C_n^i B_i.$$

Доказательство. Для доказательства немного изменим рассуждения, используемые при доказательстве теоремы 1.73.

Рассмотрим $(n + 1)$ -элементное множество и зафиксируем в нем элемент $n + 1$. Будем выбирать произвольные i элементов из множества $\{1, \dots, n\}$. Оставшиеся $n + 1 - i$ элементов рассматриваем как *одно* подмножество. Так как $0 \leq i \leq n$, то в последнем подмножестве может быть от 1 до $n + 1$ элементов. Для каждой выборки из i элементов существует B_i разбиений, а сами элементы можно выбрать C_n^i способами. ■

Введем в рассмотрение понятие *факториальных многочленов*.

Определение 1.54. Факториальный многочлен степени k

$$[x]_k = x(x - 1) \cdots (x - k + 1),$$

т. е. $[x]_1 = x$, $[x]_2 = x(x - 1)$, $[x]_3 = x(x - 1)(x - 2)$ и т. д.

Очевидно, что $\{1, [x]_1, [x]_2, \dots, [x]_n, \dots\}$ — базис в пространстве многочленов.

Теорема 1.75. Справедливо равенство

$$x^n = \sum_{k=0}^n S_{n,k} [x]_k,$$

т. е. числа Стирлинга второго рода — это матрица перехода от канонического базиса $\{1, x, x^2, \dots, x^n, \dots\}$ к базису факториальных многочленов $\{1, [x]_1, [x]_2, \dots, [x]_n, \dots\}$.

Упражнение 1.12. Докажите теорему 1.75.

Определим обратную связь этих базисов.

Определение 1.55. Числа Стирлинга первого рода — это матрица перехода от базиса факториальных многочленов $\{1, [x]_1, [x]_2, \dots, [x]_n, \dots\}$ к базису $\{1, x, x^2, \dots, x^n, \dots\}$, т. е. $[x]_n = \sum_{k=0}^n s_{n,k} x^k$.

Лемма 1.16. Для введенных чисел Стирлинга первого рода $s(n, k)$ справедливы следующие свойства:

- 1) $s_{n,0} = 0$, $n > 0$;
- 2) $s_{n,n} = 1$, $n \geq 0$;
- 3) $s_{n,k} = s_{n-1,k-1} - (n-1)s_{n-1,k}$, $0 < k < n$.

Доказательство. Свойства 1 и 2 очевидны. Докажем свойство 3. Имеем $[x]_n = [x]_{n-1}(x - n + 1)$, тогда, согласно определению 1.55,

$$\begin{aligned}
\sum_{k=0}^n s_{n,k}x^k &= (x - n + 1) \sum_{k=0}^{n-1} s_{n-1,k}x^k = \\
&= \sum_{k=0}^{n-1} s_{n-1,k}x^{k+1} - (n - 1) \sum_{k=0}^{n-1} s_{n-1,k}x^k = \\
&= \sum_{k=1}^{n-1} s_{n-1,k-1}x^k + s_{n-1,k-1}x^n - (n - 1) \sum_{k=1}^{n-1} s_{n-1,k}x^k - s_{n-1,0}(n - 1) = \\
&= \sum_{k=1}^{n-1} (s_{n-1,k-1} - (n - 1)s_{n-1,k})x^k + s_{n,n}x^n - (n - 1)s_{n-1,0}.
\end{aligned}$$

Заметим, что $s_{n,n} = 1$, $s_{n-1,0} = 0$. Приравняв коэффициенты при одинаковых степенях x , получаем требуемое. ■

Для введенных факториальных многочленов имеет место теорема, аналогичная теореме 1.62, для обычных многочленов.

Теорема 1.76 (факториальная теорема Вандермонда). Справедливо равенство

$$[x + y]_n = \sum_{k=0}^n C_n^k [x]_k [y]_{n-k}.$$

Упражнение 1.13. Докажите теорему 1.76.

1.6.11. Числа Каталана

Рассмотрим несколько задач, связанных с известной комбинаторной структурой — числами Каталана¹, и на их примере покажем приемы, позволяющие одни комбинаторные задачи сводить к другим установлением подходящего взаимно-однозначного соответствия.

Задача 1.21. Рассмотрим правильные скобочные структуры, состоящие из n открывающих и n закрывающих скобок. Термин «правильные» означает, что для каждой открывающей скобки найдется единственная соответствующая ей закрывающая, стоящая правее ее.

Например, $(())()$ и $()(())$ — правильные скобочные структуры для $n = 3$, а $()(())$ — неправильная структура для того же значения n .

Сколько существует правильных $2n$ -скобочных структур?

¹Эжен Шарль Каталан (1814 – 1894) — бельгийский математик, иностранный член-корреспондент Российской академии наук. Эти числа были известны еще Эйлеру за 100 лет до Каталана.

Задача 1.22. Рассмотрим произведение $n + 1$ матриц общего вида: $A_0 A_1 \cdots A_n$, в котором не расставлены скобки, так как произведение матриц ассоциативно.

Сколькими способами можно вычислить это произведение, сводя его к попарному умножению матриц. Или, что то же самое, сколькими способами можно расставить $2n$ скобок в указанном произведении.

На первый взгляд кажется, что задача 1.21 и задача 1.22 одинаковы, но если сравнить два разных выражения $((AB)C)$ и $(A(BC))$, просто опустив буквы, то получатся одинаковые скобочные структуры $(())$. Обратно, для правильной скобочной структуры $()()$ нет произведения трех матриц с такой расстановкой скобок.

Задача 1.23. Сколькими способами можно разрезать неправильный выпуклый $n + 2$ угольник на n треугольников отрезками, соединяющими различные вершины (рис. 1.11)?

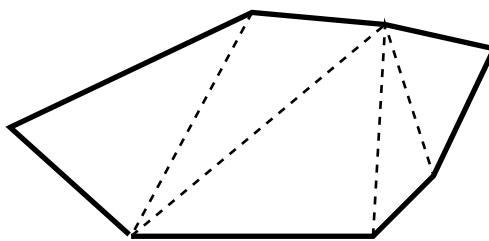


Рис. 1.11

Задача 1.24. Сколько существует корневых бинарных деревьев с $n + 1$ листьями, у которых каждая вершина, не являющаяся листом, имеет ровно двух потомков (степень каждой вершины либо 1, либо 3)?

Задача 1.25. Человек с завязанными глазами стоит за один шаг до края обрыва. Он может делать либо шаг по направлению к обрыву (не зная этого), либо шаг в обратную сторону. Каково число вариантов, сделав $2(n - 1)$ шагов, не упасть с обрыва?

Задача 1.26. Сколько существует различных корневых деревьев на плоскости, имеющих n вершин?

Задача 1.27. Сколько существует последовательностей из $2n - 1$ неотрицательных целых чисел, начинающихся с 1 и заканчивающихся на 1, у которых соседние числа отличаются на 1?

Например, для $n = 3$ к таким относятся последовательности 1 2 1 2 3 2 1 и 1 2 3 2 1 2 1.

Покажем, что все эти задачи имеют одинаковую комбинаторную структуру.

Сведение задачи 1.22 к задаче 1.21

Оставим открытые скобки на месте, буквы удалим, а каждый знак операции умножения заменим закрывающей скобкой. Получим взаимно-однозначное соответствие. Например, выражениям $((AB)C)$ и $(A(BC))$ будут соответствовать скобочные структуры $(())$ и $()()$. То, что это соответствие взаимно-однозначно, будет следовать из теоремы о префиксной форме арифметического выражения (см. гл. 3). Сейчас заметим только, что префиксные записи, однозначно определяющие арифметические выражения, для наших примеров выглядят так: $\times \times ABC$ и $\times A \times BC$. Если отбросить последнюю букву, заменить знаки умножения на открывающие скобки, а буквы — на закрывающие, то получатся упомянутые скобочные структуры.

Упражнение 1.14.

1. Придумайте другое взаимно-однозначное соответствие, сводящие задачи 1.21 и 1.22 друг к другу.
2. Как по скобочной структуре восстановить произведение матриц?

Сведение задачи 1.23 к задаче 1.22.

Выделим одну из сторон многоугольника, а по периметру, начиная с соседней стороны, сопоставим имена переменных (матриц) остальным сторонам многоугольника.

Будем рассматривать различные разбиения многоугольника на треугольники, сопоставляя каждому вновь проведенному отрезку произведение выражений, сопоставленных ранее другим его сторонам, начиная со сторон многоугольника и учитывая естественный порядок множителей (рис. 1.12,*a*).

Полученное сопоставление является искомым взаимно-однозначным соответствием.

Упражнение 1.15. По заданному порядку выполнения в умножении матриц постройте разбиение многоугольника на треугольники. Например, для многоугольника на рис. 1.12,*a* постройте разбиение по выражению $((AB)(CD))$.

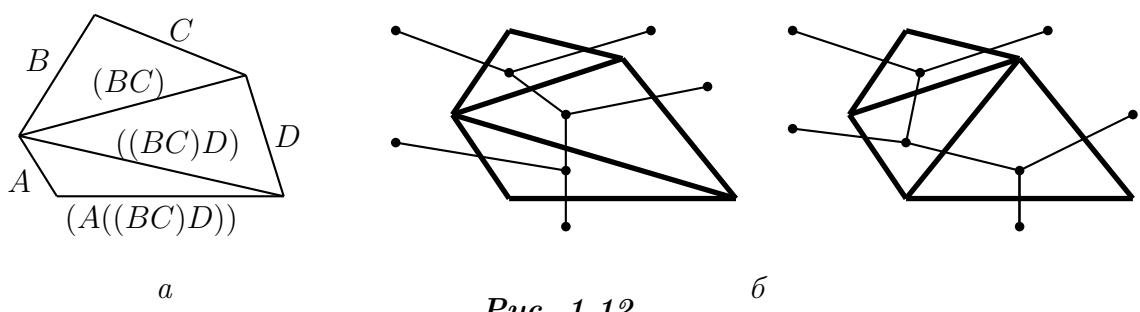
Связь задачи 1.24 и задачи 1.23.

На рис. 1.12,*b* показана связь между разбиением многоугольника на треугольники и корневыми бинарными деревьями.

Связь задачи 1.24 и задачи 1.22.

Обратим внимание, что для каждого арифметического выражения можно построить соответствующее синтаксическое дерево, которое и будет требуемым бинарным деревом (рис. 1.13).

Сведение задачи 1.25 к задаче 1.21.



Puc. 1.12

Закодируем каждый шаг человека от обрыва открывающей скобкой, к обрыву — закрывающей. Тогда, для того, чтобы человек не дешел до конца обрыва, любой начальный отрезок скобочной структуры должен содержать больше открывающих скобок, чем закрывающих. Возвращение человека на исходную позицию означает совпадение числа открывающих и числа закрывающих скобок. Эти два условия являются необходимыми и достаточными для правильности скобочной структуры. Взаимно-однозначное соответствие доказано.

Сведение задачи 1.26 к задаче 1.21.

На рис. 1.14 показано искомое взаимно-однозначное соответствие: сделяем обход дерева, помечая движение вверх открывающей скобкой, вниз — закрывающей.

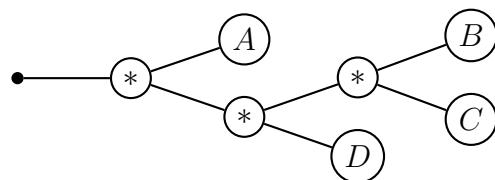
Например, указанному на рис. 1.14 дереву соответствует правильная скобочная структура (((()(())())()).

Упражнение 1.16. Сведите задачу 1.27 к одной из предыдущих задач.

В заключение решим одну из задач, что автоматически даст решение для остальных. Наибольший интерес представляет формулировка задачи 1.25, так как в ее решении идея взаимно-однозначного соответствия используется еще раз.

Нарисуем график положений человека в дискретные моменты времени, соответствующие состояниям после очередного шага. Для наглядности точки соединим отрезками. Например, на рис. 1.15 изображен путь, который может быть также описан скобочной структурой, соответствующей рис. 1.14.

Любой путь из $2(n - 1)$ шагов, не приводящий к падению человека с обрыва изображается графиком, соединяющим точки $(0; 1)$ и $(2(n - 1); 1)$ и не пересекающим ось абсцисс.



Pyc. 1.13

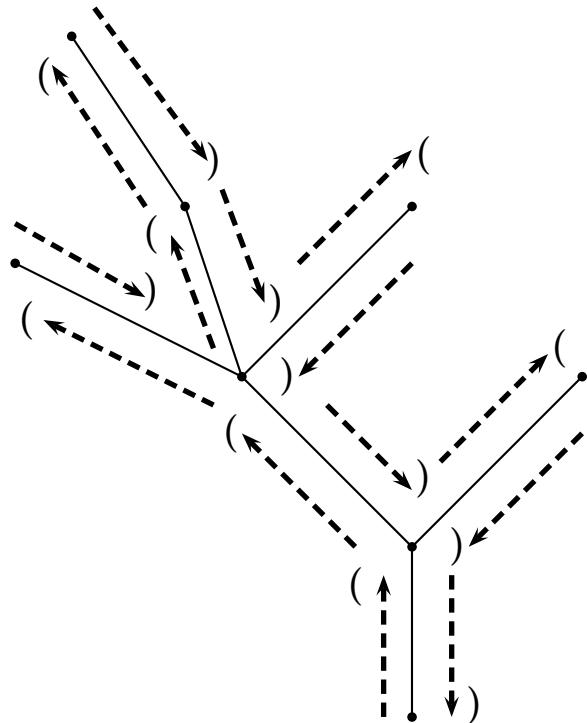


Рис. 1.14

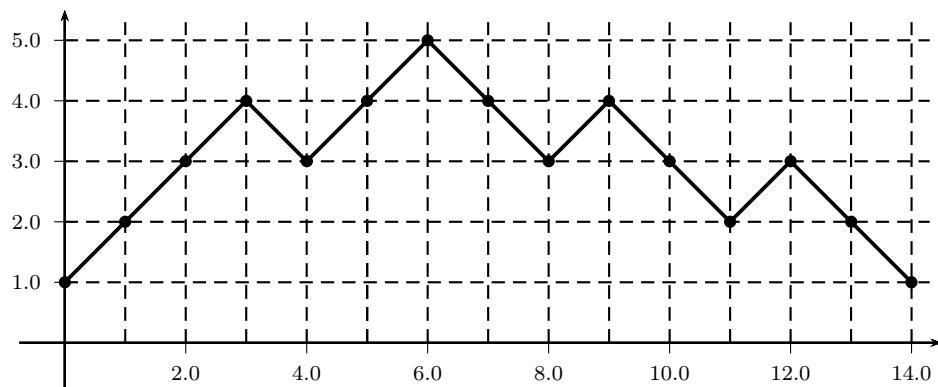


Рис. 1.15

Чтобы найти число таких графиков, найдем число графиков, соединяющих эти точки, но пересекающих ось абсцисс.

Каждому графику, соединяющему точки $(0; 1)$ и $(2(n - 1); 1)$ и пересекающему ось абсцисс можно взаимно-однозначно сопоставить график, соединяющий точки $(0; -1)$ и $(2(n - 1); 1)$. Для этого надо часть графика до первого пересечения отразить относительно оси абсцисс (рис. 1.16).

График состоит из $2(n - 1)$ звеньев, из которых возрастающих звеньев на два больше, чем убывающих (так как разница ординат начала и конца графика равна 2): n — возрастающих и $n - 2$ убывающих. Значит, общее число графиков будет равно числу способов выбрать из $2(n - 1)$ звеньев n возрастающих: C_{2n-2}^n .

Аналогично, общее число графиков, соединяющих две точки $(0; 1)$ и

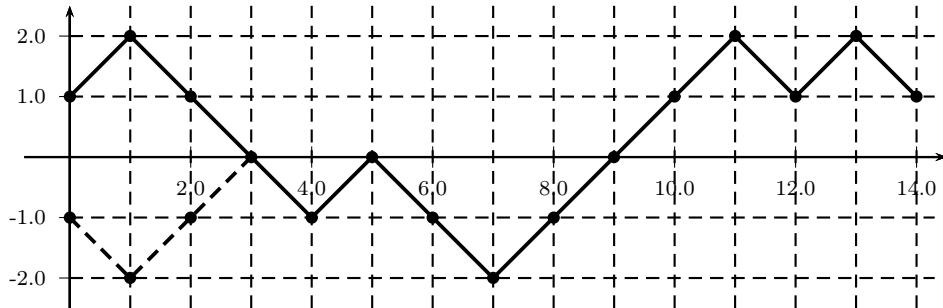


Рис. 1.16

$(2(n-1); 1)$ будет равно: C_{2n-2}^{n-1} . Таким образом, искомое число путей равно

$$\begin{aligned} C_{2n-2}^{n-1} - C_{2n-2}^n &= \frac{(2n-2)!}{(n-1)!(n-1)!} - \frac{(2n-2)!}{n!(n-2)!} = \\ &= \frac{(2n-2)!}{(n-1)!(n-2)!} \left(\frac{1}{n-1} - \frac{1}{n} \right) = \\ &= \frac{(2n-2)!}{(n-1)!(n-2)!} \frac{1}{n(n-1)} = \frac{1}{n} \frac{(2n-2)!}{(n-1)!(n-1)!} = \frac{1}{n} C_{2n-2}^{n-1}. \end{aligned}$$

Определение 1.56. Числа $\frac{1}{n} C_{2n-2}^{n-1}$ называют *числами Каталана*.

1.6.12. Разбиения чисел

Рассмотрим обобщение разачи 1.7. Пусть n — произвольное натуральное число.

Определение 1.57. *Разбиением* числа n называют набор чисел

$$\{a_1, \dots, a_k\}, \quad a_i \in \mathbb{Z}, \quad a_1 \geq \dots \geq a_k > 0 \text{ и } n = a_1 + a_2 + \dots + a_k.$$

Будем обозначать разбиение числа n , как $n = (a_1, \dots, a_k)$.

Пример 1.101. Пусть $n = 7$. Запишем все разбиения в обратном лексикографическом порядке (см. определение 1.50, стр. 145):

$(7), (6, 1), (5, 2), (5, 1, 1), (4, 3), (4, 2, 1), (4, 1, 1, 1), (3, 3, 1), (3, 2, 1, 1), (3, 1, 1, 1, 1), (2, 2, 2, 1), (2, 2, 1, 1, 1), (2, 1, 1, 1, 1, 1), (1, 1, 1, 1, 1, 1, 1).$

Определение 1.58. Для каждого разбиения (a_1, \dots, a_k) можно построить *диаграмму Ферре* (*граф Ферре*). Она представляет собой таблицу из k строк, i -я строка которой содержит последовательность из a_i точек. Если заменить точки квадратами, то получим *диаграмму Юнга*.

Любому разбиению соответствует *сопряженное разбиение*, получаемое транспозицией диаграммы Ферре.

Наибольшую квадратную поддиаграмму диаграммы Ферре, называют *квадратом Дюрфи* соответствующего разбиения. Если квадрат Дюрфи совпадает со всей диаграммой, такое разбиение называют *самосопряженным*.

Очевидно, что транспозиция диаграммы Ферре определяет взаимно-однозначное соответствие между разбиением числа n на k слагаемых и его же разбиением с наибольшим слагаемым, равным k . Справедливо утверждение.

Лемма 1.17. Пусть $n = (a_1, \dots, a_k)$. Тогда число элементов, равных m , в сопряженном разбиении равно $a_m - a_{m+1}$ (считаем, что $a_{k+1} = 0$).

Лемма 1.17 дает удобный способ вычисления сопряженного разбиения, не рисуя диаграмму Ферре.

Пример 1.102. Диаграмма Ферре и диаграмма Юнга для разбиения $12 = (4, 3, 3, 2)$ приведены на рис. 1.17 *а* и *б*. Диаграмма Ферре для соответствующего сопряженного разбиения $12 = (4, 4, 3, 1)$ показана на рис. 1.17, *в*.

Квадраты Дюрфи выделены на рис. 1.17 *а*, *в* утолщенными штриховыми линиями, на диаграмме Юнга квадрат Дюрфи заштрихован.

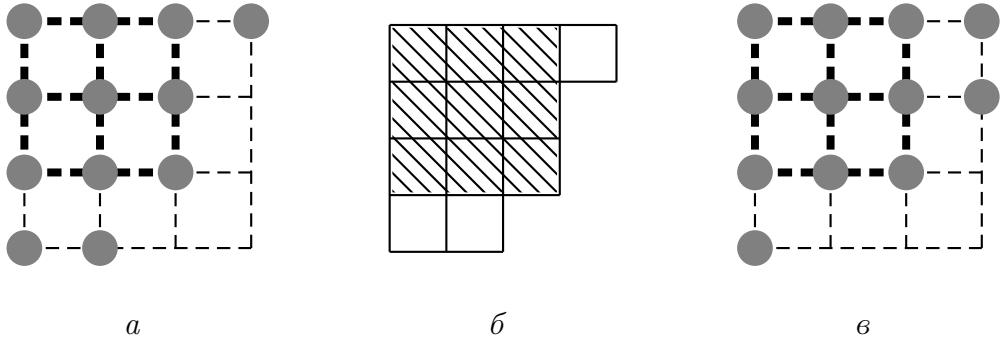


Рис. 1.17

Теорема 1.77. Число разбиений n на попарно различные слагаемые равно числу разбиений n на нечетные слагаемые.

Доказательство. Рассмотрим разбиение числа n на нечетные слагаемые b_1, \dots, b_p , где слагаемое b_i появляется в разбиении r_i раз:

$$n = \underbrace{b_1 + b_1 + \cdots + b_1}_{r_1} + \cdots + \underbrace{b_p + b_p + \cdots + b_p}_{r_p} \quad (1.97)$$

Запишем двоичное представление r_i

$$r_i = 2^{q_1} + \cdots + 2^{q_s}, \quad q_1 > \cdots > q_s \geq 0.$$

Заменим r_i слагаемых b_i в (1.97) на попарно различные слагаемые

$$b_i 2^{q_1}, \dots, b_i 2^{q_s}.$$

Очевидно, что при этом сохраняется сумма разбиения. Повторив эту операцию для всех $i = 1, \dots, p$ и упорядочив слагаемые, получим требуемое представление, так как $b_i 2^{q_i} \neq b_j 2^{q_m}$ в силу нечетности b_i, b_j . ■

Пример 1.103.

$$26 = 7 + 5 + 5 + 3 + 3 + 1 + 1 + 1 = 7 \cdot 2^0 + 5 \cdot 2^1 + 3 \cdot 2^1 + 1 \cdot (2^1 + 2^0) = 10 + 7 + 6 + 2 + 1.$$

Упражнение 1.17. Пусть $E(n)$ — число разбиений числа n с четным числом четных слагаемых, $O(n)$ — соответственно число разбиений с нечетным числом четных слагаемых и $S(n)$ — число самосопряженных разбиений. Покажите, что

$$E(n) - O(n) = S(n).$$

Построим алгоритм для генерации разбиений числа n . Генерацию будем производить в обратном лексикографическом порядке, т. е. разбиение $n = c_1 + c_2 + \dots + c_l$ идет за разбиением $n = a_1 + a_2 + \dots + a_k$ тогда и только тогда, когда существует $p \leq \min\{l, k\}$: $c_i = a_i$, для $i < p$; $c_p < a_p$.

Очевидно, что первое разбиение $n = n$, последнее $n = \underbrace{1 + 1 + \dots + 1}_n$.

Пусть $n = a_1 + a_2 + \dots + a_k$ — некое промежуточное разбиение. Какое разбиение будет следующим? Будет искать разбиение, имеющее самое большое число начальных слагаемых, равных слагаемым предыдущего разбиения — обозначим их a_1, a_2, \dots, a_{t-1} . Очевидно, что $t = \max\{i | a_i > 1\}$, т. е. предыдущее разбиение имеет вид:

$$n = a_1 + a_2 + \dots + a_{t-1} + a_t + \underbrace{1 + 1 + \dots + 1}_{k-t}.$$

Тогда оставшиеся слагаемые текущего легко определяются. Введем обозначение $s = a_t + \underbrace{1 + 1 + \dots + 1}_{k-t}$, т. е. s — сумма «остатка» предыдущего разбиения. Представим эту сумму s в виде: $s = \underbrace{l + l + \dots + l}_{[s/l]} + < s >_l$,

где $l = a_t - 1$, т. е. набираем s максимально возможными элементами, меньшими a_t . Тогда текущее разбиение имеет вид:

$$n = a_1 + a_2 + \dots + a_{t-1} + \underbrace{l + l + \dots + l}_{[s/l]} + < s >_l.$$

Запишем этот процесс в виде алгоритма. Введем следующие обозначения:

a_i — элементы текущего разбиения,

r_i — кратность i -го элемента в разбиении,

j — число элементов в текущем разбиении без учета кратности, значение переменной s описано ранее.

Алгоритм 1.36 (генерация разбиений).

```

 $a_1 \leftarrow n$ 
 $r_1 \leftarrow 1$ 
 $j \leftarrow 1$                                  $\triangleright$  первое разбиение
while  $a_1 > 1$  do
     $s \leftarrow 0$ 
    if  $a_j = 1$  then                   $\triangleright$  учет единичных слагаемых предыдущего разбиения
         $s \leftarrow s + r_j$ 
         $j \leftarrow j - 1$ 
    end if
     $s \leftarrow s + a_j$   $\triangleright$  добавляем в  $s$  последний неединичный элемент предыдущего разложения
     $r_j \leftarrow r_j - 1$                        $\triangleright$  уменьшаем его кратность в разложении
     $l \leftarrow a_j - 1$                          $\triangleright$  запоминаем новый элемент
    if  $r_j > 0$  then
         $j \leftarrow j + 1$   $\triangleright$  в текущем разложении есть еще элементы  $a_j$ , увеличиваем индекс для
        нового элемента
    end if
     $a_j \leftarrow l$                              $\triangleright$  строим «остаток» нового разбиения
     $r_j \leftarrow s \div l$ 
     $d \leftarrow \langle s \rangle_l$ 
    if  $d > 0$  then                   $\triangleright$  добавляем остаток  $d$  последним элементом в разбиение
         $j \leftarrow j + 1$ 
         $a_j \leftarrow d$ 
         $r_j \leftarrow 1$ 
    end if
end while

```

Пример 1.104. Применим алгоритм 1.36 для $n = 9$. Протокол работы алгоритма дан в табл. 1.43.

Таблица 1.43

Шаг	Текущее разбиение	Шаг	Текущее разбиение
1	$9 = 9$	8	$9 = 5 + 4$
2	$9 = 8 + 1$	9	$9 = 5 + 3 + 1$
3	$9 = 7 + 2$	10	$9 = 5 + 2 + 2$
4	$9 = 7 + 1 + 1$	11	$5 + 2 + 1 + 1$
5	$9 = 6 + 3$	12	$9 = 5 + 1 + 1 + 1 + 1$
6	$9 = 6 + 2 + 1$	13	$9 = 4 + 4 + 2$
7	$9 = 6 + 1 + 1 + 1$

Последнее разбиение $9 = 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1$.

1.6.13. Введение в динамическое программирование. Задача о рюкзаке

Идея динамического программирования заключается в следующем: исходная задача представляется в виде набора «подзадач» разного размера, решаемых одна за другой в *правильном* порядке. Правильный порядок означает, что когда нужно решать некоторую подзадачу набора, для этого уже все готово, то есть те задачи, от которых зависит ее решение, уже решены.

Применим динамическое программирование для решения *задачи о рюкзаке*. Рассмотрим классическую постановку задачи.

Имеется рюкзак вместимости W и набор из n предметов, каждый из которых имеет два положительных параметра: вес и стоимость. Задача заключается в том, чтобы собрать рюкзак с максимальной стоимостью предметов внутри, соблюдая при этом ограничение рюкзака на суммарный вес. Введем обозначения: $\{w_1, \dots, w_n\}$ — веса и $\{v_1, \dots, v_n\}$ — стоимости предметов.

Существует большое количество постановок задачи, в зависимости от условий, наложенных на рюкзак, предметы или их выбор. Подробное описание и разбор можно найти в [43].

Рассмотрим сначала одну из наиболее популярных постановок: каждый предмет имеется в неограниченном количестве¹. При использовании метода динамического программирования очень важно правильно выбрать подзадачи. В этом качестве будем рассматривать рюкзаки меньшей вместимости. Обозначим $C(w)$ — максимально возможная стоимость для рюкзака вместимости w .

Попробуем выразить $C(w)$ через ответы для меньших подзадач. Пусть в оптимальное заполнение рюкзака емкости w входит предмет i , тогда если его убрать, мы получим оптимальное заполнение рюкзака вместимости $w - w_i$. Тогда $C(w)$ — это $C(w - w_i) + v_i$ для некоторого i . Поскольку индекс предмета i неизвестен, нужно перебрать все возможные варианты. Имеем:

$$C(w) = \max_i \{C(w - w_i) + v_i \mid w_i \leq w\}.$$

Запишем алгоритм решения.

Алгоритм 1.37 (Задача о неограниченном рюкзаке).

```

 $C(0) \leftarrow 0$ 
for  $w \leftarrow 1, W$  do
     $C(w) = \max \{C(w - w_i) + v_i \mid w_i \leq w\}$ 
end for
return  $C(W)$ 

```

¹Неограниченный рюкзак (англ. Unbounded Knapsack Problem)

Пример 1.105. Пусть имеется 4 предмета. Данные об их весе и стоимости даны в таблице 1.44. Решим задачу алгоритмом 1.37 для $W = 9$.

Таблица 1.44

i	1	2	3	4
v_i	5	7	6	3
w_i	2	3	4	2

Протокол работы алгоритма запишем в таблице 1.45.

Таблица 1.45

i	$C(i)$	Решение	i	$C(i)$	Решение
0	0		5	12	$C(5) = C(5 - w_2) + v_2$
1	0		6	15	$C(6) = C(6 - w_1) + v_1$
2	5	$C(2) = C(2 - w_1) + v_1$	7	17	$C(7) = C(7 - w_2) + v_2$
3	7	$C(3) = C(3 - w_2) + v_2$	8	20	$C(8) = C(8 - w_1) + v_1$
4	10	$C(4) = C(4 - w_1) + v_1$	9	22	$C(9) = C(9 - w_1) + v_1$

Таким образом, получаем оптимальное решение:

$$C(9) = v_1 + v_1 + v_1 + v_2 = 22.$$

Рассмотрим другую постановку задачи: каждый предмет имеется в одном экземпляре². Предыдущий подход использовать не удастся, так как на каждом шаге нужно учитывать предметы, которые уже взяты ранее.

Обозначим $C(w, i)$ — максимально возможная стоимость для рюкзака вместимости w и при этом разрешено брать предметы $1, \dots, i$. Как и в предыдущем случае нужно научиться выражать $C(w, i)$ через результаты для предыдущих подзадач.

Возможны два случая: в оптимальном заполнении предмет i либо участвует, либо нет. Тогда

$$C(w, i) = \max \{C(w, i - 1), C(w - w_i, i - 1) + v_i \mid w_i \leq w\},$$

где второй элемент учитывается только при $w_i \leq w$.

Запишем алгоритм решения.

Алгоритм 1.38 (Задача о рюкзаке 0 – 1).

for $i \leftarrow 0, n$ do

²Рюкзак 0 – 1 (англ. 0 – 1 Knapsack Problem)

```

 $C(0, i) \leftarrow 0$ 
end for
for  $w \leftarrow 0, W$  do
     $C(w, 0) \leftarrow 0$ 
end for
for  $i \leftarrow 1, n$  do
    for  $w \leftarrow 1, W$  do
        if  $w_i > w$  then
             $C(w, i) \leftarrow C(w, i - 1)$ 
        else
             $C(w, i) \leftarrow \max \{C(w, i - 1), C(w - w_i, i - 1) + v_i\}$ 
        end if
    end for
end for
return  $C(W, n)$ 

```

Пример 1.106. Пусть имеется 5 предметов. Данные об их весе и стоимости даны в таблице 1.46. Решим задачу алгоритмом 1.38 для $W = 6$.

Таблица 1.46

i	1	2	3	4	5
v_i	1	2	3	1	1
w_i	2	3	2	4	1

Протокол работы алгоритма запишем в таблице 1.47.

Таблица 1.47

i	$C(1, i)$	$C(2, i)$	$C(3, i)$	$C(4, i)$	$C(5, i)$	$C(6, i)$
1	0	$v_1 = 1$	1	1	1	1
2	0	$v_1 = 1$	$v_2 = 2$	2	$v_1 + v_2 = 3$	3
3	0	$v_3 = 3$	3	$v_1 = v_3 = 4$	$v_2 + v_3 = 5$	5
4	0	$v_3 = 3$	3	$v_1 = v_3 = 4$	$v_2 + v_3 = 5$	5
5	0	$v_3 = 3$	$v_3 + v_5 = 4$	4	$v_2 + v_3 = 5$	$v_2 + v_3 + v_5 = 6$

Таким образом, получаем оптимальное решение:

$$C(6, 5) = v_2 + v_3 + v_5 = 6.$$

1.6.14. Принцип включения–исключения

Рассмотрим свойства $p(1), \dots, p(n)$, которыми обладают элементы множества S , состоящего из N элементов. Предположим, что для каждого

элемента можно однозначно определить, обладает он свойством $p(i)$ или нет. Число элементов, обладающих свойствами $p(i_1), \dots, p(i_r)$, обозначим через $N_{i_1 \dots i_r}$.

Замечание 1.71. Здесь не рассматриваются другие свойства, т. е. допускается что некоторые из элементов, быть может, обладают и другими свойствами.

Число элементов, которые не обладают ни одним из рассматриваемых свойств, обозначим через $N(0)$. Число элементов, которые обладают *ровно* r ($1 \leq r \leq n$) свойствами из n возможных, будем обозначать $N(r)$. Через $N[r]$ обозначим число элементов, обладающих *не менее* r свойствами. Тогда

$$N[r] = \sum_{i_1 < \dots < i_r} N_{i_1 \dots i_r} = \sum_{i=r}^n N(i). \quad (1.98)$$

В самом простом случае ($n = 2$) очевидно, что

$$N(0) = N - (N_1 + N_2) + N_{12} = N[0] - N[1] + N[2].$$

Действительно, каждый элемент, обладающий свойствами 1 и 2 одновременно учитывается и в N_1 и в N_2 , а поэтому при вычислении разности $N - (N_1 + N_2)$ он вычитается из N дважды. Следовательно, если к этой разности добавить N_{12} , то получится $N(0)$ — число элементов, не обладающих ни одним из двух рассматриваемых свойств. Обобщением этого правила является следующий принцип включения–исключения.

Теорема 1.78 (принцип включения–исключения¹).

$$N(0) = N[0] - N[1] + N[2] - \dots + (-1)^s N[s] + \dots + (-1)^n N[n]. \quad (1.99)$$

Доказательство. Элементы, не обладающие ни одним из свойств, входят в правую часть равенства (1.99) ровно по одному разу, а именно в первое слагаемое — N . Рассмотрим элементы, которые обладают ровно r свойствами $p(j_1), \dots, p(j_r)$ и слагаемые из (1.99) вида

$$(-1)^s N[s], s \leq r.$$

Так как совокупность индексов i_1, \dots, i_s из множества j_1, \dots, j_r можно выбрать C_r^s способами, вклад каждого из рассматриваемых элементов в правую часть (1.99) будет

$$1 - C_r^1 + C_r^2 + \dots + (-1)^s C_r^s + \dots + (-1)^r C_r^r = 0.$$

Воспользовались свойством биномиальных коэффициентов (следствие 1.22 к теореме 1.62). ■

¹ Впервые формулу (1.99) опубликовал португальский математик Даниэль да Сильва в 1854 году. Но еще в 1713 году Николай Бернулли использовал этот метод для решения задачи Монмора (пример 1.108). Другие названия: символический метод, принцип перекрестной классификации, метод решетки.

В качестве применения принципа включения–исключения докажем теорему Лежандра.

Теорема 1.79 (Лежандр). Пусть a_1, \dots, a_k набор попарно взаимно простых натуральных чисел. Тогда количество натуральных чисел, не превышающих некоторого n и взаимно простых с любым из a_i равно

$$n - \sum_{1 \leq i \leq k} \left\lfloor \frac{n}{a_i} \right\rfloor + \sum_{1 \leq i < j \leq k} \left\lfloor \frac{n}{a_i a_j} \right\rfloor - \sum_{1 \leq i < j < s \leq k} \left\lfloor \frac{n}{a_i a_j a_s} \right\rfloor + \dots + (-1)^k \left\lfloor \frac{n}{a_1 \cdots a_k} \right\rfloor \quad (1.100)$$

Доказательство. Пусть S — множество натуральных чисел $\{1, \dots, n\}$. Обозначим через $p(i)$ свойство элемента множества S делиться на a_i . Тогда $N_{i_1 \dots i_r}$ — это количество чисел, не превышающих n и делящихся нацело на каждое из чисел a_{i_1}, \dots, a_{i_r} . Так как все a_i попарно взаимно просты, то

$$N_{i_1 \dots i_r} = \left\lfloor \frac{n}{a_{i_1} \cdots a_{i_r}} \right\rfloor.$$

Подставляя последнее выражение в (1.99) и учитывая (1.98), получаем требуемое. ■

Следствие 1.27. Из теоремы Лежандра легко получить уже известную нам формулу (1.54) для функции Эйлера.

Доказательство. В теореме Лежандра заменим a_i на p_i — простые делители числа n . Тогда

$$N_{i_1 \dots i_r} = \frac{n}{p_{i_1} \cdots p_{i_r}}.$$

Подставим последнее выражение в (1.100):

$$\begin{aligned} \varphi(n) &= n - \sum_{1 \leq i \leq k} \frac{n}{p_i} + \sum_{1 \leq i < j \leq k} \frac{n}{p_i p_j} - \sum_{1 \leq i < j < s \leq k} \frac{n}{p_i p_j p_s} + \dots + (-1)^k \frac{n}{p_1 \cdots p_k} \\ &= n \left(1 - \frac{1}{p_1}\right) \left(1 - \frac{1}{p_2}\right) \cdots \left(1 - \frac{1}{p_k}\right), \end{aligned}$$

получаем (1.54). ■

Теорему 1.78 можно обобщить для числа элементов, обладающих ровно r свойствами.

Теорема 1.80.

$$\begin{aligned} N(r) &= C_r^r N[r] + \cdots + (-1)^i C_{r+i}^r N[r+i] + \cdots + (-1)^{n-r} C_n^r N[n] = \\ &= \sum_{i=0}^{n-r} (-1)^i C_{r+i}^r N[r+i]. \end{aligned} \tag{1.101}$$

Доказательство. Покажем, что каждый элемент, обладающий ровно r свойствами, учитывается в правой части (1.101) ровно один раз. Действительно, элементы, обладающие числом свойств меньше r , очевидно не учитываются. Элемент, обладающий фиксированными $s = r+j$ свойствами ($j \geq 0$), учитывается во внутренней сумме (1.101) ровно C_{r+j}^{r+i} раз. Используя свойства биномиальных коэффициентов (теорема 1.65 и следствие 1.22 к теореме 1.62), имеем

$$\sum_{i=0}^{n-r} (-1)^i C_{r+i}^r C_{r+j}^{r+i} = \sum_{i=0}^{n-r} (-1)^i C_{r+j}^r C_j^i = C_{r+j}^r \sum_{i=0}^j (-1)^i C_j^i = \begin{cases} 1 \text{ при } j = 0, \\ 0 \text{ при } j > 0. \end{cases}$$

Таким образом, в (1.101) элементы, обладающие ровно r свойствами, учитываются ровно по одному разу, а прочие элементы не учитываются. ■

Пример 1.107. Из 50 студентов 30 изучают английский язык, 25 — немецкий, 25 — французский, 15 — английский и немецкий, 20 — английский и французский, 10 — немецкий и французский, 5 — английский, французский и немецкий.

1. Сколько студентов не изучает ни один из перечисленных языков?
2. Сколько студентов изучает ровно два языка?
3. Сколько студентов изучает не менее двух языков?

Обозначим признаки изучения английского, немецкого и французского языков соответственно $p(1), p(2), p(3)$. Согласно принципу включения–исключения (теорема 1.78), число студентов, не изучающих ни один язык, равно

$$N(0) = N[0] - N[1] + N[2] - N[3] = 50 - (30 + 25 + 25) + (15 + 20 + 10) - 5 = 10.$$

Для ответа на второй вопрос воспользуемся теоремой 1.80. Тогда

$$N(2) = C_2^2 N[2] - C_3^2 N[3] = (15 + 20 + 10) - 3 \cdot 5 = 30.$$

Ответ на третий вопрос следует из формулы (1.98):

$$N[2] = N(2) + N(3) = 30 + 5 = 35.$$

Определение 1.59. Перестановку из n целых чисел $1, 2, \dots, n$, в которой ни один элемент не занимает своего естественного места, называют *беспорядком*. Например, при $n = 3$ перестановка $(3, 1, 2)$ является беспорядком, а $(3, 2, 1)$ не является, так как элемент 2 занимает в ней свое естественное место.

Пример 1.108 (задача о числе беспорядков). Найдем с помощью принципа включения–исключения общее число беспорядков. Обозначим его d_n .

Замечание 1.72. Задачу о числе беспорядков называют также *задачей Монмора*² или *задачей о встречах*.

Обозначим через $p(i)$ свойство перестановки, состоящее в том, что i -я позиция в ней занимается числом i . Так как $P_n = n!$, то $N = n!$. Очевидно, что $N_{i_1 \dots i_r}$ — число перестановок, в которых позиции i_1, \dots, i_r должны быть заняты соответствующими числами, а остальные $n - r$ позиций могут быть заполнены оставшимися $n - r$ числами произвольным образом. Тогда $N_{i_1 \dots i_r} = (n - r)!$.

Имеется C_n^r способов выбрать r позиций i_1, \dots, i_r из n возможных, так что

$$N[r] = C_n^r (n - r)! = \frac{n!}{r!}.$$

Согласно принципу включения–исключения (см. теорему 1.78),

$$d_n = n! \left[1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \dots + (-1)^r \frac{1}{r!} + \dots + (-1)^n \frac{1}{n!} \right] = n! \sum_{r=0}^n \frac{(-1)^r}{r!}.$$

Заметим, что $d_n/n!$ равно сумме первых $(n + 1)$ членов ряда Тейлора при разложении $e^{-1} = 0.36806\dots$. При этом скорость сходимости к e^{-1} достаточно велика: $d_6/6! = 0.36788\dots$

Величину d_n часто называют *субфакториалом*. Субфакториал обладает свойствами, похожими на свойства обычного факториала. Действительно, справедливы рекуррентные соотношения

$$\begin{aligned} n! &= (n - 1) [(n - 1)! + (n - 2)!], \\ d_n &= (n - 1) [d_{n-1} + d_{n-2}]. \end{aligned} \tag{1.102}$$

Рекуррентную формулу (1.102) для d_n легко преобразовать к виду

$$d_n = nd_{n-1} + (-1)^n. \tag{1.103}$$

Используя формулу (1.103), вычислим значения первых субфакториалов (табл. 1.48).

Упражнение 1.18. Докажите соотношения (1.102).

²Пьер Монмор — французский математик (1678–1719). Настоящая фамилия — Ремон, Монмор — название его поместья.

Таблица 1.48

n	1	2	3	4	5	6	7	8	9	10	11
d_n	0	1	2	9	44	255	1784	14273	128456	1284561	14130170

1.6.15. Обращение Мёбиуса

Найденная в предыдущем подразделе формула включений–исключений (1.78) относится к так называемым формулам обращения. Еще одна знаменитая формула обращения связана с суммированием по множеству делителей числа n . Эта формула использует введенную ранее функцию Мёбиуса $\mu(n)$ (определение 1.25) и имеет обобщение на случай произвольных частично упорядоченных множеств, о которых речь пойдет во второй части.

Историческая справка. А.Ф.Мёбиус (A.F.Mobius) (1790–1868) — немецкий математик, работал ассистентом Гаусса. Им был сделан значительный вклад в геометрию, в частности топологию (лента Мёбиуса названа его именем). В 1840 г. им впервые была сформулирована «проблема четырех красок».

Рассмотрим функцию, заданную на множестве делителей натурального числа n следующим образом

$$g(n) = \sum_{d|n} f(d).$$

Пример 1.109. Рассмотрим натуральное число n , представленное в каноническом виде (1.6)

$$n = p_1^{k_1} p_2^{k_2} \cdots p_k^{k_l}, \quad k_i \geq 1, \quad i \in 1 : l.$$

1. Пусть $N(n)$ — число делителей числа n . Тогда $N(n) = \sum_{d|n} 1$, т. е. $f(d) = 1$. Значения этой функции для небольших значений аргумента:

$$N(2) = 2, \quad N(6) = 4, \quad N(12) = 6, \quad N(36) = 9, \quad N(60) = 12.$$

Заметим, что в общем случае число делителей находится по формуле

$$N(n) = (k_1 + 1)(k_2 + 1) \cdots (k_l + 1).$$

2. Пусть $S(n)$ — сумма всех делителей числа n . Тогда $S(n) = \sum_{d|n} d$, т. е. $f(d) = d$. Значения этой функции для небольших значений аргумента:

$$\begin{aligned} S(2) &= 3, \quad S(4) = 7, \quad S(6) = 12, \quad S(10) = 18, \quad S(12) = 28, \\ S(20) &= 42, \quad S(30) = 72, \quad S(36) = 91, \quad S(60) = 168. \end{aligned}$$

Можно ли решить обратную задачу и найти функцию $f(n)$ через значения функции $g(d)$ на множестве делителей числа n ?

Попробуем найти эту формулу, применив принцип включений–исключений. Не умаляя общности в качестве функции $g(n)$ рассмотрим количество всех делителей числа $N(n)$. Тогда задача будет состоять в том, чтобы через количество делителей всех делителей данного числа выразить само число.

Обозначим $N'(i_1, \dots, i_k)$ — число всех делителей числа $n/(p_{i_1} \cdots p_{i_k})$, где p_{i_1}, \dots, p_{i_k} — различные простые делители числа n . По формуле включений–исключений (1.99)

$$N'(0) = N(n) - \sum_{p_i|n} N'(i) + \sum_{i_1 < i_2} N'(i_1, i_2) - \cdots + (-1)^k \sum_{i_1 < \cdots < i_k} N'(i_1, \dots, i_k), \quad (1.104)$$

где $N'(0)$ — число делителей числа n , которые не являются делителями других делителей числа n , таким числом является только само n : $N'(0) = 1$.

Учитывая определение 1.25 функции Мёбиуса $\mu(n)$, данное выражение можно переписать так:

$$1 = \sum_{d|n} \mu(d) N\left(\frac{n}{d}\right). \quad (1.105)$$

Действительно, если $d = 1$, то $\mu(d) = 1$ и получается первый член суммы; если d имеет кратные простые делители, то $\mu(d) = 0$ и соответствующие члены в сумме будут отсутствовать. Оставшиеся члены в точности совпадают с членами суммы (1.104).

Пример 1.110. Подставим в формулу (1.105) данные из примера 1.109:

$$1 = N(12) - (N(6) + N(4)) + N(2),$$

$$1 = N(60) - (N(30) + N(20) + N(12)) + (N(10) + N(6) + N(4)) - N(2).$$

Заметим, что в правой части встречаются суммы не всех делителей, например, в первую сумму не входит $N(3)$, во вторую — $N(15)$.

Упражнение 1.19. Проверьте полученную формулу на функции $S(n)$ из примера 1.109.

В общем случае получаем такой результат.

Теорема 1.81 (формула обращения Мёбиуса).

$$\text{Если } g(n) = \sum_{d|n} f(d), \text{ то } f(n) = \sum_{d|n} \mu(d) g\left(\frac{n}{d}\right).$$

Рассмотрим применение этой формулы для решения следующей комбинаторной задачи.

Задача 1.28. Сколько существует циклических последовательностей длины n , составленных из алфавита, в котором r различных символов?

Замечание 1.73. Описанную в задаче 1.28 последовательность можно интерпретировать как бесконечную, полученную повторением начального куска длины n в обе стороны, тогда конечные последовательности

$$[a_1, a_2, \dots, a_n], [a_2, a_3, \dots, a_n, a_1], \dots, [a_n, a_1, \dots, a_{n-1}]$$

следует считать одинаковыми, так как они соответствуют одной бесконечной в обе стороны последовательности. Трудность подсчета состоит в том, что в самой конечной последовательности могут быть циклы различной длины.

Пример 1.111. Рассмотрим циклические последовательности длины 6 над алфавитом из двух символов $\{A, B\}$. Их четырнадцать:

$$\begin{array}{ccccc} [AAAAAA] & [AAAAAB] & [AAAABB] & [AAABAB] & [AABAAB] \\ [AAABBB] & [AABABB] & [ABAABB] & [ABABAB] & [BBBABBA] \\ [BBBABA] & [BBBBAA] & [BBBBBA] & [BBBBBB] & \end{array}$$

Эти последовательности можно разбить на группы по длине периода:

Период длины 1: $[AAAAAA]$, $[BBBBBB]$.

Период длины 2: $[ABABAB]$.

Период длины 3: $[AABAAB]$, $[BBBABBA]$.

Период длины 6: $[AAAAAB]$, $[AAAABB]$, $[AAABAB]$, $[AAABBB]$,
 $[AABABB]$, $[ABAABB]$, $[BBBABA]$, $[BBBBAA]$,
 $[BBBBBA]$.

Из конечной последовательности с периодом длины k циклическим сдвигом периода можно получить $k - 1$ других конечных последовательностей, представляющих ту же самую циклическую последовательность.

Пример 1.112. Из последовательности $[AABAAB]$ циклическим сдвигом периода можно получить еще две последовательности $[BAABAA]$ и $[ABAABA]$.

Обозначим число циклических последовательностей длины n над алфавитом из r символов с периодом k как $M(k)$.

Замечание 1.74. Более аккуратно это число надо обозначать как $M(n, r, k)$, так как оно зависит от количества символов в алфавите и длины конечной последовательности.

Добавив к каждой конечной последовательности с периодом длины k остальные $k - 1$ конечных последовательностей, полученных циклическим сдвигом периода, получим $kM(k)$ последовательностей. Если их суммировать по всем k , являющихся делителями n , то получим множество всех слов из n букв над алфавитом из r символов. Число таких слов равно r^n .

Тогда:

$$r^n = \sum_{k|n} kM(k). \quad (1.106)$$

Пример 1.113. Подставив в формулу (1.106) данные из примера 1.110, получим

$$2^6 = 1 \cdot 2 + 2 \cdot 1 + 3 \cdot 2 + 6 \cdot 9,$$

что истинно.

Применив к формуле (1.106) формулу обращения Мебиуса (см. теорему 1.81), имеем:

$$nM(n) = \sum_{k|n} \mu(k)r^{\frac{n}{k}}, \text{ откуда } M(n) = \frac{1}{n} \sum_{k|n} \mu(k)r^{\frac{n}{k}}. \quad (1.107)$$

Пример 1.114. Подставив в формулу (1.107) $n = 6, r = 2$, получим

$$\begin{aligned} M(6) &= \frac{1}{6} \sum_{k|6} \mu(k)2^{\frac{6}{k}} = \frac{1}{6} (1 \cdot 2^6 - 1 \cdot 2^3 - 1 \cdot 2^2 + 1 \cdot 2^1) = \\ &= \frac{1}{6} (64 - 8 - 4 + 2) = \frac{54}{6} = 9 \end{aligned}$$

число циклических последовательностей с периодом длины 6.

Таким образом, общее число циклических последовательностей длины n над алфавитом из r букв вычисляется по формуле:

$$M = \sum_{d|n} M(d) = \sum_{d|n} \frac{1}{d} \sum_{k|d} \mu(k)r^{\frac{d}{k}}.$$

Задача 1.29. На основе формулы Мебиуса выведите формулу для подсчета числа циклических последовательностей длины n , которые можно получить из набора символов, в котором ровно b_i элементов i -го рода и $b_1 + b_2 + \dots + b_r = n$.

Пример 1.115. Рассмотрим циклические последовательности длины 6, построенные из четырех букв A ($b_1 = 4$) и двух букв B ($b_2 = 2$). Таких последовательностей три: $[AAAABB]$, $[AAABAB]$, $[AABAAB]$, из них первые две имеют период длины 6, а последняя — период длины 3.

1.6.16. Понятие о группе: теорема Бернсайда

□ Самосовмещения правильного тетраэдра

Рассмотрим правильный тетраэдр — треугольную пирамиду, грани которой правильные треугольники. Поставим тетраэдр на лист бумаги и обведем основание. Теперь поднимем тетраэдр, повернем и поставим обратно какой-либо гранью на начертанный треугольник. При этом тетраэдр займет то же место в пространстве, но его положение изменится (говорят, что мы сделали самосовмещение тетраэдра движением первого рода, т. е., движением состоящим из поворотов и параллельных переносов).

Сколько существует различных движений такого сорта?

Начнем перечисление, упорядочив перебор вариантов. Переберем сначала преобразования, сохраняющие положение какой-нибудь вершины (повороты тетраэдра вокруг его высот: рис. 1.18 *б, в, г, д*).

Далее будем действовать более формально. Обозначим начальные положения вершин тетраэдра цифрами 1, 2, 3, 4 (рис. 1.18,*а*), тогда преобразование тетраэдра запишем в виде так называемой подстановки.

Например,

$$a_1 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 2 & 4 \end{pmatrix}, \quad a_2 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 1 & 4 \end{pmatrix}$$

описывают повороты тетраэдра относительно высоты, проходящей через вершину 4. Читать такую запись, например, для a_1 , можно так: вершина 1 переходит на место вершины 3, вершина 2 — на место вершины 1 и т. д. Подстановочная запись очень удобна, если необходимо узнать результат нескольких последовательных преобразований. Например, найдем, как будет выглядеть положение тетраэдра после последовательного выполнения поворотов

$$a_1 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 2 & 4 \end{pmatrix}, \quad a_4 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 3 & 2 \end{pmatrix}.$$

Проследим за перемещением вершин тетраэдра:

$$\begin{array}{ll} 1 \mapsto 3 \text{ (после } a_1) & 3 \mapsto 3 \text{ (после } a_4) \\ 2 \mapsto 1 \text{ (после } a_1) & 1 \mapsto 4 \text{ (после } a_4) \\ 3 \mapsto 2 \text{ (после } a_1) & 2 \mapsto 1 \text{ (после } a_4) \\ 4 \mapsto 4 \text{ (после } a_1) & 4 \mapsto 2 \text{ (после } a_4) \end{array}$$

Эти перемещения тоже описываются подстановкой, которую называют произведением подстановок a_1, a_4 и обозначают $a_1 \circ a_4$. Этой подстановке соответствует самосовмещение тетраэдра, обозначенное на рисунке рис. 1.18

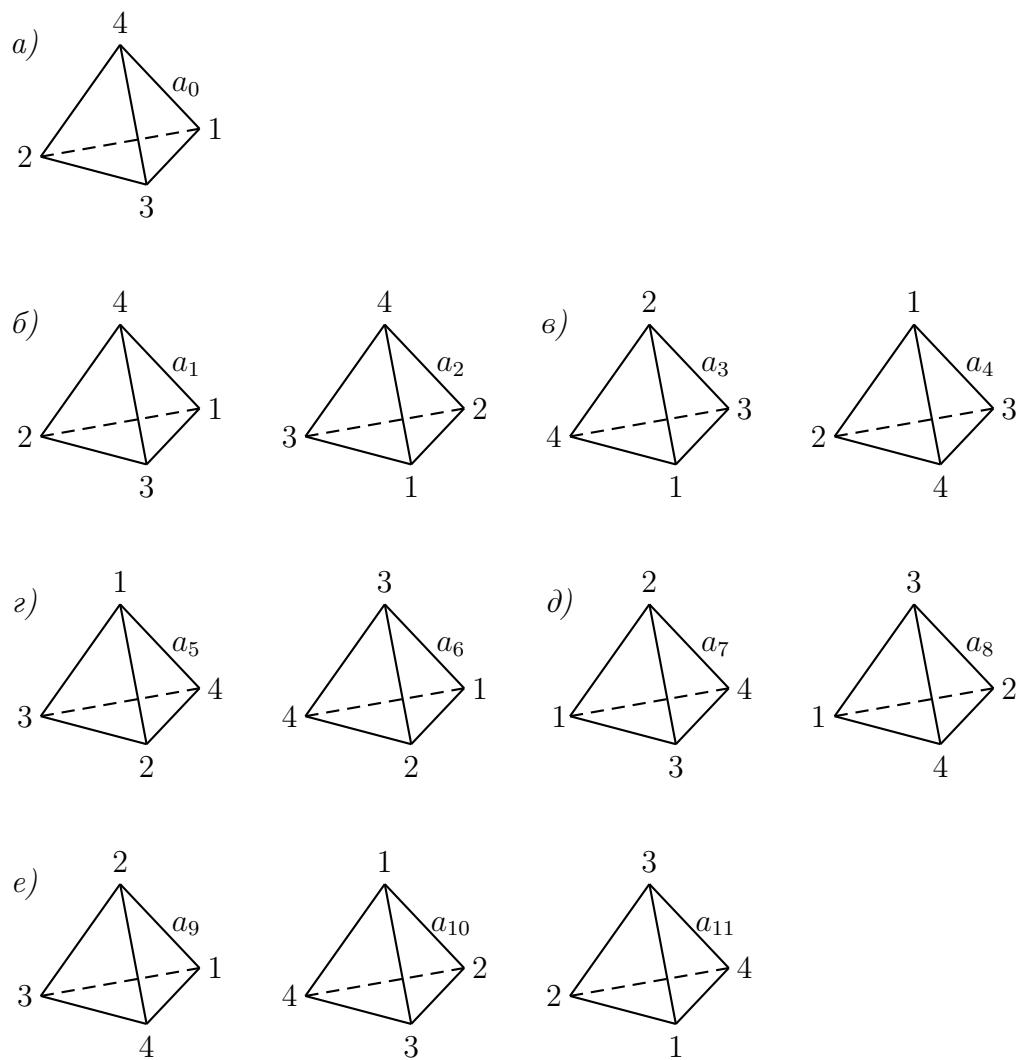


Рис. 1.18

как a_9 :

$$a_9 = a_1 \circ a_4 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \end{pmatrix}$$

Заметим, что это преобразование попарно меняет местами положения вершин 1, 3 и 2, 4 (рис. 1.18, e). Можно наглядно представить это преобразование как поворот на 180° относительно середин противоположных ребер. Обратим внимание, что если выполнить преобразования a_1, a_4 в другом порядке, то получится другой результат:

$$a_4 \circ a_1 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{pmatrix} = a_{10}$$

Для подстановок есть другой способ их записи, который называют записью

в виде произведения независимых циклов.

$$a_9 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \end{pmatrix} = (1\ 3)(2\ 4), \quad a_1 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 2 & 4 \end{pmatrix} = (1\ 3\ 2)(4)$$

Последняя запись означает, что вершины 1, 2, 3 циклически переходят друг в друга (см. рис. 1.18,б), а a_4 остается на месте (переходит в себя). Кроме a_9 есть еще два преобразования, не сохраняющие положения ни одной из вершин (см. рис. 1.18,д). Обратим внимание на различные закономерности, которым удовлетворяют поворотные самосовмещения тетраэдра, например:

$$a_1 \circ a_3 = a_6, \quad a_1^2 = a_1 \circ a_1 = a_2, \quad a_9^2 = a_0.$$

а среди них — на произведения, которые в результате дают преобразование a_0 , т. е. возвращают тетраэдр в исходное состояние, например,

$$a_1 \circ a_2 = a_0, \quad a_3 \circ a_4 = a_0, \dots a_0^2 = a_0, \quad a_9^2 = a_0.$$

Геометрически это означает, что a_2 является обратным преобразованием к a_1 . Аналогично, a_4 является обратным к a_3 , a_0 — к a_0 и a_9 — к a_9 . Это обозначают следующим образом:

$$a_2 = a_1^{-1}, \quad a_4 = a_3^{-1}, \quad a_0 = a_0^{-1}, \quad a_9 = a_9^{-1}.$$

Сформулируем основные свойства описанных преобразований тетраэдра (подстановок):

1) существует нейтральный элемент

$$a_0 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{pmatrix}$$

не меняющий положения тетраэдра;

2) у каждого элемента (некоторого преобразования тетраэдра) из полученного множества есть обратный (возвращающий тетраэдр в исходное состояние);

3) произведением двух подстановок множества является подстановка, соответствующая двум последовательным преобразованиям тетраэдра;

4) произведение подстановок (преобразований тетраэдра) обладает свойством ассоциативности, но не обладает, вообще говоря, свойством коммутативности.

Эти свойства определяют на множестве подстановок математическую структуру, которую называют *некоммутативной группой умножения подстановок* (см. определение 1.17).

□ **Подгруппы и классы смежности**

Определение 1.60. Если множество G конечно, то группу называют *конечной*, а число ее элементов — *порядком* группы. Обозначают порядок группы как $|G|$.

Пример 1.116. Порядок описанной группы поворотных самосовмещений тетраэдра равен 12.

Поставим теперь такую задачу: можно ли получить все положения тетраэдра, используя не все, а одно или несколько различных преобразований, если при этом их можно осуществлять любое количество раз в любых комбинациях, перемножая и обращая.

Пример 1.117. Возьмем преобразование

$$a_1 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 2 & 4 \end{pmatrix}.$$

Преобразование a_2 можно получить повторным выполнением преобразования a_1 :

$$a_2 = a_1 \circ a_1 = a_1^2,$$

или обращением преобразования a_1 :

$$a_2 = a_1^{-1}.$$

Далее, если умножить a_1 на себя трижды или, что то же самое, умножить a_1 на уже полученное a_2 , то результатом будет a_0 :

$$a_1^3 = a_1 \circ a_2 = a_0.$$

Как бы мы ни старались комбинировать эти подстановки, новых преобразований не получим. Таким образом, выразить все элементы группы G через a_1 нельзя.

В примере 1.117 мы построили группу G_1 меньшего размера, являющуюся частью группы всех поворотов: $G_1 = \{a_0, a_1, a_2\}$. Аналогично можно построить группы $G_2 = \{a_0, a_3, a_4\}$, $G_3 = \{a_0, a_5, a_6\}$ и т. д., которые называют *подгруппами* исходной группы G . Порядок каждой из подгрупп G_1, G_2, G_3 равен 3.

Определение 1.61. Если $G_1 \subset G$ и G_1, G являются группами, то G_1 называют *подгруппой* G .

Определение 1.62. Множества вида $aG_1 = \{a \circ g \mid g \in G_1\}$, где a — фиксированный элемент группы G , называют *классами смежности* G по подгруппе G_1 .

Очевидно, если элемент a принадлежит подгруппе G_1 , то $aG_1 = G_1$.

Пример 1.118. Рассмотрим классы смежности G по G_1 , не совпадающие с G_1 :

$$\begin{aligned} a_3G &= \{a_3 \circ a_0, a_3 \circ a_1, a_3 \circ a_2\} &= \{a_3, a_7, a_9\}; \\ a_4G &= \{a_4 \circ a_0, a_4 \circ a_1, a_4 \circ a_2\} &= \{a_4, a_{10}, a_5\}; \\ a_5G &= \{a_5 \circ a_0, a_5 \circ a_1, a_5 \circ a_2\} &= \{a_5, a_4, a_{10}\}; \\ a_6G &= \{a_6 \circ a_0, a_6 \circ a_1, a_6 \circ a_2\} &= \{a_6, a_{11}, a_8\}; \\ a_7G &= \{a_7 \circ a_0, a_7 \circ a_1, a_7 \circ a_2\} &= \{a_7, a_9, a_3\}; \\ a_8G &= \{a_8 \circ a_0, a_8 \circ a_1, a_8 \circ a_2\} &= \{a_8, a_6, a_{11}\}; \\ a_9G &= \{a_9 \circ a_0, a_9 \circ a_1, a_9 \circ a_2\} &= \{a_9, a_3, a_7\}; \\ a_{10}G &= \{a_{10} \circ a_0, a_{10} \circ a_1, a_{10} \circ a_2\} &= \{a_{10}, a_5, a_4\}; \\ a_{11}G &= \{a_{11} \circ a_0, a_{11} \circ a_1, a_{11} \circ a_2\} &= \{a_{11}, a_8, a_6\}. \end{aligned}$$

Из примера 1.118 можно отметить закономерность — классы смежности либо совпадают, либо не пересекаются. Число классов смежности равно частному от деления порядка группы на порядок подгруппы.

Теорема 1.82. Если G_1 подгруппа группы G , и $p = |G|$, $p_1 = |G_1|$ — порядки групп G и G_1 , то

- 1) p делится на p_1 ;
- 2) множество классов смежности $\{gG_1 \mid g \in G\}$ состоит из p/p_1 непересекающихся множеств, каждое из которых имеет p_1 элементов.

Доказательство.

1) Докажем, что классы смежности не пересекаются. От противного: пусть найдутся два совпадающие элемента из разных классов смежности: $c \in aG_1$ и $c \in bG_1$.

Это означает существование элементов g_1 и g_2 подгруппы G_1 таких, что $a \circ g_1 = b \circ g_2$, откуда следует

$$a = b \circ g_2 \circ g_1^{-1} = b \circ f, \text{ где } f = g_2 \circ g_1^{-1} \in G_1.$$

Таким образом элемент a принадлежит классу bG_1 , но тогда и все элементы класса aG_1 попадут в класс bG_1 .

Действительно, любой элемент класса aG_1 записывается как

$$a \circ g = (b \circ f) \circ g = b \circ (f \circ g) = b \circ h \in bG_1, \text{ где } f, g, h \in G_1.$$

Следовательно, предположение о непустом пересечении привело к совпадению классов.

2) Докажем, что количество элементов одного класса смежности равно числу элементов подгруппы. Также от противного: пусть $a \circ g_1 = a \circ g_2$ для разных элементов подгруппы, тогда, умножив обе части на элемент, обратный a , получим:

$$a^{-1} \circ (a \circ g_1) = a^{-1} \circ (a \circ g_2) \text{ или } (a^{-1} \circ a) \circ g_1 = (a^{-1} \circ a) \circ g_2,$$

откуда $g_1 = g_2$, что и дает противоречие с условием.

Таким образом, мы разбили все p элементов исходной группы G на k непересекающихся классов, в каждом из которых число элементов равно числу элементов p_1 подгруппы G_1 . Значит, $p = kp_1$. ■

Вернемся к задаче о порождающих элементах группы.

Определение 1.63. Набор элементов группы называют *системой образующих*, если любой элемент группы можно выразить как некоторое произведение образующих или им обратных.

Если система образующих состоит из одного элемента, то такую группу называют циклической.

Пример 1.119. Определенная в предыдущих примерах подгруппа G_1 вращений относительно высоты, проходящей через вершину 4, задается одним образующим элементом a_1 :

$$a_2 = a_1^2, \quad a_0 = a_1^3.$$

Если в качестве образующих взять элементы a_1 и a_3 , то получим все элементы группы поворотов тетраэдра.

Задача 1.30. Выразите все элементы группы G через a_1 и a_3 , используя операции умножения и, если надо, обращения. Можно ли обойтись только операциями умножения?

Определение 1.64. Порядком элемента группы g называют наименьший показатель степени p , в которую надо возвести элемент (p раз перемножить элемент g), чтобы получить нейтральный элемент e :

$$\underbrace{g \circ g \circ \cdots \circ g}_p = g^p = e.$$

Замечание 1.75. Обратите внимание, что когда образующий элемент один, т. е. когда подгруппа циклическая, порядок подгруппы G_1 равен порядку образующего ее элемента.

Пример 1.120. Порядок преобразования a_1 , описывающего поворот тетраэдра на 120° относительно высоты, проходящей через вершину 4, равен 3.

Задача 1.31. Попробуйте в качестве образующих использовать другие элементы, например,

$$a_9 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \end{pmatrix} \text{ и } a_{10} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{pmatrix}.$$

Все ли повороты тетраэдра будут порождаться этими элементами?

Задача 1.32.

1. Найдите другие подгруппы G .
2. Существуют ли у группы поворотов тетраэдра подгруппы порядка 1, 4, 6? Если да, то постройте их.

Изоморфные группы

Группу можно описать и по-другому — таблицей умножения, которую в теории групп называют *таблицей Кэли*.

Пример 1.121. Для описанных выше подгрупп G_1 и G_2 таблицы Кэли имеют вид (табл. 1.49 *a* и *b*):

Таблица 1.49

<i>a</i>			<i>b</i>				
G_1	a_0	a_1	a_2	G_2	a_0	a_3	a_4
a_0	a_0	a_1	a_2	a_0	a_0	a_3	a_4
a_1	a_1	a_2	a_0	a_3	a_3	a_4	a_0
a_2	a_2	a_0	a_1	a_4	a_4	a_0	a_3

Подгруппы G_1, G_2 — это группы самосовмещений тетраэдра при поворотах вокруг его высот.

Группы G_1, G_2 изоморфны. Это означает, что можно переименовать элементы одной группы так, что таблица умножения одной группы перейдет в таблицу умножения другой. В нашем примере это достигается таким переименованием:

$$a_1 \mapsto a_3, \quad a_2 \mapsto a_4.$$

Оказывается, подстановками можно описать любую конечную группу. Имеет место следующий результат.

Теорема 1.83. Каждая конечная группа изоморфна некоторой группе подстановок.

Группа симметрий тетраэдра

В заключение рассмотрим группу, определяемую всевозможными подстановками на множестве из четырех элементов. Подсчитаем число таких подстановок: первый элемент можно переместить в любое из четырех положений, второй — в любое из трех оставшихся, третий — в любое из двух оставшихся, четвертый же поставить на единственное свободное место. Тогда с каждым из четырех способов перемещения первого элемента комбинируется три способа перемещения второго и два способа перемещения третьего, что дает $4 \cdot 3 \cdot 2 = 24$ комбинации.

Из них 12 подстановок соответствуют поворотам тетраэдра. Какой геометрический смысл имеют оставшиеся 12? Возьмем, например, такую подстановку:

$$a_{12} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 2 & 4 \end{pmatrix}.$$

Она означает такое преобразование тетраэдра, при котором две вершины поменяются местами, а две другие останутся в прежнем положении. Нетрудно увидеть, что такой поворот невозможен.

Однако если зеркально отобразить тетраэдр (рис. 1.19), то после этого такое совмещение вершин станет возможным.

Таким образом, если к преобразованиям поворота добавить преобразования зеркального отражения относительно сечений тетраэдра, каждое из которых проходит через ребро и соответствующую высоту (см. рис. 1.19), то все подстановки будут соответствовать некоторым преобразованиям тетраэдра. Порядок этой самой большой группы преобразований будет 24, а группа поворотов станет подгруппой этой группы. Обратите внимание, что исходная группа поворотных самосовмещений является подгруппой группы симметрий тетраэдра и, соответственно, ее порядок является делителем порядка вновь построенной группы (24 делится на 12).

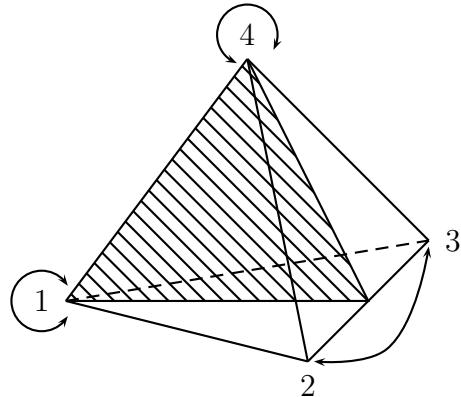


Рис. 1.19

□ Теорема Бернсайда

Поставим задачу: сколькими способами можно раскрасить грани правильного тетраэдра n красками, если считать разными те раскраски, которые не получаются друг из друга самосовмещениями тетраэдра при поворотах.

Пример 1.122. Рассмотрим раскраску граней тетраэдра двумя красками, например, красной и белой.

Перечислим все пять возможных раскрасок:

- все грани белые,
- одна грань красная,
- две грани красные (всего одна возможная раскраска, так как любые две грани имеют общее ребро, и значит, любые две такие раскраски перейдут друг в друга при повороте, совмещающем эти ребра),
- три грани красные,
- все грани красные.

Заметим, что если не отождествлять раскраски, совпадающие при поворотных самосовмещениях, то число раскрасок будет $2^4 = 16$.

В общем случае задачу можно поставить так. Имеется множество X (множество раскрасок), на котором определена группа преобразований G (переводящих одни раскраски в другие). Требуется найти количество орбит, т. е., количество подмножеств, на которые разбивается все множество раскрасок по заданной группе G . В одну орбиту попадают все элементы множества, которые переводятся друг в друга некоторым преобразованием группы G .

Пример 1.123. Показанные на рис. 1.20 раскраски тетраэдра входят в одну орбиту относительно группы поворотных самосовмещений тетраэдра. Преобразования a_7 и a_8 , описанные ранее, переводят их друг в друга соответственно.

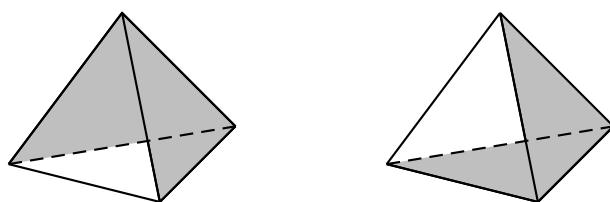


Рис. 1.20

Теорема Бернсайда позволяет свести нахождение числа орбит к нахождению для каждого преобразования группы числа раскрасок, которые оно не меняет.

Пример 1.124. Преобразование a_1 , описанное ранее, переводит в себя следующие раскраски тетраэдра двумя красками:

- все грани одного цвета (две раскраски),

- основание одного цвета, а все боковые грани — другого (две раскраски).

Теорема 1.84 (теорема Бернсайда). Число орбит получается, если для каждого преобразования найти число раскрасок, которое оно не меняет, результаты сложить и поделить на число элементов группы:

$$|S| = \frac{1}{|G|} \sum_{g \in G} \eta(g). \quad (1.108)$$

Замечание 1.76. Традиционно это утверждение называют леммой Бернсайда.

Пример 1.125 (к применению теоремы Бернсайда). Для задачи раскраски тетраэдра двумя красками получаем:

$\eta(a_0) = 16$ (тождественное преобразование сохраняет все раскраски);

$$\eta(a_1) = \dots = \eta(a_8) = 4; \quad \eta(a_9) = \eta(a_{10}) = \eta(a_{11}) = 4, \quad |G| = 12;$$

$$|S| = \frac{16 + 4 \cdot 11}{12} = \frac{60}{12} = 5.$$

Пример 1.126 (к доказательству теоремы Бернсайда). Основная идея доказательства видна из табл. 1.50, столбцы которой соответствуют различным преобразованиям, а строки - раскраскам.

Таблица 1.50

Продолжение табл. 1.50

	a_0	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}
1101	+					+	+					
1110	+							+	+			
1111	+	+	+	+	+	+	+	+	+	+	+	+

Здесь раскраска граней тетраэдра закодирована следующим образом: 1 — красная, 0 — белая, первый элемент набора — цвет основания, второй — цвет левой грани, третий — цвет задней грани, четвертый — цвет правой грани. В табл. 1.50 знаком «+» отмечены те клетки, для которых соответствующее преобразование (столбец) не меняет соответствующей раскраски (строка).

Объединим в табл. 1.50 клетки тех строк, которые соответствуют одной раскраске (на рис. 1.20 отмечена одна из таких групп). Заметим, что в каждой группе будет ровно 12 отмеченных клеток, а наша группа имеет 12 элементов. Это мы и должны обосновать в общем случае.

Доказательство теоремы Бернсайда.

Рассмотрим множество пар $(g; x)$ таких, что преобразование g сохраняет раскраску x , т. е. $g(x) = x$. Зафиксируем раскраску x , и рассмотрим все преобразования, переводящие ее в себя. Тогда будет верно следующее утверждение.

Лемма 1.18. Множество элементов группы g , обладающих свойством $g(x) = x$ является подгруппой группы G .

Доказательство. Обозначим $G_1 = \{g \mid g(x) = x\}$. Достаточно доказать, что произведение элементов G_1 также является элементом G_1 . Тогда умножение не будет нас выводить за пределы множества G_1 , а выполнимость групповых свойств для элементов G_1 будет иметь место, так как $G_1 \subset G$:

Пусть $g_1, g_2 \in G_1$. Рассмотрим $g_1 \circ g_2 : g_1 \circ g_2(x) = g_1(g_2(x)) = g_1(x) = x$. ■

Лемма 1.19. Если раскраска x может быть переведена преобразованием f группы G в раскраску y , то

$$|G_1| = |G_2|, \text{ где } G_1 = \{g \mid g(x) = x\}, G_2 = \{g \mid g(y) = y\}.$$

Доказательство. Между G_1 и G_2 существует взаимно-однозначное соответствие, определяемое через f следующим образом. Если $g \in G_1$, то $f \circ g \circ f^{-1} \in G_2$, и наоборот, если $g \in G_2$, то $f^{-1} \circ g \circ f \in G_1$.

Докажем, к примеру, первое утверждение:

$$\begin{aligned} g \in G_1 &\Rightarrow g(x) = x \Rightarrow g(f^{-1}(y)) = f^{-1}(y) \Rightarrow f(g(f^{-1}(y))) = y \Rightarrow \\ &\Rightarrow f \circ g \circ f^{-1}(y) = y. \end{aligned} \quad \blacksquare$$

Упражнение 1.20. Докажите самостоятельно, что указанное соответствие взаимно-однозначно.

Перейдем теперь к доказательству собственно теоремы Бернсайда. Подсчитаем количество пар $\{(g; x) \mid g(x) = x\}$ двумя способами.

Первый способ состоит в том, чтобы для каждого преобразования найти все раскраски x , которые оно сохраняет (суммирование по столбцам). Обозначим для каждого преобразования g множество таких раскрасок $\eta(g)$. Тогда количество искомых пар запишется как $\sum_{g \in G} \eta(g)$.

Второй способ состоит в том, чтобы просуммировать преобразования для раскрасок, входящих в одну орбиту (суммирование по группам строк). Докажем, что в каждой такой сумме будет одинаковое число слагаемых, равное порядку группы $|G|$.

Действительно, преобразования, которые сохраняют данную раскраску (строку), образуют подгруппу G_1 группы G (см. лемму 1.18). Рассмотрим классы смежности группы G по подгруппе $G_1 : fG_1$. По теореме 1.82 о классах смежности классы разбивают все множество преобразований на равные по количеству части, в каждой из которых $|G_1|$ элементов. С другой стороны, существует взаимно-однозначное соответствие между подгруппами преобразований, сохраняющими отдельные элементы орбиты и классами смежности: по лемме 1.19 преобразование вида $f \circ g \circ f^{-1}$ переводит элементы одной подгруппы в другую. Преобразование $f \circ g \in fG_1$, а значит преобразование f^{-1} сопоставляет классу смежности fG_1 подгруппу G_2 . По соображениям, высказанным ранее это сопоставление взаимно-однозначно.

Итак, второй способ подсчета дает: $|S| \times |G|$, где S — множество орбит.

Приравняв полученные в обоих способах вычисления результаты, получим искомую формулу (1.108). ■

Пример 1.127. Применим теорему Бернсайда для подсчета количества циклических последовательностей длины $n = 6$ над алфавитом из двух символов $\{A, B\}$, которые были подсчитаны ранее на основе обращения Мебиуса (см. пример 1.111).

Рассмотрим циклическую группу порядка 6. Ее образующий элемент — циклический сдвиг последовательности на одну позицию — задается подстановкой:

$$a_1 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 3 & 4 & 5 & 6 & 1 \end{pmatrix}.$$

«Раскраска» в данном случае означает конкретную строку из шести символов A и B , «орбита» — раскраски переходящие друг в друга при циклических сдвигах.

Подсчитаем для каждого преобразования число сохраняемых им раскрасок.

Тождественное преобразование

$$a_0 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 2 & 3 & 4 & 5 & 6 \end{pmatrix} = (1)(2)(3)(4)(5)(6)$$

сохраняет все 2^6 последовательностей.

Преобразования

$$a_1 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 3 & 4 & 5 & 6 & 1 \end{pmatrix} = (123456), \quad a_5 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 6 & 1 & 2 & 3 & 4 & 5 \end{pmatrix} = (165432)$$

сохраняют по две последовательности каждой (последовательности, состоящие из одинаковых символов).

Преобразование

$$a_3 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 4 & 5 & 6 & 1 & 2 & 3 \end{pmatrix} = (14)(25)(36)$$

сохраняет 2^3 последовательностей (символы на местах, входящих в один цикл подстановки, должны быть одинаковы).

Преобразования

$$a_2 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 4 & 5 & 6 & 1 & 2 \end{pmatrix} = (135)(246), \quad a_4 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 5 & 6 & 1 & 2 & 3 & 4 \end{pmatrix} = (153)(264)$$

сохраняют по 2^2 последовательности каждой.

Подставив результаты в формуле Бернсайда (1.108), получим:

$$|S| = \frac{1}{6} (2^6 + 2 \cdot 2 + 2^3 + 2 \cdot 2^2) = \frac{84}{6} = 14.$$

1.7. Производящие функции и рекуррентные уравнения

1.7.1. Производящие функции

Рассмотрим сначала несколько примеров.

Пример 1.128. Дано полоска шириной 2 и длины n . Заполняем ее домино — прямоугольниками 1×2 или 2×1 (рис. 1.21).

Поставим задачу перебрать все варианты заполнения полоски костяшками домино для последовательных значений n .

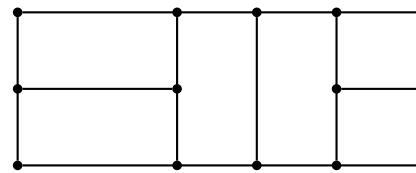
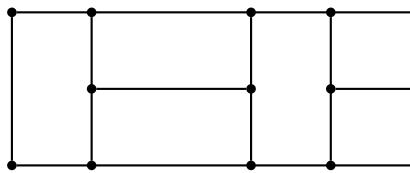


Рис. 1.21

Можно заметить, что горизонтально расположенные домино встречаются только парами. Поэтому, обозначив вертикальное домино буквой x , а пару горизонтальных — y , каждому заполнению полоски можно сопоставить алгебраическое выражение.

Например,

$$xyxy \dots, \quad yxxy \dots$$

Все такие комбинации можно записать в виде суммы (ряда, если рассматривать полоски произвольной длины):

$$x + y + xx + xy + yx \dots$$

Теперь введем новую интерпретацию перемножения алгебраических выражений, соответствующих заполнению полосок. Будем рассматривать умножение двух заполнений полосок одинаковой ширины как простое соединение этих заполненных полосок в одну с большей длиной.

Например,

$$(xxy)(yx) = xxyyx.$$

Теперь перепишем ряд следующим образом:

$$x(1 + x + y + xx + xy + yx + \dots) + y(1 + x + y + xx + xy + yx + \dots).$$

Это корректно, так как если отделить левое вертикальное домино от всех заполнений бесконечной полоски, которые с него начинаются, то опять получим все возможные заполнения бесконечной полоски домино! То же можно сказать и про отделение двух горизонтально расположенных домино.

Что означает 1 в этой сумме? «Пустое домино» — означает, что добавление его к полоске не меняет ее заполнения.

Таким образом, если обозначить все возможные полоски ширины 2, полученные соединением домино (включая пустую), как S , то можно записать уравнение:

$$S = 1 + xS + yS.$$

Используя формально привычные алгебраические операции, получим

$$S = \frac{1}{1 - x - y}.$$

Далее, опять же формально, применим формулу для бесконечно убывающей геометрической прогрессии

$$\frac{1}{1-z} = 1 + z + z^2 + z^3 + \dots$$

и после подстановки $(x+y)$ вместо z будем иметь:

$$\frac{1}{1-x-y} = 1 + (x+y) + (x+y)^2 + (x+y)^3 + \dots$$

Как интерпретировать этот результат?

Например, если нас интересует число полосок, которые можно составить из трех вертикальных и четырех (двух пар) горизонтальных домино, то надо взять слагаемое $(x+y)^{3+2}$ и найти коэффициент при x^3y^2 . Как известно, это биномиальный коэффициент $C_5^3 = 10$.

Таким образом, произведя формальные операции, мы сопоставили бесконечному ряду величину S , являющуюся функцией от x и y

$$S(x, y) = \frac{1}{1-x-y} = \sum_{k=0}^{\infty} (x+y)^k.$$

Пример 1.129. Рассмотрим соотношения:

$$1 + 2 + \dots + n = \frac{n(n+1)}{2}, \quad (1.109)$$

$$1^2 + 2^2 + \dots + n^2 = \frac{n(n+1)(2n+1)}{6}. \quad (1.110)$$

Если обозначить сумму в левой части (1.109) через f_n , а сумму в левой части (1.110) через g_n , то будут иметь место соотношения

$$f_{n+1} - f_n = n + 1; \quad g_{n+1} - g_n = (n+1)^2. \quad (1.111)$$

Взяв в качестве начальных значений $f_1 = g_1 = 1$, с помощью (1.111) можно определить f_2 и g_2 , затем — f_3 и g_3 и т. д. Таким образом последовательности $\{f_n\}$ и $\{g_n\}$ однозначно определяются соотношениями (1.111) и начальными условиями.

Обобщением (1.111) является *линейное рекуррентное¹ уравнение с постоянными коэффициентами*

$$f_{n+r} + a_1 f_{n+r-1} + a_2 f_{n+r-2} + \dots + a_r f_n = \varphi(n), \quad (1.112)$$

выполненное для всех неотрицательных целых чисел n . Коэффициенты a_1, a_2, \dots, a_r — фиксированные числа, причем $a_r \neq 0$, $\varphi(n)$ — заданная функция от n . Число r называют *порядком* уравнения.

¹От лат. *recurrens* — возвращающийся.

Замечание 1.77. Если $a_r = 0$, то уравнение (1.112) можно привести к уравнению меньшего порядка.

Если зафиксировать значения f_0, f_1, \dots, f_{r-1} и рассматривать их как начальные условия, то с помощью рекуррентного уравнения (1.112) можно шаг за шагом однозначно определять значения f_r, f_{r+1}, \dots и таким образом определить всю последовательность $\{f_n\}$.

Поставим в соответствие последовательности $\{f_n\}$ степенной ряд

$$F(z) = f_0 + f_1 z + f_2 z^2 + \dots, \quad (1.113)$$

который называют *производящей функцией* для последовательности $\{f_n\}$.

Замечание 1.78. Последовательность чисел однозначно определяет производящую функцию, но обратное утверждение верно не всегда. Из теории аналитических функций известно, что если указанный степенной ряд сходится в круге радиуса $R > 0$, то коэффициенты $f_0, f_1 \dots$ определяются по $F(z)$ однозначно. В дальнейшем нам будет неважно, при каких значениях z сходится соответствующий ряд. Мы не будем вычислять значение производящей функции в конкретной точке z , будем только выполнять некоторые формальные операции на таких рядах, а затем сравнивать коэффициенты при отдельных степенях z .

Рассмотрим некоторые свойства производящих функций. Пусть

$$G(z) = g_0 + g_1 z + g_2 z^2 + \dots$$

— производящая функция для последовательности $\{g_n\}$, а c и d — произвольные фиксированные числа. Поскольку

$$cF(z) + dG(z) = (cf_0 + dg_0) + (cf_1 + dg_1)z + \dots,$$

то последовательности $\{cf_n + dg_n\}$ отвечает производящая функция

$$cF(z) + dG(z).$$

Возьмем произведение производящих функций

$$H(z) = F(z)G(z) = h_0 + h_1 z + h_2 z^2 + \dots,$$

тогда последовательность чисел $\{h_n\}$ может быть получена из последовательностей $\{f_n\}$ и $\{g_n\}$ с помощью соотношений

$$h_n = f_0 g_n + f_1 g_{n-1} + \dots + f_{n-1} g_1 + f_n g_0. \quad (1.114)$$

Пример 1.130. Найдем производящие функции для последовательностей чисел $\{1\}$ и $\{n\}$. Для первой последовательности получаем

$$1 + z + z^2 + \dots = \frac{1}{1 - z}. \quad (1.115)$$

Последовательности $\{n\}$ соответствует ряд

$$z + 2z^2 + 3z^3 + \cdots = z(1 + 2z + 3z^2 + \cdots). \quad (1.116)$$

Поскольку выражение в скобках в правой части (1.116) получается дифференцированием ряда (1.115), то

$$z(1 + 2z + 3z^2 + \cdots) = z \frac{d}{dz} \left(\frac{1}{1-z} \right) = \frac{z}{(1-z)^2}.$$

Пример 1.131. Обобщая предыдущий пример для последовательности (1.115), легко найти производящую функцию для последовательности $\{a^k\}$.

$$1 + az + a^2z^2 + a^3z^3 + \cdots = \frac{1}{1-az}. \quad (1.117)$$

Пример 1.132. Найдем производящую функцию для разбиений числа n на не более, чем k слагаемых. Обозначим число таких разбиений $p_k(n)$, а саму производящую функцию — $P_k(z)$. В силу свойств сопряженных разбиений $p_k(n)$ равно числу разбиений числа n , с наибольшим слагаемым, не превосходящим k .

Такие разбиения можно рассматривать как решения уравнения

$$\alpha_1 + 2\alpha_2 + \cdots + k\alpha_k = n, \quad \alpha_i \in \mathbb{Z}, \quad \alpha_i \geq 0,$$

где α_i — кратность слагаемого i в разбиении. Тогда согласно (1.113), имеем

$$\begin{aligned} P_k(z) &= p_k(0) + p_k(1)z + p_k(2)z^2 + \cdots = \sum_{n=0}^{\infty} p_k(n)z^n = \sum_{n=0}^{\infty} \sum_{\alpha_1+\cdots+k\alpha_k=n} z^n = \\ &= \sum_{\alpha_1 \geq 0} \cdots \sum_{\alpha_k \geq 0} z^{\alpha_1+\cdots+k\alpha_k} = \left(\sum_{\alpha_1 \geq 0} z^{\alpha_1} \right) \cdots \left(\sum_{\alpha_k \geq 0} z^{k\alpha_k} \right). \end{aligned}$$

Применив (1.115), получим

$$P_k(z) = \frac{1}{1-z} \cdots \frac{1}{1-z^k} = \frac{1}{(1-z) \cdots (1-z^k)} = \prod_{i=1}^k (1-z^i)^{-1}.$$

Используем теперь производящие функции для решения линейного рекуррентного уравнения.

Пример 1.133. Требуется решить уравнение $f_{n+2} - f_{n+1} - f_n = 0$ с начальными условиями $f_0 = f_1 = 1$.

Обозначим через $F(z)$ производящую функцию последовательности чисел $\{f_n\}$. Умножим обе части рекуррентного уравнения на z^{n+2} , получим

$$f_{n+2}z^{n+2} - zf_{n+1}z^{n+1} - z^2f_nz^n = 0, \quad n = 0, 1, 2, \dots$$

Сложив эти равенства для всех n от 0 до ∞ , имеем

$$\sum_{n=0}^{\infty} f_{n+2} z^{n+2} - z \sum_{n=0}^{\infty} f_{n+1} z^{n+1} - z^2 \sum_{n=0}^{\infty} f_n z^n = 0. \quad (1.118)$$

Заметим, что первая сумма в (1.118) равна разности функции $F(z)$ и первых двух членов ее разложения $f_0 + f_1 z = 1 + z$, вторая сумма — разности $F(z)$ и первого члена $f_0 = 1$, а третья сумма — в точности $F(z)$. Поэтому можно записать следующее равенство:

$$[F(z) - (1 + z)] - z[F(z) - 1] - z^2 F(z) = 0.$$

Откуда находим

$$F(z) = \frac{1}{1 - z - z^2}. \quad (1.119)$$

Разложим правую часть (1.119) на простейшие дроби. Для этого решим уравнение $1 - z - z^2 = 0$. Имеем

$$1 - z - z^2 = -(z_1 - z)(z_2 - z) = -z_1 z_2 \left(1 - \frac{z}{z_1}\right) \left(1 - \frac{z}{z_2}\right),$$

где $z_1 = -\frac{1}{2} + \frac{\sqrt{5}}{2}$, $z_2 = -\frac{1}{2} - \frac{\sqrt{5}}{2}$.

Положим

$$\alpha_1 = \frac{1}{z_1} = \frac{1 + \sqrt{5}}{2}, \quad \alpha_2 = \frac{1}{z_2} = \frac{1 - \sqrt{5}}{2}.$$

Заметим, что $z_1 z_2 = -1$. Тогда $1 - z - z^2 = (1 - \alpha_1 z)(1 - \alpha_2 z)$ и, учитывая (1.117), окончательно получаем

$$F(z) = \frac{1}{\sqrt{5}} \left(\frac{\alpha_1}{1 - \alpha_1 z} - \frac{\alpha_2}{1 - \alpha_2 z} \right) = \sum_{n=0}^{\infty} \frac{1}{\sqrt{5}} (\alpha_1^{n+1} - \alpha_2^{n+1}) z^n.$$

Таким образом

$$f_n = \frac{1}{\sqrt{5}} (\alpha_1^{n+1} - \alpha_2^{n+1}), \quad n = 0, 1, 2, \dots \quad (1.120)$$

Формула (1.120) носит название *формулы Бине*¹. Члены последовательности $\{f_n\}$ уже появлялись в разделе 1.2.7, далее были подробно рассмотрены в разделе 1.2.12 и известны, как числа Фибоначчи. При больших n члены приближенно равны $\alpha_1^{n+1} \sqrt{5}$. Как было указано в том же подразделе, коэффициент $\alpha_1 = 1.6180337\dots$ известен в геометрии как «золотое сечение» или число Фидия.

¹Формула впервые опубликована Даниэлем Бернулли в 1728 г. и вновь открыта Жаком Бине в 1843 г.

1.7.2. Решение однородного линейного рекуррентного уравнения

Найдем общее решение однородного линейного рекуррентного уравнения, получающегося из (1.112) при $\varphi(n) = 0$. Метод решения является обобщением решения примера 1.133.

Предположим, что последовательность чисел $\{f_n\}$ удовлетворяет следующему однородному линейному рекуррентному уравнению:

$$f_{n+r} + a_1 f_{n+r-1} + a_2 f_{n+r-2} + \cdots + a_r f_n = 0, \quad (1.121)$$

где a_1, a_2, \dots, a_r — заданные постоянные коэффициенты и $a_r \neq 0$.

Ограничимся случаем $r = 2$

$$f_{n+2} + a_1 f_{n+1} + a_2 f_n = 0. \quad (1.122)$$

Для задания начальных условий фиксируем значения f_0, f_1 . Как и в примере 1.133 обозначим через $F(z)$ производящую функцию последовательности f_0, f_1, \dots . По заданным постоянным коэффициентам уравнения (1.122) построим многочлен

$$K(z) = 1 + a_1 z + a_2 z^2.$$

Этот многочлен можно считать производящей функцией последовательности $\{1, a_1, a_2, 0, \dots\}$. Рассмотрим произведение производящих функций $C(z) = K(z)F(z)$. Согласно (1.114), коэффициент c_{n+2} при z^{n+2} определяется соотношением

$$c_{n+2} = f_0 \cdot 0 + f_1 \cdot 0 + \cdots + f_{n-1} \cdot 0 + f_n a_2 + f_{n+1} a_1 + f_{n+2} = f_n a_2 + f_{n+1} a_1 + f_{n+2}.$$

В силу (1.122) $c_{n+2} = 0$ для $n = 0, 1, \dots$. Таким образом, производящая функция $C(z)$ имеет вид $C(z) = c_0 + c_1 z$. Тогда

$$F(z) = \frac{C(z)}{K(z)} = \frac{c_0 + c_1 z}{1 + a_1 z + a_2 z^2}.$$

Определение 1.65. Характеристическим многочленом линейного рекуррентного уравнения (1.121) называют многочлен степени r

$$G(z) = z^r + a_1 z^{r-1} + a_2 z^{r-2} + \cdots + a_r.$$

Корни этого многочлена называют *характеристическими*. При $r = 2$ многочлен имеет вид

$$G(z) = (z - \alpha_1)(z - \alpha_2).$$

Характеристический многочлен $G(z)$ и многочлен $K(z)$ связаны между собой соотношением

$$K(z) = z^2 G\left(\frac{1}{z}\right),$$

откуда получаем

$$K(z) = (1 - \alpha_1 z)(1 - \alpha_2 z), \text{ а } F(z) = \frac{c_0 + c_1 z}{(1 - \alpha_1 z)(1 - \alpha_2 z)}.$$

Возможны два случая:

1) корни характеристического многочлена $G(z)$ различны ($\alpha_1 \neq \alpha_2$), тогда

$$F(z) = \frac{c_0 + c_1 z}{(1 - \alpha_1 z)(1 - \alpha_2 z)} = \frac{b_0}{1 - \alpha_1 z} + \frac{b_1}{1 - \alpha_2 z};$$

2) характеристический многочлен $G(z)$ имеет корень кратности 2 ($\alpha_1 = \alpha_2$), тогда

$$F(z) = \frac{c_0 + c_1 z}{(1 - \alpha_1 z)^2} = \frac{b_0}{1 - \alpha_1 z} + \frac{b_1}{(1 - \alpha_1 z)^2}.$$

В примере 1.131 было показано, что

$$1 + \alpha z + \alpha^2 z^2 + \alpha^3 z^3 + \dots = \sum_{n=0}^{\infty} (\alpha z)^n = \frac{1}{1 - \alpha z}.$$

Продифференцировав последнее равенство и разделив обе части на α , получаем

$$1 + 2\alpha z + 3\alpha^2 z^2 + 4\alpha^3 z^3 + \dots = \sum_{n=0}^{\infty} (\alpha z)^n (n+1) = \frac{1}{(1 - \alpha z)^2}.$$

Тогда для случая различных корней получаем

$$F(z) = b_0 \sum_{n=0}^{\infty} (\alpha_1 z)^n + b_1 \sum_{n=0}^{\infty} (\alpha_2 z)^n = \sum_{n=0}^{\infty} (b_0 \alpha_1^n + b_1 \alpha_2^n) z^n \text{ и}$$

$$f_n = b_0 \alpha_1^n + b_1 \alpha_2^n, \quad n = 0, 1, \dots$$

Если же корень один, кратности 2, то

$$F(z) = b_0 \sum_{n=0}^{\infty} (\alpha_1 z)^n + b_1 \sum_{n=0}^{\infty} (\alpha_1 z)^n (n+1) = \sum_{n=0}^{\infty} (b_0 \alpha_1^n + b_1 (n+1) \alpha_1^n) z^n \text{ и}$$

$$f_n = (b_0 + b_1 + b_1 n) \alpha_1^n, \quad n = 0, 1, \dots$$

Полученные результаты можно обобщить на случай линейного рекуррентного уравнения порядка r .

1) Если все характеристические корни $\alpha_1, \alpha_2, \dots, \alpha_r$ являются простыми, то общее решение однородного уравнения (1.121) имеет вид

$$f_n = c_1\alpha_1^n + c_2\alpha_2^n + \dots + c_r\alpha_r^n, \quad (1.123)$$

2) Если характеристические корни $\alpha_1, \alpha_2, \dots, \alpha_s$ имеют соответственно кратности e_1, e_2, \dots, e_s , т. е.

$$G(z) = (z - \alpha_1)^{e_1}(z - \alpha_2)^{e_2} \cdots (z - \alpha_s)^{e_s}, \text{ и } e_1 + e_2 + \dots + e_s = r,$$

то общее решение однородного уравнения (1.121) имеет вид

$$f_n = \sum_{i=1}^s p_i(n)\alpha_i^n, \text{ где } p_i(n) = c_{i,0} + c_{i,1}n + \dots + c_{i,e_i-1}n^{e_i-1}$$

— есть многочлен степени $e_i - 1$.

Для определения r неопределенных постоянных используются r граничных условий, а именно значения f_0, f_1, \dots, f_{r-1} .

Пример 1.134 (числа Фибоначчи). Описанным методом найдем общий член последовательности чисел Фибоначчи (см. пример 1.133).

Уравнение $f_{n+2} - f_{n+1} - f_n = 0$ имеет характеристический многочлен

$$z^2 - z - 1 = (z - \alpha_1)(z - \alpha_2), \text{ где } \alpha_1 = \frac{1 + \sqrt{5}}{2}, \alpha_2 = \frac{1 - \sqrt{5}}{2}.$$

Так как α_1 и α_2 являются простыми корнями, то $f_n = c_1\alpha_1^n + c_2\alpha_2^n$, где c_1, c_2 — неопределенные постоянные. Подставив $f_0 = f_1 = 1$, получим систему уравнений

$$c_1 + c_2 = 1, \quad c_1\alpha_1 + c_2\alpha_2 = 1.$$

Решив эти уравнения, находим

$$c_1 = \frac{\alpha_2 - 1}{\alpha_2 - \alpha_1} = \frac{1}{\sqrt{5}}\alpha_1; \quad c_2 = \frac{\alpha_1 - 1}{\alpha_1 - \alpha_2} = -\frac{1}{\sqrt{5}}\alpha_2.$$

Отсюда получаем формулу Бине (1.120).

Пример 1.135 (полиномы Чебышева). Рассмотрим функции $T_n(x)$, определяемые однородным рекуррентным уравнением второго порядка:

$$T_{n+2}(x) = 2xT_{n+1}(x) - T_n(x), \quad T_0(x) = 1, \quad T_1(x) = x. \quad (1.124)$$

Характеристический многочлен для уравнения (1.124) имеет два корня:

$$\alpha_1 = x + \sqrt{x^2 + 1}, \quad \alpha_2 = x - \sqrt{x^2 + 1}.$$

Согласно (1.123), общее решение (1.124) можно представить в виде

$$T_n(x) = c_1 \left(x + \sqrt{x^2 + 1} \right)^n + c_2 \left(x - \sqrt{x^2 + 1} \right)^n. \quad (1.125)$$

Используя начальные условия (1.124), находим, что $c_1 = c_2 = \frac{1}{2}$.

Тогда (1.125) запишем в следующем виде:

$$T_n(x) = \frac{1}{2} \left[\left(x + \sqrt{x^2 + 1} \right)^n + \left(x - \sqrt{x^2 + 1} \right)^n \right]. \quad (1.126)$$

Замечание 1.79. Многочлены $T_n(x)$, определенные в (1.126), называют *ортогональными полиномами Чебышева первого рода*. Первые полиномы Чебышева имеют вид:

$$T_0(x) = 1, T_1(x) = x, T_2(x) = 2x^2 - 1, T_3(x) = 4x^3 - x, T_4(x) = 8x^4 - 8x^2 + 1.$$

Полиномы Чебышева можно задать в аналитическом виде:

$$T_n(x) = \cos(n \arccos(x)).$$

Если в уравнении (1.124) изменить начальные условия: $T_0(x) = 1, T_1(x) = 2x$, то его решения (обозначим их $U_n(x)$) называют *ортогональными полиномами Чебышева второго рода*.

Упражнение 1.21. Выведите рекуррентные формулы для полиномов Чебышева второго рода $U_n(x)$.

Пример 1.136. Вычислим определитель матрицы размером $n \times n$ (обозначим его через f_n)

$$f_n = \begin{vmatrix} 1 & 1 & 0 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 1 & 1 & 0 & \cdots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \cdots & 1 & 1 & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 & 1 & 1 \\ 0 & 0 & \cdots & 0 & 0 & 1 & 1 \end{vmatrix}$$

Разложив определитель по первому столбцу, получим следующее рекуррентное уравнение для f_n :

$$f_{n+2} = f_{n+1} - f_n, \text{ при этом очевидно, что } f_1 = 1, f_2 = 0.$$

Исходя из этого, положим $f_0 = 1$. Данное уравнение имеет характеристический многочлен

$$G(z) = z^2 - z + 1 = (z - \alpha_1)(z - \alpha_2), \text{ где}$$

$$\alpha_1 = \frac{1}{2} + i\frac{\sqrt{3}}{2} = e^{i(\pi/3)}, \quad \alpha_2 = \frac{1}{2} - i\frac{\sqrt{3}}{2} = e^{-i(\pi/3)}.$$

Общее решение имеет вид:

$$f_n = c_1 \alpha_1^n + c_2 \alpha_2^n = c_1 e^{i(\pi/3)n} + c_2 e^{-i(\pi/3)n}.$$

Подставив $f_0 = f_1 = 1$, получим систему уравнений

$$c_1 + c_2 = 1, \quad c_1 \alpha_1 + c_2 \alpha_2 = 1.$$

Отсюда, в свою очередь, имеем

$$c_1 = \frac{1}{2} - i \frac{1}{2\sqrt{3}}, \quad c_2 = \frac{1}{2} + i \frac{1}{2\sqrt{3}}.$$

Следовательно,

$$\begin{aligned} f_n &= \left(\frac{1}{2} - i \frac{1}{2\sqrt{3}} \right) \left(\cos \frac{\pi}{3} n + i \sin \frac{\pi}{3} n \right) + \left(\frac{1}{2} + i \frac{1}{2\sqrt{3}} \right) \times \\ &\quad \times \left(\cos \frac{\pi}{3} n - i \sin \frac{\pi}{3} n \right) = \cos \frac{\pi}{3} n + \frac{1}{\sqrt{3}} \sin \frac{\pi}{3} n. \end{aligned}$$

Пример 1.137. Найдем решение уравнения $f_{n+2} - 4f_{n+1} + 4f_n = 0$ с граничными условиями $f_0 = 1, f_1 = 4$.

Так как характеристический многочлен $z^2 - 4z + 4 = (z - 2)^2$ имеет корень 2 кратности 2, то общее решение $f_n = (c_0 + c_1 n)2^n$. С помощью граничных условий находим

$$c_0 = 1, \quad 2(c_0 + c_1) = 4, \quad \text{т. е. } c_1 = 1.$$

Таким образом, решение рассматриваемого уравнения имеет вид

$$f_n = (1 + n)2^n.$$

Пример 1.138. Найдем решение однородного рекуррентного уравнения

$$2f_{n+2} + 5f_{n+1} - 3f_n = 0$$

с граничными условиями $f_0 = 0, f_1 = 7$.

Характеристический многочлен этого уравнения

$$2z^2 + 5z - 3 = (2z - 1)(z + 3)$$

имеет два корня $\alpha_1 = \frac{1}{2}, \alpha_2 = -3$. Тогда общее решение уравнения имеет вид

$$f_n = b_0 \left(\frac{1}{2} \right)^n + b_1 (-3)^n.$$

Используя граничные условия получаем систему уравнения относительно b_0 и b_1 .

$$b_0 + b_1 = 0, \quad \frac{1}{2}b_0 - 3b_1 = 7.$$

Отсюда $b_0 = 2, b_1 = -2$. Таким образом, решение рассматриваемого уравнения имеет вид

$$f_n = \left(\frac{1}{2}\right)^{n-1} - 2(-3)^n.$$

Пример 1.139. Найдем решение однородного рекуррентного уравнения

$$9f_{n+2} - 6f_{n+1} + f_n = 0$$

с граничными условиями $f_0 = 1, f_1 = 1$.

Характеристический многочлен этого уравнения

$$9z^2 - 6z + 1 = (3z - 1)^2$$

имеет один корень $\alpha = \frac{1}{3}$ кратности 2. Тогда общее решение уравнения имеет вид

$$f_n = (b_0 + b_1 n) \left(\frac{1}{3}\right)^n.$$

Используя граничные условия получаем систему уравнения относительно b_0 и b_1 .

$$b_0 = 1, \quad \frac{1}{3}(b_0 + b_1) = 1.$$

Отсюда $b_0 = 1, b_1 = 2$. Таким образом, решение рассматриваемого уравнения имеет вид

$$f_n = (1 + 2n) \left(\frac{1}{3}\right)^n.$$

Упражнение 1.22. Найти решение однородного рекуррентного уравнения с граничными условиями:

- 1) $3f_{n+2} - 8f_{n+1} + 4f_n = 0, \quad f_0 = 2, f_1 = 2,$
- 2) $4f_{n+2} - 8f_{n+1} + 3f_n = 0, \quad f_0 = 0, f_1 = -1,$
- 3) $9f_{n+2} - 12f_{n+1} + 4f_n = 0, \quad f_0 = 1, f_1 = -2,$
- 4) $9f_{n+2} + 6f_{n+1} + f_n = 0, \quad f_0 = -1, f_1 = -1.$

Ответы:

- 1) $f_n = \frac{1}{2} \left(\frac{2}{3}\right)^n + \frac{3}{2} 2^n,$

$$2) f_n = -\left(\frac{3}{2}\right)^n + \left(\frac{1}{2}\right)^n,$$

$$3) f_n = (1 - 4n)\left(\frac{2}{3}\right)^n,$$

$$4) f_n = (-1 + 4n)\left(-\frac{1}{3}\right)^n.$$

1.7.3. Решение неоднородного линейного рекуррентного уравнения

Справедлива следующая теорема

Теорема 1.85. Пусть f_n — общее решение однородного уравнения, соответствующего уравнению (1.112). Тогда общее решение неоднородного уравнения g_n имеет вид:

$$g_n = f_n + h_n,$$

где h_n — любое частное решение уравнения (1.112).

Для произвольного уравнения (1.112) найти его частное решение h_n достаточно трудно, однако, в частном случае, когда правая часть $\varphi(n)$ имеет специальный вид:

$$\varphi(n) = c_0 + c_1 n + c_2 n^2 + \cdots + c_m n^m$$

метод нахождения дает следующая теорема.

Теорема 1.86. Пусть 1 — корень кратности s ($s \geq 0$) характеристического уравнения для (1.112). Тогда h_n имеет вид

$$h_n = b_0 n^s + b_1 n^{s+1} + \cdots + b_m n^{s+m}$$

где b_i можно найти методом неопределенных коэффициентов.

Пример 1.140. Рассмотрим линейное неоднородное уравнение первого порядка (1.111)

$$f_{n+1} - f_n = (n + 1)^2.$$

Его характеристическое уравнение имеет один корень 1. Согласно теореме 1.86, частное решение ищем в виде

$$h_n = b_0 n + b_1 n^2 + b_2 n^3.$$

Подставим h_n в уравнение, тогда

$$b_0(n + 1) + b_1(n + 1)^2 + b_2(n + 1)^3 - b_0 n - b_1 n^2 - b_2 n^3 = (n + 1)^2$$

Приведем подобные члены, тогда

$$b_0 + b_1 + b_2 + (2b_1 + 3b_2)n + 3b_2n^2 = (n+1)^2.$$

Приравняв в последнем уравнении коэффициенты при одинаковых степенях n , находим

$$b_0 = \frac{1}{6}, \quad b_1 = \frac{1}{2}, \quad b_2 = \frac{1}{3}, \quad h_n = \frac{n}{6}(1 + 3n + 2n^2) = \frac{n(n+1)(2n+1)}{6}.$$

Пример 1.141. Рассмотрим уравнение

$$f_{n+2} - 5f_{n+1} + 6f_n = n^2 + 2n - 1.$$

Его характеристическое уравнение имеет два корня 2 и 3. Согласно теореме 1.86, частное решение ищем в виде

$$h_n = b_0 + b_1n + b_2n^2.$$

Подставим h_n в уравнение, тогда

$$\begin{aligned} & b_0 + b_1(n+2) + b_2(n+2)^2 - 5(b_0 + b_1(n+1) + b_2(n+1)^2) + \\ & + 6(b_0 + b_1n + b_2n^2) = 2b_0 - 3b_1 - b_2 + (2b_1 - 6b_2)n + 2b_2n^2 = n^2 + 2n - 1. \end{aligned}$$

Приравняв в этом уравнении коэффициенты при одинаковых степенях n , находим

$$b_0 = \frac{7}{2}, \quad b_1 = \frac{5}{2}, \quad b_2 = \frac{1}{2}, \quad h_n = \frac{7 + 5n + n^2}{2}.$$

Пример 1.142. Найдем общее решение неоднородного рекуррентного уравнения

$$f_{n+2} - 3f_{n+1} + 2f_n = n^2 + 1.$$

Характеристический многочлен этого уравнения

$$z^2 - 3z + 2 = (z-1)(z-2)$$

имеет два корня $\alpha_1 = 1, \alpha_2 = 2$. Тогда общее решение однородного уравнения имеет вид

$$f_n = b_0 + b_12^n.$$

Частное решение неоднородного уравнения будем искать в виде:

$$h_n = a_0n + a_1n^2 + a_2n^3.$$

Подставляя в исходное уравнение имеем

$$\begin{aligned} & a_0(n+2) + a_1(n+2)^2 + a_2(n+2)^3 - 3(a_0(n+1) + a_1(n+1)^2 + a_2(n+1)^3) + \\ & + 2(a_0n + a_1n^2 + a_2n^3) = n^2 + 1. \end{aligned}$$

Приводя подобные члены, получаем

$$-a_0 + a_1 + 5a_2 + (-2a_1 + 3a_2)n + (3a_1 - 3a_2)n^2 = n^2 + 1.$$

Приравнивая коэффициенты при одинаковых степенях многочленов, имеем систему уравнений относительно a_0, a_1, a_2 :

$$-a_0 + a_1 + 5a_2 = 1, \quad -2a_1 + 3a_2 = 0, \quad 3a_1 - 3a_2 = 1.$$

Отсюда

$$a_0 = -\frac{8}{15}, \quad a_1 = -\frac{1}{5}, \quad a_2 = -\frac{2}{15}.$$

Тогда общее решение неоднородного уравнения имеет вид

$$g_n = f_n + h_n = b_0 + b_1 2^n - \frac{8n}{15} - \frac{n^2}{5} - \frac{2n^3}{15}.$$

Упражнение 1.23. Найти общее решение неоднородного рекуррентного уравнения:

- 1) $f_{n+2} - 4f_{n+1} + 4f_n = 2n^2 - 1,$
- 2) $f_{n+2} - f_{n+1} - 2f_n = n^2 + 3n - 3,$
- 3) $f_{n+2} - 3f_{n+1} + 2f_n = 2n + 3,$
- 4) $f_{n+2} - f_n = 3n - 2.$

Ответы:

- 1) $g_n = (b_0 + b_1 n) 2^n + 15 + 8n + 2n^2,$
- 2) $g_n = b_0 2^n + b_1 (-1)^n - \frac{1}{4} - 2n - \frac{n^2}{2},$
- 3) $g_n = b_0 + b_1 2^n - 4n - n^2,$
- 4) $g_n = b_0 + b_1 (-1)^n - \frac{5n}{2} + \frac{3n^2}{4},$

1.7.4. Производящая функция для чисел Каталана

Определим бесконечный степенной ряд, взяв в качестве его коэффициентов числа Каталана (см. определение 1.56).

Для нахождения производящей функции для чисел Каталана используем рекуррентные соотношения. Найдем рекуррентную формулу для чисел Каталана. Рассмотрим для этого модель с расстановкой скобок в произведении матриц: $A_0 A_1 \cdots A_n$ (см. задачу 1.22). При любой расстановке скобок найдем последнюю операцию, определяемую это парой, и разобьем произведение на две части $(A_0 A_1 \cdots A_k)(A_{k+1} A_{k+2} \cdots A_n)$, соответственно —

первый и второй множители. Тогда задача сводится к аналогичной задаче меньшей «размерности» — расставить скобки в каждом из множителей, а затем поставить завершающую — внешнюю — пару скобок. Поскольку в первом множителе скобки можно расставить k способами, а во втором $(n - k)$ способами, то, комбинируя различные расстановки в них, получим по правилу умножения $k(n - k)$ способов.

Далее, так как граница разбиения на два множителя может оказаться в любом месте между первым и $(n + 1)$ -м множителем, получаем n различных вариантов, которые надо просуммировать. Итак, получаем формулу:

$$u_n = \sum_{k=0}^n u_k u_{n-k}, \quad \text{где } u_0 = 0, u_1 = 1.$$

Определим производящую функцию для чисел Каталана — согласно определению 1.113, бесконечный степенной ряд, с коэффициентами — числами Каталана:

$$u(z) = \sum_{k=0}^{\infty} u_k z^k = u_0 + u_1 z + \cdots + u_n z^n + \cdots = z + z^2 + 2z^3 + \cdots + u_n z^n + \cdots$$

Найдем квадрат функции $u(x)$:

$$\begin{aligned} u(z)^2 &= (u_0 + u_1 z + \cdots + u_n z^n + \cdots)(u_0 + u_1 z + \cdots + u_n z^n + \cdots) = \\ &= u_0^2 + (u_0 u_1 + u_1 u_0)z + (u_0 u_2 + u_1 u_1 + u_2 u_0)z^2 + \cdots + \\ &+ \sum_{k=0}^n u_k u_{n-k} z^n + \cdots = u(z) - z. \end{aligned}$$

Решив полученное уравнение как квадратное относительно $u(z)$, получим

$$u(z)^2 - u(z) + z = 0, \text{ откуда } u(z)_{1,2} = \frac{1 \pm \sqrt{1 - 4z}}{2}. \quad (1.127)$$

Так как $u(0) = 0$, в полученной формуле (1.127) надо выбрать знак «минус».

Разложим в степенной ряд $\sqrt{1 - 4z}$:

$$\begin{aligned}\sqrt{1 - 4z} &= 1 + \sum_{n=1}^{\infty} (-4)^n \frac{\frac{1}{2} \left(\frac{1}{2} - 1 \right) \left(\frac{1}{2} - 2 \right) \cdots \left(\frac{1}{2} - n + 1 \right)}{n!} z^n = \\ &= 1 + \sum_{n=1}^{\infty} (-1)^n 4^n \frac{\frac{1}{2^n} \cdot 1 \cdot (1-2)(1-4) \cdots (1-2n+2)}{n!} z^n = \\ &= 1 + \sum_{n=1}^{\infty} (-1)^n 2^n \frac{(-1)^{n-1} \cdot 1 \cdot 1 \cdot 3 \cdots (2n-3)}{n!} z^n = \\ &= 1 - \sum_{n=1}^{\infty} 2^n \frac{1 \cdot 2 \cdot 3 \cdot 4 \cdots (2n-3)(2n-2)}{n! \cdot 2 \cdot 4 \cdots (2n-2)} z^n = \\ &= 1 - \sum_{n=1}^{\infty} 2^n \frac{(2n-2)!}{2^{n-1}(n-1)!n!} z^n = 1 - 2 \sum_{n=1}^{\infty} \frac{(2n-2)!}{(n-1)!n!} z^n.\end{aligned}$$

Подставив полученный ряд в выражение (1.127) для $u(z)$, получим

$$u_n = \frac{(2n-2)!}{(n-1)!n!} = \frac{1}{n} \frac{(2n-2)!}{(n-1)!(n-1)!} = \frac{1}{n} C_{2n-2}^{n-1}.$$

1.7.5. Число многочленов заданной степени, не приводимых над полем вычетов

Использование производящих функций позволяет решать и более сложные комбинаторные задачи. Одной из них является нахождение числа неприводимых многочленов над полем вычетов по модулю p .

В дальнейшем будем рассматривать только нормализованные многочлены, первый коэффициент которых равен 1, что не умаляет общности рассмотрения. Всего многочленов степени n над полем \mathbb{Z}_p имеется p^n , так как многочлен n -й степени имеет $n+1$ коэффициент, первый из которых равен 1, а остальные n могут принимать по p значений каждый. Тогда производящая функция для чисел коэффициентов задается следующим рядом:

$$1 + px + p^2x^2 + \cdots + = \sum_{n=0}^{\infty} p^n x^n = \frac{1}{1 - px} \quad (1.128)$$

по формуле суммы бесконечно убывающей прогрессии.

Вычислим теперь эту функцию другим способом. Обозначим N_d — число различных неприводимых многочленов степени d . Любой многочлен над

полем \mathbb{Z}_p может быть единственным образом записан в виде произведения степеней неприводимых многочленов:

$$f(x) = (f_1(x))^{i_1} (f_2(x))^{i_2} \dots$$

Если обозначить через d_k степени многочленов $f_k(x)$, а степень многочлена $f(x)$ через n , то получим равенство $n = i_1 d_1 + i_2 d_2 + \dots$

Заметим, что такой коэффициент у x_n можно получить при перемножении бесконечного числа следующих рядов:

$$(1 + x^{d_1} + x^{2d_1}x^{3d_1} + \dots) \cdot (1 + x^{d_2} + x^{2d_2}x^{3d_2} + \dots) \dots = \frac{1}{1 - x^{d_1}} \cdot \frac{1}{1 - x^{d_2}} \dots$$

Соберем в этом бесконечном произведении все одинаковые множители: для d_i число таких множителей обозначено ранее как N_{d_i} . Получим другое выражение для производящей функции чисел всех многочленов над полем вычетом по модулю p .

$$\prod_{d=1}^{\infty} \left(\frac{1}{1 - x^d} \right)^{N_d}. \quad (1.129)$$

Приравняем правую часть (1.128) и (1.129) и прологарифмируем полученное равенство:

$$\ln \left(\frac{1}{1 - px} \right) = \sum_{d=1}^{\infty} N_d \ln \left(\frac{1}{1 - x^d} \right).$$

Используя формулу для разложения натурального логарифма в степенной ряд

$$\ln(1 + x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \dots \text{ или } \ln \left(\frac{1}{1 - x} \right) = x + \frac{1}{2}x^2 + \frac{1}{3}x^3 + \dots,$$

получаем равенство

$$\sum_{n=1}^{\infty} \frac{(px)^n}{n} = \sum_{d=1}^{\infty} N_d \sum_{j=1}^{\infty} \frac{x^{jd}}{j}.$$

Приравняв в последнем равенстве коэффициенты при x_n , имеем

$$\frac{p^n}{n} = \sum_{d|n} N_d \frac{d}{n}, \quad (1.130)$$

так как в (1.130) справа степени x_n получаются, если j удовлетворяет условию $jd = n$, или $j = n/d$, это означает, что для всех делителей d числа n в сумме справа найдутся слагаемые n -й степени.

Преобразовав выражение (1.130), получим

$$p^n = \sum_{d|n} dN_d. \quad (1.131)$$

К равенству (1.131) можно применить формулу обращения Мебиуса (см. теорему 1.81), если заметить, что его можно представить в виде

$$g(n) = \sum_{d|n} f(d), \text{ где } g(n) = p^n, \quad f(d) = dN_d.$$

Тогда

$$f(n) = nN_n = \sum_{d|n} \mu(d)g\left(\frac{n}{d}\right) = \sum_{d|n} \mu(d)p^{\frac{n}{d}}, \quad \text{или} \quad N_n = \frac{1}{n} \sum_{d|n} \mu(d)p^{\frac{n}{d}}. \quad (1.132)$$

Пример 1.143. Найдем число неприводимых многочленов шестой степени над полем вычетом по модулю 2. По формуле (1.132) имеем

$$\begin{aligned} N_6 &= \frac{1}{6} \sum_{d|6} \mu(d)p^{\frac{6}{d}} = \frac{1}{6} (\mu(1)2^6 + \mu(2)2^3 + \mu(3)2^2 + \mu(6)2^1) = \\ &= \frac{1}{6} (2^6 - 2^3 - 2^2 + 2^1) = \frac{54}{6} = 9. \end{aligned}$$

Пример 1.144. Найдем все многочлены из примера 1.143.

Заметим, что неприводимый многочлен обязательно имеет ненулевой свободный член (иначе бы выносился множитель x) и нечетное число слагаемых (иначе значение многочлена при $x = 1$ было бы равно 0 и по теореме Безу (см. теорему 1.46) выносился множитель $x+1$). Таким образом, из всех $2^6 = 64$ многочленов остается перебрать лишь $C_5^1 + C_5^3 + C_5^5 = 5 + 10 + 1 = 16$.

Ниже приведены результаты вычислений с помощью среды Maple. Многочлены, стоящие на 1, 3, 5, 7, 8, 9, 11, 13, 14-м местах неприводимые.

```
> Factors(x^6+x^5+1) mod 2;
```

[1, [[$x^6 + x^5 + 1$, 1]]]

```
> Factors(x^6+x^4+1) mod 2
```

[1, [[$x^3 + x^2 + 1$, 2]]]

```
> Factors(x^6+x^3+1) mod 2
```

[1, [[$x^6 + x^3 + 1$, 1]]]

```

> Factors(x^6+x^2+1) mod 2
[1, [[x^3 + x + 1, 2]]]

> Factors(x^6+x+1) mod 2
[1, [[x^6 + x + 1, 1]]]

> Factors(x^6+x^5+x^4+x^3+1) mod 2
[1, [[x^4 + x + 1, 1], [x^2 + x + 1, 1]]]

> Factors(x^6+x^5+x^4+x^2+1) mod 2
[1, [[x^6 + x^5 + x^4 + x^2 + 1, 1]]]

> Factors(x^6+x^5+x^4+x+1) mod 2
[1, [[x^6 + x^5 + x^4 + x + 1, 1]]]

> Factors(x^6+x^5+x^3+x^2+1) mod 2
[1, [[x^6 + x^5 + x^3 + x^2 + 1, 1]]]

> Factors(x^6+x^5+x^3+x+2) mod 2
[1, [[x^2 + x + 1, 3]]]

> Factors(x^6+x^5+x^2+x+1) mod 2
[1, [[x^6 + x^5 + x^2 + x + 1, 1]]]

> Factors(x^6+x^4+x^3+x^2+1) mod 2
[1, [[x^4 + x^3 + x^2 + x + 1, 1], [x^2 + x + 1, 1]]]

> Factors(x^6+x^4+x^3+x+1) mod 2
[1, [[x^6 + x^4 + x^3 + x + 1, 1]]]

> Factors(x^6+x^4+x^2+x+1) mod 2
[1, [[x^6 + x^4 + x^2 + x + 1, 1]]]

> Factors(x^6+x^3+x^2+x+1) mod 2
[1, [[x^4 + x^3 + x + 1, 1], [x^2 + x + 1, 1]]]

> Factors(x^6+x^5+x^4+x^3+x^2+x+1) mod 2
[1, [[x^3 + x^2 + 1, 1], [x^3 + x + 1, 1]]]

```

1.8. Элементы дискретной теории вероятностей

Данный раздел является естественным продолжением предыдущего раздела по комбинаторике и частично написан по материалам [41].

1.8.1. Основные определения

Будем рассматривать конечное множество элементарных событий S , все его элементы считаются равноправными, равновероятными. Любое событие A считается подмножеством S и состоит из элементарных событий.

Введем некоторые естественные требования, которым должно будет удовлетворять определение вероятности.

1. При сопоставлении событий вероятность события A , содержащегося в другом событии B , должна иметь вероятность не превосходящую вероятности B (события это множества и они могут содержаться одно в другом).

2. Если событие A составляется из событий, которые не могут выполняться одновременно (такие события называются *несовместными*), то вероятность осуществления события A должна равняться сумме вероятностей составляющих событий. (Иначе говоря, если задано разбиение множества исходов A на подмножества A_i , то вероятность события A равна сумме вероятностей событий A_i).

Определение 1.66. Отношение $|A|/|S|$ будем называть *вероятностью* события A и обозначать через $P(A)$.

Событие, вероятность которого равна 1 называется *достоверным* – оно наступает всегда.

Событие, вероятность которого равна 0 называется *невозможным* – оно не наступает никогда.

Замечание 1.80. Понятие вероятности, введенное нами как *отношение числа входящих в событие (благоприятных) равновозможных исходов к числу всех исходов*, удовлетворяет всем ранее сформулированным условиям.

Пример 1.145. При бросании двух игральных костей вероятность получить не менее 11 очков равна $\frac{3}{36}$, так как всего имеется 36 равновозможных исходов и 3 из них благоприятны – $(5,6),(6,5),(6,6)$.

Было бы неверным считать, что так как может выпасть от 2 до 12 очков и благоприятными являются 2 исхода из этих 11, то искомая вероятность равна $\frac{2}{11}$. Исходы типа „числа выпавших очков“ не являются

равновозможными. Например выпадению 12 очков соответствует всего одна комбинация, а выпадению 7 очков – целых 6 комбинаций.

Упражнение 1.24. Найдите вероятность того, что на первой кости выпадет очков больше, чем на второй. Ответ: $\frac{15}{36}$.

Определение 1.67. Событие, которое составлено из элементарных событий, входящих одновременно в события A и B , называется *совмещением этих событий*. Таким образом множество элементарных событий, входящих в совмещение, – это пересечение соответствующих множеств. Очевидно, что вероятность совмещения событий не превосходит вероятности каждого из этих событий.

Оба этих определения легко переносятся на несколько событий.

Определение 1.68. Разбиение множества элементарных событий S на несовместные события S_1, S_2, \dots, S_m называется *полной системой событий*. Каждое событие A можно представить, как объединение несовместных событий – совмещение A с элементами разбиения S_i .

Очевидно, что

$$P(A) = \sum_{i=1}^m P(A \cap S_i). \quad (1.133)$$

Формула (1.133) называется *формулой полной вероятности*.

Введем понятие независимых событий.

Определение 1.69. События A и B называются *независимыми*, если вероятность их совмещения равна произведению их вероятностей:

$$P(A \cap B) = P(A) \times P(B).$$

В случае нескольких событий A_1, A_2, \dots, A_k следует различать попарную независимость этих событий и их *независимость в совокупности*. В первом случае

$$P(A_i \cap A_j) = P(A_i) \times P(A_j) \text{ при } i \neq j,$$

во втором случае (более жесткое требование)

$$P(A_1 \cap A_2 \cap \dots \cap A_k) = P(A_1) \times P(A_2) \times \dots \times P(A_k).$$

Легко показать, что независимость событий в совокупности не следует из их попарной независимости.

Пример 1.146. Рассмотрим игральную кость в виде правильного тетраэдра¹. Одна грань этого тетраэдра выкрашена в белый цвет W , вторая в черный B , третья – в красный R , а окраска четвертой грани – смешанная M , в ней есть все три цвета. При каждом бросании кость равновероятно ложится какой-то стороной вниз. Вероятность того, что на этой грани окажется белый, красный или черный цвет равна, очевидно, $1/2$, так как для этого события благоприятны два исхода из четырех возможных. Вероятность выпадения двух цветов сразу, например, красного и черного равна $1/4$ и не зависит от набора цветов. Таким образом для событий выпал красный цвет и выпал черный цвет имеем:

$$P(R \cap B) = P(R) \times P(B),$$

и по определению эти два события независимы. С другой стороны вероятность выпадения трех цветов

$$P(R \cap B \cap W) = \frac{1}{4} \neq P(R) \times P(B) \times P(W) = \frac{1}{8},$$

и независимости в совокупности нет.

1.8.2. Условные вероятности и формула Байеса

При работе с зависимыми случайными событиями необходимо учитывать их взаимное влияние на вероятностное поведение. Это влияние проявляется в том, что осуществление одного события, может изменить вероятности других событий и приходится говорить о вероятностях событий в тех или иных условиях. Введем следующее определение.

Определение 1.70. Пусть A – некоторое событие, имеющее положительную вероятность (то есть не невозможное событие). Определим для события B *условную вероятность B при условии A* как отношение

$$P(B|A) = \frac{P(B \cap A)}{P(A)}. \quad (1.134)$$

Замечание 1.81. Очевидно, что таким образом определенные условные вероятности обладают всеми свойствами обычных вероятностей: они меняются от 0 до 1, возрастают с расширением события, условная вероятность объединения несовместных событий равна сумме их условных вероятностей.

Используя понятие условной вероятности можно по другому определить понятие независимости событий (определение 1.69): события A и B независимы, если

$$P(B|A) = P(B). \quad (1.135)$$

¹Пример предложен С.Н.Бернштейном

Пример 1.147. Необходимо определить, повлиял ли просмотр рекламы на сбыт продукции. Для этого был проведен опрос, результаты которого приведены в таблице 1.51. Событие A — респондент сказал, что купит продукт, событие C — что не купит, событие B — респондент смотрел рекламу, D — не смотрел.

Таблица 1.51

	A	C
B	459	441
D	51	49

Критерием независимости события является выполнение условия (1.135). Нужно проверить, будет ли вероятность покупки $p(A)$ совпадать с вероятностью покупки при условии, что респондент смотрел рекламу, $p(A|B)$. Сосчитаем обе вероятности, исходя из определения вероятности (определение 1.66):

$$p(A) = \frac{459 + 51}{459 + 51 + 441 + 49} = 0.51; \quad p(A|B) = \frac{459}{459 + 441} = 0.51.$$

Таким образом, определили, что события A — покупка товара и B — просмотр рекламы независимы.

Пример 1.148. Из 9 вопросов экзамена студент выучил 6. Какова вероятность сдать экзамен, если для этого необходимо ответить на два случайно выбранных вопроса.

Определим события A_1 — студент ответил на первый вопрос и A_2 — студент ответил на второй вопрос, тогда $A = A_1 \cap A_2$ и

$$p(A) = p(A_1 \cap A_2) = p(A_1)p(A_2|A_1) = \frac{6}{9} \cdot \frac{5}{8} = \frac{5}{12}.$$

В терминах условной вероятности можно по другому записать формулу полной вероятности (1.133):

$$P(A) = \sum_{i=1}^m P(A|S_i)P(S_i). \quad (1.136)$$

Пример 1.149. В урне было 10 белых и 10 черных шаров. С вероятностью $1/4$ из урны удалили 5 черных шаров. Какова вероятность того, что вынутый наугад из этой урны шар окажется белым.

Мы не знаем, удалены ли шары из урны или нет и должны делать гипотезы и рассматривать условные вероятности интересующего нас события. При условии A_0 (шары не были удалены) вероятность вынуть белый

шар равна $P(B|A_0) = 10/20 = 1/2$. При условии A_1 (шары были удалены) вероятность равна $P(B|A_1) = 10/15 = 2/3$. Так как по условию задачи $P(A_1) = 1/4$ и, следовательно $P(A_0) = 3/4$, то согласно (1.136)

$$P(B) = P(B|A_0)P(A_0) + P(B|A_1)P(A_1) = \frac{1}{2} \cdot \frac{3}{4} + \frac{2}{3} \cdot \frac{1}{4} = \frac{3}{8} + \frac{1}{6} = \frac{13}{24}.$$

Пример 1.150. В магазин привезли 50 смартфонов, изготовленных тремя фирмами в количестве 15, 25 и 10 штук. Качество сборки первой фирмы — 0.8, второй и третьей соответственно 0.65 и 0.9. Найти вероятность того, что купленный смартфон окажется отличного качества.

Введем гипотезы: H_1 — смартфон изготовлен первой фирмой, H_2 — второй и H_3 — третьей. Согласно определению 1.66 вероятности гипотез:

$$P(H_1) = \frac{15}{50} = 0.3, \quad P(H_2) = \frac{25}{50} = 0.5, \quad P(H_3) = \frac{10}{50} = 0.2.$$

Введем событие A — купленный смартфон отличного качества. Согласно условиям:

$$P(A|H_1) = 0.8, \quad P(A|H_2) = 0.65, \quad P(A|H_3) = 0.9.$$

По формуле полной вероятности (1.136) имеем:

$$\begin{aligned} P(A) &= P(A|H_1)P(H_1) + P(A|H_2)P(H_2) + P(A|H_3)P(H_3) = \\ &= 0.8 \times 0.3 + 0.65 \times 0.5 + 0.9 \times 0.2 = 0.745 \end{aligned}$$

Из формулы полной вероятности (1.136) легко получить одну очень важную формулу, известную как *формула Байеса*¹. Выразим совмещение событий $P(B \cap A_i)$ с помощью условных вероятностей. Согласно (1.134) имеем

$$P(A_i|B)P(B) = P(B|A_i)P(A_i).$$

Разделив обе части формулы на $P(B)$ и подставив для $P(B)$ ее выражение из формулы полной вероятности (1.136) получим формулу Байеса

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^m P(B|A_j)P(A_j)}. \quad (1.137)$$

Эта формула имеет следующее истолкование и применение. Предположим, что случайное событие B зависит от некоторого случайного параметра, который определяется заданием полной системой событий $\{A_i\}$.

¹Thomas Bayes (1702-1761). Формула Байеса стала известной уже после его смерти. При жизни он ничего не публиковал, но тем не менее в 1742 был избран членом Лондонского Королевского общества.

Известны некоторые (*априорные* – заданные до опыта) вероятности событий A_i . Нас интересует, какое из событий осуществилось. Мы проводим эксперимент и узнаем, что осуществилось событие B . Информация об этом дает возможность пересчитать вероятности A_i – получить *апостериорные* вероятности.

Пример 1.151. Зная вероятности появления различных болезней и вероятности появления отдельных симптомов при той или иной болезни врач может делать различные анализы, которые постепенно изменяют вероятности наличия этих болезней у пациента. Пусть при заболевании гриппом с вероятностью 0.6 можно заболеть гриппом $A(GA)$, а с вероятностью 0.4 – гриппом $B(GB)$. При этом реакция иммунофлуоресценции (RI) положительна с вероятностью 0.85 в случае гриппа A и с вероятностью 0.35 в случае гриппа B . Требуется оценить вероятность того, что больной гриппом болен гриппом A , если у него реакция положительная.

Подставляя числа в формулу Байеса (1.137) получаем

$$P(A|RI^+) = \frac{0.85 \times 0.6}{0.85 \times 0.6 + 0.35 \times 0.4} = \frac{0.51}{0.51 + 0.14} \approx 0.7846.$$

и диагноз склоняется в сторону гриппа A .

Пример 1.152. Имеются две колоды карт, в 52 (обозначим ее K_{52}) и в 36 (обозначим ее K_{36}) листов. Выбирается одна из колод (с равной вероятностью) и из нее вытаскивается случайно одна карта. Меняет ли информация об этой карте наши представления о выбранной колоде?

Да! Если выбрана карта, которой нет в K_{36} , то выбрана K_{52} . Это очевидно.

А если выбрана карта, которая есть в K_{36} . Вероятности по-прежнему равны, или что-то изменилось? Давайте обсудим это.

Представьте себе, что мы много раз проводите один и тот же эксперимент — вытаскивали карту из колоды (а потом возвращали). Мы сделали это 1000 раз, и каждый раз получали карту из K_{36} . Повлиял ли этот длинный эксперимент на наши предположения? Скорее всего, мы ответим да. Не может быть (нет, маловероятно), чтобы при тысяче экспериментов ни разу не появилась карта из продолжения колоды.

А если мы сделаем только 100 попыток? Или 50? Неужели есть такой порог, при котором наши представления изменяются? Конечно же, нет. Просто при каждой попытке вероятность изменяется очень мало. Давайте посчитаем. Формула Байеса и дает нам такую возможность.

Введем следующие события:

- B_{36} — выбрана колода K_{36} ,
- B_{52} — выбрана колода K_{52} ,

- A — из колоды вынута карта 8 ♣.

Согласно (1.137) имеем:

$$P(B_{36}|A) = \frac{P(A|B_{36}) \times P(B_{36})}{P(A|B_{36}) \times P(B_{36}) + P(A|B_{52}) \times P(B_{52})} \quad (1.138)$$

Легко показать, что

$$P(B_{36}) = 1/2, \quad P(B_{52}) = 1/2, \quad P(A|B_{36}) = 1/36, \quad P(A|B_{52}) = 1/52.$$

Подставляя в (1.138), получаем

$$P(B_{36}|A) = \frac{1/2 \times 1/36}{1/2 \times 1/36 + 1/2 \times 1/52} = \frac{52}{88} = \frac{13}{22}.$$

Пример 1.153. Экзамен по дискретной математике принимают 3 злых и 9 добрых гномов. Злые гномы поставили 40% двоек и 60% пятерок, добрые – 10% двоек и 90% пятерок. Какова вероятность того, что студент:

- 1) попал на экзамен к доброму гному;
- 2) сдал экзамен;
- 3) сдавший экзамен, сдавал его доброму гному?

Введем следующие гипотезы: H_1 — студент сдавал экзамен злому гному, H_2 — студент сдавал экзамен доброму. Согласно классическому определению вероятности (определение 1.66),

$$p(H_1) = \frac{3}{12} = \frac{1}{4}; \quad p(H_2) = \frac{9}{12} = \frac{3}{4}.$$

Вероятность события A — студент сдал экзамен — зависит от того, кому студент сдавал, т. е. от гипотез H_1 и H_2 . По формуле полной вероятности (1.136),

$$p(A) = p(H_1)p(A|H_1) + p(H_2)p(A|H_2) = \frac{1}{4} \cdot 0.6 + \frac{3}{4} \cdot 0.9 = \frac{33}{40} = 0.825.$$

Вероятность того, что студент сдавал экзамен доброму гному, если известно, что экзамен сдан (вероятность гипотезы после испытания), вычисляется по формуле Байеса (1.137):

$$p(H_2|A) = \frac{p(A|H_2)p(H_2)}{p(A)} = \frac{0.9 \cdot 0.75}{0.825} = 0.8182.$$

Далее будет рассмотрен метод для фильтрации спама, основанный на применении наивного байесовского классификатора, в основе которого лежит использование формулы Байеса (1.137).

1.8.3. Дискретные случайные величины. Математическое ожидание и дисперсия

Определение 1.71. Пусть $\Omega = \{\omega\}$ – множество элементарных событий. Заданная на Ω функция ξ

$$\xi : \Omega \rightarrow X = (x_1, \dots, x_n \mid x_i \in R, 1 \leq i \leq n)$$

называется *дискретной случайной величиной*. Обозначается, как $\xi = \xi(\omega)$.

Пусть вероятности значений x_1, \dots, x_n случайной величины ξ равны p_1, p_2, \dots, p_n , то есть $P(\xi = x_i) = p_i$. Заметим, что события

$$P(\xi = x_1) = p_1, \dots, P(\xi = x_n) = p_n$$

образуют полную систему событий, поэтому

$$p_1 + \dots + p_n = 1.$$

Такое соответствие, связывающее значения случайной величины с соответствующими им вероятностями, называют *законом распределения дискретной случайной величины*. Часто ее задают табличным способом (таблица 1.52):

Таблица 1.52

ξ	x_1	\dots	x_n
p	p_1	\dots	p_n

Определение 1.72. Случайные величины ξ и η называются *независимыми*, если для любых $a, b \in R$ независимы события $\xi = a$ и $\eta = b$, то есть если

$$P(\xi = a, \eta = b) = P(\xi = a) \times P(\eta = b).$$

Определение 1.73. *Математическим ожиданием* случайной величины ξ называется величина:

$$\mathbb{E}\xi = x_1 p_1 + x_2 p_2 + \dots + x_n p_n = \sum_{k=1}^n x_k p_k. \quad (1.139)$$

Замечание 1.82. Математическое ожидание похоже на центр тяжести системы материальных точек – если считать вероятности значений случайной величины ξ массами точек, то получается в точности формула (1.139).

Теорема 1.87. Справедливы следующие свойства математического ожидания:

- 1) Если $P(\xi = a) = 1$, то $E\xi = a$.
- 2) Если ξ и η – две случайные величины, то $E(\xi + \eta) = E\xi + E\eta$.
- 3) Если $\eta = b\xi$, где $b \in R$, то $E\eta = bE\xi$.

Доказательство. Все три свойства очевидным образом следуют из определения. ■

Пример 1.154. Найти математическое ожидание дискретной случайной величины ξ , зная закон ее распределения (таблица 1.53)

Таблица 1.53

ξ	5	4	3
p	0.2	0.5	0.3

Согласно (1.139) имеем:

$$E\xi = 5 \times 0.2 + 4 \times 0.5 + 3 \times 0.3 = 3.3.$$

На практике часто приходится оценить рассеяние возможных значений случайной величины вокруг ее среднего значения.

Определение 1.74. Математическое ожидание квадрата отклонения случайной величины от ее математического ожидания называется *дисперсией* этой случайной величины:

$$D\xi = E(\xi - E\xi)^2. \quad (1.140)$$

Замечание 1.83. Дисперсия характеризует разброс случайной величины вокруг ее математического ожидания. Продолжая аналогии с механикой можно заключить, что дисперсия – это момент инерции системы материальных точек.

Если в (1.140) раскрыть скобки, то получится еще одно выражение для дисперсии:

$$D\xi = E\xi^2 - (E\xi)^2.$$

Теорема 1.88. Справедливы следующие свойства дисперсии:

- 1) Если $P(\xi = a) = 1$, то $D\xi = 0$.
- 2) Если $\eta = \xi + a$, где $a \in R$, то $D\xi = D\eta$.
- 3) Если $\eta = b\xi$, где $b \in R$, то $D\eta = b^2 D\xi$.
- 4) Если η и ξ – две независимые случайные величины, то

$$D(\xi + \eta) = D\xi + D\eta.$$

Доказательство. Первые три свойства очевидны. Докажем четвертое. По определению дисперсии

$$\begin{aligned} D(\xi + \eta) &= E(\xi + \eta - E(\xi + \eta))^2 = E(\xi - E\xi + \eta - E\eta)^2 = \\ &= E(\xi - E\xi)^2 + 2E((\xi - E\xi)(\eta - E\eta)) + E(\eta - E\eta)^2 = \\ &= D\xi + D\eta + 2(E\xi\eta - E\xi E\eta) \end{aligned}$$

Осталось показать, что $E\xi\eta - E\xi E\eta = 0$. Действительно

$$\begin{aligned} E\xi\eta &= \sum_{a,b} abP(\xi = a, \eta = b) = \sum_{a,b} abP(\xi = a)P(\eta = b) = \\ &= \sum_a aP(\xi = a) \sum_b bP(\eta = b) = E\xi E\eta \end{aligned}$$

■

Пример 1.155. Найти дисперсию случайной величины ξ , которая задана следующим законом распределения (таблица 1.54):

Таблица 1.54

ξ	1	2	5
p	0.3	0.5	0.2

Используя (1.139) имеем

$$E\xi = 1 \times 0.3 + 2 \times 0.5 + 5 \times 0.2 = 2.3.$$

Далее

$$\begin{cases} (\xi_1 - E\xi)^2 = (1 - 2.3)^2 = 1.69, \\ (\xi_2 - E\xi)^2 = (2 - 2.3)^2 = 0.09, \\ (\xi_3 - E\xi)^2 = (5 - 2.3)^2 = 7.29. \end{cases}$$

Тогда, согласно (1.140):

$$D\xi = 1.69 \times 0.3 + 0.09 \times 0.5 + 7.29 \times 0.2 = 2.01.$$

Пример 1.156. В урне 4 черных и 5 белых шаров. Наудачу извлекают два шара. Случайная величина ξ – количество черных шаров среди извлеченных. Найти математическое ожидание и дисперсию случайной величины.

Возможны три элементарных события, соответствующие извлечению двух шаров из урны: ω_0 – все извлеченные шары белые, ω_1 – извлечены

белый и черный шары, ω_2 – извлечены два черных шара. Таким образом, $0, 1, 2$ – возможные значения дискретной случайной величины ξ .

Найдем вероятности, с которыми случайная величина ξ принимает эти значения. Общее число исходов $C_9^2 = 36$. Тогда

$$\begin{aligned} P(\xi = 0) &= P(\omega_0) = p_1 = \frac{C_5^2}{36} = \frac{10}{36} \\ P(\xi = 1) &= P(\omega_1) = p_2 = \frac{C_4^1 C_5^1}{36} = \frac{20}{36} \\ P(\xi = 2) &= P(\omega_2) = p_3 = \frac{C_5^2}{36} = \frac{6}{36}. \end{aligned}$$

Согласно (1.139) и (1.140) имеем:

$$\begin{aligned} E\xi &= 0 \times \frac{10}{36} + 1 \times \frac{20}{36} + 2 \times \frac{6}{36} = \frac{8}{9} \approx 0.89, \\ D\xi &= \left(0 - \frac{8}{9}\right)^2 \times \frac{10}{36} + \left(1 - \frac{8}{9}\right)^2 \times \frac{20}{36} + \left(2 - \frac{8}{9}\right)^2 \times \frac{6}{36} \approx 0.4321. \end{aligned}$$

Задачи для самостоятельного решения

1. Натуральное число представлено в каноническом представлении (1.6) в виде произведения степеней своих простых делителей

$$n = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_k^{\alpha_k}$$

Сколько всего делителей у числа n ?

2. Разложите на простые множители $100!$

3. Представьте $\sqrt{368}$ в виде периодической цепной дроби и вычислите ее значение с точностью до $\varepsilon = 10^{-5}$.

4. Найдите общее решение диофантина уравнения $3427x - 6256y = -161$.

5. Вычислите $\frac{16}{72}$ в кольце вычетов по модулю 97.

6. Найдите наименьшее натуральное число x , удовлетворяющее условиям:

$$x \equiv 11 \pmod{12}, \quad x \equiv 20 \pmod{31}, \quad x \equiv 1 \pmod{29}, \quad x \equiv 9 \pmod{13}.$$

7. Найдите остаток от деления $36^{29^{39}}$ на 77.

8. Найдите НОД многочленов

$$3x^7 + 6x^6 - 3x^5 + 4x^4 + 14x^3 - 6x^2 - 4x + 4 \quad \text{и} \quad 3x^6 - 3x^4 + 7x^3 - 6x + 2$$

и его линейное представление над полем \mathbb{R} .

9. Найдите НОД многочленов $x^5 + x^4 + 1$ и $x^4 + x^2 + 1$ и его линейное представление над полем \mathbb{Z}_2 .

10. Найдите рациональные корни многочлена $F(x) = 24x^4 - 42x^3 - 77x^2 + 56x + 60$.

11. Постройте интерполяционный многочлен $F(x)$, удовлетворяющий условиям:

$$F(-1) = -8, F(-2) = 3, F(1) = -6, F(2) = -5, F(3) = 28.$$

12. Пусть $m = 35$ и $e = 7$ открытая часть ключа RSA. Найдите закрытую часть ключа d .

13. Поставьте электронную подпись на сообщении, заданным числом 27, используя код RSA с открытыми частями $m = 17 \cdot 23$ и $e = 125$.

14. Сформулируйте следующую геометрическую задачу в терминах многочленов. В треугольнике ABC точка D лежит на прямой AC , а точка E — на прямой BC , так что $AD = BE$. F — точка пересечения DE и AB . Докажите, что $FD \cdot AC = EF \cdot BC$.

15. Определите, можно ли выразить многочлен $q = -3x^3 - 5x^2 - x^4 + 5$ через многочлены $p_1 = x^3 - 2x - 3$ и $p_2 = x^4 - 1$. Если можно, то найдите выражение.

16. Изобразите степени мономов многочлена

$$x^3 + 3xy^2 - y^2 + 2x - xy^3 - x^2 - xy + xy^4 - 1$$

точками плоскости.

17. Упорядочьте одночлены многочлена

$$x^3 + 3xy^2 - y^2 + 2x - xy^3 - x^2 - xy + xy^4 - 1$$

по степеням.

18. Найдите конечный базис мономиального идеала, порожденного бесконечным множеством одночленов $\{xy^n; x^2; y^n \mid n — \text{натуральное}\}$. Отметьте в координатной плоскости точки, соответствующие базисным мономам.

19. Редуцируйте многочлен

$$P(x, y) = x^3 - x^2 + xy^4 - xy^3 + 3xy^2 - xy + 2x - y^2 - 1$$

по базису Гребнера из задачи 22. Редуцируется ли многочлен к нулю?

20. Постройте S -многочлен для многочленов

$$P(x, y) = 2x^2y^2 + x^2y \text{ и } Q(x, y) = 3xy^4 - xy^3.$$

21. Постройте базис Гребнера для идеала, построенного по следующим многочленам $\{x^2 - 1, (x - 1)y, (x + 1)z\}$

22. Упростите следующий базис Гребнера

$$\{x^2y^2 - y^2 - 1, x^2 + xy^3 + xy + y^2 + 1, xy^2 + x, y^4 + 2y^2 + 1\}.$$

23. Докажите, что добавление к базису многочленов, полученных редукцией, или S -многочленов не меняет идеала, определяемого этим базисом.

24. Найдите число целых слагаемых бинома $(\sqrt[3]{3} + \sqrt[5]{5})^{35}$.
25. Найдите коэффициент при x^5 в разложении $(1 + x - x^3)^6$.
26. Имеется n черных и m белых шаров. Сколькими способами можно их выложить в ряд так, чтобы никакие два черных шара не лежали рядом?
27. Будем говорить, что два числа образуют инверсию, если большее из них написано ранее, чем меньшее. Сколько инверсий во всех перестановках чисел $\{1, \dots, n\}$?
28. Определите количество двоичных 15-значных чисел, имеющих в записи 11 единиц.
29. Из колоды, содержащей 52 карты, вынули 10 карт.
1. В скольких случаях среди этих карт окажется хотя бы один туз?
 2. В скольких случаях — ровно один туз?
30. Сколько существует решений уравнения $x_1 + x_2 + \dots + x_{15} = 200$ в целых числах, где $x_i \geq 1$?
31. Сколько существует 5-значных чисел в 14-чной системе счисления, у которых не все цифры разные?
32. Порождающий многочлен линейного циклического кода длины $n = 7$, который осуществляет кодирование сообщений длины $m = 4$ имеет вид $g(x) = 1 + x + x^3$. Декодируйте сообщение $(1, 1, 1, 0, 1, 1, 1)$, если известно, что оно имеет не более одной ошибки.
33. Пусть имеется 6 кодовых символов: D, E, H, T, C, Y с частотами появления

D	E	H	T	C	Y
0.20	0.21	0.15	0.17	0.18	0.09

Постройте код Шеннона–Фэнго и зашифруйте сообщение: $C T U D E N T$.

34. Все перестановки 7 чисел $\{1, 2, 3, 4, 5, 6, 7\}$ упорядочены в лексикографическом порядке.
1. Какой по счету идет перестановка $\{4, 6, 5, 3, 7, 2, 1\}$.
 2. Найдите перестановку с номером 2546.
35. Среди 100 целых чисел 40 кратно 3, 3 кратно 11, 18 кратно 9, 2 кратно 33, 1 кратно 99. Определить, сколько из них кратно 3 или 11, но не кратно 9.
36. Каждый студент в группе знает хотя бы один иностранный язык. Шестеро знают английский, шестеро — немецкий, семеро — французский. Четверо знают английский и немецкий, трое — немецкий и французский, двое — французский и английский. Один человек знает все три языка.
1. Сколько человек в группе?
 2. Сколько из них знают только английский язык?

37. Сколькоими способами можно окрасить в три цвета:

- а) грани правильного тетраэдра;
- б) ребра правильного тетраэдра;
- в) вершины правильного тетраэдра.

Как и раньше считаем раскраски одинаковыми, если их можно получить друг из друга самосовмещениями тетраэдра при поворотах.

38. Решите ту же задачу, если, кроме указанных, одинаковыми считаются раскраски, которые переходят друг в друга при самосовмещениях посредством зеркальной симметрии.

39. Решите предыдущие две задачи для куба, октаэдра, правильной треугольной призмы и раскраски n красками.

40. Сколькоими способами можно раскрасить ребра куба в 6 красок так, чтобы одинаково окрашенных ребер было ровно по два (6 пар по 2 ребра). Одинаковыми считаются раскраски, переходящие друг в друга поворотными самосовмещениями.

41. Сколькоими способами можно собрать ожерелье из 12 бусинок, из которых 6 бусинок красного цвета, 4 зеленого и 2 синего.

42. Найдите решение однородного рекуррентного уравнения $3f_{n+2} - 8f_{n+1} + 4f_n = 0$ с начальными условиями $f_0 = 2, f_1 = 2$,

43. Найдите общее решение неоднородного рекррентного уравнения

$$f_{n+2} - 3f_{n+1} + 2f_n = n^2 + 1.$$

Глава 2

Графы и бинарные отношения

2.1. Определения графов

2.1.1. Простейшие определения и свойства

Определение 2.1. Неориентированным *графом*, или просто *графом* $\mathbf{G} = \mathbf{G}(\mathbf{V}, \mathbf{E})$, называют пару множеств: $\mathbf{V} = \{v_1, \dots, v_p\}$ — непустое, конечное множество *вершин* и $\mathbf{E} = \{(v_i, v_j) \mid v_i, v_j \in \mathbf{V}\}$ — множество неупорядоченных пар вершин, называемых *ребрами*.

Вершины и ребра графа называют его *элементами*. Число вершин графа \mathbf{G} называют его *порядком* и обозначают $|\mathbf{G}|$. Если $|\mathbf{G}| = n$, а $|\mathbf{E}| = m$, то граф называют (n, m) -графом.

Ребро (v, v) называют *петлей*.

Замечание 2.1. Иногда рассматривают более общее понятие графа, полагая что множество ребер \mathbf{E} есть семейство подмножеств множества $\mathbf{V} \times \mathbf{V}$, т. е. допускается существование кратных ребер в графе. Такой граф с кратными ребрами и петлями называют *псевдографом*. Псевдограф без петель называют *графом с кратными ребрами* или *мультиграфом*. Примерами мультиграфов могут служить структурные формулы в органической химии. Ребра, соединяющие одну и ту же пару вершин, называют *параллельными*. Мультиграф, у которого максимальное число параллельных дуг равно s , называют *s-графом*.

Граф по определению 2.1, в котором любые две его вершины соединены не более чем одним ребром (т. е. 1-граф), иногда называют *униграфом*.

Далее в качестве графов будем рассматривать униграфы без петель, хотя некоторые результаты этой главы справедливы и для мультиграфов.

Определение 2.2. Рассмотрим ребро $e = (v, u) \in \mathbf{E}$. Тогда вершины v, u называют *концами* или *концевыми точками* ребра e . При этом говорят, что ребро e *инцидентно* вершинам v, u . Вершины v, u называют *смежными*.

Множество вершин, смежных с вершиной v , называют *множеством смежности* вершины v и обозначают $\Gamma^+(v)$.

Определение 2.3. Граф \mathbf{G} называют *полным*, если любые две его вершины смежны. Полный граф с n вершинами обозначается K_n .

Определение 2.4. Звездой вершины u называют подмножество ребер $\{(v_i, v_j) \in E \mid v_i = u \text{ или } v_j = u\} = \delta(u)$. Число $|\delta(u)| = \deg(u)$ называют степенью (или валентностью) вершины u . При $\deg(u) = 0$ вершину называют изолированной, при $\deg(u) = 1$ — концевой, или висячей. Если степени всех вершин графа равны k , то граф называется регулярным, или однородным степени k . Регулярные графы третьей степени часто называют кубическими.

Теорема 2.1 (теорема Эйлера). Для произвольного графа $G(V, E)$ справедливо равенство

$$|E| = \frac{1}{2} \sum_{v \in V} \deg(v).$$

Доказательство. В графе G нет петель и каждое ребро учитывается дважды. \blacksquare

Следствие 2.1 (Лемма о рукопожатиях). У любого графа G число вершин нечетной степени — четно.

Определение 2.5. Граф $G' = (V', E')$ называют подграфом графа G , если $V' \subseteq V$, $E' \subseteq E$. В частности, подграфами G являются пустой граф и сам граф. Все остальные подграфы называют собственными подграфами. Если $V' = V$, то G' называют остовным подграфом G .

Определение 2.6. Путем длины l называют последовательность $v_0, e_1, v_1, e_2, \dots, v_{l-1}, e_l, v_l$, где e_i — ребро (v_{i-1}, v_i) , $v_i \in V$. При $v_0 \neq v_l$ путь называют открытым (незамкнутым), при $v_0 = v_l$ — циклическим (замкнутым); если $e_i \neq e_j$ ($i \neq j$) — простым. Открытый путь, где все вершины различны, называют простой цепью, или просто $(v_0 - v_l)$ -цепью. Замкнутый путь, где все вершины (кроме концевых) различны, называют простым циклом, или циклом.

Замечание 2.2. Очевидно, что если все вершины пути различны, то путь является простым.

Пример 2.1. Рис. 2.1 иллюстрирует определение 2.6.

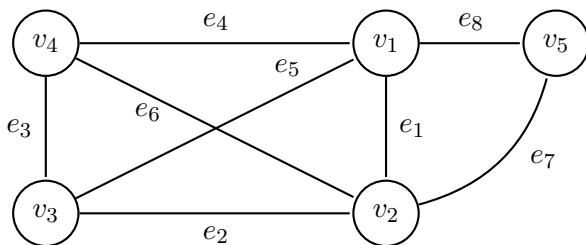


Рис. 2.1

- 1) $v_1, e_1, v_2, e_6, v_4, e_4, v_1, e_1, v_2$ — открытый путь;

- 2) $v_1, e_1, v_2, e_2, v_3, e_5, v_1, e_4, v_4$ — открытый простой путь, но не цепь;
- 3) $v_1, e_1, v_2, e_6, v_4, e_3, v_3$ — простая $(v_1 - v_3)$ -цепь;
- 4) $v_1, e_5, v_3, e_2, v_2, e_6, v_4, e_3, v_3, e_5, v_1$ — замкнутый путь;
- 5) $v_1, e_1, v_2, e_6, v_4, e_3, v_3, e_2, v_2, e_7, v_5, e_8, v_1$ — замкнутый простой путь, но не цикл;
- 6) $v_1, e_1, v_2, e_2, v_3, e_3, v_4, e_4, v_1$ — простой цикл.

Теорема 2.2. Если существует путь из вершины v в вершину u ($v \neq u$), то из ребер этого пути можно построить $(v - u)$ -цепь.

Доказательство. Пусть существует путь: $v = v_0, e_1, \dots, e_l, v_l = u$. Рассмотрим все пути из v в u , состоящие из ребер, входящих в этот путь. Среди них выберем самый короткий $v = v_0, e'_1, v'_1, \dots, v'_{k-1}, e'_k, v'_k = u$. Покажем, что $v''_i \neq v''_j$ ($i \neq j$). Пусть $v'_i = v'_j$, тогда путь

$$v = v_0, e'_1, \dots, v'_i, e'_{j+1}, v'_{j+1}, \dots, v'_k = u$$

имеет длину $k - (j - i) < k$. Противоречие с предположением о минимальности пути. ■

Теорема 2.3 (аналог теоремы 2.2 для замкнутых путей). Если существует замкнутый простой путь, то из его ребер можно составить цикл.

Замечание 2.3. Требование простоты пути в теореме 2.3 существенно: рассмотрим путь v_0, e, v_1, e, v_0 . Из ребра e цикл построить нельзя.

Упражнение 2.1. Докажите теорему 2.3.

Теорема 2.4. Пусть $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ — граф, все вершины которого имеют четную степень. Тогда для любого ребра существует замкнутый простой путь, содержащий это ребро.

Доказательство. Рассмотрим произвольное ребро $e_1 = (v_0, v_1)$. По предположению теоремы $\deg(v_1)$ — четное число. Тогда существует ребро $e_2 = (v_1, v_2)$, отличное от e_1 . В силу четности $\deg(v_2)$ найдется ребро $e_3 = (v_2, v_3)$ и т. д. Так как множество вершин \mathbf{V} конечно и каждый раз берется новое неиспользованное ребро, инцидентное v_i , то в конце концов снова придем в вершину v_0 . ■

Теорема 2.5. В предположении теоремы 2.4, если $\mathbf{E} \neq \emptyset$, то \mathbf{E} есть объединение множества ребер некоторых циклов, никакие два из которых не имеют общих ребер.

Доказательство. Возьмем произвольное ребро e_1 . Строим для него замкнутый простой путь, как в теореме 2.4: $v_0, e_1, \dots, e_l, v_l$. Вершины в этом пути могут повторяться. Допустим, что вершины v_0, v_1, \dots, v_{k-1} – различные, а $v_k = v_i$, $i \in 0 : k - 1$. Тогда ребра $\{e_{i+1}, \dots, e_k\}$ образуют цикл. Рассмотрим множество ребер $\mathbf{E}_1 = \mathbf{E} \setminus \{e_{i+1}, \dots, e_k\}$. Очевидно, что степень всех вершин графа $\mathbf{G}_1 = (\mathbf{V}, \mathbf{E}_1)$ тоже останется четной. Продолжив этот процесс, пока очередное множество $\mathbf{E}_k \neq \emptyset$, получим требуемое представление для \mathbf{E} . ■

Следствие 2.2. Если в графе \mathbf{G} существуют две различные цепи \mathbf{P}_1 и \mathbf{P}_2 , содержащие одни и те же различные вершины v и u , то из некоторого подмножества ребер \mathbf{P}_1 и \mathbf{P}_2 можно построить цикл.

Упражнение 2.2. Докажите, что в предположении теоремы 2.5 для любого ребра $e_i \in \mathbf{E}$ существует цикл, его содержащий.

Определение 2.7. Если элементами множества ребер \mathbf{E} являются упорядоченные пары вершин $[v, u]$, то граф $\mathbf{G} = \mathbf{G}(\mathbf{V}, \mathbf{E})$ называют *ориентированным*, или *орграфом*. В этом случае элементы множества \mathbf{V} называют *узлами*, а элементы \mathbf{E} – *дугами*. Для дуги $e = [v, u] \in \mathbf{E}$ узел v – *начало*, а узел u – *конец* дуги.

Для каждого узла v имеется две звезды: $\delta^+(v)$, $\delta^-(v)$ – множества выходящих и входящих в v дуг; при этом $|\delta^+(v)|$ – положительная, а $|\delta^-(v)|$ – отрицательная степени узла v .

Для орграфов имеет место теорема, аналогичная теореме Эйлера (см. теорему 2.1).

Теорема 2.6. Справедливо равенство

$$|\mathbf{E}| = \sum_{v \in \mathbf{V}} |\delta^+(v)| = \sum_{v \in \mathbf{V}} |\delta^-(v)|.$$

Доказательство. Каждому ребру нужно сопоставить его начало или его конец. ■

Определение 2.8. Направленным путем длины l называют последовательность $v_0, e_1, v_1, e_2, \dots, v_{l-1}, e_l, v_l$, где e_i – дуги $[v_{i-1}, v_i]$.

Аналогично определяют понятия открытого, замкнутого, простого пути, цепи и цикла.

Для орграфов справедливы теоремы, аналогичные теоремам 2.2 и 2.3.

Теорема 2.7. Если различные узлы v и u соединены $(v \rightarrow u)$ путем, то существует $(v \rightarrow u)$ цепь, построенная из дуг этого пути.

Теорема 2.8. Если в графе существует замкнутый ориентированный путь, то из дуг этого пути можно построить ориентированный цикл.

Замечание 2.4. Доказательства этих теорем аналогичны доказательствам теорем 2.2 и 2.3, но требование простоты для теоремы 2.3 в теореме 2.8 не нужно. Это связано с тем, что в ориентированном графе нет замкнутых путей, подобных пути: v_0, e, v_1, e, v_0 . В пути v_0, e, v_1, y, v_0 — две дуги $e = [v_0, v_1]$, $y = [v_1, v_0]$, $e \neq y$.

Теореме 2.4 соответствует следующая теорема.

Теорема 2.9. Пусть $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ — ориентированный граф и для любого узла v справедливо $|\delta^+(v)| = |\delta^-(v)|$. Тогда для любой дуги существует замкнутый простой ориентированный путь, содержащий эту дугу.

Доказательство. Следует повторить рассуждения, использованные в доказательстве теоремы 2.4. При этом всегда следует двигаться согласно ориентации дуг: пусть дуга $e = [v_0, v_1]$. Так как $|\delta^+(v_1)| = |\delta^-(v_1)|$, то существует дуга $e_2 \neq e_1$, исходящая из v_1 , и т. д. ■

Аналог теоремы 2.5 можно получить, если считать все дуги при доказательстве ориентированными.

Историческая справка. Термин «граф» впервые появился в книге венгерского математика Д. Кёнига в 1936 г., хотя начальные задачи теории графов восходят еще к Л. Эйлеру¹.

Математический термин граф с дворянским титулом «граф» связывает общее происхождение от латинского слова «графио» - пишу.

Первые задачи теории графов были связаны с решением математических развлекательных задач и головоломок, например:

- задача о кенигсбергских мостах (задача Эйлера), развитие которой привело к циклу задач об обходах графов;
- задачи о перевозках, решение которых привело к созданию эффективных методов решения транспортных задач и др.

Попытки решить сформулированную в середине XIX в. задачу четырех красок привели к появлению некоторых исследований графов, имеющих теоретическое и практическое значение. Многие результаты, относящиеся к теории графов, были получены при решении практических проблем. Так, Г. Кирхгоф при составлении полной системы уравнений для токов и напряжений в электрической схеме предложил, по существу, представлять такую схему графом и находить в нем деревья, с помощью которых выделяются линейно независимые системы контуров.

В XX в. задачи, связанные с графиками, начали возникать не только в физике, электротехнике, химии, биологии, экономике и т. д., но и внутри математики, в таких ее разделах, как алгебра, топология, теория вероятностей, теория чисел. Методы этих разделов стали успешно применяться для решения задач теории графов. В настоящее время графы эффективно используются в теории планирования и управления, теории расписаний, социологии, математической лингвистике, медицине, географии. Широкое применение находят графы в таких областях, как программирование, теория конечных автоматов, электроника.

Замечание 2.5. Наряду с термином граф в начале XX в. употреблялись в качестве синонимов и другие термины, например, карта, комплекс, диаграмма, сеть, лабиринт.

¹Euler L. Solutio problematis ad geometriam situs pertinentes. Commentarii Academiae Petropolitanae. 8. 1736. P. 128 -140.

2.1.2. Машинное представление графов

Машинное представление графов допускает большое разнообразие. Сложность получения ответа на тот или иной вопрос относительно данного графа во многом зависит от способа представления графа. Поэтому в алгоритмах на графах связь *алгоритм + структура данных* проявляется очень сильно. Один и тот же алгоритм, реализованный на различных структурах данных, очень часто приводит к совершенно разным программам. Во многих задачах на графах выбор представления является решающим для повышения эффективности алгоритма.

□ Матрица инциденций

Граф \mathbf{G} , имеющий n вершин и m ребер, представляется матрицей размером $n \times m$. Строки матрицы соответствуют вершинам, а столбцы — ребрам графа. Для неориентированного графа матрица строится следующим образом. Если u, v смежные вершины, то столбец (u, v) содержит 1 в строках u и v . Остальные элементы равны нулю.

В случае ориентированного графа, при наличии ориентированного ребра $[u, v]$ для столбца $[u, v]$ в строке u ставится -1 , в строке v соответственно 1 , а остальные элементы равны нулю.

Пример 2.2. Матрицы инциденций для неориентированного и ориентированного графов на рис. 2.2, *a* и *б* представлены соответственно в табл. 2.1 и табл. 2.2.

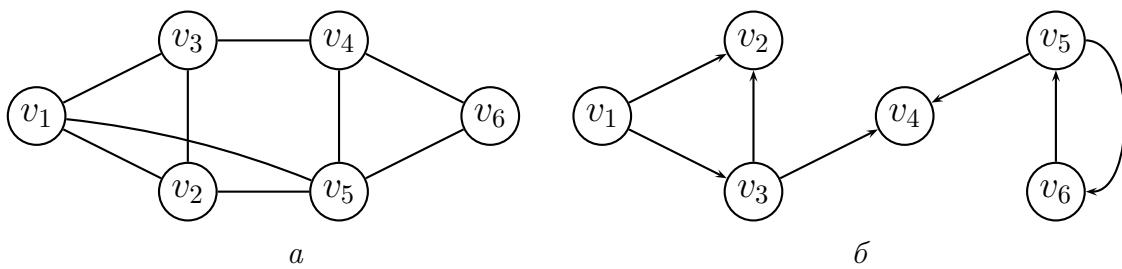


Рис. 2.2

Замечание 2.6. Недостатками такого представления графов является большой объем данных и неудобство доступа к информации. Для ответа на вопрос «существует ли в графе ребро (x, y) » в худшем случае нужно осуществить полный перебор столбцов, то есть сделать m шагов.

□ Матрица смежности

Матрица смежности для графа \mathbf{G} , имеющего n вершин, представляется $n \times n$ матрицей. Ее строки и столбцы соответствуют вершинам графа. Элемент матрицы, стоящий на пересечении строки v_i и столбца v_j равен

Таблица 2.1

	(v_1, v_2)	(v_1, v_3)	(v_1, v_5)	(v_2, v_3)	(v_2, v_5)	(v_3, v_4)	(v_4, v_5)	(v_4, v_6)	(v_5, v_6)
v_1	1	1	1	0	0	0	0	0	0
v_2	1	0	0	1	1	0	0	0	0
v_3	0	1	0	1	0	0	0	0	0
v_4	0	0	0	0	0	1	1	1	0
v_5	0	0	1	0	1	0	1	0	1
v_6	0	0	0	0	0	0	0	1	1

Таблица 2.2

	$[v_1, v_2]$	$[v_1, v_3]$	$[v_3, v_2]$	$[v_3, v_4]$	$[v_5, v_4]$	$[v_5, v_6]$	$[v_6, v_5]$
v_1	-1	-1	0	0	0	0	0
v_2	1	0	1	0	0	0	0
v_3	0	1	-1	-1	0	0	0
v_4	0	0	0	1	1	0	0
v_5	0	0	0	0	-1	-1	1
v_6	0	0	0	0	0	1	-1

1, если существует ребро (v_i, v_j) (дуга $[v_i, v_j]$ в случае ориентированного графа). Остальные элементы матрицы равны нулю.

Пример 2.3. Матрицы смежности для неориентированного и ориентированного графов из примера 2.2 соответственно приведены в табл. 2.3 *a* и *б*.

Таблица 2.3

	<i>a</i>						<i>б</i>						
	v_1	v_2	v_3	v_4	v_5	v_6		v_1	v_2	v_3	v_4	v_5	v_6
v_1	0	1	1	0	1	0	v_1	0	1	1	0	0	0
v_2	1	0	1	0	1	0	v_2	0	0	0	0	0	0
v_3	1	1	0	1	0	0	v_3	0	1	0	1	0	0
v_4	0	0	1	0	1	1	v_4	0	0	0	0	0	0
v_5	1	1	0	1	0	1	v_5	0	0	0	1	0	1
v_6	0	0	0	1	1	0	v_6	0	0	0	0	1	0

Замечание 2.7.

1) Очевидно, что для неориентированного графа матрица смежности симметрична.

2) Преимуществом такого представления является решение за один шаг вопроса: существует ли в графе ребро (x, y) ?

3) Недостатком матрицы смежности является то, что объем данных не зависит от числа ребер и всегда равен n^2 .

□ Списки инцидентности

Для каждой вершины графа $v \in \mathbf{V}$ вводится список вершин $u \in \mathbf{V}$ таких, что существует ребро (v, u) (дуга $[v, u]$ для ориентированного графа). Начало каждого списка хранится в таблице указателей НАЧАЛО. $\text{НАЧАЛО}[v]$ — это указатель на начало соответствующего списка инцидентности для вершины v . Весь такой список будем обозначать $\text{ЗАПИСЬ}[v]$. Каждый элемент списка это структура из двух полей

Вершина

Указатель на следующий элемент списка

Замечание 2.8. Для неориентированных графов каждое ребро (u, v) представлено дважды: через вершину v в списке $\text{ЗАПИСЬ}[u]$ и через вершину u в списке $\text{ЗАПИСЬ}[v]$.

Пример 2.4. Составим списки инцидентности для неориентированного и ориентированного графов из примера 2.2(рис. 2.3).

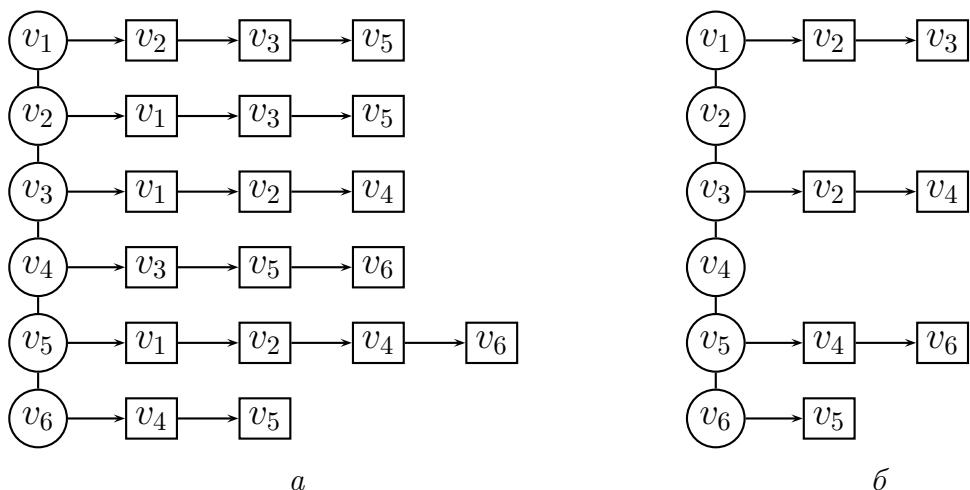


Рис. 2.3

□ Список ребер

Представляем граф в виде списка ребер или дуг как пар вершин с учетом направления дуги для ориентированных графов. Иногда, для ускорения поиска, список упорядочивают в лексиграфическом порядке.

Пример 2.5. Составим списки ребер для неориентированного и ориентированного графов из примера 2.2.

$$\begin{aligned} (v_1, v_2) \rightarrow (v_1, v_3) \rightarrow (v_1, v_5) \rightarrow (v_2, v_3) \rightarrow (v_2, v_5) \rightarrow (v_3, v_4) \rightarrow (v_4, v_5) \rightarrow \\ \rightarrow (v_4, v_6) \rightarrow (v_5, v_6) \\ (v_1, v_2) \rightarrow (v_1, v_3) \rightarrow (v_3, v_2) \rightarrow (v_3, v_4) \rightarrow (v_5, v_4) \rightarrow (v_5, v_6) \rightarrow (v_6, v_5) \end{aligned}$$

Замечание 2.9. В таком представлении удобно добавлять и удалять ребра, представлять сильно разреженные графы. Неудобно определять смежность вершин и ребер, осуществлять перебор инцидентных заданной вершине ребер.

2.1.3. Поиск в глубину и ширину в графе

Методы поиска в графе являются основой для построения алгоритмов на графах. Под поиском на графах будем понимать процесс просмотра всех вершин графа в целях отыскания вершин, удовлетворяющих некоторому условию.

Рассмотрим два подхода, базирующихся на двух основных идеях в программировании:

- 1) стеке¹ (магазине) — «первым пришел, последним ушел»,
- 2) очереди — «первым пришел, первым ушел».

Будем считать, что для описания графа используется одно из представлений предыдущего подраздела, например, списки инцидентности.

Первый алгоритм основан на идее использования стека и называется *поиск в глубину в графе*² (другие названия — *возвратный ход*, *бектрекинг*, *обход графа в глубину*).

Применение правила LIFO (Last In First Out — последним пришел, первым обслужен) приводит к следующей стратегии: идти «вглубь» графа, пока это возможно (есть непросмотренные вершины), возвращаться и искать другой путь, когда таких вершин нет. Поиск продолжается, пока не просмотрены все вершины, достижимые из исходной.

Замечание 2.10. Как правило стек реализуется в виде одностороннего списка — каждый элемент списка указывает только на следующий. При этом доступ возможен только к верхнему элементу (вершине стека).

Обозначим стек — S , вершину стека — $\text{top}(S)$, список просмотренных вершин — L .

Алгоритм 2.1 (поиск в глубину в графе «DFS»).

$S \leftarrow v_0$ ▷ записываем в стек произвольную вершину v_0 , помечаем ее как просмотренную

¹Понятие стека (англ. stack — стопка) было введено Аланом Тьюрингом в 1946г.

²DFS: Depth-first search.

```

 $L \leftarrow v_0$ 
while  $S \neq \{\emptyset\}$  do
     $v \leftarrow \text{top}(S)$ 
    if существует смежная с  $v$  вершина  $u \notin L$  then
         $S \leftarrow u$             $\triangleright$  записываем в стек вершину  $u$ , помечаем ее как просмотренную
         $L \leftarrow u$ 
    else
        Удаляем вершину  $v$  из стека
    end if
end while

```

Пример 2.6. Рассмотрим работу алгоритма на примере графа, представленного на рис. 2.4. Поиск начнем с вершины v_1 .

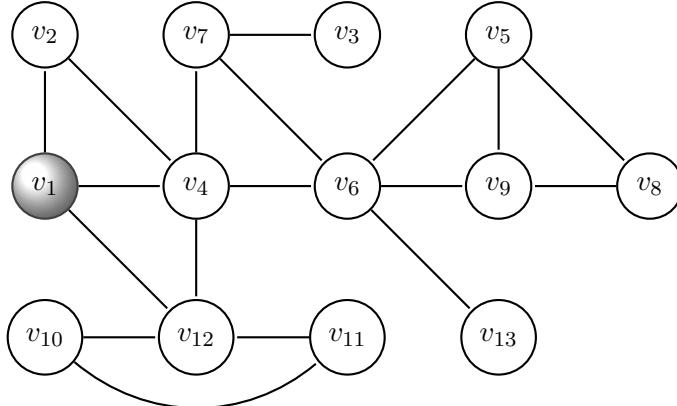


Рис. 2.4

Будем схематично изображать стек развернутым по горизонтали. Считаем, что вершина стека находится слева.

Порядок просмотра вершин следующий:

$v_1 \rightarrow v_2 \rightarrow v_4 \rightarrow v_6 \rightarrow v_5 \rightarrow v_8 \rightarrow v_9 \rightarrow v_7 \rightarrow v_3 \rightarrow v_{13} \rightarrow v_{12} \rightarrow v_{10} \rightarrow v_{11}$

Изменения состояния стека показаны в табл. 2.4.

Таблица 2.4

Текущая вершина	Действие	Состояние стека
v_1	Добавляем вершину v_1 в стек	v_1
v_2	Добавляем вершину v_2 в стек	$v_2 v_1$
v_4	Добавляем вершину v_4 в стек	$v_4 v_2 v_1$
v_6	Добавляем вершину v_6 в стек	$v_6 v_4 v_2 v_1$
v_5	Добавляем вершину v_5 в стек	$v_5 v_6 v_4 v_2 v_1$
v_8	Добавляем вершину v_8 в стек	$v_8 v_5 v_6 v_4 v_2 v_1$
v_9	Добавляем вершину v_9 в стек	$v_9 v_8 v_5 v_6 v_4 v_2 v_1$
v_9	Удаляем вершину v_9 из стека	$v_8 v_5 v_6 v_4 v_2 v_1$

Продолжение табл. 2.4

Текущая вершина	Действие	Состояние стека
v_8	Удаляем вершину v_8 из стека	$v_5 v_6 v_4 v_2 v_1$
v_5	Удаляем вершину v_5 из стека	$v_6 v_4 v_2 v_1$
v_7	Добавляем вершину v_7 в стек	$v_7 v_6 v_4 v_2 v_1$
v_3	Добавляем вершину v_3 в стек	$v_3 v_7 v_6 v_4 v_2 v_1$
v_3	Удаляем вершину v_3 из стека	$v_7 v_6 v_4 v_2 v_1$
v_7	Удаляем вершину v_7 из стека	$v_6 v_4 v_2 v_1$
v_{13}	Добавляем вершину v_{13} в стек	$v_{13} v_6 v_4 v_2 v_1$
v_{13}	Удаляем вершину v_{13} из стека	$v_6 v_4 v_2 v_1$
v_6	Удаляем вершину v_6 из стека	$v_4 v_2 v_1$
v_{12}	Добавляем вершину v_{12} в стек	$v_{12} v_4 v_2 v_1$
v_{10}	Добавляем вершину v_{10} в стек	$v_{10} v_{12} v_4 v_2 v_1$
v_{11}	Добавляем вершину v_{11} в стек	$v_{11} v_{10} v_{12} v_4 v_2 v_1$
v_{11}	Удаляем вершину v_{11} из стека	$v_{10} v_{12} v_4 v_2 v_1$
v_{10}	Удаляем вершину v_{10} из стека	$v_{12} v_4 v_2 v_1$
v_{12}	Удаляем вершину v_{12} из стека	$v_4 v_2 v_1$
v_4	Удаляем вершину v_4 из стека	$v_2 v_1$
v_2	Удаляем вершину v_2 из стека	v_1
v_1	Удаляем вершину v_1 из стека	Стек пуст

Замена стека, используемого при обходе в глубину, очередью FIFO (First In First Out — первым пришел, первым обнаружен) приводит к другому классическому алгоритму — *поиск в ширину в графе*¹ (другое название — *обход графа в ширину*).

Алгоритм поиска в ширину просматривает все вершины, достижимые из начальной. Происходит движение «вширь» графа, (сначала просматриваются все соседние вершины, затем соседи соседей и т. д.)

Замечание 2.11. Добавление вершины возможно лишь в конец очереди, выборка — только из начала очереди.

Обозначим очередь — D , начало и конец очереди — $\text{start}(D)$ и $\text{end}(D)$, список просмотренных вершин — L .

Алгоритм 2.2 (поиск в ширину в графе «BFS»).

```

 $\text{end}(D) \leftarrow v_0$                                  $\triangleright$  записываем в конец очереди произвольную вершину  $v_0$ 
 $L \leftarrow v_0$                                           $\triangleright$  помечаем ее как просмотренную
while  $D \neq \{\emptyset\}$  do
     $v \leftarrow \text{start}(D)$                                 $\triangleright v$  — вершина, находящаяся в начале очереди
    if существует смежная с  $v$  вершина  $u \notin L$  then

```

¹BFS: Breadth first search.

```

end( $D$ )  $\leftarrow u$                                  $\triangleright$  записываем в конец очереди вершину  $u$ 
 $L \leftarrow u$                                      $\triangleright$  помечаем вершину  $u$  как просмотренную
else
    Удаляем вершину  $v$  из начала очереди
end if
end while

```

Замечание 2.12. В отличие от предыдущего алгоритма, в силу конструктивных особенностей реализации очереди (замечание 2.11) здесь просматриваем все вершины одного уровня и затем переходим на следующий.

Рассмотрим работу алгоритма на графе из примера 2.6 (см. рис. 2.4).

Пример 2.7. Будем схематично изображать очередь развернутой по горизонтали. Считаем, что хвост очереди находится слева, а начало — справа.

Порядок просмотра вершин следующий:

$$v_1 \rightarrow v_2 \rightarrow v_4 \rightarrow v_{12} \rightarrow v_6 \rightarrow v_7 \rightarrow v_{10} \rightarrow v_{11} \rightarrow v_5 \rightarrow v_9 \rightarrow v_{13} \rightarrow v_3 \rightarrow v_8$$

Изменения состояния очереди показаны в табл. 2.5.

Таблица 2.5

Вершина на выходе очереди	Действие	Состояние очереди
	Добавляем вершину v_1 в очередь	v_1
v_1	Добавляем вершину v_2 в очередь	$v_2 v_1$
v_1	Добавляем вершину v_4 в очередь	$v_4 v_2 v_1$
v_1	Добавляем вершину v_{12} в очередь	$v_{12} v_4 v_2 v_1$
v_2	Удаляем вершину v_1 из очереди	$v_{12} v_4 v_2$
v_4	Удаляем вершину v_2 из очереди	$v_{12} v_4$
v_4	Добавляем вершину v_6 в очередь	$v_6 v_{12} v_4$
v_4	Добавляем вершину v_7 в очередь	$v_7 v_6 v_{12} v_4$
v_{12}	Удаляем вершину v_4 из очереди	$v_7 v_6 v_{12}$
v_{12}	Добавляем вершину v_{10} в очередь	$v_{10} v_7 v_6 v_{12}$
v_{12}	Добавляем вершину v_{11} в очередь	$v_{11} v_{10} v_7 v_6 v_{12}$
v_6	Удаляем вершину v_{12} из очереди	$v_{11} v_{10} v_7 v_6$
v_6	Добавляем вершину v_5 в очередь	$v_5 v_{11} v_{10} v_7 v_6$
v_6	Добавляем вершину v_9 в очередь	$v_9 v_5 v_{11} v_{10} v_7 v_6$
v_6	Добавляем вершину v_{13} в очередь	$v_{13} v_9 v_5 v_{11} v_{10} v_7 v_6$
v_7	Удаляем вершину v_6 из очереди	$v_{13} v_9 v_5 v_{11} v_{10} v_7$
v_7	Добавляем вершину v_3 в очередь	$v_3 v_{13} v_9 v_5 v_{11} v_{10} v_7$
v_{10}	Удаляем вершину v_7 из очереди	$v_3 v_{13} v_9 v_5 v_{11} v_{10}$
v_{11}	Удаляем вершину v_{10} из очереди	$v_3 v_{13} v_9 v_5 v_{11}$

Продолжение табл. 2.5

Вершина на выходе очереди	Действие	Состояние очереди
v_5	Удаляем вершину v_{11} из очереди	$v_3 v_{13} v_9 v_5$
v_5	Добавляем вершину v_8 в очередь	$v_8 v_3 v_{13} v_9 v_5$
v_9	Удаляем вершину v_5 из очереди	$v_8 v_3 v_{13} v_9$
v_{13}	Удаляем вершину v_9 из очереди	$v_8 v_3 v_{13}$
v_3	Удаляем вершину v_{13} из очереди	$v_8 v_3$
v_8	Удаляем вершину v_3 из очереди	v_8
	Удаляем вершину v_8 из очереди	Очередь пуста

2.1.4. Связность

Определение 2.9. Граф \mathbf{G} называют *связным*, если для любых двух различных вершин существует соединяющая их цепь. (Согласно теореме 2.2 можно говорить о пути).

Определение 2.10. Пусть \mathbf{S} — непустое подмножество вершин графа $\mathbf{G} = (\mathbf{V}, \mathbf{E})$. Обозначим через $\mathbf{E}(\mathbf{S}) \subseteq \mathbf{E}$ подмножество ребер, концевые вершины которых принадлежат \mathbf{S} . Граф $\mathbf{G}' = (\mathbf{S}, \mathbf{E}(\mathbf{S}))$ называют *подграфом* \mathbf{G} , *порожденным* \mathbf{S} .

Определение 2.11. Определим бинарное отношение \equiv на множестве вершин графа $\mathbf{G}(\mathbf{V}, \mathbf{E})$: $v \equiv u$, если существует $(v - u)$ -цепь.

Утверждение 2.1. Отношение \equiv есть отношение эквивалентности. Действительно, из определения 2.11 очевидны рефлексивность и симметричность. Далее, если $v \equiv u$, $u \equiv w$, то соединяя последовательно две цепи $(v - u)$, $(u - w)$, получаем цепь $(v - w)$. Доказана транзитивность.

Тогда множество вершин \mathbf{V} разбивается на непересекающиеся классы эквивалентности $\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_k$.

Определение 2.12. Подграфы графа \mathbf{G} , порожденные $\mathbf{V}_1, \dots, \mathbf{V}_k$: $(\mathbf{V}_1, \mathbf{E}(\mathbf{V}_1)), \dots, (\mathbf{V}_k, \mathbf{E}(\mathbf{V}_k))$, называют *связными компонентами* или *компонентами связности* графа \mathbf{G} , а k — *степенью связности* графа \mathbf{G} .

Замечание 2.13. Любая вершина и любое ребро принадлежат одной и только одной компоненте связности.

Для определения всех компонент связности неориентированного графа можно воспользоваться поиском в глубину (алгоритм 2.1).

```

while существуют непросмотренные вершины графа do
     $u \leftarrow$  первая непросмотренная вершина
    Поиск в глубину (DFS) из вершины  $u$ 
    Очередная компонента связности  $\leftarrow$  все просмотренные на данном шаге вершины
end while

```

Определение 2.13. Рассмотрим граф $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ и его ребро e . Определим подграф $\mathbf{G} - e = (\mathbf{V}, \mathbf{E} \setminus \{e\})$. Заметим, что при этом концевые точки ребра e остаются в графе $\mathbf{G} - e$.

Теорема 2.10. Если в связном графе $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ ребро e принадлежит некоторому циклу, то граф $G - e$ остается связным.

Доказательство. Пусть $v_0, e, v_1, e_2, \dots, e_l, v_l (= v_0)$ — цикл, содержащий ребро e . В любом пути, который включал удаленное ребро $e = (v_0, v_1)$, это ребро можно заменить на путь $(v_0) = v_l, e_l, \dots, e_2, v_1$. ■

Справедливо и обратное утверждение.

Теорема 2.11. Пусть $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ — связный граф, $e \in \mathbf{E}$. Если $\mathbf{G} - e$ также связный граф, то существует цикл в \mathbf{G} , содержащий ребро e .

Доказательство. Пусть $e = (v, u)$, тогда существует $(v - u)$ -цепь в графе $\mathbf{G} - e$: $(v) = v_0, e_1, v_1, e_2, \dots, v_{l-1}, e_l, v_l (= u)$. Добавив к этой цепи ребро e и вершину v , получим цикл в графе \mathbf{G} . ■

Определение 2.14. Ребро e графа $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ называют *мостом*, или *разделяющим ребром* или *ребром отделимости* или *перешейком*, если степень связности графа $\mathbf{G} - e$ больше степени связности графа \mathbf{G} .

Если при удалении вершины вместе с инцидентными ей ребрами степень связности графа увеличивается, вершина называется *точкой сочленения* или *разделяющей вершиной* или *шарниром*.

Граф называется *реберно связным*, если в нем отсутствуют мосты. В противном случае граф называется *реберно отделимым*.

При удалении всех мостов граф распадается на компоненты связности, которые называются *компонентами реберной двусвязности*.

Замечание 2.14. Если граф \mathbf{G} несвязный, то при удалении ребра e видоизменяется лишь компонента связности, содержащая e .

Опираясь на теоремы 2.10 и 2.11, легко получить критерий моста.

Теорема 2.12. Ребро e графа \mathbf{G} — мост тогда и только тогда, когда в \mathbf{G} нет циклов, содержащих это ребро.

Упражнение 2.3. Докажите самостоятельно теорему 2.12.

Пример 2.8. Рассмотрим граф, представленный на рис. 2.5.

Для данного графа мостом является каждое из ребер (v_4, v_5) , (v_5, v_6) и (v_5, v_7) . На рисунке они имеют большую «толщину». Действительно, в отличие от остальных ребер, для этих трех в графе нет циклов, их содержащих.

Вершины v_4 и v_5 – точки сочленения. При удалении всех мостов получаем четыре компоненты реберной двусвязности $\{v_1, v_2, v_3, v_4\}$, $\{v_5\}$, $\{v_6\}$ и $\{v_7\}$.

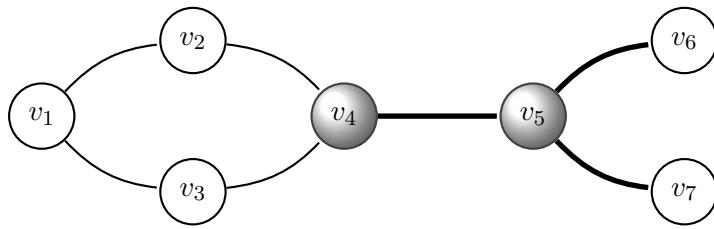


Рис. 2.5

Рассмотрим задачу нахождения всех мостов графа \mathbf{G} . Эта задача решается с помощью двух поисков в глубину.

Алгоритм 2.4 (нахождения всех мостов графа).

Шаг 1. Проводим поиск в глубину, при прохождении ребра (u, v) ориентируем его против направления движения, т. е. запрещаем прохождение по ребру в направлении от вершины v к вершине u . При обнаружении новой вершины заносим ее в список L .

Шаг 2. Проводим второй поиск в глубину с учетом ориентированности ребер. Внешний цикл по вершинам идет в таком порядке, в каком они записаны в списке L . Вершины каждой получающейся компоненты красим в свой цвет.

Шаг 3. Выбираем ребра исходного графа \mathbf{G} , соединяющие вершины разного цвета – они и будут искомыми мостами.

Пример 2.9. Рассмотрим работу алгоритма 2.4 на примере графа на рис. 2.5.

Запускаем поиск в глубину из вершины v_1 . После первого шага получаем ориентированный граф (рис. 2.6). Список L имеет вид:

$$L = \{1, 2, 4, 3, 5, 6, 7\}$$

Проводим второй поиск в глубину для графа на рис. 2.6 согласно списку L . Получаем следующие четыре компоненты: $\{v_1, v_3, v_4, v_2\}$ – цвет 1, $\{v_5\}$ – цвет 2, $\{v_6\}$ – цвет 3 и $\{v_7\}$ – цвет 4 (рис. 2.7).

Отсюда определяем мости графа: (v_4, v_5) , (v_5, v_6) , (v_5, v_7) .

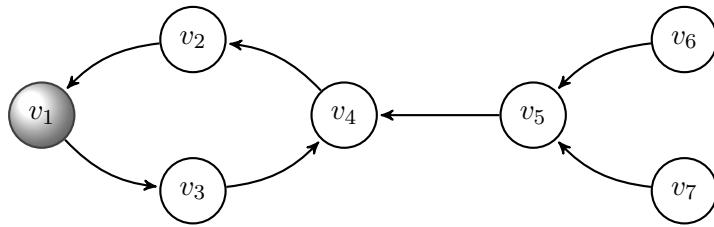


Рис. 2.6

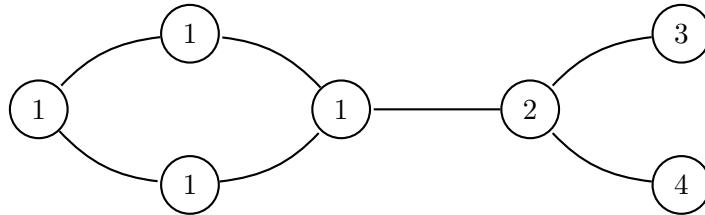


Рис. 2.7

Для ориентированных графов понятие связности вводится аналогичным образом.

Определение 2.15. Пусть $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ — ориентированный граф. Рассмотрим неориентированный граф $\mathbf{G}' = (\mathbf{V}, \mathbf{E}')$. Граф \mathbf{G}' получается из графа \mathbf{G} , если убрать ориентацию дуг. Если получившийся граф \mathbf{G}' является связным, то \mathbf{G} называют *слабосвязным* (т. е. можно попасть из любой вершины в любую, если не обращать внимания на ориентацию).

Ориентированный граф \mathbf{G} называют *сильносвязным*, если для любой пары узлов $v, u \in \mathbf{V}$ существуют $(v \rightarrow u)$ и $(u \rightarrow v)$ -цепи.

Компоненты связности сильносвязного графа иногда называют *сильными компонентами* или *бикомпонентами*.

Для ориентированных графов существует еще одно важное понятие.

Определение 2.16. Графом конденсации или *графом Герца* или *конденсацией* ориентированного графа \mathbf{G} называют такой ориентированный граф \mathbf{G}' вершинами которого служат компоненты сильной связности \mathbf{G} , а дуга в \mathbf{G}' между двумя вершинами проводится только если в графе \mathbf{G} существует хотя бы одна дуга между вершинами, входящими в соответствующие компоненты связности.

Замечание 2.15. Граф конденсации не содержит кратных ребер по построению.

Теорема 2.13. Граф конденсации не содержит циклов.

Доказательство. Вершины графа конденсации \mathbf{G}' (то есть компоненты сильной связности исходного графа \mathbf{G}) будем обозначать C_i . Предположим, что есть путь из $C_i \in \mathbf{G}'$ в $C_j \in \mathbf{G}'$. Покажем, что в этом случае нет пути из C_j в C_i .

Наличие пути в графе \mathbf{G}' из C_i в C_j означает, что в исходном графе есть путь из вершины $v \in C_i$ в вершину $u \in C_j$. Из предположения о наличии пути из C_j в C_i следует, что в исходном графе есть путь из вершины $u_1 \in C_j$ в вершину $v_1 \in C_i$.

Но пары вершин v, v_1 находятся в одной компоненте сильной связности C_i , то между ними есть путь; аналогично для u, u_1 . В итоге, объединяя пути, получаем, что есть путь от v до u и одновременно путь от u до v . Следовательно, вершины u и v должны принадлежать одной компоненте сильной связности, полученное противоречие доказывает теорему. ■

Для алгоритма построения компонент сильной связности введем следующие понятия.

Определение 2.17. Транспонированием графа $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ называется граф $\mathbf{G}^T = (\mathbf{V}, \mathbf{E}^T)$, в котором $\mathbf{E}^T = \{(u, v) : (v, u) \in \mathbf{E}\}$, т. е. граф, полученный из \mathbf{G} изменением направления каждого ребра на противоположное.

Замечание 2.16. Очевидно, что в транспонированном графе будут те же компоненты сильной связности, что и в исходном графе. Более того, граф конденсации \mathbf{G}^T совпадает с транспонированным графом конденсации исходного графа \mathbf{G} .

Первый шаг алгоритма повторяет алгоритм 2.3 — проходим по всем вершинам графа и из каждой еще не просмотренной вершины u вызываем поиск в глубину $DFS(u)$. При этом для каждой вершины u будем запоминать время выхода из вершины $T_{out}(u)$. Эти времена выхода играют ключевую роль в алгоритме. В начале алгоритма запускается таймер времени и при завершении поиска в глубину из очередной вершины время увеличивается.

Определение 2.18. Время выхода $T_{out}(C)$ из компоненты C сильной связности определим как $\max\{T_{out}(v) \mid v \in C\}$.

Справедлива следующая теорема.

Теорема 2.14. Пусть C и C' — две различные компоненты сильной связности, и в графе конденсации между ними есть дуга (C, C') . Тогда $T_{out}(C) > T_{out}(C')$.

Доказательство. Возможны два случая в зависимости от того, в какую из компонент первой зайдет поиск в глубину.

- Первой была достигнута компонента C . Это означает, что в какой-то момент времени поиск в глубину зашел в некоторую вершину $v \in C$, при этом все остальные вершины компонент C и C' еще не просмотрены. Но, т.к. по условию теоремы в графе конденсаций есть дуга (C, C') , то из вершины v есть путь не только ко всем вершинам *своей* компоненты C ,

но и до всех вершин компоненты C' . Это означает, что при запуске из вершины v обход в глубину пройдет по всем вершинам компонент C и C' , то есть

$$T_{out}(v) > T_{out}(u), \quad \forall u \in C \cup C', u \neq v.$$

- Первой была достигнута компонента C' . Аналогично предыдущему случаю, в какой-то момент времени поиск в глубину зашел в некоторую вершину $v \in C'$, причем все остальные вершины компонент C и C' еще не просмотрены. В силу существования дуги (C, C') и ацикличности графа конденсаций (теорема 2.13), не существует обратного пути из C' в C , т. е. поиск в глубину из вершины v не достигнет вершин компоненты C . Это означает, что они будут просмотрены поиском в глубину позже, и время выхода у них будет больше.

■

Следствие 2.3. Любое ребро (C, C') в графе конденсаций идет из компоненты с большей величиной T_{out} в компоненту с меньшей величиной.

Если мы отсортируем все вершины $v \in \mathbf{V}$ в порядке убывания времени выхода $T_{out}(v)$, то первой окажется некоторая вершина u , принадлежащая *корневой* компоненте сильной связности, т. е. такой, в которую не входит ни одно ребро в графе конденсаций.

Теперь нам нужно попытаться запустить такой поиск в глубину из этой вершины u , который обошел бы только эту компоненту сильной связности и не зашел ни в какую другую. Если мы научимся это делать, мы будем постепенно выделять все компоненты сильной связности: удалив из графа вершины первой выделенной компоненты, мы снова найдем среди оставшихся вершину с наибольшей величиной T_{out} , снова запустим из нее DFS, и т.д.

Чтобы научиться делать такой обход, используем понятие транспонированного графа \mathbf{G}^T . Как уже отмечалось (замечание 2.16) в этом графе будут те же компоненты сильной связности, что и в исходном графе и граф конденсации графа \mathbf{G}^T равен транспонированному графу конденсации исходного графа \mathbf{G} . Это означает, что теперь из рассматриваемой нами *корневой* компоненты уже не будут выходить дуги в другие компоненты.

Таким образом, чтобы обойти всю *корневую* компоненту сильной связности, содержащую некоторую вершину v , достаточно запустить поиск в глубину из вершины v в графе \mathbf{G}^T . Этот обход посетит все вершины этой компоненты сильной связности и только их. Как уже говорилось, дальше мы можем мысленно удалить эти вершины из графа, находить очередную вершину с максимальным значением $T_{out}(u)$ и запускать поиск в глубину на транспонированном графе из нее, и т. д.

Алгоритм 2.5 (поиск компонент сильной связности ориентированного графа).

Шаг 1. Запускаем серию поисков в глубину графа \mathbf{G}

Шаг 2. Создаем список вершин в порядке увеличения времени выхода T_{out} .

Шаг 3. Строим транспонированный граф \mathbf{G}^T .

Шаг 4. Запускаем серию поисков в глубину \mathbf{G}^T в порядке уменьшения времени выхода. Каждое множество вершин, достигнутое в результате очередного запуска обхода, и будет очередной компонентой сильной связности.

Замечание 2.17. Алгоритм 2.5 был предложен независимо Косарайю (Kosaraju) и Шариром (Sharir) в 1979г.

Пример 2.10. Рассмотрим работу алгоритма 2.5 на примере графа на рис. 2.8.

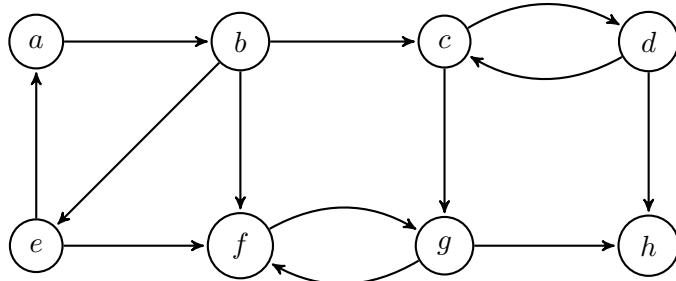


Рис. 2.8

Проводим поиск в глубину для графа 2.8. Протокол работы алгоритма показан в табл. 2.6.

Таблица 2.6

Текущая вершина	Действие	Состояние стека	T	T_{out}
a	Добавляем вершину a в стек	a	1	
b	Добавляем вершину b в стек	ba	2	
c	Добавляем вершину c в стек	cba	3	
d	Добавляем вершину d в стек	$dcba$	4	
h	Добавляем вершину h в стек	$hdcba$	5	
d	Удаляем вершину h из стека	$dcba$	6	$T_{out}(h) = 6$
c	Удаляем вершину d из стека	cba	7	$T_{out}(d) = 7$
g	Добавляем вершину g в стек	$gcba$	8	
f	Добавляем вершину f в стек	$fgcba$	9	
g	Удаляем вершину f из стека	$gcba$	10	$T_{out}(f) = 10$

Продолжение табл. 2.6

Текущая вершина	Действие	Состояние стека	T	T_{out}
c	Удаляем вершину g из стека	$c b a$	11	$T_{out}(g) = 11$
b	Удаляем вершину c из стека	$b a$	12	$T_{out}(c) = 12$
e	Добавляем вершину e в стек	$e b a$	13	
b	Удаляем вершину e из стека	$b a$	14	$T_{out}(e) = 14$
a	Удаляем вершину b из стека	a	15	$T_{out}(b) = 15$
	Удаляем вершину a из стека	Стек пуст	16	$T_{out}(a) = 16$

Запишем последовательность вершин в порядке убывания T_{out}

$$a(16), b(15), e(14), c(12), g(11), f(10), d(7), h(5)$$

Транспонируем граф 2.8 и проводим серию поисков в глубину в порядке уменьшения T_{out} . Имеем

$$\begin{aligned} a \rightarrow e \rightarrow b, \quad C_1 &= \{a, e, b\} \\ c \rightarrow d \quad C_2 &= \{c, d\} \\ g \rightarrow f \quad C_3 &= \{g, f\} \\ h \quad C_4 &= \{h\} \end{aligned}$$

Граф конденсации исходного графа 2.8 представлен на рис 2.9.

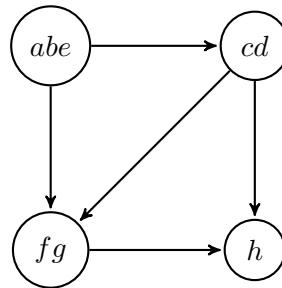


Рис. 2.9

2.1.5. Эйлеров путь в графе

Определение 2.19. *Разомкнутой* (*замкнутой*) *эйлеровой цепью* графа называют простой разомкнутый (*замкнутый*) путь, включающий все ребра графа.

Граф, имеющий замкнутую эйлерову цепь называют *эйлеровым*, а разомкнутую — *полуэйлеровым*.

Теорема 2.15. В графе $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ существует замкнутая эйлерова цепь тогда и только тогда, когда

- a) граф связный;
- б) все его вершины имеют четную степень.

Доказательство. Доказательство достаточности условия теоремы будет следствием анализа алгоритма 2.6 нахождения замкнутой эйлеровой цепи, который будет описан далее. Необходимость условия очевидна, так как появление в замкнутой эйлеровой цепи некоторой вершины k раз означает, что степень этой вершины в графе составляет $2k$. ■

Следствие 2.4. Граф $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ имеет открытую эйлерову цепь тогда и только тогда, когда

- a) \mathbf{G} — связный граф;
- б) существует ровно две вершины нечетной степени.

Доказательство. Очевидно, что нечетную степень имеют концевые вершины графа. Соединив их фиктивным ребром, попадаем в условия теоремы 2.15, следовательно, в новом графе существует замкнутая эйлерова цепь. Удалив фиктивное ребро, получим открытую эйлерову цепь для исходного графа. ■

Как и ранее вершину стека S будем обозначать как $\text{top}(S)$.

Алгоритм 2.6 (нахождение замкнутой эйлеровой цепи).

Исходные данные: связный граф $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ без вершин нечетной степени, представленный списками инцидентности $\text{ЗАПИСЬ}[v], v \in \mathbf{V}$ (см. подразд. 2.1.2)

Результат: эйлеров цикл, представленный последовательностью вершин в списке L

```

 $S \leftarrow v$                                      ▷ записываем в стек  $S$  произвольную вершину  $v$ 
 $L \leftarrow \{\emptyset\}$ 
while  $S \neq \{\emptyset\}$  do
     $v \leftarrow \text{top}(S)$ 
    if  $\text{ЗАПИСЬ}[v] \neq \{\emptyset\}$  then
         $u \leftarrow$  первая вершина списка  $\text{ЗАПИСЬ}[v]$ 
         $S \leftarrow u$                                      ▷ записываем вершину  $u$  в стек  $S$ 
         $\text{ЗАПИСЬ}[v] \leftarrow \text{ЗАПИСЬ}[v] - u$ 
         $\text{ЗАПИСЬ}[u] \leftarrow \text{ЗАПИСЬ}[u] - v$              ▷ удаляем ребро  $(v, u)$  из графа
    else
        Удаляем вершину  $v$  из стека  $S$ 
    end if
     $L \leftarrow v$                                      ▷ записываем вершину  $v$  в выходной список  $L$ 
end while

```

Обоснуем корректность алгоритма 2.6.

Принцип действия алгоритма можно объяснить следующим образом: пусть v_0 — верхний элемент стека, выбранный на шаге S2. Строим путь с началом в v_0 , причем вершины этого пути помещаются в стек S , а ребра удаляются из графа. Эти действия продолжают вплоть до того момента,

когда путь нельзя удлинить, включив в него новую вершину, т. е. когда список ЗАПИСЬ[v] пуст.

Отметим, что тогда должно быть $v = v_0$, так как в любом другом случае это означало бы, что степень вершины v нечетная. Таким образом, из графа был удален цикл, а вершины этого цикла находятся в стеке S .

Заметим, что в графе, модифицированным таким способом, степень произвольной вершины останется четной. Вершина $v = v_0$ переносится теперь из стека S в список L , а очередной вершиной v становится верхний элемент стека S .

Процесс повторяется, начиная с этой вершины (если текущий список ЗАПИСЬ[v] не пуст), в результате чего вследствие четности степени всех вершин находится и помещается в стек S некоторый цикл, проходящий через вершину v . Это продолжается до того момента, пока стек S не станет пустым.

Очевидно, что вершины, помещаемые в стек S , образуют некоторый путь, причем вершина v переносится в список L только тогда, когда список ЗАПИСЬ[v] пуст, т. е. когда все ребра, инцидентные с этой вершиной, представлены (парами соседних вершин) в стеке S или в списке L . Отсюда легко следует, что по окончании работы алгоритма список L содержит эйлеров цикл.

Пример 2.11. Рассмотрим работу алгоритма на примере графа на рис. 2.10 *a*. Будем считать, что в каждом списке инцидентности ЗАПИСЬ[v], $v \in \mathbf{V}$, вершины расположены по возрастанию номеров.

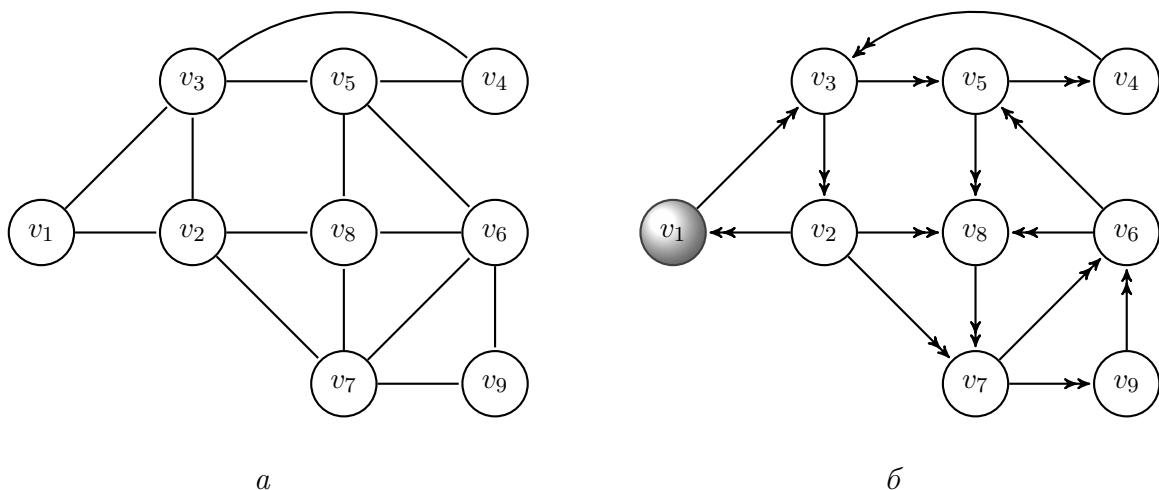


Рис. 2.10

Начнем работу алгоритма с вершины v_1 . Как и ранее будем схематично изображать стек развернутым по горизонтали. Считаем, что вершина стека находится слева. Дальнейшие шаги представлены в табл. 2.7.

Таблица 2.7

Состояние стека S	Состояние списка L	Текущее действие
v_1		Добавляем вершину v_1 в стек S
$v_2 v_1$		Добавляем вершину v_2 в стек S
$v_3 v_2 v_1$		Добавляем вершину v_3 в стек S
$v_1 v_3 v_2 v_1$		Добавляем вершину v_1 в стек S
$v_3 v_2 v_1$	v_1	Переносим вершину v_1 из стека S в список L
$v_4 v_3 v_2 v_1$	v_1	Добавляем вершину v_4 в стек S
$v_5 v_4 v_3 v_2 v_1$	v_1	Добавляем вершину v_5 в стек S
$v_3 v_5 v_4 v_3 v_2 v_1$	v_1	Добавляем вершину v_3 в стек S
$v_5 v_4 v_3 v_2 v_1$	$v_3 v_1$	Переносим вершину v_3 из стека S в список L
$v_6 v_5 v_4 v_3 v_2 v_1$	$v_3 v_1$	Добавляем вершину v_6 в стек S
$v_7 v_6 v_5 v_4 v_3 v_2 v_1$	$v_3 v_1$	Добавляем вершину v_7 в стек S
$v_2 v_7 v_6 v_5 v_4 v_3 v_2 v_1$	$v_3 v_1$	Добавляем вершину v_2 в стек S
$v_8 v_2 v_7 v_6 v_5 v_4 v_3 v_2 v_1$	$v_3 v_1$	Добавляем вершину v_8 в стек S
$v_5 v_8 v_2 v_7 v_6 v_5 v_4 v_3 v_2 v_1$	$v_3 v_1$	Добавляем вершину v_5 в стек S
$v_8 v_2 v_7 v_6 v_5 v_4 v_3 v_2 v_1$	$v_5 v_3 v_1$	Переносим вершину v_5 из стека S в список L
$v_6 v_8 v_2 v_7 v_6 v_5 v_4 v_3 v_2 v_1$	$v_5 v_3 v_1$	Добавляем вершину v_6 в стек S
$v_9 v_6 v_8 v_2 v_7 v_6 v_5 v_4 v_3 v_2 v_1$	$v_5 v_3 v_1$	Добавляем вершину v_9 в стек S
$v_7 v_9 v_6 v_8 v_2 v_7 v_6 v_5 v_4 v_3 v_2 v_1$	$v_5 v_3 v_1$	Добавляем вершину v_7 в стек S
$v_8 v_7 v_9 v_6 v_8 v_2 v_7 v_6 v_5 v_4 v_3 v_2 v_1$	$v_5 v_3 v_1$	Добавляем вершину v_8 в стек S
$v_7 v_9 v_6 v_8 v_2 v_7 v_6 v_5 v_4 v_3 v_2 v_1$	$v_8 v_5 v_3 v_1$	Переносим вершину v_8 из стека S в список L

Далее поочередно все оставшиеся вершины из стека S будут перенесены

в список L . Получившаяся эйлерова цепь:

$$\begin{aligned} v_1 &\rightarrow v_3 \rightarrow v_5 \rightarrow v_8 \rightarrow v_7 \rightarrow v_9 \rightarrow v_6 \rightarrow v_8 \rightarrow v_2 \rightarrow v_7 \rightarrow v_6 \rightarrow v_5 \rightarrow v_4 \rightarrow \\ &\rightarrow v_3 \rightarrow v_2 \rightarrow v_1 \end{aligned}$$

указана направлениями на ребрах на рис. 2.10 б.

Историческая справка. Задача существования эйлерова пути в заданном графе была решена Леонардом Эйлером в 1736 г., и представленное им необходимое и достаточное условие существования такого пути считается первой в истории теоремой теории графов.

На рис. 2.11, а изображена старинная схема мостов города Кенингсберга через реку Прегель. Буквами обозначены части города: А — Альтштадт, В — Кнайпхоф, С — Ломзе, Д — Форштадт. Можно ли прогуляться по городу, пройдя ровно один раз по каждому мосту? Размышления над этой задачей привели Леонарда Эйлера к открытию критерия существования эйлерова пути в графе.

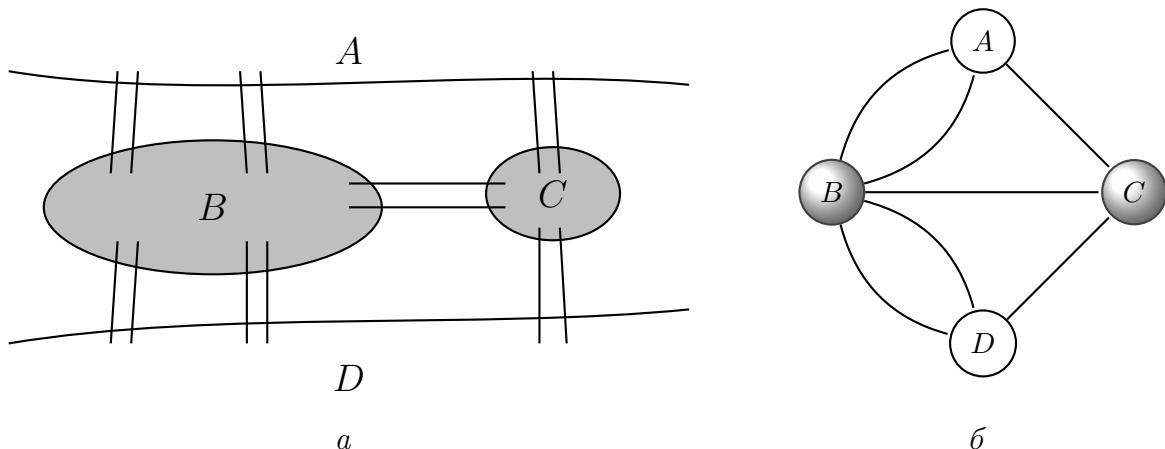


Рис. 2.11

Если каждый берег реки и острова считать вершинами графа, а каждый мост — ребром, то схему можно представить в виде мультиграфа (2-графа), как показано на рис. 2.11, б. Заметим, что теорема 2.15 справедлива и для мультиграфов. Таким образом, ответ в задаче о кенингсбергских мостах отрицательный.

Рассмотрим еще один алгоритм построения замкнутой эйлеровой цепи¹. Представление входных и выходных данных такое же, как в алгоритме 2.6.

Алгоритм 2.7 (алгоритм Флери).

Считаем, что имеется процедура, позволяющая для любого ребра графа определить: является ли это ребро мостом. Например, можно удалить ребро (u, v) из графа и проверить: существует ли в оставшемся графе путь из вершины u в вершину v или воспользоваться алгоритмом 2.4.

$u \leftarrow$ произвольная вершина графа.

$L \leftarrow u$

▷ записываем вершину u в выходной список L

¹Fleury M., "Deux problemes de geometrie de situation Journal de mathematiques elementaires (1883).

```

while E ≠ {∅} do
    W ← {v ∈ ЗАПИСЬ[u] | (u, v) не является мостом}
    if W ≠ {∅} then
        v ← любая вершина из W
    else
        v ← любая вершина из ЗАПИСЬ[u]
    end if
        ▷ ребро мост выбираем только в том случае, когда нет других возможностей.
    ЗАПИСЬ[v] ← ЗАПИСЬ[v] – u
    ЗАПИСЬ[u] ← ЗАПИСЬ[u] – v
                    ▷ удаляем ребро (v, u) из графа
    if ЗАПИСЬ[u] = ∅ then
        удаляем ЗАПИСЬ[u]
                    ▷ удаляем изолированную вершину u из графа
    end if
    L ← v
                    ▷ записываем вершину v в выходной список L
    u ← v
    end while

```

Корректность алгоритма Флери устанавливает следующая теорема.

Теорема 2.16. Алгоритм 2.7 строит замкнутую эйлерову цепь графа при выполнении условий теоремы 2.15.

Доказательство. Покажем сначала, что алгоритм корректно работает на каждом шаге.

Пусть мы достигли некоторой вершины v , начав с вершины $u, v \neq u$. Удалив ребра пути из v в u , видим, что оставшийся граф G_1 связан и содержит ровно две вершины нечетной степени v и u . Согласно следствию 2.4 из теоремы 2.15 граф G_1 имеет открытую эйлерову цепь P из v в u .

Поскольку удаление первого ребра инцидентного u цепи P либо не нарушает связности графа G_1 , либо происходит удаление изолированной вершины u и оставшийся граф G_2 связан с двумя нечетными вершинами, то отсюда получаем, что описанное в алгоритме 2.7 действие всегда возможно на каждом шаге. (Если $v = u$, то доказательство не меняется, если имеются ребра, инцидентные u).

Покажем, что данная последовательность шагов приводит к эйлерову пути. Действительно, в G не может быть ребер, оставшихся не проходимыми после использования последнего ребра, инцидентного u , поскольку в противном случае удаление ребра, смежного одному из оставшихся, привело бы к несвязному графу, что противоречит условию. ■

Пример 2.12. Рассмотрим работу алгоритма 2.7 для графа из примера 2.11 (см. рис. 2.10). Начнем работу с вершины v_1 . При этом условимся выбирать из возможных вершину с наименьшим номером. Протокол работы алгоритма запишем в виде таблицы (табл. 2.8).

Таблица 2.8

Выбираемая вершина	Комментарий
v_2	
v_3	
v_4	Вершину v_1 выбрать нельзя, так как ребро (v_3, v_1) является мостом
v_5	Ребро (v_4, v_5) является мостом, но других ребер из вершины v_4 уже нет. Удаляем изолированную вершину v_4
v_6	Вершину v_3 выбрать нельзя, так как ребро (v_5, v_3) является мостом
v_7	
v_2	
v_8	Ребро (v_2, v_8) является мостом, но других ребер из вершины v_2 уже нет. Удаляем изолированную вершину v_2
v_6	Вершину v_5 выбрать нельзя, так как ребро (v_8, v_5) является мостом. Далее все оставшиеся ребра являются мостами
v_9	Удаляем изолированную вершину v_6
v_7	Удаляем изолированную вершину v_9
v_8	Удаляем изолированную вершину v_7
v_5	Удаляем изолированную вершину v_8
v_3	Удаляем изолированную вершину v_5
v_1	Удаляем изолированную вершину v_3

Таким образом получаем тот же эйлеров цикл, что и в примере 2.11.

Пример 2.13. Классическая задача на построение эйлерова цикла — задача о домино.

Имеется N костяшек домино, на двух концах каждой костяшки записано по одному числу (от 1 до n). Требуется выложить все домино в ряд так, чтобы у любых двух соседних домино числа, записанные на их общей стороне, совпадали. Домино разрешается переворачивать.

Переформулируем задачу. Пусть числа, записанные на домино, — вершины графа, а сами костяшки — ребра этого графа (каждая домино с числами (a, b) — это ребра (a, b) и (b, a)). Тогда задача сводится к нахождению эйлерова пути в этом графе.

Для ориентированных графов понятие эйлеровой цепи вводится аналогично.

Определение 2.20. Разомкнутой (замкнутой) эйлеровой цепью ориентированного графа называют разомкнутый (замкнутый) простой ориентированный путь, содержащий все дуги графа.

Справедлива также теорема, аналогичная теореме 2.15.

Теорема 2.17. Ориентированный граф $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ имеет замкнутую эйлерову цепь тогда и только тогда, когда он слабосвязный и

$$|\delta^+(v)| = |\delta^-(v)|, \forall v \in \mathbf{V}.$$

Следствие 2.5. Ориентированный граф $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ имеет разомкнутую эйлерову цепь тогда и только тогда, когда он слабосвязный и выполнены условия:

- 1) существует и притом единственный узел $v_0 \in \mathbf{V}$:

$$|\delta^+(v_0)| = |\delta^-(v_0)| + 1;$$

- 2) существует и притом единственный узел $v_1 \in \mathbf{V}$:

$$|\delta^-(v_1)| = |\delta^+(v_1)| + 1;$$

- 3) для всех остальных узлов графа выполнено $|\delta^+(v)| = |\delta^-(v)|$.

Доказательство. Добавляем фиктивную дугу $e = [v_1, v_0]$ и применяем теорему 2.17. ■

Определение 2.21. Если граф имеет простой цикл, содержащий все вершины графа по одному разу, то такой цикл называется *гамильтоновым циклом*, а граф называется *гамильтоновым графом*. Граф, который содержит простой путь, проходящий через каждую его вершину, называется *полугамильтоновым*. Это определение можно распространить на ориентированные графы, если путь считать ориентированным.

Историческая справка. Слово *гамильтонов* в этом определении связано с именем известного ирландского математика У. Гамильтона, которым в 1859 году предложена следующая игра *Кругосветное путешествие*. Каждой из 20 вершин додекаэдра приписано название одного из крупных городов мира. Требуется, переходя от одного города к другому по ребрам додекаэдра, посетить каждый город в точности один раз и вернуться в исходный город.

Гамильтонов цикл не обязательно содержит все ребра графа. Ясно, что гамильтоновым может быть только связный граф и, что всякий гамильтонов граф является полугамильтоновым. Заметим, что гамильтонов цикл существует далеко не в каждом графе. Для гамильтоновых циклов (и путей) неизвестно никаких просто проверяемых необходимых и достаточных условий их существования, а все известные алгоритмы требуют для некоторых графов перебора большого числа вариантов. В отличии от эйлеровых графов, где имеется необходимый и достаточный критерий для графа быть эйлеровым, для гамильтоновых графов такого критерия нет. Большинство известных теорем имеет вид: *если граф G имеет достаточное количество ребер, то граф является гамильтоновым*. Приведем несколько таких теорем.

Теорема 2.18 (Оре). Если для любой пары u и v несмежных вершин графа G порядка $n \geq 3$ выполняется неравенство $\deg u + \deg v \geq n$, то G – гамильтонов граф.

Теорема 2.19 (Дирак). Если $\|G\| = n \geq 3$ и для любой вершины v графа G выполняется неравенство $\deg v \geq n/2$, то G – гамильтонов граф.

Замечание 2.18. Такой граф называется *графом Дирака*.

Пример 2.14. В графе, изображенном на рис. 2.12 *a*, гамильтоновым циклом является, например, последовательность вершин $v_1, v_2, v_3, v_5, v_4, v_1$. В графе рис. 2.12 *б*, нет гамильтоновых циклов, но есть гамильтоновы пути, например, v_2, v_1, v_3, v_5, v_4 . В графе на рис. 2.12 *в* нет и гамильтоновых путей.

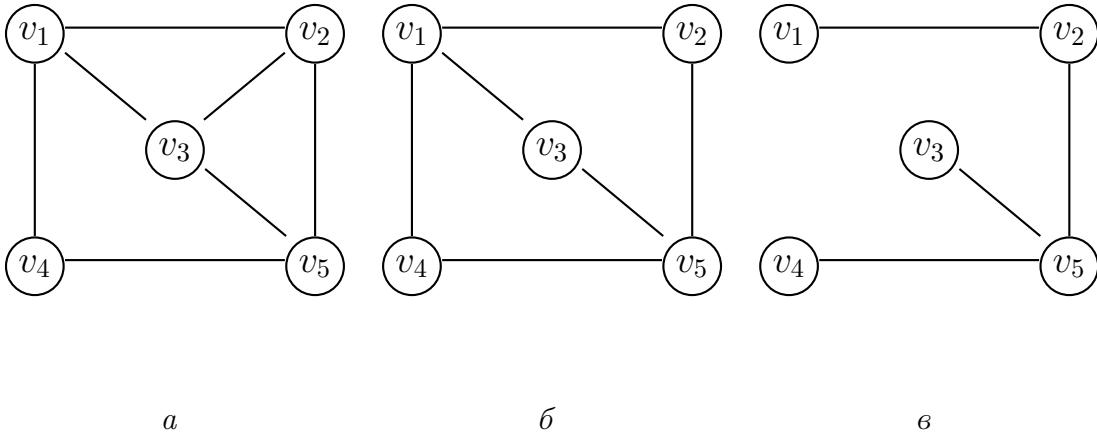


Рис. 2.12

2.1.6. Графы де Брюина

Рассмотрим несколько интересных и важных применений эйлеровых путей. Дадим необходимые определения.

Определение 2.22. Конечное множество символов $\mathcal{A} = \{\alpha_1, \dots, \alpha_n\}$ будем называть *алфавитом*, символы α_i – *буквами*, последовательность символов – *словом*. Число символов в слове w называют *длиной* слова и обозначают $|w|$. Иногда слова называют *цепочками*.

Множество слов алфавита \mathcal{A} обозначают \mathcal{A}^* .

Определение 2.23. Пусть задан алфавит \mathcal{A} , состоящий из n букв.

Графом де Брюина (обозначается $B(n, k)$) для алфавита \mathcal{A} называется ориентированный граф $\mathbf{G} = \mathbf{G}(\mathbf{V}, \mathbf{E})$. Его узлы \mathbf{V} есть множество слов длины k из \mathcal{A}^* . Дуга $[u, v] \in \mathbf{E}$ из узла $u = \{x_1 \dots x_k\}$ в узел $v = \{y_1 \dots y_k\}$ проводится, если $x_i = y_{i-1}$, $i \in 2 : k$.

Дуге $[u, v]$ сопоставляют слово w длины $k + 1$:

$$w = \{x_1 \dots x_k y_k\} = \{x_1 y_1 \dots y_k\}.$$

Узел u называют *префиксом*, а узел v — *суффиксом* слова w , символически это можно записать, как:

$$w = \{u\} + y_n = x_1 + \{v\}. \quad (2.1)$$

Очевидно, что разным дугам $[u_1, v_1]$ и $[u_2, v_2]$ соответствует два различных слова w_1 и w_2 . То есть кратных дуг в графе де Брюина нет, но очевидным образом есть петли.

Замечание 2.19. Графы названы по имени голландского математика Николаса де Брюина¹. Иногда, в литературе встречаются транскрипции «де Брайна» или «де Брёйна».

Рассмотрим некоторые свойства графов де Брюина.

Лемма 2.1. Граф $B(n, k)$ является эйлеровым графом.

Доказательство. В соответствии с теоремой 2.17 нужно доказать, что для каждого узла число входящих в него ребер равно числу исходящих. Покажем, что

$$|\delta^+(v)| = n = |\delta^-(v)|, \forall v \in \mathbf{V}. \quad (2.2)$$

Согласно (2.1) можно выбрать ровно n символов из \mathcal{A} для исходящих из узла u дуг, и столько же для входящих. Равенство (2.2) установлено. ■

Лемма 2.2. Число узлов графа $B(n, k)$ равно n^k , число дуг, соответственно, n^{k+1} .

Доказательство. Первое утверждение очевидно из определения графа де Брюина. Из теоремы 2.6 о рукопожатиях для ориентированных графов и равенства (2.2) получаем второе утверждение. ■

Пусть у нас есть алфавит $\mathcal{A} = \{\alpha_1, \dots, \alpha_n\}$ и задано число k . Рассмотрим алгоритм построения соответствующего графа де Брюина.

Алгоритм 2.8 (построения графа де Брюина).

Шаг 1. Генерируем в любом порядке, например, в лексикографическом по алгоритму 1.27, все слова длины $k + 1$ из \mathcal{A} .

Шаг 2. Для каждого слова $\{x_1 \dots x_n x_{n+1}\}$ определяем пару узлов u, v и дугу $[u, v]$ графа:

$$u = \{x_1 \dots x_n\}, v = \{x_2 \dots x_{n+1}\}.$$

¹de Bruijn N. G. A combinatorial problem // Koninklijke Nederlandse Akademie v. Wetenschappen. 1946. — v. 49. — pp. 758—764.

Пример 2.15. Пусть $\mathcal{A} = \{a, b\}$ и $k = 2$. Протокол работы алгоритма 2.8 приведен в таблице 2.9.

Таблица 2.9

Набор	Узлы	Набор	Узлы
$\{aaa\}$	$\{aa\} \rightarrow \{aa\}$	$\{baa\}$	$\{ba\} \rightarrow \{aa\}$
$\{aab\}$	$\{aa\} \rightarrow \{ab\}$	$\{bab\}$	$\{ba\} \rightarrow \{ab\}$
$\{aba\}$	$\{ab\} \rightarrow \{ba\}$	$\{bba\}$	$\{bb\} \rightarrow \{ba\}$
$\{abb\}$	$\{ab\} \rightarrow \{bb\}$	$\{bbb\}$	$\{bb\} \rightarrow \{bb\}$

Получившийся граф представлен на рисунке 2.13.

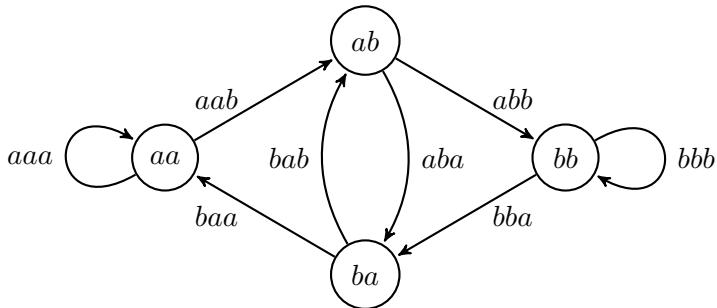


Рис. 2.13

Рассмотрим несколько применений графов де Брюина².

Пример 2.16. Пусть задан алфавит $\mathcal{A} = \{\alpha_1, \dots, \alpha_n\}$. Неизвестное слово-пароль имеет вид $\{x_1 \dots x_k\} \in \mathcal{A}^*$. Найти самое короткое слово s , которое гарантированно содержит пароль как подслово.

Построим граф де Брюина $B(n, k - 1)$. Напомним, что узлы графа – слова длины $k - 1$ из \mathcal{A}^* . В этом графе найдем эйлеров цикл. Он существует, согласно лемме 2.1.

Изначально, слово s пусто, $|s| = 0$. Возьмем произвольный узел эйлерова цикла v и добавим его слово ($k - 1$ символ) в s . Далее, переходя по дугам к следующим узлам, будем добавлять в s последнюю букву слова узла.

Согласно лемме 2.2 в графе n^k дуг, поэтому длина искомой последовательности $|s| = n^k + k - 1$.

Установим корректность решения. Очевидно, что слово меньшей длины составить нельзя: слово s содержит n^k подстрок длины k , именно столько слов такой длины можно составить из символов \mathcal{A} . Поскольку разным

²Примеры взяты с сайта <http://neerc.ifmo.ru/wiki>

дугам соответствуют разные слова длины k , получаем корректность построения.

Рассмотрим задачу в условиях примера 2.15. Построим эйлеров цикл (с учетом петель) для графа на рисунке 2.13:

$$aa \rightarrow aa \rightarrow ab \rightarrow ba \rightarrow ab \rightarrow bb \rightarrow bb \rightarrow ba \rightarrow aa$$

Тогда искомое слово:

$$s = aa + a + b + a + b + b + a + a = a a a b a b b b a a$$

Рассмотрим задачу построения генома (в упрощенном варианте). Дадим соответствующие определения.

Определение 2.24. Алфавит состоит из 4 символов $\mathcal{A} = \{A, G, C, T\}$, символы так же называются *нуклеотидами*.

Геном и его части – *риды* представляют собой слова из алфавита \mathcal{A}^* .

Пример 2.17. Задано множество R неповторяющихся ридов длины k . Известно, что все риды генома длины k входят в данное множество. Построить возможный геном.

Используем результаты предыдущего примера 2.16.

Построим граф \mathbf{G} , узлами которого будут суффиксы и префиксы длины $k - 1$ всех ридов из R . Получим подграф графа де Брюина $B(4, k - 1)$, в котором каждой дуге соответствует рид из R . Это подграф, так как в нем не обязательно есть все 4^{k-1} узлов.

Построим в \mathbf{G} эйлеров путь. Он существует, так как на геном было наложено условие о том, что все его подстроки длины k входят в множество ридов R . Этот путь и будет *возможным* ответом.

Возможным, потому, что единственno верный ответ (реальный геном) получить можно не всегда, так как не всегда в графе есть единственный эйлеров путь.

Пусть $k = 3$ и $R = \{ACT, CTG, TGA, GAC, ACG, CGA\}$. Построим возможный геном.

Строим граф \mathbf{G} подграф графа де Брюина $B(4, 2)$ (рис. 2.14)

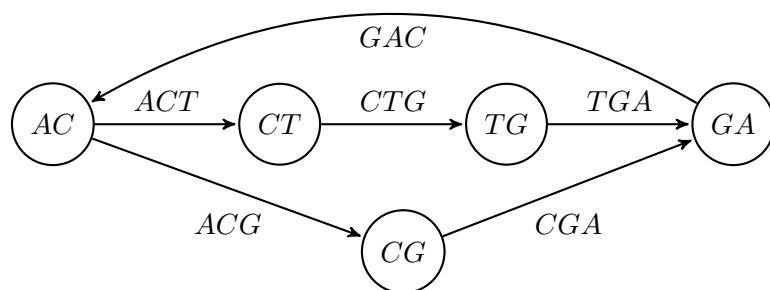


Рис. 2.14

Строим в графе \mathbf{G} эйлеров путь. Условия следствия 2.5 к теореме 2.17 выполнены:

$$\begin{aligned} |\delta^+(AC)| &= 2 = |\delta^-(AC)| + 1, \quad |\delta^-(GA)| = |\delta^+(GA)| + 1, \\ |\delta^+(v)| &= |\delta^-(v)| = 1, \quad v = CT, TG, CG. \end{aligned}$$

Тогда эйлеров путь из узла AC в узел GA имеет вид:

$$AC \rightarrow CT \rightarrow TG \rightarrow GA \rightarrow AC \rightarrow CG \rightarrow GA$$

Искомый геном: $AC + T + G + A + C + G + A = ACTGACGA$.

Замечание 2.20. Задача построения генома в примере 2.17 существенно упрощена.

1. Риды не обязательно будут неповторяющимися.
2. Данные о ридах могут быть получены лишь с некоторой вероятностью.
3. Риды не имеют фиксированной длины в силу особенностей метода их определения (*секвенирования*).

Тем не менее, рассмотренный подход используется при сборке генома или его больших ридов, но с заметными усложнениями.

2.1.7. Деревья

Определение 2.25. Связный граф, не содержащий циклов, называют *деревом*. Произвольный граф, не содержащий циклов, называют *ациклическим*, или *лесом*.

Теорема 2.20. Граф является деревом тогда и только тогда, когда между любыми двумя его различными вершинами существует одна и только одна цепь.

Доказательство. Установим необходимость. Граф \mathbf{G} является деревом, следовательно, он связный и любые две его вершины можно соединить цепью. Поскольку граф \mathbf{G} не содержит циклов, по следствию 2.2 к теореме 2.5 такая цепь только одна.

Для доказательства достаточности нужно провести те же рассуждения в обратном порядке. ■

Теорема 2.21. Дерево, содержащее не менее двух вершин, имеет по крайней мере две концевые вершины.

Доказательство. Пусть $v_0, e_1, v_1, \dots, e_l, v_l$ — цепь максимальной длины в графе. Покажем, что $\deg(v_0) = \deg(v_l) = 1$. Пусть существует ребро $e_{l+1} \neq e_l$, также инцидентное v_l , и v_{l+1} — его вторая концевая точка. Но в дереве нет циклов, тогда $v_{l+1} \neq v_i$ ($i = 0, 1, \dots$), следовательно, $v_0, e_1, v_1, \dots, e_l, v_l, e_{l+1}, v_{l+1}$ — цепь. Получаем противоречие максимальности длины цепи. Аналогичное доказательство можно провести для вершины v_0 . ■

Лемма 2.3. После удаления из дерева концевой вершины вместе с инцидентным ей ребром получается вновь дерево.

Доказательство. Пусть v — концевая вершина, а e — инцидентное ей ребро в дереве $\mathbf{G} = (\mathbf{V}, \mathbf{E})$. Рассмотрим подграф $\mathbf{G}' = (\mathbf{V} \setminus \{v\}, \mathbf{E} \setminus \{e\})$. Очевидно, что граф \mathbf{G}' не содержит циклов. Любую пару вершин графа \mathbf{G}' в исходном графе \mathbf{G} можно соединить цепью. Очевидно, что эта цепь не содержит ни вершину v , ни ребро e . Таким образом, граф \mathbf{G}' — дерево. ■

Теорема 2.22. Дерево с p вершинами имеет $p - 1$ ребро.

Доказательство. Проводим индукцию по числу вершин p . При $p = 1$ утверждение очевидно. Пусть теорема верна для всех $p \leq k$. По теореме 2.21 дерево с $k + 1 \geq 2$ вершинами имеет концевую вершину. Исключая ее вместе с инцидентным ей ребром, по лемме 2.3 получаем дерево с k вершинами, которое (по индукционному предположению) имеет $k - 1$ ребро. ■

Определение 2.26. Пусть $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ — связный граф. Дерево, являющееся подграфом \mathbf{G} и содержащее все его вершины, называют деревом, *покрывающим* граф \mathbf{G} (стягивающим деревом, каркасом, оством, оством деревом).

Теорема 2.23. В связном графе \mathbf{G} всегда существует по крайней мере одно стягивающее его дерево.

Доказательство. Если в графе \mathbf{G} нет циклов, то теорема доказана. В противном случае исключаем любое ребро e , принадлежащее циклу. По теореме 2.10 получившийся граф $\mathbf{G} \setminus \{e\}$ связный. Продолжаем эту процедуру до тех пор, пока в графе есть циклы. ■

Следствие 2.6. Пусть \mathbf{F} — подмножество ребер графа $\mathbf{G} = (\mathbf{V}, \mathbf{E})$, обладающее следующим свойством: для любого цикла графа \mathbf{G} все ребра этого цикла не лежат полностью в \mathbf{F} . Тогда существует стягивающее дерево $\mathbf{G}' = (\mathbf{V}, \mathbf{E}') : \mathbf{F} \subseteq \mathbf{E}'$.

Доказательство. При доказательстве теоремы 2.23 всегда можно исключать ребра, не лежащие в \mathbf{F} , поэтому ребра \mathbf{F} останутся в стягивающем дереве. ■

Следствие 2.7. Связный граф с p вершинами и $p - 1$ ребрами является деревом.

Доказательство. Если бы граф имел цикл, то при построении его стягивающего дерева получилось бы дерево с p вершинами, а число ребер (после исключения цикла) было бы меньше, чем $(p - 1)$. Противоречие с теоремой 2.22. ■

Следствие 2.8. Если связный граф имеет p вершин и q ребер, то

$$q - p + 1 \geq 0 \quad (q \geq p - 1).$$

Доказательство. В процессе построения стягивающего дерева (при доказательстве теоремы 2.23) при исключении m циклов ($m \geq 0$) удаляется m ребер. В получившемся стягивающем дереве остается $p - 1$ ребро. Таким образом, исходный граф имеет $q = p - 1 + m \geq p - 1$ ребер. ■

Алгоритм 2.9 (построение стягивающего дерева).

Воспользуемся алгоритмом поиска в глубину или в ширину. Добавим к нему одну операцию. При нахождении новой *непросмотренной* вершины u из текущей вершины v включаем в дерево ребро (v, u) .

Замечание 2.21. В силу применяемых алгоритмов просматриваются все вершины графа. В полученном графе нет циклов, так как последнее ребро, замыкающее цикл, должно было бы соединить уже просмотренные вершины.

Пример 2.18. Алгоритмом 2.9 построим стягивающее дерево для графа на рис. 2.15, а.

Для построения применим алгоритм 2.1 поиска в глубину. Начало работы – с вершины v_1 .

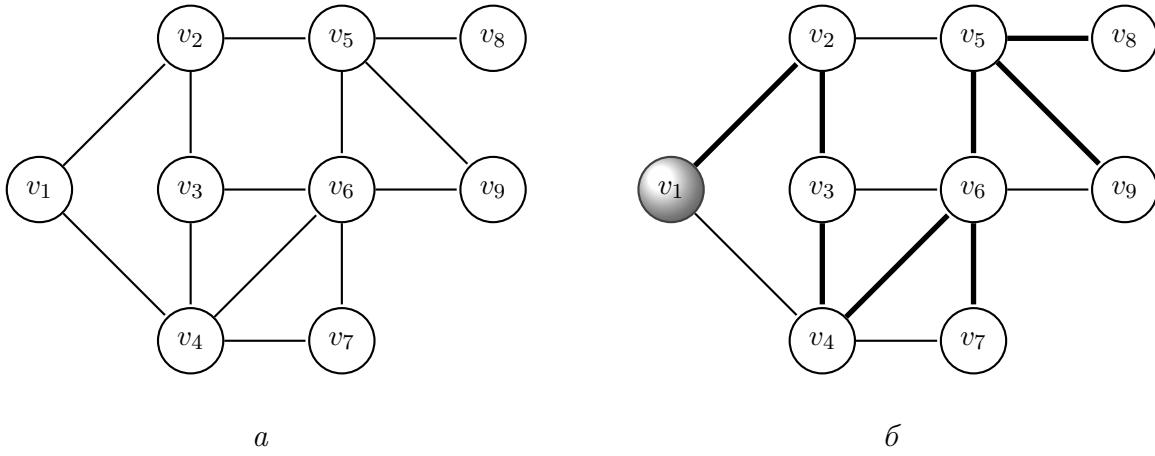


Рис. 2.15

Протокол работы алгоритма запишем в виде таблицы (табл. 2.10).

Удаляем из стека оставшиеся вершины $v_7, v_6, v_4, v_3, v_2, v_1$. Стек пуст, алгоритм заканчивает работу. Ребра, вошедшие в стягивающее дерево, имеют большую «толщину» на рис. 2.15, б.

Упражнение 2.4. Постройте стягивающее дерево для графа из примера 2.18 (см. рис. 2.15, а), используя алгоритм 2.2 поиска в ширину.

Упражнение 2.5. Предложенный алгоритм дает одно из существующих стягивающих деревьев графа. Модифицируйте алгоритм 2.9 для построения всех стягивающих деревьев данного графа.

Таблица 2.10

Состояние стека	Добавляемое ребро	Комментарий
$\{v_1\}$		Добавили в стек вершину v_1
$\{v_2 v_1\}$	(v_1, v_2)	Добавили в стек вершину v_2
$\{v_3 v_2 v_1\}$	(v_2, v_3)	Добавили в стек вершину v_3
$\{v_4 v_3 v_2 v_1\}$	(v_3, v_4)	Добавили в стек вершину v_4
$\{v_6 v_4 v_3 v_2 v_1\}$	(v_4, v_6)	Добавили в стек вершину v_6
$\{v_5 v_6 v_4 v_3 v_2 v_1\}$	(v_6, v_5)	Добавили в стек вершину v_5
$\{v_8 v_5 v_6 v_4 v_3 v_2 v_1\}$	(v_5, v_8)	Добавили в стек вершину v_8
$\{v_5, v_6, v_4 v_3 v_2 v_1\}$		Удалили из стека вершину v_8
$\{v_9 v_5 v_6 v_4 v_3 v_2 v_1\}$	(v_5, v_9)	Добавили в стек вершину v_9
$\{v_5 v_6, v_4 v_3 v_2 v_1\}$		Удалили из стека вершину v_9
$\{v_6 v_4 v_3 v_2 v_1\}$		Удалили из стека вершины v_5
$\{v_7 v_6 v_4 v_3 v_2 v_1\}$	(v_6, v_7)	Добавили в стек вершину v_7

2.1.8. Корневые деревья

Сначал введем понятие корневого дерева для неориентированных графов.

Определение 2.27. *Корневым деревом*¹ будем называть дерево с выделенной вершиной – корнем. Висячая вершина корневого дерева, не являющаяся корнем, называется *листом*. *Высота* корневого дерева – это расстояние от корня до самого удаленного листа.

Если в корневом дереве T путь, соединяющий вершину u с корнем, проходит через вершину v , то говорят, что вершина v – *предок* вершины u , а соответственно вершина u – *потомок* вершины v . В частности, каждая вершина является предком и потомком самой себя. Множество всех предков вершины v порождает путь из корня в саму вершину v . Множество всех потомков вершины v порождает дерево с корнем v , оно называется *ветвью* дерева T в вершине v .

Если предок и потомок соединены ребром, то они называются соответственно *отцом* и *сыном*.

Для ориентированных графов понятие дерева вводится аналогичным образом.

Определение 2.28. *Ориентированным деревом* называют ориентированный граф без циклов, в котором в каждый узел, кроме одного, называемого *корнем ориентированного дерева*, входит одна дуга. В корень

¹Иногда, термин *корневое дерево* заменяют на *корневое представление дерева*.

ориентированного дерева не входит ни одной дуги (его отрицательная степень равна 0). При этом для любого узла существует ориентированный путь в него из корня. Иногда, если это не приводит к неоднозначности, ориентированное дерево называют просто деревом.

Любое неориентированное дерево можно представить в виде корневого. Это представление основано на следующей лемме.

Лемма 2.4 (о построении корневого дерева). Множество вершин произвольного дерева $\mathbf{T} = (\mathbf{V}, \mathbf{E})$ можно разбить в объединение непустых попарно непересекающихся подмножеств V_0, \dots, V_k так, что будут выполнены следующие условия:

- 1) множество V_0 состоит из одной вершины;
- 2) для любого $i \in 1 : k$ никакие вершины из V_i не смежны;
- 3) для любого $i \in 1 : k$ любая вершина из V_i смежна ровно с одной вершиной из V_{i-1} и не смежна ни с одной вершиной из V_{i-2}, \dots, V_0 .

Доказательство. Рассмотрим $v_0 \in \mathbf{V}$ – произвольную вершину дерева \mathbf{T} . Обозначим $\rho(v_0, v)$ – расстояние между вершинами v_0 и v . Положим

$$m = \max \{\rho(v_0, v) \mid v \in \mathbf{V}\}$$

и рассмотрим подмножества вершин $V_i \subset \mathbf{V}$

$$V_i = \{v \in \mathbf{V} \mid \rho(v_0, v) = i\}, \quad i = 0, \dots, m.$$

Очевидно, что никакие из множеств V_i не пересекаются, и $V_0 = \{v_0\}$. Далее, каждая вершина из множества V_i имеет смежную вершину во множестве V_{i-1} . Действительно, если $\rho(v_0, v) = i$, то расстояние между v_0 и предпоследней вершиной кратчайшей $(v_0 - v)$ -цепи равно $i - 1$. Следовательно, все множества V_i непусты.

Пусть $v \in V_i$. Тогда все смежные с v вершины принадлежат одному из трех множеств V_{i-1}, V_i, V_{i+1} , поскольку расстояния от вершины v до двух смежных вершин не могут отличаться более, чем на единицу.

Осталось показать, что у v нет смежной вершины в V_i и нет двух смежных вершин в V_{i-1} . На рис. 2.16 изображены кратчайшие цепи из рассматриваемых вершин до вершины v_0 для первого случая, а на рис. 2.17 – соответственно для второго.

В первом случае, если вершина v имеет смежную вершину $u \in V_i$, то ребро (v, u) очевидно замыкает цикл в дереве \mathbf{T} .

Для второго случая, при наличии у вершины v двух смежных вершин $v_1, v_2 \in V_{i-1}$, ребра (v_1, v) и (v, v_2) замыкают цикл в дереве \mathbf{T} . Полученные в обоих случаях противоречия завершают доказательство. ■

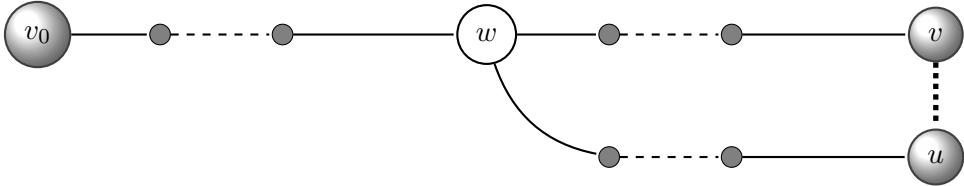


Рис. 2.16

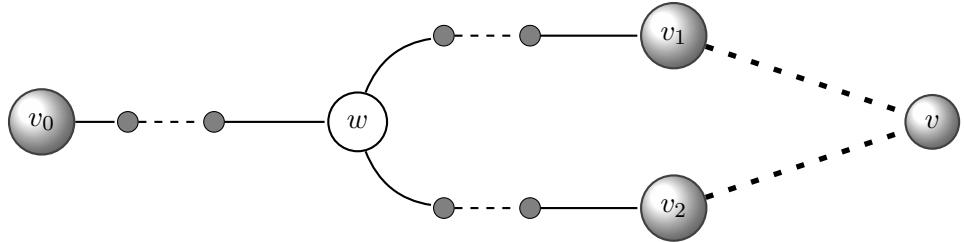


Рис. 2.17

Рассмотрим алгоритм преобразования произвольного дерева в корневое.

Пусть $\mathbf{T} = (\mathbf{V}, \mathbf{E})$ – дерево, а V_0, \dots, V_k – множества вершин, построенные в лемме 2.4, будем считать, что $V_0 = \{v_0\}$.

Помещаем вершину v_0 вверху, вершины из множества V_1 – на одном уровне и ниже, чем v_0 , вершины из множества V_2 – на одном уровне и ниже, чем вершины из V_1 , и т. д. При этом, если на i -м уровне слева направо изображены вершины v_1, v_2, \dots , то на $(i+1)$ -м уровне слева направо изображаются сначала вершины, смежные v_1 , затем смежные v_2 , и т. д. При таком построении ребра не будут пересекаться. Полученный граф, и есть корневое изображение дерева \mathbf{T} . Очевидно, что вершина v_0 есть корень построенного дерева.

Замечание 2.22. Одно и то же дерево имеет много корневых изображений, в зависимости от выбора корня (вершины v_0 в рассмотренном алгоритме).

Пример 2.19. Построим корневое дерево для каркаса из примера 2.18, изображенного на рис. 2.18, а.

В качестве корневой вершины выберем v_4 . Получаем следующую последовательность разбиений:

$$V_0 = \{v_4\}, V_1 = \{v_3, v_6\}, V_2 = \{v_2, v_5, v_7\}, V_3 = \{v_1, v_8, v_9\}.$$

На рис. 2.18, б приведено построенное корневое изображение.

В настоящее время корневые деревья применяются очень широко. Например, это деревья вариантов в алгоритмах перебора с возвратом в разделе 2.2.11. При оценке шахматной позиции (см. пример 2.4) строится дерево, корнем которого является текущая позиция в игре, сыновьями корня – позиции, возникающие после различных ее ходов, сыновьями сыновей – позиции после ответов противника на эти ходы, и т. д. На основе заранее

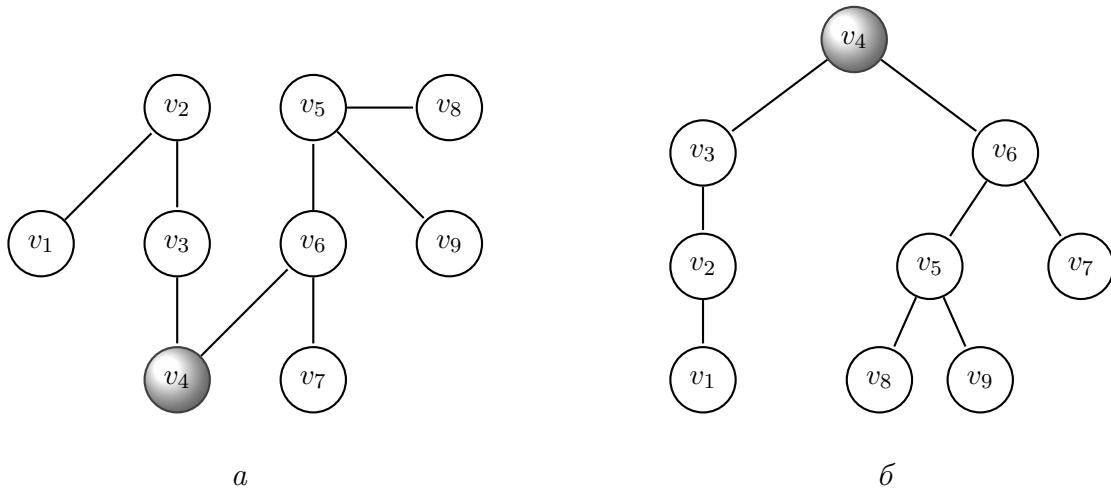


Рис. 2.18

определенной функции оценки позиции отсекаются определенные части дерева и выбирается лучший ход.

2.1.9. Код Прюфера

Рассмотрим задачу хранения деревьев. Для кодирования структуры дерева можно использовать *код Прюфера*¹. Код является оптимальным по объему. Дадим определение.

Определение 2.29. Кодом Прюфера длины $n - 2$ называется последовательность из чисел от 1 до n с повторениями.

Теорема 2.24. Существует взаимно однозначное соответствие между стягивающими деревьями для графа из n вершин и кодами Прюфера длины $n - 2$. По каждому дереву с n вершинами можно построить код Прюфера длины $n - 2$ и наоборот.

Следствие 2.9 (формула Кэли). Количество пронумерованных деревьев из n вершин равно n^{n-2} .

В качестве доказательства теоремы 2.24 приведем алгоритмы кодирования и восстановления дерева по коду Прюфера.

Алгоритм 2.10 (построение кода Прюфера).

Исходные данные: Т — дерево с множеством вершин $\{v_1, \dots, v_n\}$. Считаем, что номер вершины v_k равен k .

Результат: последовательность кода Прюфера $P = \{p_1, \dots, p_{n-2}\}$

for $i \leftarrow 1, n - 2$ **do**

¹Код первоначально был предложен Хейнцем Прюфером в 1918 г. как доказательство формулы Кэли (см. следствие 2.9 к теореме 2.24).

```

 $v_k \leftarrow$  висячая вершина с минимальным номером
 $p_i \leftarrow s =$  номер вершины  $v_s$  — единственного соседа вершины  $v_k$ 
Удаляем из дерева висячую вершину  $v_k$ 
end for

```

Пример 2.20. С помощью алгоритма 2.10 построим код Прюфера для графа на рис. 2.19. Протокол работы запишем в таблице 2.11

Таблица 2.11

i	Вершина	p_i	i	Вершина	p_i
1	v_4	1	5	v_8	1
2	v_7	5	6	v_1	6
3	v_5	2	7	v_9	6
4	v_2	6	8	v_6	3

$$P = \{1, 5, 2, 6, 1, 6, 6, 3\}.$$

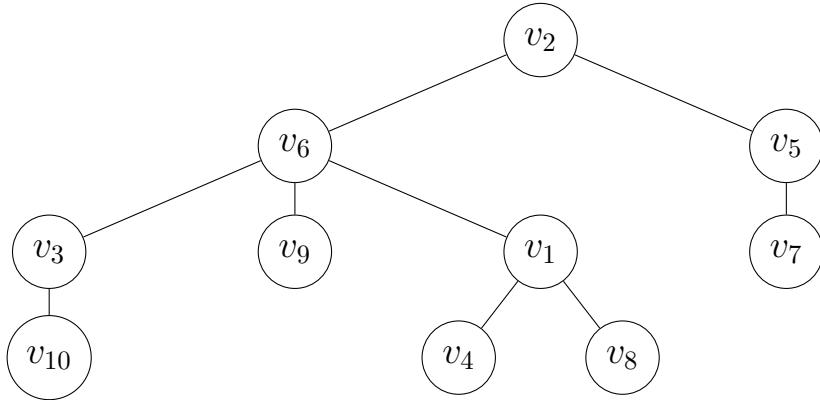


Рис. 2.19

Замечание 2.23. После завершения алгоритма 2.10 в дереве останутся неудаленными две вершины. Одной из них будет вершина с максимальным номером.

Каждая вершина встречается в коде Прюфера определенное число раз, равное ее степени минус один. Это легко понять, если заметить, что вершина удаляется из дерева в момент, когда ее степень равна единице — т. е. к этому моменту все смежные с ней ребра, кроме одного, были удалены. (Для двух оставшихся после построения кода вершин это утверждение тоже верно.)

Рассмотрим алгоритм для решения обратной задачи.

Алгоритм 2.11 (построения дерева по коду Прюфера).

Исходные данные: код Прюфера $P = \{p_1, \dots, p_{n-2}\}$, $L = \{1, \dots, n\}$

Результат: дерево T с n вершинами, пронумерованными от 1 до n

▷ Как и в алгоритме 2.10 считаем, что номер вершины v_k равен k

```

for  $i \leftarrow 1, |P| = n - 2$  do
     $k \leftarrow \min \{j \mid j \in L, j \notin P\}$ 
    Соединяем вершины  $v_k$  и  $v_{p_i}$ 
    Удаляем элемент  $k$  из  $L$ .
    Удаляем элемент  $p_i$  из  $P$ ,  $P = \{p_{i+1}, p_{i+2}, \dots, p_{n-2}\}$ .
end for
Соединяем две оставшиеся вершины в  $\mathbf{T}$ .

```

Пример 2.21. С помощью алгоритма 2.11 построим дерево по коду Прюфера $P = \{3, 3, 4, 5, 4, 6\}$. Протокол работы запишем в таблице 2.12

Таблица 2.12

i	P	L	k	Добавляем ребро
1	$\{3, 3, 4, 5, 4, 6\}$	$\{1, 2, 3, 4, 5, 6, 7, 8\}$	1	(1, 3)
2	$\{3, 4, 5, 4, 6\}$	$\{2, 3, 4, 5, 6, 7, 8\}$	2	(2, 3)
3	$\{4, 5, 4, 6\}$	$\{3, 4, 5, 6, 7, 8\}$	3	(3, 4)
4	$\{5, 4, 6\}$	$\{4, 5, 6, 7, 8\}$	7	(7, 5)
5	$\{4, 6\}$	$\{4, 5, 6, 8\}$	5	(5, 4)
6	$\{6\}$	$\{4, 6, 8\}$	4	(4, 6)
	$\{\}$	$\{6, 8\}$		(6, 8)

Получившееся дерево представлено на рисунке 2.20:

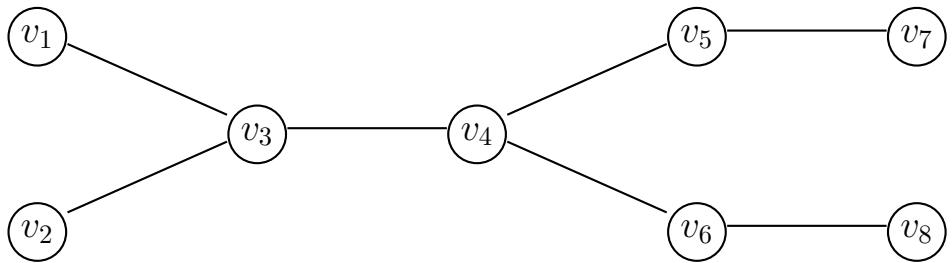


Рис. 2.20

2.1.10. Главные циклы и коциклы

Определение 2.30. Пусть граф $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ имеет p вершин и q ребер и число его компонент связности равно k . Тогда число $\varphi = p - k$ называют *корангом графа*, а $\mu = q - p + k$ — его *цикломатическим числом*.

Пусть $(\mathbf{V}_1, \mathbf{E}_1), \dots, (\mathbf{V}_k, \mathbf{E}_k)$ — компоненты связности графа, имеющие p_1, \dots, p_k вершин и q_1, q_2, \dots, q_k ребер. Таким образом, для каждой

компоненты связности

$$\varphi_i = p_i - 1, \mu_i = q_i - p_i + 1, \quad i = 1, \dots, k.$$

Очевидно, что $\varphi = \sum_{i=1}^k \varphi_i$, $\mu = \sum_{i=1}^k \mu_i$. Доказана следующая теорема.

Теорема 2.25.

- 1) $\mu \geq 0$ для любого графа \mathbf{G} ;
- 2) связный граф \mathbf{G} — дерево тогда и только тогда, когда $\mu = 0$.

Определение 2.31. Пусть $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ — связный граф с p вершинами и q ребрами и $\mathbf{T} = (\mathbf{V}, \mathbf{E}')$ — стягивающее дерево для \mathbf{G} .

Ребра $e \in \mathbf{E}'$ называют *ветвями*, ребра $y \in \mathbf{E} \setminus \mathbf{E}'$ — *хордами*.

Если к дереву \mathbf{T} добавить хорду $y \in \mathbf{E} \setminus \mathbf{E}'$: $y = (v, u)$, то y и \mathbf{T} образуют граф с циклом, который определяется единственным образом цепью, соединяющей вершины v , u . Этот цикл называют *главным циклом*, определяемым хордой y .

Замечание 2.24. Ветвей в графе $p - 1$, хорд $q - (p - 1) = q - p + 1 = \mu$, следовательно, всего существует μ главных циклов.

Определение 2.32. Введем операцию \oplus сложения по модулю 2 или симметрической разности над множествами M_1 и M_2 ребер графа $\mathbf{G} = \{\mathbf{V}, \mathbf{E}\}$:

$$M_1 \oplus M_2 = \{e \in \mathbf{E} \mid (e \in M_1 \text{ и } e \notin M_2) \text{ или } (e \notin M_1 \text{ и } e \in M_2)\}. \quad (2.3)$$

Определение 2.32 можно обобщить на произвольное число слагаемых M_1, \dots, M_k . В этом случае, в 2.3 ребра имеющие *парные* вхождения будут удаляться (согласно свойству операции \oplus). То есть ребро $u \in M_1 \oplus \dots \oplus M_k$ тогда и только тогда, когда оно принадлежит нечетному количеству из M_1, \dots, M_k .

Теорема 2.26. Любой цикл Z графа \mathbf{G} является суммой его некоторых главных циклов C_{i_1}, \dots, C_{i_k}

$$Z = C_{i_1} \oplus \dots \oplus C_{i_k}$$

Пример 2.22. Рассмотрим граф на рис. 2.21. Ребра, вошедшие в стягивающее дерево, имеют большую «толщину». Построим главные циклы графа \mathbf{G} для каждой хорды. Имеем:

$$\begin{aligned} C_1 & - (v_3, v_4) : v_3 \rightarrow v_4 \rightarrow v_5 \rightarrow v_2 \rightarrow v_3, \\ C_2 & - (v_4, v_6) : v_4 \rightarrow v_6 \rightarrow v_5 \rightarrow v_4, \\ C_3 & - (v_1, v_6) : v_1 \rightarrow v_6 \rightarrow v_5 \rightarrow v_2 \rightarrow v_1, \\ C_4 & - (v_3, v_7) : v_3 \rightarrow v_2 \rightarrow v_1 \rightarrow v_7 \rightarrow v_3 \end{aligned}$$

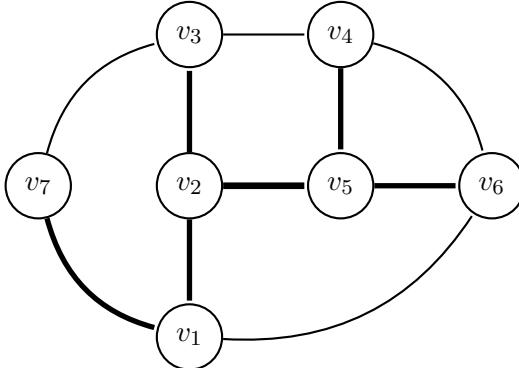


Рис. 2.21

Рассмотрим цикл $Z = \{v_1 \rightarrow v_2 \rightarrow v_3 \rightarrow v_4 \rightarrow v_6 \rightarrow v_1\}$. Тогда

$$Z = C_1 \oplus C_2 \oplus C_3.$$

Определение 2.33. Пусть x' — произвольная ветвь главного цикла, определяемого хордой x . Если к дереву \mathbf{T} добавить хорду x , то единственным простым циклом на полученном графе $(\mathbf{V}, \mathbf{E}' \cup \{x\})$ будет цикл, определяемый x . Если из него исключить ветвь x' , то образуется новое дерево $(\mathbf{V}, \mathbf{E}' \cup \{x\} \setminus \{x'\})$, стягивающее граф \mathbf{G} . Такое преобразование называют *элементарным преобразованием дерева*.

Пример 2.23. На рис. 2.22 показано элементарное преобразование дерева.

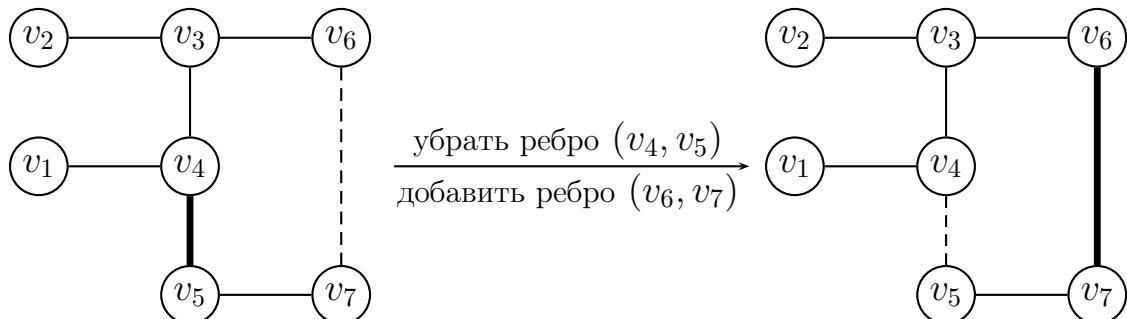


Рис. 2.22

Теорема 2.27. Пусть $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ — связный граф, $\mathbf{T} = (\mathbf{V}, \mathbf{E}')$ — стягивающее его дерево. Пусть $x' \in \mathbf{E}'$ — ветвь главного цикла, определяемого хордой $x \in \mathbf{E} \setminus \mathbf{E}'$. Тогда граф $(\mathbf{V}, \mathbf{E}' \cup \{x\} \setminus \{x'\})$ есть стягивающее дерево для \mathbf{G} .

Определение 2.34. Разрезом связного графа $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ называют множество ребер $\mathbf{C} \subseteq \mathbf{E}$, удаление которых делает граф несвязным. Простым, или минимальным, разрезом называют разрез, никакое собственное подмножество которого таким свойством не обладает.

Замечание 2.25. В этом разделе рассматриваются только простые разрезы, поэтому слово «простой» опускается.

Замечание 2.26. Определения цикла и разреза являются двойственными, поэтому разрез часто называют *коциклом*.

Пример 2.24. Рассмотрим примеры коциклов для графа на рис. 2.23.

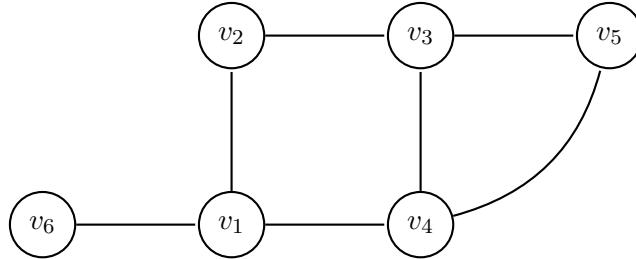


Рис. 2.23

Коциклами для данного графа, например, является одиночное ребро (v_1, v_6) , пары ребер $\{(v_1, v_2), (v_2, v_3)\}$ и $\{(v_5, v_3), (v_5, v_4)\}$. Заметим, что набор $\{(v_1, v_2), (v_1, v_6), (v_1, v_4)\}$ коциклом не является, так как его собственное подмножество $\{(v_1, v_6)\}$ обладает свойством коцикла.

Пусть $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ — связный граф, а \mathbf{C} — его коцикл. Тогда $(\mathbf{V}, \mathbf{E} \setminus \mathbf{C})$ — несвязный граф. Предположим, что его степень связности больше двух. Пусть $(\mathbf{V}_1, \mathbf{E}_1), (\mathbf{V}_2, \mathbf{E}_2), (\mathbf{V}_3, \mathbf{E}_3), \dots$ — компоненты связности $(\mathbf{V}, \mathbf{E} \setminus \mathbf{C})$. Так как граф \mathbf{G} связан, существует $x \in \mathbf{C}$, соединяющее вершины из \mathbf{V}_i и \mathbf{V}_j ($i \neq j$). Для простоты считаем, что $i = 1, j = 2$. Тогда $\mathbf{C}' = \mathbf{C} \setminus \{x\}$ есть собственное подмножество \mathbf{C} , но граф $(\mathbf{V}, \mathbf{E} \setminus \mathbf{C}')$ несвязен. Действительно, вершины из \mathbf{V}_1 и \mathbf{V}_3 не связаны между собой. Противоречие с минимальностью множества \mathbf{C} , следовательно, если \mathbf{C} — коцикл, то степень связности графа $(\mathbf{V}, \mathbf{E} \setminus \mathbf{C})$ равна двум.

Пусть \mathbf{C} — коцикл связного графа $\mathbf{G} = (\mathbf{V}, \mathbf{E})$, а $(\mathbf{V}_1, \mathbf{E}_1), (\mathbf{V}_2, \mathbf{E}_2)$ — две компоненты связности $(\mathbf{V}, \mathbf{E} \setminus \mathbf{C})$. Очевидно, что в \mathbf{C} нет ребер, которые соединяли бы две вершины в \mathbf{V}_1 или в \mathbf{V}_2 .

Введем обозначение:

$$\mathbf{E}(\mathbf{V}_1, \mathbf{V}_2) = \{(v_i, v_j) \in \mathbf{E} \mid v_i \in \mathbf{V}_1, v_j \in \mathbf{V}_2\}.$$

Тогда $\mathbf{C} = \mathbf{E}(\mathbf{V}_1, \mathbf{V}_2)$. Таким образом, доказана следующая теорема.

Теорема 2.28. Если \mathbf{C} — коцикл связного графа $\mathbf{G} = (\mathbf{V}, \mathbf{E})$, то степень связности графа $(\mathbf{V}, \mathbf{E} \setminus \mathbf{C}) = 2$. Пусть $(\mathbf{V}_1, \mathbf{E}_1), (\mathbf{V}_2, \mathbf{E}_2)$ — компоненты связности графа $(\mathbf{V}, \mathbf{E} \setminus \mathbf{C})$, тогда $\mathbf{C} = \mathbf{E}(\mathbf{V}_1, \mathbf{V}_2)$.

Следствие 2.10. Любой цикл и любой коцикл связного графа имеют четное число общих ребер.

Доказательство. Согласно теореме 2.28 произвольный коцикл можно представить в виде $\mathbf{C} = \mathbf{E}(\mathbf{V}_1, \mathbf{V}_2)$. Если некий цикл проходит только по вершинам \mathbf{V}_1 или только по вершинам \mathbf{V}_2 , то он не имеет общих ребер

с коциклом \mathbf{C} . Если же цикл проходит как по вершинам \mathbf{V}_1 , так и по вершинам \mathbf{V}_2 , то он имеет четное число ребер из \mathbf{C} (выйдя из некоторой вершины цикл должен в нее вернуться). ■

Определение 2.35. Пусть $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ — связный граф, а $\mathbf{T} = (\mathbf{V}, \mathbf{E}')$ — его стягивающее дерево. Рассмотрим $x \in \mathbf{E}'$ — ветвь этого дерева. Так как (по теореме 2.12) x — мост, то граф $(\mathbf{V}, \mathbf{E}' \setminus \{x\})$ несвязен. Таким образом, множество $\{x\}$ — это коцикл дерева \mathbf{T} .

Пусть $(\mathbf{V}_1, \mathbf{E}'_1), (\mathbf{V}_2, \mathbf{E}'_2)$ — компоненты связности $(\mathbf{V}, \mathbf{E}' \setminus \{x\})$. Обозначим через \mathbf{E}'' ребра, которые являются хордами графа \mathbf{G} и соединяют вершины из \mathbf{V}_1 и \mathbf{V}_2 , т. е. $\mathbf{E}'' = (\mathbf{E} \setminus \mathbf{E}') \cap \mathbf{E}(\mathbf{V}_1, \mathbf{V}_2)$. Очевидно, что $\{x\} \cup \mathbf{E}''$ — коцикл графа \mathbf{G} . Легко видеть, что \mathbf{E}'' — это множество хорд, главные циклы которых включают ветвь x .

Этот коцикл $\{x\} \cup \mathbf{E}''$ называют *главным коциклом*, определяемым ребром x . Так как число ветвей $\varphi = p - k$ (где p — множество вершин, а k — степень связности), существует φ главных коциклов.

Пример 2.25. Рассмотрим граф на рис. 2.24. Ветви графа отмечены большей «толщиной». Построим главные коциклы графа \mathbf{G} для каждой ветви. Имеем:

$$\begin{aligned} (v_1, v_2) : & \{(v_3, v_4), (v_2, v_5), (v_1, v_2)\}, \\ (v_2, v_3) : & \{(v_3, v_4), (v_1, v_2), (v_2, v_3)\}, \\ (v_1, v_6) : & \{(v_3, v_4), (v_2, v_5), (v_1, v_6)\}, \\ (v_5, v_6) : & \{(v_4, v_5), (v_2, v_5), (v_5, v_6)\}, \\ (v_4, v_6) : & \{(v_3, v_4), (v_4, v_5), (v_4, v_6)\}. \end{aligned}$$

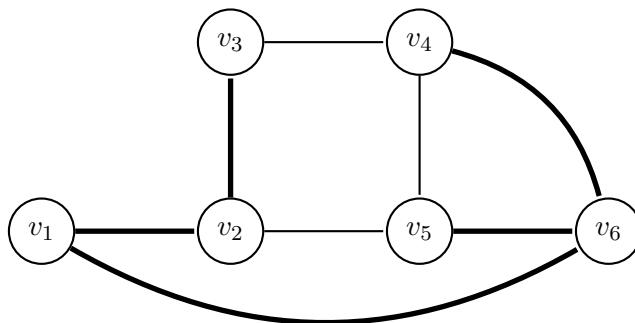


Рис. 2.24

Определение 2.36. В теореме 2.5 было введено понятие объединения совокупности ребер циклов, никакие два из которых не имеют общих ребер.

Непустое объединение множеств ребер некоторого числа циклов, никакие два из которых не имеют общих ребер, называют *границей*.

Непустое объединение множеств ребер некоторого числа коциклов, никакие два из которых не имеют общих ребер, называют *кограницей*.

Лемма 2.5. Пусть $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ — связный граф и множество ребер $\mathbf{B} \subseteq \mathbf{E}$ не содержит целиком ни одного коцикла. Тогда существует дерево, стягивающее \mathbf{G} , которое содержит только ребра из $\mathbf{E} \setminus \mathbf{B}$.

Доказательство. Рассмотрим $\mathbf{T} = (\mathbf{V}, \mathbf{E}')$ — стягивающее дерево графа \mathbf{G} , содержащий наибольшее число ребер из $\mathbf{E} \setminus \mathbf{B}$. Покажем, что $\mathbf{E}' \subset \mathbf{E} \setminus \mathbf{B}$. Пусть существует ребро $x \notin \mathbf{E}' \setminus \mathbf{B}$. Рассмотрим главный коцикл, определенный x (x является ветвью \mathbf{T}).

Так как нет коциклов, полностью лежащих в \mathbf{B} (по предположению), существует хорда $x' \in \mathbf{E} \setminus \mathbf{B}$, которая входит в этот коцикл. Другими словами: главный цикл, определяемый хордой x' , содержит ветвь x .

Сделаем элементарное преобразование дерева \mathbf{T} (см. определение 2.33) — включим хорду x' и выключим ветвь x . Тогда у нового дерева число ребер из $\mathbf{E} \setminus \mathbf{B}$ на одно больше, чем у \mathbf{T} . Противоречие, следовательно, $\mathbf{E}' \subset \mathbf{E} \setminus \mathbf{B}$. ■

Теорема 2.29. Пусть \mathbf{B} — непустое подмножество ребер графа \mathbf{G} , имеющее четное число общих ребер с любым его циклом. Тогда \mathbf{B} — объединение множеств некоторых коциклов, никакие два из которых не имеют общих ребер (т. е. \mathbf{B} — кограница).

Доказательство. Пусть множество \mathbf{B} не содержит целиком ни одного коцикла, тогда по лемме 2.5 существует дерево $\mathbf{T} = (\mathbf{V}, \mathbf{E}')$, стягивающее \mathbf{G} , такое, что $\mathbf{E}' \cap \mathbf{B} = \emptyset$. Возьмем ребро $x \in \mathbf{B}$ ($\mathbf{B} \neq \emptyset$). Так как $x \notin \mathbf{E}'$, то x — хорда.

Рассмотрим главный цикл, определенный x , тогда все остальные ребра цикла есть ветви \mathbf{T} (по определению главного цикла). Таким образом, этот главный цикл не имеет с \mathbf{B} других ребер, кроме x (так как $\mathbf{E}' \cap \mathbf{B} = \emptyset$). Противоречие. Следовательно, существует коцикл \mathbf{C}_1 , целиком лежащий в \mathbf{B} . Рассмотрим множество $\mathbf{B}_1 = \mathbf{B} \setminus \mathbf{C}_1$. Если множество $\mathbf{B}_1 \neq \emptyset$, то по следствию 2.10 к теореме 2.28 множество \mathbf{B}_1 обладает тем же свойством, что и \mathbf{B} . Продолжая этот процесс, пока очередное множество $\mathbf{B}_k \neq \emptyset$, получаем требуемое представление для \mathbf{B} . ■

Следствие 2.11. Звезда $\delta(a)$ произвольной вершины является кограницей.

Доказательство. Звезда $\delta(a)$ и любой цикл графа либо не пересекаются, либо имеют два общих ребра. ■

Теорема 2.30. Пусть \mathbf{B} — непустое подмножество ребер связного графа \mathbf{G} , имеющее четное число общих ребер с любым его коциклом. Тогда

В есть объединение множеств некоторых циклов, никакие два из которых не имеют общих ребер (т. е. **В** — граница).

Упражнение 2.6. Докажите теорему 2.30.

2.1.11. Двудольные графы

Определение 2.37. *Двудольным графом (биграфом, четным графом)* называют граф $\mathbf{G}(\mathbf{V}, \mathbf{E})$, такой, что множество его вершин \mathbf{V} есть объединение двух непересекающихся, непустых множеств \mathbf{V}_1 и \mathbf{V}_2 . При этом каждое ребро графа соединяет вершину из множества \mathbf{V}_1 с вершиной из множества \mathbf{V}_2 .

Множества \mathbf{V}_1 и \mathbf{V}_2 называют долями двудольного графа. Если двудольный граф содержит все ребра, соединяющие множества \mathbf{V}_1 и \mathbf{V}_2 , то это полный двудольный граф. При этом, если $|\mathbf{V}_1| = n$, а $|\mathbf{V}_2| = m$, то такой граф обозначают как $K_{n,m}$.

Замечание 2.27. Пусть $\mathbf{G}(\mathbf{V}, \mathbf{E})$ — двудольный граф. Его всегда можно изобразить так, чтобы вершины доли \mathbf{V}_1 лежали слева, а доли \mathbf{V}_2 — справа. Иногда вершины доли \mathbf{V}_1 будем обозначать литературой L , а доли \mathbf{V}_2 — литературой R .

Пример 2.26. Любое дерево — двудольный граф. Для доказательства достаточно расположить вершины по уровням. Нечетные уровни образуют множество $\mathbf{V}_1(L)$, а четные — множество $\mathbf{V}_2(R)$ (или наоборот) (рис. 2.25, *a*).

Любая прямоугольная «решетка» размера $n \times m$ является двудольным графом. Для доказательства посмотрите разметку вершин на рис. 2.25, *b*.

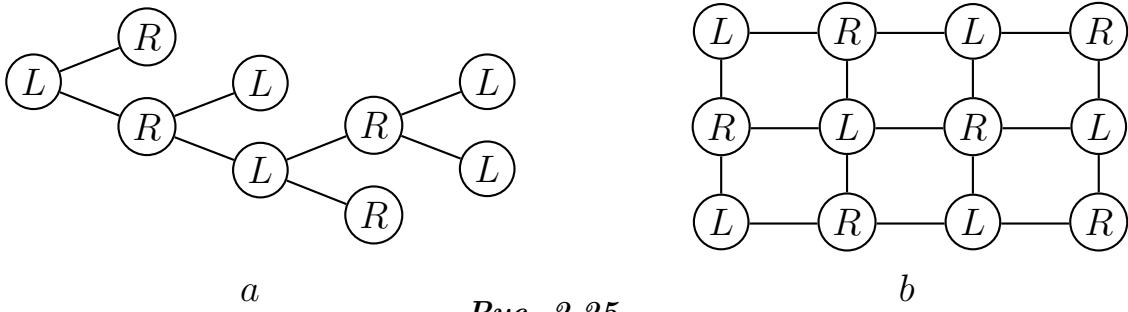


Рис. 2.25

Введем следующие обозначения. Рассмотрим $\mathbf{G}'(\mathbf{V}', \mathbf{E}')$ — подграфа графа $\mathbf{G}(\mathbf{V}, \mathbf{E})$. Если множество ребер \mathbf{E}' совпадает с множеством всех ребер графа \mathbf{G} , оба конца которых принадлежат \mathbf{V}' , то \mathbf{G}' называют *подграфом, порожденным множеством вершин \mathbf{V}'* , и обозначается $\mathbf{G}(\mathbf{V}')$.

Теорема 2.31 (теорема Кенига). Для двудольности графа необходимо и достаточно, чтобы он не содержал циклов нечетной длины.

Рассмотрим алгоритм распознавания двудольности графа.

Алгоритм 2.12 (распознавание двудольности графа). Применим для просмотра вершин графа алгоритм 2.2 поиска в ширину. При этом будем производить нумерацию просматриваемых вершин.

Шаг 1. Первой вершине v_0 приписываем номер 0. До тех пор, пока вершина v_0 не удалена из очереди, каждой просматриваемой вершине приписываем номер 1.

Шаг 2. Пусть текущая вершина на выходе очереди имеет номер n . Каждой просматриваемой и еще незанумерованной вершине присваиваем номер $n + 1$.

Шаг 3. Повторяем шаг 2, пока все вершины не будут просмотрены и занумерованы.

Шаг 4. Разобьем множество вершин графа $\mathbf{G}(\mathbf{V}, \mathbf{E})$ на две части — \mathbf{V}_{even} и \mathbf{V}_{odd} , отнеся к \mathbf{V}_{even} вершины с четными номерами, а к \mathbf{V}_{odd} — остальные вершины, и рассмотрим порожденные этими множествами подграфы $\mathbf{G}(\mathbf{V}_{even})$ и $\mathbf{G}(\mathbf{V}_{odd})$. Если оба они пусты, то $\mathbf{G}(\mathbf{V}, \mathbf{E})$ — двудольный граф. В противном случае граф \mathbf{G} не является двудольным.

Пример 2.27. Применим алгоритм 2.12 для определения двудольности графа на рис. 2.26, а.

Нумерация вершин после завершения шага 3 показана на рис. 2.26, б. Тогда множества вершин

$$\mathbf{V}_{even} = \{v_1, v_2, v_3\}, \quad \mathbf{V}_{odd} = \{v_4, v_5, v_6\}.$$

Очевидно, что множества $\mathbf{G}(\mathbf{V}_{even})$ и $\mathbf{G}(\mathbf{V}_{odd})$ пусты. Следовательно, граф на рис. 2.26, а — двудольный.

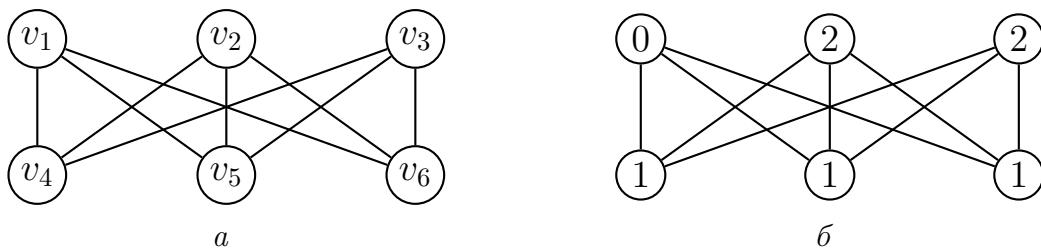


Рис. 2.26

Замечание 2.28. Граф на рис. 2.26, а, двудольность которого была определена выше, это граф $K_{3,3}$, согласно определению 2.37.

Теорема 2.32 (О сумме степеней двудольного графа). Суммы степеней вершин долей двудольного графа равны.

Доказательство. Пусть v_1, \dots, v_k – вершины одной доли, а u_1, \dots, u_p – вершины второй доли. Пусть $d(v)$ – валентность вершины v . Тогда из первой доли выходит $d(v_1) + \dots + d(v_k)$ ребер, а из второй – $d(u_1) + \dots + d(u_p)$ ребер. Так как это одни и те же ребра, получаем:

$$\sum_{i=1}^k d(v_i) = \sum_{j=1}^p d(u_j).$$

■

2.1.12. Планарность

Часто встречаются ситуации, когда важно выяснить, возможно ли нарисовать граф на плоскости так, чтобы его ребра не пересекались. Например, в радиоэлектронике при изготовлении микросхем печатным способом электрические цепи наносят на плоскую поверхность изоляционного материала. А так как проводники не изолированы, то они не должны пересекаться. Аналогичная задача возникает при проектировании железнодорожных и других путей, где нежелательны пересечения. Таким образом возникает понятие плоского графа.

Для дальнейшего нам понадобиться одно из важнейших понятий в теории графов – изоморфизм графов.

Определение 2.38. Графы $\mathbf{G}_1 = (\mathbf{V}_1, \mathbf{E}_1)$ и $\mathbf{G}_2 = (\mathbf{V}_2, \mathbf{E}_2)$ называют *изоморфными* (обозначение: $\mathbf{G}_1 \sim \mathbf{G}_2$), если между ними существует взаимно-однозначное отображение $\varphi: \mathbf{G}_1 \rightarrow \mathbf{G}_2$ ($\mathbf{V}_1 \rightarrow \mathbf{V}_2$, $\mathbf{E}_1 \rightarrow \mathbf{E}_2$), которое сохраняет соответствие между ребрами (дугами) графов, т. е. для любого ребра (дуги) $e = (v, u)$ верно:

$$e' = \varphi(e) = (\varphi(v), \varphi(u)), \quad (e \in \mathbf{E}_1, e' \in \mathbf{E}_2).$$

Отображение φ есть изоморфное отображение.

Характеристики графов, инвариантные относительно изоморфизмов графов (т. е. принимающие одинаковые значения на изоморфных графах), называют инвариантами графов.

Таким образом, изоморфные графы различаются только обозначением вершин. К сожалению, установить изоморфизм двух графов визуально достаточно трудно, что показывает следующий пример.

Пример 2.28. На рис. 2.27, *a*, *b* изображены два изоморфных графа. Приведем одно из возможных изоморфных отображений:

$$\begin{aligned} v_0 &\rightarrow u_0, \quad v_1 \rightarrow u_3, \quad v_2 \rightarrow u_5, \quad v_3 \rightarrow u_6, \quad v_4 \rightarrow u_7, \quad v_5 \rightarrow u_2, \quad v_6 \rightarrow u_1, \quad v_7 \rightarrow u_4, \\ v_8 &\rightarrow u_9, \quad v_9 \rightarrow u_8, \quad v_{10} \rightarrow u_{10}. \end{aligned}$$

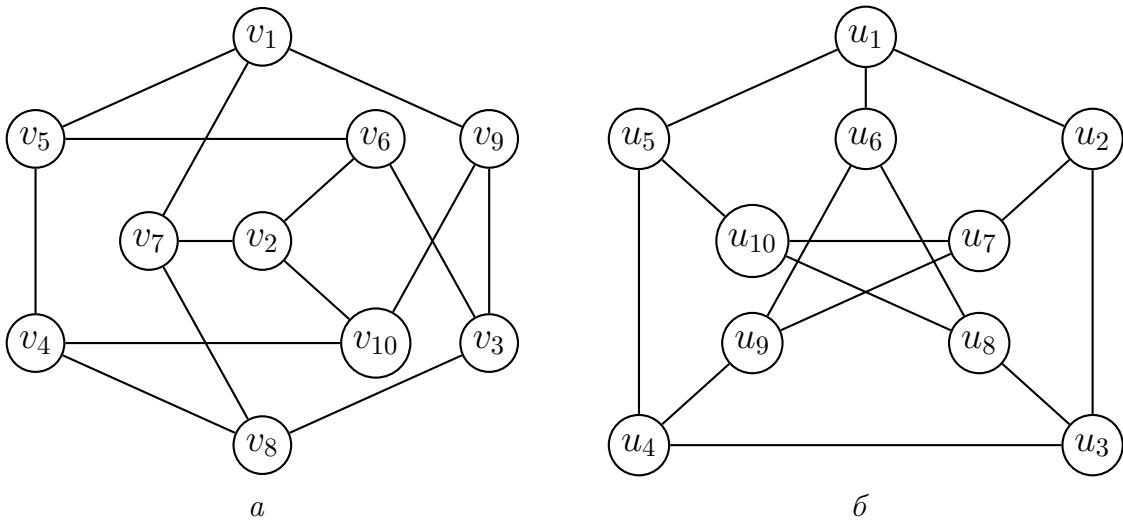


Рис. 2.27

Определение 2.39. Плоским графом называют граф, вершины которого являются точками плоскости, а ребра — непрерывными плоскими линиями без самопересечений, соединяющими соответствующие вершины так, что никакие два ребра не имеют общих точек, кроме инцидентной им обоим вершины.

Любой граф, изоморфный плоскому графу, называют *планарным*. О планарных графах говорят, что они укладываются на плоскости (имеют плоскую укладку).

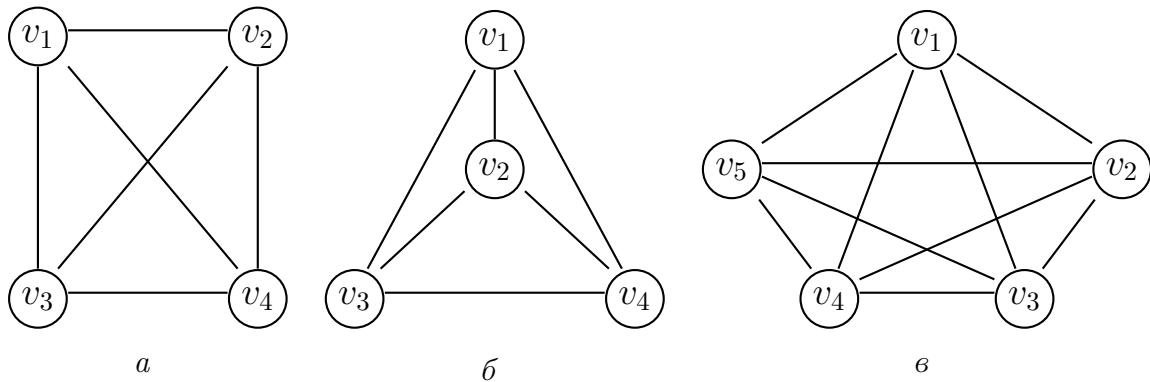


Рис. 2.28

Пример 2.29. Граф на рис. 2.28,а, не является плоским, однако он планарный, поскольку изоморден плоскому полному графу K_4 (рис. 2.28,б). Заметим, что полный граф K_5 (рис. 2.28,в) не является планарным.

Далее будет строго доказано, что граф K_5 не планарен. Легко установить справедливость следующей теоремы.

Теорема 2.33. Каждый граф можно уложить в трехмерном евклидовом пространстве \mathbf{R}^3 .

Доказательство. Рассмотрим произвольный граф $\mathbf{G} = (\mathbf{V}, \mathbf{E})$. Все его вершины разместим в различных точках координатной оси x . Рассмотрим пучок плоскостей, проходящих через x , и зафиксируем $|\mathbf{E}|$ различных таких плоскостей. Теперь каждое ребро $(u, v) \in \mathbf{E}$ изобразим полуокружностью, проходящей в соответствующей плоскости через вершины u и v . Очевидно, что различные ребра не будут пересекаться, кроме как в общих вершинах. ■

Очевидны следующие свойства планарных графов:

- 1) всякий подграф планарного графа планарен;
- 2) граф планарен тогда и только тогда, когда каждая его связная компонента — планарный граф.

Введем следующие определения.

Определение 2.40. Рассмотрим плоский граф. Каждый цикл такого графа ограничивает две части плоскости: одна — внутренняя (ограниченная), другая — внешняя (неограниченная). Если по части плоскости, ограниченной циклом плоского графа, не проходит ни одна цепь этого графа, начало и конец которой принадлежат обсуждаемому циклу, то эту часть плоскости называют *гранью* плоского графа. Если грань неограниченная, то ее называют *внешней*; ограниченную грань плоского графа называют *внутренней*.

Границей грани будем считать множество вершин и ребер, принадлежащих данной грани. Отметим, что всякий плоский граф имеет одну, и притом единственную, неограниченную грань. Такую грань называют *внешней*, а все остальные — *внутренними*.

Замечание 2.29. Будем считать, что если в графе нет циклов, то внешняя грань ограничена всеми ребрами графа.

Пример 2.30. На рис. 2.29 отмечены грани плоского графа: 1, 2, 3 — внутренние, 4 — внешняя.

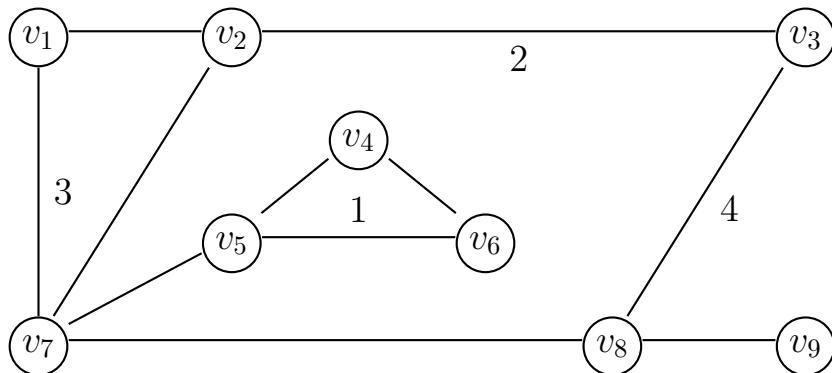


Рис. 2.29

Оказывается, что число граней плоского графа связано с числом вершин и ребер постоянным соотношением, установленным Эйлером.

Теорема 2.34 (теорема Эйлера). Пусть $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ — плоский связный граф, имеющий p вершин, q ребер и f граней. Тогда $p - q + f = 2$.

Доказательство. Зафиксируем число вершин графа и проведем доказательство индукцией по числу ребер. Поскольку граф \mathbf{G} связан, то $q \geq p - 1$. Пусть $q = p - 1$, тогда в силу связности граф \mathbf{G} является деревом, в нем нет циклов и, следовательно, $f = 1$ (существует одна внешняя грань). База индукции установлена.

Пусть теорема справедлива для всех q , таких, что $p - 1 \leq q < q_1$. Докажем ее для q_1 . Поскольку $q_1 > p - 1$, то граф \mathbf{G} содержит цикл. Пусть $e \in \mathbf{V}$ — некоторое ребро этого цикла. Тогда ребро e принадлежит разным граням. Удалим ребро e из графа \mathbf{G} . Тогда в новом графе \mathbf{G}_1 эти две грани сольются в одну, но при этом граф \mathbf{G}_1 останется связан, так как удаленное ребро принадлежало циклу. Граф \mathbf{G}_1 имеет p вершин, $q_1 - 1$ ребро и $f - 1$ грань. По индукционному предположению справедливо соотношение

$$p - (q_1 - 1) + (f - 1) = 2.$$

Отсюда следует, что $p - q_1 + f = 2$. Это доказывает теорему. ■

Следствие 2.12. В условиях теоремы 2.34 при $p > 3$ справедливо равенство

$$3p - q \geq 6. \quad (2.4)$$

Доказательство. Обозначим через φ_k число граней, ограниченных k ребрами. Так как в графе \mathbf{G} нет петель и параллельных ребер, то в силу замечания 2.29

$$\varphi_0 = \varphi_1 = \varphi_2 = 0.$$

Запишем очевидные соотношения:

$$f = \varphi_3 + \varphi_4 + \cdots, \quad 2q = 3\varphi_3 + 4\varphi_4 + \cdots. \quad (2.5)$$

В первом соотношении просуммированы все грани, во втором — ребра, ограничивающие каждую грань, при этом каждое ребро учитывается дважды.

Из равенств (2.5) легко получить, что $2q \geq 3f$. По теореме Эйлера $p - q + f = 2$. Отсюда получаем требуемое неравенство. ■

Используя теорему Эйлера, можно установить непланарность некоторых конкретных графов. Рассмотрим полный граф K_5 из примера 2.29. Для него $p = 5, q = 10$. Если бы граф K_5 был плоским, то, согласно (2.4), $15 - 10 \geq 6$. Полученное противоречие и доказывает непланарность графа K_5 .

Легко доказывается непланарность полного двудольного графа $K_{3,3}$, представленного на рис. 2.26, а. Действительно, для этого графа $p = 6$, $q = 9$. Если бы граф $K_{3,3}$ был плоским, то число граней $f = 9 - 6 + 2 = 5$. Для двудольного графа $K_{3,3}$ нет циклов длины 3, поэтому $\varphi_3 = 0$, следовательно, должно быть выполнено

$$5 = \varphi_4 + \varphi_5 + \dots, \quad 18 = 4\varphi_4 + 5\varphi_5 + \dots$$

Отсюда легко получить противоречие.

Имеется несколько критериев непланарности графа. Приведем без доказательства *критерий Понtryгина–Куратовского*. Для этого введем понятие *гомеоморфизма* графа.

Определение 2.41. Операцией *разбиения ребра* $e = (v, u)$ называют операцию замены ребра e двумя ребрами $e_1 = (v, w)$ и $e_2 = (w, u)$, где w — новая вершина. Два графа называют гомеоморфными, если они могут быть получены из одного графа с помощью операций разбиения ребер.

Пример 2.31. Из графа на рис. 2.30, а можно получить два гомеоморфных графа (рис. 2.30, б и в):

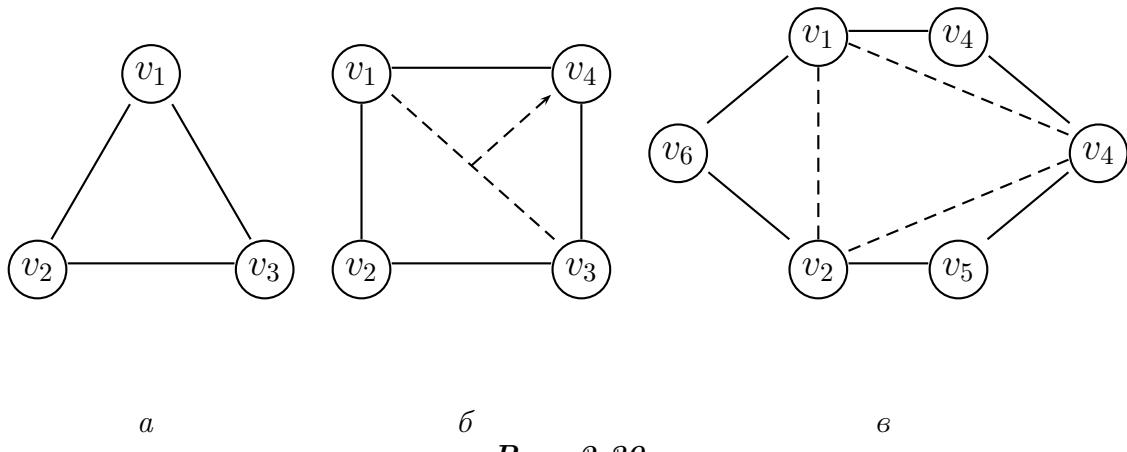


Рис. 2.30

Разбиения ребер показаны на рис. 2.30, б и в штриховыми линиями. Справедлива следующая теорема.

Теорема 2.35 (теорема Понtryгина–Куратовского). Граф планарен тогда и только тогда, когда он не содержит подграфов, гомеоморфных $K_{3,3}$ или K_5 .

2.1.13. Раскраска графов

Определение 2.42. Пусть $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ — некоторый граф. Пусть $\{1, 2, \dots, k\}$ — множество «цветов». Отображение $f : \mathbf{V} \rightarrow \{1, 2, \dots, k\}$ называют *k-раскраской* вершин графа \mathbf{G} . Такую *k*-раскраску называют *правильной*, если для любого ребра $(v_1, v_2) \in \mathbf{E}$ справедливо $f(v_1) \neq f(v_2)$, т. е. смежные вершины получают различную окраску.

Граф \mathbf{G} называют k -раскрашиваемым, если для него существует правильная k -раскраска.

Наименьшее k , для которого граф \mathbf{G} является k -раскрашиваемым, называют *хроматическим числом* графа \mathbf{G} (обозначение: $\chi(\mathbf{G})$). Если $\chi(\mathbf{G}) = k$, то граф \mathbf{G} называют k -хроматическим.

Замечание 2.30. Пусть граф \mathbf{G} получен из мультиграфа $\widehat{\mathbf{G}}$ (см. замечание 2.1) заменой всех ребер, соединяющих две данные вершины, на одно ребро. Очевидно, что раскраска мультиграфа $\widehat{\mathbf{G}}$ правильная тогда и только тогда, когда соответствующая раскраска является правильной для графа \mathbf{G} .

Существует ряд задач, которые приводят к раскраске графа (задачи составления расписаний, обслуживания и др.)

Заметим сначала, что для любого натурального n существует граф \mathbf{G} , такой, что $\chi(\mathbf{G}) = n$. Примером такого графа является K_n . Очевидно, что $\chi(K_n) = n$.

Ясно, что 1-хроматические графы это графы, состоящие из изолированных вершин, 2-хроматические графы — это двудольные графы и только они.

В настоящее время нет описания k -хроматических графов при $k > 3$. Нет также эффективных алгоритмов нахождения хроматического числа графа. Однако, имеются хорошие оценки хроматического числа.

Теорема 2.36 (теорема Визинга). Обозначим через $\Delta(\mathbf{G})$ — максимальную степень вершин графа $\mathbf{G} = (\mathbf{V}, \mathbf{E})$. Тогда справедливо неравенство

$$\chi(\mathbf{G}) \leq \Delta(\mathbf{G}) + 1.$$

Доказательство. Проводим индукцию по n — числу вершин графа. Выберем произвольную вершину $v \in \mathbf{V}$ и удалим ее вместе с инцидентными ей ребрами. Получим граф \mathbf{G}' , для которого $\Delta(\mathbf{G}') \leq \Delta(\mathbf{G})$. Число вершин у графа \mathbf{G}' равно $n-1$ и, следовательно, по индукционному предположению для \mathbf{G}' имеется $(\Delta(\mathbf{G}') + 1)$ -раскраска, а значит и $(\Delta(\mathbf{G}) + 1)$ -раскраска. Тогда $(\Delta(\mathbf{G}) + 1)$ -раскраску для \mathbf{G} можно получить так: окрасим вершину v в цвет, отличный от цветов вершин, смежных с ней, число которых не больше $\Delta(\mathbf{G})$. ■

Для планарных графов данную оценку можно уточнить.

Теорема 2.37 (теорема Хивуда). Для любого планарного графа $\chi(\mathbf{G}) \leq 5$.

Доказательство. Покажем, что в планарном графе существует вершина степени не выше 5. Предположим, что все вершины имеют степень не меньшую, чем 6. Согласно теореме Эйлера (см. теорему 2.34),

$\sum \deg(v) = 2q$, следовательно, $2q \geq 6p$ или $q \geq 3p$. Последнее неравенство противоречит (2.4).

Теперь докажем утверждение теоремы индукцией по числу вершин графа $m = |\mathbf{V}|$.

Для графа с числом вершин, не превосходящим пяти, утверждение очевидно. Считая теорему доказанной для графов с числом вершин, не превосходящим m , рассмотрим граф с $m + 1$ вершиной. По доказанному выше, у графа есть вершина v_0 степени не более пяти. Рассмотрим два случая.

1. Степень вершины v_0 меньше пяти. Удалим из графа эту вершину вместе со всеми входящими в нее ребрами; получим граф с m вершинами. По предположению индукции, для него существует правильная раскраска в 5 или менее цветов. По этой раскраске можно построить правильную раскраску исходного графа с тем же свойством – достаточно раскрасить вершину в v_0 в цвет, отличный от цветов соседних вершин (которых не более четырех).

2. Степень вершины v_0 равна пяти. Рассмотрим смежные с ней вершины. Среди них найдутся две, не соединенные ребром, иначе граф содержал бы полный граф K_5 , что противоречит предположению о планарности графа.

Обозначим эти вершины v_1 и v_2 . Удалим из графа вершину v_0 вместе с пятью смежными с ней ребрами, получим граф \mathbf{G}_1 . Соединим в нем вершины v_1 и v_2 в одну. Получаем граф \mathbf{G}_2 с $m - 1$ вершиной. По индукционному предположению для него существует правильная раскраска не более чем в 5 цветов. Заметим, что получившийся граф \mathbf{G}_2 не содержит петель, так как вершины v_1 и v_2 в исходном графе не были смежны.

При соединении вершин v_1 и v_2 в одну, при наличии у них общих смежных вершин, могут появиться кратные ребра (рис. 2.31), но согласно замечанию 2.30 это не влияет на существование правильной раскраски.

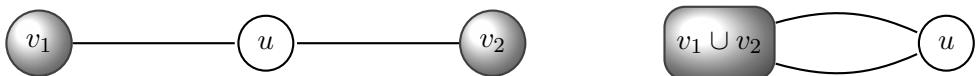


Рис. 2.31

Очевидно, что из этой раскраски получается правильная раскраска графа \mathbf{G}_1 , так как вершины v_1 и v_2 не смежны, то их можно раскрасить в один цвет. Далее, по этой раскраске легко получить раскраску для исходного графа – смежные с v_0 вершины раскрашены не более чем в четыре цвета.

Замечание 2.31. Легко привести пример планарного графа, для которого $\chi(\mathbf{G}) = 4$, например полный граф K_4 (рис. 2.28,б).

Попытки построить планарный граф, для которой недостаточно четырех цветов, длительное время не приводили к успеху. Поэтому естественно возникла гипотеза четырех красок. Она была сформулирована А. Кэли в 1878 году и утверждала, что $\chi(\mathbf{G}) = 4$.

Рассмотренная выше теорема Хивуда была доказана в 1890 году. Подтвердить гипотезу о четырех красках, не удавалось еще почти 90 лет.

Решение было предложено в 1976 г. Т. Аппелем и М. Хакеном. Соответствующее доказательство было получено с применением ЭВМ и пока оно не имеет проверки в традиционном понимании.

Задачи определения хроматического числа и построения минимальной правильной раскраски для произвольного графа является достаточно сложной. Понятно, что для того чтобы найти хроматическое число n -вершинного графа нужно рассмотреть конечное число ситуаций. Действительно, всего имеется k^n различных способов раскраски графа \mathbf{G} в k цветов. Перебрав их, мы либо найдем правильную k -раскраску графа \mathbf{G} , либо убедимся, что ее не существует. Однако, даже для сравнительно небольших n и k такой перебор нельзя провести за приемлемое время.

Рассмотрим простой эвристической алгоритм построения правильной раскраски, близкой к оптимальной, часто его называют алгоритмом *последовательной* раскраски.

Алгоритм 2.13 (Алгоритм последовательной раскраски).

Шаг 1. Упорядочиваем вершины графа в список L в порядке невозрастания степеней.

Шаг 2. Полагаем цвет окраски $p = 1$. Окрашиваем первую вершину списка L в цвет p . Удаляем ее из списка.

Шаг 3. Пока список L не пуст (не окрашены все вершины), повторяем:

- окрашиваем в цвет p все вершины списка L , которые не смежны с вершинами, уже окрашенными в цвет p ;
- удаляем их из списка L ;
- выбираем следующий цвет $p = p + 1$.

Замечание 2.32. Рассмотренный выше алгоритм можно модифицировать. Для этого после каждого шага нужно заново упорядочить список еще неокрашенных вершин по невозрастанию их степеней.

При работе алгоритма предполагалось, что если две вершины имеют одинаковые степени, то порядок раскраски таких вершин случаен. Такие вершины можно также упорядочить, но уже по *двойным* степеням. Двойную степень вершины можно определить как сумму степеней инцидентных ей вершин. Этот процесс можно продолжать и далее.

Пример 2.32. Применим алгоритм 2.13 для раскраски графа, изображенного на рис. 2.32,*a*.

1. Упорядочим вершины графа в список в порядке невозрастания степеней (в скобках – степень вершины):

$$L = \{v_4(5), v_2(4), v_5(4), v_8(4), v_6(3), v_1(2), v_3(2), v_7(2)\}.$$

2. Окрашиваем в цвет 1 вершину v_4 . Удаляем ее из списка.
 3. Окрашиваем в цвет 1 вершину v_3 . Удаляем ее из списка, переходим к цвету 2.
- Замечание 2.33.** Вершину v_6 на этом шаге нельзя покрасить в цвет 1 после окраски в цвет 1 смежной с ней вершины v_3 .
4. Окрашиваем в цвет 2 вершину v_2 и, не смежную с ней, еще не окрашенную вершину v_6 . Удаляем их из списка, переходим к цвету 3.
 5. Окрашиваем в цвет 3 вершину v_5 и, не смежную с ней, еще не окрашенную вершину v_1 . Удаляем их из списка, переходим к цвету 4.
 6. Окрашиваем в цвет 4 вершину v_8 . Удаляем ее из списка. Список пуст (все вершины графа окрашены), алгоритм заканчивает работу.

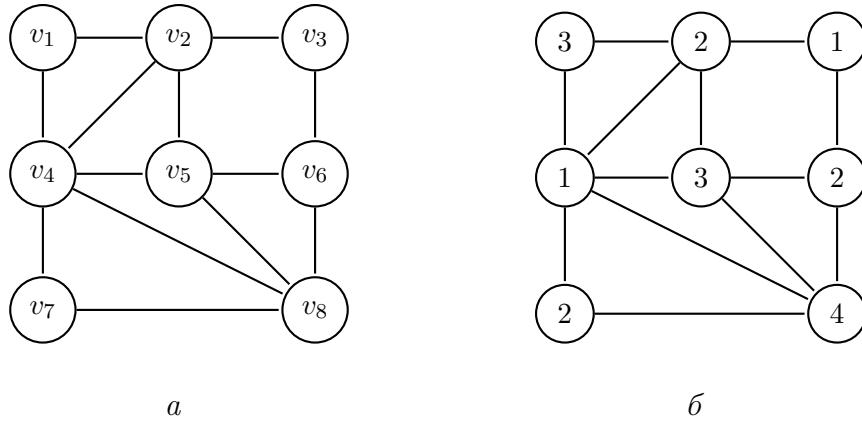


Рис. 2.32

Легко заметить, что в данном примере построена минимальная раскраска ($\chi(\mathbf{G}) = 4$), так как $\deg(v_4) = 5$. Результат работы алгоритма представлен на рис. 2.32, б.

Рассмотрим пример задачи, для решения которой нужно найти хроматическое число и наименьшую правильную раскраску графа. Это задача составления оптимального расписания.

Нужно прочесть ряд лекций группам студентов. Некоторые из лекций не могут читаться одновременно, например, потому, что их читает один и тот же лектор, или их надо читать одной и той же группе студентов. Требуется составить расписание занятий так, чтобы чтение всех лекций заняло минимально возможное время.

С точки зрения раскраски графов нужно построить граф \mathbf{G} , в котором вершины соответствуют лекциям. Вершины графа смежны тогда и только тогда, когда соответствующие лекции не могут читаться одновременно.

Очевидно, что любая правильная k -раскраска графа \mathbf{G} определяет допустимое расписание. При этом вершины, раскрашенные i -м цветом, дают

список лекций, которые надо читать на i -й паре. И наоборот, любое допустимое расписание определяет правильную раскраску графа \mathbf{G} . Таким образом, составление оптимального расписания сводится к нахождению наименьшей раскраски построенного графа.

Пример 2.33. Необходимо провести занятия по математике (M), физике (F), иностранному языку (L) и истории России (I) в группах G_1 и G_2 . Занятия проводятся преподавателями T_1 и T_2 – по одному занятию по каждому предмету. Таблица 2.13 указывает, какие занятия надо провести в группах и какими преподавателями они могут быть проведены:

Таблица 2.13

Лекции	Группы		Преподаватели	
	G_1	G_2	T_1	T_2
Математика	+		+	
Физика	+	+	+	
Иностранный язык		+		+
История России	+	+		+

Требуется найти минимальное число пар, в которые можно уложить все занятия, и составить соответствующее расписание.

Построим граф с вершинами M_1 , F_1 , F_1 , I_1 , I_2 и L_2 (буква соответствует обозначению предмета, а цифра – номеру группы). Соединим ребрами вершины, соответствующие парам, которые нельзя проводить одновременно. Получим граф, изображенный на рис. 2.33,а. Применим для его раскраски алгоритм 2.13.

1. Упорядочим вершины графа в список в порядке невозрастания степеней:

$$L = \{I_1(4), F_2(4), F_1(3), M_1(3), L_2(3), I_2(3)\}.$$

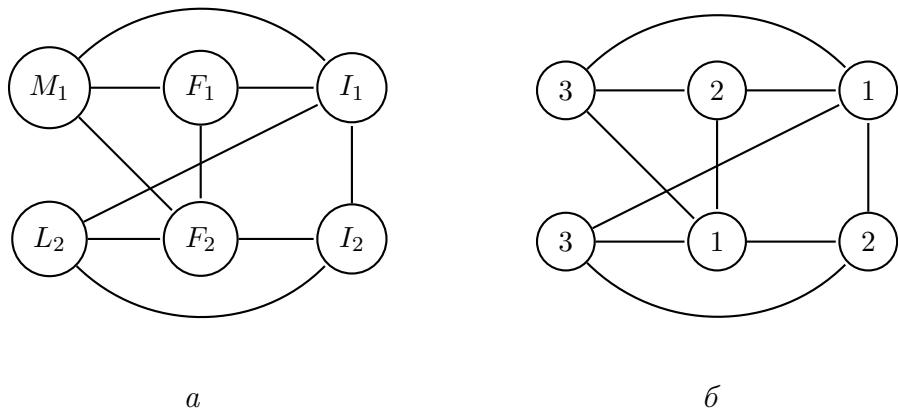
2. Окрашиваем в цвет 1 вершины I_1 и F_2 .

3. Окрашиваем в цвет 2 вершины F_1 и I_2 .

4. Окрашиваем в цвет 3 оставшиеся вершины M_1 и L_2 . Все вершины графа окрашены, алгоритм заканчивает работу.

Результат работы алгоритма представлен на рис. 2.33,б.

Соответственно, оптимальное расписание содержит три пары занятий (табл. 2.14):



Puc. 2.33

Таблица 2.14

Номер пары	Группа G_1	Группа G_2
1	История России (T_2)	Физика (T_1)
2	Физика (T_1)	История России (T_2)
3	Математика (T_1)	Иностранный язык (T_2)

2.2. Алгоритмы на графах

2.2.1. Построение наибольшего паросочетания

Рассмотрим задачу, которую называют задачей о супружеских парах. Неформальная постановка задачи такая. Имеется n юношей и m девушек. Про каждого юношу и каждую девушку известно, нравятся ли они друг другу. Нужно переженить как можно больше симпатизирующих друг другу пар.

Перед тем как поставить задачу формально, дадим необходимые определения.

Определение 2.43. Паросочетанием двудольного графа называют подмножество его ребер, никакие два из которых не являются смежными (не инциденты одной вершине).

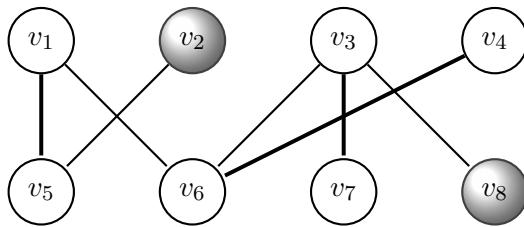
Будем считать, что множество из одного ребра является паросочетанием.

Паросочетание P в графе \mathbf{G} называется *максимальным*, если в \mathbf{G} нет паросочетаний, число ребер в которых больше, чем в P .

Вершина v графа \mathbf{G} называется *насыщенной* в паросочетании P , если в P существует ребро, инцидентное v , в противном случае вершина называется *свободной*.

Паросочетание P называется *совершенным*, если все вершины графа G насыщены в P .

Очевидно, что каждое совершенное паросочетание является максимальным, но обратное неверно. Например, в графе на рис. 2.34 паросочетание $\{(v_1, v_2), (v_3, v_4), (v_5, v_6)\}$ выделенное жирными линиями, является максимальным, но не совершенным. Вершины $v_1, v_3, v_4, v_5, v_6, v_7$ — насыщенные, v_2, v_8 — (выделенные серым цветом) свободные.

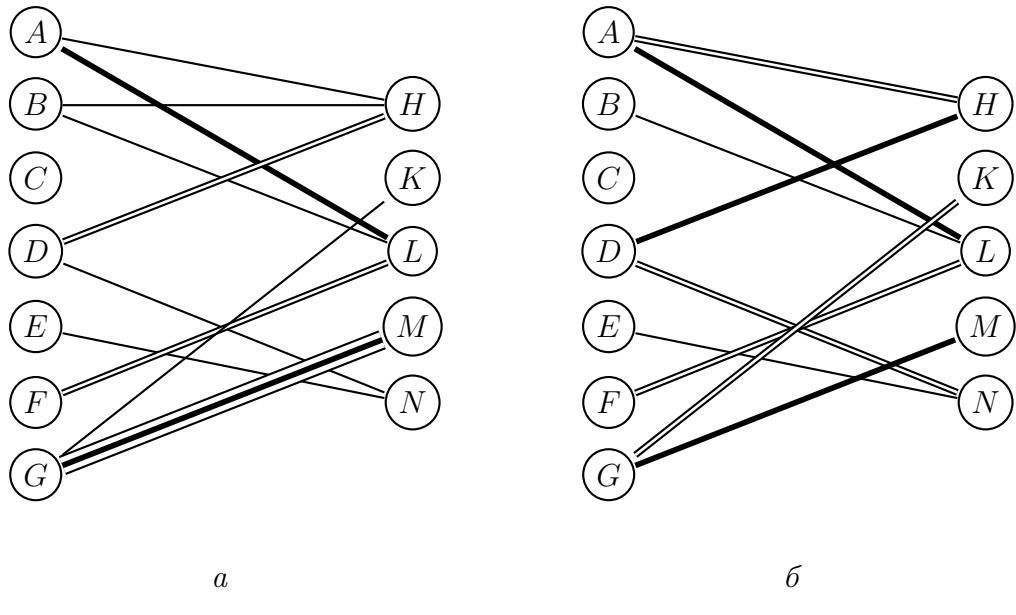


Puc. 2.34

Формальная постановка задачи. В заданном двудольном графе найти максимальное паросочетание (или все максимальные паросочетания).

Для нахождения максимального паросочетания применим идею последовательного улучшения любого паросочетания. Рассмотрим сначала два произвольных паросочетания X и Y , второе из которых содержит большее число ребер: $|X| < |Y|$ (на рис. 2.35,*a* и *б* ребра паросочетания X выделены жирной линией, а паросочетания Y — двойной).

Если исключить общие для обоих паросочетаний ребра (на рис. 2.35,а это ребро GM), то оставшиеся ребра разбиваются на цепочки.



Puc. 2.35

Причем, цепочки, имеющие два или более ребер, состоят из чередующихся ребер различных паросочетаний. Действительно, ребра одного па-

росочетания не могут быть смежными по определению. Так как ребер у паросочетания Y больше, то обязательно найдется цепочка, начинающаяся и заканчивающаяся ребром из паросочетания Y (на рис. 2.35,*a* это единственное ребро DH , на рис. 2.35,*б* — это цепочка $FLAHDN$).

Таким образом, переход от паросочетания X к Y может осуществляться изменением принадлежности ребер вдоль такой цепочки: все ее ребра, принадлежащие паросочетанию X из нее надо удалить, заменив оставшимися ребрами этой цепочки.

Чтобы алгоритмизировать процесс нахождения таких цепочек, сделаем следующее: ориентируем ребра паросочетания X справа налево, а ребра, не вошедшие в X слева направо. Тогда движение вдоль цепочки соответствует движению по направлениям ребер, начиная с вершины левой части двудольного графа (рис. 2.36).

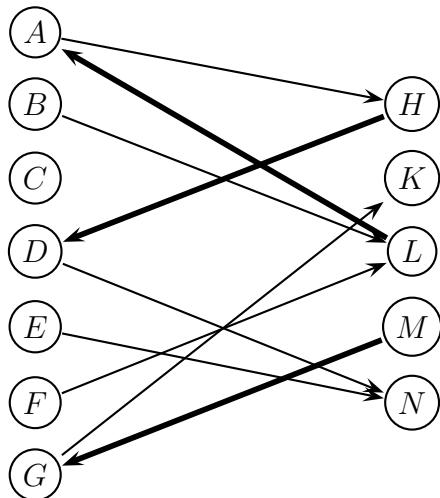


Рис. 2.36

Далее, чтобы применить для нахождения улучшающих цепочек алгоритмы обхода графа, введем две фиктивные вершины: S (начальную) и T (конечную). Соединим вершину S с вершинами левой части L , не входящими в построенное паросочетание X , выбрав направления ребер из вершины S . Аналогично, все «свободные» вершины правой части соединим с вершиной F , направив ребра в нее (рис. 2.37).

Тогда, если удастся найти путь из вершины S в вершину T , то вдоль него можно улучшить паросочетание X , поменяв ориентацию вдоль пути (что и означает замену ребер старого паросочетания ребрами улучшающего паросочетания) и исключив фиктивные ребра. На рис. 2.37 таким путем является

$$S \rightarrow F \rightarrow L \rightarrow A \rightarrow H \rightarrow D \rightarrow N \rightarrow T.$$

Проводя указанное действие, получаем следующий график (рис. 2.38): Если на очередном шаге пути из S в T нет, то максимальное паросо-

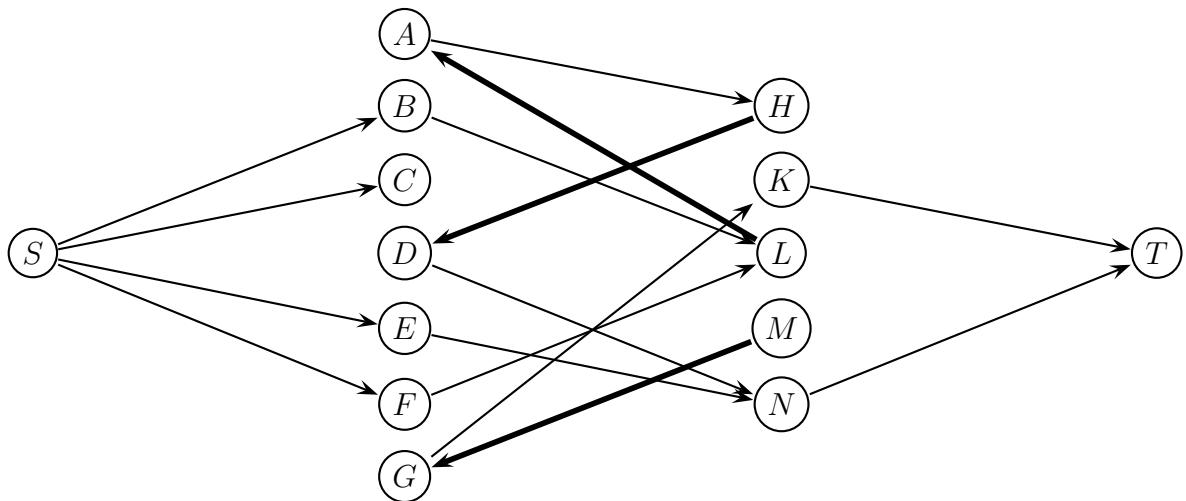


Рис. 2.37

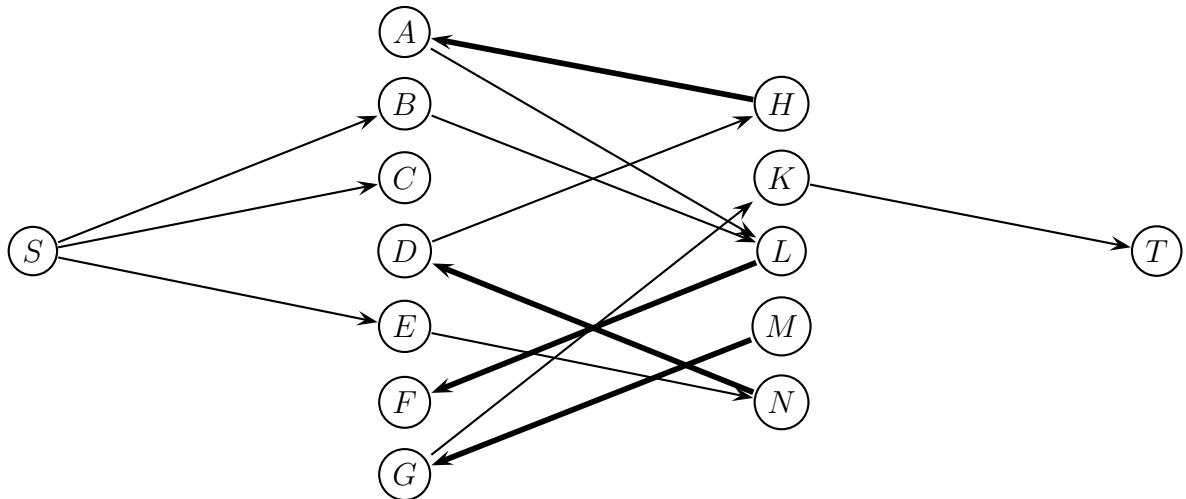


Рис. 2.38

чтение найдено.

Замечание 2.34. Обычно существует несколько различных путей из S в T . Если стоит задача отыскания всех паросочетаний, то нужно организовать перебор вариантов, в противном случае достаточно выбора произвольного пути. В разобранном примере (см. рис. 2.37) имелся еще один улучшающий путь $S \rightarrow E \rightarrow N \rightarrow T$, который приводит к другому максимальному паросочетанию.

Сравните: $X_1 = \{AH; DN; FL; GM\}$ и $X_2 = \{AL; DH; EN; GM\}$.

Построенные алгоритмом максимальные паросочетания могут быть различными (в зависимости от процедуры выбора очередного пути), но число ребер в них всегда будет одинаково и является инвариантом алгоритма.

Сформулируем эти рассуждения в виде алгоритма.

Алгоритм 2.14 (построения максимального паросочетания).

Инициализация:

Шаг 1. Ввести фиктивные вершины S и T , соединив S с вершинами левой части двудольного графа L ребрами, ориентированными от вер-

шины S . Соединить вершины правой части двудольного графа R с вершиной T , ориентировав ребра к вершине T .

Шаг 2. Ориентировать все ребра двудольного графа слева направо (от вершин множества L к вершинам множества R).

Основная часть:

ЦИКЛ-ПОКА существует путь от S к T

Найти путь от S к T

Изменить ориентацию ребер вдоль этого пути

Удалить начальное и конечное ребро этого пути

КЦ

Выход алгоритма:

Множество ребер, ориентированных справа налево (от вершин множества R к вершинам множества L), и дает искомое паросочетание.

Пример 2.34. С помощью алгоритма 2.14 построим максимальное паросочетание для двудольногографа \mathbf{G} . Левая доля V_1 состоит из вершин $\{v_1, v_2, v_3, v_4, v_5, v_6, v_7\}$, правая доля V_2 из вершин $\{u_1, u_2, u_3, u_4, u_5, u_6, u_7\}$. Ребра в графе заданы следующим списком:

$(v_1, u_5), (v_1, u_6), (v_1, u_7), (v_2, u_3), (v_3, u_2), (v_4, u_2), (v_4, u_3), (v_5, u_1), (v_5, u_3), (v_5, u_4), (v_5, u_6), (v_6, u_7), (v_7, u_2)$.

Проводим инициализацию алгоритма. Получившийся ориентированный граф изображен на рис. 2.39.

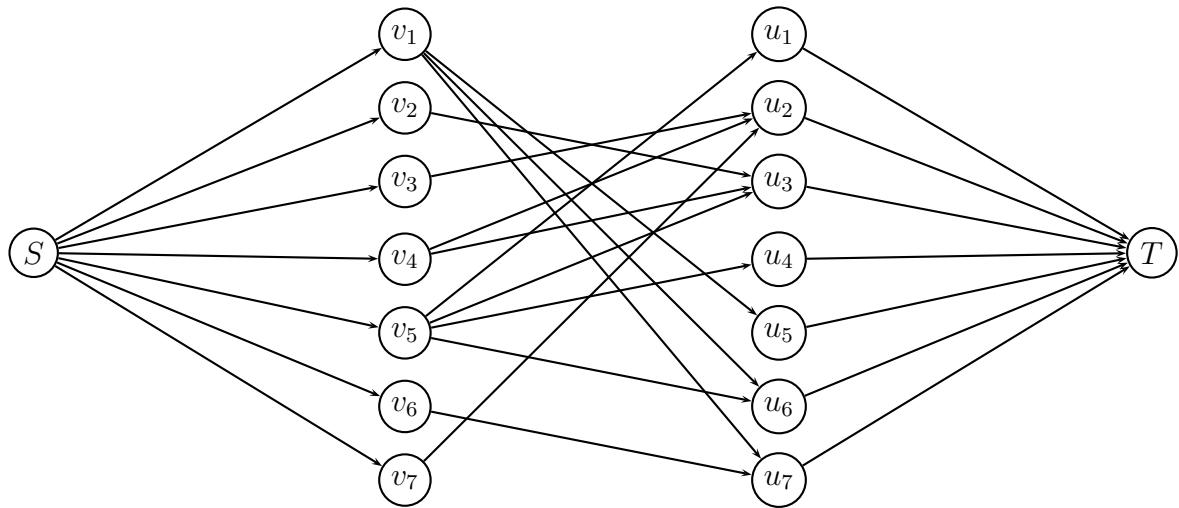


Рис. 2.39

Будем перебирать возможные пути, для определенности соединяя вершины с меньшими номерами. Обозначим искомое паросочетание P . Запишем протокол работы алгоритма (табл. 2.15):

Таким образом, $P = \{(v_1, u_5), (v_2, u_3), (v_3, u_2), (v_5, u_1), (v_6, u_7)\}$.

Таблица 2.15

Путь	Фиктивные ребра	Добавляем в P
$S \rightarrow v_1 \rightarrow u_5 \rightarrow T$	$(S, v_1), (u_5, T)$	(v_1, u_5)
$S \rightarrow v_2 \rightarrow u_3 \rightarrow T$	$(S, v_2), (u_3, T)$	(v_2, u_3)
$S \rightarrow v_3 \rightarrow u_2 \rightarrow T$	$(S, v_3), (u_2, T)$	(v_3, u_2)
$S \rightarrow v_5 \rightarrow u_1 \rightarrow T$	$(S, v_5), (u_1, T)$	(v_5, u_1)
$S \rightarrow v_6 \rightarrow u_7 \rightarrow T$	$(S, v_6), (u_7, T)$	(v_6, u_7)

Экспериментальная работа. Сравните работу алгоритмов поиска в ширину и глубину для поиска максимальных паросочетаний в графах — решетках размера $n \times m$ (см. пример 2.26, рис. 2.25,б).

2.2.2. Построение минимальных стягивающих деревьев

Определение 2.44. Граф $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ называют *нагруженным*, или *взвешенным*, если для каждого ребра (v_i, v_j) определена его длина (или вес) c_{ij} . Матрицу $\{c_{ij}\}$ называют *матрицей весов ребер*.

Веса несуществующих ребер полагают равными ∞ или 0 в зависимости от приложений. Заметим, что матрица весов является простым обобщением матрицы смежности.

Пусть \mathbf{G} — связный нагруженный граф. Задача построения минимального стягивающего дерева заключается в том, чтобы из множества стягивающих деревьев найти дерево, у которого сумма длин ребер минимальна. Приведем типичные случаи, когда возникает необходимость построения минимального стягивающего дерева графа.

1. Нужно соединить n городов железнодорожными линиями (автомобильными дорогами, линиями электропередач, сетью трубопроводов и т. д.) так, чтобы суммарная длина линий или стоимость была минимальной.

2. Требуется построить схему электрической сети, в которой клеммы должны быть соединены с помощью проводов наименьшей общей длины.

Решение этой задачи «слепым» перебором вариантов потребовало бы чрезвычайно больших вычислений даже при относительно малых n (можно доказать, что полный граф K_n содержит n^{n-2} различных стягивающих деревьев). Однако для ее решения имеются эффективные алгоритмы. Опишем два из них — алгоритмы Прима и Краскала, применяемые к произвольному связному графу \mathbf{G} с n вершинами.

Историческая справка. Исторически первый алгоритм построения минимального стягивающего дерева был предложен Отакаром Борувкой (Otakar Boruvka) в 1926 г., задолго до появления первых компьютеров и современной теории графов. Борувка применил его в качестве метода нахождения оптимальной электрической сети в Моравии. Несколько раз он был переоткрыт различными авторами, поэтому его иногда называют алгоритмом Соллина, особенно в литературе по параллельным вычислениям.

Рассматриваемые алгоритмы разработаны практически одновременно Дж. Краскалом¹ (Joseph Kruskal) и Р. Примом (Robert Prim) в 1956 г. В 1958 г. Э. Дейкстра (Edsger Dijkstra) независимо переоткрыл алгоритм Прима, но название осталось прежним, так как алгоритм Дейкстры уже существовал (см. алгоритм 2.19 поиска кратчайших путей в графе).

Замечание 2.35. Рассматриваемые далее алгоритмы Прима и Краскала, как и алгоритм Хаффмена (см.замечание 1.62) являются примерами *жадных* алгоритмов.

Алгоритм 2.15 (алгоритм Прима или ближайшего соседа).

Построение начнем с произвольной вершины s .

Шаг 1. Выбираем ребро $e_1 = (s, u)$ минимального веса и строим дерево \mathbf{T}_1 , полагая $\mathbf{V}_{T_1} = \{s, u\}$, $\mathbf{E}_{T_1} = \{e_1\}$.

Шаг $i + 1$. Если дерево \mathbf{T}_i уже построено и $i < n - 1$, то среди ребер, соединяющих вершины этого дерева с вершинами графа \mathbf{G} , не входящими в \mathbf{T}_i , выбираем ребро e_{i+1} минимального веса. Строим дерево \mathbf{T}_{i+1} , присоединяя к \mathbf{T}_i ребро e_{i+1} вместе с его не входящим в \mathbf{T}_i концом.

Проведем обоснование корректности алгоритма Прима.

Лемма 2.6. Пусть веса всех ребер связного графа различны. Разобьем множество \mathbf{V} его вершин на два непересекающихся подмножества \mathbf{V}_1 и \mathbf{V}_2 . Рассмотрим подмножество ребер $\widehat{\mathbf{E}} = \mathbf{E}(\mathbf{V}_1, \mathbf{V}_2)$. Выберем в $\widehat{\mathbf{E}}$ ребро e_1 минимального веса. Тогда это ребро обязательно войдет в минимальное стягивающее дерево \mathbf{T} .

Доказательство. От противного. Пусть это ребро не входит в минимальное стягивающее дерево: $e_1 \notin \mathbf{T}$. Тогда должно найтись другое ребро e_2 из множества $\widehat{\mathbf{E}}$, входящее в стягивающее дерево, иначе нарушилась бы связность. По предположению о различности весов ребер, вес e_2 больше веса e_1 . Добавим к дереву \mathbf{T} хорду e_1 , тогда в полученном графе появится главный цикл, определяемым хордой e_1 . В этот цикл обязательно должно войти еще одно ребро множества $\widehat{\mathbf{E}}$. Единственно возможное ребро — ребро e_2 (остальные ребра \mathbf{E}_1 не входят в $\mathbf{T} \cap e_1$).

Сделаем элементарное преобразование дерева \mathbf{T} (см. определение 2.33) — включим хорду e_1 и выключим ветвь e_2 . Новое дерево очевидно будет также стягивающим, но меньшего веса. Полученное противоречие доказывает лемму. ■

¹Иногда говорят Крускал.

Замечание 2.36. Если среди ребер есть ребра одинакового веса, то формулировка леммы 2.6 несколько изменится. Заключительная часть ее будет звучать так «найдется минимальное стягивающее дерево, содержащее это ребро». Основная часть доказательства сохранится, только надо обратить внимание на то, что при замене ребер можно перейти к дереву равного веса.

Лемма 2.7. Рассмотрим подграф, содержащий все вершины, построенные на k -м шаге алгоритма и соединяющие их ребра. Тогда дерево, построенное на k -м шаге, является минимальным стягивающим деревом этого подграфа.

Доказательство. Лемма доказывается аналогично лемме 2.6. Обратим внимание на обоснование индукционного перехода. Пусть мы добавили к текущему оствовному дереву минимальное ребро из множества ребер, соединяющих его вершины с остальными. При этом мы расширили множество вершин графа (добавив второй конец ребра). Минимальное оствовное дерево нового подграфа обязательно должно включать одно ребро, соединяющее новую вершину с остальными. Таких ребер не было в оствовном дереве, построенном ранее. Следовательно, добавление нового ребра, инцидентного данной вершине, не может оказаться на выборе ребер на предыдущих этапах. ■

Справедлива следующая теорема.

Теорема 2.38. Граф T_{n-1} является стягивающим деревом минимального веса в графе G .

Доказательство. Рассмотрим случай, когда все веса ребер различны. Докажем ее по индукции.

База индукции. При инициализации алгоритма выбирается произвольная вершина s связного неориентированного графа. Далее в дерево включается минимальное ребро, соединяющее эту вершину с остальными. Обозначив $V_1 = \{s\}$, $V_2 = V \setminus \{s\}$, по лемме 2.6 можно утверждать, что выбранное ребро обязательно войдет в минимальное оствовное дерево.

Индукционный переход. Пусть после k шагов алгоритма получено k ребер, принадлежащих минимальному стягивающему дереву. Обозначив $V_1 = \{s, v_1, \dots, v_k\}$, $V_2 = V \setminus V_1$ можно применить лемму 2.6 еще раз и утверждать, что очередное выбранное ребро принадлежит стягивающему дереву. Итак, для случая различных весов ребер теорема доказана.

В общем случае теорема следует из леммы 2.7 очевидным способом, когда k становится равным числу, на 1 меньше числа вершин графа. ■

Проиллюстрируем работу алгоритма 2.15 на примере.

Пример 2.35. Построим стягивающее дерево минимального веса для графа на рис. 2.40, начиная с вершины H .

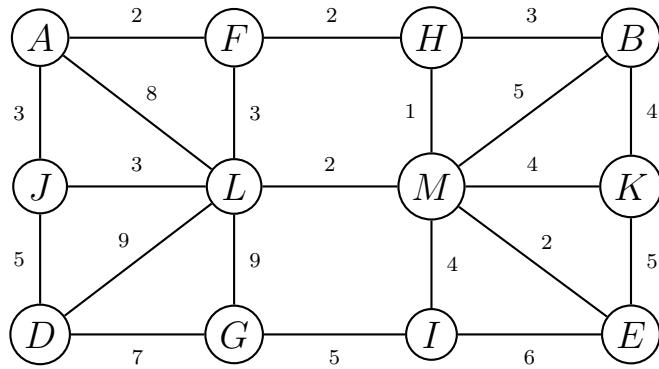


Рис. 2.40

Согласно алгоритму 2.15, получаем последовательность включаемых ребер:

$$\begin{aligned} HM(1) \rightarrow HF(2) \rightarrow FA(2) \rightarrow ML(2) \rightarrow ME(2) \rightarrow LJ(3) \rightarrow HB(3) \rightarrow \\ BK(4) \rightarrow MI(4) \rightarrow IG(5) \rightarrow JD(5) \end{aligned}$$

Таким образом, вес построенного дерева T_{11} равен 33. Полученное дерево выглядит следующим образом (рис. 2.41).

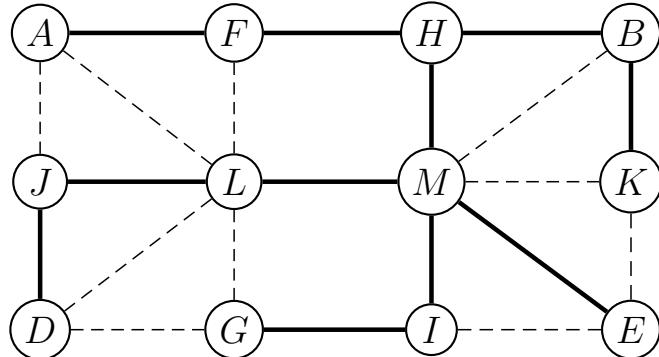


Рис. 2.41

Замечание 2.37. Приведенные рассуждения нигде не опирались на знаки весов ребер и, следовательно, алгоритм Прима применим для графов с произвольными (положительными, отрицательными или нулевыми) весами ребер. Отсюда вытекает, что стягивающее дерево максимального веса можно также построить алгоритмом Прима. Следует только изменить знаки весов на противоположные или всюду минимальный вес заменить максимальным.

Все сказанное справдливо и для алгоритма Краскала.

Алгоритм 2.16 (алгоритм Краскала).

Шаг 1. Выбираем ребро e_1 минимального веса в графе и рассматриваем подграф \mathbf{G}_1 , состоящий из концов ребра e_1 и самого ребра.

Шаг $i + 1$. Выбираем в графе ребро e_{i+1} , имеющее минимальный вес среди ребер графа \mathbf{G} , не входящих в подграф \mathbf{G}_i и не составляющих цикла с ребрами из \mathbf{G}_i . Строим подграф $\mathbf{G}_{i+1} := \mathbf{G}_i \cup \{e_{i+1}\}$.

Относительно алгоритма Краскала справедлива теорема, аналогичная теореме 2.38.

Теорема 2.39. Подграф \mathbf{G}_{n-1} является минимальным стягивающим деревом графа \mathbf{G} .

Доказательство. От противного. Пусть существует дерево $\widehat{\mathbf{T}}$ с весом, меньшим чем вес дерева \mathbf{T}_K , построенного по алгоритму Краскала. Сравним эти деревья по совпадению ребер в том порядке, как ребра вставлялись в дерево \mathbf{T}_K . Найдем первое из ребер e дерева $\widehat{\mathbf{T}}$, не вошедших в \mathbf{T}_K . Добавим это ребро в дерево \mathbf{T}_K . В полученном графе появится цикл, содержащий ребро e . В этом цикле должно найтись ребро w , вес которого больше либо равен весу e . Действительно, ребра меньшего веса вместе с e не могут образовать цикла, так как входят в дерево \mathbf{T}_K . Заменив ребро w на e , получим новое оствовное дерево с весом, либо меньшим веса $\widehat{\mathbf{T}}$ (что сразу приводит к противоречию), либо с равным весу $\widehat{\mathbf{T}}$. В последнем случае после замены ребер можно продолжить процесс сравнения и использовать те же рассуждения. Так как по предположению вес $\widehat{\mathbf{T}}$ меньше веса \mathbf{T}_K , то рано или поздно вес нового оствовного дерева должен стать меньше веса $\widehat{\mathbf{T}}$, что и приводит к противоречию. ■

Проиллюстрируем работу алгоритма 2.16 на примере графа на рис. 2.40 (см. пример 2.35).

Пример 2.36. Применив алгоритм 2.16, получим одну из возможных последовательностей включаемых ребер:

$$\begin{aligned} HM(1) \rightarrow FA(2) \rightarrow ME(2) \rightarrow ML(2) \rightarrow HF(2) \rightarrow AJ(3) \rightarrow HB(3) \rightarrow \\ MK(4) \rightarrow MI(4) \rightarrow JD(5) \rightarrow IG(5) \end{aligned}$$

Таким образом, вес построенного дерева \mathbf{G}_{11} равен 33. Полученное дерево выглядит следующим образом (рис. 2.42).

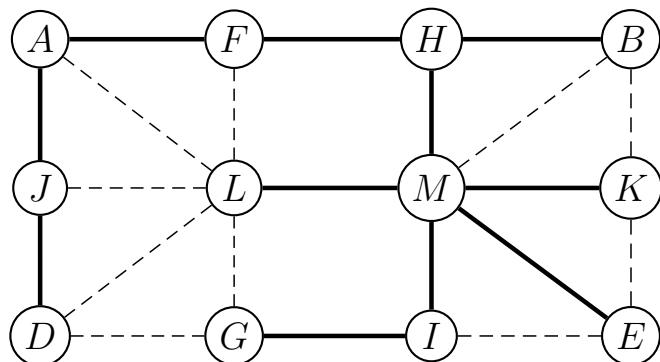


Рис. 2.42

Замечание 2.38. Построенный алгоритм 2.16 включает процедуру определения наличия цикла в графе, что само по себе — трудная задача. Поэтому алгоритм Краскала можно переписать в другой форме, используя идею раскраски вершин.

Алгоритм 2.17 (модифицированный алгоритм Краскала).

Упорядочим все ребра по возрастанию (неубыванию) весов.

Присвоим вершинам графа различные номера (раскрасим вершины в разные цвета).

ЦИКЛ-ПОКА не кончились ребра

Рассмотреть очередное ребро

ЕСЛИ концы ребра окрашены в разные цвета, ТО

добавить ребро в стягивающее дерево и

перекрасить все вершины, номер которых совпадает с номером

одного конца добавленного ребра в цвет второго конца этого ребра

(для определенности можно перекрашивать в цвет вершины с большим номером)

КЕ

КЦ

Пример 2.37. Рассмотрим протокол перекраски вершин графа на рис. 2.43, а.

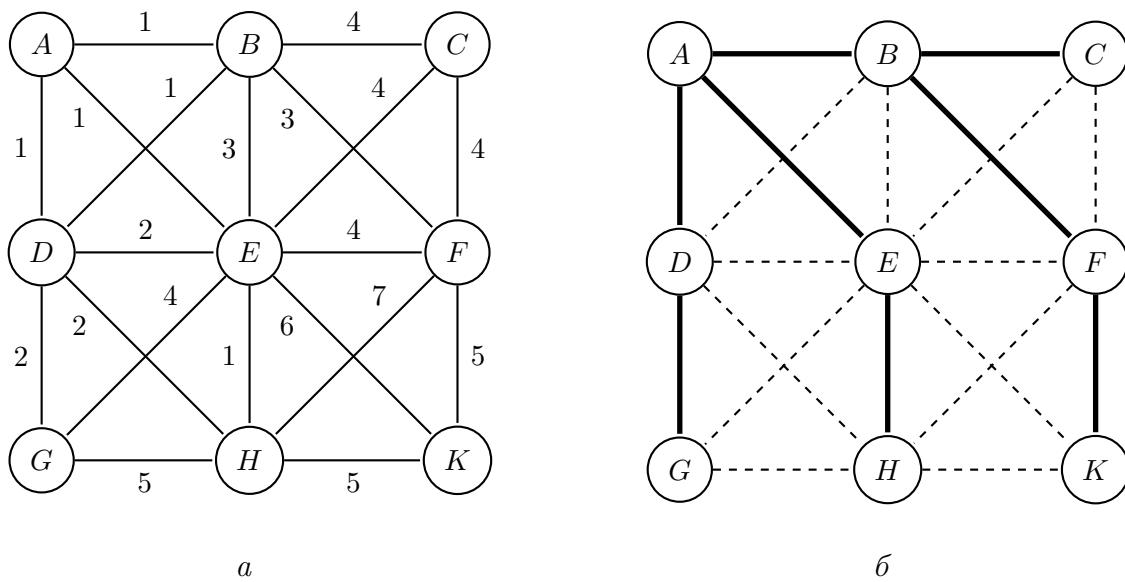


Рис. 2.43

Последовательность ребер, упорядоченных по возрастанию (неубыванию) весов; ребра одного веса упорядочены в лексикографическом порядке (можно было взять любой другой порядок):

$AB \rightarrow AD \rightarrow AE \rightarrow BD \rightarrow EH \rightarrow DE \rightarrow DG \rightarrow DH \rightarrow BE \rightarrow BF \rightarrow \rightarrow BC \rightarrow CE \rightarrow CF \rightarrow EF \rightarrow EG \rightarrow FK \rightarrow GH \rightarrow HK \rightarrow EK \rightarrow \rightarrow FH.$

На рис. 2.43,*б* представлено стягивающее дерево, построенное в результате работы алгоритма 2.17. Хорды дерева обозначены штриховыми линиями.

Протокол работы модифицированного алгоритма Краскала показан в табл. 2.16. Полужирным шрифтом выделены номера вершин текущего ребра.

Таблица 2.16

Текущее ребро	Номера вершин									Выбранное ребро
	A	B	C	D	E	F	G	H	K	
AB	1	2	3	4	5	6	7	8	9	AB
AD	2	2	3	4	5	6	7	8	9	AD
AE	4	4	3	4	5	6	7	8	9	AE
BD	5	5	3	5	5	6	7	8	9	
EH	5	5	3	5	5	6	7	8	9	EH
DE	8	8	3	8	8	6	7	8	9	
DG	8	8	3	8	8	6	7	8	9	DG
DH	8	8	3	8	8	6	8	8	9	
BE	8	8	3	8	8	6	8	8	9	
BF	8	8	3	8	8	6	8	8	9	BF
BC	8	8	3	8	8	8	8	8	9	BC
CE	8	8	8	8	8	8	8	8	9	
CF	8	8	8	8	8	8	8	8	9	
EF	8	8	8	8	8	8	8	8	9	
EG	8	8	8	8	8	8	8	8	9	
FK	8	8	8	8	8	8	8	8	9	FK
GH	9	9	9	9	9	9	9	9	9	
HK	9	9	9	9	9	9	9	9	9	
EK	9	9	9	9	9	9	9	9	9	
FH	9	9	9	9	9	9	9	9	9	

Замечание 2.39. В примере 2.37 видно, что последние четыре шага лишние: дерево построено, и все вершины графа раскрашены в один цвет. Оба этих признака можно использовать для проверки окончания цикла (например, при числе вершин графа n цикл можно закончить, когда число выбранных ребер равно $n - 1$).

Задача 2.1. Докажите корректность модифицированного алгоритма Краскала.

2.2.3. Задача нахождения кратчайших путей в графе

Рассмотрим ориентированный или неориентированный взвешенный граф. Сформулируем три различные постановки задачи на нахождение

кратчайших путей.

1. Заданы две вершины графа: S и F . Найти длину кратчайшего пути, соединяющего вершину S с вершиной F .
2. Задана вершина графа S . Найти длины кратчайших путей, соединяющих вершину S с остальными вершинами графа.
3. Найти длины кратчайших путей, соединяющих все пары вершин графа.

Попытаемся оценить сложность каждой из задач и упорядочить их для последовательного решения.

Первая мысль — оставить порядок решения таким, как он здесь представлен. Действительно, решив первую задачу, можно перебрать все возможные вершины F графа, сохранив начальную S , и, тем самым, решить вторую задачу (сделать цикл по вершинам графа, используя в теле цикла процедуру решения первой задачи). Затем можно «освободить» первую вершину и с помощью еще одного цикла решить третью задачу.

Тем самым определена трудоемкость указанных алгоритмов: если трудоемкость алгоритма решения первой задачи обозначить $T(n)$ (n — число вершин графа), то трудоемкость второго будет $nT(n)$, третьего — $n^2T(n)$.

Такой путь возможен, но не рационален. Почему?

Для ответа на вопрос представим, как искать кратчайший путь между двумя фиксированными вершинами S и F : для того чтобы убедиться, что он кратчайший, надо сравнить его длину с длинами других возможных путей. Но для этого надо знать кратчайшие расстояния от вершины S до остальных вершин графа! Получается, что вторая задача по отношению к первой является базовой и решение надо начинать с нее. Действительно, следующие два алгоритма, которые будут рассмотрены (алгоритм Форда–Беллмана и алгоритм Дейкстры), решают именно вторую задачу. В то же время, очевидно, что при решении второй задачи одновременно решается и первая, так что выделять ее особо не требуется.

Из изложенного вытекает и основная идея алгоритма, находящего кратчайшие пути: надо сравнивать уже найденные варианты «кратчайших путей» с теми, которые используют проход через другие вершины. Это можно записать более формально, если ввести для каждой вершины A ее пометку — вещественное число $D(A)$, которое по окончании работы алгоритма означает кратчайшее расстояние, а на промежуточных шагах имеет иной смысл (который будет ясен из дальнейшего анализа алгоритмов, так как для каждого алгоритма его смысл на промежуточных шагах свой). Тогда новое значение пометки можно найти как

$$D(A) := \min\{D(A); D(B) + P(B; A)\},$$

где минимум берется по всем (промежуточным) вершинам графа B , а $P(B; A)$ — это вес ребра (в алгоритме Форда–Беллмана) или вес пути (в алгоритме Флойда), соединяющего B с A . Для удобства пронумеруем вершины графа, тогда описанную операцию можно записать так:

$$\begin{aligned} &\text{ЦИКЛ ПО } k \text{ от 1 до } n \\ &D(i) := \min\{D(i); D(k) + P(k; i)\} \quad (2.6) \\ &\text{КЦ} \end{aligned}$$

Назовем внутреннюю часть этого цикла модификацией пометки.

2.2.4. Алгоритм Форда–Беллмана

Постановка задачи. Имеется ориентированный или неориентированный взвешенный граф, веса которого задаются произвольными вещественными числами. Требуется найти длины всех кратчайших путей от вершины S (источника) до остальных вершин.

Первоначально проанализируем корректность постановки задачи. Ответим на два вопроса:

1. Что делать в случае, когда пути вообще нет, в частности, если граф оказывается несвязным?
2. Не может ли оказаться, что при наличии отрицательных ребер кратчайших путей нет вовсе?

Ответ на первый вопрос прост. Введем понятие бесконечно большой длины и будем присваивать «бесконечность» пометкам вершин, обозначающим длины кратчайших путей от источника. Ответ на второй вопрос более сложен.

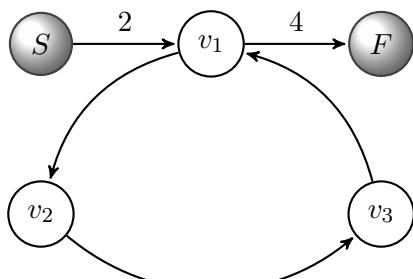


Рис. 2.44

На рис. 2.44 показан пример графа, в котором не существует кратчайшего пути из вершины S в вершину F . Это связано с наличием цикла отрицательной длины ($v_1 \rightarrow v_2 \rightarrow v_3 \rightarrow v_1$) — «бегая» по нему, можно «накрутить» сколь угодно малую (отрицательную, но большую по абсолютному значению) длину, т. е. наименьшего пути между S и F не существует

в том смысле, что для любого пути можно найти путь меньшей длины. Нетрудно понять, что отсутствие цикла отрицательной длины и является необходимым и достаточным условием для корректности поставленной задачи. Докажите самостоятельно следующую теорему.

Теорема 2.40. Во взвешенном графе с вещественными весами кратчайшие пути между любыми двумя вершинами существуют тогда и только тогда, когда в графе нет циклов отрицательной длины.

Заметим, что это условие несколько избыточно для исходной задачи, в которой начальная вершина фиксирована. Так появляется интересное задание для исследования.

Задание 2.1. Может ли при наличии цикла отрицательной длины существовать решение задачи о кратчайших путях от фиксированной вершины S до остальных. Найдите поясняющий пример. Сформулируйте необходимые и достаточные условия для этого.

Основная идея алгоритма. Сопоставим вершинам пометки, смысл которых — кратчайшие пути, состоящие из одного ребра, т. е. из матрицы весов возьмем строчку, соответствующую начальной вершине (веса ребер, выходящих из начальной вершины). Повторим для каждой вершины графа цикл (2.6), модифицирующий пометки одной вершины через все возможные промежуточные вершины. Нетрудно понять, что при этом к путям из одного ребра добавятся пути из двух ребер и смысл пометки станет иным: «кратчайшее расстояние от источника до всех остальных вершин на множестве путей, состоящих либо из одного ребра, либо из двух». Каждое следующее повторение будет расширять множество путей, добавляя пути, состоящие из большего количества ребер. Но этот процесс не бесконечен. Простой путь в графе из n вершин не может иметь более $(n - 1)$ ребра (заметим, что здесь как раз и играет роль отсутствие циклов отрицательной длины — ведь рассматривая пути с повторяющимися вершинами, т. е. с циклами, в этом случае можно было бы уменьшать длину пути и далее). Поэтому, повторив цикл $(n - 2)$ раза, наверняка получим искомый результат.

После сделанных замечаний нетрудно сформулировать алгоритм, который называется алгоритмом Форда–Беллмана.

Алгоритм 2.18 (алгоритм Форда–Беллмана).

Инициализация:

```

ЦИКЛ ПО  $i$  от 1 до  $n$ 
     $D(i) := P(s; i)$ 
КЦ

```

Здесь в качестве начальных значений пометок берутся веса ребер, соединяющих начальную вершину s с остальными (напомним, что при отсутствии таких ребер пометка берется равной бесконечности).

Основной алгоритм:

```

ЦИКЛ ПО  $t$  от 2 до  $n - 1$ 
    ЦИКЛ ПО  $i$  от 1 до  $n$ 
        ЦИКЛ ПО  $k$  от 1 до  $n$ 
             $D(i) := \min\{D(i); D(k) + P(k; i)\}$ 
        КЦ
    КЦ
КЦ

```

Здесь переменная t — просто счетчик и не имеет отношения к номерам вершин i и k .

Доказательство корректности алгоритма 2.18. Доказательство корректности циклических алгоритмов использует общий прием, который называется выделением инварианта цикла. Инвариант цикла — это утверждение, зависящее от параметра цикла (в данном случае t), которое остается верным (при любом допустимом значении параметра t) после исполнения тела цикла при условии, что оно было верно до этого.

Для данного циклического алгоритма таким инвариантом является содержательный смысл пометки $D(i)$. Сформулируем этот результат в виде леммы.

Лемма 2.8. После выполнения t -го шага алгоритма (точнее, внешнего цикла алгоритма) пометка $D(i)$ обозначает кратчайшее расстояние между источником и вершиной i на множестве путей, состоящих не более чем из t ребер.

Доказательство. Докажем утверждение по индукции.

База индукции. Первым шагом алгоритма является инициализация, которая и обеспечивает верность утверждения для $t = 1$. База доказана.

Индукционный переход. Пусть утверждение верно для $t = t$ перед выполнением тела цикла, докажем, что оно будет верно и для $t = t + 1$ после выполнения тела цикла. То есть, надо доказать, что если до выполнения тела цикла пометка $D(i)$ показывала кратчайшее расстояние от источника до вершины i на множестве путей, состоящих не более чем из t ребер, то после его выполнения она будет показывать кратчайшее расстояние на множестве путей, состоящих не более чем из $(t + 1)$ ребра. Но любой путь до вершины i из не более чем $(t + 1)$ ребра — это путь из не более чем t ребер (для таких путей кратчайшее расстояние уже известно по индукционному предположению) или путь ровно из $(t + 1)$ ребра. В последнем случае этот путь можно разбить на два: путь до предпоследней вершины

k из t ребер (наименьшее значение которого известно по индукционному предположению) и ребро из вершины k в вершину i . Остается из этих двух вариантов выбрать наилучший:

$$D(i) := \min\{D(i); D(k) + P(k; i)\}.$$

Но это и есть то самое правило модификации, которое лежит в основе алгоритма! ■

Перейдем к доказательству корректности алгоритма. Посмотрим теперь на наш инвариант:

« $D(i)$ — кратчайшее расстояние на множестве путей не более чем из m ($m \leq n$) ребер»

С другой стороны, число m меняется от 1 до $(n-1)$, а утверждение остается верным. Значит, после окончания работы алгоритма получим утверждение:

« $D(i)$ — кратчайшее расстояние на множестве путей не более чем из $(n-1)$ ребра».

Но так как циклов отрицательной длины в графе нет, то последнее утверждение равносильно утверждению:

« $D(i)$ — кратчайшее расстояние на множестве всех путей».

Корректность алгоритма установлена.

Замечание 2.40. Полезно обратить внимание на два содержательных шага в доказательстве:

1) использование метода математической индукции; причем база индукции обеспечивалась соответствующей инициализацией, а индукционный переход — соответствующей модификацией;

2) использование инварианта цикла — утверждения, которое, с одной стороны, отражает содержание введенной переменной, с другой — позволяет формально анализировать корректность циклического алгоритма.

Это общие приемы, которые будут повторяться при выводе и обосновании других алгоритмов.

Замечание 2.41. В доказательстве предполагалось, что на очередном шаге внешнего цикла используются пометки предыдущего шага, на самом деле пометки меняются не только при изменении параметра внешнего цикла m , но и при изменении внутреннего i (при переходе от одной вершины к другой). Разумеется, от этого «скорость» нахождения длин кратчайших путей только увеличивается.

2.2.5. Алгоритм Форда–Беллмана для нахождения дерева кратчайших путей

Алгоритм Форда–Беллмана позволяет находить длины кратчайших путей от источника до остальных вершин, но не сами пути. В то же время, нетрудно видеть, что при поиске длин кратчайших путей имеется вся необходимая информация для восстановления путей. Небольшая доработка алгоритма Форда–Беллмана, предусматривающая введение вторых пометок

вершин графа (для хранения тех вершин, через которые осуществлялась модификация длины пути), решает эту задачу. Заметим, что объединение всех кратчайших путей от источника до остальных вершин обязательно дает дерево, ориентированное естественным образом (от корня к листьям), которое называют *деревом кратчайших путей*. Обоснование этому дает следующая лемма.

Лемма 2.9. Если построен кратчайший путь от вершины S до вершины F , а M — произвольная промежуточная вершина этого пути, то кратчайший путь от S до M будет частью пути от S до F .

Доказательство. Докажем от противного. Пусть существует путь от S до M , не полностью совпадающий с начальной частью пути от S до F (рис. 2.45). Тогда построим новый путь от S до F , состоящий из нового пути от S до M и старого — от M до F . По предположению, он будет короче, что противоречит тому, что старый путь от S до F является кратчайшим. Противоречие доказывает теорему. ■

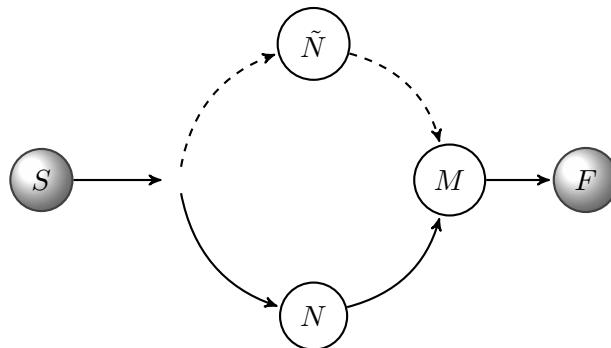


Рис. 2.45

Задание 2.2. Верно ли, что для связного неориентированного взвешенного графа дерево кратчайших путей будет:

- остовным деревом графа;
- минимальным остовным деревом графа.

Ответ обоснуйте.

Приведем теперь сам алгоритм.

Инициализация:

```

ЦИКЛ ПО  $i$  от 1 до  $n$ 
     $D(i) := P(s; i); \quad G(i) := s$ 
    КЦ
  
```

Вычисление пометок:

```

ЦИКЛ ПО  $m$  от 2 до  $n - 1$ 
    ЦИКЛ ПО  $i$  от 1 до  $n$ 
        ЦИКЛ ПО  $k$  от 1 до  $n$ 
            ЕСЛИ  $D(k) + P(k; i) < D(i)$ 
                ТО       $D(i) := D(k) + P(k; i); G(i) := k$ 
            КЦ
        КЦ
    КЦ

```

Построение дерева кратчайших путей T :

```

 $T = \emptyset$ 
ЦИКЛ ПО  $i$  от 1 до  $n$ 
     $T := T \cup (G(i); i)$ 
КЦ

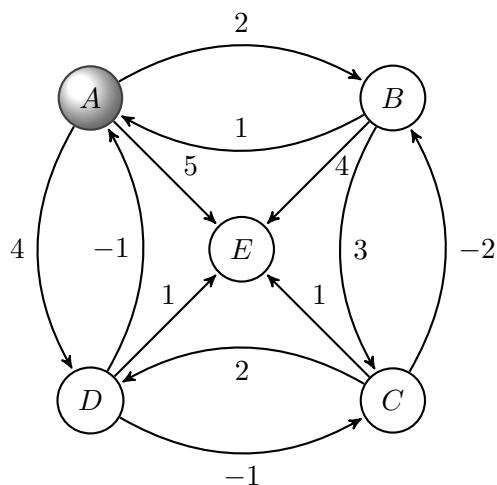
```

где $(G(i); i)$ — очередное ребро дерева кратчайших путей.

Замечание 2.42. В этом алгоритме в дерево добавляется петля $(s; s)$, где s — источник. Этого легко избежать организовав последний цикл по всем вершинам, кроме вершины s .

Рассмотрим работу алгоритма на примере.

Пример 2.38. Граф и матрица его весов¹ представлены на рис. 2.46, а и б. Источником является вершина A .



a

	A	B	C	D	E
A	0	2	∞	4	5
B	1	0	3	∞	4
C	∞	-2	0	2	1
D	-1	∞	-1	0	1
E	∞	∞	∞	∞	0

б

Рис. 2.46

¹Символом ∞ в таблице обозначается отсутствие соответствующей дуги.

Методические замечания. При «прогонке» алгоритма полезно не смотреть на рисунок, а использовать более формальное представление графа, например матрицу весов. Результат работы алгоритма, наоборот, рекомендуется проиллюстрировать графически.

Протокол оформим в виде таблицы, где каждая строка — значения пометок на очередном шаге внешнего цикла. Для экономии места будем следить за обеими пометками одновременно. Выполним инициализацию. В качестве начальной вершины S (источника) возьмем вершину A . Тогда первой строкой протокола будет первая строка матрицы весов, в которую добавлены пометки «предшествующих» вершин. В начале работы алгоритма — это источник, в данном случае — вершина A .

Номер шага m	A	B	C	D	E
1	$0/A$	$2/A$	∞	$4/A$	$5/A$

Далее поочередно для всех вершин, начиная с вершины B , проверим возможность улучшить ее пометку через все соседние вершины (пометку вершины A можно не модифицировать, так как при отсутствии циклов отрицательной длины пометка источника измениться не может). К вершине B можно перейти из вершин A и C (см. столбец B матрицы весов, рис. 2.46, δ). Вершину A по указанным ранее причинам можно не рассматривать. Для вершины C суммарный вес пути будет

$$D(C) + P(C; B) = \infty - 2 = \infty,$$

что не улучшает старую пометку, которую, таким образом, следует сохранить. К вершине C можно перейти из вершин B и D (см. столбец C матрицы весов, рис. 2.46, δ), соответствующие веса путей будут:

$$D(B) + P(B; C) = 2 + 3 = 5; \quad D(D) + P(D; C) = 4 - 1 = 3.$$

Вторая сумма лучше и первой и старой пометки (которая равна ∞), поэтому поменяем эту пометку и сохраним имя вершины (пометка $G(C)$), через которую пришли в C . Далее аналогично. Заметим, что пометку в вершине E улучшим через C

$$D(C) + P(C; E) = 3 + 1 = 4,$$

которая сама была улучшена на предыдущем шаге. Об этом «ускорении» отмечалось в замечании 2.41.

Номер шага m	A	B	C	D	E
1	$0/A$	$2/A$	∞	$4/A$	$5/A$
2	$0/A$	$2/A$	$3/D$	$4/A$	$4/C$

На следующем шаге внешнего цикла произойдет модификация пометки $D(B)$ через вершину C :

$$D(C) + P(C; B) = 3 - 2 = 1.$$

Очередной шаг не принесет изменений, что неудивительно, так как самые короткие пути необязательно должны состоять из наибольшего числа ребер. Итоговое состояние протокола:

Номер шага m	A	B	C	D	E
1	$0/A$	$2/A$	∞	$4/A$	$5/A$
2	$0/A$	$2/A$	$3/D$	$4/A$	$4/C$
3	$0/A$	$1/C$	$3/D$	$4/A$	$4/C$
4	$0/A$	$1/C$	$3/D$	$4/A$	$4/C$

Для построения дерева кратчайших путей достаточно последней строки протокола:

Номер шага m	A	B	C	D	E
4	$0/A$	$1/C$	$3/D$	$4/A$	$4/C$

Собрав пары «пометка вершины — вершина», получим дерево кратчайших путей:

$$(C; B), (D; C), (A; D), (C; E).$$

Около каждой вершины можно указать кратчайшее расстояние до нее:

$$A - 0, B - 1, C - 3, D - 4, E - 4.$$

На рис. 2.47 показано дерево кратчайших путей с кратчайшими расстояниями до вершин.

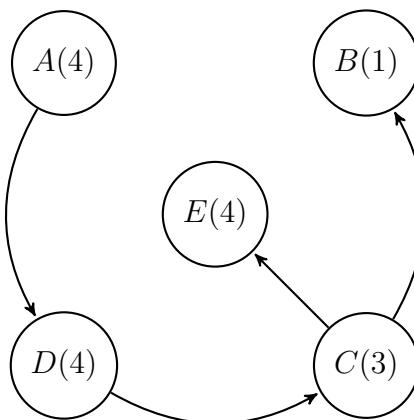


Рис. 2.47

Задача 2.2. Примените алгоритм Форда—Беллмана для построения дерева кратчайших путей для того же графа, но других начальных вершин:
а) B , б) C , в) D .

2.2.6. Алгоритм Дейкстры

Постановка задачи. В отличие от предыдущей, в данной задаче веса ребер неотрицательны. Тем самым, циклы отрицательной длины автоматически исключаются и задача всегда будет иметь решение. Кроме того, это условие позволяет на порядок уменьшить трудоемкость алгоритма (об этом далее).

Основная идея алгоритма. На каждом шаге отыскивается вершина, кратчайшее расстояние до которой уже найдено. Например, на первом шаге это будет вершина — конец наименьшего ребра, инцидентного источнику. Пометки остальных вершин модифицируются через эту вершину, после чего ее можно не рассматривать.

Алгоритм 2.19 (алгоритм Дейкстры). Обозначим W — множество вершин, кратчайшее расстояние до которых уже найдено, и назовем его множеством обработанных вершин.

Инициализация: такая же, как в алгоритме Форда–Беллмана.

Основной алгоритм:

```

 $W := \{s\}$ 
ЦИКЛ-ПОКА  $V \setminus W \neq \emptyset$  (есть необработанные вершины)
    найти вершину  $w \in V \setminus W$  с наименьшей пометкой  $D(w)$ 
    перенести найденную вершину в множество обработанных:
     $W := W \cup \{w\}$ 
    модифицировать пометки необработанных вершин через  $w$ :
        ЦИКЛ ПО всем вершинам  $u \in V \setminus W$ , смежным с  $w$ 
             $D(u) := \min\{D(u); D(w) + P(w; u)\}$ 
        КЦ
    КЦ

```

Доказательство корректности алгоритма 2.19. Для доказательства выясним смысл пометок вершин.

Лемма 2.10. Пометка $D(w)$ каждой обработанной вершины — кратчайшее расстояние от источника до вершины w . Пометка $D(u)$ каждой необработанной вершины — кратчайшее расстояние от источника до этой вершины на множестве путей, у которых все вершины, кроме последней, обработаны.

Доказательство. Докажем лемму по индукции. База индукции обеспечена инициализацией: источник — обработанная вершина, и расстояние до нее (равное 0) найдено, пометки других вершин определяются длинами ребер из источника в эти вершины, их можно рассматривать как пути из одного ребра, начальная вершина которого обработана, а конец нет.

Индукционный переход. Предположим, утверждение леммы верно для $m = k$, докажем, что после выполнения одного шага внешнего цикла оно будет верно для $m = k + 1$. Сначала докажем, что значение пометки необработанной вершины w с минимальной пометкой — это расстояние до нее. Докажем от противного. Пусть существует более короткий путь до w . Тогда часть его вершин обработана (по крайней мере, первая вершина), а часть — нет (по крайней мере, последняя). Рассмотрим первую необработанную вершину u (рис. 2.48). Длина пути от источника до этой вершины должна быть не больше длины пути от источника до вершины w (в силу неотрицательности ребер), но тогда пометка вершины u будет меньше пометки вершины w , что противоречит правилу выбора вершины w ! Противоречие доказывает утверждение. Теперь докажем справедливость утверждения леммы для необработанных вершин. Но оно следует непосредственно из алгоритма модификации — новая пометка строится так, чтобы она была минимальным расстоянием на множестве путей, у которых все вершины, кроме последней, обработаны. ■

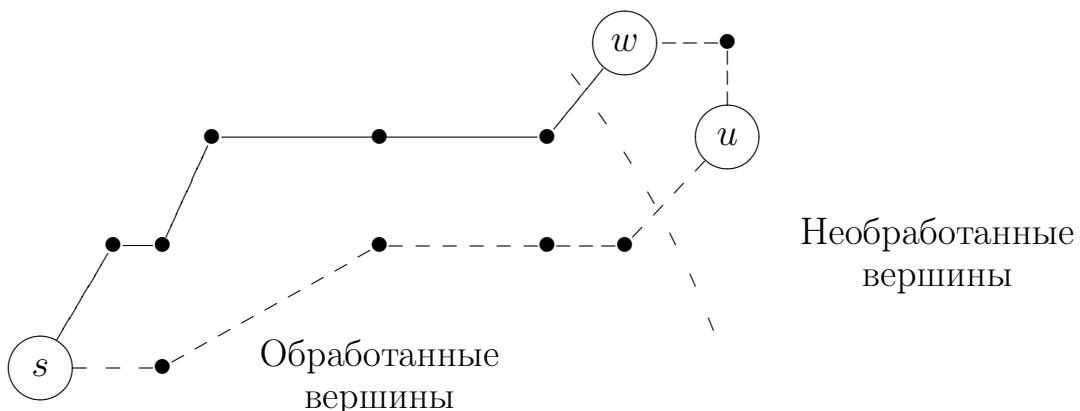


Рис. 2.48

Теорема 2.41. Алгоритм Дейкстры корректно решает задачу о нахождении кратчайших длин путей от источника до всех вершин графа.

Доказательство. Согласно лемме 2.10, на каждом шаге алгоритма одна вершина переходит в множество обработанных, следовательно, за конечное число шагов алгоритм свою работу заканчивает, а все вершины при этом оказываются обработанными. Из леммы также следует, что пометки обработанных вершин являются кратчайшими расстояниями до них. ■

Рассмотрим пример протокола работы алгоритма Дейкстры.

Пример 2.39. Найдем кратчайшие расстояния от вершины E (источника) связного неориентированного графа до остальных его вершин. Граф и его матрица весов изображены на рис. 2.49, a и b .

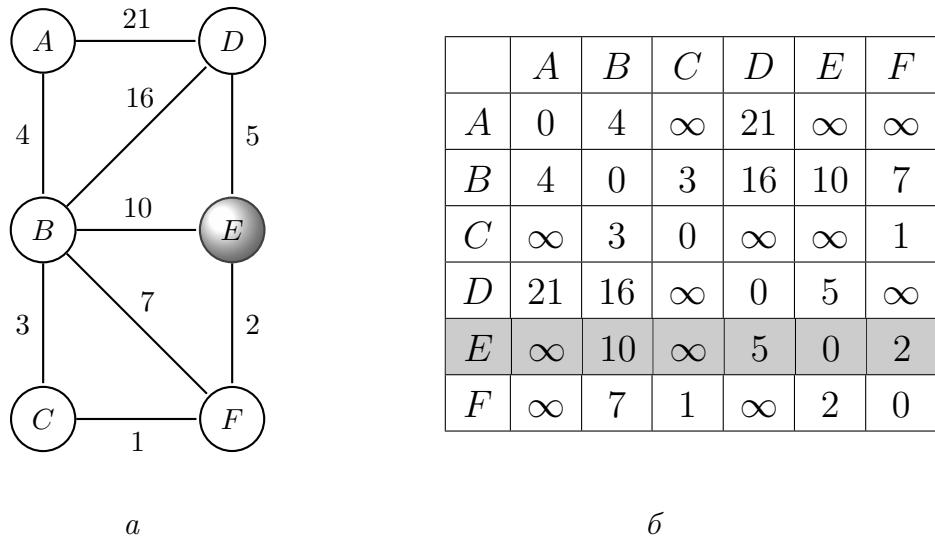


Рис. 2.49

Протокол работы алгоритма будем вести так же, как при «прогонке» алгоритма Форда–Беллмана (см. пример 2.38). Первая строка протокола — строка *E* матрицы весов, показывающая длины ребер, выходящих из *E*. (Вершина *E* уже «обработана», поэтому ее пометка не пересчитывается, что выделено серым цветом).

	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>
<i>E</i>	∞	10	∞	5	0	2

Минимальная из этих пометок длин — пометка вершины *F* — есть длина кратчайшего пути от *E* до *F*. Переведем эту вершину в разряд обработанных (что в протоколе выделим серым цветом) и пересчитаем пометки вершин, смежных с вершиной *F*.

Для этого обратим внимание на значения весов ребер в строке *F*:

$$D(F) + P(F; B) = 2 + 7 = 9 < 10,$$

поэтому пометка вершины *B* модифицируется на 9 через вершину *F*;

$$D(F) + P(F; C) = 2 + 1 = 3 < \infty,$$

поэтому пометка вершины *C* также модифицируется через *F*. Теперь выбираем минимальную из получившихся пометок — пометку вершины *C* и продолжаем до тех пор, пока все вершины не будут обработаны. Полностью протокол показан в табл. 2.17.

Числовые пометки обработанных вершин показывают кратчайшие расстояния до них, а символьные — предшествующие вершины в кратчайших путях (табл. 2.18). Таким образом, дерево кратчайших путей состоит из ребер: *BA, CB, FC, ED, EF*.

Таблица 2.17

Номер шага m	A	B	C	D	E	F
1	∞	$10/E$	∞	$5/E$	0	2/E
2	∞	$9/F$	3/F	$5/E$		
3	∞	$6/C$		5/E		
4	$26/D$	6/C				
5	10/B					

Таблица 2.18

A	B	C	D	E	F
$10/B$	$6/C$	$3/F$	$5/E$	0	$2/E$

Используя символьные пометки (табл. 2.18), мы можем получить и сами кратчайшие маршруты. Начиная от данной вершины нужно каждый раз брать символьную пометку от текущей вершины, пока не придем в стартовую вершину – так мы получим искомый кратчайший путь, но записанный в обратном порядке:

$$\begin{aligned} A &\rightarrow B \rightarrow C \rightarrow F \rightarrow E \\ B &\rightarrow C \rightarrow F \rightarrow E \\ C &\rightarrow F \rightarrow E \\ F &\rightarrow E \end{aligned}$$

Задание 2.3. Найдите дерево кратчайших путей из источника D . Убедитесь, что дерево в данном случае вырождается в путь, проходящий через все вершины графа по одному разу (*гамильтонов путь*).

2.2.7. Сравнительный анализ трудоемкости алгоритмов Форда–Беллмана и Дейкстры

При нахождении длин кратчайших путей каждый из описанных алгоритмов многократно проверяет возможность модификации вершины и, если надо, осуществляет ее. Посчитаем количество таких проверок в зависимости от n — числа вершин графа.

Алгоритм Форда–Беллмана повторяет $(n - 2)$ раза операцию, которая включает в себя попытку модификации каждой из вершин через все (соседние) вершины графа. (При задании графа матрицей весов отсутствуют специальные указатели на соседние вершины и приходится де-

лать проверку для всех вершин.) Таким образом, трудоемкость алгоритма Форда–Беллмана

$$T(n) = (n - 2)n \cdot n \sim n^3.$$

Здесь знак \sim означает асимптотическую эквивалентность, т. е. предел отношения соединенных этих знаком функций на бесконечности равен константе, отличной от нуля. Следовательно, главное в оценке — показатель степени n , или, в общем случае, характер роста функции трудоемкости. Использование асимптотических оценок трудоемкости оправдано, поскольку небольшие вариации алгоритма (в данном случае можно было бы исключить операции с источником и число сравнений сразу бы изменилось на $(n - 2)(n - 1)(n - 1)$) или подсчет иного множества операций (в данном случае, например, можно считать не только операции сравнения, но и арифметические операции) привели бы к изменению формулы трудоемкости, но не изменили бы порядок ее роста.

Итак, трудоемкость алгоритма Форда–Беллмана имеет порядок n^3 .

Алгоритм Дейкстры делает n шагов, связанных с обработкой вершин. В рамках каждого шага ищется минимальная пометка и через нее пересчитываются пометки соседних вершин. Поиск минимальной пометки в простейшем случае осуществляется перебором всех n вершин (на самом деле, на каждом новом шаге просматривать надо на одну вершину меньше), после чего просматриваются все (необработанные) вершины и их пометки (если надо) модифицируются. Нетрудно видеть, что порядок трудоемкости для алгоритма Дейкстры

$$T(n) \sim n^2.$$

Замечание 2.43. Поиск минимальной пометки можно ускорить за счет иного способа структурирования данных — на этом основаны различные модификации алгоритма Дейкстры. Понятно, что для таких алгоритмов и пересчет пометок надо организовать по-другому, храня информацию о соседях каждой вершины, иначе просмотр всех вершин при модификации «сведет на нет» все выгоды, полученные за счет быстрого нахождения минимума.

Итак, какой выигрыш во времени дает замена алгоритма Форда–Беллмана алгоритмом Дейкстры?

Для $n = 100$ получим стократный выигрыш, т. е., если программа по второму алгоритму решит задачу за 1 с, то по первому только за 100 с. Еще более заметен выигрыш для $n = 1000$, что для задач на графах — реальная ситуация. В этом случае, на компьютере, который решит задачу с помощью алгоритма Дейкстры за 1 с, с помощью алгоритма Форда–Беллмана ее придется решать более 15 мин.

2.2.8. Алгоритм Флойда

Алгоритм Флойда решает более общую задачу — нахождение всех кратчайших расстояний между всеми парами вершин графа. Но главное

достоинство этого алгоритма — простота. Как известно, с позиций надежности программирования «прозрачность» алгоритма часто имеет большее значение, чем его трудоемкость.

Постановка задачи. Найти кратчайшие расстояния между всеми парами вершин ориентированного или неориентированного взвешенного графа, не имеющего циклов отрицательной длины.

Идея алгоритма. Расширение множества промежуточных вершин для путей, по которым ищутся кратчайшие расстояния. Алгоритм начинает работу с матрицы весов, рассматриваемых как длины кратчайших путей, не имеющих ни одной промежуточной вершины. Далее поочередно каждую вершину представляют в роли промежуточной и через нее осуществляют модификацию пометки, т. е. на каждом шаге одну из вершин добавляют к множеству промежуточных вершин.

Опишем сам алгоритм.

Алгоритм 2.20 (алгоритм Флойда).

Инициализация:

```

ЦИКЛ ПО  $i$  от 1 до  $n$ 
    ЦИКЛ ПО  $j$  от 1 до  $n$ 
         $D(i; j) := P(i; j)$ 
    КЦ
КЦ

```

Здесь матрица пометок заполняется весами ребер.

Основной алгоритм:

```

ЦИКЛ ПО  $k$  от 1 до  $n$ 
    ЦИКЛ ПО  $i$  от 1 до  $n$ 
        ЦИКЛ ПО  $j$  от 1 до  $n$ 
             $D(i; j) := \min\{D(i; j); D(i; k) + D(k; j)\}$ 
        КЦ
    КЦ
КЦ

```

Здесь внешний цикл идет по промежуточным вершинам k , внутри него для каждой пары вершин проверяется возможность улучшения их пометки через k .

Обоснуем корректность приведенного алгоритма.

Лемма 2.11. После выполнения k -го шага (внешнего цикла) алгоритма Флойда пометки $D(i; j)$ — суть длины кратчайших путей, все промежуточные вершины которых принадлежат множеству $\{1, \dots, k\}$.

Доказательство. Докажем по индукции. Как уже отмечалось, база индукции обеспечена инициализацией пометок.

Индукционный переход. Предположим, что после шага k пометки $D(i; j)$ равны кратчайшим длинам путей между парами вершин i и j на множестве путей, у которых все промежуточные вершины принадлежат множеству $\{1, \dots, k\}$. Докажем, что после выполнения $(k+1)$ -го шага пометки будут равны длинам кратчайших путей с промежуточными вершинами из множества $\{1, \dots, k, k+1\}$. Действительно, каждый из таких путей либо не проходит через вершину $(k+1)$, и тогда его длина $D(i; j)$ нам известна, либо проходит, и тогда его длина будет равна

$$D(i; k+1) + D(k+1; j).$$

(Заметим, что каждая вершина может входить в кратчайший путь не более одного раза, так как граф не имеет циклов отрицательной длины). Модификация пометки организована так, чтобы из этих чисел выбрать меньшее. Таким образом, индукционный переход доказан. ■

Теорема 2.42. Алгоритм Флойда корректно решает задачу нахождения всех кратчайших расстояний между всеми парами вершин графа.

Доказательство. После окончания работы внешнего цикла все вершины попадут во множество промежуточных и по лемме 2.11 пометки $D(i; j)$ будут равны длинам кратчайших путей, у которых промежуточные вершины принадлежат множеству $\{1, \dots, n\}$, т. е. множеству всех вершин графа. ■

Рассмотрим пример работы алгоритма Флойда.

Пример 2.40. Применим алгоритм Флойда к графу, который рассматривался при обсуждении алгоритма Форда–Беллмана (см. рис. 2.46, пример 2.38).

Как уже отмечалось, полезно следить за работой алгоритма, не глядя на график. Поэтому не будем повторять рисунок графа, а протокол работы представим последовательными состояниями матриц пометок, которые получаются после выполнения очередного шага внешнего цикла (добавление новой промежуточной вершины). Начальное состояние матрицы пометок — матрица весов графа.

	A	B	C	D	E
A	0	2	∞	4	5
B	1	0	3	∞	4
C	∞	-2	0	2	1
D	-1	∞	-1	0	1
E	∞	∞	∞	∞	0

Рассмотрим A в качестве промежуточной вершины. Обратим внимание на начальные вершины ребер, входящих в A (столбец A), в дальнейшем назовем их входящими вершинами, и на конечные вершины ребер, выходящих из A (строка A). Например, в A можно попасть из B по пути «длиной» 1, а из A в D — по пути «длиной» 4. Итого, получаем путь длиной 5, что лучше старой пометки (∞).

Аналогичную операцию проводим с другими парами вершин:

$$\begin{aligned} B \text{ и } E : D(B; A) + D(A; E) &= 1 + 5 > 4, \\ D \text{ и } E : D(D; A) + D(A; E) &= -1 + 5 > 1, \\ D \text{ и } B : D(D; A) + D(A; B) &= -1 + 2 < \infty. \end{aligned}$$

Как видим, на этом шаге осуществляется еще одна модификация — пометки $D(D; B)$. На следующем состоянии матрицы пометок отметим клетки с модифицированными пометками. Обратим внимание на пометку пары одинаковых вершин:

$$D \text{ и } D : D(D; A) + D(A; D) = -1 + 4 > 0.$$

Если бы удалось улучшить такую пометку, это бы означало, что найден цикл отрицательной длины. Такая особенность алгоритма Флойда позволяет применять его для графов, у которых заранее неизвестно существование кратчайших путей. Если следить за числами на диагонали матрицы пометок, то появление на ней отрицательного числа означает, что найден цикл отрицательной длины, а значит, алгоритм не применим. Это другая замечательная особенность алгоритма Флойда — проверять собственную применимость.

Далее в качестве промежуточной вершины рассматриваем вершину B :

	A	B	C	D	E
A	0	2	∞	4	5
B	1	0	3	5	4
C	∞	-2	0	2	1
D	-1	1	-1	0	1
E	∞	∞	∞	∞	0

Жирным шрифтом выделены те пометки, которые будут модифицированы на этом шаге.

Следующей промежуточной вершиной будет вершина C .

	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>
<i>A</i>	0	2	5	4	5
<i>B</i>	1	0	3	5	4
<i>C</i>	-1	-2	0	2	1
<i>D</i>	-1	1	-1	0	1
<i>E</i>	∞	∞	∞	∞	0

На этом шаге модифицируются три пометки.

Следующий шаг — модификация через вершину *D*. На этом шаге модифицируются две пометки.

	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>
<i>A</i>	0	2	5	4	5
<i>B</i>	1	0	3	5	4
<i>C</i>	-1	-2	0	2	1
<i>D</i>	-2	-3	-1	0	0
<i>E</i>	∞	∞	∞	∞	0

На последнем шаге рассмотрим вершину *E* как промежуточную.

	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>
<i>A</i>	0	1	5	4	4
<i>B</i>	1	0	3	5	4
<i>C</i>	-1	-2	0	2	1
<i>D</i>	-2	-3	-1	0	1
<i>E</i>	∞	∞	∞	∞	0

На этом шаге модификаций пометок нет.

Рассмотрим итоговую таблицу (табл. 2.19), добавив в нее символьные пометки, обозначающие вершину, через которую произошла модификация пометки.

Как пользоваться символьными пометками для построения кратчайших путей?

Найдем кратчайший путь от *D* до *A*. Для этого определим промежуточную вершину на пути от *D* к *A*: $G(D; A) = C$. Теперь известно, что путь проходит через вершины *D, C, A*. Далее определим промежуточные вершины для подпутей: $G(D; C) = D$ и $G(C; A) = B$. То, что промежуточная вершина подпути *DC* совпала с его началом *D*, означает, что других промежуточных вершин на этом под пути нет. Рассмотрим промежуточные вершины подпутей *CB* и *BA*: $G(C; B) = C$, $G(B; A) = B$. Видно, что все промежуточные вершины найдены и кратчайший путь от *D* до *C*: *DCBA*.

Таблица 2.19

	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>
<i>A</i>	0	$1/D$	$3/B$	$4/A$	$4/D$
<i>B</i>	$1/B$	0	$3/B$	$5/A$	$4/B$
<i>C</i>	$-1/B$	$-2/C$	0	$2/C$	$1/C$
<i>D</i>	$-1/C$	$1/C$	$-1/D$	0	$1/C$
<i>E</i>	∞	∞	∞	∞	0

Алгоритм 2.21 (построение кратчайшего пути по пометкам G между вершинами S и F).

Вспомогательная рекурсивная процедура:

```

ПРОЦЕДУРА ShortPath( $X; Y$ )
ЕСЛИ  $G(X; Y) \neq X$  ТО
    вставить между  $X$  и  $Y$  вершину  $G(X; Y)$ 
    ShortPath( $X; G(X; Y)$ )
    ShortPath( $G(X; Y); Y$ )

```

Основная программа

```
ShortPath( $S; F$ )
```

Задание 2.4. Сравните алгоритм Уоршелла (алгоритм 2.29) для построения транзитивного замыкания и алгоритм Флойда. При каких значениях весов алгоритм Флойда превращается в алгоритм Уоршелла?

2.2.9. Волновой алгоритм

Рассмотрим еще одну специальную задачу поиска пути в графе – поиск кратчайшего пути в лабиринте. Для ее решения был предложен *волновой алгоритм (алгоритм Ли)*. Алгоритм основан на методе поиска в ширину. В основном используется при разводке печатных плат, поиске кратчайшего расстояния на карте в компьютерных играх.

Историческая справка. Волновой алгоритм для поиска пути в лабиринте был предложен Э. Ф. Муром. Ли независимо открыл этот же алгоритм при формализации задачи разводки печатных плат в 1961 году.

Алгоритм работает на ограниченной замкнутой линией фигуре (не обязательно прямоугольной), разбитой на прямоугольные ячейки, в частном случае – квадратные. Множество всех ячеек разбивается на подмножества: *проходимые* (свободные), т. е при поиске пути их можно проходить, *непроходимые* (препятствия), путь через эту ячейку запрещен, стартовая ячейка

(источник) и финишная (приемник) между которыми нужно найти кратчайший путь, если это возможно, либо, при отсутствии пути, выдать сообщение о непроходимости.

Определение 2.45. Соседними ячейками для данной ячейки по фон Нейману считаются только 4 ячейки по вертикали и горизонтали, по Муру – все 8 ячеек, включая диагональные.

Работа алгоритма включает в себя два этапа: распространение волны и восстановление пути.

Алгоритм 2.22 (Распространение волны).

Шаг 1. Помечаем стартовую ячейку $d = 0$.

Шаг 2. Для каждой ячейки, помеченной числом d помечаем все соседние (по Муру или по фон Нейману) свободные непомеченные ячейки числом $d + 1$. Шаг 2 повторяется до тех пор, пока не помечена финишная ячейка и есть возможность помечать новые ячейки.

Если финишная ячейка не помечена, то пути не существует.

Восстановление кратчайшего пути происходит в обратном направлении: при выборе ячейки от финишной ячейки к стартовой на каждом шаге выбирается ячейка, имеющая расстояние от стартовой (d) на единицу меньше текущей ячейки. Очевидно, что таким образом находится кратчайший путь. Некоторые модификации волнового алгоритма предполагают сохранение информации о том, из какой вершины волна перешла в данную, что ускоряет генерацию пути, но занимает больше памяти.

Замечание 2.44. Путей с минимальной числовой длиной пути, как при поиске пути по Муру, так и по фон Нейману может существовать несколько. Выбор окончательного пути в приложениях диктуется другими соображениями, находящимися вне этого алгоритма. Однако, длина пути в любом случае остается одной, вне зависимости от выбранной стратегии.

Преимущества волнового алгоритма в том, что с его помощью можно найти трассу в любом лабиринте и с любым количеством препятствий. Единственным недостатком волнового алгоритма является, то, что при построении пути требуется большой объем памяти.

Пример 2.41. Рассмотрим работу алгоритма на примере лабиринта на рис. 2.50.

На рис. 2.51 показан этап распространения волны. Заметим, что в данном примере существует несколько кратчайших путей.

Пример 2.42. В этом примере (рис. 2.52 и рис. 2.53) такой путь единственный. Отмечен красным цветом.

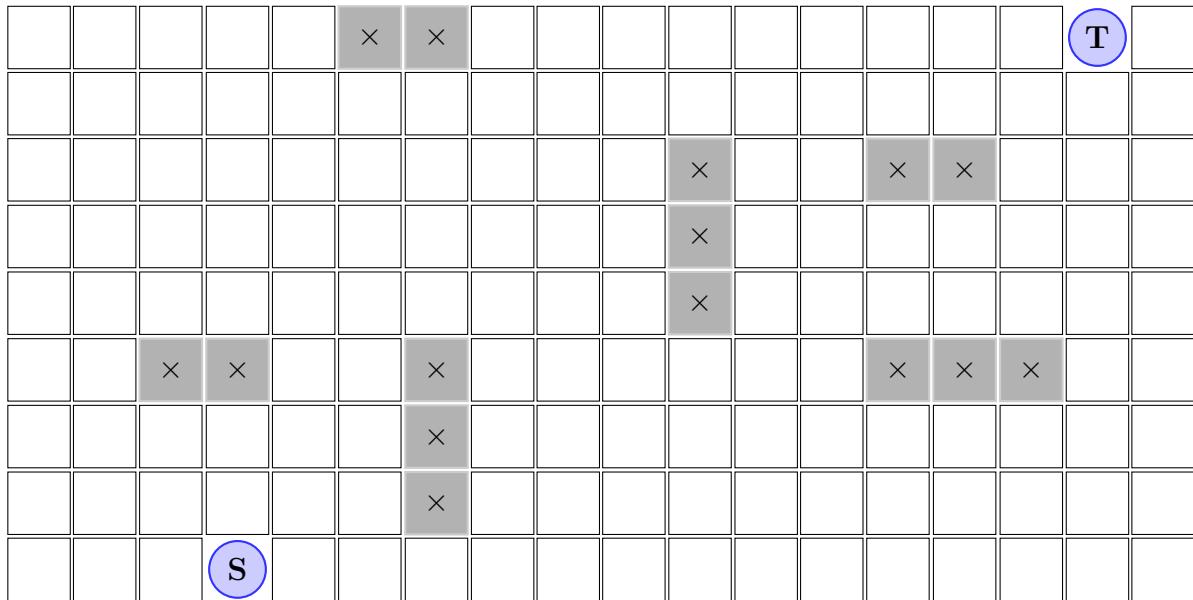


Рис. 2.50

11	10	11	10	9	x	x	12	13	14	15	16	17	18	19	20	T	
10	9	10	9	8	9	10	11	12	13	14	15	16	17	18	19	20	21
9	8	9	8	7	8	9	10	11	12	x	14	15	x	x	18	19	20
8	7	8	7	6	7	8	9	10	11	x	13	14	15	16	17	18	19
7	6	7	6	5	6	7	8	9	10	x	12	13	14	15	16	17	18
6	5	x	x	4	5	x	7	8	9	10	11	12	x	x	x	16	17
5	4	3	2	3	4	x	6	7	8	9	10	11	12	13	14	15	16
4	3	2	1	2	3	x	5	6	7	8	9	10	11	12	13	14	15
3	2	1	S	1	2	3	4	5	6	7	8	9	10	11	12	13	14

Рис. 2.51

2.2.10. Алгоритм Форда–Фалкерсона

Историческая справка. Впервые задача о максимальном потоке была решена Джорджем Данцигом в ходе подготовки воздушного моста во время блокады Западного Берлина, происходившей в 1948–1949 году. Позднее, в 1951 году он сформулировал задачу в общем виде.

В 1955 году, Лестер Форд и Делберт Фалкерсон первыми разработали алгоритм, специально предназначенный для решения этой задачи. Он получил название *алгоритм Форда–Фалкерсона*. В дальнейшем решение задачи многократно улучшалось различными авторами.

Определение 2.46. Транспортной сетью (или просто сетью) S называют ориентированный взвешенный граф \mathbf{G} с положительными весами

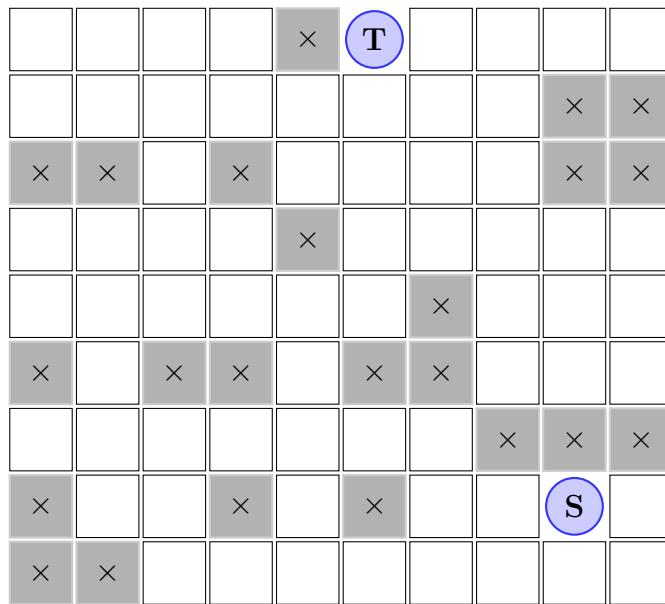


Рис. 2.52

				x	T			
		12		12	11	12	x	x
x	x	11	x	11	10	11	12	x
12	11	10	9	x	9	10	11	12
11	10	9	8	7	8	x	12	
x	9	x	x	6	x	x		
9	8	7	6	5	4	3	x	x
x	9	8	x	6	x	2	1	S
x	x	7	6	5	4	3	2	1

Рис. 2.53

ребер, в котором существует ровно одна вершина s (*источник*), в которую не входит ни одно ребро, и ровно одна вершина t (*сток*), из которой не выходит ни одного ребра.

Вес ребра (u, v) будем обозначать $c(u, v)$.

Смысл весов ребер в данной задаче — пропускные способности ребер.

Пример 2.43. Для сети автомобильных дорог пропускная способность ребра — количество автомобилей, проезжающих в единицу времени, для электрической сети — проводимость (величина обратная сопротивлению) участка цепи.

Определение 2.47. Потоком в сети называют функцию, сопоставляющую каждому ребру неотрицательное число (величину потока через ребро) так, чтобы:

- 1) оно не превышало пропускной способности ребра;
- 2) для каждой промежуточной вершины (т. е. любой вершины, кроме s и t) выполнялось следующее условие: сумма потоков по входящим в вершину ребрам была равна сумме потоков по исходящим из нее ребрам.

Более формально, функция $f : E \rightarrow \mathbb{R}_+$, определенная на множестве ребер графа $G = (V, E)$, называется потоком в сети S , если она удовлетворяет следующим условиям:

- 1) для любого ребра $(u, v) : 0 \leq f(u, v) \leq c(u, v);$
- 2) для любой вершины $v \in V \setminus \{s, t\} :$

$$\sum_{u \in V | (v, u) \in E} f(v, u) - \sum_{u \in V | (u, v) \in E} f(u, v) = 0. \quad (2.7)$$

Пример 2.44. На рис. 2.54 показан пример транспортной сети и нескольких потоков на ней (веса ребер — пропускные способности ребер, а величины в скобках — значения потоков через ребра). Заметим, что ребро (u, v) неориентированное. Его можно заменить двумя по-разному ориентированными ребрами с одинаковыми пропускными способностями.

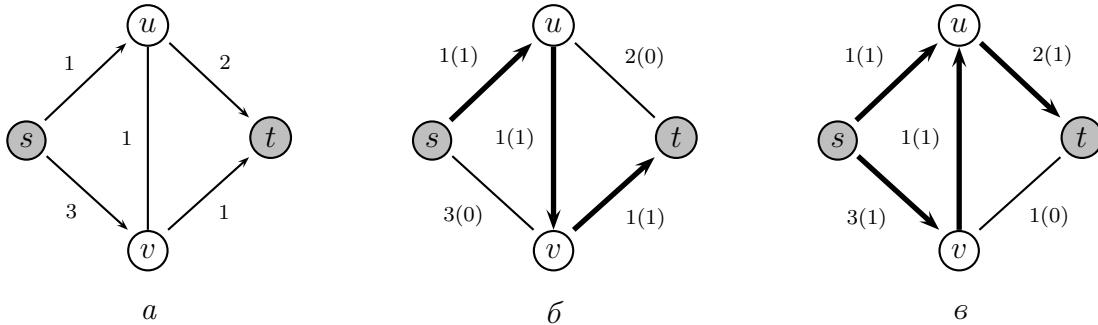


Рис. 2.54

Определение 2.48. Разность (2.7) суммарных потоков через входящие и выходящие в вершину v ребра называют дивергенцией потока в данной вершине и обозначают $\text{div}(f, v)$.

Равенство дивергенции нулю в данной вершине, означает, что она не является ни источником, ни стоком.

Пример 2.45. Равенство дивергенции нулю в промежуточных вершинах транспортной сети для автомобильных потоков означает невозможность приостановления движения и отправки новых автомобилей с промежуточных пунктов маршрута.

Упражнение 2.7. Вычислите $\text{div}(s)$ и $\text{div}(t)$.

Определение 2.49. $s - t$ разрезом транспортной сети P называется множество ребер, соединяющих вершины множеств V_1 и V_2 , где V_1, V_2 – разбиение множества вершин \mathbf{V} графа на два непересекающихся подмножества, обладающее свойством: $s \in V_1$ и $t \in V_2$.

Более формально, множество ребер P графа \mathbf{G} называется $s - t$ разрезом, если оно состоит из всех ребер (u, v) , таких что:

$$u \in V_1, v \in V_2, \text{ или } u \in V_2, v \in V_1, \text{ где } \mathbf{V} = V_1 \cup V_2, V_1 \cap V_2 = \emptyset, s \in V_1, t \in V_2.$$

В дальнейшем, где это не приводит к неоднозначности, $s - t$ разрез будем называть разрезом. Ребра разреза P с началом в V_1 обозначим P^+ , а с началом в V_2 – P^- .

Определение 2.50. Пропускной способностью $C(P)$ разреза P назовем сумму пропускных способностей ребер, образующих разрез:

$$C(P) = \sum_{(u,v) \in P} c(u, v).$$

Обозначим $F(E_1)$ сумму потоков ребер для данного подмножества ребер E_1 :

$$F(E_1) = \sum_{(u,v) \in E_1} f(u, v).$$

Определение 2.51. Величиной потока $w(P)$ через разрез P назовем разность $F(P^+) - F(P^-)$.

Пример 2.46. Рассмотрим разрез $V_1 = \{s\}$, $V_2 = \mathbf{V} \setminus \{s\}$. Величина потока через этот разрез равна сумме потоков через исходящие из начальной вершины ребра.

Упражнение 2.8. Вычислите величину потока через разрез

$$V_1 = \mathbf{V} \setminus \{t\}, V_2 = \{t\}.$$

Теорема 2.43. Величина потока через разрез не превышает пропускной способности разреза.

Доказательство.

$$\begin{aligned} w(P) &= F(P^+) - F(P^-) \leq F(P^+) = \sum_{(u,v) \in P^+} f(u, v) \leq \sum_{(u,v) \in P} f(u, v) \leq \\ &\leq \sum_{(u,v) \in P} c(u, v) = C(P) \end{aligned}$$

■

Теорема 2.44. Величина потока через разрез не зависит от выбора разреза.

Доказательство. Рассмотрим произвольный разрез графа P , определяемый разбиением \mathbf{V} на V_1 и V_2 и просуммируем дивергенции по вершинам V_1 . По определению потока дивергенции в промежуточных вершинах равны нулю, поэтому

$$\sum_{v \in V_1} \text{div}(v) = \text{div}(s).$$

Вычислим эту сумму по-другому, используя определение дивергенции.

$$\begin{aligned} \sum_{v \in V_1} \text{div}(v) &= \sum_{v \in V_1} \left(\sum_{u|(u,v) \in \mathbf{E}} f(u, v) - \sum_{u|(v,u) \in \mathbf{E}} f(v, u) \right) = \\ &= \sum_{v \in V_1} \sum_{u|(u,v) \in \mathbf{E}} f(u, v) - \sum_{v \in V_1} \sum_{u|(v,u) \in \mathbf{E}} f(v, u). \end{aligned}$$

Заметим, что двойное суммирование можно рассматривать, как суммирование по ребрам, а каждое ребро, концом которого является промежуточная вершина, входит один раз в первую сумму и один раз во вторую. Таким образом, после упрощения в сумме остаются только члены, соответствующие ребрам разреза:

$$\sum_{v \in V_1} \sum_{u \in V_2} f(u, v) - \sum_{v \in V_1} \sum_{u \in V_2} f(v, u) = F(P^+) - F(P^-) = F(P).$$

Таким образом, для любого разреза P величина $F(P)$ потока через разрез равна $\text{div}(s)$. ■

Определение 2.52. Величину $\text{div}(s)$ называют величиной потока P .

Пример 2.47. В терминах сети автомобильных дорог это определение означает, что величина потока определяется как количество автомобилей, выезжающих из начального пункта транспортной сети в единицу времени.

Лемма 2.12. Максимальная величина потока не превышает минимальной пропускной способности разреза.

Доказательство. Величина потока равна величине потока через любой разрез, значит, величина любого потока не превышает минимальной пропускной способности разреза, это верно и для максимального потока. ■

Можно предположить, что верно и более сильное утверждение.

Теорема 2.45 (теорема Форда–Фалкерсона). Максимальная величина потока равна минимальной пропускной способности разреза.

Замечание 2.45. Теорему 2.45 будем доказывать для сетей с целочисленных пропускными способностями ребер.

Лемма 2.13 (об увеличивающей цепи). Если величина потока меньше минимальной пропускной способности разреза, то существует цепь от начальной вершины до конечной, обладающая свойством: величина потока через каждое ребро этой цепи

- a) меньше пропускной способности этого ребра (такие ребра называются ненасыщенными), если направление ребра согласуется с направлением движения от начальной вершины к конечной и
- б) ненулевая, если направление ребра противоположное.

Доказательство. Докажем от противного. Пусть такой цепи не существует. Введем пометки ребер. Сначала пометим вершину s . Далее будем помечать вершины графа следующим образом: если выходящие из помеченных вершин ребра не насыщены, то концы этих ребер помечаем, также помечаем начальные вершины всех входящих ребер с ненулевым потоком.

Вершина t по предположению не должна оказаться помеченной. Тогда, если обозначить V_1 — множество помеченных вершин, а $V_2 = \mathbf{V} \setminus V_1$ — множество не помеченных, то ребра, соединяющие V_1 с V_2 определят разрез сети. Этот разрез не содержит ни одного ненасыщенного ребра, идущего от V_1 в V_2 и ни одного ребра, ведущего из V_2 в V_1 с положительной величиной потока, значит, поток через этот разрез совпадает с его пропускной способностью, что противоречит условию и доказывает лемму. ■

Дадим конструктивное доказательство теоремы 2.45, которое называют *алгоритмом Форда–Фалкерсона*. В качестве начального потока возьмем нулевой. Если этот поток меньше минимальной пропускной способности разреза, то по лемме 2.13 найдем увеличивающую цепь. Вдоль этой цепи величину потока можно увеличить на 1, добавляя к величинам потока по согласованным ребрам 1 и вычитая по 1 из остальных ребер. Так будем повторять до тех пор, пока величина потока не достигнет минимальной пропускной способности разреза. Процесс обязательно остановится, так как величина потока на каждом шаге увеличивается на 1.

Замечание 2.46. Перенос доказательства на случай вещественных пропускных способностей предполагает увеличение потока вдоль улучшающего пути на величину, равную минимальной разности между пропускными способностями ребер и величинами потока через ребра этого пути. Эта величина на каждом шаге может уменьшаться, что не гарантирует в общем случае сходимость процесса.

Сформулируем алгоритм Форда–Фалкерсона более формально.

Алгоритм 2.23 (алгоритм Форда–Фалкерсона).

{инициализация потока нулевыми значениями}

ЦИКЛ по всем ребрам графа $(u, v) \in E$ $f(u, v) := 0$

КЦ

{пометки вершин для нахождения пути, вдоль которого можно увеличить поток;

 s — начальная вершина, t — конечная вершина, $m(v)$ — пометка вершины v }

ЦИКЛ

 $m(s) := s; V_1 := \{s\}; V_2 := V \setminus V_1$ ЦИКЛ по вершинам $u \in V_1$ ЦИКЛ по вершинам $w \in V_2$, смежным с u ЕСЛИ $f(u, w) < c(u, w)$ ИЛИ $f(w, u) > 0$ ТО $m(w) := u; V_1 := V_1 \cup \{w\}; V_2 := V_2 \setminus \{w\}$

КЕ

КЦ

КЦ

{увеличение потока на 1}

ЕСЛИ $t \in V_1$ ТО $v := t$ ЦИКЛ-ПОКА $v \neq t$ $z := v; v := m(v)$ ЕСЛИ $f(v, z) < c(v, z)$ ТО $f(v, z) := f(v, z) + 1$

ИНАЧЕ

ЕСЛИ $f(z, v) > 0$ ТО $f(z, v) := f(z, v) - 1$

КЦ

КЕ

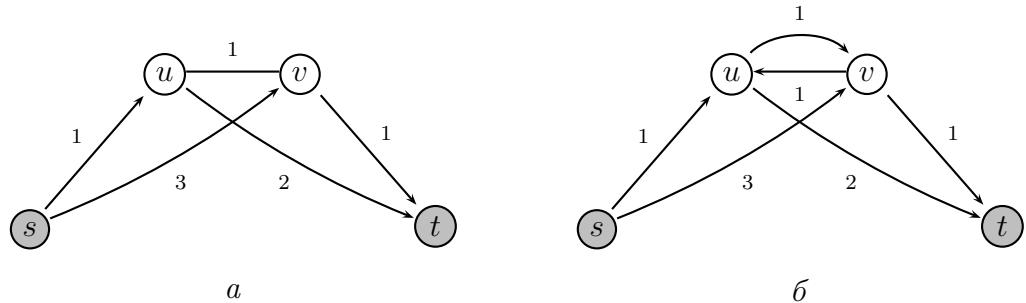
ДО ТЕХ ПОР ПОКА $t \notin V_1$

{вычисление максимального потока}

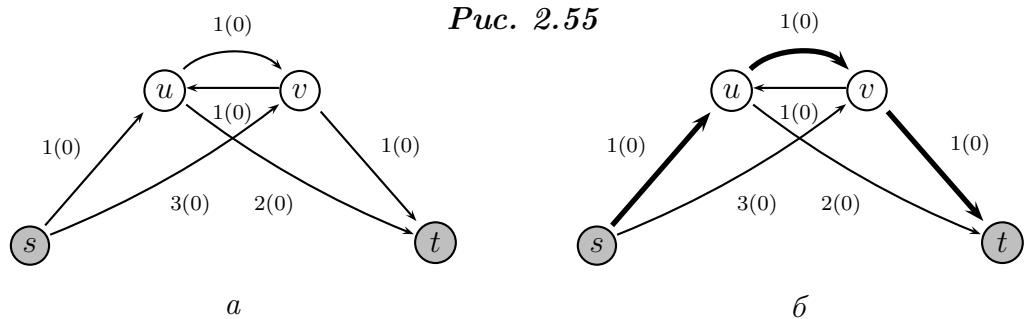
 $F := 0$ ЦИКЛ по вершинам $w \in V$ смежным с s $F := F + f(s, w)$

КЦ

Пример 2.48. На рис. 2.55 – 2.58 показана работа алгоритма Форда–Фалкерсона для транспортной сети с источником s и стоком t .Неориентированное ребро (u, v) на рис. 2.55,*a* заменяется двумя ориентированными ребрами (см. рис. 2.55,*b*).Поток F инициализируется нулевыми значениями, далее осуществляется постановка пометок и выбор увеличивающей цепи.Величина потока $F = 1$. Показаны два шага построения улучшающей цепи и увеличения потока.Величина потока F становится равной 2, затем 3. На очередном шаге пометить конечную вершину t не удается. Значит, найден максимальный



Puc. 2.55



Puc. 2.56

поток. Пометки вершин позволяют построить минимальный разрез: помеченные вершины s и v образуют множество V_1 , а непомеченные u и t — множество V_2 .

Ребра (s, u) , (u, v) и (v, t) образуют минимальный разрез. Его пропускная способность равна 3.

Задача 2.3. Докажите, что алгоритм 2.24 решает задачу нахождения максимального потока плоской транспортной сети.

Идея алгоритма в том, чтобы на каждом шаге находить самый «верхний» из возможных путей и загружать его как можно больше.

Алгоритм 2.24 (нахождения максимальной транспортной сети).

ПОЛУЧАЕМ величину потока T равной нулю.

находим самый верхний путь и пропускаем через него поток, величина которого равна минимальной пропускной способности ребра на этом пути:

уменьшаем пропускные способности ребер найденного пути на величину C ; удаляем из графа ребра с нулевой пропускной способностью:

$$F := F + C$$

KII

Пример 2.49. Применим данный алгоритм к транспортной сети из примера 2.48, изобразив ее в виде плоского графа. На рис. 2.59 показан протокол работы алгоритма для плоской транспортной сети с источником s и стоком t :

Коментарии к работе алгоритма приведены в табл. 2.20.

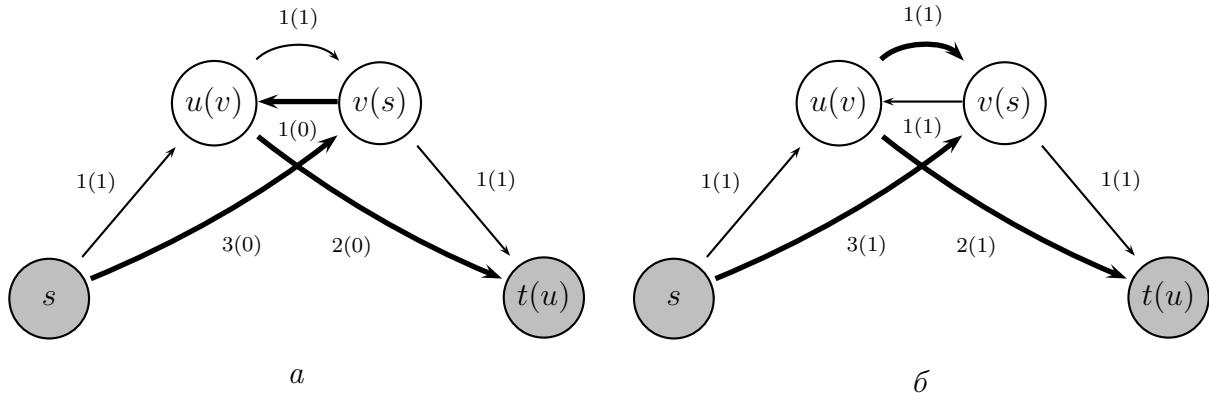


Рис. 2.57

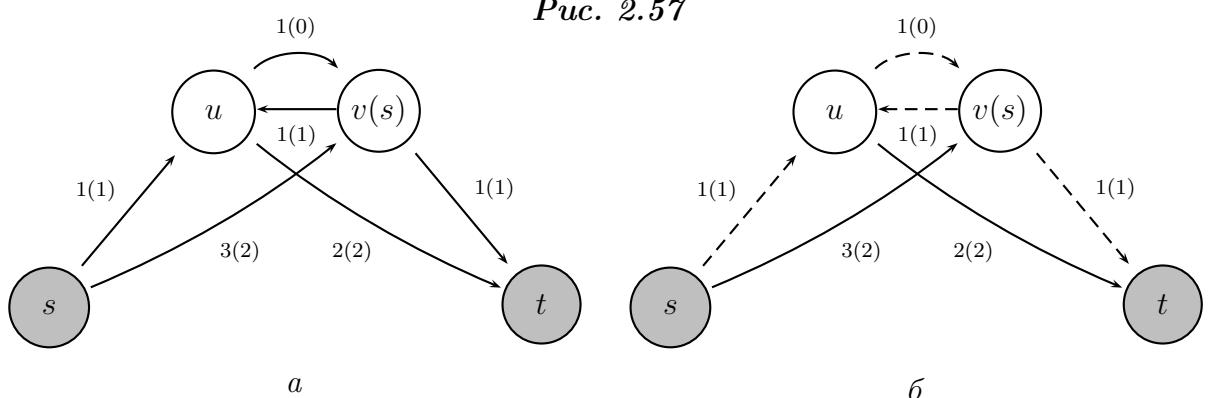


Рис. 2.58

Таблица 2.20

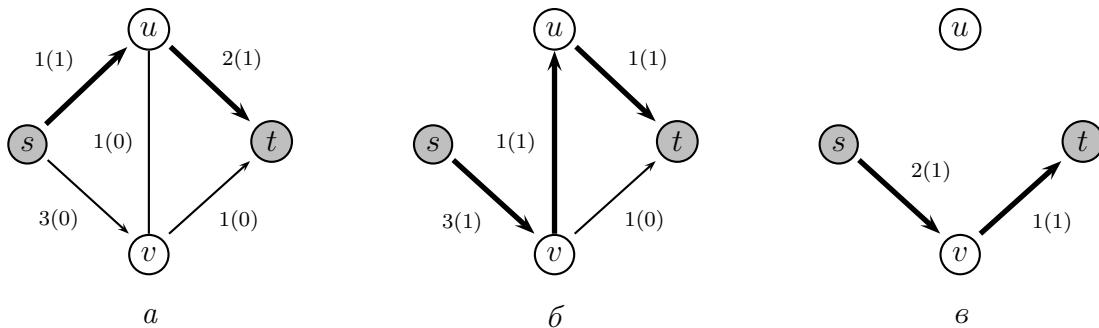
Шаг	Иллюстрация	Самый верхний путь	Значение C	Значение F
1	рис. 2.59, а	$s \rightarrow u \rightarrow t$	1	1
2	рис. 2.59, б	$s \rightarrow v \rightarrow u \rightarrow t$	1	2
3	рис. 2.59, в	$s \rightarrow v \rightarrow t$	1	3

2.2.11. Перебор с возвратом

Для множества задач дискретной математики эффективных алгоритмов не найдено и, скорее всего, не существует. Рассмотрим структуру переборных алгоритмов с возвратом, осуществляющих последовательный перебор всех допустимых комбинаций. Аналогичные перечислительные задачи были разобраны в комбинаторике. Однако, если там структура перебираемого множества была определена заранее, то здесь она определяется динамически с помощью функций расширения частичных решений.

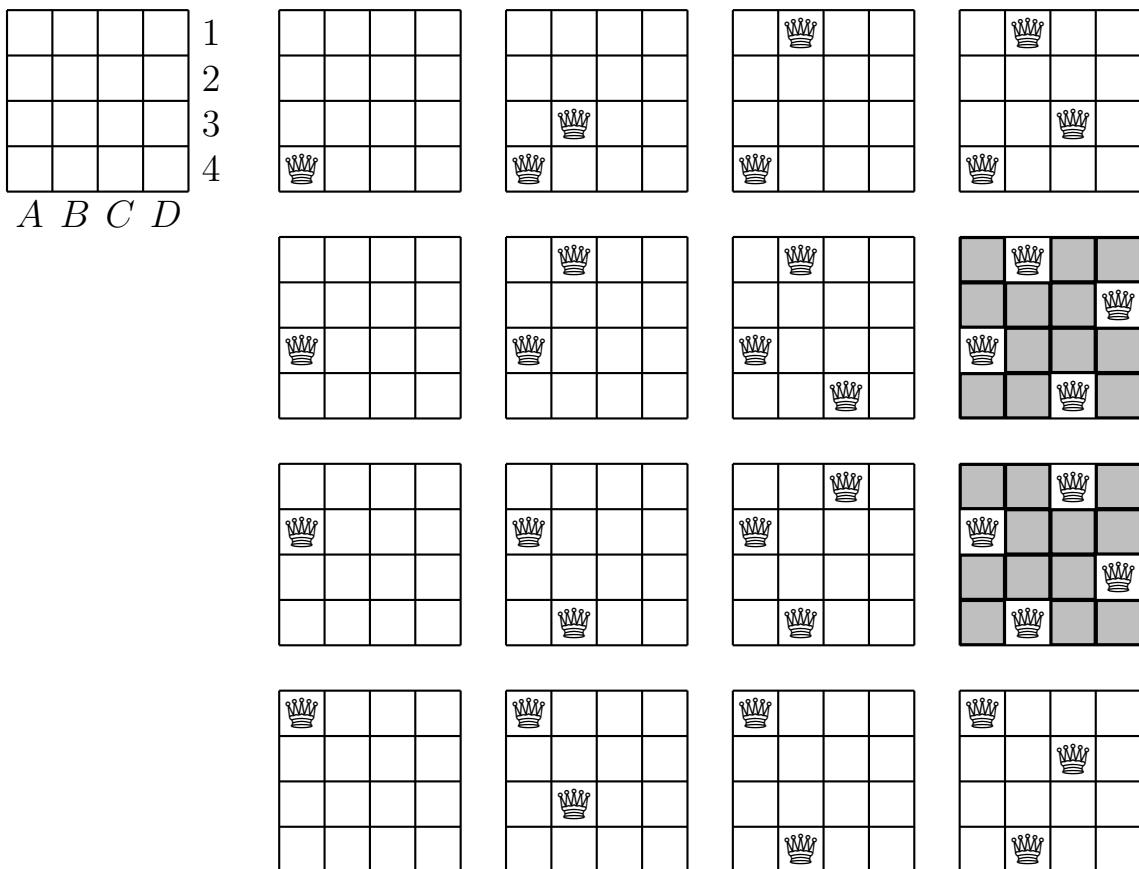
Приведем основные понятия, связанные с алгоритмом перебора с возвратом на примере известной задачи Эйлера.

Задача 2.4. Расставить на шахматной доске восемь ферзей так, чтобы они не били друг друга.



Puc. 2.59

Для простоты рассмотрим доску размера 4×4 и, соответственно, четыре ферзя (рис. 2.60).



Puc. 2.60

Для осуществления перебора заметим, что каждый ферзь должен стоять на отдельной вертикали. Поставим 1-го ферзя на первую клетку первой вертикали и запомним частичное решение как $(A1)$. Теперь будем пытаться поставить 2-го ферзя на вторую вертикаль так, чтобы он и 1-й не были друг друга. Первая и вторая позиции вертикали не подходят, поэтому ферзя ставим на $B3$. Полученное частичное решение запомним как $(A1; B3)$. Попробуем расширить полученное частичное решение, но все клетки третьей вертикали находятся под боем поставленных ферзей. Таким образом,

мы зашли в тупик.

В этом случае, используя хранящееся частичное решение, делаем шаг назад и пробуем следующее расширение первого частичного решения: $(A1; B4)$. Можно сделать еще один шаг: $(A1; B4; C2)$, однако дальнейшее расширение невозможно. Возвращаемся на шаг назад, но других расширений для $(A1; B4)$ нет. Делаем еще шаг назад, но у $(A1)$ также нет других расширений, поэтому делаем еще шаг назад и меняем первый ход на $(A2)$. Весь перебор представлен в табл. 2.21.

Таблица 2.21

$(A1)$	$(A1; B3)$	$(A1; B4)$	$(A1; B4; C2)$
$(A2)$	$(A2; B4)$	$(A2; B4; C1)$	$(A2; B4; C1; D3)$
$(A3)$	$(A3; B1)$	$(A3; B1; C4)$	$(A3; B1; C4; D2)$
$(A4)$	$(A4; B1)$	$(A4; B1; C3)$	$(A4; B2)$

На рис. 2.60 показаны последовательные положения фигур на шахматной доске.

На рис. 2.61 изображено дерево решений, которое строится динамически в процессе перебора с возвратом. Ход решения можно представить как обход этого дерева в глубину.

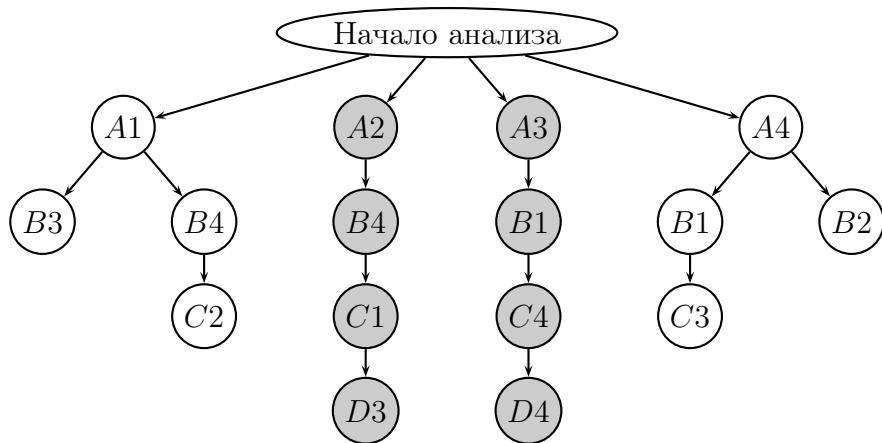


Рис. 2.61

Сформулируем алгоритм в общем виде.

Пусть множество вариантов, которые надо перебрать, описывается множеством упорядоченных наборов $x = (a_1, a_2, \dots, a_k)$. Каждый следующий элемент набора a_{k+1} определяется выбором из множества $\text{ext}(x)$, которое динамически вычисляется с помощью заданной функции расширения (ext). Расширение набора будем обозначать операцией конкатенации

(приписывание символа к строке):

$$x = (a_1, a_2, \dots, a_k), \quad a_{k+1} \in \text{ext}(x) \Rightarrow x \circ a_{k+1} = (a_1, a_2, \dots, a_k, a_{k+1}).$$

Такие упорядоченные наборы будем называть частичными решениями задачи. Для всех частичных решений $x \in X$ определена функция $\text{sol}(x)$, которая устанавливает, является ли данное частичное решение решением задачи.

Множество всех решений обозначим P .

Алгоритм 2.25 (рекурсивный алгоритм «перебор с возвратом»).

```

 $x \leftarrow \Lambda$                                  $\triangleright$  начальное значение частичного решения — пустой набор
 $P \leftarrow \emptyset$                              $\triangleright P$  — множество решений
function backtrack( $x, P$ )
    if  $\text{sol}(x)$  then
         $P \leftarrow P \cup \{x\}$ 
    else
        while  $y \in \text{ext}(x)$  do                 $\triangleright$  цикл по всем расширениям  $x$ 
            backtrack( $x \circ y, P$ )                   $\triangleright$  расширение частичного решения  $x$ 
        end while
    end if
end function

```

Замечание 2.47. Частичное решение имеет структуру стека: расширение его происходит дописыванием нового элемента справа, а возврат к предыдущим частичным решениям — удалением элементов справа. При выполнении рекурсивного алгоритма создается стек вызовов, который взаимно-однозначно соответствует частичному решению.

Задача 2.5. Используя перебор с возвратом, найдите гамильтонов путь (путь, проходящий через все вершины графа ровно по одному разу, см. определение 2.21) с началом в вершине A для графа на рис. 2.62.

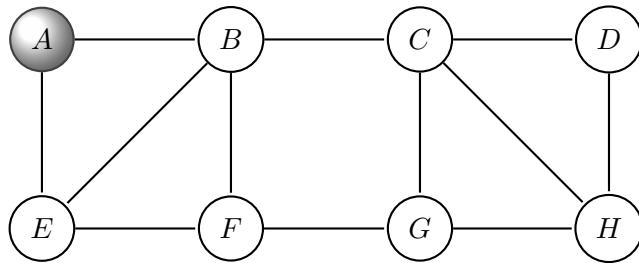


Рис. 2.62

Замечание 2.48. Алгоритм 2.25 на самом деле описывает группу алгоритмов. Действительно, функция $\text{ext}(x)$, определяющая порядок перебора частичных решений, может быть выбрана по-разному. Особенно важен правильный выбор функции $\text{ext}(x)$, если нас интересует любое (первое встретившееся) решение. В этом случае полезны различные эмпирические соображения для выбора очередного расширения. В частности, здесь могут использоваться «жадные» алгоритмы.

Пример 2.50. Рассмотрим популярную задачу, связанную с поиском гамильтонова пути в графе. Требуется обойти конем все клетки шахматной доски заданной формы, пройдя каждую ровно по одному разу.

Существует следующий алгоритм для выбора расширения частичного решения: нужно всегда первой идти на ту клетку, из которой существует минимум ходов (стратегия «идти в тупик»). Известно, что для многих конфигураций шахматных досок этот алгоритм дает решение. В общем же случае, очевидно, это оправданный способ упорядочивания частичных решений.

2.2.12. Метод ветвей и границ

Пусть на перебираемом множестве (множестве решений) P задана целевая вещественная функция f . Рассмотрим задачу нахождения элемента множества, для которого значение функции f минимально. В этом случае объем перебора можно существенно уменьшить, если ввести на множестве частичных решений оценочную функцию $b(x)$. Функция $b(x)$ подбирается так, чтобы давать нижнюю оценку для значения целевой функции на решении, получающимся последовательным расширением данного частичного решения.

Определение 2.53. Функцию b называют *оценочной* для целевой функции f , если $b(x) \leq f(p)$ для любого решения p и любого частичного решения x , приводящего к p , т. е., если существует последовательность расширений y_1, y_2, \dots, y_k , такая, что

$$p = x \circ y_1 \circ y_2 \circ \cdots \circ y_k, \text{ где } y_i \in \text{ext}(x \circ y_1 \circ y_2 \circ \cdots \circ y_{i-1}).$$

Пример 2.51. Рассмотрим задачу нахождения минимального гамильтонова пути в неориентированном графе с неотрицательными весами ребер — задачу коммивояжера. Определим целевую функцию f , как сумму весов ребер, входящих в гамильтонов путь.

Рассмотрим один из вариантов (не лучший) выбора частичного решения и оценочной функции. В качестве частичного решения возьмем простой путь в графе, расширения частичного решения — ребра, которыми можно дополнить путь, а в качестве функции b для частичного решения — сумму весов всех ребер этого пути. Очевидно, функция будет оценочной, так как для построения гамильтонова пути по предложенной схеме на каждом шаге добавляются ребра неотрицательной длины.

Модернизируем алгоритм 2.25 перебора с возвратом для получения алгоритма, реализующего метод ветвей и границ (branches and bounds).

Алгоритм 2.26 (рекурсивный алгоритм для метода ветвей и границ).

```

 $x := \emptyset$                                  $\triangleright$  начальное значение частичного решения — пустой набор
 $P \leftarrow \emptyset$                            $\triangleright P$  — множество решений
 $r \leftarrow \infty$      $\triangleright r$  — рекорд, минимальное значение целевой функции на найденных решениях
function bandb( $x, P, r$ )
  if sol( $x$ ) then
    if  $f(x) < r$  then                                 $\triangleright$  изменение рекорда
       $r \leftarrow f(x)$ 
       $P \leftarrow \{x\}$ 
    else
      if  $f(x) = r$  then                             $\triangleright$  сохранение нового решения
         $P \leftarrow P \cup \{x\}$ 
      end if
    end if
    return
  else
    if  $b(x) \leq r$  then                       $\triangleright$  отсечение ветвей дерева перебора
      while  $y \in \text{ext}(x)$  do                   $\triangleright$  цикл по всем расширениям  $x$ 
        bandb( $x \circ y, P, r$ )
      end while
    end if
  end if
end function

```

Пример 2.52. Найдем по предложенной выше схеме кратчайший гамильтонов путь в графе на рис. 2.63. Построение начнем с вершины A .

Порядок в каждом множестве $\text{ext}(x)$ определим как лексикографический (по вершинам ребер, выходящих из конца построенного пути).

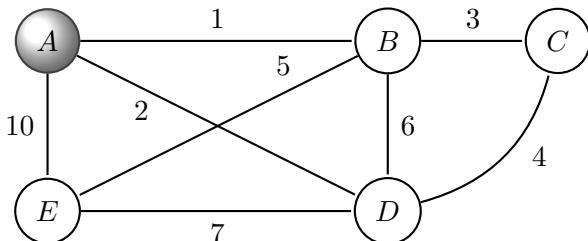


Рис. 2.63

Первым будет построен гамильтонов путь $ABCDE$ (четырьмя последовательными расширениями пустого решения: AB, BC, CD, DE). При этом значение рекорда будет $r = 15$.

Следующие частичные пути будут $ABDC, ABDE$, которые дальше не удастся продолжить, не проходя повторно через некоторые вершины. Затем будет найден второй гамильтонов путь $ABEDC$, который не будет сохранен, так как значение целевой функции для него $f(ABEDC) = 17 > r$. Далее просмотрим тупиковые пути DBC и $ADBE$, а затем найдем третий гамильтонов путь $ADCBE$, который имеет меньшую длину, поэтому рекорд поменяется: $r = 14$.

Далее будет опять найден гамильтонов путь с $f(ADEBC) = 17$, кото-

рый не будет сохранен. Заметим, что если бы было известно, что в графе нет ребер нулевого веса (или если нас не интересуют все возможные минимальные решения), то в алгоритме нестрогое неравенство $b(x) \leq r$ можно было заменить на строгое $b(x) < r$. В этом случае при рассмотрении частичного решения $ADEB$, значение $b(ADEB) = 14$ и дальнейший перебор не проводится.

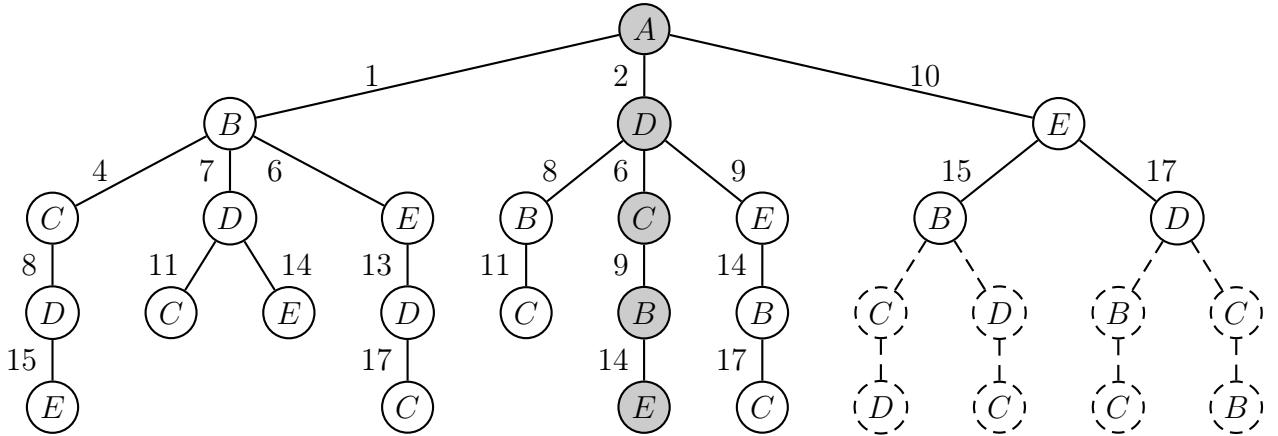


Рис. 2.64

Далее будут рассмотрены только частичные решения AEB и AED , которые будут отброшены (и значит, дальнейшее расширение по этим ветвям проводиться не будет), в силу того, что $b(AEB) = 15 > r$ и $b(AED) = 17 > r$.

На рис. 2.64 показано дерево решений данной задачи методом ветвей и границ, для сравнения штриховыми линиями отмечены ветви, отброшенные оценочной функцией.

□ Задача коммивояжера (второй алгоритм)

Опишем менее очевидный метод построения частичных решений, который, как правило, дает лучшие результаты, чем предыдущий.

В качестве частичного решения рассмотрим теперь не путь, который является частью искомого гамильтонова пути, а набор ребер, которые не входят в него. В качестве оценочной функции b выберем вес минимального остовного дерева, построенного для графа, из которого уже выброшены ребра частичного решения x (если после отбрасывания граф становится несвязным, можно считать вес бесконечным и обрывать расширение данного частичного решения). Гамильтоновы пути являются подмножеством множества остовных деревьев, а минимум на более широком множестве будет меньше или равен минимуму на более узком. Поэтому $b(x) \leq f(p)$, для любого подмножества ребер исходного графа $E \setminus \{x\}$, из которых можно построить гамильтонов путь.

В качестве расширений частичного решения рассмотрим ребра, выхо-

дящие из вершины остовного дерева степени 3 или более (такой вершины не бывает в гамильтоновом пути, поэтому одно из ребер надо удалять; если таких вершин несколько, достаточно взять одну из них).

2.3. Бинарные отношения

2.3.1. Введение. Постановка задачи

Бинарные отношения уже встречались в школьном курсе математики. Примеры таких отношений — отношения неравенства, равенства, подобия, параллельности, делимости и пр. Бинарное отношение каждым двум объектам сопоставляет логическое значение «да», если объекты находятся в этом отношении, и «нет» в ином случае. Другими словами, множество пар объектов разбивается на два подмножества, пары первого подмножества находятся в данном отношении, а второго — не находятся. Это свойство можно положить в основу определения бинарного отношения.

Определение 2.54. Пусть задано множество M . Рассмотрим декартово произведение этого множества на себя $M \times M$. Подмножество R множества $M \times M$ называют *бинарным отношением* R на множестве M . Если пара $(x; y)$ принадлежит множеству R , то говорят, что элемент x находится в отношении R с элементом y , и записывают xRy .

Пример 2.53. Введем отношение сравнимости R : x сравнимо с y по модулю m тогда и только тогда, когда x и y имеют одинаковые остатки от деления на m , т. е. $x \equiv y \pmod{m}$.

Рассмотрим введенное отношение R для случая $m = 3$ на множестве $M = \{1; 2; 3; 4; 5; 6\}$, тогда

$$M \times M = \left[\begin{array}{cccccc} (1; 1); & (1; 2); & (1; 3); & (1; 4); & (1; 5); & (1; 6); \\ (2; 1); & (2; 2); & (2; 3); & (2; 4); & (2; 5); & (2; 6); \\ (3; 1); & (3; 2); & (3; 3); & (3; 4); & (3; 5); & (3; 6); \\ (4; 1); & (4; 2); & (4; 3); & (4; 4); & (4; 5); & (4; 6); \\ (5; 1); & (5; 2); & (5; 3); & (5; 4); & (5; 5); & (5; 6); \\ (6; 1); & (6; 2); & (6; 3); & (6; 4); & (6; 5); & (6; 6) \end{array} \right].$$

Отношение R определяется множеством таких пар:

$$R = \begin{bmatrix} (1; 1); & & (1; 4); & \\ & (2; 2); & & (2; 5); \\ & & (3; 3); & (3; 6); \\ (4; 1); & & (4; 4); & \\ & (5; 2); & & (5; 5); \\ & & (6; 3); & (6; 6) \end{bmatrix}.$$

Пример 2.54. Рассмотрим в качестве $M = \mathbf{R}$ — множество вещественных чисел, или, иными словами, множество точек вещественной прямой. Тогда $M \times M = \mathbf{R}^2$ — множество точек координатной плоскости. Отношение неравенства $<$ определяется множеством пар $R = \{(x; y) | x < y\}$. Это координаты точек, лежащих выше прямой $y = x$.

Упражнение 2.9.

1. На множестве вещественных чисел задано отношение: xRy тогда и только тогда, когда одно из чисел вдвое больше другого. Изобразите на плоскости множество точек, определяющее это отношение.

2. На множестве $M = \{1; 2; 3; 4; 5; 6\}$ задано отношение делимости: xRy тогда и только тогда, когда x делится на y . Сколько пар содержит это отношение? Перечислите эти пары.

3. Введем на множестве $M = \{1; 2; 3; 4; 5; 6\}$ отношение взаимной простоты, т. е. xRy тогда и только тогда, когда x и y взаимно просты: $D(x; y) = 1$. Сколько пар содержит это отношение? Перечислите эти пары.

2.3.2. Свойства бинарных отношений

Определение 2.55. Бинарное отношение R на множестве M называют *рефлексивным*, если каждый элемент этого множества находится в отношении с самим собой: xRx для любого $x \in M$.

Пример 2.55.

1. Отношение сравнимости рефлексивно (при любом натуральном t и на любом множестве целых чисел).

2. Отношение строгого неравенства на множестве вещественных чисел не рефлексивно.

3. Отношение делимости рефлексивно (на любом множестве целых чисел, не содержащем нуля).

Определение 2.56. Бинарное отношение R на множестве M называют *антирефлексивным* (*иррефлексивным*), если ни один элемент этого множества не находится в отношении с самим собой: для любого $x \in M$ неверно, что xRx .

Пример 2.56.

1. Отношение строгого неравенства на множестве вещественных чисел антирефлексивно.
2. Отношение взаимной простоты антирефлексивно на любом множестве целых чисел, не содержащем 1 и -1 , рефлексивно на множествах $\{1\}$, $\{-1\}$, $\{-1; 1\}$ и не является ни рефлексивным, ни антирефлексивным в ином случае.

Определение 2.57. Бинарное отношение R на множестве M называют *симметричным*, если вместе с каждой парой $(x; y)$ в отношение входит и симметричная пара $(y; x)$: для любых $x, y \in M$ $xRy \Rightarrow yRx$.

Пример 2.57.

1. Отношение сравнимости симметрично при любом натуральном t и на любом множестве целых чисел.
2. Отношение строгого неравенства на множестве вещественных чисел не симметрично.
3. Отношение делимости симметрично на множестве попарно взаимно простых целых чисел, не содержащем единицу. Например, на множестве простых чисел.
4. Отношение взаимной простоты симметрично на любом множестве целых чисел.

Определение 2.58. Бинарное отношение R на множестве M называют *асимметричным*, если ни одна пара не входит в отношение вместе с симметричной ей: для любых $x, y \in M$, если xRy , то неверно, что yRx .

Пример 2.58.

1. Отношение строгого неравенства на множестве вещественных чисел асимметрично.
2. Отношение делимости не является асимметричным ни на каком множестве целых чисел, не содержащем нуля, вследствие рефлексивности.

Определение 2.59. Бинарное отношение R на множестве M называют *антисимметричным*, если никакая пара, состоящая из разных элементов, не входит в отношение вместе с симметричной ей: для любых $x, y \in M$, если xRy и yRx , то $x = y$.

Пример 2.59.

1. Отношение нестрогого неравенства на множестве вещественных чисел антисимметрично.
2. Отношение делимости является антисимметричным на любом множестве целых чисел.

Упражнение 2.10.

1. Верно ли, что асимметричное отношение всегда антирефлексивно? Докажите.
2. Верно ли, что симметричное отношение всегда рефлексивно? Докажите.
3. Верно ли, что асимметричное отношение всегда антисимметрично? Докажите.
4. Верно ли, что отношение асимметрично тогда и только тогда, когда оно антирефлексивно и антисимметрично? Докажите.

Определение 2.60. Бинарное отношение R на множестве M называют *транзитивным*, если вместе с парами $(x; y)$ и $(y; z)$ в отношение входит и пара (x, z) , т. е. для любых $x, y, z \in M$, если xRy и yRz , то xRz .

Замечание 2.49. Свойство транзитивности хорошо иллюстрируется отношением достижимости: если пункт y достижим из пункта x , а пункт z — из пункта y , то пункт z достижим из пункта x .

Пример 2.60.

1. Отношение сравнимости транзитивно при любом натуральном t и на любом множестве целых чисел.
2. Отношение строгого (нестрогого) неравенства транзитивно на любом подмножестве вещественных чисел.
3. Отношение делимости транзитивно на любом множестве целых чисел.
4. Отношение взаимной простоты не является транзитивным на любом множестве целых чисел. Например, 2 взаимно просто с 3, 3 взаимно просто с 4, но 2 и 4 не взаимно просты.

Упражнение 2.11. Верно ли, что транзитивное и симметричное отношение всегда рефлексивно? Докажите.

2.3.3. Способы задания отношений

Кроме явного перечисления пар, определяющих бинарное отношение, возможны следующие способы задания отношений.

- **Задание процедурой проверки.** Рассмотрим этот способ на примере.

Пример 2.61.

1. Отношение взаимной простоты проверяется процедурой нахождения наибольшего общего делителя: если $D(x; y) = 1$, то $(x; y)$ входит в отношение взаимной простоты.

2. Отношение делимости проверяется процедурой деления с остатком: если $x \equiv 0 \pmod{y}$, то $(x; y)$ входит в отношение делимости.

3. Той же процедурой проверяется отношение равенства остатков при делении на m : если $x - y \equiv 0 \pmod{m}$, то $(x; y)$ входит в отношение.

Для отношений на конечных множествах (которые являются основными для дискретной математики) используются также следующие способы задания и описания отношений.

- **Задание матрицей смежности.** Определим матрицу A размера $|M| \times |M|$, где $|M|$ — количество элементов множества M . Пронумеруем элементы множества M . Тогда $a_{ij} = 1$, если элемент с номером i находится в отношении с элементом с номером j (iRj), и $a_{ij} = 0$ иначе.

Пример 2.62. Матрица смежности для отношения делимости на множестве $M = \{1; 2; 3; 4; 5; 6\}$ выглядит так:

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

- **Задание графом.** Элементы множества изображаются точками плоскости и образуют множество вершин графа. Отношение представляют дугами (ребрами) графа: если $(x; y)$ входит в отношение, то из вершины x проводится ориентированная дуга в y .

Пример 2.63. Граф для отношения сравнимости по модулю три на множестве $M = \{1; 2; 3; 4; 5; 6; 7; 8\}$ выглядит, как показано на рис. 2.65 а. Заметим, что он состоит из трех компонент связности: $\{1; 4; 7\}$, $\{3; 6\}$ и $\{2; 5; 8\}$.

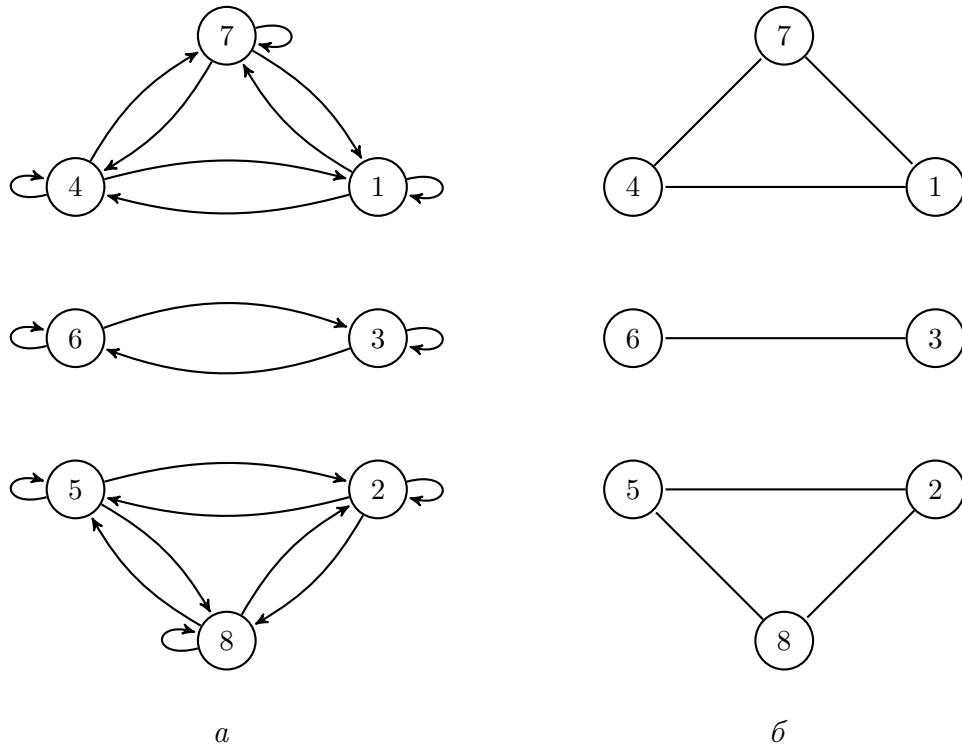


Рис. 2.65

Замечание 2.50. Для симметричных и рефлексивных отношений петли обычно не изображают, а пары ориентированных дуг, соединяющих данные вершины, заменяют одним неориентированным ребром. Например, граф бинарного отношения из примера 2.63 будет выглядеть так, как показано на рис. 2.65 б.

• **Задание списком смежности.** Для каждого элемента множества перечисляются элементы, находящиеся с ним в данном отношении.

Пример 2.64. Список смежности для отношения взаимной простоты на множестве $M = \{1; 2; 3; 4; 5; 6\}$ выглядит так:

$$\begin{array}{lll} 1 \rightarrow 2; 3; 4; 5; 6 & 2 \rightarrow 1; 3; 5 & 3 \rightarrow 1; 2; 4; 5 \\ 4 \rightarrow 1; 3; 5 & 5 \rightarrow 1; 2; 3; 4; 6 & 6 \rightarrow 1; 5 \end{array}$$

Дадим интерпретацию свойств бинарных отношений на описывающих их графах и матрицах.

Теорема 2.46. Справедливы следующие утверждения:

- 1) диагональ матрицы смежности рефлексивного отношения состоит из единиц;
- 2) у симметричного отношения матрица смежности симметрична;
- 3) граф рефлексивного отношения имеет петли в каждой вершине;
- 4) граф симметричного отношения вместе с дугой, соединяющей x с y , содержит дугу, соединяющую y с x ;

5) граф транзитивного отношения обладает следующим свойством: если из вершины x , двигаясь вдоль дуг, можно попасть в вершину y , то в графе должна быть дуга, непосредственно соединяющая x с y .

Упражнение 2.12.

1. Опишите свойства матрицы смежности: а) антирефлексивного отношения; б) асимметричного отношения; в) антисимметричного отношения; г) транзитивного отношения.
2. Опишите свойства графа: а) антирефлексивного отношения; б) асимметричного отношения; в) антисимметричного отношения.

2.3.4. Отношения эквивалентности и толерантности

Определение 2.61. Бинарное отношение, обладающее свойствами рефлексивности, симметричности и транзитивности, называют отношением **эквивалентности**.

Пример 2.65. Отношение сравнимости (по любому модулю) является отношением эквивалентности.

Сопоставим каждому элементу множества M все элементы, находящиеся с ним в данном отношении эквивалентности: $M_x = \{y \in M \mid xRy\}$. Справедлива следующая теорема.

Теорема 2.47. Множества M_x и M_y либо не пересекаются, либо совпадают.

Доказательство. Все элементы одного класса эквивалентны между собой, т. е. если $x, y \in M_z$, то xRy . Действительно, пусть $x, y \in M_z$, следовательно xRz и yRz . По симметричности отношения R имеем zRy . Тогда, в силу транзитивности, из xRz и zRy получаем xRy .

Пусть классы M_x и M_y имеют общий элемент z , тогда для любых элементов $x' \in M_x$ и $y' \in M_y$ по доказанному выше $x'Rz$ и zRy' , откуда $x'Ry'$. Таким образом, если классы пересекаются, то они совпадают. ■

Определение 2.62. Множества M_x называют *классами эквивалентности* множества M по отношению R .

Следствие 2.13. Отношение эквивалентности можно интерпретировать как отношение достижимости на неориентированном графе. Тогда классам эквивалентности будут соответствовать компоненты связности графа. Это подграфы, на которые распадается граф: любые вершины из разных компонент связности не достижимы друг из друга, а вершины одной компоненты связности достижимы.

Пример 2.66. Рассмотрим отношение сравнимости целых чисел по модулю m . Классами эквивалентности этого отношения будут классы вычетов по модулю m .

Как построить разбиение конечного множества M на классы эквивалентности для отношения R ? Эту задачу решает следующий алгоритм.

Алгоритм 2.27 (метод раскраски).

Шаг 1. (Инициализация) Пронумеруем элементы множества $M : m[i]$, $1 \leq i \leq n$ и «раскрасим» их в разные цвета. Можно считать, что цвета заданы номерами элементов. Тогда в начале работы алгоритма цвет элемента совпадает с его номером $\text{color}[i] = i$.

Шаг 2. Далее будем перебирать элементы $m[i]$ множества от второго до последнего, сравнивая их с предыдущими: если элемент $m[i]$ будет находиться в отношении R с одним из предыдущих $m[j]$, то его надо «перекрасить» в цвет $m[j]$.

После окончания работы алгоритма элементы каждой компоненты связности будут раскрашены в свой цвет.

Отношение эквивалентности является частным случаем более широкого понятия.

Определение 2.63. Отношением *толерантности* (или просто *толерантностью*) на множестве M называется бинарное отношение, удовлетворяющее свойствам *рефлексивности* и *симметричности*, но не обязательно являющееся транзитивным.

В отличие от отношения эквивалентности, дающего разбиение множества M на непересекающиеся подмножества, отношение толерантности дает *покрытие* множества M , т. е. семейство множеств, таких, что их объединение содержит множество M .

Пример 2.67. Отношение знакомства между людьми является отношением толерантности. В этом случаях свойство транзитивности не предполагается обязательно быть выполненным. В самом деле, X может быть знаком с Y , Y – с Z , но при этом X и Z могут быть не знакомы между собой.

Пример 2.68. Пусть $\{M_i \mid i \in I\}$ – некоторое семейство непустых подмножеств множества M . Тогда отношение пересечения есть отношение толерантности.

Замечание 2.51. На содержательном уровне толерантность означает следующее. Любой объект идентичен сам с собой (свойство рефлексивности), а сходство двух объектов не зависит от того, в каком порядке они сравниваются (свойство симметричности). Однако, если один объект сходен с другим, а этот другой – с третьим, то это совсем не значит, что все три объекта схожи между собой (таким образом, свойство транзитивности может не выполняться).

2.3.5. Отношения предпорядка и порядка

Определение 2.64. Бинарное отношение называют отношением *предпорядка*, если оно рефлексивно и транзитивно.

Отношение предпорядка, обладающее свойством антисимметричности, называют отношением *порядка*.

Замечание 2.52. Иногда определение порядка уточняют так: отношение порядка, определенное выше, называют отношением *нестрогого порядка*, а асимметричное транзитивное отношение — отношением *строгого порядка*.

Отношение симметричного предпорядка является отношением эквивалентности.

Пример 2.69. Отношение R заданное на множестве $M = \{a, b, c\}$

$$R = \{(a, a), (b, b), (c, c), (a, b), (b, a), (b, c), (a, c)\}$$

является предпорядком. Не выполняется свойство антисимметрии (aRb и bRa , но $a \neq b$), однако оно рефлексивно и транзитивно, значит является предпорядком.

Пример 2.70.

1. На множестве вещественных чисел отношение нестрогого неравенства является отношением (нестрогого) порядка, а отношение строгого неравенства — строгого порядка.

2. Отношение делимости на любом множестве целых чисел, не содержащем нуля, является отношением (нестрогого) порядка.

Определение 2.65. Отношение порядка может обладать или не обладать следующим свойством: для любых $x, y \in M$ либо xRy , либо yRx .

Если отношение порядка обладает этим свойством, то его называют отношением *линейного порядка*, в противном случае — отношением *частичного порядка*.

Пример 2.71.

1. Отношение нестрогого неравенства есть отношение линейного порядка, так как для любых двух вещественных чисел одно будет не меньше другого.

2. Отношение делимости на множестве целых чисел M может являться (в зависимости от множества) отношением частичного порядка или отношением линейного порядка. Например, если $M = \{1; 2; 3; 4\}$, то это отношение будет отношением частичного порядка, так как ни 3 не делится на 4, ни 4 на 3. Если же $M = \{1; 2; 4; 8\}$, то отношение будет отношением линейного порядка.

Пример 2.72. Пусть множество M состоит из n -мерных векторов, причем для каждого элемента вектора определены отношения $>$ и \geqslant . Будем считать, что

$$X = (x_1, \dots, x_n) > Y = (y_1, \dots, y_n), \text{ если } x_i \geqslant y_i, i = 1, \dots, n$$

и существует $1 \leqslant k \leqslant n : x_k > y_k$.

Такое отношение называется *отношением Парето*. Данное отношение транзитивно и асимметрично, то есть это отношение строгого порядка. Отношение используется в многокритериальных задачах принятия решений.

Вектор X называется *Парето-оптимальным решением*, если нет такого вектора Y , что $Y > X$ по Парето.

Рассмотрим следующую задачу. Любая иерархическая система определяет отношение подчинения, которое является отношением порядка, но, как правило, только частичного.

Например, из рис. 2.66 видно, что, например, для элементов E и C нельзя определить, кто из них «главнее». Это не всегда удобно (например, отношение подчинения желательно определить для любых двух военнослужащих). Решение этой задачи сводится к построению отношения линейного порядка, согласованного с заданным отношением частичного порядка.

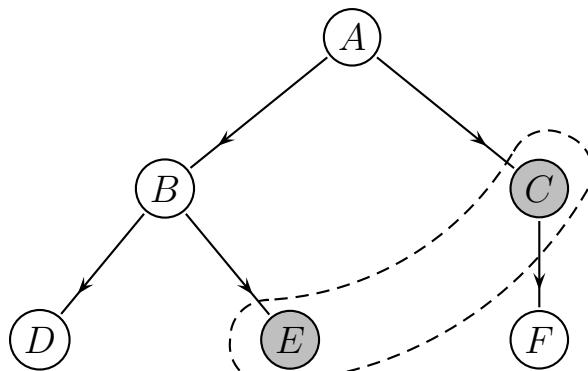


Рис. 2.66

Определение 2.66. Отношение L называют *согласованным* с отношением R , если любые элементы множества, находящиеся в отношении R , будут находиться и в отношении L .

Определение 2.67. Элемент x множества M называют минимальным по отношению R , если не существует элемента «меньше его», т. е. не существует $y \in M$, такого, что yRx .

Теорема 2.48. Любое отношение порядка (в том числе, частичного), заданное на конечном множестве M , имеет минимальный элемент.

Доказательство. Докажем от противного: пусть для каждого элемента M существует меньший его. Тогда получается цепочка (строится справа

налево): $\dots vRz zRy yRx$. Эта цепочка не может состоять из бесконечного количества различных элементов ввиду конечности множества M . Не умоляя общности, будем считать, что в ней встретится элемент x , такой, что цепочка имеет вид $\dots xRw \dots vRz zRy yRx$. Применим свойство транзитивности к этой цепочке от первого до предпоследнего звена: получим xRy . Сравним этот результат с последним звеном: для отношения нестрогого порядка отсюда (по антисимметричности) получим $x = y$, что противоречит условию. Для строгого порядка эти отношения (по асимметричности) не могут существовать одновременно, что также приводит к противоречию. ■

2.3.6. Алгоритм топологической сортировки

Алгоритм топологической сортировки по заданному отношению частичного порядка строит согласованное с ним отношение линейного порядка.

В качестве отношения линейного порядка будем рассматривать отношение неравенства на множестве номеров элементов множества M (n — количество элементов M). Нумерация элементов описывается алгоритмом:

Алгоритм 2.28 (топологической сортировки).

ЦИКЛ ПО i ОТ 1 ДО n

Найти минимальный элемент множества и присвоить ему номер i .

Удалить этот элемент из множества.

КЦ

Обоснуем корректность работы алгоритма 2.28.

1. Отношение L , определенное как отношение неравенства на номерах элементов, является отношением линейного порядка.

2. Отношение L согласовано с отношением R . Действительно, выбранный на i -м шаге элемент является минимальным на множестве всех элементов M , за исключением уже выбранных, поэтому он не может быть «больше» (не существует $y \in M$ такого, что yRx) ни одного из них. По вновь введенному отношению этот элемент будет «больше» всех выбранных и «меньше» всех не выбранных, что подтверждает согласованность.

3. Наконец, алгоритм обязательно (за n шагов) закончит свою работу.

Пример 2.73. Для иерархической системы, (см. рис. 2.66), можно построить несколько согласованных линейных отношений порядка.

На первом шаге минимальных элементов три — D, E, F .

Выберем один из них, например D , и присвоим ему номер 1. После выбрасывания этого элемента из графа отношения (и всех дуг, входящих в

него), минимальными будут два элемента — E и F . Выберем один из них, например F , и присвоим ему следующий номер — 2 (рис. 2.67).

Далее минимальных элементов будет снова два — E и C и т. д.

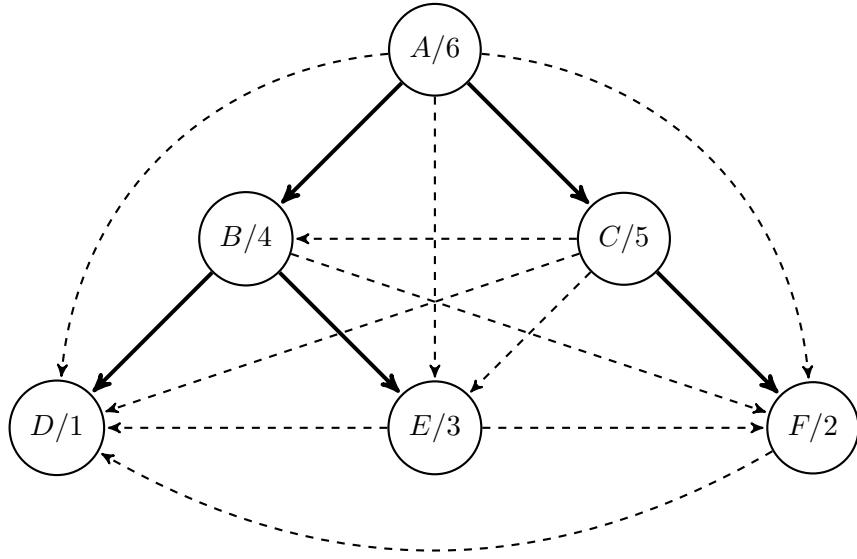


Рис. 2.67

Добавленные для согласованного отношения L связи на рис. 2.67 изображены пунктиром. Протокол работы алгоритма 2.28 представлен в таблице 2.22.

Таблица 2.22

Вершина	D	F	E	B	C	A
Номер	1	2	3	4	5	6

Упражнение 2.13.

1. Дайте определение максимального элемента.
2. Сформулируйте алгоритм топологической сортировки на основе выбора максимального элемента.
3. Примените построенный алгоритм к примеру 2.73.
4. Сколько в этом примере можно построить различных отношений линейного порядка, согласованных с заданным отношением частичного порядка?

Следующая диаграмма (рис. 2.68) показывает связь между различными классами бинарных отношений.

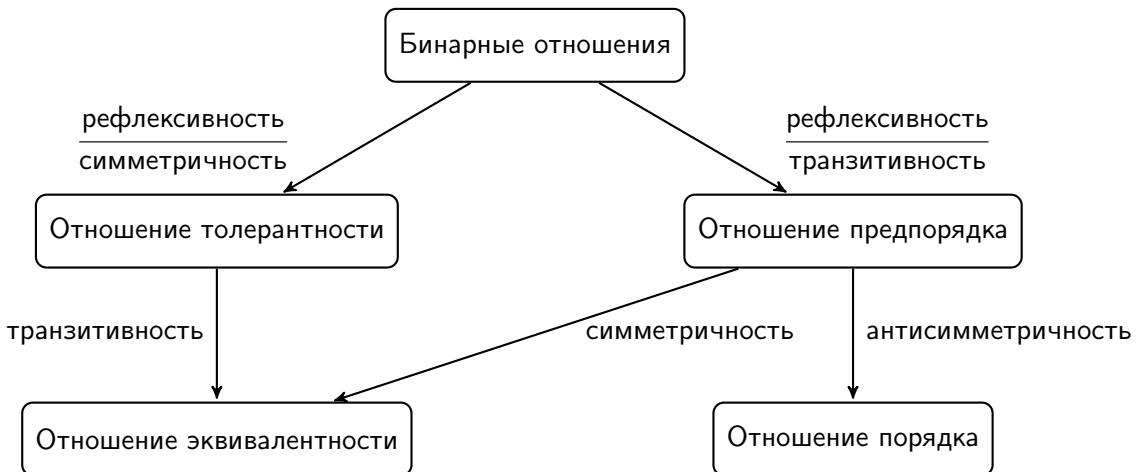


Рис. 2.68

2.3.7. Частично упорядоченные множества. Диаграмма Хассе

Определение 2.68. Множество M , на котором задано отношение частичного порядка, называется *частично упорядоченным*. Если на множестве задано отношение линейного порядка, множество называется *линейно упорядоченным*.

Обратимся к графическому представлению упорядоченных множеств. Упорядоченное множество M , имеющее конечное число элементов удобно представлять с помощью ориентированного графа $H(M)$, который называют *диаграммой Хассе* и строят следующим образом. Вершины $H(M)$ однозначно соответствуют элементам M . При этом вершину, соответствующую элементу a , называют просто вершиной a . Если элемент b непосредственно следует за элементом a (и только в этом случае), вершину b соединяют дугой с вершиной a , ориентированной от b к a .

Определение 2.69. Говорят, что элемент b непосредственно следует за элементом a , если $a \leq b$, и не существует элемента $c \in M$, такого, что $a \leq c \leq b$.

Историческая справка. Впервые систематически такого рода визуализация описана Биркгофом в 1948 году, название было дано им в честь использовавшего подобные диаграммы Хельмута Хассе¹.

Обычно в диаграммах Хассе большие элементы помещают над меньшими.

Теорема 2.49. Диаграмма Хассе обладает следующими очевидными свойствами:

¹Хельмут Хассе (Гельмут Гассе, нем. Helmut Hasse) — немецкий математик. Работал в Галле, Марбурге, Гётtingене и Берлине.

- 1) соотношение $a \leq b$ эквивалентно существованию ориентированного пути из вершины b в вершину a на диаграмме Хассе;
- 2) если на диаграмме существует ребро e , идущее из вершины b в вершину a , то никакого другого пути из b в a быть не может;
- 3) диаграмма Хассе не имеет циклов.

Пример 2.74. Рассмотрим отношение включения определенное на множестве всех подмножеств четырехэлементного множества $\{a, b, c, d\}$. Это частично упорядоченное множество (строгий порядок). Построим соответствующую диаграмму Хассе (рис. 2.69).

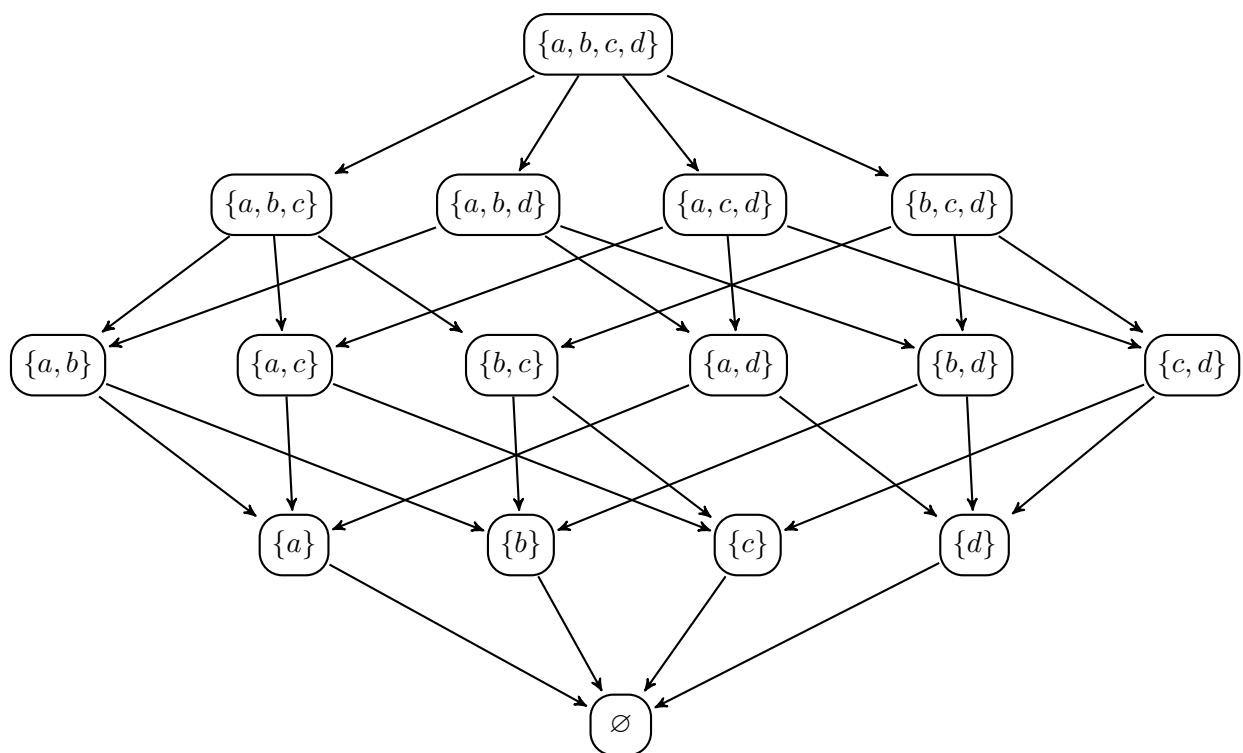


Рис. 2.69

2.3.8. Транзитивное замыкание бинарного отношения. Алгоритм Уоршелла

Рассмотрим сначала общую постановку задачи.

Пусть R есть некоторое бинарное отношение на множестве M , и это отношение не обладает свойством P . Логично в такой ситуации расширить отношение R до \widehat{R} , чтобы \widehat{R} уже обладало свойством P . Так как R есть подмножество декартового произведения $M \times M$, то под расширением будем понимать добавление в R новых пар элементов множества M .

Ясно, что исходное множество R будет подмножеством в \widehat{R} . В таком случае, если вновь построенное множество \widehat{R} будет минимальным среди

всех расширений R с выделенным свойством, то говорят, что \widehat{R} является замыканием R относительно данного свойства. Дадим формальное определение.

Определение 2.70. Отношение \widehat{R} называется замыканием отношения R относительно свойства P , если:

1) \widehat{R} обладает свойством P ;

2) $R \subseteq \widehat{R}$;

3) \widehat{R} является подмножеством любого другого отношения, содержащего R и обладающего свойством P .

Пример 2.75. Пусть $R = \{(a, b) | a, b \in N, a > b\}$. Симметрично ли это отношение? Ответ отрицательный. Построим замыкание этого отношения относительно свойства симметричности. Это означает, что если пара (a, b) принадлежит R , то пара (a, b) также будет принадлежать \widehat{R} . Кроме того, пара (b, a) также будет принадлежать \widehat{R} . Т. е. в нашем случае

$$\widehat{R} = \{(a, b) | a, b \in N, a > b\} \cup \{(a, b) | a, b \in N, a < b\} = \{(a, b) | a, b \in N, a \neq b\}.$$

Вывод: симметричным замыканием отношения „меньше“ на множестве натуральных чисел является отношение „не равно“.

Перейдем к рассмотрению важного частного случая.

Рассмотрим следующую задачу. Пусть на конечном множестве M задано бинарное отношение R , не являющееся, вообще говоря, транзитивным. Как построить согласованное с ним транзитивное бинарное отношение?

Замечание 2.53.

1. Эта задача имеет содержательную интерпретацию на графике: надо дополнить граф следующими дугами. Если из вершины x можно построить путь до вершины y , то надо построить дугу $(x; y)$.

2. На матрице смежности задачу можно сформулировать так: по матрице смежности построить матрицу достижимости.

В теореме 2.50 будет получен алгоритм построения матрицы достижимости (транзитивного замыкания) с помощью возведения матрицы смежности в степень.

Заменим обычные арифметические операции, участвующие в определении умножения матриц, следующими: операцию сложения двух чисел заменим операцией выбора наибольшего значения (а операцию умножения сохраним прежней). Тогда верна теорема.

Теорема 2.50. Пусть A — матрица смежности рефлексивного отношения R , заданного на конечном множестве M , а n — число элементов M . Тогда A^{n-1} — матрица достижимости.

Доказательство. Рассмотрим умножение матрицы смежности A на себя. Элемент, стоящий в i -й строке и j -м столбце результирующей матрицы, по определению умножения матриц (со сделанными уточнениями) покажет, можно ли достичь из i -й вершины j -ю, используя не более одной промежуточной вершины. Соответственно, элементы $(n - 1)$ -й степени матрицы A , покажут попарные достижимости вершин с использованием не более $(n - 2)$ промежуточных вершин. Но на графе из n вершин любой путь без циклов будет иметь максимум n вершин, а значит не более $(n - 2)$ промежуточных. ■

Пример 2.76. Рассмотрим граф рефлексивного отношения на множестве из трех элементов. Пусть A — матрица смежностей этого отношения, $n=3$.

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix} \text{ тогда, после работы алгоритма } A^{n-1} = A^2 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

На рис. 2.70 представлены, соответственно, исходный граф и граф транзитивного замыкания.



Рис. 2.70

Замечание 2.54. Граф, соответствующий транзитивному замыканию, является полным трехвершинным графом — все его вершины соединены дугами.

Упражнение 2.14. Примените алгоритм умножения матриц к антирефлексивному отношению. Какой смысл будет иметь матрица A^{n-1} ?

Оценим трудоемкость предложенного алгоритма. При вычислении одного элемента матрицы, полученной умножением матриц размера $n \times n$ на себя требуется n умножений и $(n - 1)$ сложение. Для вычисления всех элементов потребуется порядка n^3 действий, а для возведения матрицы в $(n - 2)$ степень — порядка n^4 действий.

Возникает вопрос, нет ли алгоритма с меньшим порядком трудоемкости. Таким алгоритмом является *алгоритм Уоршелла*, трудоемкость которого имеет порядок n^3 .

- **Алгоритм Уоршелла.** Идея алгоритма Уоршелла состоит в расширении множества промежуточных вершин по следующему правилу: на

каждом шаге в рассмотрение добавляется одна новая вершина, после чего достижимости вершин пересчитываются «через нее». Если w — промежуточная вершина, то достижимость вершины v из вершины u через w пересчитывается по правилу

$$D[u; v] := \max\{D[u; v], D[u; w] \cdot D[w; v]\},$$

т. е. не меняется, если v достижима из u , и меняется (с 0 на 1), если достижимости до введения промежуточной вершины w не было, а w достижима из u и v достижима из w .

Таким образом, после k шагов будут соединены те вершины, которые достижимы по путям, проходящим только через первые k вершин (кроме первой и последней).

Замечание 2.55. Так как элементы матрицы смежности по существу логические значения (0 и 1), правило модификации матрицы D естественнее записать в логических операциях

$$D[u; v] := D[u; v] \text{ ИЛИ } (D[u; w] \text{ И } D[w; v]).$$

Таким образом, алгоритм Уоршелла можно записать так.

Алгоритм 2.29 (алгоритм Уоршелла).

```

ДЛЯ ВСЕХ  $w \in M$ 
    ДЛЯ ВСЕХ  $u \in M$ 
        ДЛЯ ВСЕХ  $v \in M$ 
             $D[u; v] := D[u; v] \text{ ИЛИ } (D[u; w] \text{ И } D[w; v])$ 
        КЦ
    КЦ
КЦ

```

Легко видеть, что трудоемкость полученного алгоритма n^3 .

Замечание 2.56.

1. Построение алгоритма 2.29 осуществлялось на основе теоретических рассуждений, которые, фактически, составляют доказательство корректности полученного алгоритма. Заметим, что после выполнения k -го шага внешнего цикла значения $D[u; v]$ показывают достижимость вершины v из вершины u , используя первые k вершин в качестве промежуточных.

2. Оба алгоритма построения транзитивного отношения, согласованного с данным, строят одно и то же отношение. Оно называется транзитивным замыканием исходного отношения, так как является минимальным (остальные согласованные с исходным транзитивные отношения его содержат) среди всех таких расширений исходного отношения.

3. Алгоритм Уоршелла является частным случаем алгоритма Флойда.

Пример 2.77. С помощью алгоритма Уоршелла построим матрицу достижимости для графа на рис. 2.71,*a*. Матрица смежности графа представлена на рис. 2.71,*b*.

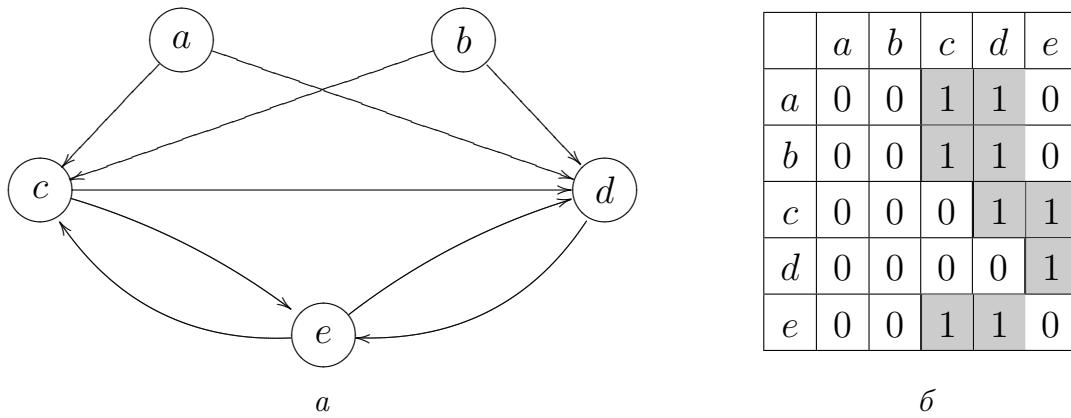


Рис. 2.71

Проиллюстрируем работу алгоритма Уоршелла по шагам. В качестве номера шага будет фигурировать соответствующая вершина внешнего цикла. Матрица достижимости D в начале работы совпадает с матрицей смежности.

Замечание 2.57. При работе с промежуточной вершиной u не модифицируются элементы строки и столбца u матрицы достижимости.

Шаг a. Пересчет матрицы достижимости идет через вершину a :

$$D[u; v] := \max\{D[u; v], (D[u; a] \cdot D[a; v])\}.$$

В силу того, что $D[u; a] = 0$ для всех вершин, то на этом шаге матрица достижимости не меняется.

Шаг b. Пересчет матрицы достижимости идет через вершину b :

$$D[u; v] := \max\{D[u; v], (D[u; b] \cdot D[b; v])\}.$$

Так как $D[u; b] = 0$ для всех вершин, то на этом шаге матрица достижимости также не меняется.

Шаг c. Пересчет матрицы достижимости идет через вершину c :

$$D[u; v] := \max\{D[u; v], (D[u; c] \cdot D[c; v])\}.$$

Матрица D модифицируется следующим образом:

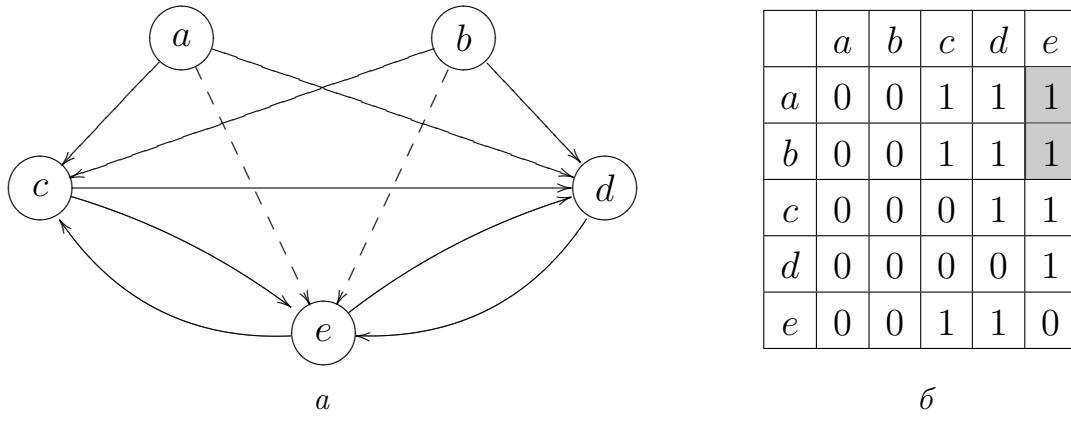
$$\begin{aligned} D[a; e] &:= D[a; c] \cdot D[c; e] = 1, \\ D[b; e] &:= D[b; c] \cdot D[c; e] = 1. \end{aligned}$$

Граф и матрица достижимости после шага c показаны на рис. 2.72. Добавленные дуги отмечены штриховыми линиями.

Шаг d. Пересчет матрицы достижимости идет через вершину d :

$$D[u; v] := \max\{D[u; v], (D[u; d] \text{ И } D[d; v])\}.$$

На этом шаге матрица достижимости не меняется.



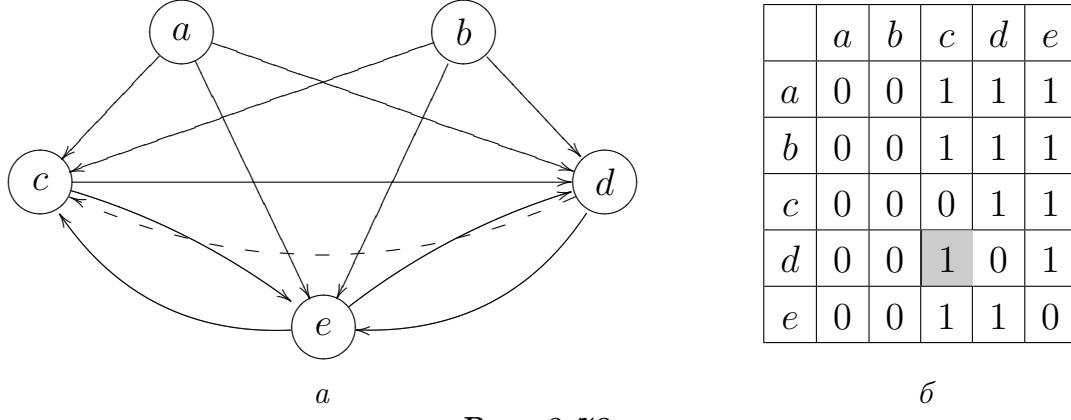
Puc. 2.72

Шаг e. Пересчет матрицы достижимости идет через вершину e :

$$D[u;v] := \max\{D[u;v], (D[u;e] \text{ } \amalg \text{ } D[e;v])\}.$$

Матрица D модифицируется следующим образом:

Граф и матрица достижимости представлены на рис. 2.73. Добавленная дуга отмечена штриховой линией.



Puc. 2.73

2.3.9. Примеры и задачи на бинарные отношения

Пример 2.78. Пусть $M = \{1, 2, 3, 4, 5\}$ и

$$R = \{(x, y) \mid x, y \in M, x + 1 < y\}.$$

Очевидно, что отношение R антирефлексивно, ассиметрично и транзитивно.

Пример 2.79. Отношение на множестве слов русского языка определено следующим образом: aRb тогда и только тогда, когда они имеют хотя бы одну общую букву. Исследовать отношение на множестве

$$M = \{\text{лошадь, корова, вагон, нить, пир, липа}\}$$

Отношение R рефлексивно, симметрично, но не транзитивно:
 $\text{лошадь } R \text{ корова, корова } R \text{ пир,}$ но неверно, что $\text{лошадь } R \text{ пир.}$

Таким образом, отношение R – отношение толерантности. С помощью алгоритма 2.29 можно построить транзитивное замыкание отношения.

Пример 2.80. Рассмотрим отношение включения множеств $A \subseteq B$, определенное на множестве всех подмножеств четырехэлементного множества $\{a, b, c, d\}$.

Отношение рефлексивно, антисимметрично и транзитивно. Это отношение порядка, но частичного, так как, например, $\{a, b\} \not\subseteq \{a, c\}$. С помощью алгоритма топологической сортировки легко вложить его в полное отношение.

Пример 2.81. Построить бинарное отношение R на некотором множестве A :

- a) рефлексивное, симметричное, не транзитивное;
- b) рефлексивное, транзитивное, не симметричное;
- c) симметричное, транзитивное, не рефлексивное.

Отношение R будем описывать явным перечислением входящих в него элементов:

- a) $A = \{1, 2, 3\}, R = \{(1, 1), (2, 2), (3, 3), (1, 2), (2, 1), (2, 3), (3, 2)\}$
- b) $A = \{1, 2\}, R = \{(1, 1), (2, 2), (1, 2)\}$
- c) $A = \{1, 2\}, R = \{(1, 1)\}$

Пример 2.82. Пусть $M = \{2, 3, 4, 5, 6, 7, 8, 9\}$ и

$$R = \{(x, y) \mid x, y \in M, D(x, y) = 1\}.$$

Очевидно, что отношение R антирефлексивно, симметрично и транзитивно.

Пример 2.83. Пусть $A = \{1, 2, 3\}$, а отношение R на A задано упорядоченными парами:

$$R = \{(1, 1), (1, 2), (1, 3), (3, 1), (2, 3)\}.$$

Оно не рефлексивно, не симметрично и не транзитивно. Построим соответствующие замыкания.

1. Замыкание по рефлексивности: $\hat{R} = R \cup \{(1, 1), (2, 2), (3, 3)\}$.
2. Замыкание по симметричности: $\hat{R} = R \cup \{(2, 1), (3, 2)\}$.

3. Замыкание по транзитивности можно построить с помощью алгоритма 2.29.

Задача 2.6. Показать, что отношение R определенное на множестве $A = \{1, 2, 3, 4, 5, 6\}$:

$$R = \{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6), (1, 2), (1, 4), (2, 1), (2, 4), (3, 5), (5, 3), (4, 1), (4, 2)\}$$

есть отношение эквивалентности. Построить классы эквивалентности.

Задача 2.7. Исследовать отношение R , заданное на множестве M :

$$M = \{-4, -3, -2, -1, 0, 1, 2, 3\}, \quad R = \{(x, y) \mid x, y \in M, xy > 0\}.$$

Задачи для самостоятельного решения

1. Докажите, что если число ребер графа с n вершинами ($n > 2$) больше C_{n-1}^2 , то граф связан.

2. Докажите, что если степень каждой вершины графа не меньше двух, то граф содержит цикл.

3. Докажите, что в каждом графе с n вершинами ($n \geq 2$) найдутся две вершины с одинаковыми степенями.

4. Определите число ориентированных графов с n вершинами (без петель), в которых каждая пара различных вершин соединена:

- а) не более чем одним ребром;
- б) точно одним ребром.

5. Используя алгоритм Косарайю и Ширира (алгоритм 2.5) определите компоненты сильной связности в ориентированном графе на рис. 2.74 и постройте соответствующий граф конденсации.

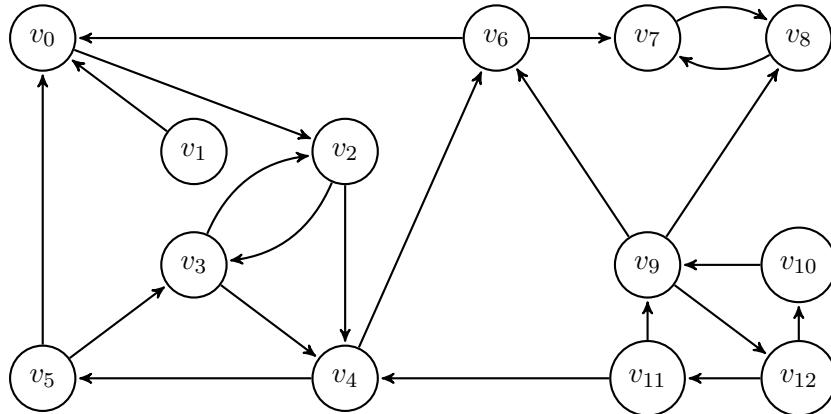


Рис. 2.74

6. Докажите, что число стягивающих деревьев в графе K_n при $n > 1$ равно n^{n-2} .

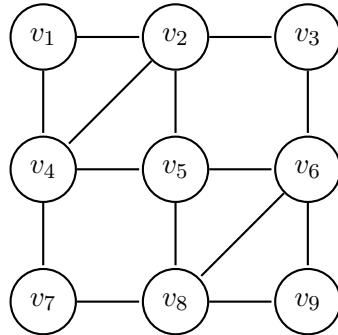
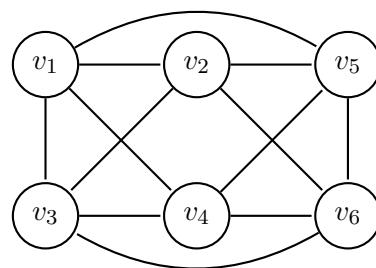
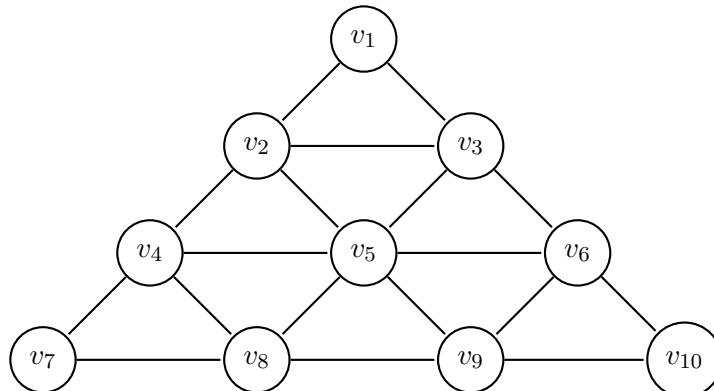
7. Докажите, что лес из k деревьев, содержащий n вершин, имеет ровно $n - k$ ребер.

8. Докажите, что если у дерева есть, по крайней мере, одно ребро, то у него существует висячая вершина.

9. Найдите стягивающие деревья в графах K_5 и $K_{3,3}$.

10. Определите при каких m и n графы K_n и $K_{m,n}$ являются эйлеровыми.

11. Постройте замкнутую эйлерову цепь для графов на рис 2.75 *a*,*b* и рис 2.76.

*a**b**Рис. 2.75**Рис. 2.76*

12. Модифицируйте алгоритм Флери для нахождения *всех* замкнутых эйлеровых цепей графа.

13. Постройте главные циклы и коциклы графов K_5 и $K_{3,3}$.

14. Перечислите все попарно неизоморфные графы с четырьмя вершинами.

15. Определите изоморфны ли графы на рис. 2.77.

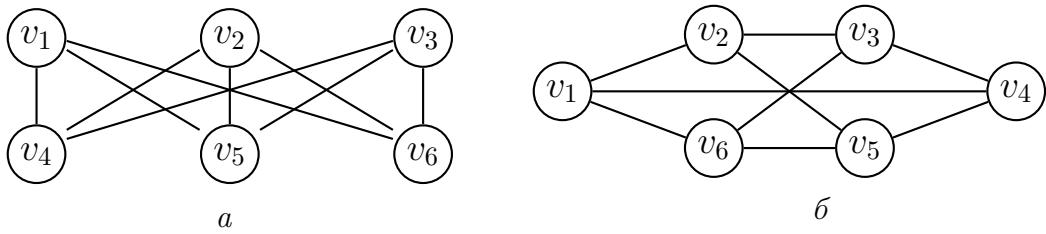


Рис. 2.77

16. Выясните, какие из графов, изображенных на рис. 2.78, изоморфны друг другу.

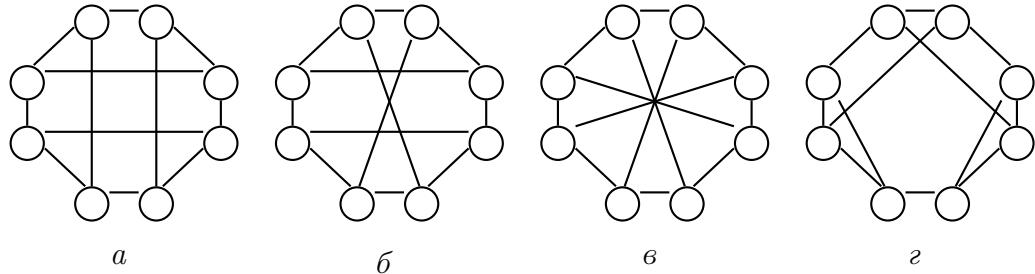


Рис. 2.78

17. Покажите, что формула Эйлера (см. теорему 2.34) следующим образом обобщается на случай плоского *несвязного* (n, m) -графа:

$$n - m + f = k + 1,$$

где k — число компонент связности графа.

18. Определите какое наименьшее количество ребер нужно удалить из графа K_6 , чтобы получить планарный граф.

19. На основе алгоритма Краскала сформулируйте алгоритм, определяющий связность графа. Докажите корректность алгоритма.

20. Для заданного на рис. 2.79 графа постройте минимальное остовное дерево, применяя:

- а) алгоритм Прима (построение начинать с вершины K);
- б) модифицированный алгоритм Краскала.

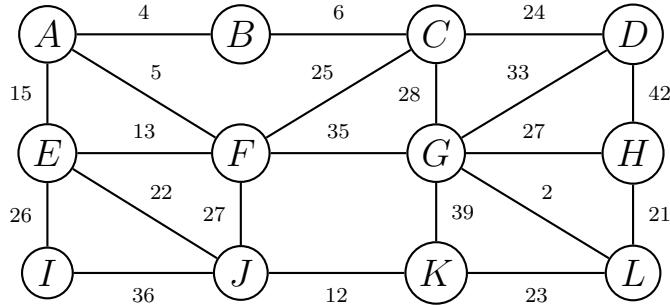


Рис. 2.79

21. Каким свойством должен обладать граф и исходная вершина, чтобы алгоритм Прима построения минимального остовного дерева давал ту же последовательность ре-

бер, что и алгоритм Краскала (для простоты будем считать длины всех ребер различными).

22. Нарисуйте двудольный граф по следующим данным. Множество вершин левой доли $V_L = \{6, 7, 8, 9, 10, 15\}$. Множество вершин правой доли $V_R = \{1, 2, 3, 4, 5, 6\}$. Множество ребер определяется отношением делимости числа из левой доли на число из правой. Например, вершина 10 из левой доли соединяется с вершинами 1, 2, 5 правой доли. Постройте наибольшее паросочетание для данного двудольного графа.

23. На основе алгоритма Форда–Беллмана сформулируйте алгоритм, находящий минимальные длины путей, состоящих ровно из k ребер от источника до остальных вершин. Докажите корректность алгоритма.

24. В графе с положительными весами ребер известно, что расстояния от источника до остальных вершин различны. Сформулируйте алгоритм, находящий путь от источника до k -й по дальности вершины. Обоснуйте его корректность.

25. Для графа на рис. 2.80:

- вычислите длины кратчайших путей от вершины A до остальных вершин с помощью алгоритма Дейкстры;
- с помощью алгоритма Флойда определите кратчайшие пути между всеми парами вершин. Выпишите кратчайший путь от вершины C до вершины D и его длину;
- с помощью алгоритма Уоршелла постройте граф и матрицу достижимости.

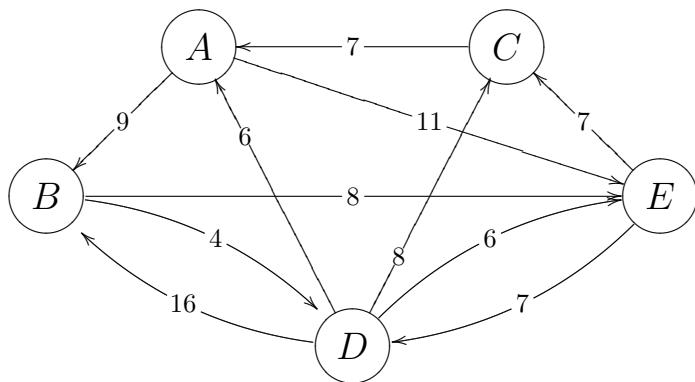


Рис. 2.80

26. На основе алгоритма Флойда сформулируйте алгоритм, определяющий наличие циклов отрицательной длины. Докажите корректность алгоритма.

27. Все вершины графа раскрашены в два цвета (для простоты будем считать, что цвет определяется четностью номера вершины). На основе алгоритма Флойда постройте алгоритм для нахождения длин кратчайших путей, у которых промежуточными могут быть только четные вершины. Докажите корректность алгоритма.

28. Используя алгоритм перебора с возвратом расставьте на шахматной доске восемь ферзей так, чтобы они не били друга.

29. Приведите примеры отношений:

- не рефлексивного, но симметричного и транзитивного;
- не транзитивного, но рефлексивного, и симметричного.

30. На множестве плоскостей в пространстве задано отношение:

- а) параллельности плоскостей;
- б) перпендикулярности плоскостей.

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

31. На множестве чисел $\{53, 43, 54, 42, 44, 60, 50, 20\}$ заданы отношения:

- 1) xRy тогда и только тогда, когда $x \equiv y \pmod{3}$;
- 2) xQy тогда и только тогда, когда в наборе имеется элемент, больший x , но меньший y .

Для каждого из этих отношений:

- а) проверьте, является ли отношение рефлексивным, антирефлексивным, симметричным, асимметричным, антисимметричным, транзитивным;
- б) постройте матрицы и графы этих отношений;
- в) определите, являются ли эти отношения отношениями эквивалентности, частичного порядка, линейного порядка;
- г) для отношения эквивалентности постройте классы эквивалентности;
- д) для отношения частичного порядка примените алгоритм топологической сортировки и получите согласованное отношение линейного порядка.

Глава 3

Математическая логика и теория алгоритмов

3.1. Логика высказываний

Историческая справка. Формальную систему логики впервые разработал великий греческий ученый Аристотель (ученик Платона, воспитатель Александра Македонского). В своем логическом своде «Органон» он создал раздел формальной логики — силлогистику. Его труды оказали влияние на развитие логической науки во всем мире. В Европе до XVII в. вся логика развивалась на основе учения Аристотеля.

Первые значительные попытки превращения логики в математическую науку сделал великий немецкий ученый и политический деятель Готфрид Вильгельм Лейбниц (1646–1716). Однако решающего успеха в этом направлении в 1847 г. добился английский математик Джордж Буль (1815–1864), построив алгебру логики, названную в его честь булевой.

Современный вид математическая логика приобрела в 80-е годы XIX в. в трудах немецкого логика, математика и философа Готлоба Фреге (1848–1925). Он привел первую аксиоматику логики высказываний и предикатов и сделал попытку свести математику к логике.

3.1.1. Основные понятия логики высказываний

Логика высказываний рассматривает предложения языка не с точки зрения их смысла, содержания, а только с точки зрения их истинности или ложности.

Высказыванием будем называть повествовательное предложение, которое является или может оказаться либо истинным, либо ложным. Причем оно не может быть истинным и ложным одновременно. В дальнейшем истину будем обозначать цифрой 1, а ложь — цифрой 0.

Пример 3.1 (задача Кислера [17]). Браун, Джонсон и Смит обвиняются в подделке сведений о доходах. Они дают под присягой следующие показания:

Браун: Джонсон виновен, а Смит не виновен.

Джонсон: Если Браун виновен, то виновен и Смит.

Смит: Я невиновен, но хотя бы один из них виновен..

Требуется ответить на вопросы:

1. Совместимы ли показания всех троих?
2. Если все трое не виновны, то кто лжесвидетельствовал?
3. Предполагая, что все показания истинны, выяснить виновного.

Введем следующие обозначения: B — *Браун виновен*, J — *Джонсон виновен*, S — *Смит виновен*.

Проанализируем показания обвиняемых с точки зрения их «внутреннего строения». Высказывания B, J, S можно назвать простыми, а, например, высказывание Джонсона (если B , то S) — составным, полученным из простых высказываний B и S . Этот пример показывает способ построения составных (сложных) высказываний из простых. Эти способы будем называть *операциями*. В качестве основных выделим пять операций.

Определение 3.1. Пусть X и Y — некоторые высказывания. Тогда высказывание:

- 1) « X и Y » — *конъюнкция* высказываний X и Y ;
- 2) « X или Y » — *дизъюнкция* высказываний X и Y ;
- 3) «**не** X » — *отрицание* высказывания X ;
- 4) «**если** X , **то** Y » — *импликация* высказываний X и Y ;
- 5) « X тогда и только тогда, когда Y » — *эквиваленция* высказываний X и Y .

Например, высказывание Джонсона из примера 3.1 является импликацией высказываний B и S .

Введем следующие обозначения для операций: \wedge — конъюнкция; \vee — дизъюнкция; \neg — отрицание; \rightarrow — импликация; \leftrightarrow — эквиваленция. Символы $\wedge, \vee, \neg, \rightarrow, \leftrightarrow$ называют *связками*.

Следующая таблица определяет, какие значения принимают высказывания, полученные с помощью этих операций, если исходные высказывания X и Y принимают значения 0 или 1. Такую таблицу называют *таблицей истинности* (табл. 3.1).

Более точно, табл. 3.1 содержит пять таблиц истинности, — по одной для каждой связки, которые для удобства объединены в одну.

Для операций \vee и \rightarrow введенные значения требуют дополнительных пояснений.

В русском языке союз «или» имеет два значения: *разделительное* — или то, или другое, но не оба, например, («Пан или пропал», что равносильно «Или пан, или пропал»); *соединительное* — или то, или другое, или

Таблица 3.1

X	Y	$X \wedge Y$	$X \vee Y$	$\neg X$	$X \rightarrow Y$	$X \leftrightarrow Y$
0	0	0	0	1	1	1
0	1	0	1	1	1	0
1	0	0	1	0	0	0
1	1	1	1	0	1	1

оба. Как видно из табл. 3.1, союз «или» будем понимать в соединительном смысле.

Таблица истинности для импликации построена на основе принципов логики Аристотеля:

- *Из лжи следует все, что угодно* (см. первую и вторую строки табл. 3.1).
- *Истина не нуждается в обосновании* (вторая и четвертая строки табл. 3.1).

Пример 3.2. Если n кратно 6, то n кратно 3. Эта импликация истинна при любом натуральном n .

Например, при $n = 6$ это определяет четвертая строка табл. 3.1, при $n = 3$ — вторая строка, а при $n = 5$ — первая.

С помощью логических операций можно строить не только конкретные высказывания, но и выражения, которые имеют только форму высказывания, а конкретное содержание в них можно вкладывать любое. Такие выражения называют *пропозициональными* (*высказывательными*) формулами. Дадим точные определения.

Определение 3.2. Символы истины 1 и лжи 0 называют *логическими константами* (или в дальнейшем, просто константами).

*Логическими переменными*¹ (в дальнейшем просто переменными) называют буквы X, Y, Z, \dots с индексами и без них, значениями которых могут быть 1, или 0.

Определение 3.3. *Формулами* логики высказываний называют:

- 1) логические переменные и константы;
- 2) выражения вида $(F) \wedge (G)$, $(F) \vee (G)$, $\neg(F)$, $(F) \rightarrow (G)$, $(F) \leftrightarrow (G)$, где F и G — формулы логики высказываний.

¹ Другие названия этого понятия: пропозициональные переменные, атомарные формулы или атомы логики высказываний.

Формулу F , в которую входят переменные X_1, X_2, \dots, X_n , будем обозначать $F(X_1, X_2, \dots, X_n)$. Более того, будем пользоваться последним обозначением, даже если некоторые из переменных отсутствуют в записи формулы F , т. е. являются *фиктивными* переменными.

Замечание 3.1. Определение 3.3 относится к так называемым индуктивным² определениям. Такие определения вводят сначала базовые объекты (в нашем случае это введенные в первом пункте логические переменные и константы), а затем способы конструирования новых объектов из уже полученных (в нашем случае это заданные во втором пункте возможности применения логических операций).

Пример 3.3. Рассмотрим выражение $((X) \wedge (Y)) \rightarrow (\neg(Z))$. Является ли данное выражение формулой?

Буквы X, Y, Z — переменные в силу первого пункта определения 3.3, Применяя второй пункт, получаем, что выражение $(X) \wedge (Y)$ — формула, $(\neg(Z))$ — формула, следовательно, $((X) \wedge (Y)) \rightarrow (\neg(Z))$ — также формула.

Если следовать строго определению 3.3, то в формуле надо писать много скобок. Это неудобно для восприятия формулы. Чтобы уменьшить количество скобок условимся переменные и константы в скобки не заключать и введем приоритет (силу связывания) для связок.

Будем считать, что \neg имеет наивысший приоритет, затем в порядке уменьшения приоритета следуют связки $\wedge, \vee, \rightarrow$ и самый маленький приоритет имеет связка \leftrightarrow . Связки одного старшинства применяются в порядке их следования слева направо.

Пример 3.4. Используя эти соглашения, формулу из примера 3.3 $((X) \wedge (Y)) \rightarrow (\neg(Z))$ можно записать в виде $X \wedge Y \rightarrow \neg Z$.

Замечание 3.2. Для дальнейшего упрощения записи формул введем следующие соглашения.

1. Будем опускать символ конъюнкции \wedge , т. е. вместо $X \wedge Y$ писать просто XY , если его присутствие будет очевидно из контекста.

2. Вместо символа отрицания $\neg X$ будем писать \bar{X} , например, $X \bar{Y} Z \vee X \bar{Z}$ вместо $X \wedge \neg Y \wedge Z \vee X \wedge \neg Z$.

Пример 3.5. Вернемся к задаче Кислера (см. пример 3.1). Запишем показания свидетелей в виде формул. Имеем

$$\text{Браун: } J\bar{S}, \quad \text{Джонсон: } B \rightarrow S, \quad \text{Смит: } \bar{S}(B \vee J). \quad (3.1)$$

Задача 3.1. Разработайте алгоритм синтаксического анализа текста в алфавите исходных символов, который решает следующую задачу: по тексту дает ответ на вопрос: является текст формулой логики высказываний или нет.

Можно ли решить эту задачу за один просмотр текста?

²От лат. *inductio* — выведение, наведение

В дальнейшем нам понадобится понятие подформулы. Дадим индуктивное определение, аналогичное определению 3.3 для формул.

Определение 3.4.

- 1) Подформулой переменной или константы является она сама.
- 2) Подформулами формулы \overline{F} являются она сама и все подформулы формулы F .
- 3) Подформулами формул FG , $F \vee G$, $F \rightarrow G$, $F \leftrightarrow G$ являются они сами и все подформулы формул F и G .

Пример 3.6. Формула $F = XY \rightarrow X \vee Z$ имеет шесть подформул: X , Y , Z , XY , $X \vee Z$, $XY \rightarrow X \vee Z$.

Определение 3.3 позволяет определять соответствие формулы синтаксису логики высказываний (проводить синтаксический анализ). Введем понятие значения (смысла) формулы, т. е. семантику логики высказываний.

Рассмотрим формулу $F(X_1, X_2, \dots, X_n)$. Если задать значения всех входящих в формулу переменных,

$$X_1 = a_1, X_2 = a_2, \dots, X_n = a_n, \quad a_i \in \{0, 1\}, i = 1, 2, \dots, n, \quad (3.2)$$

то используя таблицу истинности (см. табл. 3.1) можно вычислить значение всей формулы на этом наборе:

$$F(a_1, a_2, \dots, a_n) = b, \quad b \in \{0, 1\}.$$

В этом случае говорят, что задана *интерпретация формулы* F на наборе (3.2). Такие наборы из нулей и единиц называют *булевыми наборами*.

Таким образом, чтобы определить значение формулы, нужно задать ее интерпретацию на всех возможных булевых наборах значений входящих в нее переменных.

Очевидна следующая теорема.

Теорема 3.1. Число различных булевых наборов формулы от n переменных равно 2^n .

Доказательство. Каждый набор $a = (a_1, \dots, a_n)$ можно интерпретировать как двоичную запись некоторого натурального числа

$$N(a) = (a_1, \dots, a_n)_2 = a_n + a_{n-1}2 + \cdots + a_k2^{n-k} + \cdots + a_12^{n-1}. \quad (3.3)$$

Число $N(a)$ будем называть *номером набора*. Согласно результатам примера 1.77 (см. гл. 1) число таких двоичных чисел равно 2^n . ■

Условимся в дальнейшем упорядочивать булевы наборы в лексикографическом порядке, т. е. в порядке возрастания номеров наборов. Таким образом семантическое значение формулы логики высказываний $F(X_1, X_2, \dots, X_n)$ можно задать следующей таблицей (табл. 3.2).

Таблица 3.2

X_1	X_2	\dots	X_{n-1}	X_n	$F(X_1, X_2, \dots, X_{n-1}, X_n)$
0	0	\dots	0	0	$F(0, 0, \dots, 0, 0)$
0	0	\dots	0	1	$F(0, 0, \dots, 0, 1)$
\dots	\dots	\dots	\dots	\dots	\dots
1	1	\dots	1	0	$F(1, 1, \dots, 1, 0)$
1	1	\dots	1	1	$F(1, 1, \dots, 1, 1)$

Эту таблицу называют *таблицей истинности* для формулы F . Каждая строка табл. 3.2 соответствует интерпретации формулы F на данном булевом наборе.

Пример 3.7. Ответим на вопросы, поставленные в задаче Кислера (см. пример 3.1). Согласно (3.1) составим совместную таблицу истинности для всех свидетельских показаний (табл. 3.3).

Таблица 3.3

B	J	S	$J\bar{S}$	$B \rightarrow S$	$\bar{S}(B \vee J)$	B	J	S	$J\bar{S}$	$B \rightarrow S$	$\bar{S}(B \vee J)$
0	0	0	0	1	0	1	0	0	0	0	1
0	0	1	0	1	0	1	0	1	0	1	0
0	1	0	1	1	1	1	1	0	1	0	1
0	1	1	0	1	0	1	1	1	0	1	0

Совместны ли показания? Этот вопрос уточняется так: можно ли придать переменным B, J, S такие логические значения, чтобы формулы (3.1) были истинными. Из табл. 3.3 находим, что показания совместны при $B = 0, J = 1, S = 0$ (третья строка таблицы). Таким образом доказано, что если принять гипотезу об истинности всех показаний, то виновные найдены: \bar{B}, J, \bar{S} .

Рассмотрим другую модель дополнительной гипотезы: невиновный говорит правду, а виновный — неизвестно что. Тогда истинными должны быть формулы:

$$\bar{B} \rightarrow J\bar{S}, \quad \bar{J} \rightarrow (B \rightarrow S), \quad \bar{S} \rightarrow \bar{S}(B \vee J). \quad (3.4)$$

Для решения задачи нужно заполнить таблицу, аналогичную табл. 3.3 и выделить в ней строки, где все формулы (3.4) истинны.

Задача 3.2. Рассмотрите еще одну гипотезу в задаче Кислера: виновный говорит правду, а невиновный лжет.

В задаче Кислера по синтаксической записи формул с помощью таблицы истинности логических связок (см. табл. 3.1) были получены их семантические значения. Рассмотрим пример обратной задачи: имея таблицу истинности для некоторой формулы, получить синтаксическую запись этой формулы (т. е. синтезировать текст с заданным смыслом).

Пример 3.8 (задача о путешественнике). Путешественник попал в один из двух городов и желает знать, в какой именно. Он имеет право задать один вопрос, причем ответом на него может быть «да» или «нет». Будем считать, что ему не обязательно говорят правду.

Введем две переменные X и Y :

$$X = \begin{cases} 1, & \text{Санкт-Петербург,} \\ 0, & \text{Москва.} \end{cases} \quad Y = \begin{cases} 1, & \text{собеседник правдив,} \\ 0, & \text{в противном случае.} \end{cases}$$

Путешественник должен задать вопрос: *верно ли, что $F(X, Y)$?* Таблица истинности для формулы $F(X, Y)$ очевидно выглядит следующим образом (см. табл. 3.4).

Таблица 3.4

X	Y	$F(X, Y)$	X	Y	$F(X, Y)$
0	0	1	1	0	0
0	1	0	1	1	1

Задача построения формулы $F(X, Y)$ по таблице истинности (см. табл. 3.4) будет решена в подразд. 3.1.5 (см. пример 3.22).

3.1.2. Равносильность формул логики высказываний

Нетрудно привести примеры формул, которые являясь синтаксически различными, имеют одинаковые значения. Например, формулы $X \vee Y$ и $Y \vee X$. Подобные формулы будем называть *равносильными*. Дадим точное определение.

Определение 3.5. Формулы $F(X_1, \dots, X_n)$ и $G(X_1, \dots, X_n)$ называют *равносильными* (символически обозначают $F = G$), если на любом булевом наборе (a_1, \dots, a_n) выполняется равенство $F(a_1, \dots, a_n) = G(a_1, \dots, a_n)$.

Замечание 3.3. Благодаря введению фиктивных переменных можно считать, что наборы переменных у формул F и G совпадают.

Пример 3.9. Установим, что две формулы логики высказываний $F = X \rightarrow Y$ и $G = \bar{X} \vee Y$ равносильны. Построим совместную таблицу истинности для F и G (табл. 3.5).

Таблица 3.5

X	Y	$X \rightarrow Y$	$\bar{X} \vee Y$	X	Y	$X \rightarrow Y$	$\bar{X} \vee Y$
0	0	1	1	1	0	0	0
0	1	1	1	1	1	1	1

Из табл. 3.5 видно, что столбцы формул F и G совпадают. Это означает, что формулы F и G равносильны.

Используя определение равносильности легко доказать справедливость следующей теоремы.

Теорема 3.2. Отношение равносильности на множестве формул логики высказываний есть отношение эквивалентности.

Близким к понятию равносильности является понятие тождественной истинности.

Определение 3.6. Формулу $F(X_1, \dots, X_n)$ называют *тождественно истинной*, или *тавтологией*, если на любом булевом наборе (a_1, \dots, a_n) справедливо $F(a_1, \dots, a_n) = 1$.

Пример 3.10. Покажем, что формула $F = XY \rightarrow X$ является тождественно истинной. Составим таблицу истинности формулы F (табл. 3.6).

Таблица 3.6

X	Y	$XY \rightarrow X$	X	Y	$XY \rightarrow X$
0	0	1	1	0	1
0	1	1	1	1	1

Видно, что столбец формулы F состоит из одних единиц. Это означает, что формула F тождественно истинна.

Двойственным к понятию тождественной истинности является понятие тождественной ложности.

Определение 3.7. Формулу $F(X_1, \dots, X_n)$ называют *тождественно ложной*, или *противоречием*, если на любом булевом наборе (a_1, \dots, a_n) справедливо $F(a_1, \dots, a_n) = 0$.

Понятия равносильности и тождественной истинности связывает следующая теорема.

Теорема 3.3. Формулы F и G равносильны тогда и только тогда, когда формула $F \leftrightarrow G$ является тождественно истинной.

Доказательство. Используя определение равносильности и таблицу истинности связок (см. табл. 3.1) для операции \leftrightarrow получаем требуемое. ■

В логике высказываний часто приходится преобразовывать формулы, сохраняя их равносильность. Для таких преобразований используются так называемые *законы логики высказываний*. Приведем некоторые из этих законов.

Теорема 3.4. Пусть F, G и H — некоторые формулы логики высказываний. Тогда имеют место следующие равносильности (законы).

- | | |
|---|--|
| 1) $FF = F;$ | 2) $F \vee F = F;$ |
| 3) $FG = GF;$ | 4) $F \vee G = G \vee F;$ |
| 5) $F(GH) = (FG)H;$ | 6) $F \vee (G \vee H) = (F \vee G) \vee H;$ |
| 7) $F(G \vee H) = FG \vee FH;$ | 8) $F \vee GH = (F \vee G)(F \vee H);$ |
| 9) $\overline{FF} = 0;$ | 10) $F \vee \overline{F} = 1;$ |
| 11) $\overline{FG} = \overline{F} \vee \overline{G};$ | 12) $\overline{F \vee G} = \overline{F} \overline{G};$ |
| 13) $\overline{\overline{F}} = F;$ | 14) $F \rightarrow G = \overline{F} \vee G;$ |
| 15) $F \leftrightarrow G = (F \rightarrow G)(G \rightarrow F);$ | |

Приведенные равносильности легко доказать с помощью таблиц истинности. Например, в примере 3.9 была установлена равносильность 14.

Прокомментируем список законов (равносильностей).

Законы 1 и 2 называют *идемпотентностью*, 3 и 4 — *коммутативностью*, 5 и 6 — *ассоциативностью*, соответственно, конъюнкции и дизъюнкции. Ассоциативность конъюнкции означает, что в конъюнкции трех формул скобки можно ставить как угодно, а следовательно, вообще не ставить. Из этого утверждения следует, что в конъюнкции четырех, пяти и т. д. (любого конечного числа) формул скобки можно ставить как угодно и поэтому вообще не ставить. Аналогичное замечание можно сделать и для дизъюнкций.

Закон 7 называют *дистрибутивностью* конъюнкции относительно дизъюнкции, а закон 8 — *дистрибутивностью* дизъюнкции относительно конъюнкции. Для применения этих законов в преобразованиях формул удобно иметь в виду следующий аналог. В законе 7 заменим формулы F, G и H , соответственно, буквами a, b и c , конъюнкцию — умножением $*$, дизъюнкцию — сложением $+$. Получим известное числовое тождество

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

Данное тождество есть дистрибутивность умножения чисел относительно сложения. В средней школе применение этого равенства слева направо называют раскрытием скобок, а справа налево — вынесением общего множителя. Отличие операций над высказываниями \wedge и \vee от числовых операций $*$ и $+$ состоит в том, что для высказываний выполняются обе дистрибутивности, а для чисел только одна. Сложение не дистрибутивно относительно умножения.

Закон 9 называют *законом противоречия*, закон 10 — *законом исключенного третьего*¹, законы 11 и 12 — *законами де Моргана*², закон 13 — *снятием двойного отрицания или инволютивностью*, а установленный в примере 3.9 закон 14 — *законом контрапозиции*.

Лемма 3.1 (лемма о подстановке). Если в формуле F заменить подформулу G равносильной ей формулой \hat{G} , то получим формулу \hat{F} , равносильную исходной формуле F .

Доказательство. При любой подстановке констант в формулы F и \hat{F} начиная с некоторого места процессы их вычисления совпадают, так как результаты вычисления формул G и \hat{G} совпадают по условию леммы. ■

Используя лемму 3.1 и законы логики высказываний теоремы 3.4, получаем еще один способ доказательства равносильности двух формул. Этот способ состоит в переходе от одной формулы к другой с помощью тавтологий. Проиллюстрируем второй способ на примере.

Пример 3.11. Доказать равносильность формул

$$X(Z \rightarrow Y) \vee (X \rightarrow Z)Y \text{ и } (X \vee Y)(Y \vee \bar{Z}).$$

Запишем цепочку равносильностей. Над знаком равносильности ($=$) будем ставить номера использованных тавтологий из теоремы 3.4.

$$\begin{aligned} X(Z \rightarrow Y) \vee (X \rightarrow Z)Y &\stackrel{14}{=} X(\bar{Z} \vee Y) \vee (\bar{X} \vee Z)Y \stackrel{8}{=} \\ &\stackrel{8}{=} (X(\bar{Z} \vee Y) \vee \bar{X} \vee Z) (X(\bar{Z} \vee Y) \vee Y) \stackrel{8}{=} \\ &\stackrel{8}{=} (X \vee \bar{X} \vee Z)(\bar{Z} \vee Y \vee \bar{X} \vee Z)(X \vee Y)(\bar{Z} \vee Y \vee Y) \stackrel{2,4,10}{=} \\ &\stackrel{2,4,10}{=} (X \vee Y)(Y \vee \bar{Z}). \end{aligned}$$

3.1.3. Принцип двойственности

В этом подразделе будем рассматривать формулы, содержащие только логические связки \wedge , \vee и \neg .

¹Равносильности 9 и 10 называют также законами дополнительности.

²Августус Морган (1806-1871) — шотландский математик, первый президент Лондонского математического общества, один из основателей формальной логики.

Заметим, что большинство равносильностей, приведенных в теореме 3.4, существует в двух похожих формах. Рассмотрим логические законы 1 – 12. Если конъюнкцию заменить дизъюнкцией, дизъюнкцию — конъюнкцией, 0 на 1, а 1 на 0, то равносильность, записанная в левом столбце, перейдет в равносильность, записанную в правом столбце.

Дадим формальные определения.

Определение 3.8. Операцию конъюнкции называют *двойственной* для операции дизъюнкции, а операцию дизъюнкции называют *двойственной* для операции конъюнкции.

Формулы F и F^* называют *двойственными*, если формула F^* получается из формулы F путем замены в ней каждой операции на двойственную.

Для булевого набора $a = (a_1, \dots, a_n)$ определим *двойственный* набор $a^* = (b_1, \dots, b_n)$, где $b_i = \neg a_i$, $i \in 1 : n$.

Пример 3.12. Для формулы $F = X(Y \vee Z)$ двойственной является $F^* = X \vee YZ$.

Для формулы $G = \overline{XY}$ двойственной является $G^* = \overline{X \vee Y}$.

Прежде чем ввести принцип двойственности, рассмотрим вспомогательное утверждение.

Лемма 3.2. Формула $F(X_1, \dots, X_n)$ принимает значение 1 на булевом наборе $a = (a_1, \dots, a_n)$ тогда и только тогда, когда двойственная формула $F^*(X_1, \dots, X_n)$ принимает значение 0 на двойственном наборе a^* , т. е. справедливо равенство

$$\overline{F(X_1, \dots, X_n)} = F^*(\overline{X_1}, \dots, \overline{X_n}). \quad (3.5)$$

Доказательство. Проведем индукцию по k — числу логических связок формулы F .

Если $k = 0$, то $F = F^* = X_i$ и утверждение очевидно.

Предположим, что утверждение леммы справедливо при числе логических связок меньших k и докажем его справедливость для k . Тогда формула F должна иметь вид: $\neg G$, $G \vee H$ или $G \wedge H$. Рассмотрим, например, второй случай $F = G \vee H$. Тогда $F^* = G^* \wedge H^*$. По индукционному предположению утверждение леммы справедливо для G^* и H^* . Имеем

$$\begin{aligned} \overline{F(X_1, \dots, X_n)} &= \overline{G(X_1, \dots, X_n) \vee H(X_1, \dots, X_n)} \stackrel{12}{=} \\ &\stackrel{12}{=} \overline{G(X_1, \dots, X_n)} \wedge \overline{H(X_1, \dots, X_n)} = \\ &= G^*(\overline{X_1}, \dots, \overline{X_n}) \wedge H^*(\overline{X_1}, \dots, \overline{X_n}) = F^*(\overline{X_1}, \dots, \overline{X_n}). \end{aligned}$$

Аналогично доказываются два других случая. ■

Пример 3.13. Рассмотрим формулу $F = X(Y \vee Z)$ из примера 3.12. Имеем

$$\overline{F(X, Y, Z)} = \overline{X(Y \vee Z)} \stackrel{11}{=} \overline{X} \vee \overline{(Y \vee Z)} \stackrel{12}{=} \overline{X} \vee \overline{Y} \overline{Z} = F^*(\overline{X}, \overline{Y}, \overline{Z}).$$

Следствие 3.1. Равенство (3.5) можно переписать в виде:

$$F^*(X_1, \dots, X_n) = \overline{F(\overline{X_1}, \dots, \overline{X_n})}. \quad (3.6)$$

Это соотношение может служить другим равносильным определением двойственной формулы. Из равенства (3.6) очевидно следует тождество

$$(F^*)^* = F.$$

Из леммы 3.2 и определения равносильности очевидным образом вытекает *принцип (закон) двойственности*.

Теорема 3.5 (принцип двойственности). Если формулы F и G равносильны, то равносильны и двойственные им формулы, т. е. формула F^* равносильна формуле G^* .

В качестве примера можно привести уже упоминавшиеся в начале подраздела «парные» равносильности из теоремы 3.4.

Замечание 3.4. В дальнейшем принцип двойственности и соотношения (3.5) и (3.6) будут использоваться неоднократно (см. подразд. 3.1.5 и 3.2.6). Например, установив справедливость некоторых представлений для дизъюнкции, легко составить аналогичные соотношения и для конъюнкции.

3.1.4. Логическое следствие

Рассмотрим вопрос: является ли данное утверждение следствием других. Введем необходимые понятия.

Определение 3.9. Формулу $G(X_1, \dots, X_n)$ называют *логическим следствием* формул $F_1(X_1, \dots, X_n), \dots, F_k(X_1, \dots, X_n)$, если для любой интерпретации (a_1, \dots, a_n) из того, что

$$F_1(a_1, \dots, a_n) = 1, \dots, F_k(a_1, \dots, a_n) = 1$$

следует, что $G(a_1, \dots, a_n) = 1$.

Пример 3.14. На складе совершено хищение. Подозрение пало на трех человек: Брауна, Джонсона и Смита, они были доставлены для допроса. Установлено следующее:

1. *Никто, кроме Брауна, Джонсона и Смита, не был замешан в деле.*

2. *Брауна никогда не ходит на дело без, по крайней мере, одного соучастника.*

3. *Смит не виновен.*

Ответим на вопрос: *виновен ли Джонсон?*

Обозначим через X утверждение — «Браун виновен», через Y — «Джонсон виновен», через Z — «Смит виновен». Запишем факты 1–3 с помощью формул исчисления высказываний:

$$F_1 = X \vee Y \vee Z, \quad F_2 = X \rightarrow Y \vee Z, \quad F_3 = \overline{Z}.$$

Предполагаемый ответ: «Джонсон - виновен» обозначим через G . Проверим, является ли формула G логическим следствием формул F_1, F_2, F_3 . Построим совместную таблицу истинности формул F_1, F_2, F_3 и G (табл. 3.7).

Таблица 3.7

X	Y	Z	F_1	F_2	F_3	G	X	Y	Z	F_1	F_2	F_3	G
0	0	0	0	1	1		1	0	0	1	0	1	
0	0	1	1	1	0		1	0	1	1	1	0	
0	1	0	1	1	1	1	1	1	0	1	1	1	1
0	1	1	1	1	0		1	1	1	1	1	0	

Рассматриваем только те интерпретации, в которых все формулы F_1, F_2, F_3 истинны. Таких интерпретаций только две (в табл. 3.7 они выделены серым цветом). В этих интерпретациях значение формулы G также истинно. Следовательно, формула G — логическое следствие, т. е. Джонсон виновен.

Пример 3.15. Докажем, что формула $G = Y \rightarrow X$ не является логическим следствием формул $F_1 = X \vee Y, F_2 = X \rightarrow Y, F_3 = Y$. Для этого построим совместную таблицу истинности формул F_1, F_2, F_3 и G (табл. 3.8).

Таблица 3.8

X	Y	F_1	F_2	F_3	G	X	Y	F_1	F_2	F_3	G
0	0	0	1	0	1	1	0	1	0	0	1
0	1	1	1	1	0	1	1	1	1	1	1

Рассмотрим интерпретацию $X = 0, Y = 1$. Из табл. 3.8 видно, что $F_1(0, 1) = F_2(0, 1) = F_3(0, 1) = 1$, но $G(0, 1) = 0$. Следовательно, формула G не является логическим следствием формул F_1, F_2, F_3 .

Понятие логического следствия тесно связано с понятием выполнимости.

Определение 3.10. Множество формул

$$\{F_1(X_1, \dots, X_n), F_2(X_1, \dots, X_n), \dots, F_k(X_1, \dots, X_n)\}$$

называют *выполнимым*, если существует интерпретация (a_1, \dots, a_n) такая, что

$$F_1(a_1, \dots, a_n) = F_2(a_1, \dots, a_n) = \dots = F_k(a_1, \dots, a_n) = 1.$$

Проверить выполнимость множества формул $\{F_1, F_2, \dots, F_k\}$ можно построением совместной таблицы истинности этих формул. Если найдется хотя бы одна строка, в которой в столбцах формул F_1, F_2, \dots, F_k стоят единицы, то это множество формул выполнимо. Если такой строки нет, то множество формул невыполнимо.

Так, множество формул $\{F_1, F_2, F_3, G\}$ из примера 3.15 выполнимо, поскольку в табл. 3.8 в последней строке в столбцах этих формул стоят единицы.

В дальнейшем будет необходимо следующее утверждение.

Теорема 3.6. Формула $G(X_1, \dots, X_n)$ является логическим следствием формул $F_1(X_1, \dots, X_n), F_2(X_1, \dots, X_n), \dots, F_k(X_1, \dots, X_n)$ тогда и только тогда, когда множество формул

$$L = \{F_1(X_1, \dots, X_n), F_2(X_1, \dots, X_n), \dots, F_k(X_1, \dots, X_n), \overline{G(X_1, \dots, X_n)}\}$$

невыполнимо.

Доказательство. Пусть формула G является следствием множества формул F_1, \dots, F_k . Предположим, от противного, что множество L выполнимо. Это означает, что существует интерпретация (a_1, \dots, a_n) такая, что

$$F_1(a_1, \dots, a_n) = \dots = F_k(a_1, \dots, a_n) = \overline{G(a_1, \dots, a_n)} = 1. \quad (3.7)$$

Так как G — логическое следствие формул F_1, \dots, F_k из (3.7) следует, что $G(a_1, \dots, a_n) = 1$. Полученное противоречие доказывает, что множество формул $\{F_1, \dots, F_k, \overline{G}\}$ невыполнимо.

Пусть теперь множество формул L невыполнимо. Рассмотрим интерпретацию (a_1, \dots, a_n) такую, что

$$F_1(a_1, \dots, a_n) = \dots = F_k(a_1, \dots, a_n) = 1. \quad (3.8)$$

Поскольку L невыполнимо, то $\overline{G(a_1, \dots, a_n)} = 0$. Тогда $G(a_1, \dots, a_n) = 1$. Следовательно, из равенств (3.8) следует равенство $G(a_1, \dots, a_n) = 1$. Это означает, что G — логическое следствие множества формул F_1, \dots, F_k . ■

Следствие 3.2. Формула G является логическим следствием формул F_1, \dots, F_k тогда и только тогда, когда формула $F_1 \wedge \dots \wedge F_k \wedge \overline{G}$ есть противоречие.

Пример 3.16. На некотором острове, населенном рыцарями и лжецами (рыцари говорят только правду, лжецы — только ложь), разнесся слух о том, что на нем зарыты сокровища. Путешественник прибывает на остров и спрашивает одного из местных жителей: *Есть ли на острове сокровища?*. В ответ на вопрос путешественника он заявляет: *Сокровища на этом острове есть в том и только в том случае, если я рыцарь*.

Можно ли определить, есть ли сокровища на острове?

Обозначим через X утверждение «Местный житель — рыцарь», через Y — «Сокровища на острове есть». Тогда высказывание местного жителя «Сокровища на острове есть в том и только в том случае если, я — рыцарь» запишется формулой: $X \leftrightarrow Y$. Тогда, если он — рыцарь, то его высказывание истинно, если лжец, то ложно.

Таким образом, вопрос сводится к следующему: является ли формула Y логическим следствием формулы $X \leftrightarrow (X \leftrightarrow Y)$. Для проверки этого утверждения достаточно на основании следствия 3.2 к теореме 3.6 доказать, что формула $(X \leftrightarrow (X \leftrightarrow Y))\overline{Y}$ есть противоречие. Построим ее таблицу истинности (табл. 3.9).

Таблица 3.9

X	Y	$(X \leftrightarrow (X \leftrightarrow Y))\overline{Y}$	X	Y	$(X \leftrightarrow (X \leftrightarrow Y))\overline{Y}$
0	0	0	1	0	0
0	1	0	1	1	0

Формула $(X \leftrightarrow (X \leftrightarrow Y))\overline{Y}$ ложна в любой интерпретации, следовательно, она есть противоречие, и поэтому верно утверждение, что сокровища на острове есть.

Замечание 3.5. Примеры 3.14 и 3.16 взяты из [45].

3.1.5. Нормальные формы в логике высказываний

Во многих случаях удобно рассматривать вместо формул общего вида формулы в некоторой канонической форме. Дадим необходимые определения.

Определение 3.11. Литералом называют переменную или ее отрицание, элементарной конъюнкцией — литерал или конъюнкцию литералов.

Определение 3.12. Формулу, в которой знак отрицания находится только перед переменными, называют формулой с *тесными отрицаниями*.

Пример 3.17. Формулы $X \leftrightarrow \bar{Y}$, $\bar{X} \rightarrow \bar{Y}$ являются формулами с тесными отрицаниями, а формулы $\bar{X}\bar{Y}$, $X \vee \bar{Y} \vee Z$ — нет.

Определение 3.13. Формула G имеет *дизъюнктивную нормальную форму* (ДНФ), если она является дизъюнкцией элементарных конъюнкций.

Пример 3.18. Формулы X , \bar{Y} , $X\bar{Y}$, $(X\bar{Y}) \vee (\bar{X}Z)$ имеют ДНФ, а формулы $\bar{X}\bar{Y}$, $X \vee Y \vee 0$, $X \rightarrow Y$ — нет.

Теорема 3.7. Для любой формулы F существует формула G , равносильная F и имеющая дизъюнктивную нормальную форму.

Доказательство. Следует из рассмотрения алгоритма 3.1, который по данной формуле F выдает формулу G , удовлетворяющую условию теоремы. ■

Алгоритм 3.1 (приведение к ДНФ).

Шаг 1. Используя законы 15 и 14, исключаем из исходной формулы эквиваленцию и импликацию.

Шаг 2. С помощью законов 11 и 12 преобразуем формулу к формуле с тесными отрицаниями.

Шаг 3. Если формула содержит подформулу вида $F(G \vee H)$, то согласно закону 7 (см. теорему 3.4) заменяем ее на равносильную формулу $FG \vee FH$.

Пример 3.19. Применение алгоритма 3.1 проиллюстрируем на примере формулы $F = (X \leftrightarrow Y)X$. Как и ранее (см. пример 3.11) в цепочке равносильностей будем ставить номера использованных тавтологий из теоремы 3.4.

Выполним первый шаг.

$$\overline{(X \leftrightarrow Y)}X \stackrel{15}{=} \overline{(X \rightarrow Y)(Y \rightarrow X)}X \stackrel{14}{=} \overline{(\bar{X} \vee Y)(\bar{Y} \vee X)}X.$$

Перейдем ко второму шагу.

$$\overline{(\bar{X} \vee Y)(\bar{Y} \vee X)}X \stackrel{11}{=} \left(\overline{(\bar{X} \vee Y)} \vee \overline{(\bar{Y} \vee X)} \right) X \stackrel{12,13}{=} (X\bar{Y} \vee Y\bar{X})X.$$

Выполнение третьего шага заключается в применении дистрибутивности.

$$(X\bar{Y} \vee Y\bar{X})X \stackrel{7}{=} X\bar{Y}X \vee Y\bar{X}X = G.$$

Алгоритм на этом завершен. Формула G имеет ДНФ. Но эту формулу можно упростить. Действительно, формула $Y\bar{X}X = 0$, а формула $X\bar{Y}X = X\bar{Y}$. Тогда

$$X\bar{Y}X \vee Y\bar{X}X = X\bar{Y} = \hat{G}.$$

Формула \hat{G} , как и G , имеет ДНФ и равносильна исходной формуле F .

Пример 3.19 показывает, что может существовать несколько равносильных формул имеющих разное представление в дизъюнктивной нормальной форме. Иногда это обстоятельство является неудобным. Чтобы его исключить, вводится более узкое понятие, чем ДНФ.

Определение 3.14. Формула $F(X_1, \dots, X_n)$ имеет *совершенную дизъюнктивную нормальную форму* (СДНФ), если выполнены следующие условия:

- 1) F имеет дизъюнктивную нормальную форму;
- 2) каждая элементарная конъюнкция ДНФ содержит один и только один из литералов X_i или \bar{X}_i для любого $i = 1, \dots, n$;
- 3) F не содержит одинаковых элементарных конъюнкций.

Пример 3.20. Формулы X , $\bar{X}Y$, $\bar{X}Y \vee X\bar{Y}$ имеют СДНФ.

Формулы $\bar{X}\bar{Y}$, $XY \vee \bar{X}Z$, $XY \vee X\bar{Y} \vee YX$ не имеют СДНФ. Для первой формулы не выполняется первое условие, для второй — второе условие, для третьей формулы — последнее условие из определения СДНФ.

Теорема 3.8. Для любой выполнимой формулы F существует равносильная ей формула G , имеющая совершенную дизъюнктивную нормальную форму.

Доказательство. Как и теорема 3.7, доказательство данной теоремы следует из алгоритма 3.2, который по формуле F выдает формулу G , удовлетворяющую условию теоремы. ■

Алгоритм 3.2 (первый алгоритм приведения к СДНФ).

Шаг 1. Применяя алгоритм 3.1, приводим формулу к ДНФ.

Шаг 2. Если в полученной ДНФ какая-то элементарная конъюнкция C не содержит ни переменной X_i , ни ее отрицания для $i \in 1 : n$, то равносильным преобразованием заменяем C на две элементарные конъюнкции

$$C = C(X_i \vee \bar{X}_i) = CX_i \vee C\bar{X}_i. \quad (3.9)$$

Шаг 3. Если элементарная конъюнкция C содержит два вхождения одного литерала, то одно из них вычеркиваем. Если же конъюнкция C содержит X_i и \overline{X}_i для $i \in 1 : n$, то вычеркиваем всю элементарную конъюнкцию.

Шаг 4. Если полученная формула содержит одинаковые элементарные конъюнкции, то оставляем одну из них, а остальные вычеркиваем.

Замечание 3.6. Преобразование (3.9) является равносильным, в силу очевидных равенств $X_i \vee \overline{X}_i = 1$ и $C \wedge 1 = C$.

Пример 3.21. Применим алгоритм 3.2 к формуле $F = XY \vee X\overline{Z}$.

Эта формула уже имеет ДНФ, поэтому выполнение алгоритма начинается с шага 2. Применив преобразование (3.9) к XY и $X\overline{Z}$, имеем

$$XY \vee X\overline{Z} = XYZ \vee XY\overline{Z} \vee X\overline{Z}Y \vee X\overline{Z}\overline{Y}.$$

Однаковых и противоположных литералов в элементарных конъюнкциях полученной формулы нет, поэтому шаг 3 не выполняется.

Формула содержит одинаковые элементарные конъюнкции — вторую и третью. При выполнении четвертого шага будет оставлена одна из них. Получается формула G

$$G = XYZ \vee XY\overline{Z} \vee X\overline{Z}\overline{Y}.$$

Формула G равносильна F и имеет СДНФ.

Замечание 3.7. Требование выполнимости формулы F в формулировке теоремы 3.8 существенно.

Покажем, что если формула $F(X_1, \dots, X_n)$ невыполнима, то после ее приведения к ДНФ каждая элементарная конъюнкция будет содержать хотя бы одну пару противоположных литералов X и \overline{X} .

Предположим, что в полученной ДНФ найдется элементарная конъюнкция, не содержащая противоположных литералов. Пусть это конъюнкция $X_{i_1} \wedge \dots \wedge X_{i_k}$. Тогда для интерпретации (a_1, \dots, a_n) , где

$$a_i = \begin{cases} 1, & i = i_1, \dots, i_k, \\ 0, & \text{иначе} \end{cases}$$

справедливо равенство $F(a_1, \dots, a_n) = 1$, что противоречит невыполнимости формулы F .

Так как каждая элементарная конъюнкция будет содержать хотя бы одну пару противоположных литералов, то на шаге 3 алгоритма 3.2 все элементарные конъюнкции будут вычеркнуты.

Имеется еще один способ приведения к СДНФ, основанный на построении таблицы истинности исходной формулы. Введем обозначения.

$$X^\alpha = \begin{cases} X, & \text{если } \alpha = 1, \\ \overline{X}, & \text{если } \alpha = 0. \end{cases}$$

Алгоритм 3.3 (второй алгоритм приведения к СДНФ).

Шаг 1. Составим таблицу истинности формулы $F(X_1, \dots, X_n)$.

Шаг 2. Выделим строки таблицы, в которых в столбце F стоит 1. (Хотя бы одна такая строка должна быть, так как формула F выполнима.)

Шаг 3. Каждой выделенной строке (a_1, \dots, a_n) поставим в соответствие элементарную конъюнкцию $X_1^{a_1} \cdots X_n^{a_n}$.

Шаг 4. Составим дизъюнкцию, определенных на шаге 3 элементарных конъюнкций. Полученная формула G равносильна F и имеет СДНФ.

Замечание 3.8. Очевидно, что каждая элементарная конъюнкция, построенная на шаге 3, принимает значение 1 только на наборе (a_1, \dots, a_n) . Поэтому формула G принимает значение 1 только на интерпретациях, выделенных на шаге 2 алгоритма. Отсюда следует, что таблицы истинности F и G совпадают.

Элементарную конъюнкцию $X_1^{a_1} \cdots X_n^{a_n}$ называют *конституентой единицы* набора (a_1, \dots, a_n) (символически обозначают K_{a_1, \dots, a_n}^1).

Пример 3.22. Вернемся к задаче о путешественнике (см. пример 3.8). Как было установлено, таблица истинности формулы $F(X, Y)$ имеет вид (табл. 3.10).

Таблица 3.10

X	Y	$F(X, Y)$	X	Y	$F(X, Y)$
0	0	1	1	0	0
0	1	0	1	1	1

Применив алгоритм 3.3, имеем $F(X, Y) = XY \vee \overline{X} \overline{Y}$. Вопрос, который должен задать путешественник звучит так: «Правда ли, что это Санкт-Петербург и ты правдив, или это Москва и ты лжешь».

Пример 3.23. С помощью алгоритма 3.3 построим СДНФ для формулы $F(X_1, X_2, X_3) = X_1(X_2 \rightarrow X_3)$.

Составим таблицу истинности формулы F (табл. 3.11).

Таблица 3.11

X_1	X_2	X_3	$X_1(X_2 \rightarrow X_3)$	X_1	X_2	X_3	$X_1(X_2 \rightarrow X_3)$
0	0	0	0	1	0	0	1
0	0	1	0	1	0	1	1
0	1	0	0	1	1	0	0
0	1	1	0	1	1	1	1

Применив алгоритм 3.3 получим искомое представление:

$$F(X_1, X_2, X_3) = X_1(X_2 \rightarrow X_3) = X_1 X_2 X_3 \vee X_1 \overline{X_2} X_3 \vee X_1 \overline{X_2} \overline{X_3}.$$

Рассмотрим еще одну каноническую форму записи формул логики высказываний — конъюнктивную нормальную форму. Она получается из ДНФ заменой \wedge на \vee и \vee на \wedge . Дадим точные определения.

Определение 3.15. Элементарной дизъюнкцией, или дизъюнктом называют литерал или дизъюнкция литералов.

Замечание 3.9. Очевидно, что элементарная дизъюнкция является двойственной к элементарной конъюнкции. Это свойство позволяет свести определяемые новые понятия к ДНФ и СДНФ.

Определение 3.16. Формула G имеет конъюнктивную нормальную форму (КНФ), если она является конъюнкцией элементарных дизъюнкций.

Замечание 3.10. Используя понятие двойственной формулы можно дать и другое равносильное определение КНФ. Формула G имеет КНФ, если двойственная ей формула G^* определена (т. е. не содержит связок \rightarrow и \leftrightarrow) и находится в ДНФ.

Пример 3.24. Формулы X , \overline{Y} , $X \vee \overline{Y}$, $X\overline{Y}$, $(X \vee \overline{Y})(X \vee Z)$ имеют КНФ, а формулы $X \rightarrow Y$, $\overline{X} \vee \overline{Y}$, $X\overline{Y} \vee X\overline{Z}$ — нет.

Для КНФ справедливо утверждение, аналогичное теореме 3.7.

Теорема 3.9. Для любой формулы F существует формула G , равносильная F и имеющая конъюнктивную нормальную форму.

Доказательство. Доказательство теоремы легко следует из анализа алгоритма приведения к КНФ. Этот алгоритм легко получить из алгоритма 3.1 приведения к ДНФ, если на шаге 3 вместо закона дистрибутивности 7 (см. теорему 3.4) использовать двойственную равносильность 8, т. е. если формула содержит подформулу вида $F \vee GH$, то заменить ее на формулу $(F \vee G)(F \vee H)$. ■

Как и для случая ДНФ введем более узкое понятие, нежели КНФ.

Определение 3.17. Формула $F(X_1, \dots, X_n)$ имеет совершенную конъюнктивную нормальную форму (СКНФ), если выполнены следующие условия:

- 1) F имеет КНФ;
- 2) каждая элементарная дизъюнкция содержит один и только один из литералов X_i или $\neg X_i$ для любого $i \in 1 : n$;
- 3) F не содержит одинаковых элементарных дизъюнкций.

Замечание 3.11. Как и для КНФ можно дать другое равносильное определение СКНФ используя понятие двойственности. Формула G имеет СКНФ, если двойственная ей формула G^* находится в СДНФ.

Пример 3.25. Формулы X , $\overline{X} \vee Y$, $(\overline{X} \vee Y)(X \vee \overline{Y})$ имеют СКНФ.

Формулы $\overline{X \vee Y}$, $(X \vee Y)(\overline{X} \vee Z)$, $(X \vee Y)(X \vee \overline{Y})(Y \vee X)$ не имеют СКНФ. Для первой формулы не выполняется первое условие, для второй формулы — второе условие, для третьей формулы — последнее условие из определения СКНФ.

Теорема 3.10. Для любой не тождественно истинной формулы F существует равносильная ей формула G , имеющая совершенную конъюнктивную нормальную форму.

Доказательство. Как и в теореме 3.8, доказательство этой теоремы легко следует из алгоритма 3.4, который по формуле F выдает формулу G , удовлетворяющую условию теоремы. ■

Алгоритм 3.4 (первый алгоритм приведения к СКНФ).

Шаг 1. Применив первые два шага алгоритма 3.1 и теорему 3.9, приведем формулу к КНФ.

Шаг 2. Если в полученной на первом шаге КНФ какая-либо элементарная дизъюнкция D не содержит ни переменной X_i , ни ее отрицания для некоторого $i \in 1 : n$, то равносильным преобразованием заменяем D на две элементарные дизъюнкции

$$D \vee X_i \overline{X}_i = (D \vee X_i)(D \vee \overline{X}_i). \quad (3.10)$$

Шаг 3. Если элементарная дизъюнкция D содержит два вхождения одного литерала, то одно из них вычеркиваем. Если же дизъюнкция D содержит X_i и \overline{X}_i для некоторого $i \in 1 : n$, то вычеркиваем всю элементарную дизъюнкцию.

Шаг 4. Если полученная формула содержит одинаковые элементарные дизъюнкции, то оставляем одну из них, а остальные вычеркиваем.

Замечание 3.12. На шаге 2 алгоритма 3.2 приведения к СДНФ используется равносильное преобразование (3.9). В силу двойственности СДНФ и СКНФ на том же шаге алгоритма 3.4 применяется двойственное равносильное преобразование (3.10).

Преобразование (3.10) является равносильным, в силу очевидных равенств

$$X_i \overline{X}_i = 0, \quad D \vee 0 = D.$$

Пример 3.26. В качестве примера работы алгоритма 3.4 рассмотрим формулу $F(X, Y, Z) = (X \vee Z)(X \rightarrow Y)$.

Приводим формулу к КНФ.

$$\overline{(X \vee Z)}(X \rightarrow Y) \stackrel{14}{=} \overline{(X \vee Z)}(\overline{X} \vee Y) \stackrel{12}{=} \overline{X} \overline{Z}(\overline{X} \vee Y).$$

Эта формула имеет КНФ, но не имеет СКНФ. Построим ее СКНФ. Применя преобразование (3.10), имеем:

$$\begin{aligned} \overline{X} \overline{Z}(\overline{X} \vee Y) &= (\overline{X} \vee Y \overline{Y} \vee Z \overline{Z})(X \overline{X} \vee Y \overline{Y} \vee \overline{Z})(\overline{X} \vee Y \vee Z \overline{Z}) = \\ &= (\overline{X} \vee Y \vee Z)(\overline{X} \vee \overline{Y} \vee Z)(\overline{X} \vee Y \vee \overline{Z}) \wedge \\ &\quad \wedge (\overline{X} \vee \overline{Y} \vee \overline{Z})(X \vee Y \vee \overline{Z})(X \vee \overline{Y} \vee \overline{Z}). \end{aligned}$$

Задание 3.1. Докажите, что требование не тождественной истинности формулы F в формулировке теоремы 3.10 существенно.

Имеется еще один способ приведения к СКНФ, основанный на построении таблицы истинности исходной формулы. Он очевидным образом следует из алгоритма 3.3 с учетом двойственности СДНФ и СКНФ и соотношения (3.5).

Алгоритм 3.5 (Второй алгоритм приведения к СКНФ).

Шаг 1. Составляем таблицу истинности формулы $F(X_1, \dots, X_n)$.

Шаг 2. Выделяем строки таблицы, в которых в столбце F стоит 0. (Хотя бы одна такая строка должна быть, так как формула F не тавтология).

Шаг 3. Каждой выделенной строке (a_1, \dots, a_n) поставим в соответствие элементарную дизъюнкцию $\overline{X_1^{a_1}} \vee \dots \vee \overline{X_n^{a_n}}$.

Шаг 4. Составим конъюнкцию, определенных на шаге 3 элементарных дизъюнкций. Полученная формула G равносильна F и имеет СКНФ.

Замечание 3.13. Очевидно, что каждая элементарная дизъюнкция, построенная на шаге 3, принимает значение 0 только на наборе (a_1, \dots, a_n) . Поэтому формула G принимает значение 0 только на интерпретациях, выделенных на шаге 2 алгоритма. Отсюда следует, что таблицы истинности F и G совпадают.

Элементарную дизъюнкцию $\overline{X_1^{a_1}} \vee \dots \vee \overline{X_n^{a_n}}$ называют *конституентой нуля* набора (a_1, \dots, a_n) (символически обозначают K_{a_1, \dots, a_n}^0).

Очевидно, что $\overline{K_{a_1, \dots, a_n}^0} = K_{a_1, \dots, a_n}^1$.

Пример 3.27. С помощью алгоритма 3.5 построим СКНФ, для формулы $F(X_1, X_2, X_3) = (\overline{X}_1 \rightarrow X_3) \rightarrow (\overline{X}_2 \rightarrow \overline{X}_1)$.

Составим таблицу истинности формулы F (табл. 3.12).

Таблица 3.12

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

Применив алгоритм 3.5 получим искомое представление:

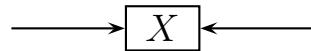
$$F(X_1, X_2, X_3) = (X_1 \vee X_2 \vee \overline{X}_3)(X_1 \vee \overline{X}_2 \vee \overline{X}_3)(\overline{X}_1 \vee \overline{X}_2 \vee X_3)(\overline{X}_1 \vee \overline{X}_2 \vee \overline{X}_3)$$

3.1.6. Контактные схемы

В этом подразделе рассмотрим применение логики высказываний к анализу так называемых контактных схем.

Определение 3.18. *Контактом* будем называть устройство, которое в процессе работы может быть в двух состояниях: контакт замкнут или разомкнут.

Контакт X на чертеже будем изображать следующим образом:



Контакты можно соединять между собой различными способами (рис. 3.1).

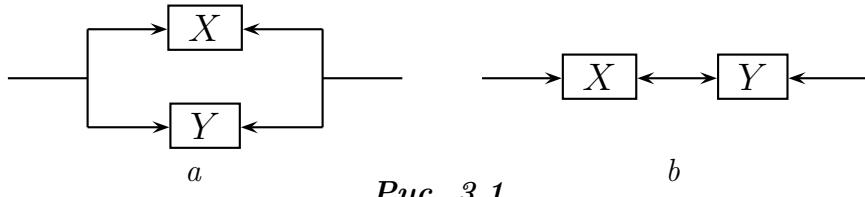


Рис. 3.1

Первое соединение контактов (рис. 3.1, а) называют *параллельным*, второе (рис. 3.2, б) — *последовательным*. Контакты, соединенные между собой, называют *контактной схемой*. Будем предполагать наличие у схемы двух выделенных точек входа и выхода. Схему назовем замкнутой, если существует последовательность замкнутых контактов X_1, X_2, \dots, X_n такая, что X_i соединен с X_{i+1} , X_1 — с входом, X_n — с выходом. Схему, не являющуюся замкнутой, назовем разомкнутой.

Каждому контакту поставим в соответствие переменную логики высказываний, которая равна 1 тогда и только тогда, когда контакт замкнут.

Переменную и контакт будем обозначать одной буквой. Пусть схема S построена из контактов X_1, X_2, \dots, X_n с помощью параллельного и последовательного соединений. Тогда по схеме S можно построить формулу логики высказываний F_S так, что параллельному соединению соответствует дизъюнкция, последовательному — конъюнкция. Например, схеме S на рис. 3.2 соответствует формула

$$F_S = X(Z \vee \bar{Y}) \vee \bar{X}Z \vee (X \vee \bar{Y})\bar{Z}.$$

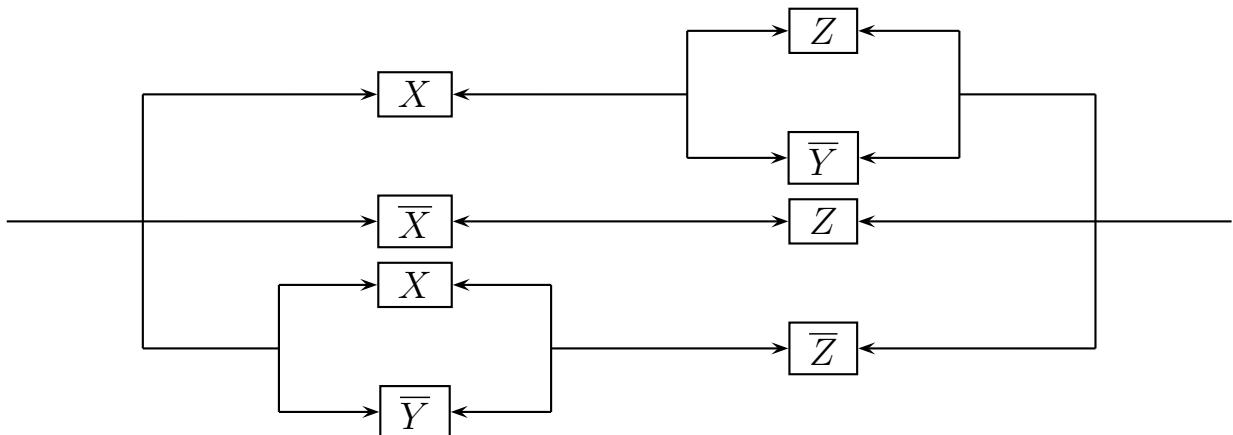


Рис. 3.2

Через \bar{X} обозначают контакт, который замкнут тогда и только тогда, когда X разомкнут. Формула F_S «представляет схему» в следующем смысле: схема S замкнута в том и только в том случае, если F_S принимает значение 1.

В силу определения контактным схемам соответствуют формулы в дизъюнктивной или конъюнктивной нормальных формах. Нетрудно понять, что по всякой такой формуле F можно восстановить схему, которую формула F «представляет».

Пусть схемам S и T соответствуют формулы F_S и F_T в описанном ранее смысле. Тогда, если схемы S и T эквивалентны (т. е. замкнуты и разомкнуты одновременно), то F_S и F_T равносильны, и обратно. Этот факт используется для решения задачи минимизации контактных схем, которая состоит в том, чтобы по данной схеме S найти схему T , эквивалентную S и содержащую меньше контактов. Один из путей решения этой задачи состоит в переходе к формуле F_S и в отыскании формулы G , равносильной F_S и содержащей меньше вхождений переменных.

Например, формула F_S равносильна формуле $X \vee \bar{X}Z \vee \bar{Y}\bar{Z}$. Следовательно, схема, приведенная на рис. 3.2 эквивалентна схеме, которая имеет на три контакта меньше (рис. 3.3).

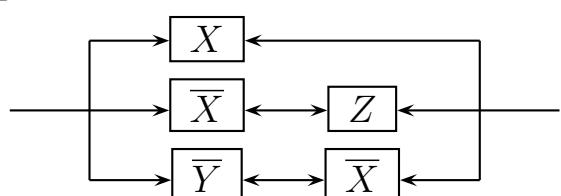


Рис. 3.3

Пример 3.28. Требуется, чтобы включение света в комнате осуществлялось с помощью трех различных выключателей таким образом, чтобы нажатие на любой из них приводило к включению света, если он был выключен, и выключению, если он был включен. Построить контактную схему, удовлетворяющую этому требованию.

Обозначим выключатели переменными X_1, X_2 и X_3 , а формулу для контактной схемы $F(X_1, X_2, X_3)$. Построим ее таблицу истинности. Очевидно, что $F(0, 0, 0) = 0$. Далее заметим, что если значения переменных X_1, X_2 и X_3 в каждой следующей строке отличаются от предыдущей только в одной позиции, то значение формулы F меняется на противоположное. Это отражает требование о том, чтобы при нажатии на любой из выключателей свет выключался, если он был включен, и включался, если он был выключен.

Составим таблицу истинности для данного порядка строк (табл. 3.13).

Таблица 3.13

X	Y	Z	$F(X_1, X_2, X_3)$	X	Y	Z	$F(X_1, X_2, X_3)$
0	0	0	0	1	0	1	0
0	1	0	1	1	0	0	1
0	1	1	0	1	1	0	0
0	0	1	1	1	1	1	1

По таблице истинности, используя алгоритм 3.3, построим СДНФ для формулы $F(X_1, X_2, X_3)$:

$$F(X_1, X_2, X_3) = XYZ \vee X\bar{Y}\bar{Z} \vee \bar{X}\bar{Y}Z \vee \bar{X}YZ.$$

Преобразуем формулу F к равносильной формуле G , более удобной для реализации схемы:

$$G = X(YZ \vee \bar{Y}\bar{Z}) \vee \bar{X}(\bar{Y}Z \vee Y\bar{Z}).$$

По формуле G построим искомую схему (рис. 3.4).

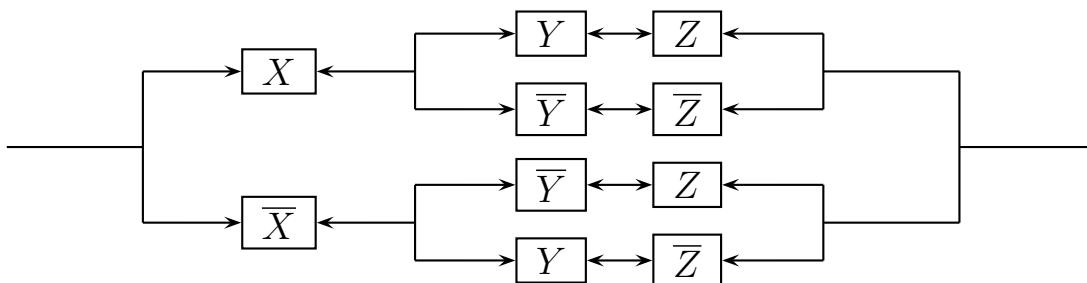


Рис. 3.4

Задача 3.3. Пусть каждый из трех членов комитета голосует «за», нажимая кнопку. Постройте по возможности более простую цепь, которая была бы замкнута тогда и только тогда, когда не менее двух членов комитета голосуют «за».

3.1.7. Метод минимизирующих карт

Определение 3.19. Дизъюнктивную нормальную форму называют *минимальной*, если она содержит наименьшее общее число вхождений 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 вхождений переменных. Однако при большом числе переменных такой перебор практически невыполним. Существуют эффективные способы нахождения минимальной ДНФ. Рассмотрим один из них — *метод минимизирующих карт*¹. Хотя он и не является самым эффективным, зато прост для изложения и не требует введения дополнительных понятий.

Пусть формула задана таблицей истинности или СДНФ. Составим следующую карту (табл. 3.14).

Таблица 3.14

\bar{X}_1	\bar{X}_2	...	\bar{X}_n	$\bar{X}_1 \bar{X}_2$	$\bar{X}_1 \bar{X}_3$...	$\bar{X}_{n-1} \bar{X}_n$...	$\bar{X}_1 \cdots \bar{X}_{n-1} \bar{X}_n$
\bar{X}_1	\bar{X}_2	...	X_n	$\bar{X}_1 X_2$	$\bar{X}_1 X_3$...	$\bar{X}_{n-1} X_n$...	$\bar{X}_1 \cdots \bar{X}_{n-1} X_n$
...
X_1	\bar{X}_2	...	\bar{X}_n	$X_1 \bar{X}_2$	$X_1 \bar{X}_3$...	$\bar{X}_{n-1} X_n$...	$X_1 \cdots \bar{X}_{n-1} \bar{X}_n$
...
X_1	X_2	...	\bar{X}_n	$X_1 X_2$	$X_1 X_3$...	$X_{n-1} \bar{X}_n$...	$X_1 \cdots X_{n-1} \bar{X}_n$
X_1	X_2	...	X_n	$X_1 X_2$	$X_1 X_3$...	$X_{n-1} X_n$...	$X_1 \cdots X_{n-1} X_n$

Замечание 3.14. Каждую строку табл. 3.14 определяет один из булевых наборов

$$\alpha_1, \alpha_2, \dots, \alpha_n, \quad \alpha_i = 0, 1, \quad i \in 1 : n. \quad (3.11)$$

Таким образом, всего в табл. 3.14 — 2^n строк, столько же, сколько в таблице истинности для формулы от n переменных.

¹ Вариант такого подхода известен под названием *метод неопределенных коэффициентов* или *гарвардский метод*.

Столбцы карты заполняют сначала одиночными литералами, соответствующими набору (3.11): $X_1^{\alpha_1}, X_2^{\alpha_2}, \dots, X_n^{\alpha_n}$ — их всего C_n^1 . Затем добавляют всевозможные конъюнкции из двух различных литералов: $X_1^{\alpha_1}X_2^{\alpha_2}, \dots, X_{n-1}^{\alpha_{n-1}}X_n^{\alpha_n}$ — их всего C_n^2 , далее — все конъюнкции из трех различных литералов и т. д. В последний столбец записывают конституенту единицы набора (3.11) — $X_1^{\alpha_1}X_2^{\alpha_2}\dots X_n^{\alpha_n}$. Всего в табл. 3.14

$$C_n^1 + C_n^2 + \dots + C_n^n = 2^n - 1$$

столбцов. При этом размеры и содержание карты *не зависит* от вида формулы и определяется только количеством ее переменных.

Строки табл. 3.14 будем записывать в обычном лексикографическом порядке, т. е. в порядке возрастания номеров булевых наборов (3.11).

Теорема 3.11. Если элементарная конъюнкция $X_1^{\alpha_1}X_2^{\alpha_2}\dots X_n^{\alpha_n}$, принадлежащая j -й строке (см. табл. 3.14, не входит в СДНФ, выражающую формулу $F(X_1, X_2, \dots, X_n)$, то любая конъюнкция j -й строки не входит ни в какую ДНФ, выражающую исходную формулу F .

Доказательство. Действительно, если конъюнкция $X_1^{\alpha_1}X_2^{\alpha_2}\dots X_n^{\alpha_n}$ не входит в СДНФ, выражающую формулу $F(X_1, X_2, \dots, X_n)$, то, согласно второму алгоритму построения СДНФ $F(\alpha_1, \alpha_2, \dots, \alpha_n) = 0$. Если бы какая-то конъюнкция j -й строки вошла в некоторую ДНФ, выращивающую формулу F , то $F(\alpha_1, \alpha_2, \dots, \alpha_n) = 1$. ■

Согласно теореме 3.11 опишем способ построения минимальной ДНФ.

Алгоритм 3.6 (метод минимизирующих карт).

Шаг 1. Отмечаем в табл. 3.14 строки, в которых соответствующая им конъюнкция последнего столбца $X_1^{\alpha_1}X_2^{\alpha_2}\dots X_n^{\alpha_n}$ не принадлежит СДНФ. Вычеркиваем *все* конъюнкции в этих строках.

Шаг 2. Конъюнкции, вычеркнутые в указанных на *шаге 1* строках, вычеркиваем также во всех остальных строках таблицы.

Шаг 3. В каждой строке выбираем из оставшихся конъюнкций лишь те, которые имеют наименьшее число вхождений переменных а остальные вычеркнем.

Шаг 4. В каждой строке выбираем по одной оставшейся конъюнкции и составляем из них ДНФ.

Шаг 5. Из всех ДНФ, полученных на предыдущем шаге, выбираем минимальную.

Замечание 3.15. Заметим, что на последнем шаге алгоритма предусматривается перебор различных ДНФ, для нахождения минимальной из них. Однако при этом вариантов перебора, как правило значительно меньше, чем в случае, когда перебираются все равносильные ДНФ.

Пример 3.29. Пусть

$$F(X_1, X_2, X_3) = X_1 X_2 \bar{X}_3 \vee X_1 \bar{X}_2 X_3 \vee X_1 \bar{X}_2 \bar{X}_3 \vee \bar{X}_1 \bar{X}_2 X_3.$$

Составим соответствующую таблицу (табл. 3.15).

Таблица 3.15

\bar{X}_1	\bar{X}_2	\bar{X}_3	$\bar{X}_1 \bar{X}_2$	$\bar{X}_1 \bar{X}_3$	$\bar{X}_2 \bar{X}_3$	$\bar{X}_1 \bar{X}_2 \bar{X}_3$
\bar{X}_1	\bar{X}_2	X_3	$\bar{X}_1 \bar{X}_2$	$\bar{X}_1 X_3$	$\bar{X}_2 X_3$	$\bar{X}_1 \bar{X}_2 X_3$
\bar{X}_1	X_2	\bar{X}_3	$\bar{X}_1 X_2$	$\bar{X}_1 \bar{X}_3$	$X_2 \bar{X}_3$	$\bar{X}_1 X_2 \bar{X}_3$
\bar{X}_1	X_2	X_3	$\bar{X}_1 X_2$	$\bar{X}_1 X_3$	$X_2 X_3$	$\bar{X}_1 X_2 X_3$
X_1	\bar{X}_2	\bar{X}_3	$X_1 \bar{X}_2$	$X_1 \bar{X}_3$	$\bar{X}_2 \bar{X}_3$	$X_1 \bar{X}_2 \bar{X}_3$
X_1	\bar{X}_2	X_3	$X_1 \bar{X}_2$	$X_1 X_3$	$\bar{X}_2 X_3$	$X_1 \bar{X}_2 X_3$
X_1	X_2	\bar{X}_3	$X_1 X_2$	$X_1 \bar{X}_3$	$X_2 \bar{X}_3$	$X_1 X_2 \bar{X}_3$
X_1	X_2	X_3	$X_1 X_2$	$X_1 X_3$	$X_2 X_3$	$X_1 X_2 X_3$

Отметим серым цветом вычеркнутые строки таблицы. После применения шагов 1–3 таблица примет следующий вид (табл. 3.16).

Таблица 3.16

					$\bar{X}_2 X_3$	
			$X_1 \bar{X}_2$	$X_1 \bar{X}_3$		
			$X_1 \bar{X}_2$		$\bar{X}_2 X_3$	
				$X_1 \bar{X}_3$		

После применения шагов 4 и 5 получим минимальную ДНФ, равносильную исходной формуле $F: X_1 \bar{X}_3 \vee \bar{X}_2 X_3$.

Задача 3.4. Построить минимальную ДНФ для формулы, заданной в виде СДНФ:

$$F(X_1, X_2, X_3) = \bar{X}_1 \bar{X}_2 \bar{X}_3 \vee \bar{X}_1 X_2 \bar{X}_3 \vee \bar{X}_1 X_2 X_3 \vee X_1 X_2 \bar{X}_3 \vee X_1 X_2 X_3$$

Строим для данной формулы минимизирующую карту и работаем с ней по алгоритму 3.6. Отметим справа от последнего столбца те конъюнкции, которые входят в СДНФ данной формулы. Вычеркнем неотмеченные

строки, затем вычеркнем в остальных строках (действуя по столбцам) те элементы, которые попали в вычеркнутые строки.

После применения шагов 1–3 алгоритма карта примет следующий вид (табл. 3.17).

Таблица 3.17

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

Таким образом, получаем $F(X_1, X_2, X_3) = X_2 \vee \bar{X}_1\bar{X}_3$.

Задача 3.5. Построить минимальную ДНФ для формулы, заданной в виде СДНФ:

$$F(X_1, X_2, X_3) = \bar{X}_1\bar{X}_2\bar{X}_3 \vee \bar{X}_1X_2\bar{X}_3 \vee X_1\bar{X}_2\bar{X}_3 \vee X_1X_2X_3$$

Как и в примере 3.4, строим для данной формулы минимизирующую карту и применяем алгоритм 3.6. Карта примет следующий вид (табл. 3.18).

Таблица 3.18

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

Таким образом, $F(X_1, X_2, X_3) = \overline{X}_2\overline{X}_3 \vee \overline{X}_1\overline{X}_3 \vee X_1X_2X_3$.

3.1.8. Метод Куайна – Мак-Класки

Рассмотрим еще один метод построения минимальной ДНФ с помощью сокращения множества перебираемых элементарных конъюнкций. Метод основан на очевидных тождествах:

$$XY \vee X\overline{Y} = X, \quad (X \vee Y)(X \vee \overline{Y}) = X.$$

Определение 3.20. Говорят, что элементарная конъюнкция K покрывает элементарную конъюнкцию L (обозначается $K \succ L$), если любой литерал из K входит в L . Например, $X_1X_3 \succ X_1\overline{X}_2X_3$.

Алгоритм 3.7 (Куайна – Мак-Класки).

Шаг 1. Все конъюнкции из исходной СДНФ записываем в двоичном представлении литералов

$$X^{\alpha_1}, X^{\alpha_2} \dots X^{\alpha_n}, \alpha_i \in \{0, 1\} \rightarrow \alpha_1\alpha_2 \dots \alpha_n,$$

(например: $X_1\overline{X}_2X_3\overline{X}_4 \rightarrow 1010$).

Шаг 2. Полученные двоичные вектора разбиваем на непересекающиеся множества (классы) по числу единиц имеющихся в их двоичной записи, (например: 1010 – класс 2).

Шаг 3. Производим склеивание между векторами соседних классов. Склейваем только элементы соседних классов, чтобы исключить сравнения, заведомо не дающие склеивания (в отличие от метода более раннего метода Куайна). Такие склеенные наборы называют *импликантами*. В склененных векторах символом * отмечаем скленные позиции переменных. Склейваем вектора, отличающиеся только в одной позиции. Это отличие может быть только в том, что в некотором разряде одного вектора стоит 0, в этом же разряде второго вектора стоит 1, а все остальные разряды совпадают полностью, с учетом символов *. Например, 01011, 11011 \rightarrow *1011. Импликанты не участвующие ни в одной склейке называются *простыми*.

Шаг 4. Повторяем шаг 3 до тех пор, пока существуют склеиваемые наборы. Отмечаем простые импликанты.

Шаг 5. Строим матрицу для минимизации: столбцы матрицы представляют собой двоичные вектора из исходной СДНФ, строки – простые импликанты. В каждом столбце отмечаем те строки, которые покрывают вектор столбца.

Шаг 6. Выбираем минимальное число строк, покрывающих все столбцы, начиная решение задачи с выбора столбцов, содержащих единственную отметку.

Рассмотрим работу алгоритма 3.7 на следующем примере. Простые импликанты отмечены серым цветом.

Пример 3.30. Пусть

$$\begin{aligned} F(X_1, X_2, X_3, X_4) = & \overline{X}_1 \overline{X}_2 X_3 X_4 \vee \overline{X}_1 X_2 \overline{X}_3 \overline{X}_4 \vee \overline{X}_1 X_2 \overline{X}_3 X_4 \vee \\ & \overline{X}_1 X_2 X_3 X_4 \vee X_1 \overline{X}_2 \overline{X}_3 X_4 \vee X_1 X_2 \overline{X}_3 X_4 \vee \\ & X_1 X_2 X_3 \overline{X}_4 \vee X_1 X_2 X_3 X_4. \end{aligned}$$

После шага 1 получаем множество двоичных векторов:

$$0011, 0100, 0101, 0111, 1001, 1101, 1110, 1111.$$

Разбиваем полученные двоичные наборы на четыре класса:

$$\underbrace{(0100)}_1 \quad \underbrace{(0011, 0101, 1001)}_2 \quad \underbrace{(0111, 1101, 1110)}_3 \quad \underbrace{(1111)}_4$$

Процесс склеивания векторов представлен на рис. 3.5.

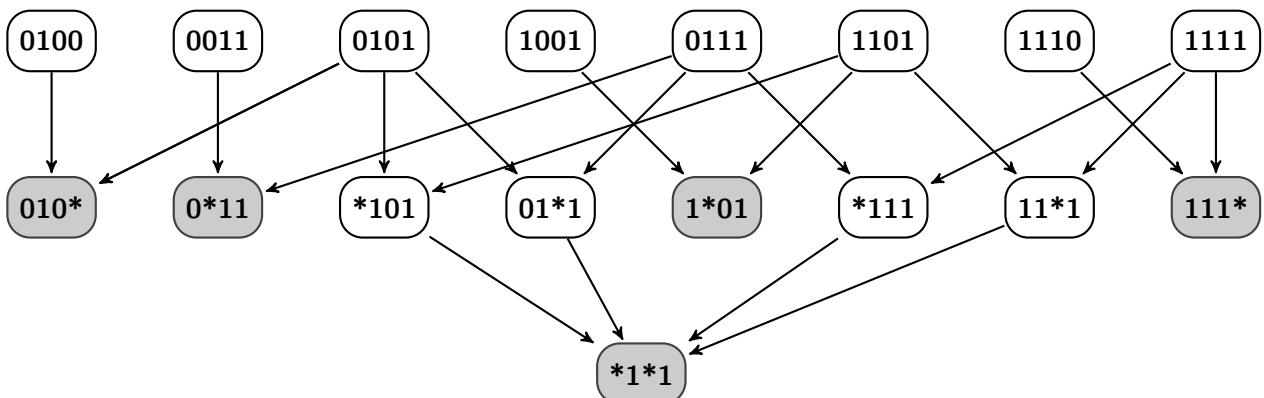


Рис. 3.5

На следующем этапе строим матрицу для минимизации (табл. 3.19).

Необходимо выбрать минимальное число строк, покрывающих все столбцы. Начинаем с выбора столбцов, содержащих единственную отметку. Такими являются столбцы 0011, 0100, 1001, 1110. Имеем:

- 1) 0^*11 , покрываем 0011 и 0111;
- 2) 010^* , покрываем 0011 и 0111;
- 3) 1^*01 , покрываем 1001 и 1101;
- 4) 111^* , покрываем 1110 и 1111.

Таблица 3.19

	0011	0100	0101	0111	1001	1101	1110	1111
010*								
0*11								
1*01								
111*								
*1*1								

Таким образом:

$$F(X_1, X_2, X_3, X_4) = \overline{X}_1 X_3 X_4 \vee \overline{X}_1 X_2 \overline{X}_3 \vee X_1 \overline{X}_3 X_4 \vee X_1 X_2 X_3.$$

Заметим, что в этом примере самая короткая конъюнкция $X_2 X_4$, покрывающая наибольшее число исходных конъюнкций, в решение не вошло.

Замечание 3.16. Время работы метода растет экспоненциально с увеличением входных данных. Можно показать, что для функции от n переменных верхняя граница количества основных импликант $3^n/n$. Если $n = 32$ их может быть больше чем $6.5 * 10^{15}$.

Задача 3.6. Построить минимальную ДНФ для функции

$$\begin{aligned} F(X_1, X_2, X_3, X_4) = & \overline{X}_1 \overline{X}_2 X_3 X_4 \vee \overline{X}_1 X_2 \overline{X}_3 \overline{X}_4 \vee \overline{X}_1 X_2 \overline{X}_3 X_4 \vee \\ & \overline{X}_1 X_2 X_3 X_4 \vee X_1 \overline{X}_2 \overline{X}_3 X_4 \vee X_1 \overline{X}_2 X_3 X_4 \vee \\ & X_1 X_2 \overline{X}_3 \overline{X}_4 \vee X_1 X_2 \overline{X}_3 X_4. \end{aligned}$$

Получаем множество двоичных векторов:

$$0011, 0100, 0101, 0111, 1001, 1011, 1100, 1101.$$

Проводим склеивание векторов (рис. 3.6).

Матрица для минимизации имеет вид (табл. 3.20).

Начинаем с выбора столбцов, содержащих минимальное число отметок (в данном случае одну). Это столбец 0100. Имеем:

- 1) *10*, покрываем 0100, 0101, 1100 и 1101;
- 2) 0*11, покрываем 0011 и 0111;
- 3) 10*1, покрываем 1001 и 1011.

Таким образом:

$$F(X_1, X_2, X_3, X_4) = X_2 \overline{X}_3 \vee \overline{X}_1 X_3 X_4 \vee X_1 \overline{X}_2 X_4.$$

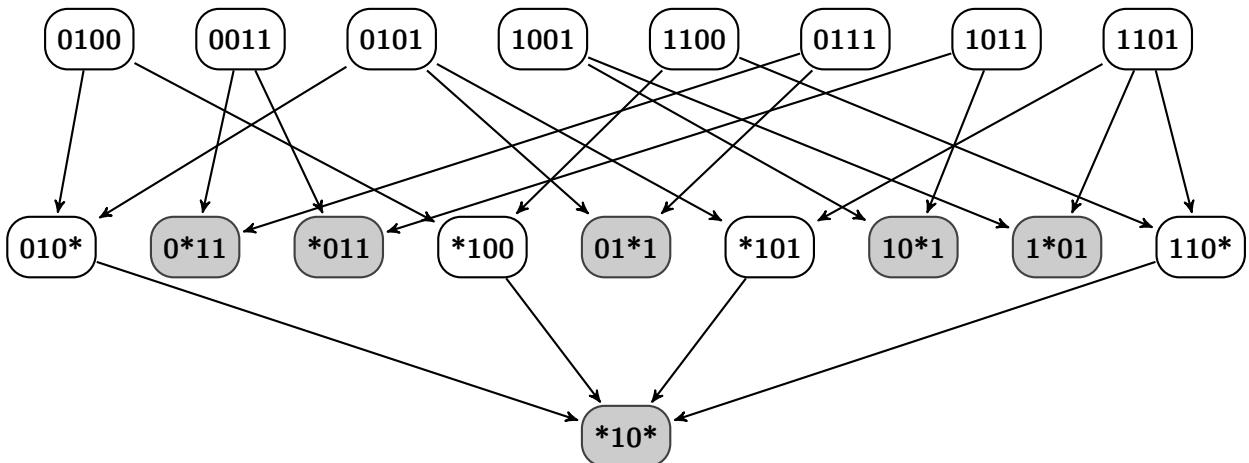


Рис. 3.6

Таблица 3.20

	0100	0011	0101	1001	1100	0111	1011	1101
0*11								
*011								
01*1								
10*1								
1*01								
10								

Задача 3.7. Построить минимальную ДНФ для функции

$$\begin{aligned}
 F(X_1, X_2, X_3, X_4) = & \overline{X}_1 \overline{X}_2 X_3 \overline{X}_4 \vee \overline{X}_1 X_2 \overline{X}_3 X_4 \vee \overline{X}_1 X_2 X_3 \overline{X}_4 \vee \\
 & \overline{X}_1 X_2 X_3 X_4 \vee X_1 \overline{X}_2 X_3 \overline{X}_4 \vee X_1 X_2 \overline{X}_3 \overline{X}_4 \vee \\
 & X_1 X_2 \overline{X}_3 X_4 \vee X_1 X_2 X_3 \overline{X}_4.
 \end{aligned}$$

Множество двоичных векторов имеет вид:

$$0010, 0101, 0110, 0111, 1010, 1100, 1101, 1110.$$

Далее определяем простые импликанты (рис. 3.7).

Составляем матрицу для минимизации (табл. 3.21).

Начинаем с выбора столбцов, содержащих минимальное число отметок (в данном случае одну). Это столбец 0010. Имеем:

- 1) **10, покрываем 0010, 0110, 1010 и 1110;
- 2) 01*1, покрываем 0101 и 0111;
- 3) 110*, покрываем 1100 и 1101.

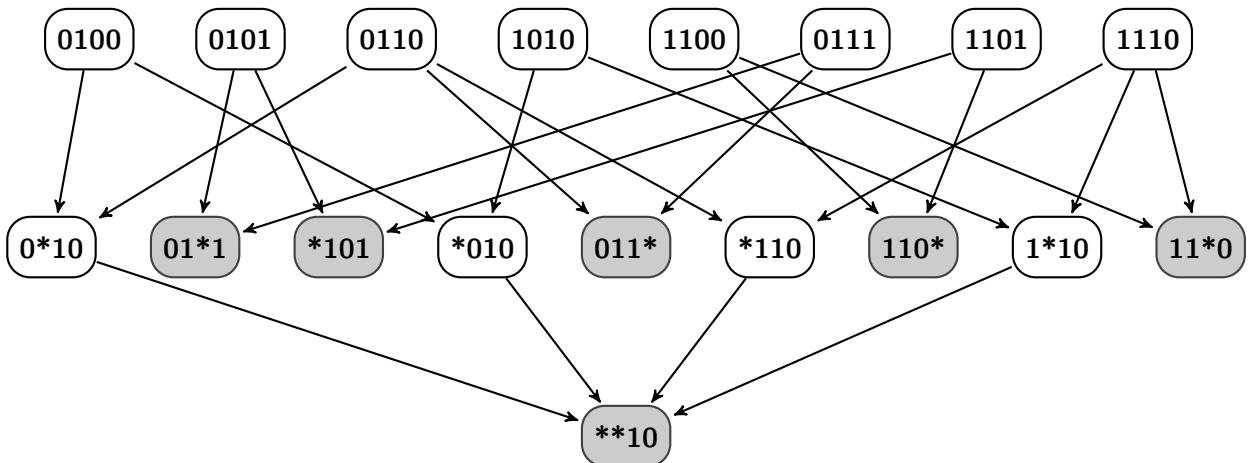


Рис. 3.7

Таблица 3.21

	0010	0101	0110	1010	1100	0111	1101	1110
01*1								
*101								
011*								
110*								
11*0								
**10								

Таким образом:

$$F(X_1, X_2, X_3, X_4) = X_3\bar{X}_4 \vee \bar{X}_1X_2X_4 \vee X_1X_2\bar{X}_3.$$

Задача 3.8. Используя методы минимизирующих карт и Куайна – Мак-Класси для схемы на рис. 3.8 постройте эквивалентную ей более простую цепь.

Замечание 3.17. По вышеизложенным причинам ряд известных методов минимизации, таких как методы Блейка, карт Карнапа, остались вне рамок данной книги. Более подробно познакомиться с этой темой можно, например, в [51].

3.1.9. Метод резолюций в логике высказываний

Рассмотрим еще один метод доказательства того, что формула G является логическим следствием формул F_1, F_2, \dots, F_k , который называют *методом резолюций*. Особенность метода состоит в том, что он оперирует не с произвольными формулами, а с дизъюнктами.

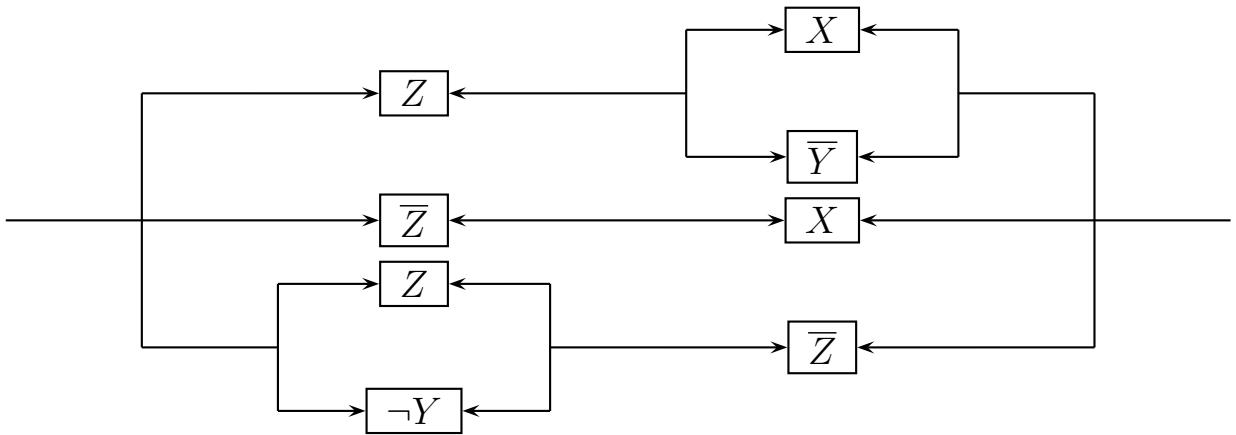


Рис. 3.8

Замечание 3.18. Условимся не различать дизъюнкты, которые получаются один из другого с помощью равносильностей, установленных в теореме 3.4: коммутативности и ассоциативности дизъюнкции (законы 4 и 6), а также идемпотентности (закон 2). Последнее означает, что дизъюнкты $X \vee \bar{Y} \vee X \vee Z$ и $Z \vee X \vee \bar{Y}$ равны.

Введем в рассмотрение *пустой* дизъюнкт, т. е. дизъюнкт, не содержащий литералов. В методе резолюций его обычно обозначают \square . Считаем, что пустой дизъюнкт ложен при любой интерпретации, т. е. равен константе 0. Например, формула $F \wedge \square$ равносильна \square , а формула $F \vee \square$ равносильна F .

Определение 3.21. Литералы L и \bar{L} называют *противоположными*, или *контрарными*.

Определение 3.22. Правилом *резолюций* в логике высказываний называют следующее правило: из дизъюнктов $X \vee F$ и $\bar{X} \vee G$ выводим дизъюнкт $F \vee G$.

Дизъюнкт $F \vee G$ называют *резольвентой* дизъюнктов $X \vee F$ и $\bar{X} \vee G$.

Пример 3.31. Из дизъюнктов $\bar{X} \vee Y \vee Z$ и $X \vee \bar{Y}$ выводим дизъюнкт $Y \vee Z \vee \bar{Y}$.

Замечание 3.19. Заметим, что в первых двух дизъюнктах последнего примера есть еще одна пара противоположных литералов.

Применять правило резолюций не обязательно к самым левым литералам (поскольку не различаются дизъюнкты, отличающиеся порядком записи литералов). Тогда правило резолюций, примененное к литералам Y и \bar{Y} этих дизъюнктов, даст $\bar{X} \vee Z \vee X$.

В дизъюнктах не будем писать повторяющиеся литералы и пустой дизъюнкт \square , если есть другие литералы.

Лемма 3.3. Резольвента есть логическое следствие породивших ее дизъюнктов.

Доказательство. Рассмотрим дизъюнкты $X \vee F$ и $\bar{X} \vee G$. Составим совместную таблицу истинности дизъюнктов и резольвенты (табл. 3.22).

Таблица 3.22

X	F	G	$X \vee F$	$\bar{X} \vee G$	$F \vee G$	X	F	G	$X \vee F$	$\bar{X} \vee G$	$F \vee G$
0	0	0	0	1	1	1	0	0	1	0	1
0	0	1	0	1	1	1	0	1	1	1	1
0	1	0	1	1	1	1	1	0	1	0	1
0	1	1	1	1	1	1	1	1	1	1	1

Согласно определению 3.9 формула $F \vee G$ есть логическое следствие формул $X \vee F$ и $\bar{X} \vee G$ ■

Определение 3.23. Пусть S — множество дизъюнктов. *Резолютивным выводом*, или просто *выводом* из S , называют последовательность дизъюнктов D_1, D_2, \dots, D_n такая, что каждый дизъюнкт этой последовательности принадлежит S или следует из предыдущих по правилу резолюций.

Дизъюнкт D выводим из S , если существует вывод из S , последним дизъюнктом которого является D .

Вывод пустого дизъюнкта из S называют *опровержением* S .

Пример 3.32. Пусть $S = \{\bar{X} \vee Y, Z \vee X \vee \bar{Y}, Y\}$, тогда последовательность

$$D_1 = Z \vee X \vee \bar{Y}, D_2 = Y, D_3 = Z \vee X, D_4 = \bar{X} \vee Y, D_5 = Y \vee Z$$

является выводом из S . Дизъюнкт $Y \vee Z$ выводим из S .

Определение 3.24. Дизъюнкт \hat{D} зависит от дизъюнкта D , если при выводе \hat{D} использовалась резольвента D .

Применение метода резолюций основано на следующем утверждении.

Теорема 3.12 (о полноте метода резолюций). Множество дизъюнктов логики высказываний S невыполнимо тогда и только тогда, когда из S выводим пустой дизъюнкт.

Доказательство. Пусть из S выводим пустой дизъюнкт. Предположим противное: множество S выполнимо, т. е. существует интерпретация, при которой все дизъюнкты из S истинны. Рассмотрим вывод пустого дизъюнкта $D_1, \dots, D_n = \square$. Если дизъюнкт $D_j \in S$, то по предположению, на этой интерпретации $D_j = 1$. Если же он получается из предыдущих по правилу резолюций, то, согласно лемме 3.3, также $D_j = 1$. При $i = n \square = 1$. Противоречие доказывает достаточность.

Докажем необходимость. Доказательство проведем индукцией по следующему параметру: $d(S)$ есть сумма числа вхождений литералов в дизъюнкты из S минус число дизъюнктов.

Пусть множество дизъюнктов S невыполнимо. Если пустой дизъюнкт принадлежит S , то он очевидно выводим из S . Будем считать, что $\square \notin S$. Тогда каждый дизъюнкт из S содержит хотя бы один литерал, следовательно $d \geq 0$.

База индукции: $d(S) = 0$. Тогда все дизъюнкты состоят из одного литерала. Поскольку множество S невыполнимо, то в нем должна найтись пара контрапротивных литералов, например, X и \bar{X} . В этом случае очевидно, что пустой дизъюнкт выводим из S .

Шаг индукции: $d(S) > 0$. Предположим, что для любого множества дизъюнктов \tilde{S} , такого, что $d(\tilde{S}) < d(S)$ необходимость теоремы доказана. Пусть $S = \{D_1, \dots, D_k\}$. Так как $d(S) > 0$, в S существует хотя бы один дизъюнкт, содержащий более одного литерала. Не умаляя общности, считаем, что это дизъюнкт $D_k = L_k \vee F_k$, где L_k — литерал и $F_k \neq \square$.

Рассмотрим два множества дизъюнктов

$$S_1 = S \setminus \{D_k\} \cup \{L_k\}, \quad S_2 = S \setminus \{D_k\} \cup \{F_k\}.$$

Из невыполнимости S очевидно, что S_1 и S_2 также невыполнимы и при этом $d(S_1) < d(S)$ и $d(S_2) < d(S)$. Тогда по индукционному предположению из S_1 и S_2 выводим пустой дизъюнкт. Запишем выводы пустого дизъюнкта из S_1 и S_2 соответственно

$$A_1, \dots, A_i, \dots, A_l = \square, \tag{3.12}$$

$$B_1, \dots, B_j, \dots, B_n = \square. \tag{3.13}$$

Если в (3.12) не содержится дизъюнкта L_k , или в (3.13) нет дизъюнкта F_k , такая последовательность дизъюнктов будет выводом из S и необходимость теоремы доказана. Считаем, что в (3.12) содержится L_k , например $L_k = A_i$, а в (3.13) — $F_k = B_j$.

Преобразуем (3.13) следующим образом: к дизъюнкту B_j и всем зависимым от него дизъюнктам, добавим литерал L_k . Получим следующую последовательность дизъюнктов

$$B_1, \dots, B_{j-1}, \hat{B}_j = F_k \vee L_k, \hat{B}_{j+1}, \dots, \hat{B}_m \tag{3.14}$$

По определению (3.14) является выводом из S . Если дизъюнкт B_m не зависит от B_j , то очевидно, что $\hat{B}_m = \square$. В этом случае необходимость теоремы доказана. Предположим, что B_m зависит от B_j . Тогда $\hat{B}_m = L_k$.

В вывод (3.12) вместо дизъюнкта $A_i = L_k$ подставим последовательность (3.14). Получим последовательность

$$A_1, \dots, A_{i-1}, \underbrace{B_1, \dots, B_{j-1}, \hat{B}_j, \dots, \hat{B}_m}_{A_i} = L_k, A_{i+1}, \dots, A_l = \square. \tag{3.15}$$

Последовательность (3.15) есть вывод пустого дизъюнкта из S . Теорема доказана. \blacksquare

Используя теорему 3.12, построим алгоритм для определения того, является ли формула G логическим следствием множества формул F_1, \dots, F_k .

Алгоритм 3.8 (метод резолюций).

Шаг 1. Составляем множество формул $T = \{F_1, \dots, F_k, \overline{G}\}$.

Шаг 2. Каждую из формул множества T приводим к КНФ и в полученных формулах зачеркиваем знаки конъюнкции. Получаем множество дизъюнктов S .

Шаг 3. Строим опровержение S .

Если на *шаге 3* построено опровержение S , то формула G есть логическое следствие формул F_1, \dots, F_k . Иначе G не является логическим следствием F_1, \dots, F_k .

Обоснованием алгоритма 3.8 служит следующая теорема.

Теорема 3.13. Множество формул T и множество дизъюнктов S , полученное на шаге 2 алгоритма 3.8, одновременно выполнимы (или невыполнимы).

Доказательство. Пусть формула G_i есть КНФ формулы $F_i \in T$. Тогда $G_i = D_{i_1} \cdots D_{i_n}$. В силу определения конъюнкции, очевидно, что выполнимость G_i равносильна выполнимости множества дизъюнктов $\{D_{i_1}, \dots, D_{i_n}\}$. \blacksquare

Пример 3.33. Применим алгоритм 3.8 для решения задачи о хищении на складе из примера 3.14. Множество формул T имеет вид:

$$\{F_1 = X \vee Y \vee Z, F_2 = X \rightarrow Y \vee Z, F_3 = \overline{Z}, \overline{G} = \overline{Y}\}.$$

После выполнения шага 2 получаем множество дизъюнктов

$$S = \{D_1 = X \vee Y \vee Z, D_2 = \overline{X} \vee Y \vee Z, D_3 = \overline{Z}, D_4 = \overline{Y}\}.$$

Строим опровержение S .

$$D_1, D_2, D_5 \stackrel{D_2, D_1}{=} Y \vee Z, D_3, D_6 \stackrel{D_5, D_3}{=} Y, D_4, D_7 \stackrel{D_6, D_4}{=} \square.$$

Следовательно, формула G — логическое следствие, т. е. Джонсон виновен.

Пример 3.34. Петров, Иванов и Сидоров сдавали экзамен по «математической логике и теории алгоритмов». Если Петров не сдал экзамен на «отлично», то и Иванов не сдал на «отлично». Сидоров и еще один из друзей сдали экзамен на «отлично». Следует ли отсюда, что не верно, что Петров сдал экзамен не на «отлично», а Иванов на «отлично»?

Для ответа на вопрос применим метод резолюций. Запишем данные в виде формул логики высказываний. Введем обозначения:

1. X – Петров сдал экзамен на «отлично».
2. Y – Иванов сдал экзамен на «отлично».
3. Z – Сидоров сдал экзамен на «отлично».

Получаем следующие формулы:

$$F_1(X, Y) = \overline{X} \rightarrow \overline{Y}, \quad F_2(X, Y, Z) = Z \wedge (X \vee Y), \quad G(X, Y) = \overline{\overline{X} \wedge Y}.$$

Приводим формулы к КНФ и составляем множество дизъюнктов S :

$$S = \{D_1 = X \vee \overline{Y}, \quad D_2 = Z, \quad D_3 = X \vee Y, \quad D_4 = \overline{X}, \quad D_5 = Y\}$$

Строим опровержение S .

$$D_1, \quad D_4, \quad D_6 \stackrel{D_1, D_4}{=} \overline{Y}, \quad D_5, \quad D_7 \stackrel{D_5, D_6}{=} \square.$$

Пример 3.35. Доказать логичность следующих рассуждений используя метод резолюций.

1. Если команда Зенит выигрывает, то город Санкт-Петербург торжествует: $A \rightarrow A_1 = \overline{A} \vee A_1$.
2. Если команда Спартак выигрывает, то торжествует город Москва: $B \rightarrow B_1 = \overline{B} \vee B_1$.
3. Выигрывает либо команда Зенит, либо команда Спартак: $A \vee B$.
4. Если команда Зенит выигрывает, то город Москва не торжествует: $A \rightarrow \overline{B_1} = \overline{A} \vee \overline{B_1}$.
5. Если выигрывает команда Спартак, то город Санкт-Петербург не торжествует: $B \rightarrow \overline{A_1} = \overline{B} \vee \overline{A_1}$.
6. Таким образом, город Санкт-Петербург торжествует тогда и только тогда, когда не торжествует город Москва: $A_1 \overline{B_1} \vee \overline{A_1} B_1$

$$\overline{A_1 \overline{B_1}} \vee \overline{\overline{A_1} B_1} = \overline{A_1 \overline{B_1}} \wedge \overline{\overline{A_1} B_1} = (\overline{A_1} \vee B_1) (\overline{A_1} \vee \overline{B_1}).$$

Приводим формулы к КНФ и составляем множество дизъюнктов S :

$$S = \{D_1 = \overline{A} \vee A_1, D_2 = \overline{B} \vee B_1, D_3 = A \vee B, D_4 = \overline{A} \vee \overline{B_1}, \\ D_5 = \overline{B} \vee \overline{A_1}, D_6 = A_1 \vee \overline{B_1}, D_7 = \overline{A_1} \vee B_1.\}$$

Строим опровержение S .

$$D_1, D_3, D_8 \stackrel{D_1, D_3}{=} A_1 \vee B, D_9 \stackrel{D_2, D_6}{=} \overline{B} \vee A_1, D_{10} \stackrel{D_8, D_9}{=} A_1, \\ D_{11} \stackrel{D_3, D_5}{=} A \vee \overline{A_1}, D_{12} \stackrel{D_4, D_7}{=} \overline{A} \vee \overline{A_1}, D_{13} \stackrel{D_{11}, D_{12}}{=} \overline{A_1}, D_{14} \stackrel{D_{10}, D_{13}}{=} \square$$

Задача 3.9. Доказать логичность следующих рассуждений используя метод резолюций.

1. Если работа выполнена, то начальство отпустит на рыбалку ($A \rightarrow B$).
2. Если отпустят на рыбалку, то обязательно возьмут на нее и сына ($B \rightarrow C$).
3. Если берут сына, значит надо брать лодку ($C \rightarrow D$).
4. Если брать с собой лодку, то поедут все вместе ($D \rightarrow E$).
5. Таким образом, если работа выполнена, то все вместе едут на рыбалку ($A \rightarrow E$).

Введем следующие обозначения:

- a) A – работа выполнена;
- b) B – отпуск на рыбалку;
- c) C – взять на рыбалку сына;
- d) D – рыбалку провести с лодкой;
- e) E – на рыбалку поедут все вместе.

Имеем множество формул:

$$\{F_1 = A \rightarrow B, F_2 = B \rightarrow C, F_3 = C \rightarrow D, F_4 = D \rightarrow E, G = A \rightarrow E\}$$

Приводим их к КНФ и составляем множество дизъюнктов S :

$$S = \{D_1 = \overline{A} \vee B, D_2 = \overline{B} \vee C, D_3 = \overline{C} \vee D, D_4 = \overline{D} \vee E, D_5 = A, \\ D_6 = \overline{E}\}$$

Строим опровержение S .

$$D_1, D_5, D_7 \stackrel{D_1, D_5}{=} B, D_8 \stackrel{D_2, D_7}{=} C, D_9 \stackrel{D_3, D_8}{=} D, D_{10} \stackrel{D_4, D_9}{=} E, D_{11} \stackrel{D_6, D_{10}}{=} \square$$

Задача 3.10. Доказать логичность рассуждений используя метод резолюций.

1. Я могу надеть на себя брезентовые штаны, или шерстяное платье, или пиджак и юбку с разрезом ($A \vee B \vee C$).
2. Я буду выглядеть великолепно, если надену пиджак и юбку с разрезом и при этом возьму с собой сумку ($CD \rightarrow E$).
3. И наоборот, я буду выглядеть ужасно, если надену на себя брезентовые штаны или шерстяное платье ($A \vee B \rightarrow \overline{E}$).
4. Однако сумку надо брать обязательно, если надеть пиджак и юбку с разрезом ($C \rightarrow D$).
5. Итак, чтобы выглядеть великолепно, я выбираю последнее, т.е. надену на себя пиджак и юбку с разрезом ($C \rightarrow E$).

Введем обозначения:

- a) A — надеть брезентовые штаны;
- b) B — надеть шерстяное платье;
- c) C — надеть пиджак и юбку с разрезом;
- d) D — взять с собой сумку;
- e) E — великолепно смотрится.

Имеем множество формул:

$$\{F_1 = A \vee B \vee C, F_2 = CD \rightarrow E, F_3 = A \vee B \rightarrow \overline{E}, F_4 = C \rightarrow D, G = C \rightarrow E\}$$

Приводим их к КНФ и составляем множество дизъюнктов S :

$$S = \{D_1 = A \vee B \vee C, D_2 = \overline{C} \vee \overline{D} \vee E, D_3 = \overline{A} \vee \overline{E}, D_4 = \overline{B} \vee \overline{E}, D_5 = \overline{C} \vee D, D_6 = C, D_7 = \overline{E}\}$$

Легко построить опровержение S .

Задача 3.11. Методом резолюций решите задачу о сокровище на острове рыцарей и лжецов (см. пример 3.16).

3.1.10. Стратегии метода резолюций

В множестве дизъюнктов часто существует не одна пара дизъюнктов, к которым можно применить правило резолюций. Способ выбора дизъюнктов и литералов в них, к которым применяется правило резолюций для получения резольвенты, называют *стратегией метода*. Рассмотрим две стратегии: стратегию насыщения уровней и стратегию вычеркивания.

Далее считаем, что множество дизъюнктов S упорядочено.

Алгоритм 3.9 (стратегия насыщения уровней). Наиболее простой способ выбора дизъюнктов из S для получения резольвенты состоит в организации полного перебора возможных вариантов.

Шаг 1. Положим $S_0 = S$. Пусть D_2 пробегает по порядку множество дизъюнктов S_0 , начиная со второго. В качестве D_1 рассматриваются последовательно дизъюнкты из S_0 , предшествующие D_2 начиная с первого, и формируем последовательность S_1 , состоящую из всевозможных резольвент дизъюнктов D_1 и D_2 .

Шаг n. Пусть получены последовательности S_0, S_1, \dots, S_{n-1} . Строим S_n следующим образом. В качестве D_2 берутся по порядку дизъюнкты из S_{n-1} , а в качестве D_1 — дизъюнкты из $S_0 \cup S_1 \cup \dots \cup S_{n-1}$, предшествующие D_2 . Последовательность S_n будет состоять из всевозможных резольвент дизъюнктов D_1 и D_2 .

Процесс порождения резольвент прекращается, как только получается пустой дизъюнкт.

Замечание 3.20. Уровнями называют последовательности дизъюнктов $S_0, S_1, S_2, \dots, S_n, \dots$

Пример 3.36. Построим опровержение следующего множества дизъюнктов

$$S = \{D_1 = X \vee Y, D_2 = \overline{X} \vee \overline{Y}, D_3 = X \vee Z, D_4 = \overline{X} \vee Z, D_5 = \overline{Z}\}$$

используя стратегию насыщения уровней.

Шаг 1. Строим уровень S_1 :

$$\begin{aligned} S_1 = \{ & D_6 \stackrel{D_2, D_1}{=} Y \vee \overline{Y}, D_7 \stackrel{D_2, D_1}{=} X \vee \overline{X}, D_8 \stackrel{D_3, D_2}{=} \overline{Y} \vee Z, D_9 \stackrel{D_4, D_1}{=} Y \vee Z, \\ & D_{10} \stackrel{D_4, D_3}{=} Z, D_{11} \stackrel{D_5, D_3}{=} X, D_{12} \stackrel{D_5, D_4}{=} \overline{X}\}. \end{aligned}$$

Шаг 2. Строим уровень S_2 :

$$\begin{aligned} S_2 = \{ & D_{13} \stackrel{D_6, D_1}{=} X \vee Y, D_{14} \stackrel{D_6, D_2}{=} \overline{X} \vee \overline{Y}, D_{15} \stackrel{D_7, D_1}{=} X \vee Y, D_{16} \stackrel{D_7, D_2}{=} \overline{X} \vee \overline{Y}, \\ & D_{17} \stackrel{D_7, D_3}{=} X \vee Z, D_{18} \stackrel{D_7, D_4}{=} \overline{X} \vee Z, D_{19} \stackrel{D_8, D_1}{=} X \vee Z, D_{20} \stackrel{D_8, D_5}{=} \overline{Y}, \\ & D_{21} \stackrel{D_8, D_6}{=} \overline{Y} \vee Z, D_{22} \stackrel{D_9, D_2}{=} \overline{X} \vee Z, D_{23} \stackrel{D_9, D_6}{=} Y \vee Z, D_{24} \stackrel{D_9, D_8}{=} Z, \\ & D_{25} \stackrel{D_{10}, D_5}{=} \square\}. \end{aligned}$$

Замечание 3.21. Из примера 3.36 видно, что стратегия насыщения уровней порождает много лишних дизъюнктов. Например, дизъюнкты D_6 и D_7 — тождественно истинны. Удаление или добавление тождественно истинного дизъюнкта не влияет на выполнимость S , поэтому такие дизъюнкты должны быть удалены из вывода.

Некоторые дизъюнкты порождаются в выводе неоднократно, например,

$$D_{13} = D_{15} = X \vee Y, D_8 = D_{20} = \overline{Y} \vee Z.$$

Оптимизируем алгоритм 3.9 с учетом сделанных замечаний.

Определение 3.25. Дизъюнкт D называют *расширением* дизъюнкта F , если $D = F \vee G$ (с учетом возможной перестановки литералов). Например, $X \vee Y \vee Z$ есть расширение $X \vee Z$.

Алгоритм 3.10 (стратегия вычеркивания). Данная стратегия является модификацией алгоритма 3.9.

Оптимизация заключается в следующем: после того, как получена очередная резольвента D дизъюнктов D_1 и D_2 проверяется, является ли она тавтологией или расширением некоторого дизъюнкта $F \in S_0 \cup \dots \cup S_{n-1}$, и в случае положительного ответа D вычеркивается, т. е. не включается в последовательность S_n .

Пример 3.37. Применим стратегию вычеркивания ко множеству дизъюнктов S из примера 3.36.

Шаг 1. Строим уровень S_1 :

$$S_1 = \{D_6 \stackrel{D_3, D_2}{=} \overline{Y} \vee Z, D_7 \stackrel{D_4, D_1}{=} Y \vee Z, D_8 \stackrel{D_4, D_3}{=} Z, D_9 \stackrel{D_5, D_3}{=} X, D_{10} \stackrel{D_5, D_4}{=} \overline{X}\}.$$

Шаг 2. Строим уровень S_2 :

$$S_2 = \{D_{11} \stackrel{D_6, D_5}{=} \overline{Y}, D_{12} \stackrel{D_7, D_5}{=} Y, D_{13} \stackrel{D_8, D_5}{=} \square\}.$$

Задача 3.12. Рассмотрим следующее множество дизъюнктов:

$$S = \{D_1 = \overline{A} \vee C, D_2 = \overline{B} \vee C, D_3 = \overline{A} \vee B \vee C, D_4 = \overline{C} \vee D, \\ D_5 = A \vee D, D_6 = \overline{D}\}.$$

Построить опровержение S используя стратегию вычеркивания (алгоритм 3.10).

Шаг 1. Строим уровень S_1 :

$$S_1 = \{D_7 \stackrel{D_1, D_4}{=} \overline{A} \vee D, D_8 \stackrel{D_1, D_5}{=} C \vee D, D_9 \stackrel{D_2, D_4}{=} \overline{B} \vee D, \\ D_{10} \stackrel{D_3, D_4}{=} \overline{A} \vee B \vee D, D_{11} \stackrel{D_4, D_6}{=} \overline{C}, \\ D_{12} \stackrel{D_5, D_6}{=} A\}.$$

Шаг 2. Строим уровень S_2 :

$$S_2 = \{D_{13} \stackrel{D_1, D_{12}}{=} \overline{A}, D_{14} \stackrel{D_1, D_{13}}{=} C, D_{15} \stackrel{D_2, D_{12}}{=} \overline{B}, D_{16} \stackrel{D_4, D_8}{=} D, \\ D_{17} \stackrel{D_6, D_{17}}{=} \square\}.$$

Задача 3.13. Построить опровержение следующего множества дизъюнктов S , используя стратегию вычеркивания (алгоритм 3.10).

$$S = \{D_1 = A \vee B \vee C, D_2 = \overline{A} \vee B \vee C, D_3 = B \vee \overline{C}, D_4 = \overline{B} \vee C, \\ D_5 = \overline{C} \vee A, D_6 = \overline{C}\}.$$

Теорема 3.14. Если множество дизъюнктов S невыполнимо, то используя алгоритмы 3.9 и 3.10 можно построить опровержение S .

Доказательство. Для стратегии насыщения уровней это очевидно. Рассмотрим стратегию вычеркивания. Заметим, что если D и F — дизъюнкты из S и D есть расширение F , то множество S невыполнимо в том и только в том случае, когда невыполнимо множество $S \setminus \{D\}$. ■

3.1.11. Линейная резолюция

Линейная резолюция соответствует следующей схеме: выбирается некоторый дизъюнкт из исходного множества дизъюнктов, к нему и к другому дизъюнкту применяется правило резолюции. К полученной резольвенте и какому-либо другому дизъюнкту опять применяем правило резолюции, получаем следующий дизъюнкт и т. д. Процесс повторяется до тех пор, пока не будет получен пустой дизъюнкт, или для текущего дизъюнкта будет невозможно применить правило резолюции ни с одним другим дизъюнктом.

В последнем случае осуществляется откат к тому дизъюнкту, где был возможен множественный выбор дизъюнкта для резольвирования.

Определение 3.26. Рассмотрим множество дизъюнктов S . Возьмем дизъюнкт $C_0 \in S$. *Линейный вывод* дизъюнкта C из множества дизъюнктов S это вывод, в котором выполняются условия:

1. Для $i = 0, \dots, k - 1$ дизъюнкт C_{i+1} есть резольвента дизъюнкта C_i (называемого *центральным дизъюнктом*) и B_t (называемого *боковым*).
2. Каждый B_t либо принадлежит S , либо есть C_j для некоторого $j < i$.

Общая схема линейной резолюции изображена на рис. 3.9.

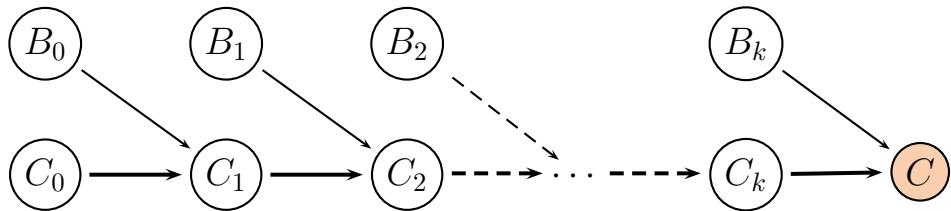


Рис. 3.9

Пример 3.38. Формируется экипаж для кругосветного путешествия. Браун, Джонсон и Смит проходят тестирование. Окончательное слово остается за командиром экипажа. После беседы с ними командир высказал свое мнение:

- 1) я возьму в путешествие одного из трех кандидатов;
- 2) если я возьму Брауна, но не возьму Джонсона, то возьму также и Смита;
- 3) членами экипажа станут либо Смит и Джонс, либо ни один из них;
- 4) если в экипаж будет принят Смит, то также будет принят и Браун.

Считая, что высказывания командира экипажа истинны, нужно определить: отправится ли в кругосветное путешествие Смит?

Введем следующие обозначения:

- 1) B – Браун отправится в путешествие;
- 2) J – Джонс отправится в путешествие;
- 3) S – Смит отправится в путешествие.

Запишем в виде формулам высказывания командира экипажа и вопрос о Смите (обозначим его через G):

$$F_1 = B \vee J \vee S; F_2 = B \wedge \overline{J} \rightarrow S; F_3 = J \wedge S \vee \overline{J} \wedge \overline{S}; F_4 = S \rightarrow B; G = S.$$

Приводим формулы к КНФ и строим множество дизъюнктов:

$$\begin{aligned} D_1 &= B \vee J \vee S, \quad D_2 = \overline{B} \vee J \vee S, \quad D_3 = \overline{J} \vee S, \quad D_4 = J \vee \overline{S}, \quad D_5 = \overline{S} \vee B, \\ D_6 &= \overline{S}. \end{aligned}$$

Применяем линейную резолюцию, начиная с D_6 (рис. 3.10).

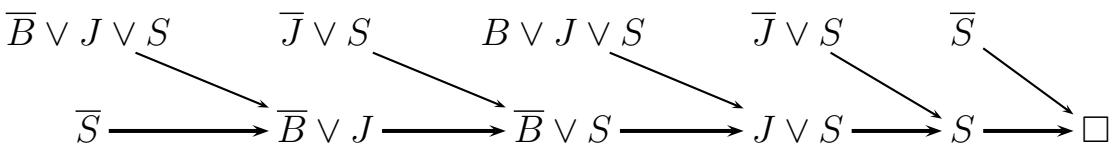


Рис. 3.10

3.1.12. Решетки, булевы алгебры, кольца

Вернемся еще раз к логическим законам, установленным в теореме 3.4. Доказанные для формул логики высказываний, эти равносильности позволяют определить новые алгебраические структуры, в дополнение к уже рассмотренным в главе 1 понятиям группы, кольца и поля (см. подразд. 1.3.1) При этом множество формул логики высказываний будет одним из примеров реализации новых понятий.

Дадим соответствующие определения. Как и прежде, запись $(\mathfrak{M}, \vee, \wedge)$ будет означать, что на непустом множестве \mathfrak{M} заданы операции \vee и \wedge .

Определение 3.27. Множество $(\mathfrak{M}, \vee, \wedge)$ называют *решеткой*, если для любых элементов $a, b, c \in \mathfrak{M}$ выполнены следующие условия:

- 1) ассоциативность: $(a \vee b) \vee c = a \vee (b \vee c); (a \wedge b) \wedge c = a \wedge (b \wedge c);$
- 2) коммутативность: $a \vee b = b \vee a; a \wedge b = b \wedge a;$
- 3) законы поглощения: $(a \vee b) \wedge a = a; (a \wedge b) \vee a = a.$

Пример 3.39. Рассмотрим примеры решеток.

1) $(\mathbb{N}, D(x, y), M(x, y))$ — множество натуральных чисел с операциями вычисления наибольшего общего делителя и наименьшего общего кратного;

2) $(\mathbb{R}, \max\{x, y\}, \min\{x, y\})$ — множество вещественных чисел с операциями вычисления большего и меньшего числа из двух данных чисел.

Замечание 3.22. Аксиомы решетки двойственны относительно операций \vee и \wedge . Поэтому из любого утверждения, доказанного с помощью аксиом, можно получить *двойственное* предложение, поменяв местами в исходном утверждении знаки \vee и \wedge .

Лемма 3.4. Для любого элемента a решетки \mathfrak{M} выполняются законы *идемпотентности*¹:

$$a \vee a = a, \quad a \wedge a = a. \quad (3.16)$$

Доказательство. Применив второй закон поглощения для элементов a и $a \vee b$ имеем

$$a = (a \wedge (a \vee b)) \vee a = ((a \vee b) \wedge a) \vee a = a \vee a.$$

Согласно замечанию 3.22 заменив в рассуждениях операцию \vee на \wedge получаем второе равенство в (3.16) ■

В произвольной решетке легко ввести отношение нестрогого частично-го порядка (см. определение 2.65 на стр. 341).

Определим это отношение R следующим образом: aRb если $a \wedge b = a$. Справедлива теорема.

¹Иногда их включают в определение решетки.

Теорема 3.15. Введенное отношение R является отношением частичного порядка.

Доказательство. Согласно второму закону идемпотентности (3.16) отношение R рефлексивно. Докажем его антисимметричность. Пусть aRb и bRa . Тогда

$$a \stackrel{aRb}{=} a \wedge b = b \wedge a \stackrel{bRa}{=} b.$$

Установим транзитивность отношения. Пусть aRb и bRc . Покажем, что тогда aRc . Имеем.

$$a \wedge c \stackrel{aRb}{=} (a \wedge b) \wedge c = a \wedge (b \wedge c) \stackrel{bRc}{=} a \wedge b \stackrel{aRb}{=} a. \quad \blacksquare$$

Определение 3.28. Решетку $(\mathfrak{M}, \vee, \wedge)$, в которой операции \vee и \wedge дистрибутивны одна относительно другой, т. е. для любых элементов $a, b, c \in \mathfrak{M}$ выполняются следующие тождества:

$$a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c), \quad a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c)$$

называют *дистрибутивной решеткой*.

Пример 3.40. Рассмотрим множество всех подмножеств произвольного множества M с операциями объединения и пересечения подмножеств. Такое множество называют *булеаном* и обозначают $B(M)$ или 2^M . Очевидно, что $(2^M, \cup, \cap)$ — дистрибутивная решетка.

Определение 3.29. Нейтральные элементы решетки $(\mathfrak{M}, \vee, \wedge)$ по операциям \vee и \wedge называют соответственно *нулем* или *нижней границей* (обозначают 0) и *единицей*, или *верхней границей* (обозначают 1), т. е для любого элемента $a \in \mathfrak{M}$ выполняются тождества

$$0 \vee a = a, \quad 1 \wedge a = a. \quad (3.17)$$

Решетку, для которой существуют нижняя и верхняя грани, называют *ограниченной*.

Лемма 3.5. Если в решетке существуют ноль и единица, то они единственны.

Доказательство. Пусть $0'$ и $0''$ — нули решетки. Тогда используя первое равенство в (3.17) для $0'$ и $0''$ и коммутативность операции \vee , получаем

$$0'' = 0' \vee 0'' = 0'' \vee 0' = 0'.$$

Аналогично доказывается единственность единицы. ■

Лемма 3.6. Для любого элемента решетки $a \in \mathfrak{M}$ выполняются тождества

$$0 \wedge a = 0, \quad 1 \vee a = 1. \quad (3.18)$$

Доказательство. Используя (3.17) и аксиому поглощения получаем

$$0 \wedge a = 0 \wedge (0 \vee a) = 0.$$

Второе тождество в (3.18) доказывается аналогично. ■

Определение 3.30. Элемент $\bar{a} \in \mathfrak{M}$ называют *дополнением* элемента a в ограниченной решетке $(\mathfrak{M}, \vee, \wedge)$ (обозначают $\neg a$), если

$$a \vee \neg a = 1, \quad a \wedge \neg a = 0.$$

Если для каждого элемента ограниченной решетки существует дополнение, то решетку называют *решеткой с дополнением*.

Дистрибутивную ограниченную решетку с дополнением называют *булевой алгеброй* и обозначают $(\mathfrak{M}, \vee, \wedge, \neg)$.

Лемма 3.7. Если дополнение элемента существует, то оно единственное.

Доказательство. Предположим, что для элемента a существуют два дополнения a_1 и a_2 . Имеем

$$\begin{aligned} a_1 &= a_1 \wedge 1 = a_1 \wedge (a \vee a_2) = (a_1 \wedge a) \vee (a_1 \wedge a_2) = (a_2 \wedge a) \vee (a_2 \wedge a_1) = \\ &= a_2 \wedge (a \vee a_1) = a_2 \wedge 1 = a_2. \end{aligned} \quad \blacksquare$$

Пример 3.41. Приведем примеры булевых алгебр.

1. Самая простая булева алгебра содержит два элемента 0 и 1, а действия в ней определяются таблицей истинности связок для \vee и \wedge (см. табл. 3.1). При этом 0 есть дополнение 1, а 1 — дополнение 0.

2. Решетка $(2^M, \cup, \cap)$ из примера 3.40 является булевой алгеброй. При этом 0 есть пустое множество, а 1 — само множество M . Дополнением произвольного элемента $A \subseteq M$ является множество $M \setminus A$. Эта булева алгебра носит название *алгебры Кантора*.

3. Множество всех формул исчисления высказываний есть булева алгебра, если отождествить равносильные формулы, т. е в качестве элементов множества рассматривать классы эквивалентности, определяемые отношением равносильности. Дополнение при этом определяется логической связкой отрицания \neg . Единицей является класс тавтологий, а нулем — класс противоречий. Эту алгебру часто называют *алгеброй Линденбаума–Тарского*.

4. Рассмотрим произвольное натуральное число n . В качестве \mathfrak{M} возьмем множество всех его натуральных делителей. Для элементов $x, y \in \mathfrak{M}$ введем операции

$$x \wedge y = D(x, y); \quad x \vee y = M(x, y); \quad \neg x = \frac{n}{x}.$$

Нулем алгебры является число 1, а единицей — n .

Установим справедливость некоторых равносильностей теоремы 3.4 для произвольной булевой алгебры.

Теорема 3.16. Пусть $(\mathfrak{M}, \vee, \wedge, \neg)$ — булева алгебра. Тогда для любых элементов $a, b \in \mathfrak{M}$ справедливы законы де Моргана и снятия двойного отрицания:

$$\neg(a \wedge b) = \neg a \vee \neg b; \quad \neg(a \vee b) = \neg a \wedge \neg b; \quad \neg(\neg a) = a. \quad (3.19)$$

Доказательство. Последнее равенство в (3.19) очевидно в силу коммутативности операций \vee, \wedge и единственности дополнения (см. лемму 3.7)). Докажем, например, второе тождество в (3.19). Имеем

$$\begin{aligned} (a \vee b) \vee (\neg a \wedge \neg b) &= ((a \vee b) \vee \neg a) \wedge ((a \vee b) \vee \neg b) = \\ &= ((a \vee \neg a) \vee b) \wedge ((b \vee \neg b) \vee a) = \\ &= (1 \vee b) \wedge (1 \vee a) = 1. \end{aligned}$$

Аналогично доказывается, что $(a \vee b) \wedge (\neg a \wedge \neg b) = 0$. Таким образом, элемент $\neg a \wedge \neg b$ есть дополнение элемента $a \vee b$. ■

В теории графов было введено понятие изоморфизма графов, как взаимно-однозначное отображение, сохраняющее соответствие между ребрами графов (см. определение 2.38). Применим эти соображения для булевых алгебр.

Определение 3.31. Взаимно-однозначное отображение φ между двумя булевыми алгебрами \mathfrak{M} и \mathfrak{N} называют *изоморфиzmом*, если ноль и единица алгебры \mathfrak{M} переходят соответственно в ноль и единицу алгебры \mathfrak{N} и для любых элементов $a, b \in \mathfrak{M}$

$$\varphi(\neg a) = \neg \varphi(a), \quad \varphi(a \vee b) = \varphi(a) \vee \varphi(b), \quad \varphi(a \wedge b) = \varphi(a) \wedge \varphi(b).$$

Булевые алгебры \mathfrak{M} и \mathfrak{N} в этом случае называют *изоморфными*.

Как уже отмечалось, аксиомы булевой алгебры соответствуют логическим законам формул логики высказываний, установленным в теореме 3.4. Согласно принципу двойственности логики высказываний, если конъюнкцию заменить дизъюнкцией, дизъюнкцию — конъюнкцией, 0 — на 1, а 1 — на 0, то должна получиться та же система аксиом. Это рассуждение приводит к очень важному примеру изоморфизма булевых алгебр.

Пример 3.42. Пусть $(\mathfrak{M}, \vee, \wedge, \neg)$ — булева алгебра. Построим изоморфную булеву алгебру \mathfrak{M}^* , состоящую из элементов алгебры \mathfrak{M} , но с другими операциями. Изоморфизм осуществляется отображение $\varphi : a \rightarrow \neg a$. То, что отображение φ является изоморфизмом, очевидным образом следует из теоремы 3.16.

Определение 3.32. Булеву алгебру \mathfrak{M}^* называют *двойственной* к булевой алгебре \mathfrak{M} .

Упражнение 3.1. Докажите, что $(\mathfrak{M}^*)^* = \mathfrak{M}$.

В общем случае вопрос о существовании изоморфизма решает следующая теорема.

Теорема 3.17. Любая конечная булева алгебра изоморфна булевой алгебре всех подмножеств некоторого множества (алгебре Кантора).

Следствие 3.3. Количество элементов в любой конечной булевой алгебре есть степень двойки.

С булевой алгеброй тесно связано еще одно понятие.

Определение 3.33. Рассмотрим ассоциативное кольцо с единицей (см. определение 1.18), в котором для каждого элемента выполнен закон идемпотентности по умножению: $aa = a$. Такое кольцо называют *булевым*.

Булево кольцо легко можно преобразовать в булеву алгебру. Введем операции \vee , \wedge и \neg следующим образом:

$$a \vee b = a + b + ab, \quad a \wedge b = ab, \quad \neg a = a + 1. \quad (3.20)$$

Обратно, если \mathfrak{M} — булева алгебра, то ее можно превратить в булево кольцо, полагая

$$ab = a \wedge b, \quad a + b = (a \wedge \neg b) \vee (b \wedge \neg a). \quad (3.21)$$

Упражнение 3.2. Докажите, что преобразованиями (3.20) и (3.21) можно получить из булевой алгебры булево кольцо и, соответственно, из булева кольца — булеву алгебру.

Историческая справка. Джордж Буль родился в 1815 г. в г. Линкольне. Сын сапожного мастера, увлекающегося математикой, он окончил только начальную школу и дальнейшие знания приобретал самостоятельно. На жизнь Буль зарабатывал как учитель. В 1849 г. стал профессором математики в Куинсколледже (Ирландия), где преподавал до конца жизни. Одна из его пяти дочерей — Этель Лилиан, в замужестве Войнич, стала автором известного романа «Овод».

3.2. Булевы функции

3.2.1. Основные понятия

Определение 3.34. Функцию $f(x_1, x_i, \dots, x_n)$, аргументы и значение которой определены на множестве $\{0, 1\}$ называют *булевой функцией*.

Булевы функции также называют *переключательными, логическими функциями* или функциями алгебры логики.

Множество $\{0, 1\}$ будем в дальнейшем обозначать B .

Для задания булевой функции $f(x_1, x_2, \dots, x_n)$ достаточно определить, какое значение функции соответствует каждому из наборов значений аргументов, т. е. выписать таблицу (табл. 3.23).

Таблица 3.23

x_1	x_2	\cdots	x_{n-1}	x_n	$f(x_1, x_2, \dots, x_{n-1}, x_n)$
0	0	\cdots	0	0	$f(0, 0, \dots, 0, 0)$
0	0	\cdots	0	1	$f(0, 0, \dots, 0, 1)$
\cdots	\cdots	\cdots	\cdots	\cdots	\cdots
1	1	\cdots	1	0	$f(1, 1, \dots, 1, 0)$
1	1	\cdots	1	1	$f(1, 1, \dots, 1, 1)$

Замечание 3.23. Заметим, что табл. 3.23 и таблица истинности для формулы логики высказываний функции (см. табл. 3.2) идентичны. Таким образом, формулы логики высказываний и булевы функции это два различных подхода к интерпретации таблицы истинности.

Формулу логики высказываний можно считать синтаксической записью таблицы истинности, а для булевой функции это табличное задание функциональной зависимости.

Пусть $F(X_1, \dots, X_n)$ и $G(X_1, \dots, X_n)$ — формулы логики высказываний. Зададим в формулах прописные буквы X_i на строчные x_i ($1 \leq i \leq n$). Получим булевые функции $f(x_1, \dots, x_n)$ и $g(x_1, \dots, x_n)$. Тогда формулы $F(X_1, \dots, X_n)$ и $G(X_1, \dots, X_n)$ равносильны в том и только в том случае, когда функции $f(x_1, \dots, x_n)$ и $g(x_1, \dots, x_n)$ равны.

Замечание 3.23 позволяет перенести на булевы функции результаты предыдущего подраздела (законы логики высказываний, нормальные формы и алгоритмы приведения к ним).

Справедлива следующая теорема.

Теорема 3.18. Число различных булевых функций от n переменных равно 2^{2^n} .

Доказательство. Зафиксируем число аргументов n . Очевидно, что число различных булевых функций от n переменных равно числу различных таблиц истинности. При лексикографическом упорядочении строк число различных таблиц равно числу способов заполнения последнего столбца (значений функции). Число способов заполнения этого столбца равно 2^h , где h — высота столбца, но $h = 2^n$. Таким образом, число различных булевых функций от n переменных равно 2^{2^n} . ■

Упражнение 3.3. Найдите число булевых функций от n переменных, которые на фиксированных $m \leq n$ наборах принимают значение 0.

Рассмотрим примеры булевых функций. Основную роль здесь играют функции одной (табл. 3.24,*a*) и двух (табл. 3.24,*b*) переменных.

Замечание 3.24. Для обозначения функций наряду с префиксной записью $f(x_1, x_2)$ будем применять и инфиксную $x_1 f x_2$.

Таблица 3.24

a

x	0	1	x	\bar{x}
0	0	1	0	1
1	0	1	1	0

b

x	y	xy	$x \vee y$	$x \rightarrow y$	$x \leftrightarrow y$	$x \oplus y$	$x y$	$x \downarrow y$
0	0	0	0	1	1	0	1	1
0	1	0	1	1	0	1	1	0
1	0	0	1	0	0	1	1	0
1	1	1	1	1	1	0	0	0

Условимся также об обозначении некоторых функций.

Для функций $x \vee y$, $x \rightarrow y$, $x \leftrightarrow y$ сохраним названия, идущие из логики высказываний: дизъюнкция, импликация и эквиваленция. Функции xy и $x \oplus y$, — естественно, умножение и сложение; умножение можно было бы назвать конъюнкцией, сложение иногда называют сложением по модулю 2, или функцией неравнозначности. Функция $x | y$ — штрих Шеффера, $x \downarrow y$ — стрелка Пирса, стрелка Лукасевича или функция Вебба).

Определение булевой функции позволяет рассматривать функции только от фиксированного числа аргументов. Аналогично формулам логики высказываний введем понятие *фиктивных* переменных.

Определение 3.35. Булева функция $f(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n)$ существенно зависит от переменной x_i , если существует такой булевый набор значений переменных $(\alpha_1, \dots, \alpha_{i-1}, \alpha_{i+1}, \dots, \alpha_n)$, что

$$f(\alpha_1, \dots, \alpha_{i-1}, 0, \alpha_{i+1}, \dots, \alpha_n) \neq f(\alpha_1, \dots, \alpha_{i-1}, 1, \alpha_{i+1}, \dots, \alpha_n). \quad (3.22)$$

В этом случае переменную x_i называют *существенной*. Если x_i не является существенной переменной, то ее называют *фиктивной*.

Пусть для функции $f(x_1, \dots, x_n)$ переменная x_i является фиктивной. По таблице функции $f(x_1, \dots, x_n)$ построим новую таблицу, вычеркнув все строки вида $(\alpha_1, \dots, \alpha_{i-1}, 1, \alpha_{i+1}, \dots, \alpha_n)$ и столбец переменной x_i . Полученная таблица определяет новую функцию $g(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$.

Говорят, что функция g получена из функции f путем удаления фиктивной переменной x_i , а также, что функция f получается из функции g путем введения фиктивной переменной x_i .

Определение 3.36. Функции $f(x_1, \dots, x_n)$ и $g(x_1, \dots, x_m)$ называют *равными*, если одна из них получается из другой с помощью введения или вычеркивания некоторых фиктивных переменных.

Пример 3.43. Пусть функция $f(x_1, x_2, x_3)$ задана таблицей значений (табл. 3.25). Определим ее существенные и фиктивные переменные.

Таблица 3.25

x_1	x_2	x_3	$f(x_1, x_2, x_3)$	x_1	x_2	x_3	$f(x_1, x_2, x_3)$
0	0	0	0	1	0	0	0
0	0	1	0	1	0	1	0
0	1	0	1	1	1	0	1
0	1	1	1	1	1	1	1

Замечание 3.25. Как и для формул логики высказываний в цепочке равенств булевых функций будем ставить номера использованных равносильностей из теоремы 3.4.

Неравенство (3.22) выполняется только для x_2 : $f(0, 0, 0) \neq f(0, 1, 0)$. Переменная x_2 — существенная, а переменные x_1, x_3 — фиктивные. Действительно, если используя табл. 3.25, по алгоритму 3.3 построить СДНФ для $f(x_1, x_2, x_3)$, то получим

$$\begin{aligned} f(x_1, x_2, x_3) &= \bar{x}_1 x_2 \bar{x}_3 \vee \bar{x}_1 x_2 x_3 \vee x_1 x_2 \bar{x}_3 \vee x_1 x_2 x_3 \stackrel{7}{=} \\ &\stackrel{7}{=} \bar{x}_1 x_2 (\bar{x}_3 \vee x_3) \vee x_1 x_2 (\bar{x}_3 \vee x_3) \stackrel{10}{=} \bar{x}_1 x_2 \vee x_1 x_2 \stackrel{7}{=} \\ &\stackrel{7}{=} x_2 (\bar{x}_1 \vee x_1) \stackrel{10}{=} x_2. \end{aligned}$$

Таким образом, $f(x_1, x_2, x_3) = g(x_2) = x_2$.

Упражнение 3.4. Докажите, что для любой булевой функции, отличной от константы 0 или 1, существует равная ей, у которой все переменные существенные.

Замечание 3.26. При лексикографическом упорядочении строк таблицы истинности булеву функцию иногда задают вектором значений функции. Например, булеву функцию из примера 3.43 можно записать как $f(x_1, x_2, x_3) = (00110011)$.

Задача 3.14. Определите существенные и фиктивные переменные функции $f(x_1, x_2, x_3) = (11110011)$.

Упражнение 3.5. Найдите все булевые функции двух и трех переменных, у которых все переменные существенны.

С табличным заданием функции непосредственно связан такой ее параметр, как *вес*.

Определение 3.37. Множество двоичных наборов $\{a_1, \dots, a_n\}$, на которых булева функция $f(x_1, \dots, x_n)$ принимает значение 1, называют *областью истинности* функции f и обозначают E_f .

Мощность области истинности функции f называют *весом* функции f и обозначают $\|f\|$.

Очевидно, что $0 \leq \|f\| \leq 2^n$, причем равенства достигаются лишь для функций-констант 0 и 1.

Определение 3.38. Булеву функцию от $2l + 1$ переменной, равную 1, если $m \geq l + 1$ переменных принимают значение 1, и 0 — в противном случае, называют *мажоритарной* и обозначают Maj_{2l+1} .

Пример 3.44. Покажем, что $\|Maj_{2l+1}\| = 2^{2l}$.

Действительно, рассмотрим двоичный набор α

$$\alpha = (\alpha_1, \dots, \alpha_{2l+1}), \quad \alpha_1 + \dots + \alpha_{2l+1} \geq l + 1, \quad Maj_{2l+1}(\alpha_1, \dots, \alpha_{2l+1}) = 1.$$

Каждому такому набору α однозначно соответствует набор $\bar{\alpha}$

$$\bar{\alpha} = (\bar{\alpha}_1, \dots, \bar{\alpha}_{2l+1}), \quad \bar{\alpha}_1 + \dots + \bar{\alpha}_{2l+1} \leq l, \quad Maj_{2l+1}(\bar{\alpha}_1, \dots, \bar{\alpha}_{2l+1}) = 0.$$

Таким образом $Maj_{2l+1}(x_1, \dots, x_{2l+1})$ принимает единичные значения на половине наборов.

Для булевых функций используются и другие способы задания. Рассмотрим *геометрический способ*.

Под геометрическим способом задания булевой функции $f(x_1, \dots, x_n)$ понимается выделение тех вершин n -мерного двоичного куба, на наборах координат которых функция принимает единичное значение.

На рис. 3.11 показаны геометрические представления для функций $f(x_1) = \bar{x}_1$ (рис. 3.11, а), $g(x_1, x_2) = x_1 \oplus x_2$ (рис. 3.11, б) и мажоритарной функции $Maj_3(x_1, x_2, x_3) = x_1x_2 \vee x_1x_3 \vee x_2x_3$ (рис. 3.11, в).

Как видно из рис. 3.11 геометрический способ представления булевых функций n переменных эффективен для небольших размерностей $n = 1, 2, 3$.

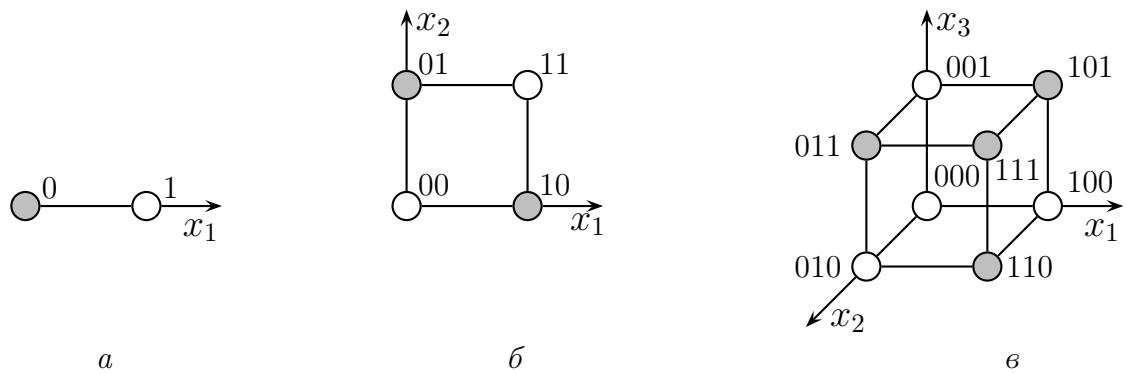


Рис. 3.11

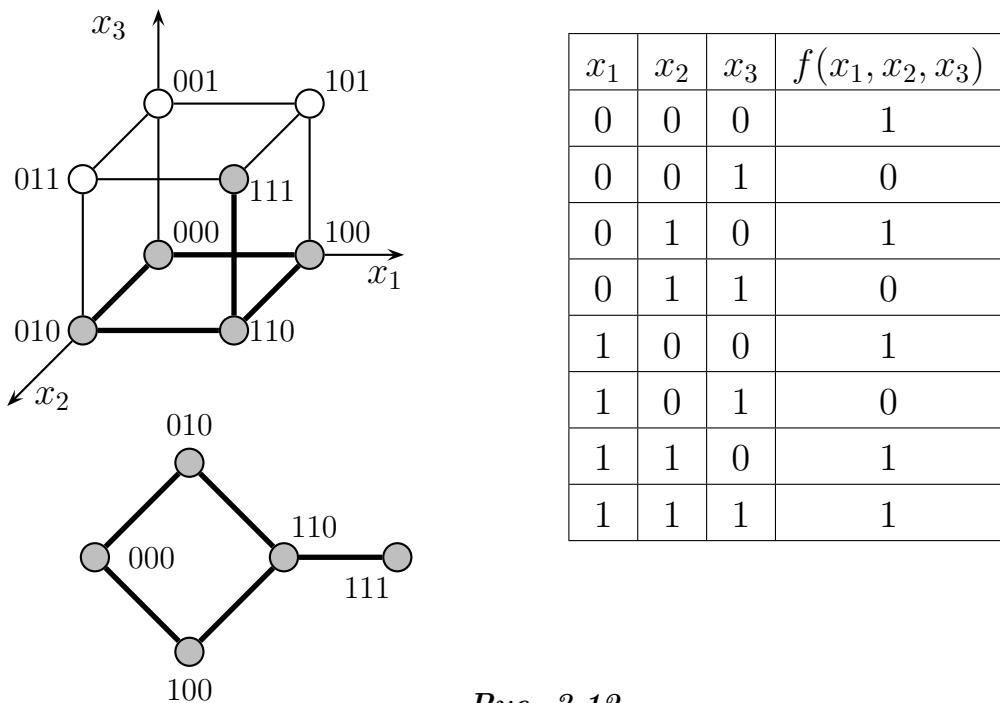


Рис. 3.12

Часто по геометрическому заданию функции строят *граф связности* вершин n -мерного куба, соответствующий данной функции. Для этого сначала отмечают те ребра, у которых оба конца выделены, т. е. соответствующие вершины лежат в области истинности. Затем все остальные ребра и вершины, не лежащие в области истинности, отбрасывают.

На рис. 3.12 изображены геометрическое задание и соответствующий граф связности для булевой функции $f(x_1, x_2, x_3)$, заданной таблицей значений.

3.2.2. Замкнутость и полнота

Определение 3.39. Пусть имеется некоторый набор булевых функций K . Суперпозицией функций из этого набора называют булевые функции, полученные с помощью конечного числа применения двух операций:

- 1) переименование любой переменной, входящей в функцию из K ;
- 2) подстановка функции из K или полученной ранее суперпозиции вместо любой переменной.

Суперпозицию иначе называют *сложной функцией*.

Пример 3.45. Пусть $x \oplus y \in K$. Переименованием переменной y получаем функцию $0 = x \oplus x$.

Пусть $x \vee y, \bar{x} \in K$. Тогда подстановкой получаем функцию xy , поскольку

$$xy = \overline{\bar{x} \vee \bar{y}}. \quad (3.23)$$

Определение 3.40. Замыканием набора функций из K (обозначают $[K]$) называют множество всех суперпозиций этого набора.

Класс функций K называют *замкнутым*, если его замыкание совпадает с ним самим.

Лемма 3.8. Справедливы следующие свойства замыкания:

1. $K \subseteq [K]$.
2. $[[K]] = [K]$.
3. Если $F \subseteq G$, то $[F] \subseteq [G]$.
4. $[F \cap G] \subseteq [F] \cap [G]$.
5. $[F] \cup [G] \subseteq [F \cup G]$.

Доказательство очевидным образом следует из определения 3.40.

Определение 3.41. Функцию $f(x_1, \dots, x_n)$ называют *сохраняющей ноль*, если $f(0, \dots, 0) = 0$.

Через T_0 обозначим класс всех функций, сохраняющих 0, т. е.

$$T_0 = \{f(x_1, \dots, x_n) \mid f(0, \dots, 0) = 0\}.$$

Лемма 3.9. Класс T_0 не пуст и не совпадает со всем множеством булевых функций B_F .

Доказательство. Функции $xy, x \vee y, x \oplus y \in T_0$, следовательно $T_0 \neq \emptyset$, а функции $x \leftrightarrow y, x \mid y, x \downarrow y \notin T_0$, т. е. $T_0 \neq B_F$. ■

Справедлива следующая теорема.

Теорема 3.19. Класс T_0 замкнут.

Доказательство. Если $f(x_1, \dots, x_n)$ сохраняет 0 и y_1, \dots, y_n — новые переменные, то очевидно, что $f(y_1, \dots, y_n)$ также сохраняет ноль. Это означает, что T_0 замкнут относительно операции переименования переменных. Пусть $f(x_1, \dots, x_k), g_1(x_1, \dots, x_n), \dots, g_k(x_1, \dots, x_n)$ сохраняют 0. Тогда

$$f(g_1(0, \dots, 0), \dots, g_k(0, \dots, 0)) = f(0, \dots, 0) = 0.$$

Это означает, что T_0 замкнут относительно операции подстановки. Таким образом доказано, что T_0 — замкнутый класс. ■

Упражнение 3.6. Обозначим через T_0^n множество всех булевых функций от n переменных, сохраняющих 0. Докажите, что

$$|T_0^n| = 2^{2^n - 1}.$$

Определение 3.42. Функцию $f(x_1, \dots, x_n)$ называют *сохраняющей единицу*, если $f(1, \dots, 1) = 1$.

Через T_1 обозначим класс всех функций, сохраняющих 1, т. е.

$$T_1 = \{f(x_1, \dots, x_n) \mid f(1, \dots, 1) = 1\}.$$

Лемма 3.10. Класс T_1 не пуст и не совпадает со всем множеством булевых функций B_F .

Доказательство. Функции $xy, x \vee y, x \rightarrow y \in T_1$, т. е. $T_1 \neq \emptyset$, а функции $x \oplus y, x \mid y, x \downarrow y \notin T_1$, следовательно $T_1 \neq B_F$. ■

Теорема 3.20. Класс T_1 является замкнутым.

Доказательство теоремы 3.20 аналогично приведенному в теореме 3.19.

Упражнение 3.7. Обозначим через T_1^n множество всех булевых функций от n переменных, сохраняющих 1. Докажите, что

$$|T_1^n| = 2^{2^n - 1}.$$

Обозначим через B_F класс всех булевых функций.

Определение 3.43. Класс булевых функций K называют *полным*, если $[K] = B_F$.

Другими словами, класс K полон, если любую булеву функцию можно получить из K конечным числом операций подстановки и переименования переменных.

Очевидно следующее утверждение.

Лемма 3.11 (о моделировании). Если K — полный класс и любая функция класса K выражается (с помощью операций подстановки и переименования переменных) через функции класса L , то L — также полный класс.

Теорема 3.21. Следующие классы функций являются полными

- | | |
|---------------------------------------|--------------------------------------|
| а) $K_1 = \{xy, x \vee y, \bar{x}\};$ | б) $K_2 = \{xy, \bar{x}\};$ |
| в) $K_3 = \{x \vee y, \bar{x}\};$ | г) $K_4 = \{xy, x \oplus y, 1, 0\}.$ |

Доказательство. Докажем сначала полноту класса K_1 . Рассмотрим произвольную булеву функцию $f \in B_F$. По таблице значений функции f (используя алгоритм 3.3) построим равную ей функцию g в виде СДНФ.

Но, согласно определению СДНФ, в записи функции g использованы только функции из класса K_1 . Следовательно, f может быть получена из функций класса K_1 операциями подстановки и переименования переменных, т. е. $f \in [K_1]$. Следовательно, любая булева функция принадлежит замыканию $[K_1]$. Таким образом, K_1 — полный класс.

Из равносильности 12 теоремы 3.4 получаем равенство

$$x \vee y = \bar{x} \bar{y}. \quad (3.24)$$

Таким образом, функции K_2 выражаются через функции K_1 . По лемме 3.11 K_2 — полный класс.

Аналогично, используя (3.23) и применяя лемму 3.11, получаем полноту класса K_3 . Полнота класса K_4 следует из полноты класса K_2 и очевидного равенства $\bar{x} = x \oplus 1$. ■

3.2.3. Полиномы Жегалкина и линейные функции

Определение 3.44. Рассмотрим полный класс $K_4 = \{xy, x \oplus y, 1, 0\}$ из теоремы 3.21. Очевидно, что его замыкание содержит многочлены (от любого числа переменных) над полем \mathbb{Z}_2 .

Эти многочлены называют *полиномами Жегалкина*¹. Поскольку для элементов $a \in \mathbb{Z}_2$ справедливо тождество $a^2 = a$, то полином Жегалкина можно записать в виде

$$\sum_{\{i_1, \dots, i_k\} \subseteq \{1, \dots, n\}} \bigoplus a_{i_1, \dots, i_k} x_{i_1} \cdots x_{i_k}, \quad a_{i_1, \dots, i_k} \in B, \quad (3.25)$$

и суммирование ведется по всем подмножествам множества $\{1, \dots, n\}$.

Число переменных в самом длинном слагаемом представления (3.25) называют *степенью нелинейности* булевой функции $\deg(f)$.

Пример 3.46. Рассмотрим булеву функцию от двух переменных

$$f(x_1, x_2) = 1 \cdot x_1 x_2 \oplus 0 \cdot x_1 \oplus 1 \cdot x_2 \oplus 1.$$

¹Полином был предложен Иваном Жегалкиным в 1927 году.

Это полином Жегалкина от двух переменных x_1 и x_2 , с коэффициентами $a_{12} = 1$, $a_2 = 1$, $a_1 = 0$, $a_0 = 1$. Данный полином можно записать и так:

$$f(x_1, x_2) = x_1 x_2 \oplus x_2 \oplus 1, \quad \deg(f) = 2.$$

Теорема 3.22. Для любой булевой функции $f(x_1, \dots, x_n)$ существует единственное представление в виде полинома Жегалкина.

Доказательство. Существование полинома Жегалкина следует из полноты класса K_4 . Докажем единственность.

В сумме (3.25) для полинома Жегалкина от n переменных всего 2^n слагаемых (число всех подмножеств множества из n элементов), т. е. каждый полином определяется булевым набором 2^n коэффициентов. Отсюда следует, что всего полиномов от n переменных 2^{2^n} — столько же, сколько и булевых функций от n переменных (см. теорему 3.18). Но разным булевым функциям (с различными таблицами значений) очевидно соответствуют разные полиномы Жегалкина. Получаем взаимно-однозначное соответствие между B_F и полиномами Жегалкина. ■

Замечание 3.27. В англоязычной литературе представление булевой функции в виде (3.25) иногда называют *алгебраической нормальной формой*.

Пример 3.47. Для функций $x_1 \rightarrow x_2$ и $x_1 \vee x_2$ справедливы равенства:

$$x_1 \rightarrow x_2 = x_1 x_2 \oplus x_1 \oplus 1, \quad x_1 \vee x_2 = x_1 x_2 \oplus x_1 \oplus x_2.$$

Упражнение 3.8. Докажите, что функция, представленная полиномом Жегалкина, существенно зависит от всех входящих в него переменных.

Рассмотрим три алгоритма представления булевой функции полиномом Жегалкина.

Первый из них можно было бы определить как *метод равносильных преобразований*.

Алгоритм 3.11 (метод равносильных преобразований).

Шаг 1. По таблице значений функции f (используя алгоритм 3.3) построим равную ей функцию g в виде СДНФ.

Шаг 2. Используя соотношения

$$\bar{x} = x \oplus 1, \quad x \vee y = \bar{x} \bar{y} = (x \oplus 1)(y \oplus 1) \oplus 1 = xy \oplus x \oplus y, \quad (3.26)$$

заменяем отрицание и дизъюнкцию в формуле g на «умножение» (конъюнкцию) и «сложение» (\oplus).

Шаг 3. «Приводим подобные члены» согласно равенству $x \oplus x = 0$.

Замечание 3.28. Полезно использовать следующее равенство: если $xy = 0$, то $x \vee y = x \oplus y$.

Пример 3.48. Пусть функция $f(x_1, x_2, x_3)$ задана табл. 3.26.

Таблица 3.26

x_1	x_2	x_3	$f(x_1, x_2, x_3)$	x_1	x_2	x_3	$f(x_1, x_2, x_3)$
0	0	0	0	1	0	0	0
0	0	0	1	1	0	1	1
0	0	1	0	1	1	0	1
0	1	1	0	1	1	1	0

После шага 1 получаем представление

$$f(x_1, x_2, x_3) = g(x_1, x_2, x_3) = \bar{x}_1 \bar{x}_2 x_3 \vee x_1 \bar{x}_2 x_3 \vee x_1 x_2 \bar{x}_3.$$

Продолжим преобразования функции $g(x_1, x_2, x_3)$, используя (3.26). С учетом замечания 3.28 имеем:

$$\begin{aligned} g(x_1, x_2, x_3) &= \bar{x}_1 \bar{x}_2 x_3 \oplus x_1 \bar{x}_2 x_3 \oplus x_1 x_2 \bar{x}_3 = \\ &= (x_1 \oplus 1)(x_2 \oplus 1)x_3 \oplus x_1(x_2 \oplus 1)x_3 \oplus x_1 x_2(x_3 \oplus 1) = \\ &= x_1 x_2 x_3 \oplus x_1 x_2 \oplus x_2 x_3 \oplus x_3. \end{aligned}$$

Второй способ можно назвать *методом неопределенных коэффициентов*. Изложим этот способ для функции $f(x_1, x_2, x_3)$ из примера 3.48.

Пример 3.49. Рассмотрим функцию $f(x_1, x_2, x_3)$, заданную табл. 3.26. Тогда последовательно находим:

- 1) $a_0 = f(0, 0, 0) = 0;$
- 2) $a_0 \oplus a_1 = f(1, 0, 0) = 0, a_1 = 0;$
- 3) $a_0 \oplus a_2 = f(0, 1, 0) = 0, a_2 = 0;$
- 4) $a_0 \oplus a_3 = f(0, 0, 1) = 1, a_3 = 1;$
- 5) $a_0 \oplus a_1 \oplus a_2 \oplus a_{12} = f(1, 1, 0) = 1, a_{12} = 1;$
- 6) $a_0 \oplus a_1 \oplus a_3 \oplus a_{13} = f(1, 0, 1) = 1, a_{13} = 0;$
- 7) $a_0 \oplus a_2 \oplus a_3 \oplus a_{23} = f(0, 1, 1) = 0, a_{23} = 1;$
- 8) $a_0 \oplus a_1 \oplus a_2 \oplus a_3 \oplus a_{12} \oplus a_{13} \oplus a_{23} \oplus a_{123} = f(1, 1, 1) = 0, a_{123} = 1.$

Подставив найденные коэффициенты в представление (3.25), получим

$$f(x_1, x_2, x_3) = x_1 x_2 x_3 \oplus x_1 x_2 \oplus x_2 x_3 \oplus x_3.$$

Задача 3.15. Методом неопределенных коэффициентов найдем полином Жегалкина для булевой функции трех переменных $f(x_1, x_2, x_3)$, заданной лексикографически упорядоченным столбцом значений таблицы истинности: $(0, 1, 1, 0, 1, 0, 1, 1, 1)$.

Согласно представлению (3.25) для функции от трех переменных полином Жегалкина с неопределенными коэффициентами имеет вид:

$$\begin{aligned} f((x_1, x_2, x_3)) = & a_0 \oplus a_1 x_1 \oplus a_2 x_2 \oplus a_3 x_3 \oplus a_{12} \oplus x_1 x_2 \oplus a_{13} x_1 x_3 \oplus \\ & \oplus a_{23} x_2 x_3 \oplus a_{123} x_1 x_2 x_3. \end{aligned}$$

Выпишем систему уравнений для неизвестных коэффициентов и последовательно решим ее:

- 1) $a_0 = f(0, 0, 0) = 0;$
- 2) $a_0 \oplus a_1 = f(1, 0, 0) = 1, a_1 = 1;$
- 3) $a_0 \oplus a_2 = f(0, 1, 0) = 1, a_2 = 1;$
- 4) $a_0 \oplus a_3 = f(0, 0, 1) = 1, a_3 = 1;$
- 5) $a_0 \oplus a_1 \oplus a_2 \oplus a_{12} = f(1, 1, 0) = 1, a_{12} = 1;$
- 6) $a_0 \oplus a_1 \oplus a_3 \oplus a_{13} = f(1, 0, 1) = 0, a_{13} = 0;$
- 7) $a_0 \oplus a_2 \oplus a_3 \oplus a_{23} = f(0, 1, 1) = 0, a_{23} = 0;$
- 8) $a_0 \oplus a_1 \oplus a_2 \oplus a_3 \oplus a_{12} \oplus a_{13} \oplus a_{23} \oplus a_{123} = f(1, 1, 1) = 1, a_{123} = 1.$

Подставив найденные коэффициенты в представление (3.25), получим

$$f(x_1, x_2, x_3) = x_1 x_2 x_3 \oplus x_1 x_2 \oplus x_3 \oplus x_2 \oplus x_1. \quad (3.27)$$

Третий алгоритм часто называют *методом треугольника*. Впервые был предложен в [46], более подробно описан в [47].

Алгоритм 3.12 (метод треугольника).

Шаг 1. Строим таблицу значений функции (как и ранее строки в таблице идут в порядке возрастания двоичных кодов). Вектор значений функции выписываем напротив первой строки таблицы.

Шаг 2. Далее заполняем треугольник Паскаля, складывая попарно соседние булевые значения текущей строки ($\alpha_i \oplus \alpha_{i+1}$). Результат сложения выписываем в следующей строке. Продолжаем вычисления, пока в очередной строке не останется лишь одна цифра.

Шаг 3. Числа на левой стороне треугольника есть коэффициенты полинома при конъюнкциях, соответствующих наборам значений переменных.

Конъюнкции строятся по булевым наборам в левой части таблицы по следующему принципу: если напротив переменной x_i стоит 1, то переменная входит в конъюнкцию; в противном случае переменная отсутствует в конъюнкции. Набору $(0, 0, 0)$ соответствует константа 1.

Шаг 4. В состав полинома Жегалкина входят только конъюнкции из строк с единицами на левой стороне треугольника.

Пример 3.50. Рассмотрим работу алгоритма 3.12 для функции из примера 3.48. Протокол работы записан в таблице 3.27.

Строки с единичными значениями функции выделены цветом, значения левой стороны треугольника – „жирным“ шрифтом.

Таблица 3.27

x_1	x_2	x_3	f																
0	0	0	0	0		1		0		0		0		1		1		0	
0	0	1	1		1		1		0		0		1		0		1		x_3
0	1	0	0			0		1		0		1		1		1			
0	1	1	0				1		1		1		0		0				x_2x_3
1	0	0	0					0		0		1		0					
1	0	1	1						0		1		1						
1	1	0	1							1		0							x_1x_2
1	1	1	0								1								$x_1x_2x_3$

Получаем представление функции f полиномом Жегалкина, совпадающее с (3.27).

Приведем обоснование алгоритма 3.12 для $n = 2$. В общем случае доказательство можно провести аналогично. Запишем общий вид полинома Жегалкина для функции двух переменных. Имеем

$$f(x_1, x_2) = a_0 \oplus a_1x_1 \oplus a_2x_2 \oplus a_{12}x_1x_2.$$

Используя метод неопределенных коэффициентов, получаем:

$$\begin{cases} a_0 = f(0, 0), \\ a_1 = f(0, 0) \oplus f(1, 0), \\ a_2 = f(0, 0) \oplus f(0, 1), \\ a_{12} = f(0, 0) \oplus f(0, 1) \oplus f(1, 0) \oplus f(1, 1). \end{cases} \quad (3.28)$$

Значения в правой части (3.28) совпадают со значениями левой стороны треугольника Паскаля, построенного для случая $n = 2$.

Определение 3.45. Функцию $f(x_1, \dots, x_n)$ называют *линейной*, если она представима в виде линейного полинома Жегалкина, т. е. если существуют $a_0, a_1, \dots, a_n \in B$ такие, что

$$f(x_1, \dots, x_n) = a_0 \oplus a_1x_1 \oplus \dots \oplus a_nx_n.$$

Класс всех линейных функций обозначим буквой L .

Лемма 3.12. Класс L не пуст и не совпадает со всем множеством булевых функций B_F .

Доказательство. Функции $0, 1, \bar{x}, x \oplus y \in L$, следовательно $L \neq \emptyset$, а функции $xy, x \vee y, x \rightarrow y \notin L$, т. е. $L \neq B_F$. ■

Замечание 3.29. Любая булева функция от одной переменной линейна.

Очевидна следующая теорема.

Теорема 3.23. Класс L является замкнутым.

Упражнение 3.9. Обозначим через L^n множество всех линейных булевых функций от n переменных. Докажите, что

$$|L^n| = 2^{n+1}.$$

Следующее утверждение называют *леммой о нелинейной функции*.

Лемма 3.13. Пусть $f(x_1, \dots, x_n)$ — нелинейная функция. Тогда замыкание класса $K = \{f(x_1, \dots, x_n), 0, 1, \bar{x}\}$ содержит произведение xy .

Доказательство. Так как $f(x_1, \dots, x_n)$ — нелинейная функция, то ее полином Жегалкина содержит хотя бы одно нелинейное слагаемое вида: $x_{i_1}x_{i_2} \cdots x_{i_k}, k \geq 2$. Среди них выберем самое короткое. Для переменных, входящих в конъюнкцию, положим $x_{i_3} = \dots = x_{i_k} = 1$, а те, которые в данное слагаемое не входят, положим равными нулю. Введем новые обозначения: $x = x_{i_1}$, $y = x_{i_2}$. Тогда функция f примет вид

$$f(x, y) = xy \oplus \alpha x \oplus \beta y \oplus \gamma, \quad \alpha, \beta, \gamma \in B.$$

Составим следующую таблицу (табл. 3.28):

Таблица 3.28

α	β	γ	$f(x, y)$	xy
0	0	0	xy	$xy = f(x, y)$
0	0	1	$xy \oplus 1 = \bar{xy}$	$xy = \overline{f(x, y)}$
0	1	0	$xy \oplus y = \bar{xy}$	$xy = f(\bar{x}, y)$
0	1	1	$xy \oplus y \oplus 1 = \bar{\bar{xy}}$	$xy = \overline{f(\bar{x}, y)}$
1	0	0	$xy \oplus x = x\bar{y}$	$xy = f(x, \bar{y})$
1	0	1	$xy \oplus x \oplus 1 = \bar{x}\bar{y}$	$xy = \overline{f(x, \bar{y})}$
1	1	0	$xy \oplus x \oplus y = \bar{x}\bar{y}$	$xy = \overline{f(\bar{x}, \bar{y})}$
1	1	1	$xy \oplus x \oplus y \oplus 1 = \bar{x}\bar{y}$	$xy = f(\bar{x}, \bar{y})$

Последний столбец табл. 3.28 содержит представление функции xy через отрицание, нелинейную функцию $f(x_1, \dots, x_n)$ и подстановку в нее констант 0 и 1. ■

Пример 3.51. Получим функцию xy с помощью констант, отрицания и нелинейной функции $f(x_1, x_2, x_3) = x_1x_2x_3 \oplus x_1x_2 \oplus x_1$. Воспользуемся леммой 3.13 о нелинейной функции.

Заметим, что требование выбора самого короткого нелинейного слагаемого является существенным. Возьмем, например, слагаемое $x_1x_2x_3$, тогда (полагая $x_3 = 1$) получим $f(x_1, x_2, 1) = x_1x_2 \oplus x_1x_2 \oplus x_1 = x_1$ и лемма неприменима.

Сделаем правильный выбор и рассмотрим слагаемое x_1x_2 . Полагая $x_3 = 0$, получаем $f(x_1, x_2, 0) = x_1x_2 \oplus x_1$. Тогда $xy = f(x, \bar{y}, 0)$.

Задача 3.16. Какие из следующих функций линейны:

- a) $x \downarrow y$; b) $\overline{x \leftrightarrow \bar{y}}$; c) $(x \leftrightarrow y) \leftrightarrow z$; d) $x \rightarrow (y \rightarrow x)$.

3.2.4. Быстрое вычисление полинома Жегалкина

По значениям $f[0], \dots, f[2^n - 1]$ булевой функции нетрудно вычислить коэффициенты соответствующего полинома Жегалкина. Программа вычислений выглядит так:

```

for( $k = 0; k < 2^n; k++$ ) (3.29)
     $a[k] = f[k];$ 
for( $i = 1; i <= n; i++$ )
{
    for( $s = 0; s < 2^{n-i}; s++$ )
    {
        for( $l = 0; l < 2^{i-1}; l++$ )
             $a[2^i s + 2^{i-1} + l] = a[2^i s + l] \oplus a[2^i s + 2^i - 1 + l];$ 
    }
}

```

После работы схемы (3.29) в массиве $a[0 : 2^n - 1]$ будут находиться коэффициенты полинома Жегалкина.

Разберёмся в схеме (3.29) более детально. При $i = 1 (l = 0)$ вычисляем

```

for( $s = 0; s < 2^n - 1; s++$ )
     $a[2s + 1] = a[2s] \oplus a[2s + 1];$ 

```

При $i = 2$

```

for( $s = 0; s < 2^{n-2} - 1; s++$ )
{
    for( $l = 0; l < 2; l++$ )
         $a[4s + 2 + l] = a[4s + l] \oplus a[4s + 2 + l];$ 
}

```

и так далее. При $i = n(s = 0)$

```

for( $l = 0; l < 2^{n-1} - 1; l++$ )
     $a[2^{n-1} + l] = a[l] \oplus a[2^{n-1} + l];$ 

```

Схема (3.29) называется *быстрым алгоритмом вычисления коэффициентов полинома Жегалкина*. В ней используется только операция сложения по модулю 2 в количестве

$$\sum_{i=1}^n 2^{n-i} 2^{i-1} = n 2^{n-1} = \frac{1}{2} N \log_2 N$$

операций.

В таблице (3.29) приведены результаты вычислений по схеме (3.29) для конкретной булевой функции трёх переменных. В последнем столбце перечислены все мономы от трёх переменных, а в предпоследнем – найденные значения коэффициентов при этих мономах.

Таблица 3.29

№	x_1	x_2	x_3	$f(x_1, x_2, x_3)$	$i = 1$	$i = 2$	$i = 3$	мономы
0	0	0	0	0	1	1	1	1
1	1	0	0	0	1	1	1	x_1
2	0	1	0	1	1	0	0	x_2
3	1	1	0	0	1	0	0	$x_1 x_2$
4	0	0	1	0	0	0	1	x_3
5	1	0	1	1	1	1	0	$x_1 x_3$
6	0	1	1	1	1	1	1	$x_2 x_3$
7	1	1	1	0	1	0	0	$x_1 x_2 x_3$

Получаем представление

$$f(x_1, x_2, x_3) = 1 \oplus x_1 \oplus x_3 \oplus x_2 x_3 \quad (3.30)$$

Рассмотрим обратную задачу – вычисление значений полинома Жегалкина. Она решается просто, так как соотношения (3.29) обратимы. Удивительным является тот факт, что значения полинома Жегалкина будут найдены, если воспользоваться самой схемой (3.29), поменяв в ней местами идентификаторы f и a . В таблице (3.30) приведены результаты вычисления значений полинома (3.30) по схеме (3.29). В последнем столбце указаны значения полинома (3.30).

Таблица 3.30

№	x_1	x_2	x_3	a	$i = 1$	$i = 2$	$i = 3$
0	0	0	0	1	1	1	1
1	1	0	0	1	0	0	0
2	0	1	0	0	0	1	1
3	1	1	0	0	0	0	0
4	0	0	1	1	1	1	0
5	1	0	1	0	1	1	1
6	0	1	1	1	1	0	1
7	1	1	1	0	1	0	0

3.2.5. Булевые функции в криптографии

В традиционных системах шифрования, переводящих открытое сообщение в зашифрованное с помощью секретного ключа, важную роль играют булевые функции. Рассмотрим схему поточного шифрования, когда каждый поступающий символ тут же преобразуется в символ шифртекста.

Исходный текст $\{x_i\}$, ключ $\{y_i\}$, шифртекст $\{z_i\}$ – бинарные наборы одинаковой длины. Схема поточного шифрования следующая:

$$z_i = x_i \oplus y_i, \quad i \in 0 : n$$

При дешифровании схема та же, что и при шифровании, только исходный текст и шифртекст меняются местами. Шифр носит название *шифр Вернама* или *одноразовый шифрблокнот* (*one time pad*).

Историческая справка. Шифр назван в честь телеграфиста Гильберта Вернама, построившего в 1917 году телеграфный аппарат, который выполнял эту операцию автоматически – надо было только подать на него ленту с ключом. Не будучи шифровальщиком, тем не менее, Вернам верно заметил важное свойство своего шифра – каждая лента должна использоваться только один раз и после этого уничтожаться.

В американском кинофильме 2001 года «Пароль „Рыба-меч“» хакер (которого играет Хью Джекмен) взламывает именно шифр Вернама.

Обычно дважды такой шифр не используют: при сложении двух шифртекстов, соответствующих одному ключу, получается сумма исходных текстов, что даёт много информации об исходных текстах и даже часто позволяет их прочесть. На практике шифровальщик при личной встрече снабжается блокнотом, каждая страница которого содержит ключ. Такой же блокнот есть и у принимающей стороны. Использованные страницы уничтожаются.

В 1949 году Клод Шенон опубликовал работу, в которой доказал, что при совершенно случайном ключе, используемом один раз, шифр Вернама является абсолютно стойкой криптосистемой, то есть перехват шифртекста не даёт никакой информации о переданном сообщении. Других шифров с этим свойством не существует. Это по сути означает, что шифр Вернама является самой безопасной криптосистемой из всех возможных.

Пример 3.52.

Таблица 3.31

<i>a</i> (шифрование)						<i>b</i> (десифрование)					
Исходное	0	1	0	1	1	1	Шифрованное	0	1	1	0
Ключ	0	0	1	1	1	0	Ключ	0	0	1	1
Шифрованное	0	1	1	0	0	1	Исходное	0	1	0	1

3.2.6. Самодвойственные функции

Перенесем принцип двойственности логики высказываний на булевы функции. Для определения воспользуемся соотношением (3.6).

Определение 3.46. Булеву функцию $g(x_1, \dots, x_n)$ называют *двойственной* к $f(x_1, \dots, x_n)$, если выполняется равенство

$$g(x_1, \dots, x_n) = \overline{f(\bar{x}_1, \dots, \bar{x}_n)}.$$

Пример 3.53. Функция xy двойственна $x \vee y$, и наоборот. Это следует из равенств (3.24) и (3.23).

Функция $x \rightarrow y$ двойственна функции $\bar{x}y$, поскольку

$$\bar{x} \rightarrow \bar{y} \stackrel{14}{=} \overline{\bar{x} \vee \bar{y}} \stackrel{13}{=} \overline{x \vee y} \stackrel{12}{=} \bar{x}y.$$

Двойственной к функции $\max(x_1, x_2, \dots, x_n) = x_1 \vee x_2 \vee \dots \vee x_n$ является функция

$$\overline{x_1 \vee x_2 \vee \dots \vee x_n} = x_1 x_2 \cdots x_n = \min(x_1, x_2, \dots, x_n).$$

Задача 3.17. Определите, является ли функция f двойственной к функции g :

$$a) f = x \leftrightarrow y, g = \bar{x}y \vee x\bar{y}; \quad b) f = (x \oplus y)z, g = (x \oplus y)(z \oplus 1);$$

Определение 3.47. Булеву функцию $f(x_1, \dots, x_n)$ называют *самодвойственной*, если выполняется равенство

$$f(x_1, \dots, x_n) = \overline{f(\bar{x}_1, \dots, \bar{x}_n)}. \quad (3.31)$$

Другими словами, функция самодвойственна, если она совпадает со своей двойственной. Заметим, что равенство (3.31) для определения самодвойственности равносильно равенству

$$\overline{f(x_1, \dots, x_n)} = f(\bar{x}_1, \dots, \bar{x}_n). \quad (3.32)$$

Пример 3.54. Покажем, что $f(x_1, x_2, x_3) = x_1x_2 \vee x_1x_3 \vee x_2x_3$ есть самодвойственная функция. Имеем

$$\begin{aligned} \overline{f(\bar{x}_1, \bar{x}_2, \bar{x}_3)} &= \overline{\bar{x}_1 \bar{x}_2 \vee \bar{x}_1 \bar{x}_3 \vee \bar{x}_2 \bar{x}_3} \stackrel{11}{=} \overline{x_1 \vee x_2} \vee \overline{x_1 \vee x_3} \vee \overline{x_2 \vee x_3} \stackrel{12,13}{=} \\ &\stackrel{12,13}{=} (x_1 \vee x_2)(x_1 \vee x_3)(x_2 \vee x_3) \stackrel{7,1}{=} x_1x_2 \vee x_1x_3 \vee x_2x_3 = \\ &= f(x_1, x_2, x_3). \end{aligned}$$

Из последнего примера видно, что определять самодвойственность булевой функции на основании определения иногда достаточно сложно. Эта задача легко решается с помощью следующей теоремы.

Теорема 3.24. При лексикографическом упорядочивании строк таблицы значений самодвойственной функции последний столбец (значения функции) антисимметричен относительно середины таблицы.

Доказательство. Пусть $(\alpha_1, \dots, \alpha_n)$ — произвольная строка таблицы значений самодвойственной булевой функции, а $N(\alpha)$ ее номер, определенный согласно (3.3) в теореме 3.1. Рассмотрим соответствующую двойственную строку $(\bar{\alpha}_1, \dots, \bar{\alpha}_n)$. Номер этой строки обозначим через $\bar{N}(\alpha)$. Легко показать, что

$$\bar{N}(\alpha) = 2^n - 1 - N(\alpha).$$

Следствие 3.4. Самодвойственная функция равна 1 на половине наборов своей таблицы значений.

Следствие 3.5. Самодвойственную функцию полностью определяет первая (вторая) половина столбца значений таблицы истинности.

Упражнение 3.10. Докажите, что среди булевых функций двух переменных нет самодвойственных, существенно зависящих от обеих переменных.

Класс всех самодвойственных функций обозначим буквой S .

Лемма 3.14. Класс S не пуст и не совпадает со всем множеством булевых функций B_F .

Доказательство. Функции $x, x_1x_2 \vee x_1x_3 \vee x_2x_3 \in S$, т. е. $S \neq \emptyset$, а функции $xy, x \rightarrow y, x \oplus y \notin S$, следовательно $S \neq B_F$. ■

Справедлива теорема.

Теорема 3.25. S — замкнутый класс.

Доказательство. Поскольку равенство (3.31) выполняется для всех значений переменных, класс S замкнут относительно операции переименования переменных.

Пусть функции $f(x_1, \dots, x_k), g_1(x_1, \dots, x_n), \dots, g_k(x_1, \dots, x_n)$ принадлежат S . Тогда, используя равенство (3.32), получаем, что

$$\begin{aligned} \overline{f(x_1, \dots, x_k)} &= f(\bar{x}_1, \dots, \bar{x}_k), \\ \overline{g_1(x_1, \dots, x_n)} &= g_1(\bar{x}_1, \dots, \bar{x}_n), \dots, \overline{g_k(x_1, \dots, x_n)} = g_k(\bar{x}_1, \dots, \bar{x}_n). \end{aligned}$$

Тогда, если $h(x_1, \dots, x_n) = f(g_1(x_1, \dots, x_n), \dots, g_k(x_1, \dots, x_n))$, то

$$\begin{aligned} h(\bar{x}_1, \dots, \bar{x}_n) &= f(g_1(\bar{x}_1, \dots, \bar{x}_n), \dots, g_k(\bar{x}_1, \dots, \bar{x}_n)) = \\ &= f(\overline{g_1(x_1, \dots, x_n)}, \dots, \overline{g_k(x_1, \dots, x_n)}) = \\ &= \overline{f(g_1(x_1, \dots, x_n), \dots, g_k(x_1, \dots, x_n))} = \overline{h(x_1, \dots, x_n)}. \end{aligned}$$

В силу равенства (3.32), $h(x_1, \dots, x_n)$ — самодвойственная функция. Следовательно, класс S замкнут относительно операции подстановки. ■

Упражнение 3.11. Обозначим через S^n множество всех самодвойственных булевых функций от n переменных. Докажите, что

$$|S^n| = 2^{2^{n-1}}.$$

Лемма 3.15 (о несамодвойственной функции). Пусть $f(x_1, \dots, x_n)$ — несамодвойственная функция. Тогда замыкание класса

$$K = \{f(x_1, \dots, x_n), \bar{x}\}$$

содержит константы 0, 1.

Доказательство. Так как $f(x_1, \dots, x_n)$ — несамодвойственная функция, существуют $a_1, \dots, a_n \in B$, такие, что

$$\overline{f(a_1, \dots, a_n)} \neq f(\bar{a}_1, \dots, \bar{a}_n).$$

Множество B содержит только два элемента. Поэтому из этого неравенства следует равенство

$$f(a_1, \dots, a_n) = f(\bar{a}_1, \dots, \bar{a}_n).$$

Для удобства обозначений (не умаляя общности рассуждений) предположим, что $a_1 = \dots = a_k = 0$, $a_{k+1} = \dots = a_n = 1$. Тогда последнее равенство можно записать так:

$$f(\underbrace{0, \dots, 0}_k, 1, \dots, 1) = f(\underbrace{1, \dots, 1}_k, 0, \dots, 0),$$

Рассмотрим функцию $g(x) = f(x, \dots, x, \bar{x}, \dots, \bar{x})$. Заметим, что $g(x)$ принадлежит $[K]$. Имеем равенства

$$\begin{aligned} g(0) &= f(\underbrace{0, \dots, 0}_k, \bar{0}, \dots, \bar{0}) = f(\underbrace{0, \dots, 0}_k, 1, \dots, 1) = \\ &= f(\underbrace{1, \dots, 1}_k, 0, \dots, 0) = f(\underbrace{1, \dots, 1}_k, \bar{1}, \dots, \bar{1}) = g(1). \end{aligned}$$

Следовательно, $g(x)$ — одна из констант, принадлежащая $[K]$. Поскольку K содержит отрицание, то и другая константа принадлежит $[K]$. ■

Пример 3.55. Является ли функция $f(x_1, x_2) = x_2(x_2 \rightarrow x_1)$ самодвойственной? Для получения ответа на вопрос согласно теореме 3.24 составим ее таблицу значений (табл. 3.32).

Таблица 3.32

x_1	x_2	$f(x_1, x_2)$	x_1	x_2	$f(x_1, x_2)$
0	0	0	1	0	0
0	1	0	1	0	0

Заметим, что столбец значений функции не является антисимметричным относительно середины: равенство (3.31) не выполняется, например, при $x_1 = 0, x_2 = 1$. Следовательно, функция $f(x_1, x_2)$ самодвойственной не является.

Задача 3.18. Определить, можно ли получить константы из функции $f(x_1, x_2, x_3) = x_3 \rightarrow x_1 x_2$ и функции отрицания.

Составим ее таблицу значений (табл. 3.33). Заметим, что $f \notin S$, т. к. нарушается условие самодвойственности на первом и последнем наборах: $f(0, 0, 0) = f(1, 1, 1) = 1$. Таким образом

$$f(x, x, x) = x \rightarrow xx = \bar{x} \vee x = 1; \quad f(\bar{x}, \bar{x}, \bar{x}) = \bar{x} \rightarrow \bar{x}\bar{x} = x \vee \bar{x} = 1.$$

Используя отрицание, получаем вторую константу.

Таблица 3.33

x_1	x_2	x_3	$f(x_1, x_2, x_3)$	x_1	x_2	x_3	$f(x_1, x_2, x_3)$
0	0	0	1	1	0	0	1
0	0	1	0	1	0	1	0
0	1	0	1	1	1	0	1
0	1	1	0	1	1	1	1

Задача 3.19. Определите самодвойственны или нет следующие функции:

$$f(x_1, x_2, x_3) = (x_1 \oplus x_2)(x_2 \oplus x_3); \quad g(x_1, x_2, x_3) = (x_1 \vee x_2)(x_1 \vee x_3)(x_2 \vee x_3).$$

Упражнение 3.12. Докажите, что булева функция четности

$$p_n(x_1, x_2, \dots, x_n) = x_1 \oplus x_2 \oplus \dots \oplus x_n$$

является самодвойственной тогда и только тогда, когда n нечетно.

Упражнение 3.13. Определите, какие из линейных функций являются самодвойственными.

3.2.7. Монотонные функции

Введем отношение частичного нестрогого порядка на множестве булевых наборов длины n .

Определение 3.48. Будем считать, что

$$\alpha = (\alpha_1, \dots, \alpha_n) \leqslant \beta = (\beta_1, \dots, \beta_n), \text{ если } \alpha_1 \leqslant \beta_1, \dots, \alpha_n \leqslant \beta_n, \alpha_i, \beta_i \in B.$$

Например, $(1, 0, 0, 1) \leqslant (1, 0, 1, 1)$, $(1, 0, 1, 0) \leqslant (1, 0, 1, 1)$. В то же время неверно, что $(1, 0, 0, 1) \leqslant (1, 0, 1, 0)$.

Определение 3.49. Булеву функцию $f(x_1, \dots, x_n)$ называют *монотонной*, если для любых двух наборов $\alpha = (\alpha_1, \dots, \alpha_n)$, $\beta = (\beta_1, \dots, \beta_n)$ из условия $\alpha \leqslant \beta$ следует, что $f(\alpha) \leqslant f(\beta)$.

Класс всех монотонных функций обозначим буквой M .

Лемма 3.16. Класс M не пуст и не совпадает со всем множеством булевых функций B_F .

Доказательство. Функции $1, 0, xy, x \vee y \in M$, т. е. $M \neq \emptyset$, а функции $\bar{x}, x \rightarrow y, x \oplus y \notin M$, следовательно $M \neq B_F$. ■

Проверка монотонности булевой функции по определению достаточно трудоемка. Используем для этого рассмотренный ранее подход графического представления упорядоченных множеств – диаграмму Хассе (раздел 2.3.7).

Пример 3.56. Проверим монотонность функции

$$f(x_1, x_2, x_3) = x_1 \oplus x_2 x_3.$$

Составим таблицу значений функции (табл. 3.34).

Таблица 3.34

x_1	x_2	x_3	$x_1 \oplus x_2 x_3$	x_1	x_2	x_3	$x_1 \oplus x_2 x_3$
0	0	0	0	1	0	0	1
0	0	1	0	1	0	1	1
0	1	0	0	1	1	0	1
0	1	1	1	1	1	1	0

Начертим диаграмму Хассе частично упорядоченного множества булевых наборов длины 3 (рис. 3.13).

В вершинах этого графа пропишем значения исследуемой функции согласно табл. 3.34, а справа от вершин – соответствующие наборы. Монотонность функции означает, что при движении по любому ориентированному пути в диаграмме Хассе, значение функции не должно увеличиваться.

Это свойство нарушается при переходе от вершины $(1, 1, 1)$, к вершинам $(1, 1, 0)$, $(1, 0, 1)$ и $(0, 1, 1)$.

Соответствующие дуги в графе на рис. 3.13 показаны штриховыми линиями.

Таким образом функция f не монотонна.

Задача 3.20. Определите монотонность следующих функций:

$$a) xy \oplus y \oplus x; \quad b) x \rightarrow (x \rightarrow y); \quad c) xy \vee xz \vee yz.$$

Упражнение 3.14. Докажите монотонность функций:

$$\max(x_1, \dots, x_n) \text{ и } \min(x_1, \dots, x_n),$$

определенных в примере 3.53.

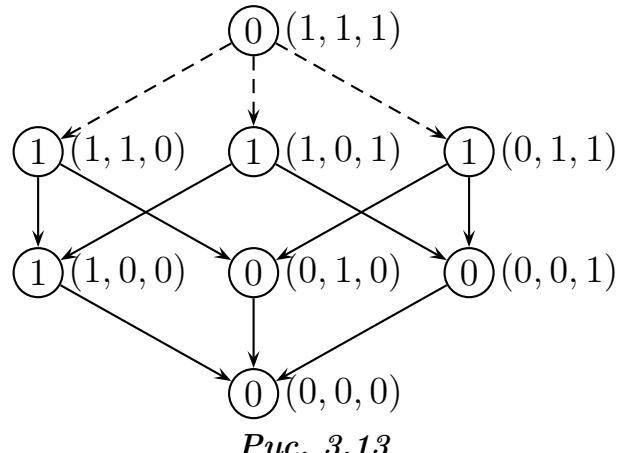


Рис. 3.13

Упражнение 3.15. Определите, какие из линейных функций являются монотонными.

Теорема 3.26. M — замкнутый класс.

Доказательство. Пусть $f(x_1, \dots, x_n) \in M$ и y_1, \dots, y_n — новые переменные. Возьмем два набора значений переменных y_1, \dots, y_n :

$$\alpha = (\alpha_1, \dots, \alpha_n) \leqslant \beta = (\beta_1, \dots, \beta_n). \quad (3.33)$$

Но эти векторы будут наборами значений переменных x_1, \dots, x_n , и поэтому выполняется неравенство $f(\alpha) \leqslant f(\beta)$. Это означает, что класс M замкнут относительно операции переименования переменных.

Пусть $f(x_1, \dots, x_k), g_1(x_1, \dots, x_n), \dots, g_n(x_1, \dots, x_n)$ — монотонные функции и α и β — два набора значений переменных x_1, \dots, x_n , удовлетворяющих (3.33). Рассмотрим подстановку

$$h(x_1, \dots, x_n) = f(g_1(x_1, \dots, x_n), \dots, g_n(x_1, \dots, x_n)).$$

Введем обозначения:

$$\gamma_1 = g_1(\alpha), \dots, \gamma_k = g_k(\alpha), \quad \zeta_1 = g_1(\beta), \dots, \zeta_k = g_k(\beta).$$

В силу монотонности функций g_1, \dots, g_k имеем, что $\gamma_1 \leqslant \zeta_1, \dots, \gamma_k \leqslant \zeta_k$. Тогда

$$\begin{aligned} h(\alpha_1, \dots, \alpha_n) &= f(g_1(\alpha_1, \dots, \alpha_n), \dots, g_k(\alpha_1, \dots, \alpha_n)) = f(\gamma_1, \dots, \gamma_k) \leqslant \\ &\leqslant f(\zeta_1, \dots, \zeta_k) = f(g_1(\beta_1, \dots, \beta_n), \dots, g_k(\beta_1, \dots, \beta_n)) = \\ &= h(\beta_1, \dots, \beta_n). \end{aligned}$$

Знак неравенства поставлен в силу монотонности функции f . Таким образом, доказано, что класс M замкнут относительно операции подстановки. ■

Замечание 3.30. В случаях изучения классов T_0, T_1, L, S достаточно легко можно было определить число функций класса от n переменных (см. упр. 3.6, 3.7, 3.9, 3.11). Вопрос о числе монотонных функций от n переменных не решен до настоящего времени. Существуют лишь оценки этого числа.

Упражнение 3.16. Докажите, что функция, двойственная монотонной, сама монотонна.

Упражнение 3.17. Докажите, что монотонными являются те и только те функции, которые или являются константами, или допускают представление в КНФ или ДНФ, не содержащей отрицаний.

Упражнение 3.18. Докажите, что монотонная функция, не сохраняющая ноль (единицу), равна тождественно единице (нулю).

Лемма 3.17 (о немонотонной функции). Пусть $f(x_1, \dots, x_n)$ — немонотонная функция. Тогда замыкание класса

$$K = \{f(x_1, \dots, x_n), 0, 1\}$$

содержит \bar{x} .

Доказательство. Поскольку $f(x_1, \dots, x_n)$ немонотонна, то существуют наборы значений переменных $\alpha = (\alpha_1, \dots, \alpha_n)$, $\beta = (\beta_1, \dots, \beta_n)$ такие, что $\alpha < \beta$, но $f(\alpha) > f(\beta)$. Неравенство $f(\alpha) > f(\beta)$ означает, что $f(\alpha) = 1$ и $f(\beta) = 0$. Пусть наборы α и β отличаются k компонентами ($k \geq 1$). Тогда в α на этих местах стоят нули, а в β — единицы. Заменив указанные компоненты по одной (по старшинству), получим цепочку наборов:

$$\alpha < \alpha^1 < \alpha^2 < \dots < \alpha^{k-1} < \beta,$$

где соседне наборы различаются только одной компонентой. Очевидно, что найдутся соседние наборы α^j , α^{j+1} , такие, что $f(\alpha^j) = 1$, $f(\alpha^{j+1}) = 0$. Пусть они различаются i -й компонентой. Тогда $\alpha_i^j = 0$, $\alpha_i^{j+1} = 1$, а остальные компоненты одинаковы. Введем функцию

$$g(x) = f(\alpha_1^j, \alpha_2^j, \dots, \alpha_{i-1}^j, x, \alpha_{i+1}^j, \alpha_{i+2}^j, \dots, \alpha_n^j).$$

По условиям леммы $g(x) \in [K]$. Тогда

$$g(0) = g(\alpha_i^j) = f(\alpha^j) = 1, \quad g(1) = g(\alpha_i^{j+1}) = f(\alpha^{j+1}) = 0,$$

т. е. $g(x) = \bar{x}$. ■

Пример 3.57. Построим функцию отрицания из немонотонной функции и констант. Пусть функция $f(x_1, x_2, x_3)$ задана своей таблицей значений (табл. 3.35).

Таблица 3.35

x_1	x_2	x_3	$f(x_1, x_2, x_3)$	x_1	x_2	x_3	$f(x_1, x_2, x_3)$
0	0	0	0	0	0	1	1
0	1	0	0	0	1	1	1
1	0	0	1	1	0	1	0
1	1	0	0	1	1	1	1

Очевидно, что функция $f(x_1, x_2, x_3)$ не является монотонной. Действительно, монотонность нарушается на наборах $(0, 0, 1)$ и $(1, 0, 1)$. Эти наборы различаются только первой компонентой. Тогда, согласно доказательству леммы 3.17 определяем функцию $g(x) = f(x, 0, 1)$. Нетрудно убедиться в том, что $g(0) = 1$, а $g(1) = 0$, т. е. функция $g(x)$ является отрицанием.

Задача 3.21. Выяснить, можно ли построить функцию \bar{x} из функции $f(x_1, x_2, x_3) = x_1 \oplus x_2 x_3$ и констант.

Функция f немонотонна, т. к. $f(1, 0, 0) = 1$, $f(1, 1, 1) = 0$, тогда по лемме о немонотонной функции из этой функции можно получить функцию отрицания.

Выберем пару соседних наборов, на которых нарушено условие монотонности. Имеем, $f(1, 1, 0) = 1$, $f(1, 1, 1) = 0$. Выбирая соответствующую замену переменных, получаем, что

$$f(1, 1, x) = 1 \oplus x = \bar{x}.$$

Замечание 3.31. Монотонные булевы функции были определены, как *монотонно неубывающие* (определение 3.49). При этом определении класс M является замкнутым (см. теорему 3.26). Возникает вопрос: можно ли было определить этот класс, как множество *монотонно невозрастающих* булевых функций и при этом сохранить свойство замкнутости?

Отрицательный ответ на этот вопрос дает рассмотрение функции $f(x) = \bar{x}$. Отрицание \bar{x} — монотонно убывающая функция. Подстановка функции в себя $f(f(x)) = \bar{\bar{x}} = x$ монотонно возрастает.

3.2.8. Теорема Поста

Установим необходимый и достаточный критерий полноты класса булевых функций, который называют теоремой Поста.

Определение 3.50. Классы T_0, T_1, L, S, M называют *основными замкнутыми классами (классами Поста)* булевых функций.

Теорема 3.27 (теорема Поста). Класс булевых функций K является полным тогда и только тогда, когда он не содержится ни в одном из основных замкнутых классов.

Доказательство. Установим необходимость. Пусть K — полный класс. Предположим, что K содержится в одном из основных замкнутых классов, например, $K \subseteq T_0$. Тогда из леммы 3.11 следует, что класс T_0 также полный, т. е. $[T_0] = B_F$.

Но, согласно теореме 3.19 справедливо равенство $T_0 = [T_0]$, а по лемме 3.9 имеем $T_0 \neq B_F$. Противоречие показывает, что K не содержится в T_0 . Аналогично доказывается, что K не содержится и в других основных замкнутых классах.

Перейдем к доказательству достаточности. Пусть K не содержится ни в одном из классов T_0, T_1, S, M и L . Тогда класс K содержит функции

$$f_0 \notin T_0, f_1 \notin T_1, f_2 \notin S, f_3 \notin M, f_4 \notin L. \quad (3.34)$$

Докажем, что $[K]$ содержит \bar{x} и xy . Так как $f_0 \notin T_0$, то $f_0(0, \dots, 0) = 1$. Относительно значения $f_0(1, \dots, 1)$ рассмотрим два случая.

1. Пусть $f_0(1, \dots, 1) = 0$. Рассмотрим функцию $g(x) = f_0(x, \dots, x)$. Тогда $g(0) = f_0(0, \dots, 0) = 1$ и $g(1) = f_0(1, \dots, 1) = 0$, т. е. $g(x) = \bar{x}$. Это означает, что замыкание класса K содержит отрицание. Классу K принадлежит несамодвойственная функция f_2 . По лемме 3.15 о несамодвойственной функции замыкание класса K содержит константы. Поскольку K содержит нелинейную функцию f_4 , то в силу леммы 3.13 о нелинейной функции $[K]$ содержит произведение xy . Мы доказали, что в первом случае $[K]$ содержит \bar{x} и xy .

2. Предположим, что $f_0(1, \dots, 1) = 1$. Тогда, если $g(x) = f_0(x, \dots, x)$ и $h(x) = f_1(g(x), \dots, g(x))$, то

$$g(0) = f_0(0, \dots, 0) = 1 = f_0(1, \dots, 1) = g(1),$$

$$h(0) = f_1(g(0), \dots, g(0)) = f_1(1, \dots, 1) = 0 = f_1(g(1), \dots, g(1)) = h(1),$$

поскольку f_1 не сохраняет 1. Это означает, что константы принадлежат $[K]$.

Так как K содержит немонотонную функцию, то по лемме 3.17 о немонотонной функции, $[K]$ принадлежит отрицание. Далее, f_4 — нелинейная функция из K , поэтому по лемме о нелинейной функции $[K]$ содержит и xy .

Итак, в обоих случаях, замыканию $[K]$ принадлежат \bar{x} и xy . Поскольку по теореме 3.21 $\{\bar{x}, xy\}$ — полный класс, то K , согласно лемме 3.9, также полный класс. ■

Приведенное доказательство теоремы Поста иллюстрирует рис. 3.14. Непрерывными линиями показан первый, рассмотренный при доказательстве случай, штриховые линии соответствуют второму случаю.

Следствие 3.6. Каждый полный класс содержит полный подкласс, состоящий не более чем из четырех функций.

Доказательство. Пусть K — полный класс. Тогда по теореме 3.27 класс K не содержится ни в одном из основных замкнутых классов. Согласно доказательству теоремы 3.27 в K найдутся функции f_0, f_1, f_2, f_3, f_4 , удовлетворяющие (3.34).

При рассмотрении первого случая были использованы три функции f_0, f_2 и f_4 , т. е. $\{f_0, f_2, f_4\}$ — полный подкласс класса K . Во втором случае использованы четыре функции: f_0, f_1, f_3 и f_4 . Тогда $\{f_0, f_2, f_3, f_4\}$ — полный подкласс класса K . Итак, в обоих случаях в K найдется полный подкласс, содержащий не более четырех функций. ■

Пример 3.58. Существует полный подкласс, состоящий точно из четырех функций. В качестве примера рассмотрим класс

$$K = \{0, 1, xy, x \oplus y \oplus z\}.$$

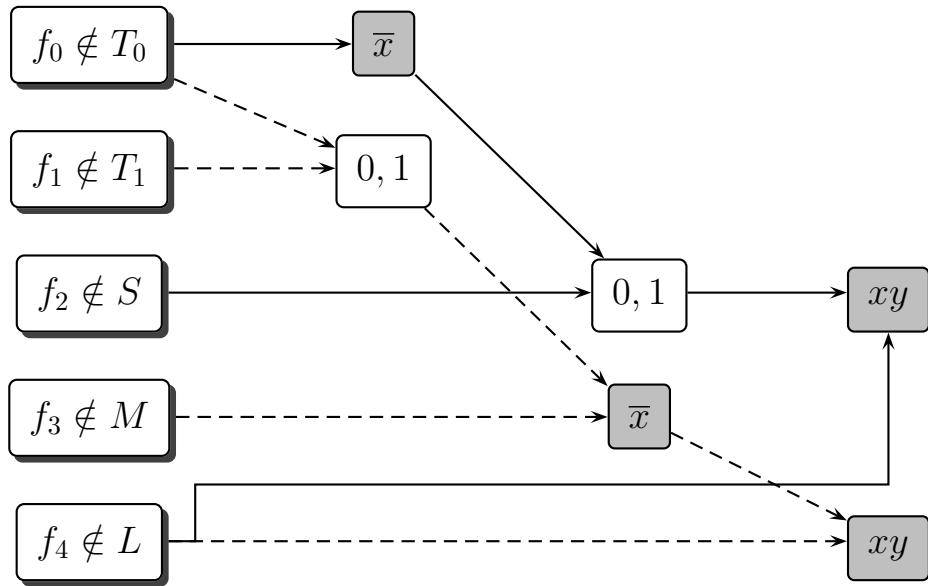


Рис. 3.14

Очевидно, что $0 \notin T_1$, $1 \notin T_0$. Функция xy не является ни самодвойственной, ни линейной. Нетрудно проверить, что $x \oplus y \oplus z$ — немонотонная функция. Следовательно, по теореме Поста K — полный класс.

В то же время, любой его собственный подкласс полным не является. Действительно, если удалить функцию 0, то оставшиеся функции будут сохранять 1, если удалить 1, то будут сохранять 0. Класс функций K без функции xy состоит из линейных функций, а без последней функции — из монотонных.

Определение 3.51. Замкнутый класс K называют *предполным* если K — неполный класс, но для любой функции $f \notin K$ класс $K \cup \{f\}$ является полным.

Теорема 3.28. Предполными являются классы T_0 , T_1 , S , M и L и только они.

Доказательство. Необходимость докажем на примере класса T_0 . В нем содержатся несамодвойственная и нелинейная функция xy , немонотонная функция $x \oplus y$ и функция 0, не сохраняющая 1. Пусть $f \notin T_0$. Рассмотрим класс $\widehat{T}_0 = \{0, xy, x \oplus y\} \cup \{f\}$. Легко показать, что \widehat{T}_0 не содержится ни в одном из основных замкнутых классов и, следовательно, является полным.

Поскольку $\widehat{T}_0 \subseteq T_0 \cup \{f\}$, то по лемме 3.11 класс $T_0 \cup \{f\}$ также является полным.

Докажем достаточность. Пусть K — предполный класс. Поскольку класс K не является полным, то K содержится в одном из основных замкнутых классов. Для определенности предположим, что $K \subseteq T_0$. Если $K \neq T_0$, то существует функция f , такая, что $f \notin K$ и $f \in T_0$. Тогда

$K \cup \{f\} \subseteq T_0$ и класс $K \cup \{f\}$ не может быть полным. Следовательно, $K = T_0$. \blacksquare

Упражнение 3.19. Булеву функцию называют *полной (обобщенной функцией Шеффера)*, если класс, единственным элементом которого является эта функция, будет полным. Докажите, что штрих Шеффера и стрелка Пирса — единственные полные функции среди всех булевых функций двух переменных.

Определение 3.52. Систему функций Ψ называют *базисом* замкнутого класса K , если $[\Psi] = K$, но замыкание любой собственной подсистемы Ψ уже не совпадает с K .

Пример 3.59. Рассмотрим полный класс $K_4 = \{xy, x \oplus y, 1, 0\}$ из теоремы 3.21. Согласно определению 3.44 это полиномы Жегалкина. Но это не базис, так как $1 \oplus 1 = 0$ и базис K_4 состоит из трех функций $\{xy, x \oplus y, 1\}$.

Упражнение 3.20. Докажите, что система $\Psi = \{0, 1, xy, x \vee y\}$ есть базис класса M .

Упражнение 3.21. Докажите, что если система булевых функций полна, то будет полной и система, состоящая из двойственных функций.

Применим теорему Поста для исследования полноты системы булевых функций $K = \{f_1, \dots, f_n\}$. Для проверки условий теоремы составим таблицу, которую называют *таблицей Поста* (табл. 3.36).

Таблица 3.36

	T_0	T_1	L	S	M
f_1					
\dots	\dots	\dots	\dots	\dots	\dots
f_n					

В клетки таблицы ставится знаки: «+», если функция, стоящая в строке принадлежит классу Поста, стоящему в данном столбце, «-» — в противном случае. Для полноты системы K необходимо и достаточно (в силу теоремы Поста), чтобы в каждом столбце стоял хотя бы один знак «-».

Пример 3.60. Исследовать на полноту системы булевых функций $K_1 = \{x \oplus y, xy, 1, 0\}$ и $K_2 = \{x \oplus y \oplus z, xy, x \rightarrow y\}$. Составим таблицы Поста для системы K_1 (табл. 3.37, а) и системы K_2 (табл. 3.37, б).

Система K_1 полна, но не является базисом. Базис образует подсистема $\{x \oplus y, xy, 1\}$.

Таблица 3.37

	<i>Таблица 3.37</i>										
	<i>a</i>	<i>b</i>									
	<i>T</i> ₀	<i>T</i> ₁	<i>L</i>	<i>S</i>	<i>M</i>		<i>T</i> ₀	<i>T</i> ₁	<i>L</i>	<i>S</i>	<i>M</i>
<i>x</i> \oplus <i>y</i>	+	-	+	-	-		+	+	+	+	-
<i>xy</i>	+	+	-	-	+		+	+	-	-	+
1	-	+	+	-	+		-	+	-	-	-
0	+	-	+	-	+						

Система K_2 не является полной, так как $K_2 \subset T_1$. Для получения полной системы нужно добавить любую функцию не сохраняющую 1, (например, \bar{x}).

Пример 3.61. Покажем полноту системы функций

$$G = \{f(x, y, z) = xy \rightarrow z, g(x, y) = x \oplus y\}.$$

Проиллюстрируем поэтапное доказательство теоремы Поста, т. е. выразим константы, отрицание и конъюнкцию через функции системы G .

Построим таблицу Поста (табл. 3.38) для системы функций G . Полином Жегалкина для $f(x, y, z)$ имеет вид

$$xy \rightarrow z = xyz \oplus xy \oplus 1. \quad (3.35)$$

Принадлежность функций f, g остальным классам Поста достаточно очевидна.

Таблица 3.38

	<i>T</i> ₀	<i>T</i> ₁	<i>L</i>	<i>S</i>	<i>M</i>
<i>xy</i> \rightarrow <i>z</i>	-	+	-	-	-
<i>x</i> \oplus <i>y</i>	+	-	+	-	-

По теореме Поста делаем вывод, что система функций G полна.

1. Получение констант. Имеем,

$$f(0, 0, 0) = f(1, 1, 1) = 1 : f(x, x, x) = xx \rightarrow x = 1, \quad (3.36)$$

Вторую константу 0 получим из функции

$$g(x, y) \notin T_1 : g(f(x, x, x), f(x, x, x)) = (xx \rightarrow x) \oplus (xx \rightarrow x) = 0. \quad (3.37)$$

2. Получение отрицания. По лемме о немонотонной функции

$$g(1, x) = 1 \oplus x = \bar{x}. \quad (3.38)$$

Подставим $f(x, x, x)$ вместо 1 в (3.38), согласно (3.36) получим

$$g(f(x, x, x), x) = (xx \rightarrow x) \oplus x = \bar{x}.$$

3. Получение конъюнкции. По лемме о нелинейной функции, используя (3.35) для функции f , имеем

$$xy = \overline{f(x, y, 0)}.$$

Далее, используя (3.36), (3.37) и (3.38) получаем цепочку равенств:

$$\begin{aligned} xy &= f(x, y, g(f(x, x, x), f(x, x, x))) \oplus 1 = \\ &= (xy \rightarrow ((xx \rightarrow x) \oplus (xx \rightarrow x))) \oplus (xx \rightarrow x) \end{aligned}$$

Таким образом, константы, отрицание и конъюнкцию выразили через функции системы G .

Задача 3.22. Выразить функцию $h = x \downarrow y$ через функции системы G из примера 3.61.

Выразим функцию h через конъюнкцию и отрицание: $h = \overline{xy}$. Используя выражения для отрицания и конъюнкции через функции системы G , которые были получены в примере 3.61, получаем

$$\begin{aligned} h &= g(f(x, x, x), f(g(f(x, x, x), x), y, g(f(x, x, x), f(x, x, x)))) = \\ &= (xx \rightarrow x) \oplus (((xx \rightarrow x) \oplus x)y \rightarrow ((xx \rightarrow x) \oplus (xx \rightarrow x))). \end{aligned}$$

Задача 3.23. Значения булевых функций $f(x_1, x_2, x_3)$ и $g(x_1, x_2, x_3)$ заданы табл. 3.39. Исследовать на полноту систему булевых функций $\{f, g\}$, в случае полноты найти все базисы данной системы.

Таблица 3.39

x_1	x_2	x_3	$f(x_1, x_2, x_3)$	$g(x_1, x_2, x_3)$
0	0	0	1	0
0	0	1	0	1
0	1	0	1	0
0	1	1	1	1
1	0	0	1	0
1	0	1	1	1
1	1	0	0	0
1	1	1	1	0

Составим таблицу Поста для системы $\{f, g\}$ (табл. 3.40).

Таким образом, система $\{f, g\}$ — полна и является базисом. Ни одну из функций нельзя удалить из системы, не теряя при этом свойство полноты.

Таблица 3.40

	T_0	T_1	L	S	M
$f(x_1, x_2, x_3)$	—	+	—	—	—
$g(x_1, x_2, x_3)$	+	—	—	—	—

Задача 3.24. Исследовать на полноту систему булевых функций

$$\begin{cases} f_1(x, y) = x \oplus y, \\ f_2(x, y, z) = xy \oplus z, \\ f_3(x, y, z) = x \oplus y \oplus z \oplus 1, \\ f_4(x, y, z) = xy \oplus yz \oplus xz \end{cases}$$

и в случае полноты найти все базисы данной системы.

Составим таблицу Поста для данного случая (табл. 3.41).

Таблица 3.41

	T_0	T_1	L	S	M
f_1	+	—	+	—	—
f_2	+	—	—	—	—
f_3	—	—	+	+	—
f_4	+	+	—	+	+

Из анализа таблицы получаем, что система полна и имеет два собственных базиса: $\{f_2, f_3\}$ и $\{f_1, f_3, f_4\}$.

Задача 3.25. Будут ли полными следующие системы функций:

a) $K_1 = \{x \oplus y, x \vee y \vee \bar{z}\}$; б) $K_2 = \{\bar{x}, xy \vee xz \vee yz\}$.

3.2.9. Разложение Шеннона

Определение 3.53. *Разложением Шеннона*¹ или *декомпозицией Шеннона* по переменной x_i называется метод представления булевой функции от n переменных в виде суммы двух подфункций от $(n - 1)$ остальных переменных:

¹Клод Элвуд Шеннон (Shannon) (1916 — 2001) — американский инженер и математик. Является основателем теории информации, нашедшей применение в современных высокотехнологических системах связи. Шеннон внес огромный вклад в теорию вероятностных схем, теорию автоматов и теорию систем управления — области наук, входящие в понятие «кибернетика». В 1948 году предложил использовать слово *бит* для обозначения наименьшей единицы информации.

$$\begin{aligned} f(x_1, \dots, x_n) &= x_i f(x_1, \dots, x_{i-1}, 1, x_{i+1}, \dots, x_n) \vee \\ &\vee \overline{x_i} f(x_1, \dots, x_{i-1}, 0, x_{i+1}, \dots, x_n). \end{aligned} \quad (3.39)$$

Тождество (3.39) остается справедливым и при замене дизъюнкции на \oplus , так как оба слагаемых ортогональны – никогда не принимают истинное значение одновременно.

$$\begin{aligned} f(x_1, \dots, x_n) &= x_i f(x_1, \dots, x_{i-1}, 1, x_{i+1}, \dots, x_n) \oplus \\ &\oplus \overline{x_i} f(x_1, \dots, x_{i-1}, 0, x_{i+1}, \dots, x_n). \end{aligned}$$

Для каждой из оставшихся функций от меньшего числа переменных можно продолжить разложение (3.39) по одной из оставшихся переменных.

Функция $f(x_1, \dots, x_{i-1}, 1, x_{i+1}, \dots, x_n)$ в (3.39) называется положительным дополнением, а функция $f(x_1, \dots, x_{i-1}, 0, x_{i+1}, \dots, x_n)$ – отрицательным дополнением.

Замечание 3.32. Хотя этот метод часто приписывают Клоду Шеннону, но Буль доказал его гораздо раньше, а сама возможность такого разложения по любой из переменной непосредственно вытекает из возможности определения любой булевой функции с помощью таблицы истинности.

Разложение (3.39) по переменной x_i основано на том, что таблицу истинности для булевой функции от n переменных можно разбить на две части таким образом, чтобы в первой части оказались только те булевые наборы, на которых переменная x_i всегда принимает значение 1, а во второй части остались только наборы, на которых переменная x_i всегда принимает значение 0.

Пример 3.62. Получим разложение Шеннона для булевой функции $f(x_1, x_2, x_3)$, заданной лексикографически упорядоченным вектором значений:

$$f(x_1, x_2, x_3) = \{1, 1, 0, 1, 0, 1, 0, 1\}$$

по переменной x_1 .

Запишем булеву функцию $f(x_1, x_2, x_3)$ в виде СДНФ:

$$f(x_1, x_2, x_3) = \overline{x_1} \overline{x_2} \overline{x_3} \vee \overline{x_1} \overline{x_2} x_3 \vee \overline{x_1} x_2 x_3 \vee x_1 \overline{x_2} x_3 \vee x_1 x_2 x_3.$$

Из последнего равенства применяя дистрибутивность по x_1 и $\overline{x_1}$ получаем требуемое разложение:

$$\begin{aligned} f(x_1, x_2, x_3) &= x_1 (\overline{x_2} x_3 \vee x_2 x_3) \vee \overline{x_1} (\overline{x_2} \overline{x_3} \vee \overline{x_2} x_3 \vee x_2 x_3) = \\ &= x_1 f(1, x_2, x_3) \vee \overline{x_1} f(0, x_2, x_3). \end{aligned}$$

Аналогично выполняется разложение функции $f(x_1, x_2, x_3)$ по переменной x_2 или x_3 .

$$\begin{aligned} f(x_1, x_2, x_3) &= x_2(\overline{x_1}x_3 \vee x_1x_3) \vee \overline{x_2}(\overline{x_1}\overline{x_3} \vee \overline{x_1}x_3 \vee x_1x_3) = \\ &= x_3(\overline{x_1}x_2 \vee x_1\overline{x_2} \vee x_1x_2) \vee \overline{x_3}(\overline{x_1}\overline{x_2} \vee \overline{x_1}x_3). \end{aligned}$$

3.2.10. Бинарные диаграммы решений

В современных задачах, решаемых с применением булевой алгебры, применяются все более сложные булевые функции. Отсюда, возникает задача удобного представления булевой функции. Отсюда возникают следующие требования к такому представлению:

- 1) оно должно быть каноническим;
- 2) оно должно быть минимальным (по числу переменных);
- 3) этим представлением можно было легко манипулировать.

Из курса нам известно несколько канонических представлений булевой функции (например, ДНФ и КНФ, таблица истинности). Но данные представления перестали удовлетворять выше изложенным требованиям в связи со сложностью самих булевых функций. Рассмотрим новый достаточно эффективный метод представления булевой функции – бинарную диаграмму решений.

Основной идеей для создания такой структуры данных послужило рассмотренное ранее разложение Шеннона (3.39). Любую булеву функцию по одной из входных переменных можно разделить на две подфункции, называемых положительным и отрицательным дополнением, из которых по принципу *if-then-else* выбирается только одна подфункция в зависимости от значения входной переменной (по которой выполнялось разложение Шеннона). Представляя каждую такую подфункцию в виде поддерева и продолжая разложение по оставшимся входным переменным можно получить бинарную диаграмму решений.

Определение 3.54. *Бинарная диаграмма решений*¹ – это представление булевой функции в виде корневого ориентированного дерева.

Из каждого узла дерева идут по два ребра: одно соответствует нулевому значению соответствующей переменной (будем его изображать штриховой линией), а другое – единичному значению этой переменной (оно изображается сплошной линией). Таким образом, с каждой ветвью от корня

¹Binary Decision Diagrams, BDD.

до листа сопоставляется набор значений переменных, а листья помечаются соответствующими значениями функции.

Введем следующие обозначения.

Каждая его нетерминальная (не висячая) вершина графа v помечена переменной $index(v)$ (обычно, это переменные заголовка булевой функции — x_i) и имеет две исходящих дуги к двум сыновьям: левую — $left(v)$ (будем ее изображать штриховой линией) в случае, когда переменной x_i присваивается значение 0, и правую — $right(v)$ (показана сплошной линией) в случае, когда переменной x_i присваивается значение 1. Каждая терминальная (висячая) вершина графа помечена либо 0, либо 1.

Для заданной таблицы истинности значение функции, реализованной бинарной диаграммой решений, определяется с помощью продвижения по пути от корня к терминальной вершине по ветвям, соответствующим присваиваемым переменным значениям. В качестве значения функции берется метка найденной терминальной вершины.

Дуги на диаграмме обычно упорядочены таким образом, что значения терминальных вершин (слева направо) соответствуют их вхождениям в таблицу истинности (сверху вниз).

Обычно рассматривают *упорядоченные бинарные диаграммы решений*², где задан линейный порядок на множество переменных и требуется, чтобы для любой вершины u и любого ее преемника v , не являющегося терминальной вершиной, соответствующие им переменные были упорядочены по возрастанию индексов переменных булевой функции $f(x_1, \dots, x_n)$, например $x_1 < x_2 < \dots < x_n$.

Замечание 3.33. Дональд Кнут назвал бинарные диаграммы решений *одной из действительно фундаментальных структур данных, которые появились за последние двадцать пять лет*.

В настоящее время OBDD широко используются в системах автоматизированного проектирования (САПР) для синтеза логических схем, формальной верификации, задачах распознавания и классификации текстов.

На рис. 3.15 изображено представление булевой функции, заданное лексикографически упорядоченным вектором значений:

$$f(x_1, x_2, x_3) = \{0, 0, 0, 1, 0, 1, 0, 1\}.$$

Определяются три правила преобразования OBDD (редукции графа), не изменяющих значение представляемой булевой функции.

Удаление дублирующих терминальных переменных. Удаляются все терминальные вершины с заданной меткой (значением булевой функции), кроме одной, а все дуги, ведущие к удаленным вершинам, перенаправляются в оставшуюся вершину.

² В англоязычной литературе такое представление называется Ordinary Binary Decision Diagrams, или сокращенно OBDD.

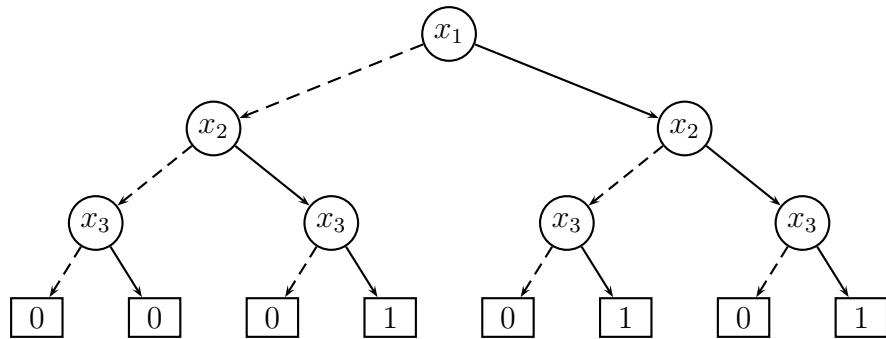


Рис. 3.15

Рис. 3.16 иллюстрирует применение первого правила к дереву решений, представленного на рис. 3.15. Применение правила преобразования сокращает количество терминальных вершин с 8 до 2.

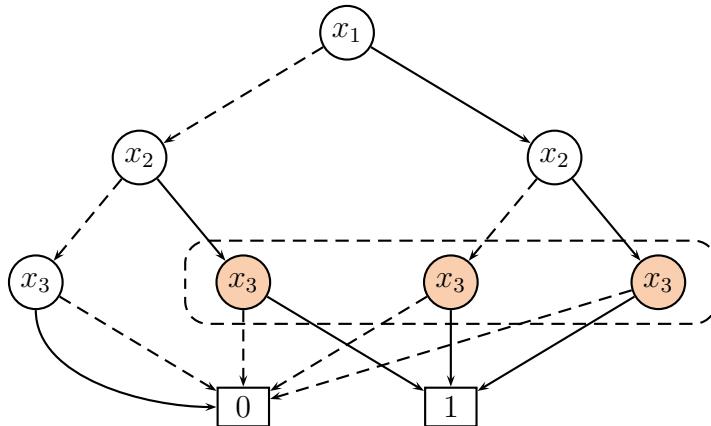


Рис. 3.16

Удаление дублирующих нетерминальных вершин. Если для нетерминальных вершин u и v выполняются следующие условия:

$$\text{index}(u) = \text{index}(v), \text{left}(u) = \text{left}(v), \text{right}(u) = \text{right}(v),$$

то одна из этих вершин удаляется, а все дуги, входящие в нее, перенаправляются во вторую вершину.

На рис. 3.17 *a* показано применение второго правила к дереву решений, после шага 1 (рис. 3.16). Применение правила удаляет две вершины с переменной x_3 (три вершины с меткой x_3 , помеченные цветом на рис. 3.16 сливаются в одну) и дуги к терминальным вершинам с метками 0 и 1.

Удаление избыточных проверок. Если для нетерминальной вершины v верно $\text{left}(v) = \text{right}(v)$, то вершина v удаляется, а все дуги, входящие в нее, перенаправляются в $\text{left}(v)$.

На рис. 3.17 *b* показано применение третьего правила к дереву решений, после шага 2 (рис. 3.17 *a*). Применение третьего правила удаляет две

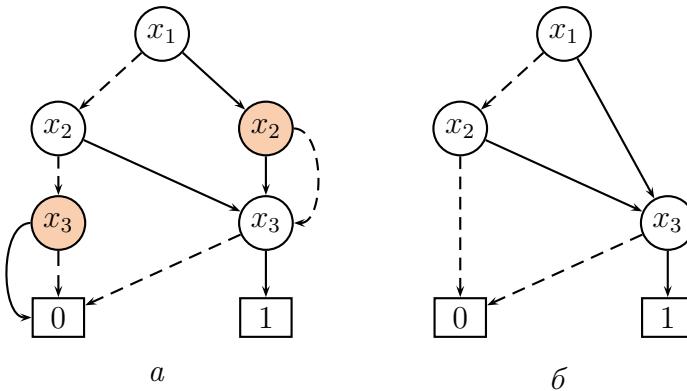


Рис. 3.17

вершины: одну с переменной x_3 и одну с переменной x_2 (помеченные цветом на рис. 3.17 а).

Замечание 3.34. В общем случае правила преобразования должны применяться неоднократно, поскольку каждое применение преобразования может привести к появлению новых возможностей для других преобразований.

Теорема 3.29. Относительно OBDD-представления очевидны следующие утверждения.

1. Любая булева функция при любом заданном порядке на переменных имеет OBDD-представление.
2. Для любого заданного порядка на переменных две упорядоченные бинарные диаграммы решений одной и той же функции изоморфны.

Из теоремы 3.29 этого вытекают несколько важных следствий.

Легко может быть проверена выполнимость. Функция является выполнимой в том и только том случае, когда ее OBDD-представление не сводится к одиночной терминальной вершине с меткой 0.

Любая тавтологическая функция должна иметь терминальную вершину с меткой 1 в качестве своего OBDD-представления.

Если функция не зависит от переменной x , то ее OBDD-представление не может иметь вершин, помеченных x .

Таким образом, как только построены представления функций в виде OBDD, многие их характеристики становятся легко проверяемыми. Многие логические операции (конъюнкция, дизъюнкция, отрицание) могут быть выполнены непосредственно над OBDD с помощью алгоритмов, выполняющих манипуляции над графами за полиномиальное время.

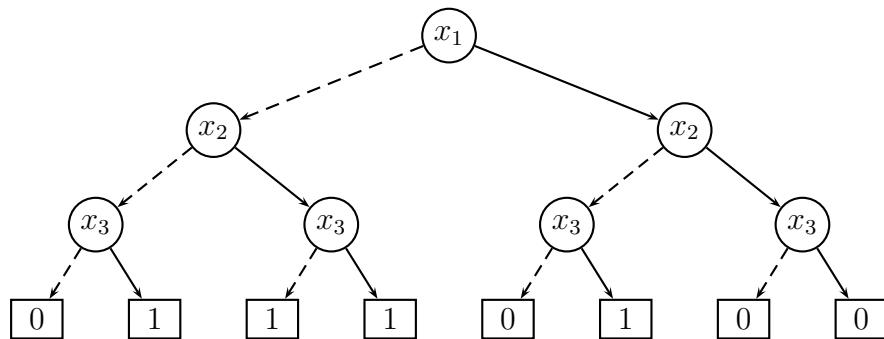
Замечание 3.35. Размер OBDD определяется как булевой функцией, так и выбором порядка входных переменных. При составлении графа для булевой функции $f(x_1, \dots, x_n)$ количество узлов в лучшем случае может линейно зависеть от n , а в худшем случае зависимость может быть экспоненциальной при неудачном выборе порядка входных переменных.

Выбор порядка переменных является критически важным при использовании таких структур данных на практике. Проблема нахождения лучшего порядка переменных является NP-полной задачей (см. определение 3.109).

Пример 3.63. Построим OBDD-представление булевой функции, заданное лексикографически упорядоченным вектором значений:

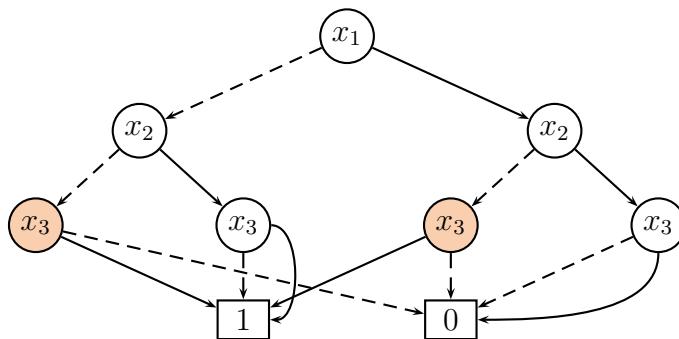
$$f(x_1, x_2, x_3) = \{0, 1, 1, 1, 0, 1, 0, 0\}.$$

На рис. 3.18 представлена исходная бинарная диаграмма решений данной функции.



Puc. 3.18

После применения первого правила получаем следующее дерево (рис. 3.19).



Puc. 3.19

Применение второго правила показано на рис. 3.20 а. Две вершины с меткой x_3 , помеченные цветом на рис. 3.19 сливаются в одну.

Применение третьего правила (рис 3.20 *b*) удаляет две вершины с переменной x_3 (помеченные цветом на рис. 3.20 *a*).

Пример 3.64. Построим OBDD-представление булевой функции, заданное лексикографически упорядоченным вектором значений:

$$f(x_1, x_2, x_3) = \{1, 0, 0, 1, 1, 0, 1, 0\}.$$

На рис. 3.21 представлена исходная бинарная диаграмма решений данной функции.

После применения первого правила получаем следующее дерево (рис. 3.22).

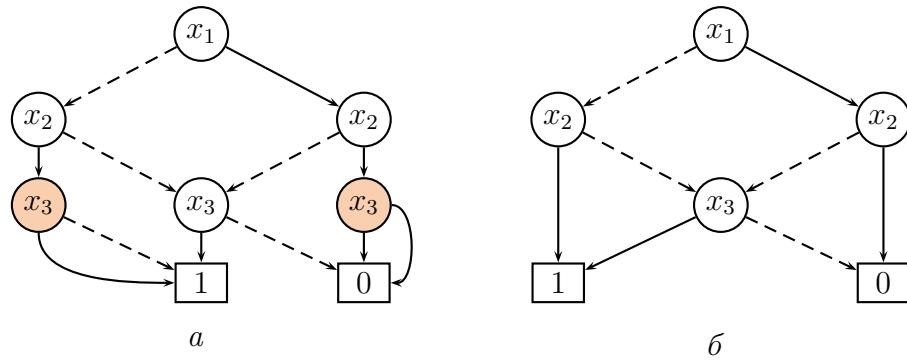


Рис. 3.20

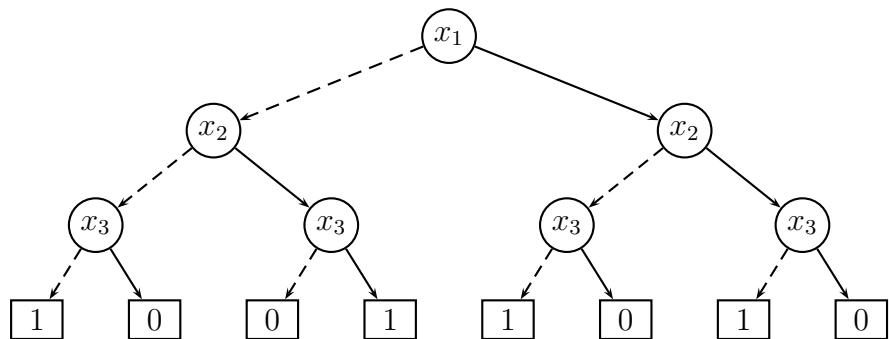


Рис. 3.21

Применение второго правила показано на рис. 3.23 *a*. Три вершины с меткой x_3 , помеченные цветом на рис. 3.22 сливаются в одну левую вершину с меткой x_3 на рис. 3.23 *a*. Все эти вершины имеют на рис. 3.22 в качестве левого потомка терминальную вершину 1, а в качестве правого – терминальную вершину 0.

3.3. Логика предикатов

Средствами логики нельзя выразить даже очень простые с математической точки зрения рассуждения. Рассмотрим, например, следующее умозаключение. *Любой человек смертен, Сократ – человек, следовательно, Сократ смертен*. Все эти утверждения с точки зрения логики высказываний являются простыми, их структура не может быть проанализирована, поэтому нельзя доказать логичность этого рассуждения в рамках логики высказываний.

Рассмотрим расширение логики высказываний, которое называют *логикой предикатов*.

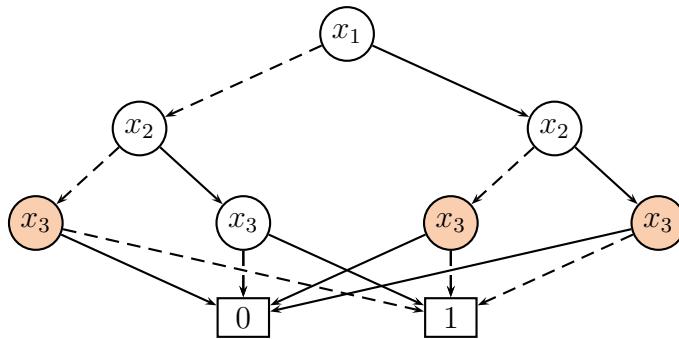


Рис. 3.22

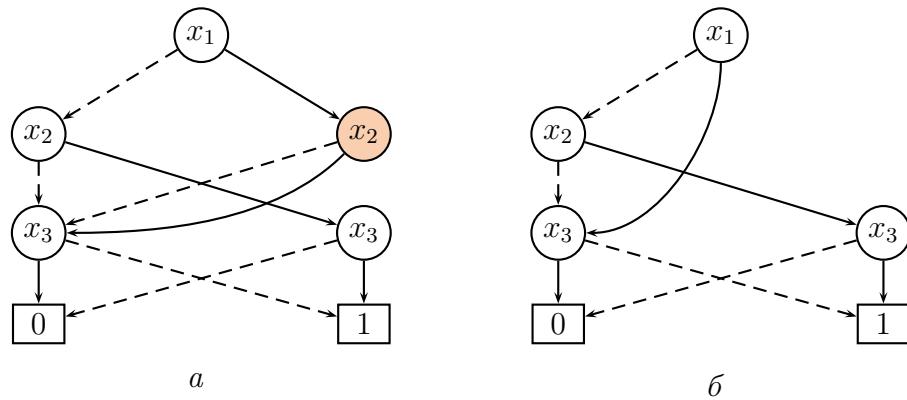


Рис. 3.23

3.3.1. Основные определения

В естественной речи часто встречаются сложные высказывания, истинность которых может изменяться при изменении объектов, о которых идет речь, хотя форма самого высказывания остается прежней. Например, высказывание «У кошки четыре ноги» истинно, «У слона четыре ноги» тоже истинно, а высказывание «У человека четыре ноги» — ложно. Все эти высказывания имеют одну форму «У объекта x четыре ноги», где x — некоторый объект. Такое переменное высказывание, истинностное значение которого зависит от параметра, и называют предикатом.

Таким образом, предикат есть функция, определенная на некотором множестве параметров и со значениями в $\{0, 1\}$. Этот подход естественным образом можно обобщить для случая n параметров.

Дадим формальное определение.

Определение 3.55. Пусть M — непустое множество. Тогда n -местным¹ предикатом² $P(x_1, \dots, x_n)$, называют функцию $P : M^n \rightarrow \{0, 1\}$.

Множество M называют *предметным множеством*, а аргументы $x_1, \dots, x_n \in M$ — *предметными переменными*.

¹Иногда местность предиката или функции будем обозначать верхним индексом P^n .

²От лат. *praedicatum* — сказанное.

Замечание 3.36. Можно считать, что высказывание это *нульместный предикат*, т. е. предикат, в котором нет переменных для замены.

Пример 3.65. Пусть множество M есть класс млекопитающих. Рассмотрим одноместный предикат $P(x) : \text{У объекта } x \text{ четыре ноги}$. Тогда $P(\text{слон}) = 1, P(\text{кошка}) = 1, P(\text{человек}) = 0$.

Пример 3.66. Пусть M — множество натуральных чисел \mathbb{N} . Рассмотрим двухместные предикаты $Q(x, y) : D(x, y) = 1, G(x, y) : x < y$. Тогда, например, $Q(5, 8) = 1, Q(6, 8) = 0$, а $G(1, 3) = 1, G(8, 5) = 0$.

Замечание 3.37. Булева функция от n переменных есть n -местный предикат, заданный на предметном множестве $M = B^n$.

Предикаты, заданные на множестве M , могут быть связаны логическими связками — *конъюнкцией*, *дизъюнкцией*, *отрицанием*, *импликацией* и *эквиваленцией*. Эти операции вводятся очевидным образом согласно табл. 3.1. Приведем в качестве примера определение конъюнкции предикатов.

Определение 3.56. Предикат $R(x_1, \dots, x_n)$ называют конъюнкцией предикатов $P(x_1, \dots, x_n)$ и $Q(x_1, \dots, x_n)$, заданных на множестве M , если на любом наборе значений предметных переменных $(\alpha_1, \dots, \alpha_n)$ из M

$$R(\alpha_1, \dots, \alpha_n) = P(\alpha_1, \dots, \alpha_n) \wedge Q(\alpha_1, \dots, \alpha_n).$$

Для предикатов вводятся две новые операции — *квантор общности* и *квантор существования*.

Определение 3.57. Пусть $P(x_1, \dots, x_n, y)$ — предикат, заданный на множестве M .

Выражение $(\forall y)P(x_1, \dots, x_n, y)$ (читается: *для всякого y выполняется $P(x_1, \dots, x_n, y)$*) есть предикат $Q(x_1, \dots, x_n)$, полученный из P навешиванием квантора *общности* на переменную y .

$$Q(\alpha_1, \dots, \alpha_n) = 1 \Leftrightarrow P(\alpha_1, \dots, \alpha_n, \beta) = 1, \text{ для любого } \beta \in M.$$

Выражение $(\exists y)P(x_1, \dots, x_n, y)$ (читается: *существует y такой, что выполняется $P(x_1, \dots, x_n, y)$*) есть предикат $R(x_1, \dots, x_n)$, полученный из P навешиванием квантора *существования* на переменную y .

$$R(\alpha_1, \dots, \alpha_n) = 1 \Leftrightarrow P(\alpha_1, \dots, \alpha_n, \beta) = 1, \text{ для некоторого } \beta \in M.$$

Операцию навешивания квантора \forall или \exists на переменную x называют еще *квантификацией переменной x* .

¹От лат. *quantum* — сколько.

Пример 3.67. В обозначениях примера 3.66 рассмотрим предикаты

$$R(x) = (\forall y)G(x, y), \quad W(y) = (\exists x)G(x, y).$$

Тогда $R(x) = 0$ для всех x , а $W(1) = 0$, $W(y) = 1$ для всех $y > 1$.

Введем понятие формулы логики предикатов. Сделаем это также, как и в логике высказываний (см. определение 3.3), т. е. сначала введем понятие атомарной формулы, а затем — формулы.

В отличие от логики высказываний, где переменные (атомарные формулы) не анализируются, атомарная формула для логики предикатов имеет структуру.

Определение 3.58. *Сигнатурой* называют множество

$$\Sigma = \Sigma_{\mathcal{F}} \cup \Sigma_{\mathcal{P}} \cup \Sigma_{\mathcal{C}}, \quad \text{где}$$

$\Sigma_{\mathcal{F}} = \{f_i^{n_i}\}_{i \in I}$ — множество функциональных символов (функций);

$\Sigma_{\mathcal{P}} = \{P_j^{n_j}\}_{j \in J}$ — множество предикатных символов;

$\Sigma_{\mathcal{C}} = \{c_k\}_{k \in K}$ — множество предметных констант (нульместных символов функций).

В определении сигнатуры можно оставить два множества $\Sigma_{\mathcal{F}}$ и $\Sigma_{\mathcal{P}}$, считая, что $\Sigma_{\mathcal{C}} \subseteq \Sigma_{\mathcal{F}}$

Определение 3.59. *Термом* называют

- 1) переменную или предметную константу;
- 2) выражение $f(t_1, \dots, t_n)$, где t_1, \dots, t_n — термы, а $f \in \Sigma_{\mathcal{F}}$.

Определение 3.60. *Атомарной формулой* называют выражение вида $P(t_1, \dots, t_n)$, где t_1, \dots, t_n — термы, а $P \in \Sigma_{\mathcal{P}}$.

Определение 3.61. *Формулой логики предикатов* называют

- 1) атомарную формулу,
- 2) выражение вида

$$(F) \wedge (G), \quad (F) \vee (G), \quad \neg(F), \quad (F) \rightarrow (G), \quad (F) \leftrightarrow (G), \quad (\forall y)F, \quad (\exists y)F,$$

где F и G — формулы логики предикатов, переменная $y \in M$.

Формулу F в двух последних выражениях называют *областью действия квантора* по переменной y .

Замечание 3.38. К соглашениям о скобках и приоритете операций, принятом в логике высказываний (см. подразд. 3.1.1), добавим следующее: кванторы имеют одинаковый приоритет, который больше приоритета всех связок.

Как и ранее вместо $\neg(F)$ будем писать \overline{F} , а $(F) \wedge (G)$ иногда заменять на FG .

Определение 3.62. Вхождение переменной в формулу называют *связанным*, если переменная стоит непосредственно за квантором или входит в область действия квантора по этой переменной. В противном случае вхождение называют *свободным*.

Переменную называют *свободной переменной* формулы, если существует хотя бы одно свободное вхождение переменной в формулу и *связной переменной* формулы, если имеется связное вхождение.

Формулу называют *замкнутой формулой* или *предложением* если любое вхождение переменной в формулу является связным.

Пример 3.68. Рассмотрим формулу

$$(\exists x)Q(x, y) \rightarrow \overline{P(g(x, y))} \wedge (\forall z)P(z).$$

Первое вхождение переменной x связано, второе — свободно. Все вхождения переменной y свободны, вхождение переменной z — связано.

Формула $(\forall x)(P(x) \rightarrow Q(x) \wedge R(x))$ является предложением, так как все вхождения переменной x связанные.

3.3.2. Семантика логики предикатов

Формула логики предикатов есть только абстрактная схема высказывания. Введем понятие значения (смысла) формулы, т. е. семантику логики предикатов. В логике высказываний это было сделано с помощью интерпретации переменных логическими значениями 0, 1. Используем аналогичный подход для формул логики предикатов.

Определение 3.63. Алгебраической системой $\mathfrak{M} = \langle M, \Sigma^{\mathfrak{M}} \rangle$ сигнатуры Σ называют непустое предметное множество M для которого задана *интерпретация* *сигнатурных символов* или просто *интерпретация*, т. е. функция, которая:

- 1) каждой предметной константе $c \in \Sigma_C$ ставит в соответствие выделенный элемент $c^{\mathfrak{M}} \in M$;
- 2) каждому символу n -местной функции $f^n \in \Sigma_F$ ставит в соответствие n -местную функцию $f^{\mathfrak{M}}$, определенную на множестве M ;
- 3) каждому символу n -местного предиката $P^n \in \Sigma_P$ ставит в соответствие n -местный предикат $P^{\mathfrak{M}}$, заданный на M .

Если сигнтура не содержит функциональных символов, то такую алгебраическую систему называют *моделью*.

В результате любая формула F получает в соответствие предикат, местность которого равна числу свободных переменных формулы F .

Замечание 3.39. Обычно, когда алгебраическая система известна и задана интерпретация, при обозначении предикатов, функций и выделенных элементов индекс \mathfrak{M} опускается.

Пример 3.69. Пусть $\mathfrak{M} = \langle \mathbb{N}, \Sigma^{\mathfrak{M}} = \{S(x, y, z), P(x, y, z)\} \rangle$, где

$$S(x, y, z) : x + y = z, \quad P(x, y, z) : xy = z.$$

1. Формула $F(x) = (\forall y)S(x, y, y)$ истинна в \mathfrak{M} только при $x = 0$.
2. Формула $Q(x) = (\forall y)P(x, y, y)$ истинна в \mathfrak{M} только при $x = 1$.
3. Формула $R(x, y) = (\forall z)(\forall u)(S(x, z, u) \rightarrow S(y, z, u))$ истинна в \mathfrak{M} только при $x = y$.
4. Формула $W(x, y) = (\exists z)S(x, z, y)$ истинна в \mathfrak{M} только при $x < y$.

Пример 3.70. Пусть $\mathfrak{M} = \langle 2^A, \Sigma^{\mathfrak{M}} = \{Q(x, y)\} \rangle$, где A — некоторое множество, $Q(x, y) : x \subseteq y$.

1. Формула

$$F(x, y, z) = Q(x, y) \wedge Q(x, z) \wedge (\forall u)(Q(u, y) \wedge Q(u, z) \rightarrow Q(u, x))$$

истинна в \mathfrak{M} только при $x = y \cap z$.

2. Формула $G(x) = (\forall y)Q(x, y)$ истинна в \mathfrak{M} только при $x = \emptyset$.

Задача 3.26. Пусть $\mathfrak{M} = \langle M, \Sigma^{\mathfrak{M}} = \{P(x), L(x), F(x), B(x, y)\} \rangle$, где M — множество точек, прямых и плоскостей трехмерного евклидова пространства, а $P(x), L(x), F(x)$ и $B(x, y)$ следующие предикаты:

$$\begin{aligned} P(x) : &x - \text{точка}; \quad L(x) : x - \text{прямая}; \quad F(x) : x - \text{плоскость}; \\ B(x, y) : &x \text{ принадлежит } y. \end{aligned}$$

Запишите следующие формулы:

- а) через две различные точки можно провести единственную прямую;
- б) через три точки, не лежащие на одной прямой, можно провести единственную плоскость.

Проиллюстрируем использование выразительных возможностей языка логики предикатов.

Пример 3.71. Рассмотрим задачу перевода на язык логики предикатов следующего рассуждения.

Любой радикал является сторонником общественного прогресса. Некоторые консерваторы недолюбливают всех сторонников общественного прогресса. Значит, некоторые консерваторы недолюбливают всех радикалов.

Выберем сигнатуру:

$$\begin{aligned} R(x) : x - \text{радикал}; \quad P(x) : x - \text{сторонник общественного прогресса}; \\ K(x) : x - \text{консерватор}; \quad L(x, y) - x \text{ недолюбливает } y. \end{aligned}$$

Тогда рассуждение запишется в виде следующей последовательности формул.

$$\begin{aligned} F_1 &= (\forall x)(R(x) \rightarrow P(x)); \quad F_2 = (\exists x)(K(x) \wedge (\forall y)(P(y) \rightarrow L(x, y))); \\ F_3 &= (\exists x)(K(x) \wedge (\forall y)(R(y) \rightarrow L(x, y))). \end{aligned}$$

Рассмотрение примера 3.71 показывает, что выразительных средств логики предикатов достаточно, чтобы записать рассуждение о радикалах и консерваторах. Естественно далее поставить вопрос, логично ли оно? Будет ли третье предложение следствием первых двух? Ответ на этот вопрос будет получен в примере 3.78 (см. подразд. 3.3.5)

3.3.3. Примеры и задачи на выразительность логики предикатов

Рассмотрим еще несколько примеров формулировки утверждений с помощью предикатов.

Пример 3.72. На предметном множестве натуральных чисел задана сигнатура:

- 1) $E(x)$ – число x – четное;
- 2) $P(x)$ – число x – простое;
- 3) $S(x, y, z)$ – число x является суммой чисел y и z .

Запишем в виде формулы логики предикатов теорему Гольдбаха: *каждое четное натуральное число является суммой двух простых чисел*.

$$F = (\forall x) [E(x) \rightarrow (\exists y)(\exists z) (P(y) \wedge P(z) \rightarrow S(y, z, x))]$$

Пример 3.73. Пусть на предметном множестве $M = \mathbb{N}$ определены предикаты:

- 1) $P(x)$ – число x – простое;
- 2) $Q(x, y)$ – число x больше числа y .

Выразим с помощью формулы логики предикатов утверждение (Евклид): *простых чисел бесконечно много*.

$$G = (\forall x)(\exists y) [P(y) \wedge Q(y, x)]$$

Пример 3.74. Пусть на предметном множестве геометрических фигур заданы следующие предикаты:

- 1) $C(x)$ – x – квадрат;
- 2) $S(x)$ – x – шар;
- 3) $B(x)$ – x фигура черного цвета;
- 4) $W(x)$ – x фигура белого цвета.

Запишем в виде формулы логики предикатов утверждение: *если все шары черные, то белых квадратов нет.*

$$F = (\forall x) (S(x) \rightarrow B(x)) \rightarrow \overline{(\exists y) (C(y) \wedge W(y))}$$

Задача 3.27. Пусть определена сигнатура:

- 1) $S(x, t)$ – я вижу предмет x в момент времени t ;
- 2) $T(x, t)$ – я беру предмет x в момент времени t .

Записать в виде формулы логики предикатов следующие утверждения:

1. Я беру всякую вещь, которую я никогда не вижу.
2. Я никогда не беру того, что я всегда вижу.
3. Всегда существуют вещи, которые я не вижу и не беру.

Задача 3.28. Подберите сигнатуру и запишите следующие рассуждения в виде последовательности формул логики предикатов.

1. Имеются прилежные студенты. Ни один студент не лишен способностей. Значит, некоторые студенты, лишенные способностей, не прилежны.

2. Тот, кто распускает этот слух, должен быть и ловким, и беспринципным. Браун не ловок. Джонсон не беспринципен. Значит, ни Браун, ни Джонсон не распускают этот слух.

3. Если бы кто-нибудь мог решить эту задачу, то и какой-нибудь математик мог бы. Смит – математик, а не может ее решить. Значит, задача неразрешима.

3.3.4. Равносильность формул логики предикатов

Определение 3.64. Формулы F и G называют *равносильными*, если для любой алгебраической системы $\mathfrak{M} = \langle M, \Sigma^{\mathfrak{M}} \rangle$ они одновременно истинны или одновременно ложны в \mathfrak{M} при любых значениях свободных переменных.

Как и в логике высказываний равносильность формул будем символически обозначать $F = G$.

В отличие от логики высказываний равносильность формул логики предикатов нельзя определить на основе построения таблиц истинности на всевозможных интерпретациях. Их может быть бесконечно много.

Первый способ определения равносильности основан на рассмотрении значений истинности обеих формул для произвольной алгебраической системы согласно определению 3.64.

Пример 3.75. Установим равносильность формул $F(x) = \overline{(\forall y)P(x,y)}$ и $G(x) = (\exists y)\overline{P(x,y)}$

Рассмотрим произвольную алгебраическую систему $\mathfrak{M} = \langle M, \Sigma^{\mathfrak{M}} \rangle$. Пусть $F^{\mathfrak{M}}(a) = 1$ для произвольного $a \in M$. Тогда найдется $b \in M$, такой, что $P^{\mathfrak{M}}(a, b) = 0$. Следовательно $\overline{P^{\mathfrak{M}}(a, b)} = 1$. Из последнего равенства получаем, что $G^{\mathfrak{M}}(a) = 1$.

Итак, доказано, что если $F^{\mathfrak{M}}(a)$ истинно, то $G^{\mathfrak{M}}(a)$ тоже истинно для произвольного $a \in M$. Обратная импликация доказывается аналогично. Следовательно, формулы $F(x)$ и $G(x)$ равносильны.

Определение 3.65. Формулу F называют *тождественно истинной* (общезначимой или тавтологией), если она истинна в любой алгебраической системе при любых значениях свободных переменных.

Прием, использованный в примере 3.75, позволяет устанавливать равносильность для достаточно простых формул и его сложно применять в общем случае. Основной способ состоит в преобразовании формул на основе установленных равносильностей. Имеет место следующая теорема.

Теорема 3.30. Пусть F, G, H — произвольные формулы логики предикатов. Тогда законы 1 – 15 логики высказываний, установленные в теореме 3.4, справедливы и для формул логики предикатов.

Доказательство. Доказательство каждого закона проводится так же, как в примере 3.75 с использованием табл. 3.1 истинности логических связок. ■

Дополним полученный список законами, специфичными для логики предикатов.

Теорема 3.31. Для произвольных формул F и G логики предикатов справедливы равносильности (законы):

$$16) \quad (\forall x)(F(x) \wedge G(x)) = (\forall x)F(x) \wedge (\forall x)G(x);$$

$$17) \quad (\exists x)(F(x) \vee G(x)) = (\exists x)F(x) \vee (\exists x)G(x);$$

$$18) \quad (\forall x)(\forall y)F(x, y) = (\forall y)(\forall x)F(x, y);$$

$$19) \quad (\exists x)(\exists y)F(x, y) = (\exists y)(\exists x)F(x, y);$$

$$20) \quad \overline{(\forall x)F(x)} = (\exists x)\overline{F(x)};$$

$$21) \quad \overline{(\exists x)F(x)} = (\forall x)\overline{F(x)}.$$

Доказательство. В примере 3.75 была установлена равносильность 20. Остальные доказываются аналогично. ■

Согласно законам 16 и 17 квантор общности можно переносить через конъюнкцию, а квантор существования — через дизъюнкцию. Однако поменять местами \wedge и \vee в этих законах нельзя.

Доказательство того, что формулы неравносильны, будем осуществлять построением алгебраической системы \mathfrak{M} , в которой одна формула истинна, другая ложна.

Следствие 3.7. Существует алгебраическая система \mathfrak{M} , в которой

$$\begin{aligned} (\forall x)(F(x) \vee G(x)) &\neq (\forall x)F(x) \vee (\forall x)G(x); \\ (\exists x)(F(x) \wedge G(x)) &\neq (\exists x)F(x) \wedge (\exists x)G(x). \end{aligned}$$

Доказательство. Рассмотрим $\mathfrak{M} = \langle \mathbb{N}, \Sigma^{\mathfrak{M}} = \{F(x), G(x)\} \rangle$, где

$$F(x) : x — четное число, \quad G(x) : x — нечетное число.$$

Для данной алгебраической системы \mathfrak{M} получаем неравенства

$$\begin{aligned} (\forall x)(F(x) \vee G(x)) = 1 &\neq (\forall x)F(x) \vee (\forall x)G(x) = 0, \\ (\exists x)(F(x) \wedge G(x)) = 0 &\neq (\exists x)F(x) \wedge (\exists x)G(x) = 1, \end{aligned}$$

которые и доказывают следствие. ■

Равносильности 18 и 19 позволяют менять местами одноименные кванторы. Однако, для разноименных кванторов, эта операция не сохраняет равносильность.

Следствие 3.8. Существует алгебраическая система \mathfrak{M} , в которой

$$(\forall x)(\exists y)(F(x, y) \neq (\exists y)(\forall x)F(x, y). \quad (3.40)$$

Доказательство. Рассмотрим систему $\mathfrak{M} = \langle \mathbb{N}, \Sigma^{\mathfrak{M}} = \{F(x, y)\} \rangle$, где $F(x, y) : x < y$. Тогда в (3.40) левая формула будет истинной, а правая — ложной. ■

В следствии 3.7 было доказано, что квантор существования нельзя переносить через конъюнкцию. Однако, если одна из формул F или G не содержит переменной x , то легко показать, что это делать можно.

Следствие 3.9. Пусть формула G не содержит x . Тогда справедливы равносильности

$$22) \quad (\forall x)(F(x) \vee G) = (\forall x)F(x) \vee G;$$

$$23) \quad (\exists x)(F(x) \wedge G) = (\exists x)F(x) \wedge G.$$

Результаты теоремы 3.31 и следствия 3.9 обобщает следующая лемма.

Лемма 3.18. Равносильности 16, 17, и 22, 23 можно записать в общем виде:

$$24) \quad (\mathbf{Q}_1x)(\mathbf{Q}_2y)(F(x) \vee G(y)) = (\mathbf{Q}_1x)F(x) \vee (\mathbf{Q}_2y)G(y);$$

$$25) \quad (\mathbf{Q}_1x)(\mathbf{Q}_2y)(F(x) \wedge G(y)) = (\mathbf{Q}_1x)F(x) \wedge (\mathbf{Q}_2y)G(y),$$

где $\mathbf{Q}_1, \mathbf{Q}_2$ — кванторы \exists или \forall ; переменная y не входит в $F(x)$; переменная x не входит в $G(y)$.

Рассмотрим еще два очевидных закона логики предикатов.

Лемма 3.19 (о переименовании связанных переменных). Пусть переменная z не входит в $F(x)$, а переменная x не входит в $F(z)$. Тогда имеют место следующие равносильности

$$26) \quad (\forall x)F(x) = (\forall z)F(z);$$

$$27) \quad (\exists x)(F(x)) = (\exists z)F(z).$$

Проиллюстрируем применение полученных равносильностей логики предикатов на примере.

Замечание 3.40. Как и ранее для логики высказываний и булевых функций в цепочке равенств формул логики предикатов будем ставить номера использованных равносильностей из теоремы 3.31, ее следствий и леммы 3.19.

Пример 3.76. Докажем равносильность формул F и G , где:

$$\begin{aligned} F &= \overline{(\exists x)((\forall y)P(x, y) \rightarrow (\forall z)(P(z, z) \vee Q(z)))}; \\ G &= (\forall x)(\forall y)(\exists z) \left(P(x, y) \wedge \overline{P(z, z)} \wedge \overline{Q(z)} \right). \end{aligned}$$

Запишем последовательность равносильных преобразований.

$$\begin{aligned} F &\stackrel{21}{=} (\forall x)\overline{((\forall y)P(x, y) \rightarrow (\forall z)(P(z, z) \vee Q(z)))} \stackrel{14}{=} \\ &\stackrel{14}{=} (\forall x)\overline{\left((\forall y)P(x, y) \vee (\forall z)(P(z, z) \vee Q(z)) \right)} \stackrel{12,13}{=} \\ &\stackrel{12,13}{=} (\forall x) \left((\forall y)P(x, y) \wedge \overline{(\forall z)(P(z, z) \vee Q(z))} \right) \stackrel{20,12}{=} \\ &\stackrel{20,12}{=} (\forall x) \left((\forall y)P(x, y) \wedge (\exists z)(\overline{P(z, z)} \wedge \overline{Q(z)}) \right) \stackrel{25}{=} \\ &\stackrel{25}{=} (\forall x)(\forall y)(\exists z) \left(P(x, y) \wedge \overline{P(z, z)} \wedge \overline{Q(z)} \right) = G. \end{aligned}$$

Задача 3.29. Докажите равносильности:

- a) $(\forall x)((\forall y)P(x, y) \rightarrow (\exists z)(P(x, z) \wedge Q(z))) = (\forall x)(\exists u)(P(x, u) \rightarrow Q(u));$
- b) $(\exists x)F(x) \rightarrow (\exists x)G(x) = (\forall x)(\exists y)(F(x) \rightarrow G(y)).$

3.3.5. Логическое следствие

Определение 3.66. Формулу G называют *логическим следствием*¹ множества формул $K = \{F_1, \dots, F_k\}$, если для любой алгебраической системы \mathfrak{M} из истинности в \mathfrak{M} всех формул из K при некоторых значениях переменных следует истинность G в \mathfrak{M} при тех же значениях переменных (символически обозначают $K \models G$).

Пример 3.77. Покажем, что формула $G = Q(c)$ есть логическое следствие формул $F_1 = (\forall x)(P(x) \rightarrow Q(x) \wedge R(x))$ и $F_2 = P(c)$.

Рассмотрим произвольную алгебраическую систему $\mathfrak{M} = \langle M, \Sigma^{\mathfrak{M}} \rangle$. Пусть $F_1^{\mathfrak{M}} = F_2^{\mathfrak{M}} = 1$. Из истинности $F_2^{\mathfrak{M}}$ следует, что $P^{\mathfrak{M}}(c^{\mathfrak{M}}) = 1$. Истинность $F_1^{\mathfrak{M}}$ означает, что $P^{\mathfrak{M}}(\alpha) \rightarrow Q^{\mathfrak{M}}(\alpha) \wedge R^{\mathfrak{M}}(\alpha) = 1$ для любого элемента $\alpha \in M$, в частности и для $\alpha = c^{\mathfrak{M}}$. Таким образом для $\alpha = c^{\mathfrak{M}}$ истинна импликация и ее посылка, следовательно, истинно и заключение, т. е. $Q^{\mathfrak{M}}(c^{\mathfrak{M}}) \wedge R^{\mathfrak{M}}(c^{\mathfrak{M}}) = 1$. Отсюда $G^{\mathfrak{M}} = Q^{\mathfrak{M}}(c^{\mathfrak{M}}) = 1$ и в силу произвольности алгебраической системы \mathfrak{M} формула G есть логическое следствие формул F_1 и F_2 .

Пример 3.78. Докажем логичность рассуждения о радикалах и консерваторах (см. пример 3.71). Это рассуждение записано в виде последовательности формул F_1 , F_2 , и F_3 .

Рассмотрим произвольную алгебраическую систему $\mathfrak{M} = \langle M, \Sigma^{\mathfrak{M}} \rangle$. Пусть $F_1^{\mathfrak{M}}$ и $F_2^{\mathfrak{M}}$ истинны. Покажем, что $F_3^{\mathfrak{M}}$ также истинна. Из истинности $F_2^{\mathfrak{M}}$ следует, что существует $\alpha \in M$, $K^{\mathfrak{M}}(\alpha) = 1$.

Возьмем произвольный элемент $\beta \in M$. Если $R^{\mathfrak{M}}(\beta) = 0$, то

$$K^{\mathfrak{M}}(\alpha) \wedge (R^{\mathfrak{M}}(\beta) \rightarrow L^{\mathfrak{M}}(\alpha, \beta)) = 1.$$

Пусть $R^{\mathfrak{M}}(\beta) = 1$. Тогда из истинности $F_1^{\mathfrak{M}}$ получаем $P^{\mathfrak{M}}(\beta) = 1$, а из истинности $F_2^{\mathfrak{M}}$ следует, что и $L^{\mathfrak{M}}(\alpha, \beta) = 1$.

В силу произвольности $\beta \in M$ формула $F_3^{\mathfrak{M}}$ истинна. Таким образом рассуждение о радикалах и консерваторах логично.

Определение 3.67. Множество формул логики предикатов

$$K = \{F_1(x_1, \dots, x_l), \dots, F_n(x_1, \dots, x_l)\}$$

выполнимо, если существует алгебраическая система $\mathfrak{M} = \langle M, \Sigma^{\mathfrak{M}} \rangle$ и набор значений свободных переменных $\alpha_1, \dots, \alpha_l \in M$ таких, что

$$F_1^{\mathfrak{M}}(\alpha_1, \dots, \alpha_l) = 1, \dots, F_n^{\mathfrak{M}}(\alpha_1, \dots, \alpha_l) = 1.$$

¹Иногда говорят, что формула G семантически следует из множества формул K .

Пример 3.79. Множество формул

$$K = \{F_1 = (\forall x)(\exists y)(P(y) \wedge Q(x, y)), F_2 = (\forall y)Q(c, y), F_3 = \overline{P(c)}\}$$

выполнимо.

Пусть $\mathfrak{M} = \langle \mathbb{N}, \Sigma^{\mathfrak{M}} = \{P(x), Q(x, y), c\} \rangle$, где

$$P(x) : x - \text{простое число}, \quad Q(x, y) : x \leq y, \quad c = 1.$$

Тогда высказывание $F_1^{\mathfrak{M}}$ означает, что для любого натурального числа x найдется простое число y , которое не меньше x ; высказывание $F_2^{\mathfrak{M}}$ означает, что 1 — наименьшее натуральное число; высказывание $F_3^{\mathfrak{M}}$ означает, что 1 — не простое число. Ясно, что все эти высказывания истинны, и поэтому множество формул K выполнимо.

Понятия логического следствия и выполнимости в логике первого порядка связаны точно так же, как и в логике высказываний.

Теорема 3.32. Формула G является логическим следствием формул F_1, \dots, F_k тогда и только тогда, когда множество формул $\{F_1, \dots, F_k, \overline{G}\}$ невыполнимо.

Доказательство. Для доказательства нужно провести рассуждения, аналогичные использованным при доказательстве теоремы 3.6. ■

3.3.6. Нормальные формы

Как и в логике высказываний, в логике предикатов рассматриваются формулы в определенной канонической форме. Приведем две из них: предваренную нормальную и сколемовскую нормальную формы.

Определение 3.68. Формула G имеет *предваренную* или *пренексную нормальную форму* (ПНФ), если

$$G = (Q_1 x_1) \dots (Q_n x_n) F, \quad (3.41)$$

где Q_1, \dots, Q_n кванторы \exists или \forall , а F — бескванторная формула.

Выражение $(Q_1 x_1) \dots (Q_n x_n)$ называют *префиксом* формулы G , а F — *матрицей* формулы.

Если все Q_i равны \forall , то форму называют \forall -формулой, или *универсальной формулой*.

Пример 3.80. Формула $(\forall x)(\exists y) (P(x, y) \wedge \overline{Q(y)})$ имеет ПНФ, а формула $(\exists x)(T(x) \wedge (\forall y)(S(y) \rightarrow L(x, y)))$ не имеет.

Формула $(\forall x)(\forall y)(\forall z)(F(x, x) \wedge F(x, z) \rightarrow (F(x, y) \vee F(y, z)))$ имеет универсальную форму.

Теорема 3.33. Для любой формулы F существует формула G , равносильная F и имеющая ПНФ.

Доказательство. Теорема следует из рассмотрения алгоритма 3.13, который по данной формуле F выдает формулу G в ПНФ. ■

Алгоритм 3.13 (приведение к ПНФ).

Шаг 1. Используя равносильности 15 и 14, исключаем из исходной формулы эквиваленцию и импликацию.

Шаг 2. Переносим отрицание к атомарным формулам, пользуясь законами 11, 12 и 20, 21.

Шаг 3. С помощью равенств 24 и 25 выносим кванторы вперед, используя при необходимости переименование связанных переменных (равносильности 26 и 27).

Пример 3.81. Пусть $F = (\forall x)P(x) \rightarrow (\forall y)(Q(y) \rightarrow (\forall z)R(y, z))$. Выполним шаг 1.

$$F \stackrel{14}{=} \overline{(\forall x)P(x)} \vee (\forall y) \left(\overline{Q(y)} \vee (\forall z)R(y, z) \right).$$

Перейдем к шагу 2.

$$F \stackrel{20}{=} (\exists x)\overline{P(x)} \vee (\forall y) \left(\overline{Q(y)} \vee (\forall z)R(y, z) \right).$$

Выполнение шага 3 запишем цепочкой равносильностей.

$$\begin{aligned} F &\stackrel{24}{=} (\exists x)(\forall y) \left(\overline{P(x)} \vee \overline{Q(y)} \vee (\forall z)R(y, z) \right) \stackrel{22}{=} \\ &\stackrel{22}{=} (\exists x)(\forall y)(\forall z) \left(\overline{P(x)} \vee \overline{Q(y)} \vee R(y, z) \right). \end{aligned}$$

Задача 3.30. Приведите к ПНФ следующие формулы:

$$a) \overline{(\exists x)P(x)} \vee (\forall x)Q(x, y); \quad b) (\forall x) \left(R(x) \rightarrow \overline{(\exists y)S(y)} \right).$$

Определение 3.69. *Литералом* в логике предикатов называют атомарную формулу или ее отрицание.

Замечание 3.41. С учетом определения 3.69, как и в логике высказываний для бескванторной формулы логики предикатов можно ввести понятие конъюнктивной нормальной формы. При таком определении полностью сохраняется утверждение теоремы 3.9 и алгоритм приведения к КНФ, полученный при доказательстве теоремы 3.9.

Определение 3.70. Говорят, что \forall -формула имеет сколемовскую нормальную форму (СНФ), если ее матрица представлена в конъюнктивной нормальной форме.

Пример 3.82. Формула $(\forall x)(P(x) \wedge (\overline{R(y)} \vee Q(x, y)))$ имеет СНФ, а формулы $(\forall x)(\exists y)(Q(x, y) \vee G(y))$ и $(\forall x)(Q(x, y) \vee (P(y) \rightarrow G(x)))$ нет.

Вторая формула не является \forall -формулой, а у третьей матрица не имеет КНФ.

Рассмотрим алгоритм приведения к СНФ. Такой процесс называют *сколемизацией*¹.

Алгоритм 3.14 (приведение к СНФ).

Шаг 1. С помощью алгоритма 3.13 приводим формулу к ПНФ.

Шаг 2. Матрицу формулы приводим к КНФ (алгоритм приведения описан в доказательстве теоремы 3.9). После выполнения *шага 2* формула имеет вид (3.41).

Шаг 3. Исключаем кванторы существования из префикса. Пусть Q_k , где $k \in 1 : n$, самый левый квантор существования в представлении (3.41).

Если $k = 1$, то выбирается новая предметная константа α , отличная от констант, входящих в матрицу формулы; все вхождения переменной x_k в матрице заменяются на α , и $(Q_k x_k)$ вычеркивается из префикса.

В случае $k > 1$ выбирается $(k - 1)$ -местный функциональный символ f , отличный от функторов, входящих в матрицу формулы; все вхождения переменной x_k в матрице заменяются на $f(x_1, \dots, x_{k-1})$ и $(Q_k x_k)$ вычеркивается из префикса.

Такое преобразование проводится для всех кванторов существования слева направо.

Замечание 3.42. На шаге 3 алгоритма 3.14 можно не выделять отдельно случай $k = 1$, так как при этом 0-местный функциональный символ есть предметная константа.

Пример 3.83. С помощью алгоритма 3.14 привести к СНФ формулу

$$F = (\exists x)(\forall y)(P(x, y) \rightarrow (\exists z)(Q(x, z) \wedge G(y))).$$

Приводим формулу к ПНФ:

$$\begin{aligned} F &\stackrel{14}{=} (\exists x)(\forall y) \left(\overline{P(x, y)} \vee (\exists z)(Q(x, z) \wedge G(y)) \right) \stackrel{17}{=} \\ &\stackrel{17}{=} (\exists x)(\forall y)(\exists z) \left(\overline{P(x, y)} \vee (Q(x, z) \wedge G(y)) \right). \end{aligned}$$

Преобразуем матрицу к КНФ:

$$F \stackrel{8}{=} (\exists x)(\forall y)(\exists z) \left((\overline{P(x, y)} \vee (Q(x, z))) \wedge (\overline{P(x, y)} \vee (G(y))) \right).$$

¹Процедура названа в честь известного логика Т. А. Сколема

Проводим процедуру сколемизации. Заменив сначала x на α , а затем z на $f(y)$, получаем СНФ формулы F

$$\begin{aligned} F &\xrightarrow{x=\alpha} (\forall y)(\exists z) \left(\left(\overline{P(\alpha, y)} \vee Q(\alpha, z) \right) \wedge \left(\overline{P(\alpha, y)} \vee G(y) \right) \right) \xrightarrow{z=f(y)} \\ &\xrightarrow{z=f(y)} (\forall y) \left(\left(\overline{P(a, y)} \vee Q(a, f(y)) \right) \wedge \left(\overline{P(a, y)} \vee G(y) \right) \right). \end{aligned}$$

Справедлива следующая теорема.

Теорема 3.34. Для любой формулы F алгоритм 3.14 строит формулу G , имеющую СНФ и одновременно выполнимую (или невыполнимую) с F .

Доказательство. Из описания алгоритма ясно, что результатом его работы является формула в СНФ. Очевидно, преобразования, проводимые на первых двух шагах алгоритма, являются равносильными, и в частности, сохраняют выполнимость (или невыполнимость) исходной формулы.

Проведем анализ шага 3 (исключение кванторов существования). Пусть \widehat{F} – формула после шага 2. Покажем, что \widehat{F} невыполнима тогда и только тогда, когда невыполнима G . Рассмотрим случай, когда \widehat{F} содержит только один квантор существования:

$$\begin{aligned} \widehat{F} &= (\forall x_1) \dots (\forall x_{k-1}) (\exists x_k) (\forall x_{k+1}) \dots (\forall x_n) Q(x_1, \dots, x_n), \\ G &= (\forall x_1) \dots (\forall x_{k-1}) (\forall x_{k+1}) \dots (\forall x_n) Q(x_1, \dots, x_{k-1}, y, x_{k+1}, \dots, x_n), \end{aligned}$$

где $y = f(x_1, \dots, x_{k-1})$.

Пусть формула \widehat{F} невыполнима. Допустим, что формула G выполнима, т. е. существует интерпретация, в которой

$$\begin{aligned} (\forall x_1) \dots (\forall x_{k-1}) (\forall x_{k+1}) \dots (\forall x_n) Q(x_1, \dots, x_{k-1}, \alpha, x_{k+1}, \dots, x_n) &= 1, \\ \alpha &= f(x_1, \dots, x_{k-1}). \end{aligned}$$

Очевидно, что \widehat{F} также выполнима в данной интерпретации.

Обратно, пусть формула G – невыполнима. Допустим, что \widehat{F} выполнима. Введем функцию $f(x_1, \dots, x_{k-1}) = x_k$. Тогда очевидно, что $G = 1$, полученное противоречие доказывает теорему.

Если в префиксе имеется $m > 1$ кванторов существования, то доказательство проводится аналогично. ■

3.3.7. Примеры и задачи на приведение к ПНФ и СНФ

Пример 3.84. Приведем к сколемовской нормальной форме следующую формулу:

$$F = (\exists x)(\forall y) (P(x) \rightarrow Q(y, z)) \rightarrow (\exists x)(\forall z) (Q(x, z) \wedge P(y))$$

Сначало приведем формулу F к предваренной нормальной форме. Запишем последовательность равносильных преобразований.

$$\begin{aligned}
 F &\stackrel{14}{=} \overline{(\exists x)(\forall y) (\overline{P(x)} \vee Q(y, z))} \vee (\exists x)(\forall z) (Q(x, z) \wedge P(y)) \stackrel{20,21}{=} \\
 &\stackrel{20,21}{=} (\forall x)(\exists y) \overline{(\overline{P(x)} \vee Q(y, z))} \vee (\exists x)(\forall z) (Q(x, z) \wedge P(y)) \stackrel{12,13}{=} \\
 &\stackrel{12,13}{=} (\forall x)(\exists y) \left(P(x) \wedge \overline{Q(y, z)} \right) \vee (\exists x)(\forall z) (Q(x, z) \wedge P(y)) \stackrel{27}{=} \\
 &\stackrel{27}{=} (\forall x)(\exists y) \left(P(x) \wedge \overline{Q(y, z)} \right) \vee (\exists w)(\forall z) (Q(w, z) \wedge P(y)) \stackrel{24}{=} \\
 &\stackrel{24}{=} (\forall x)(\exists y)(\exists w) \left(P(x) \wedge \overline{Q(y, z)} \vee (\forall z) (Q(w, z) \wedge P(y)) \right) \stackrel{26}{=} \\
 &\stackrel{26}{=} (\forall x)(\exists y)(\exists w) \left(P(x) \wedge \overline{Q(y, z)} \vee (\forall u) (Q(u, u) \wedge P(y)) \right) \stackrel{24}{=} \\
 &\stackrel{24}{=} (\forall x)(\exists y)(\exists w)(\forall u) \left(P(x) \wedge \overline{Q(y, z)} \vee Q(w, u) \wedge P(y) \right).
 \end{aligned}$$

Сколемизируем полученную формулу.

$$\begin{aligned}
 F &\xrightarrow{y=f(x)} (\forall x)(\exists w)(\forall u) \left(P(x) \wedge \overline{Q(f(x), z)} \vee Q(w, u) \wedge P(f(x)) \right) \xrightarrow{w=g(x)} \\
 &\xrightarrow{w=g(x)} (\forall x)(\forall u) \underbrace{\left(P(x) \wedge \overline{Q(f(x), z)} \vee Q(g(x), u) \wedge P(f(x)) \right)}_{M(x,u,z)}.
 \end{aligned}$$

В последнем выражении легко привести матрицу формулы в КНФ:

$$\begin{aligned}
 M(x, u, z) = & (P(x) \vee Q(g(x), u)) \wedge (P(x) \vee P(f(x))) \wedge \\
 & \wedge \left(\overline{Q(f(x), z)} \vee Q(g(x), u) \right) \wedge \left(\overline{Q(f(x), z)} \vee P(f(x)) \right).
 \end{aligned}$$

Пример 3.85. Приведем к сколемовской нормальной форме формулу уже представленную в ПНФ:

$$F = (\exists x)(\forall y)(\forall z)(\exists u)(\exists v)(\exists w) \left(P(x, y) \vee \overline{R(z, u, v)} Q(u, w) \right)$$

Проводим процедуру сколемизации.

$$\begin{aligned}
 F &\xrightarrow{x=\alpha} (\forall y)(\forall z)(\exists u)(\exists v)(\exists w) \left(P(\alpha, y) \vee \overline{R(z, u, v)} Q(u, w) \right) \xrightarrow{u=f(y,z)} \\
 &\xrightarrow{u=f(y,z)} (\forall y)(\forall z)(\exists v)(\exists w) \left(P(\alpha, y) \vee \overline{R(z, f(y, z), v)} Q(f(y, z), w) \right) \xrightarrow{v=g(y,z)} \\
 &\xrightarrow{v=g(y,z)} (\forall y)(\forall z)(\exists w) \left(P(\alpha, y) \vee \overline{R(z, f(y, z), g(y, z))} Q(f(y, z), w) \right) \xrightarrow{w=h(y,z)} \\
 &\xrightarrow{w=h(y,z)} (\forall y)(\forall z) \underbrace{\left(P(\alpha, y) \vee \overline{R(z, f(y, z), g(y, z))} Q(f(y, z), h(y, z)) \right)}_{M(y,z)}
 \end{aligned}$$

Аналогично примеру 3.84 приводим матрицу формулы в КНФ:

$$M(y, z) = \left(P(\alpha, y) \vee \overline{R(z, f(y, z), g(y, z))} \right) (P(\alpha, y) \vee Q(f(y, z), h(y, z)))$$

Задача 3.31. Привести формулу логики предикатов сначала к предваренной нормальной форме, затем к сколемовской нормальной форме.

$$F = (\forall x)(\forall y) [(\exists z) (P(x, z) \wedge P(y, z)) \rightarrow (\exists y)Q(x, y, u)]$$

Пример 3.86. Равносильными преобразованиями преобразуем формулу логики предикатов, представленную в ПНФ:

$$F = (\forall x)(\forall y)(\exists z)(\exists u) (P(z) \vee Q(x, y, u))$$

так, чтобы в ней осталось три квантора.

Имеем:

$$\begin{aligned} F &\stackrel{17}{=} (\forall x)(\forall y) ((\exists z)P(z) \vee (\exists u)Q(x, y, u)) \stackrel{23}{=} \\ &\stackrel{23}{=} (\forall x)(\forall y) ((\exists t)P(t) \vee (\exists t)Q(x, y, t)) \stackrel{19}{=} (\forall x)(\forall y)(\exists t) (P(t) \vee Q(x, y, t)) \end{aligned}$$

Пример 3.87. Используя равносильности преобразуем формулу логики предикатов, представленную в ПНФ:

$$G = (\exists x)(\exists y)(\forall z)(\forall u) (S(x, z, u) \vee L(y))$$

так, чтобы в ней осталось три квантора.

Запишем цепочку равносильных преобразований:

$$\begin{aligned} G &\stackrel{24}{=} (\exists x)(\forall z)(\forall u)S(x, z, u) \vee (\exists y)L(y) \stackrel{27}{=} (\exists t)(\forall z)(\forall u)S(t, z, u) \vee (\exists t)L(t) \stackrel{24}{=} \\ &\stackrel{24}{=} (\exists t)(\forall z)(\forall u) (S(t, z, u) \vee L(t)) \end{aligned}$$

3.3.8. Метод резолюций в логике предикатов

Обобщим метод резолюций для решения задачи логического следования в логике предикатов. Как и в логике высказываний, нужно преобразовать множество формул логики предикатов к множеству *бескванторных дизъюнктов* литералов, сохранив при этом преобразовании выполнимость (или невыполнимость).

Рассмотрим алгоритм такого преобразования для множества формул логики предикатов $T = \{F_1, \dots, F_k\}$. Множество дизъюнктов обозначим S_T .

Алгоритм 3.15 (инициализация метода резолюций).

Шаг 1. Полагаем $S_T = \emptyset$.

Шаг 2. Каждую формулу $F_i \in T$ с помощью алгоритма 3.14 приводим (сохраняя выполнимость) к формуле G_i в СНФ.

Шаг 3. Если формула G_i не является замкнутой (имеет свободные переменные x_{i_1}, \dots, x_{i_m}), то для каждой переменной x_{i_l} , заменяем все ее свободные вхождения на новую предметную константу α_{i_l} , отличную от констант, входящих в матрицу формулы G_i . Получаем замкнутую формулу H_i , одновременно выполнимую (или невыполнимую) с G_i .

Шаг 4. В формуле H_i опускаем кванторы всеобщности, подразумевая в дальнейшем, что каждая переменная связана соответствующим квантором.

Шаг 5. По построению H_i имеет вид КНФ, т. е. $H_i = D_{i_1} \cdots D_{i_n}$. В силу определения конъюнкции, очевидно, что выполнимость H_i равносильна выполнимости множества дизъюнктов $\{D_{i_1}, \dots, D_{i_n}\}$.

Шаг 6. $S_T = S_T \cup \{D_{i_1}, \dots, D_{i_n}\}$, для $i = 0, \dots, k$.

Замечание 3.43. Обоснование шага 4 алгоритма было проведено в теореме 3.13 для логики высказываний.

□ Подстановка и унификация

Так как литералы в логике предикатов имеют структуру, правило резолюций в прежнем виде не применимо. Действительно, множество дизъюнктов $S = \{P(x), \overline{P(\alpha)}\}$ невыполнимо (переменная x связана квантором общности). Но по правилу резолюций для логики высказываний нельзя построить опровержение S .

Сделаем в множестве S замену $x \rightarrow \alpha$. Получим множество дизъюнктов $\tilde{S} = \{P(\alpha), \overline{P(\alpha)}\}$. Очевидно, что множества S и \tilde{S} одновременно выполнимы (или невыполнимы). Но из \tilde{S} с помощью правила резолюций легко получить пустой дизъюнкт \square .

Таким образом, в логике предикатов правило резолюций надо дополнить возможностью делать подстановку.

Определение 3.71. Подстановкой называют множество равенств $\delta = \{x_1 = t_1, \dots, x_n = t_n\}$, где x_1, \dots, x_n — переменные; t_1, \dots, t_n — термы, причем терм t_i не содержит переменной x_i .

Действие подстановки $\delta = (x_1 = t_1, \dots, x_n = t_n)$ на множество дизъюнктов S , выражющееся в одновременной замене $x_1 \rightarrow t_1, \dots, x_n \rightarrow t_n$, обозначим $\delta(S)$; пустую подстановку, не содержащую равенств — ε ; композицию (последовательное применение) подстановок δ_1 и затем δ_2 — $\delta_2 \circ \delta_1$

Пример 3.88. Пусть

$$S = \{R(x, y) \vee P(f(y), z), Q(y, z, z)\}, \quad \delta = \{x = g(y), y = \alpha, z = h(x)\},$$

$$\text{тогда } \delta(S) = \{R(g(y), \alpha) \vee P(f(\alpha), h(x)), Q(\alpha, h(x), h(x))\}.$$

Определение 3.72. Пусть $\{L_1, \dots, L_k\}$ — множество литералов. Подстановку δ называют *унификатором* этого множества, если выполнено равенство $\delta(L_1) = \dots = \delta(L_k)$.

Множество называют *унифицируемым*, если существует унификатор этого множества.

Пример 3.89. Множество $S = \{P(a, y), P(x, f(b))\}$ унифицируемо подстановкой $\delta = \{x = a, y = f(b)\}$.

Если множество унифицируемо, то существует, как правило, не один унификатор этого множества, а несколько. Среди всех унификаторов данного множества выделяют *наиболее общий унификатор*.

Определение 3.73. Унификатор δ множества литералов называют *наиболее общим унификатором* этого множества, если для любого унификатора σ того же множества существует подстановка λ , такая, что $\sigma = \lambda \circ \delta$.

Пример 3.90. Подстановка в примере 3.89 — наиболее общий унификатор.

Рассмотрим алгоритм нахождения наиболее общего унификатора. Введем понятие множества рассогласований.

Определение 3.74. Пусть S — множество литералов. Определим первую слева позицию, в которой не для всех литералов стоит один и тот же символ. Затем в каждом литерале выпишем выражение, которое начинается символом, занимающим эту позицию. (Этими выражениями могут быть сам литерал, атомарная формула или терм). Множество полученных выражений называют *множеством рассогласований* в S .

Пример 3.91. Пусть $S = \{P(x, f(y)), P(x, \alpha), P(x, g(h(x)))\}$, тогда первая слева позиция, в которой не все литералы имеют один и тот же символ — пятая. Множество рассогласований есть $\{f(y), \alpha, g(h(x))\}$.

Алгоритм 3.16 (унификация множества литералов S).

Шаг 1. Полагаем $k = 0, S_k = S, \delta_k = \varepsilon$.

Шаг 2. Если множество S_k состоит из одного литерала, то δ_k — наиболее общий унификатор. Алгоритм заканчивает работу. Иначе ищем множество E_k рассогласований в S_k .

Шаг 3. Если в E_k существует переменная x_k и терм t_k , не содержащий x_k , то

а) полагаем $\delta_{k+1} = \{x_k = t_k\} \circ \delta_k$;

б) в множестве S_k выполняем замену $x_k = t_k$, полученное множество берем в качестве S_{k+1} ;

с) полагаем $k = k + 1$ и переходим к *шагу 2*.

В противном случае множество S неунифицируемо и алгоритм завершает работу.

Пример 3.92. Пусть $S = \{P(x, f(y), f(g(x))), P(\alpha, u, z)\}$. Протокол работы алгоритма 3.16 запишем в виде табл. 3.42.

Таблица 3.42

k	S_k	E_k	Подстановка
0	$\{P(x, f(y), f(g(x))), P(\alpha, u, z)\}$	$\{x, \alpha\}$	$x = \alpha$
1	$\{P(\alpha, f(y), f(g(\alpha))), P(\alpha, u, z)\}$	$\{u, f(y)\}$	$u = f(y)$
2	$\{P(\alpha, f(y), f(g(\alpha))), P(\alpha, f(y), z)\}$	$\{z, f(g(\alpha))\}$	$z = f(g(\alpha))$
3	$\{P(\alpha, f(y), f(g(\alpha)))\}$		

Множество S унифицируемо и наиболее общий унификатор

$$\delta_3 = \{x = \alpha, u = f(y), z = f(g(\alpha))\}.$$

Пример 3.93. Пусть $S = \{Q(f(\alpha), g(x)), Q(y, y)\}$. Протокол работы алгоритма 3.16 запишем в виде табл. 3.43.

Таблица 3.43

k	S_k	E_k	Подстановка
0	$\{Q(f(\alpha), g(x)), Q(y, y)\}$	$\{y, f(\alpha)\}$	$y = f(\alpha)$
1	$\{Q(f(\alpha), g(x)), Q(f(\alpha), f(\alpha))\}$	$\{g(x), f(\alpha)\}$	

Множество S неунифицируемо, так как множество рассогласований $E_1 = \{g(x), f(\alpha)\}$ не содержит переменной.

Замечание 3.44. Алгоритм 3.16 всегда заканчивает работу, так как S_{k+1} содержит на одну переменную меньше, чем S_k .

Обоснованием алгоритма 3.16 служит следующая теорема.

Теорема 3.35. Пусть S — множество литералов. Если S унифицируемо, то алгоритм 3.16 заканчивает работу на шаге 2 и подстановка δ_k есть наиболее общий унификатор множества S .

Доказательство. Рассмотрим σ — произвольный унификатор множества S . Покажем, что на каждом шаге k алгоритма 3.16 справедливо равенство

$$\sigma = \lambda_k \circ \delta_k, \quad (3.42)$$

где λ_k — некоторая подстановка. Доказательство проведем индукцией по k .

База индукции: $k = 0$. Тогда $\delta_0 = \varepsilon$ и $\sigma = \sigma \circ \varepsilon = \sigma \circ \delta_0$. В качестве подстановки λ_0 можно взять σ .

Шаг индукции: Предположим, что для $0 \leq k \leq m$, равенство (3.42) справедливо. Докажем его для $k = m + 1$.

Если S_m содержит один литерал, то на следующем проходе алгоритм остановится на шаге 2. Согласно (3.42), δ_m есть наиболее общий унификатор.

Пусть S_m содержит более одного литерала. Рассмотрим множество рассогласований E_m . Так как σ унификатор множества S и по индукционному предположению равенство (3.42) справедливо для $k = m$, то подстановка λ_m должна унифицировать множество E_m .

Поскольку E_m — унифицируемое множество рассогласований, в E_m существует хотя бы одна переменная x . Пусть терм $t \in E_m$, $t \neq x$. Множество E_m унифицируется подстановкой λ_m , поэтому $\lambda_m(x) = \lambda_m(t)$. Отсюда следует, что t не содержит x .

Не умаляя общности можно считать, что на шаге 3а алгоритма 3.16 для получения δ_{m+1} использовано равенство $x = t$, т. е.

$$\delta_{m+1} = \{x = t\} \circ \delta_m. \quad (3.43)$$

Из равенства $\lambda_m(x) = \lambda_m(t)$ следует, что $\{x = \lambda_m(t)\} \in \lambda_m$. Тогда

$$\lambda_{m+1} = \lambda_m \setminus \{x = \lambda_m(t)\}.$$

Так как t не содержит x , то $\lambda_{m+1}(t) = \lambda_m(t)$. Таким образом

$$\lambda_{m+1} \circ \{x = t\} = \lambda_{m+1} \cup \{x = \lambda_{m+1}(t)\} = \lambda_{m+1} \cup \{x = \lambda_m(t)\} = \lambda_m. \quad (3.44)$$

Подставив выражение для λ_m из (3.44) в (3.42) при $k = m$ с учетом (3.43), получаем

$$\sigma = \lambda_m \circ \delta_m = \lambda_{m+1} \circ \{x = t\} \circ \delta_m = \lambda_{m+1} \circ \delta_{m+1}.$$

Равенство (3.42) доказано для $k = m + 1$ и, следовательно, для любого k .

По условию теоремы множество S унифицируемо, следовательно, алгоритм должен закончить работу на шаге 2. Очевидно, что последняя подстановка δ_k будет унификатором множества S , а в силу (3.42) δ_k есть наиболее общий унификатор. ■

□ Метод резолюций

Определение 3.75. Правилом резолюций в логике предикатов называют правило: из дизъюнктов $\overline{Q(t_1, \dots, t_k)} \vee F$ и $Q(r_1, \dots, r_k) \vee G$ выводим дизъюнкт $\delta(F) \vee \delta(G)$, где δ — наиболее общий унификатор множества $\{Q(t_1, \dots, t_k), Q(r_1, \dots, r_k)\}$.

Дизъюнкт $\delta(F) \vee \delta(G)$ называют *бинарной резольвентой*.

Пример 3.94. С помощью правила резолюций из пары дизъюнктов $P(x) \vee Q(x)$ и $\overline{P(\alpha)} \vee R(y)$ подстановкой $\{x = \alpha\}$ можно вывести дизъюнкт $Q(\alpha) \vee R(y)$.

В отличие от логики высказываний, в логике предикатов существует еще одно правило.

Определение 3.76. Пусть дан дизъюнкт D , в котором два и более литералов с одинаковым знаком (*наличие или отсутствие отрицания*) имеют унифициатор δ . Тогда $\delta(D)$ называют *склейкой* дизъюнкта D .

Пример 3.95. Склейка дизъюнкта $\overline{P(x)} \vee \overline{P(f(y))} \vee Q(x, y)$ дает дизъюнкт $\overline{P(f(y))} \vee Q(f(y), y)$. При этом унифициатор $\delta = \{x = f(y)\}$.

Модифицируем определение вывода в логике предикатов с учетом нового правила склейки.

Определение 3.77. Пусть S — множество дизъюнктов. *Выводом* из множества дизъюнктов S называют последовательность дизъюнктов D_1, D_2, \dots, D_n , такая, что каждый дизъюнкт D_i или принадлежит S , или выводим либо из предыдущих дизъюнктов по правилу резолюций, либо из предыдущего дизъюнкта по правилу склейки.

Как и в логике высказываний, дизъюнкт D выводим из S , если существует вывод из S , последним дизъюнктом которого является D .

Для логики предикатов также справедлива теорема о полноте. Ее формулировка совпадает с формулировкой теоремы 3.12.

Теорема 3.36. Множество дизъюнктов S логики предикатов невыполнимо тогда и только тогда, когда из S выводим пустой дизъюнкт.

Доказательство теоремы 3.36 можно найти в [55].

Для решения вопроса о логическом следствии формулы G из множества формул $T = \{F_1, \dots, F_k\}$ метод резолюций в логике предикатов применяется так же, как и в логике высказываний.

Алгоритм 3.17 (метод резолюций).

Шаг 1. Применив алгоритм 3.15 из множества формул $T \cup \overline{G}$, получаем множество дизъюнктов S_T .

Шаг 2. Строим вывод пустого дизъюнкта из множества S_T .

Замечание 3.45. В отличие от логики высказываний, шаг 2 алгоритма 3.17 может никогда не завершиться при существовании бесконечного числа возможных резольвент, получаемых в процессе выполнения алгоритма. В 1936 г. А. Чёрчем и А. Тьюрингом было доказано, что не существует алгоритма для проверки невыполнимости множества формул логики предикатов.

Пример 3.96. Пусть

$$\begin{aligned} T &= \{F_1 = (\forall x)(E(x) \rightarrow V(x) \wedge R(x)), F_2 = (\exists x)(E(x) \wedge Q(x))\}, \\ G &= (\exists x)(Q(x) \wedge R(x)). \end{aligned}$$

Покажем, что $T \vDash G$. Используем алгоритм 3.15 для построения множества дизъюнктов. После сколемизации получаем

$$\begin{aligned}\widehat{F}_1 &= (\forall x) \left(\left(\overline{E(x)} \vee V(x) \right) \wedge \left(\overline{E(x)} \vee R(x) \right) \right); \\ \widehat{F}_2 &= E(\alpha) \wedge Q(\alpha), \quad \overline{G} = (\forall x) \left(\overline{Q(x)} \vee \overline{R(x)} \right).\end{aligned}$$

Множество дизъюнктов S_T имеет вид

$$\begin{aligned}S_T = \{D_1 &= \overline{E(x)} \vee V(x), D_2 = \overline{E(x)} \vee R(x), D_3 = E(\alpha), D_4 = Q(\alpha), \\ D_5 &= \overline{Q(x)} \vee \overline{R(x)}\}.\end{aligned}$$

Строим опровержение множества S_T .

$$D_4, D_5, D_6 \stackrel{D_4, D_5, \{x=\alpha\}}{=} \overline{R(\alpha)}, D_2, D_3, D_7 \stackrel{D_2, D_3, \{x=\alpha\}}{=} R(\alpha), D_8 \stackrel{D_6, D_7}{=} \square.$$

Пример 3.97. Установим логичность следующего рассуждения.

Таможенники обыскивают всякого, кто въезжает в страну, кроме высокопоставленных лиц. Некоторые люди, способствующие провозу наркотиков, въезжали в страну и были обысканы исключительно людьми, также способствовавшими провозу наркотиков. Никто из высокопоставленных лиц не способствовал провозу наркотиков. Следовательно, некоторые из таможенников способствовали провозу наркотиков.

Представим рассуждение в виде последовательности формул логики предикатов. Введем следующую сигнатуру.

$$\begin{aligned}E(x) - x &\text{ въезжал в страну;} & S(x, y) - x &\text{ обыскивает } y; \\ V(x) - x &\text{ - высокопоставленное лицо,} & G(x) - x &\text{ - таможенник;} \\ P(x) - x &\text{ способствовал провозу наркотиков.}\end{aligned}$$

Тогда рассуждение может быть представлено следующими формулами логики предикатов:

$$\begin{aligned}F_1 &= (\forall x) \left(E(x) \wedge \overline{V(x)} \rightarrow (\exists y)(G(y) \wedge S(y, x)) \right); \\ F_2 &= (\exists x)(E(x) \wedge P(x) \wedge (\forall y)(S(y, x) \rightarrow P(y))); \\ F_3 &= \overline{(\exists x)(V(x) \wedge P(x))}, \quad G = (\exists x)(G(x) \wedge P(x)).\end{aligned}$$

С помощью алгоритма 3.15 строим множество дизъюнктов для формул $\{F_1, F_2, F_3, \overline{G}\}$. После сколемизации получаем

$$\begin{aligned}\widehat{F}_1 &= (\forall x) \left(\left(\overline{E(x)} \vee V(x) \vee G(f(x)) \right) \wedge \left(\overline{E(x)} \vee V(x) \vee S(f(x), x) \right) \right); \\ \widehat{F}_2 &= (\forall y) \left(E(\alpha) \wedge P(\alpha) \wedge \left(\overline{S(y, \alpha)} \vee P(y) \right) \right), \quad \widehat{F}_3 = (\forall x) \left(\overline{V(x)} \vee \overline{P(x)} \right); \\ \widehat{\overline{G}} &= (\forall x) \left(\overline{G(x)} \vee \overline{P(x)} \right).\end{aligned}$$

Множество дизъюнктов S_T имеет вид

$$\begin{aligned} S_T = \{ & D_1 = \overline{E(x)} \vee V(x) \vee G(f(x)), D_2 = \overline{E(x)} \vee V(x) \vee S(f(x), x), \\ & D_3 = E(\alpha), D_4 = P(\alpha), D_5 = \overline{S(y, \alpha)} \vee P(y), D_6 = \overline{V(x)} \vee \overline{P(x)}, \\ & D_7 = \overline{G(x)} \vee \overline{P(x)} \}. \end{aligned}$$

Строим вывод пустого дизъюнкта в S_T .

$$\begin{aligned} D_2, D_3, D_8 &\stackrel{D_2, D_3, \{x=\alpha\}}{=} V(\alpha) \vee S(f(\alpha), \alpha), D_1, D_9 &\stackrel{D_1, D_3, \{x=\alpha\}}{=} V(\alpha) \vee G(f(\alpha)), \\ D_4, D_6, D_{10} &\stackrel{D_4, D_6, \{x=\alpha\}}{=} \overline{V(\alpha)}, D_{11} &\stackrel{D_9, D_{10}}{=} G(f(\alpha)), D_{12} &\stackrel{D_8, D_{10}}{=} S(f(\alpha), \alpha), \\ D_5, D_{13} &\stackrel{D_5, D_{12}, \{y=f(\alpha)\}}{=} P(f(\alpha)), D_7, D_{14} &\stackrel{D_7, D_{13}, \{x=f(\alpha)\}}{=} \overline{G(f(\alpha))}, D_{15} &\stackrel{D_{11}, D_{14}}{=} \square. \end{aligned}$$

Задача 3.32. Доказать логичность следующих рассуждений используя метод резолюций.

1. Все студенты университета – болельщики Зенита.
2. Некоторые студенты играют в футбол.
3. Следовательно, некоторые из болельщиков Зенита играют в футбол.

Введем следующие обозначения:

- a) $Z(x)$ – x болельщик Зенита;
- b) $S(x)$ – x студент университета;
- c) $F(x)$ – x играет в футбол.

Тогда вышеприведенные рассуждения можно представить формулами

$$\begin{aligned} F_1 &= (\forall x) (S(x) \rightarrow Z(x)), \quad F_2 = (\exists y) (S(y) \wedge F(y)), \\ G &= (\exists x) (Z(x) \wedge F(x)). \end{aligned}$$

Составим множество формул $\{F_1, F_2, \overline{G}\}$ и каждую из них приведем к сколемовской нормальной форме. Получим формулы:

$$\begin{aligned} H_1 &= (\forall x) \left(\overline{S(x)} \vee Z(x) \right), \quad H_2 = (S(\alpha) \wedge F(\alpha)), \\ H_3 &= (\forall x) \left(\overline{Z(x)} \vee \overline{F(x)} \right). \end{aligned}$$

Множество дизъюнктов S_T имеет вид:

$$S_T = \left\{ D_1 = \overline{S(x)} \vee Z(x), D_2 = S(\alpha), D_3 = F(\alpha), D_4 = \overline{Z(x)} \vee \overline{F(x)} \right\}.$$

Пустой дизъюнкт из множества S_T выводится очевидным образом.

Задача 3.33.

1. Существуют студенты, которые любят всех преподавателей.
2. Ни один из студентов не любит невежд.
3. Итак, ни один из преподавателей не является невеждой.

Введем сигнатуру:

- a) $S(x)$ — x студент;
- b) $T(x)$ — x преподаватель;
- c) $I(x)$ — x невежда;
- d) $L(x, y)$ — x любит y .

Запишем эти утверждения в виде формул логики предикатов:

$$\begin{aligned} F_1 &= (\exists x) (S(x) \wedge (\forall y) (T(y) \rightarrow L(x, y))), \\ F_2 &= (\forall x) \left(S(x) \rightarrow (\forall y) \left(I(y) \rightarrow \overline{L(x, y)} \right) \right), \\ G &= (\forall x) \left(T(x) \rightarrow \overline{I(x)} \right). \end{aligned}$$

С помощью алгоритма 3.15 строим множество дизъюнктов для формул $\{F_1, F_2, \overline{G}\}$. После сколемизации получаем:

$$\begin{aligned} \widehat{F}_1 &= (\forall y) \left(S(\alpha) \wedge \left(\overline{T(y)} \vee L(\alpha, y) \right) \right), \\ \widehat{F}_2 &= (\forall x)(\forall y) \left(\overline{S(x)} \vee \overline{I(y)} \vee \overline{L(x, y)} \right), \\ \widehat{\overline{G}} &= T(\alpha)I(\alpha). \end{aligned}$$

Множество дизъюнктов S_T имеет вид

$$\begin{aligned} S_T = \{D_1 &= S(\alpha), D_2 = \overline{T(y)} \vee L(\alpha, y), D_3 = \overline{S(x)} \vee \overline{I(y)} \vee \overline{L(x, y)}, \\ D_4 &= T(\alpha), D_5 = I(\alpha)\}. \end{aligned}$$

Строим вывод пустого дизъюнкта в S_T .

$$\begin{aligned} D_2, D_3, D_6 &\stackrel{D_2, D_3, \{x=\alpha\}}{=} \overline{S(\alpha)} \vee \overline{T(y)} \vee \overline{I(y)}, \\ D_1, D_7 &\stackrel{D_1, D_6}{=} \overline{T(y)} \vee \overline{I(y)}, D_4, D_8 &\stackrel{D_4, D_7, \{y=\alpha\}}{=} \overline{I(\alpha)}, \\ D_5, D_9 &\stackrel{D_5, D_8}{=} \square \end{aligned}$$

3.3.9. Метод резолюций и ПРОЛОГ

Историческая справка. В 1965 году в работе «A machine oriented logic based on the resolution principle» Дж. А. Робинсон (J.A.Pobinson) представил метод автоматического поиска доказательства теорем в логике предикатов, получивший название *принцип резолюции*. Идея данного метода была предложена Ж. Эрбаном в 1931 г., когда еще не было компьютеров. Робинсон модифицировал этот метод так, что он стал пригоден для автоматического, компьютерного использования, и, кроме того, разработал эффективный алгоритм унификации, составляющий основу его метода.

Этот метод послужил отправной точкой для создания в 1971 г. А. Колмеройер нового языка программирования со встроенной процедурой логического вывода — языка ПРОЛОГ (PROLOG). Название языка происходит от слов **ЛОГическое ПРОграммирование** (PROgramming in LOGic).

Определение 3.78. Дизъюнкт называют *хорновским*, если не более чем один литерал входит в него без отрицания. В ПРОЛОГ хорновские дизъюнкты называют обычно *предложениями* или *клозами*.

Дизъюнкты, в которых ровно один литерал без отрицания могут быть равносильным образом (использованные равносильности теоремы 3.4 указаны над равенством) преобразованы к виду

$$L \vee \overline{L}_1 \vee \dots \vee \overline{L}_n \stackrel{14,11}{=} L_1 \wedge \dots \wedge L_n \rightarrow L.$$

Такие дизъюнкты называют *правилами* и условно записываются так: $L_1, \dots, L_n \rightarrow L$. При этом дизъюнкт L называют *заголовком правила*.

Допустим случай, когда $n = 0$, тогда дизъюнкт имеет вид: L . Такие дизъюнкты называют *фактами*.

Если ни один литерал не входит в дизъюнкт без отрицания, то его равносильным образом можно преобразовать к виду

$$\overline{L}_1 \vee \dots \vee \overline{L}_n \stackrel{11}{=} \overline{L_1 \wedge \dots \wedge L_n}.$$

Подобные дизъюнкты записывают так: $? - L_1, \dots, L_n$ и называют *запросами*.

Множество хорновских дизъюнктов, содержащих ровно один запрос, называют *ПРОЛОГ-программой*, а множество хорновских дизъюнктов без запросов — *базой знаний*.

Замечание 3.46. Подобная постановка задачи с одной стороны, позволяет существенно упростить процедуру вывода методом резолюций, а с другой — достаточна для реализации большинства алгоритмических процедур.

ПРОЛОГ, как система программирования, состоит из языка хорновских дизъюнктов (в разных версиях синтаксис может быть различным) и интерпретатора, реализующего метод резолюций с той или иной стратегией выбора.

Наиболее часто используется линейная резолюция. Берется самый левый литерал цели (подцель) и первый унифицируемый с ним дизъюнкт. К нему применяется правило резолюции. Полученная резольвента добавляется в программу в качестве нового вопроса (основной дизъюнкт для линейной резолюции). И так до тех пор, пока не будет получен пустой дизъюнкт,

что будет означать успех, или до тех пор, пока очередную подцель будет невозможно унифицировать ни с одним дизъюнктом программы, что будет означать неудачу.

В последнем случае включается так называемый *бэктрекинг* – механизм возврата, который осуществляет откат программы к той точке, в которой выбирался унифицирующийся с последней подцелью дизъюнкт.

Пример 3.98. Рассмотрим следующую программу.

- 1: программирует(Иванов).
- 2: программирует(Петров).
- 3: читал(Иванов, Ирэ_Пол).
- 4: читал(Петров, Смешарики).
- 5: книга(Ирэ_Пол, программирование).
- 6: книга(Смешарики, детская).
- 7: сдавал(Иванов).
- 8: сдавал(Петров).
- 9: сдавал(X), программирует(X), знает(X) \rightarrow экзамен($X, 5$)
- 10: сдавал(X), программирует(X) \rightarrow экзамен($X, 4$)
- 11: сдавал(X) \rightarrow экзамен($X, 3$)
- 12: читал(X, Y), книга(Y , программирование) \rightarrow знает(X)
- 13: ? -- экзамен(Петров, Z).

При вычислении ответа на запрос, интерпретатор пытается унифицировать его с фактами. В нашем случае запрос не унифицируется ни с одним из фактов.

Тогда интерпретатор пытается унифицировать запрос с заголовком одного из правил. Это можно сделать с заголовком правила 9 с помощью подстановки $\{X = \text{Петров}, Z = 5\}$. Запрос принимает следующий вид

- 14: ? -- сдавал(Петров), программирует(Петров), знает(Петров)

Далее, первый литерал запроса 14 унифицируется с фактом 8, а второй с фактом 2. Последний дизъюнкт запроса 14 не унифицируется ни с одним из фактов. Унифицируем его с заголовком правила 12 следующей подстановкой $\{X = \text{Петров}\}$. Получаем запрос.

- 15: ? -- читал(Петров, Y), книга(Y , программирование)

Первый литерал запроса 15 унифицируется с фактом 4 подстановкой $\{Y = \text{Смешарики}\}$, но получившийся второй литерал не унифицируется ни с одним из фактов, ни с заголовками правил. Возвращаемся назад, к запросу 14. Ни один из его дизъюнктоов не имеет альтернативных вариантов унификации, поэтому возвращаемся еще на шаг назад к запросу 13. Правило 9 «отработано».

Интерпретатор пытается унифицировать запрос с заголовком одного из правил, следующих за правилом 9. Это можно сделать с заголовком правила 10 с помощью подстановки $\{X = \text{Петров}, Z = 4\}$. Запрос принимает вид

16: ? -- сдавал(Петров), программирует(Петров)

Как было показано, первый литерал запроса 16 унифицируется с фактом 8, а второй с фактом 2. Запрос пуст. Интерпретатор заканчивает работу. Ответ на запрос 13 положителен.

Интерес обычно представляет не столько сам факт успешного завершения программы, сколько конкретные значения переменных, при которых это возможно. В нашем случае $Z = 4$.

Замечание 3.47. Из определения 3.78 очевидно следует, что интерпретатор при поиске ответа на запрос строит вывод с помощью правила резолюций.

Пример 3.99. Рассмотрим следующие утверждения¹.

Ивана интересуют компьютеры, книги и автомобили. Петра интересует нечто, что интересует Ивана, но если это нечто является техникой и если это произведено в России. Известно, что компьютеры и автомобили – это техника. Кроме того, известно, что компьютеры производятся в Китае, а автомобили – в Америке и России. Вопрос: *Что интересует Петра?*

Формализуем задачу на Прологе.

- 1: интерес(Иван, компьютеры).
- 2: интерес(Иван, книги).
- 3: интерес(Иван, автомобили).
- 4: интерес(Иван, X), техника(X), произведено(X , Россия)
 \rightarrow интерес(Петр, X).
- 5: техника(компьютеры).
- 6: техника(автомобили).
- 7: произведено(компьютеры, Китай).
- 8: произведено(автомобили, Америка).
- 9: произведено(автомобили, Россия).

¹Пример взят с сайта Олега Акимова: <http://sceptic-ratio.narod.ru/index.htm>

10: ? -- интерес(Петр, X).

Первые три факта несопоставимы с целью; далее унифицируем запрос с заголовком правила. Запрос принимает следующий вид

11: ? -- интерес(Иван, X), техника(X), произведено(X , Россия)

Первый литерал запроса 11 унифицируется с фактом 1, а второй – с фактом 5 подстановкой $\{X = \text{компьютеры}\}$, однако третий литерал при данной подстановке не унифицируется ни с одним из фактов, ни с заголовками правил.

Пробуем унифицировать первый литерал запроса с фактом 2 подстановкой $\{X = \text{книги}\}$, но с этой подстановкой не унифицируется второй литерал. Переходим к факту 3. С подстановкой $\{X = \text{автомобили}\}$ происходит унификация всех литералов запроса. Запрос пуст и программа заканчивает работу. Ответ на запрос 10 положителен, *Петра интересуют автомобили*.

3.4. Нечеткие множества

Чем сложнее система, тем менее мы способны дать точные и в то же время имеющие практическое значение суждения об ее поведении.

– Лотфи Заде

3.4.1. Введение

Достаточно часто поиск оптимального решения практической задачи на основе классических методов математики затруднен. Причина заключается в проблеме осуществления корректного подбора приемлемого аналитического описания решаемой задачи. Но, даже в случае успешной реализации аналитического описания поставленной задачи ее решение может потребовать слишком больших ресурсов.

Как заметил фон Нейман, стремление получить точную, исчерпывающую модель для достаточно сложного объекта (процесса) не имеет смысла, поскольку сложность такого описания становится соизмеримой со сложностью самого объекта. Однако существует другой подход к решению проблемы.

Историческая справка. В 1965 году американский математик Лотфи Заде¹ (L. Zade) опубликовал статью «Нечеткие множества» (*Fuzzy sets*). Было дано новое определение понятия множества, предназначенное для описания сложных плохо определенных систем, в которых наряду с количественными данными присутствуют неоднозначные, субъективные, качественные данные. Эта статья породила новое научное направление, появились нечеткие отношения, нечеткая логика и т. д. Эти понятия в настороящее время широко используются в экспертных системах, системах искусственного интеллекта. Нечеткая логика входит в качестве составной части в понятие «мягкие вычисления» (*soft computing*), введенное Заде в 1994 году.

Началом практического применения теории нечетких множеств можно считать 1975 год, когда Мамдани и Ассилиан построили первый нечеткий контролер для управления простым паровым двигателем. Период с конца 80-х годов и до нашего времени характеризуется бумом практического применения теории нечеткой логики в разных сферах науки и техники. В начале 80-х европейские и американские инженерные и научные сообщества весьма скептически восприняли новую теорию². Зато на Востоке нечеткая логика пошла «на ура». Для людей, воспитанных на восточной философии, с ее неоднозначными и расплывчатыми категориями, нечеткая логика сразу стала своей, родной.

В бизнесе и финансах нечеткая логика получила признание после того, как в 1988 году экспертная система на основе нечетких правил для прогнозирования финансовых индикаторов **единственная** предсказала биржевой крах. В 1990 году Комитет по контролю экспорта США внес нечеткую логику в список критически важных оборонных технологий, не подлежащих экспорту потенциальному противнику.

За прошедшее время нечеткая логика прошла путь от почти антинаучной теории, практически отвергнутой в Европе и США, до ситуации конца девяностых годов, когда в Японии в широком ассортименте появились *нечеткие* бритвы, пылесосы, фотокамеры.

Сегодня нечеткая логика рассматривается как стандартный метод моделирования и проектирования. В 1997 г. язык нечеткого управления (*Fuzzy Control Language*) внесен в Международный стандарт программируемых контроллеров IEC 1131-7. Системы на нечетких множествах разработаны и успешно внедрены в таких областях, как: медицинская диагностика, техническая диагностика, финансовый менеджмент, управление персоналом, биржевое прогнозирование, распознавание образов, разведка ископаемых, выявление мошенничества, управление компьютерными сетями, управление технологическими процессами, управление транспортом, логистика, поиск информации в Интернете, радиосвязь и телевидение. Спектр приложений очень широкий – от бытовых видеокамер, пылесосов и стиральных машин до средств наведения ракет ПВО и управления боевыми вертолетами и самолетами.

В настоящее время в Японии это направление переживает настоящий бум. Здесь функционирует специально созданная лаборатория *Laboratory for International Fuzzy Engineering Research* (LIFE). Программой этой организации является создание более близких человеку вычислительных устройств. LIFE объединяет 48 компаний в числе которых находятся: *Hitachi, Mitsubishi, NEC, Sharp, Sony, Honda, Mazda, Toyota*. Из зарубежных участников LIFE можно выделить: *IBM, Fuji Xerox*, а также *NASA*³.

Как известно, одним из наиболее поразительным свойств человеческого интеллекта является способность принимать правильные решения в об-

¹Профессор технических наук Калифорнийского университета в Беркли.

²В конце 60-х годов работы Л. Заде даже были рассмотрены в Конгрессе США как яркий пример бессмысленной траты государственных средств, выделяемых на развитие науки.

³NASA стало использовать нечеткую логику в маневрах стыковки.

становке неполной и нечеткой информации. Построение моделей приближенных рассуждений человека и использование их в компьютерных системах будущих поколений представляет сегодня одну из важнейших проблем науки.

Нечеткая логика как раз и предназначена для формализации человеческих способностей к неточным или приближенным рассуждениям, которые позволяют более адекватно описывать ситуации с неопределенностью. Классическая логика по своей сути игнорирует проблему неопределенности, поскольку все высказывания и рассуждения в формальных логических системах могут иметь только два значения: 0 или 1. В отличие от этого в нечеткой логике истинность рассуждений оценивается в некоторой степени, которая может принимать и другие значения. Чтобы иметь возможность выражать неопределенные знания, необходима такая логическая система, которая позволяет некоторому высказыванию иметь истинностное значение, отличающееся от бинарного 0 или 1. Один из подходов – расширить множество истинностных значений и позволить высказываниям принимать некоторые дополнительные значения.

Одним из первых такой вариант многозначной логической системы предложил в 1930 г. польский математик Ян Лукасевич⁴. В логике Лукасевича используется три истинностных значения: {0, 0.5, 1}, где значение 0 интерпретируется как *ложь*, 1 – как *истина*, а число 0.5 – как *возможно*. В качестве высказываний с истинностным значением «возможно» могут выступать такие, которые относятся к некоторому моменту времени в будущем. Так, например, высказывание *сборная России по футболу выйдет в 1/8 финала на предстоящем чемпионате мира* до начала чемпионата не может быть оценено ни как истинное, ни как ложное. Именно по этой причине более адекватным ответом на вопрос об его истинности будет использование трехзначной логики с соответствующей интерпретацией истинности в форме значения «возможно».

Благодаря такому подходу, нечеткая логика обеспечивает разрешимость некоторых классически неразрешимых проблем. К примеру, хорошо известными являются древние парадоксы *sorites* (*куча*) и *falakros* (*лысый человек*).

Парадокс кучи: Одно пшеничное зерно не образует кучи. То же верно и для двух зерен, трех и т.д. Следовательно, куча не существует.

Парадокс лысого человека: Человек без волос или только с одним волосом – лысый. То же верно и для человека с двумя волосами и т.д. Следовательно, все люди – лысые.

⁴Ян Лукасевич – польский логик. Построил первую систему многозначной логики, а с её помощью – систему модальной логики. Разработал оригинальный язык для формализации логических выражений (уже встркавшаяся в курсе (3.9.6) *польская запись*, послужившая основой для более известной *обратной польской записи*).

Указанные парадоксы возникают тогда, когда свойство «быть кучей» и «быть лысым» понимаются точно, т.е. исключая их нечеткость. Классическая двузначная логика не способна с ними справиться. В рамках нечеткой логики подобные парадоксы не имеют места. Предлагаемое нечеткой логикой решение заключается в допущении, что импликация $S(x) \rightarrow S(x+1)$, где $S(x)$ означает, например, высказывание « x зерен не образуют кучи», истинна только в некоторой степени, близкой к 1 в зависимости от числа зерен x , скажем, $1 - \Delta(x)$, где $\Delta(x) > 0$. При этом допущении парадокс кучи, так же как и парадокс лысого, исчезает.

Замечание 3.48. Между прочим, в интернете существует специализированная поисковая система <http://nigma.ru/>, поисковый движок которой спроектирован на основе принципов нечеткого поиска. Выдаваемые этой поисковой системой результаты существенно отличаются от Yandex или Google. Иногда она находит документы, о существовании которых вышеуказанные поисковики даже не подозревают.

Перейдем к рассмотрению основ нечеткой логики, которая использует основные понятия теории нечетких множеств для формализации неточных знаний и выполнения приближенных рассуждений в той или иной проблемной области.

3.4.2. Основные понятия

Пусть M - некоторое универсальное множество¹ и A его произвольное подмножество. Для любого $x \in M$ верно $x \in A$ или $x \notin A$. На множестве M вводится *характеристическая функция* подмножества A :

$$\mu_A(x) = \begin{cases} 1, & x \in A, \\ 0, & x \notin A. \end{cases}$$

Множества, для которых характеристическая функция принимает только значения 0 или 1, называются *четкими*.

Пример 3.100. Пусть $M = \{a, b, c, d, e, f, g, h\}$ и $A = \{a, b, c, d\}$. Тогда функция $\mu_A(x)$ описана таблицей 3.44:

Таблица 3.44

x	a	b	c	d	e	f	g	h
$\mu_A(x)$	1	1	1	1	0	0	0	0

Очевидно, что для универсума $\mu_M(x) \equiv 1$, а для пустого множества $\mu_\emptyset(x) \equiv 0$. Для нечетких множеств характеристическая функция может принимать любые значения из промежутка $[0, 1]$.

¹В дальнейшем его будем называть «универсумом».

Пример 3.101. В условиях примера 3.100 определим $\mu_A(x)$ таблицей 3.45:

Таблица 3.45

x	a	b	c	d	e	f	g	h
$\mu_A(x)$	1.0	0.9	0.7	0.8	0.1	0.2	0.0	0.0

Для нечетких множеств характеристическую функцию называют *функцией принадлежности (Membership Function)*.

Определение 3.79. Нечетким множеством A , заданным на универсуме M , будем называть упорядоченное множество пар

$$A = \{(x, \mu_A(x)) \mid \forall x \in M\}$$

Используется также запись $x \in_{\mu_A(x)} A$. Например, $a \in_{1.0} A$, $b \in_{0.9} A$. Нечеткое множество A называется *пустым*, если

$$\mu_A(x) = 0, \forall x \in M.$$

Носителем нечеткого множества A , называется множество точек

$$S(A) = \{x \in M \mid \mu_A(x) > 0\}$$

Высотой нечеткого множества A называется величина

$$h(A) = \sup_{x \in A} \mu_A(x)$$

Нечеткое множество называется *нормальным*, если его высота равна единице. В противном случае нечеткое множество называется *субнормальным*.

Замечание 3.49. Субнормальное нечеткое множество всегда можно нормализовать, поделив функцию принадлежности μ_A на величину $h(A)$.

Ядром нечеткого множества A называется его четкое подмножество, элементы которого имеют степени принадлежности равные единице.

$$\text{core}(A) = \{x \in A \mid \mu_A(x) = 1\}$$

Замечание 3.50. Ядро субнормального нечеткого множества пустое.

Элементы, для которых $\mu_A(x) = 0.5$ называются *точками перехода* нечеткого множества A .

Четкое множество A^* , ближайшее к нечеткому множеству A , определяется следующим образом:

$$\mu_{A^*}(m) = \begin{cases} 0, & \text{если } \mu_A(m) < 0,5, \\ 1, & \text{если } \mu_A(m) > 0,5, \\ 0 \text{ или } 1, & \text{в противном случае.} \end{cases}$$

Кардинальным числом (мощностью) нечеткого n -элементного множества A называется величина

$$\text{card}(A) = |A| = \sum_{i=1}^n \mu_A(x_i). \quad (3.45)$$

Пример 3.102. Формализуем неточное определение «горячий чай». В качестве универсума M будет выступать шкала температуры в градусах Цельсия. Очевидно, что она будет изменяться от 0 до 100 градусов. Описание нечеткого множества для понятия «горячий чай» может выглядеть следующим образом (таблица 3.46):

Таблица 3.46

x	0	10	20	30	40	50	60	70	80	90	100
$\mu_A(x)$	0.0	0.0	0.0	0.15	0.30	0.50	0.80	0.90	1.0	1.0	1.0

Так, чай с температурой 60° принадлежит к множеству «горячий чай» со степенью принадлежности 0.80. Для одного человека чай при температуре 60° может оказаться горячим, для другого – не слишком горячим. Именно в этом и проявляется нечеткость задания соответствующего множества.

Нечеткое множество в примере является нормальным, его точка перехода – 50° .

Определение 3.80. Множество A_α будем называть *нечетким множеством α -уровня*, если оно образует совокупность таких элементов $x \in M$, степень принадлежности которых множеству A больше или равно $\alpha \in (0, 1]$.

$$A_\alpha = \{x \mid \mu_A(x) \geq \alpha, x \in M\}.$$

Пример 3.103. Пусть

$$A = \{(a, 0.2); (b, 0.8); (c, 0.5); (d, 0.1); (e, 0.25)\}.$$

Тогда

$$A_{0.25} = \{(b, 0.8); (c, 0.5); (e, 0.25)\}.$$

Замечание 3.51. Часто используются более компактная запись конечных или счетных нечетких множеств. Так, вместо приведенного выше в примере 3.103 представления множества A , его можно записать следующим образом:

$$A = 0.2/a + 0.8/b + 0.5/c + 0.1/d + 0.25/e. \quad (3.46)$$

В дальнейшем для представления нечетких множеств будем использовать обе записи.

Очевидно, справедливо следующее свойство.

Лемма 3.20. $A_{\alpha_1} \subseteq A_{\alpha_2}$ тогда и только тогда, когда $\alpha_1 \leq \alpha_2$.

Справедлива теорема о декомпозиции.

Теорема 3.37 (о декомпозиции). Любое нечеткое множество $A \subseteq M$ можно представить в виде

$$A = \bigcup_{\alpha \in [0,1]} \alpha A_\alpha, \quad (3.47)$$

где A_α означает нечеткое множество со следующей функцией принадлежности:

$$\mu_{\alpha A_\alpha}(x) = \begin{cases} \alpha, & x \in A_\alpha, \\ 0, & x \notin A_\alpha. \end{cases} \quad (3.48)$$

Разложение нечеткого множества в виде (3.47) называется *декомпозицией* нечеткого множества A .

Пример 3.104. Пусть

$$A = \{(a, 0.1); (b, 0.3); (c, 0.7); (d, 0.8); (e, 1.0)\}.$$

Тогда

$$A = 0.1 \times A_{0.1} \cup 0.3 \times A_{0.3} \cup 0.7 \times A_{0.7} \cup 0.8 \times A_{0.8} \cup 1.0 \times A_{1.0},$$

где

$$\begin{cases} A_{0.1} = \{(a, 0.1); (b, 0.1); (c, 0.1); (d, 0.1); (e, 0.1)\}, \\ A_{0.3} = \{(b, 0.3); (c, 0.3); (d, 0.3); (e, 0.3)\}, \\ A_{0.7} = \{(c, 0.7); (d, 0.7); (e, 0.7)\}, \\ A_{0.8} = \{(d, 0.8); (e, 0.8)\}, \\ A_{1.0} = \{(e, 1.0)\}. \end{cases}$$

3.4.3. Операции над нечеткими множествами

Определение 3.81. Определим для нечетких множеств отношения включения¹ (\subset , \subseteq) и равенства (=) следующим образом:

$$\begin{aligned} A \subset B &\iff \mu_A(x) < \mu_B(x), \forall x \in M, \\ A \subseteq B &\iff \mu_A(x) \leq \mu_B(x), \forall x \in M, \\ A = B &\iff \mu_A(x) = \mu_B(x), \forall x \in M. \end{aligned}$$

Пример 3.105. Пусть

$$A = \{(a, 0.0); (b, 0.1); (c, 0.5); (d, 0.9); (e, 1.0)\}.$$

Элемент a не принадлежит множеству A , элемент b принадлежит ему в малой степени, элемент c более или менее принадлежит, элемент d принадлежит в значительной степени, e является элементом A .

Определим для нечетких множеств, как и для обычных, основные логические операции. Пусть A и B – нечеткие подмножества одного универсума M .

Объединение множеств $A \cup B$ определяется функцией принадлежности

$$\mu_{A \cup B}(x) = \max\{\mu_A(x), \mu_B(x)\}, \forall x \in M.$$

Пример 3.106. Пусть нечеткие множества A и B определены следующим образом

$$\begin{aligned} A &= \{(a, 1.0); (b, 1.0); (c, 0.9); (d, 0.8); (e, 0.6); (f, 0.5); (g, 0.4); (h, 0.2)\}, \\ B &= \{(a, 0.5); (b, 1.0); (c, 0.6); (d, 0.4); (e, 0.2); (f, 0.0); (g, 0.0); (h, 0.0)\}. \end{aligned}$$

Тогда нечеткое множество $C = A \cup B$ будет равно:

$$C = \{(a, 1.0); (b, 1.0); (c, 0.9); (d, 0.8); (e, 0.6); (f, 0.5); (g, 0.4); (h, 0.2)\}.$$

Пересечение множеств $A \cap B$ определяется функцией принадлежности

$$\mu_{A \cap B}(x) = \min\{\mu_A(x), \mu_B(x)\}, \forall x \in M.$$

Пример 3.107. Рассмотрим нечеткие множества A и B из примера 3.106. Тогда нечеткое множество $D = A \cap B$ будет равно:

$$D = \{(a, 1.5); (b, 1.0); (c, 0.6); (d, 0.4); (e, 0.2); (f, 0.0); (g, 0.0); (h, 0.0)\}.$$

Разность множеств $A \setminus B$ определяется функцией принадлежности

$$\mu_{A \setminus B}(x) = \max\{\mu_A(x) - \mu_B(x), 0\}.$$

¹Иногда говорят, что B «доминирует» над A .

Пример 3.108. Рассмотрим нечеткие множества A и B из примера 3.106. Тогда нечеткое множество $E = A \setminus B$ будет равно:

$$E = \{(a, 0.5); (b, 0.0); (c, 0.3); (d, 0.4); (e, 0.4); (f, 0.5); (g, 0.4); (h, 0.2)\}.$$

Дополнение $\overline{A} = M \setminus A$ множества A определяется функцией принадлежности

$$\mu_{\overline{A}}(x) = 1 - \mu_A(x), \quad \forall x \in M.$$

Возведение в степень α множества A определяется функцией принадлежности

$$\mu_{A^\alpha}(x) = \mu_A^\alpha(x), \quad \forall x \in M.$$

Частными случаями возведения в степень являются $\alpha = 2$ – операция **концентрирования**², обозначается $\text{CON}(A)$ и $\alpha = 0.5$ – операция **размывания**, обозначается $\text{DIL}(A)$.

В результате применения операции концентрирования к множеству A снижается степень нечеткости описания, причем для элементов с высокой степенью принадлежности это уменьшение относительно мало, а для элементов с малой степенью принадлежности относительно велико. Операция размывания увеличивает степень нечеткости исходного нечеткого множества.

Пример 3.109. Пусть $A = \{(a, 0.3); (b, 0.6); (c, 1.0)\}$. Тогда

$$\begin{aligned} \text{CON}(A) &= \{(a, 0.09); (b, 0.36); (c, 1.0)\}, \\ \text{DIL}(A) &= \{(a, 0.548); (b, 0.775); (c, 1.0)\}. \end{aligned}$$

Операция *контрастной интенсификации*, обозначающаяся $\text{INT}(A)$ определяется следующим образом:

$$\mu_{\text{INT}(A)}(x) = \begin{cases} 2\mu_A(x)^2, & 0 \leq \mu_A(x) \leq 0.5, \\ 1 - 2(1 - \mu_A(x))^2, & 0.5 \leq \mu_A(x) \leq 1. \end{cases}$$

Эта операция отличается от концентрирования тем, что она увеличивает те значения $\mu_A(x)$, которое больше 0.5 и уменьшает те, которые меньше 0.5. Таким образом, контрастная интенсификация уменьшает нечеткость множества A .

Пример 3.110. Определим нечеткое множество A – множество вещественных чисел, близких к числу π , характеристической функцией

$$\mu_A(x) = \frac{1}{(1 + |x - \pi|)^m}, \quad m \geq 1.$$

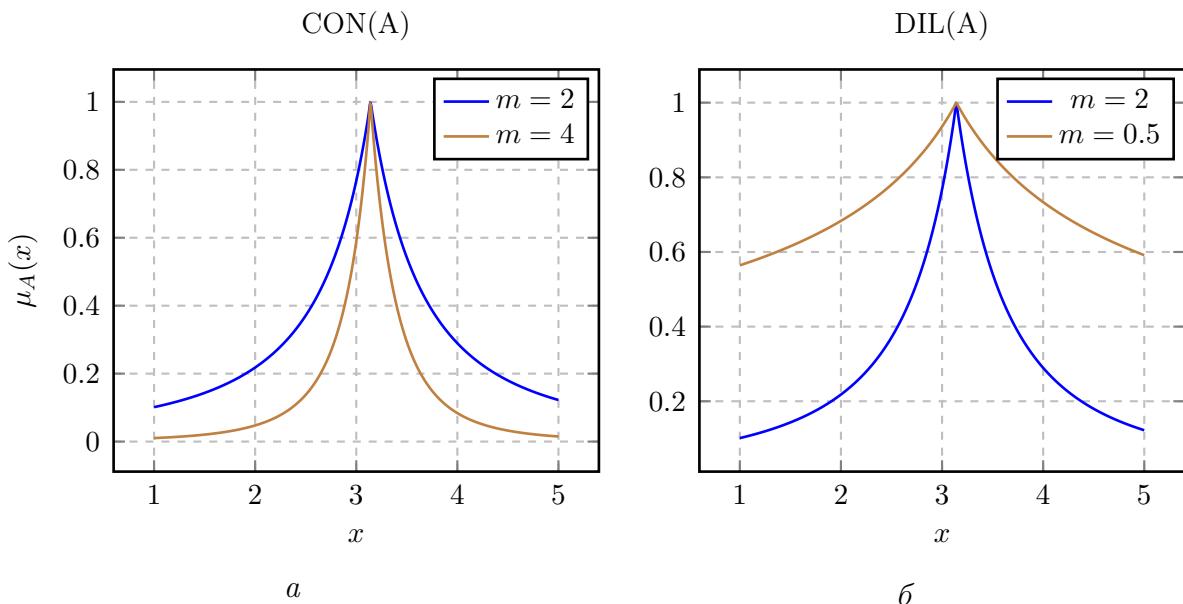


Рис. 3.24

Например, для описания множества чисел, **не очень далеких** от π , положить $m = 2$. Для описания множества чисел, **очень близких** к π , можно провести концентрирование предыдущей функции принадлежности, то есть положить $m = 4$ (см. рис. 3.24 а). Увеличение степени нечеткости исходного нечеткого множества показано на рис. 3.24 б.

Для введенных операций над нечеткими множествами справедливы **почти все** законы классической теории множеств (см. табл. 3.47):

Таблица 3.47

$A \cup (A \cap B)$	поглощение
$A \cap (A \cup B)$	
$A \cup A = A$	идемпотентность
$A \cap A = A$	
$A \cup B = B \cup A$	коммутативность
$A \cap B = B \cap A$	
$A \cup (B \cup C) = (A \cup B) \cup C$	ассоциативность
$A \cap (B \cap C) = (A \cap B) \cap C$	
$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$	дистрибутивность
$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$	
$A \cup \bar{B} = \bar{A} \cap \bar{B}$	законы де Моргана

²Аналоги в естественном языке «очень» и «не очень».

Продолжение табл. 3.47

$\overline{A \cap B} = \overline{A} \cup \overline{B}$	
$\overline{\overline{A}} = A$	двойное дополнение

Для нечетких множеств **не справедливы** законы противоречия и исключенного третьего:

$$A \cap \overline{A} \neq \emptyset, \quad A \cup \overline{A} \neq M. \quad (3.49)$$

Для доказательства (3.49) рассмотрим пример.

Пример 3.111. Рассмотрим универсум $M = \{a, b, c, d, e, f\}$. Пусть нечеткое множество A имеет вид

$$A = \{(a, 1.0); (b, 0.8); (c, 0.6); (d, 0.4); (e, 0.2); (f, 0.0)\}.$$

Тогда, по определению, множество \overline{A} имеет следующий вид

$$\overline{A} = \{(a, 0.0); (b, 0.2); (c, 0.4); (d, 0.6); (e, 0.8); (f, 1.0)\}.$$

Находим пересечение нечетких множеств

$$A \cap \overline{A} = \{(a, 0.0); (b, 0.2); (c, 0.4); (d, 0.4); (e, 0.2); (f, 0.0)\} \neq \emptyset.$$

Аналогично, находим объединение нечетких множеств

$$A \cup \overline{A} = \{(a, 1.0); (b, 0.8); (c, 0.6); (d, 0.6); (e, 0.8); (f, 1.0)\} \neq M.$$

Для нечетких множеств часто вводят понятие дизъюнктивной суммы.

Определение 3.82. *Дизъюнктивная сумма* нечетких множеств A и B , заданных на универсальном множестве M , – это нечеткое множество

$$A \oplus B = (A \setminus B) \cup (B \setminus A) = (A \cap \overline{B}) \cup (\overline{A} \cap B)$$

с функцией принадлежности, заданной следующим образом:

$$\mu_{A \oplus B}(x) = \max \{\min \{\mu_A(x); 1 - \mu_B(x)\}; \min \{1 - \mu_A(x); \mu_B(x)\}\}$$

Пример 3.112. Пусть

$$A = 0.9/a + 1/b + 0.6/d, \quad B = 0.7/a + 1/c + 0.4/d.$$

Тогда

$$A \oplus B = 0.97/a + 1/b + 1/c + 0.76/d.$$

Упражнение 3.22. Пусть задано n нечетких множеств $A_i, i \in 1 : n$ универсального множества M . Обозначим

$$B = A_1 \cup \cdots \cup A_n, \quad C = A_1 \cap \cdots \cap A_n, \quad i \in 1 : n.$$

Показать, что

$$\mu_B(x) = \max \{\mu_{A_1}(x), \dots, \mu_{A_n}(x)\}, \quad \mu_C(x) = \min \{\mu_{A_1}(x), \dots, \mu_{A_n}(x)\}.$$

Определение 3.83. Пусть A_1, \dots, A_n – нечеткие множества универсальных множеств M_1, \dots, M_n соответственно, а $\omega_1, \dots, \omega_n$ – неотрицательные числа, такие, что $\sum_{i=1}^n \omega_i = 1$. Выпуклой комбинацией нечетких множеств A_1, \dots, A_n называется нечеткое подмножество A множества $M = M_1 \times \cdots \times M_n$ с n -мерной функцией принадлежности:

$$\mu_A(x_1, \dots, x_n) = \omega_1 \mu_{A_1}(x_1) + \cdots + \omega_n \mu_{A_n}(x_n) = \sum_{i=1}^n \omega_i \mu_{A_i}(x_i).$$

Пример 3.113. Пусть для универсальных множеств $M_1 = \{a, b\}$ и $M_2 = \{b, c\}$ определены нечеткие множества A_1 и A_2 :

$$A_1 = 0.5/a + 1/b, \quad A_2 = 0.2/b + 1.0/c, \quad \omega_1 = 0.4, \omega_2 = 0.6.$$

Тогда

$$\begin{aligned} A &= (0.5 \times 0.4 + 0.2 \times 0.6)/(a; b) + (0.5 \times 0.4 + 1.0 \times 0.6)/(a; c) + \\ &\quad + (1.0 \times 0.4 + 0.2 \times 0.6)/(b; b) + (1.0 \times 0.4 + 1.0 \times 0.6)/(b; c) = \\ &= 0.32/(a; b) + 0.8/(a; c) + 0.52/(b; b) + 1.0/(b; c). \end{aligned}$$

Рассмотрим возможность увеличения нечеткости множества. Для этого предварительно определим операцию умножения нечеткого множества на число.

Пусть A – нечеткое множество, определенное на универсальном множестве M .

Определение 3.84. Рассмотрим число $\alpha > 0$, такое, что

$$\alpha \max_{x \in M} \{\mu_A(x)\} \leq 1,$$

тогда функция принадлежности нечеткого множества αA определяется как

$$\mu_{\alpha A}(x) = \alpha \mu_A(x), \quad \forall x \in M.$$

Определение 3.85. Пусть для всех $x \in M$ определены нечеткие множества $K(x)$. Совокупность всех $K(x)$ называется *ядром оператора Φ увеличения нечеткости*. Результатом действия оператора Φ на нечеткое множество A является нечеткое множество вида:

$$\Phi(A, K) = \bigcup_{x \in M} \mu_A(x)K(x).$$

Пример 3.114. Пусть

$$\begin{aligned} M &= \{a, b, c, d\}, \quad A = 0.8/a + 0.6/b + 0.2/c + 0.0/d, \\ K(a) &= 1.0/a + 0.4/b, \quad K(b) = 1.0/b + 0.4/c + 0.4/d, \\ K(c) &= 1.0/c + 0.5/d, \quad K(d) = 1.0/d. \end{aligned}$$

Тогда

$$\begin{aligned} \Phi(A, K) &= \mu_A(a)K(a) \cup \mu_A(b)K(b) \cup \mu_A(c)K(c) \cup \mu_A(d)K(d) = \\ &= 0.8(1.0/a + 0.4/b) \cup 0.6(1.0/b + 0.4/c + 0.4/d) \cup \\ &\quad \cup 0.2(1.0/c + 0.5/d) \cup 0.0(1.0/d) = \\ &= (0.8/a + 0.32/b) \cup (0.6/b + 0.24/c + 0.24/d) \cup \\ &\quad \cup (0.2/c + 0.1/d) \cup (0.0/d) = \\ &= 0.8/a + 0.6/b + 0.24/c + 0.24/d. \end{aligned}$$

3.4.4. Мера нечеткости множеств

Формализация понятия меры нечеткости множества имеет важное практическое и теоретическое значение.

Для определения степени нечеткости множества Р. Егер ввел определение меры нечеткости, сводящейся к измерению уровня различия между множеством A и его дополнением \bar{A} . Он исходил из того, что единственным коренным отличием алгебры нечетких множеств от обычной булевой алгебры является непустота пересечения множества A и \bar{A} , то есть для нечетких множеств

$$A \cap \bar{A} = B \neq \emptyset.$$

Очевидно, что чем ближе A к \bar{A} , тем больше B , и тем сильнее A отличается от четкого множества. На основании этого Р. Егер предложил для описания меры четкости нечеткого множества A следующее выражение.

Определение 3.86. *Мера Егера* степени нечеткости n -элементного множества A в метрике ρ , обозначаемая $FUZ_\rho(A)$, определяется выражением

$$FUZ_\rho(A) = 1 - \frac{D_\rho(A, \bar{A})}{n^{1/\rho}}, \quad (3.50)$$

где $D_p(A, \bar{A})$ – это мера расстояния между множествами A и \bar{A} , содержащими n элементов.

Значение $\rho = 1$ соответствует метрике Хемминга, в которой

$$D_1(A, B) = \sum_{i=1}^n |\mu_A(x_i) - \mu_B(x_i)|, \quad D_1(A, \bar{A}) = \sum_{i=1}^n |2\mu_A(x_i) - 1|, \quad (3.51)$$

а значение $\rho = 2$ соответствует метрике Евклида, в которой

$$D_2(A, B) = \sqrt{\sum_{i=1}^n (\mu_A(x_i) - \mu_B(x_i))^2}, \quad D_2(A, \bar{A}) = \sqrt{\sum_{i=1}^n (2\mu_A(x_i) - 1)^2}. \quad (3.52)$$

Пример 3.115. Пусть

$$A = \{(a, 0.1); (b, 0.5); (c, 0.8); (d, 1.0); (e, 0.8); (f, 0.5); (g, 0.1)\}.$$

Тогда

$$\bar{A} = \{(a, 0.9); (b, 0.5); (c, 0.2); (d, 0.0); (e, 0.2); (f, 0.5); (g, 0.9)\}.$$

В соответствии с определением меры Егера (3.50), используя (3.51) и (3.52) получаем

$$FUZ_1(A) = 1 - \frac{1}{7}(0.8 + 0.0 + 0.6 + 1.0 + 0.6 + 0.0 + 0.8) = 0.457,$$

$$FUZ_2(A) = 1 - \sqrt{\frac{(0.64 + 0.0 + 0.36 + 1.0 + 0.36 + 0.0 + 0.64)}{7}} = 0.347.$$

Другую меру нечеткости предложил Б. Коско. Она основана на понятии кардинального числа множества (3.45). В соответствии с этой мерой:

$$FUZ(A) = \frac{\text{card}(A \cap \bar{A})}{\text{card}(A \cup \bar{A})}. \quad (3.53)$$

Пример 3.116. Для множества из примера 3.115 получаем меру Коско, равную:

$$FUZ(A) = \frac{0.1 + 0.5 + 0.2 + 0.0 + 0.2 + 0.5 + 0.1}{0.9 + 0.5 + 0.8 + 1.0 + 0.8 + 0.5 + 0.9} = \frac{1.6}{5.4} = 0.296.$$

Замечание 3.52. Заметим, что обе меры для четких множеств дают один и тот же нулевой результат, так как в мере Коско $\text{card}(A \cap \bar{A}) = 0$, а для меры Егера $D_p(A, \bar{A}) = \sqrt[n]{n}$ и вследствие 3.50 дает в результате также $FUZ_p(A) = 0$.

3.4.5. Функции принадлежности

Введенное определение 3.79 для нечеткого множества не накладывает ограничений на выбор функции принадлежности. Однако, на практике целесообразно использовать аналитическое представление функции принадлежности $\mu_A(x)$ нечеткого множества A .

Существует свыше десятка типовых форм кривых для задания функций принадлежности. Выделяют следующие типовые функции принадлежности.

Треугольная функция принадлежности определяется тройкой чисел (a, b, c) , $a < b < c$, и ее значение в точке x вычисляется согласно выражению:

$$\mu_A(x) = \begin{cases} 1 - \frac{b-x}{b-a}, & a \leq x \leq b, \\ 1 - \frac{x-b}{c-b}, & b \leq x \leq c, \\ 0, & \text{в остальных случаях.} \end{cases}$$

При $(b-a) = (c-b)$ имеем случай симметричной треугольной функции принадлежности, которая может быть однозначно задана двумя параметрами из тройки (a, b, c) . На рис 3.25 а приведен график треугольной функции принадлежности при $a = 25, b = 50, c = 75$.

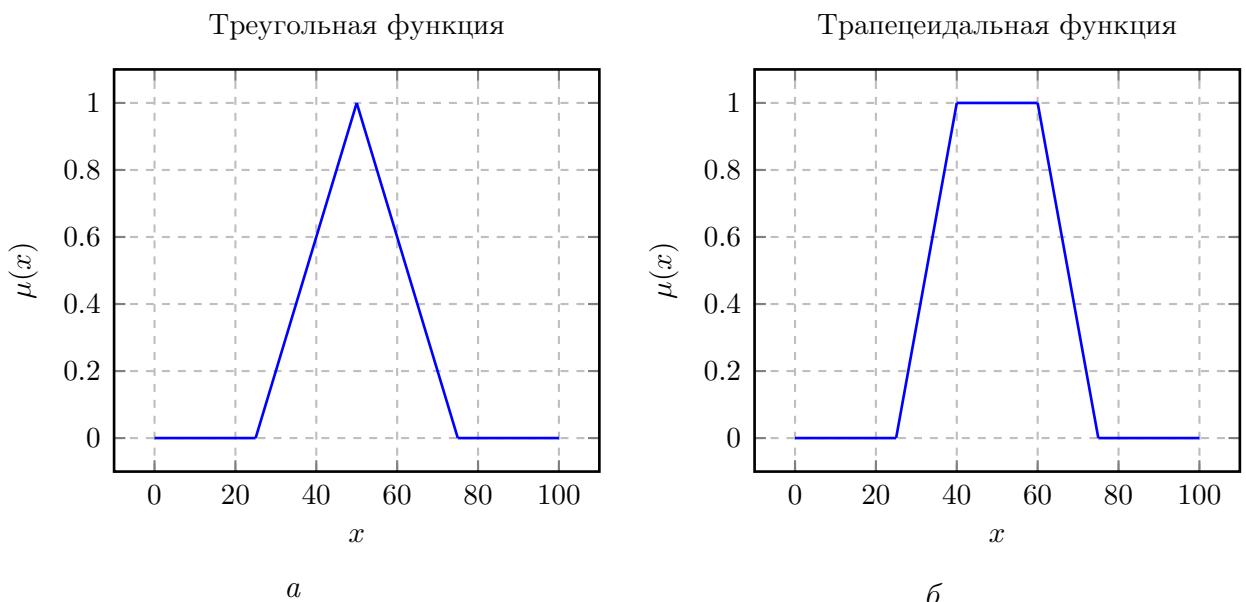


Рис. 3.25

Трапецидальная функция принадлежности. Аналогично предыдущему случаю для задания трапецидальной функции принадлежности

необходима четверка чисел (a, b, c, d) , $a < b < c < d$:

$$\mu_A(x) = \begin{cases} 1 - \frac{b-x}{b-a}, & a \leq x \leq b, \\ 1, & b \leq x \leq c, \\ 1 - \frac{x-c}{d-c}, & c \leq x \leq d, \\ 0, & \text{в остальных случаях.} \end{cases}$$

На рис 3.25 б приведен график трапецидальной функции принадлежности при $a = 25, b = 40, c = 60, d = 75$.

Треугольные и трапецидальные функции принадлежности как правило используются для задания неопределенностей типа: «приблизительно равно», «среднее значение», «расположен в интервале», «подобен объекту», «похож на предмет» и т. п.

Гауссова функция принадлежности. Функция принадлежности гауссова типа описывается выражением

$$\mu_A(x) = e^{-\left(\frac{x-c}{\sigma}\right)^2}$$

и зависит от двух параметров. Параметр c обозначает центр нечеткого множества, а параметр σ отвечает за крутизну функции. На рис 3.26 а приведен график гауссовой функции принадлежности при $c = 50, \sigma = 12.5$.

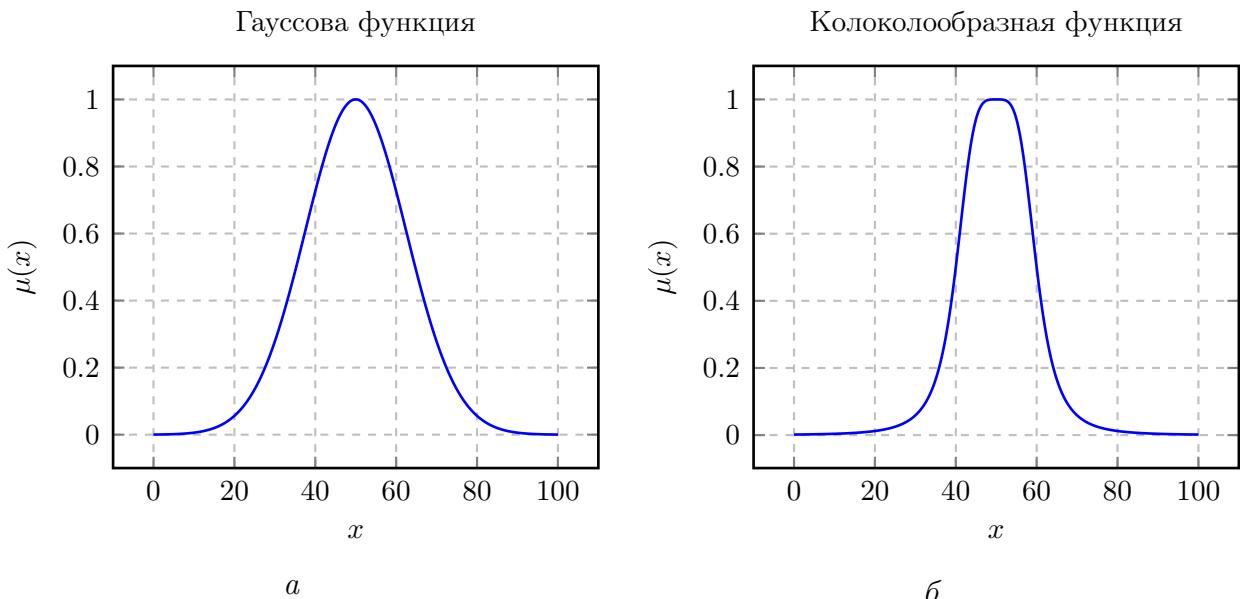


Рис. 3.26

К гауссовым функциям принадлежности часто относят также так называемую *колоколообразную (bell-shaped) функцию*, которая в общем случае задается аналитически следующим выражением:

$$\mu_A(x) = \frac{1}{1 + \left| \frac{x - c}{a} \right|^{2b}},$$

где a, b, c – числовые параметры, принимающие произвольные вещественные значения и упорядоченные отношением параметров: $a < b < c$, $b > 0$. Параметры геометрически интерпретируются следующим образом: a – коэффициент концентрации функции принадлежности, b – коэффициент крутизны функции принадлежности, c – координата максимума функции принадлежности.

На рис 3.26 b приведен график колоколообразной функции принадлежности при $a = 10, b = 2, c = 50$.

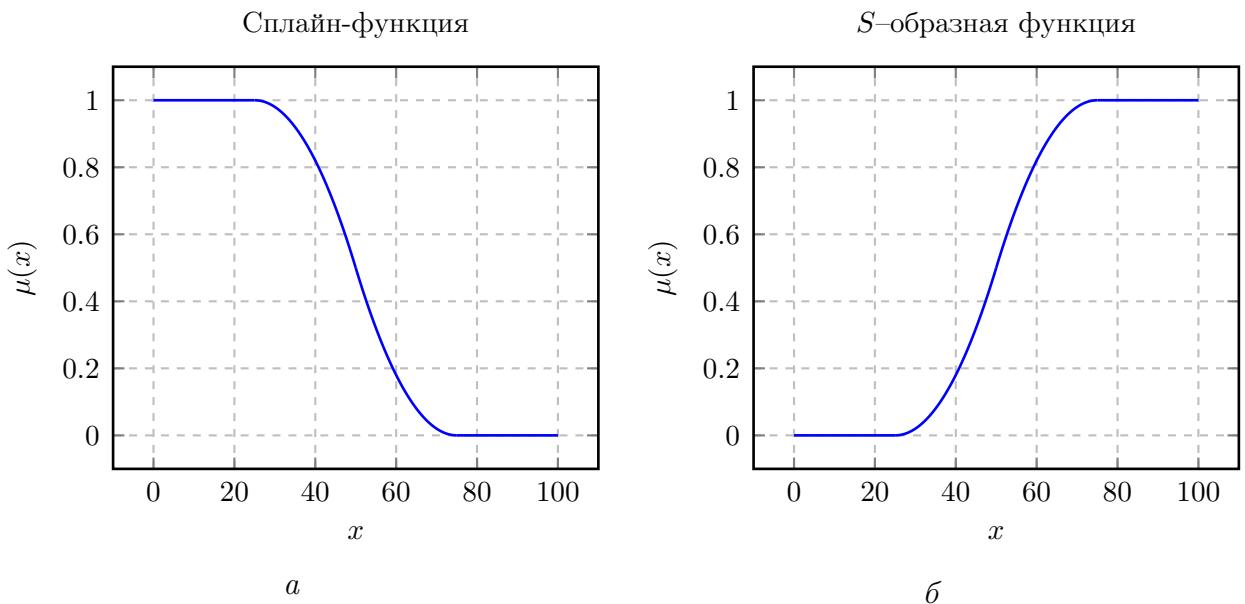
Сплайн-функция принадлежности. Сплайн-функция (иногда ее называют *Z-образной функцией*) принадлежности в общем случае аналитически задается следующим выражением:

$$\mu_A(x) = \begin{cases} 1, & x < a, \\ \frac{1}{2} + \frac{1}{2} \cos \left(\frac{x - a}{b - a} \pi \right), & a \leq x \leq b, \\ 0, & x > b, \end{cases}$$

где a, b – некоторые числовые параметры, принимающие произвольные вещественные значения и упорядоченные отношением: $a < b$. На рис 3.27 а приведен график сплайн-функции принадлежности при $a = 25, b = 75$.

Сплайн-функции принадлежности, используются для задания неопределенностей типа: «малое количество», «небольшое значение», «незначительная величина», «низкий уровень» и т.п.

S-образная функция принадлежности. S-образная функция принадлежности в общем виде может быть задана аналитически следующим



Puc. 3.27

выражением:

$$\mu_A(x) = \begin{cases} 0, & a \leq x, \\ 1 - 2 \left(\frac{x-a}{b-a} \right)^2, & a < x \leq \frac{a+b}{2}, \\ 2 \left(\frac{b-x}{b-a} \right)^2, & \frac{a+b}{2} < x < b, \\ 1, & x \geq b, \quad a < b \end{cases}$$

S-образные функции принадлежности применяются для задания неопределенностей типа: «большое количество», «большое значение», «значительная величина», «высокий уровень» и т. п. На рис 3.27 в приведен график *S*-образной функции принадлежности при $a = 25, b = 75$.

К типу S -образных и одновременно Z -образных функций принадлежности может быть отнесена так называемая *сигмоидальная* функция (сигмоид), которая в общем случае задается аналитически следующим выражением:

$$\mu_A(x) = \frac{1}{1 + e^{-a(x-b)}}$$

При этом в случае $a > 0$ может быть получена S –образная функция принадлежности, а в случае $a < 0$, соответственно – Z –образная функция принадлежности.

Существует множество других функций принадлежности нечетких множеств, заданных как композиции вышеупомянутых базовых функций

(двойная гауссова, двойная сигмоидальная и т.п.), либо как комбинации по участкам возрастания и убывания (сигмоидально-гауссова, сплайн-треугольная и т. п.)

3.4.6. Методы построения функций принадлежности

Наиболее распространенными методами построения функций принадлежности являются прямые (*метод прямого оценивания*) и косвенные (*обратный метод оценивания*) методы. При использовании прямого метода построения каждому $x \in X$ экспертом задаётся значение функции принадлежности $\mu_A(x)$.

Прямые методы характеризуются тем, что эксперт непосредственно задает правила определения значений функции принадлежности $\mu_A(x)$, характеризующей элемент x . Эти значения согласуются с его предпочтениями на множестве элементов M следующим образом:

- 1) для любых $x_1, x_2 \in M$, $\mu_A(x_1) < \mu_A(x_2)$ тогда и только тогда, когда x_2 предпочтительнее x_1 , т.е. в большей степени характеризуется свойством A ;
- 2) для любых $x_1, x_2 \in M$, $\mu_A(x_1) = \mu_A(x_2)$ тогда и только тогда, когда x_1 и x_2 безразличны относительно свойства A .

Прямые методы построения удобны при решении задач, для которых свойства физических величин могут быть измерены, например, скорость, время, расстояние, давление и т. д. Задание абсолютно точных значений функций принадлежности не требуется, достаточно определить тип функции принадлежности и характерные значения множества. При необходимости более точного определения функции принадлежности можно воспользоваться последующим анализом результатов для коррекции нечеткой модели.

Косвенные методы построения функций принадлежности используются при решении задач, для которых свойства физических величин не могут быть измерены. Эти методы основаны на экспертных оценках. В качестве простейшего рассмотрим следующий метод, называемый *частотным*.

Пусть имеется n экспертов. Допустим, что на вопрос о том, принадлежит ли элемент x нечеткому множеству A , $n_1 \leq n$ экспертов отвечают положительно. В этом случае полагаем

$$\mu_A(x) = \frac{n_1}{n}.$$

Наибольшее распространение среди косвенных методов получил метод *парных сравнений* или *парных соотношений*.

Метод основан на обработке матрицы оценок, отражающих мнение эксперта об относительной принадлежности элементов множеству или степени выраженности у них некоторого оцениваемого свойства. Потребуем, чтобы для всех элементов множества A для определяемой функции μ_A выполнялось равенство:

$$\sum_{i=1}^n \mu_A(x_i) = 1. \quad (3.54)$$

Оценку преимущества элемента x_i над элементом x_j с точки зрения свойства A обозначим через a_{ij} . Для обеспечения согласованности примем $a_{ij} = 1/a_{ji}$. Оценки a_{ij} составляют матрицу парных сравнений $S = \|a_{ij}\|$. По построению, матрица S является диагональной ($a_{ii} = 1$) и обратно симметричной.

Замечание 3.53. Оценки a_{ij} как правило определяются по девятибалльной шкале Саати (табл. 3.48):

Таблица 3.48

a_{ij}	
1	Отсутствует преимущество элемента x_i над элементом x_j
3	Имеется слабое преимущество x_i над x_j
5	Имеется существенное преимущество x_i над x_j
7	Имеется явное преимущество x_i над x_j
9	Имеется абсолютное преимущество x_i над x_j

Значения 2, 4, 6, 8 – промежуточные сравнительные оценки.

Далее найдем $W = (w_1, \dots, w_n)$ – нормированный собственный вектор матрицы S , соответствующий максимальному собственному числу λ_{max} :

$$SW = \lambda_{max} W, \quad w_1 + \dots + w_n = 1. \quad (3.55)$$

Положим $\mu_A(x_i) = w_i$, $i \in 1 : n$. В силу (3.55) условие (3.54) на μ_A будет выполнено.

Замечание 3.54. Большое число численных примеров для этого метода можно найти в [44].

3.4.7. Задачи для самостоятельного решения

Задача 3.34. Определить нечеткое множество как результат операции объединения двух нечетких множеств A и B , заданных следующим

образом:

$$\begin{aligned} A &= \{(a, 1.0); (b, 1.0); (c, 0.9); (d, 0.9); (e, 0.8); (f, 0.7); (g, 0.4); (h, 0.2)\}, \\ B &= \{(a, 0.9); (b, 1.0); (c, 0.7); (d, 0.6); (e, 0.2); (f, 0.1); (g, 0.3); (h, 0.1)\}. \end{aligned}$$

Задача 3.35. Определить нечеткое множество как результат операции пересечения двух нечетких множеств A и B , заданных следующим образом:

$$\begin{aligned} A &= \{(a, 0.4); (b, 0.2); (c, 0.4); (d, 0.9); (e, 0.8); (f, 0.7); (g, 0.4); (h, 0.2)\}, \\ B &= \{(a, 0.7); (b, 1.0); (c, 1.0); (d, 1.0); (e, 1.0); (f, 0.2); (g, 0.6); (h, 0.1)\}. \end{aligned}$$

Задача 3.36. Определить нечеткое множество как результат операции разности двух нечетких множеств A и B , заданных следующим образом:

$$\begin{aligned} A &= \{(a, 1.0); (b, 1.0); (c, 0.9); (d, 0.8); (e, 0.6); (f, 0.5); (g, 0.4); (h, 0.2)\}, \\ B &= \{(a, 0.7); (b, 0.9); (c, 0.6); (d, 0.7); (e, 0.2); (f, 0.0); (g, 0.1); (h, 0.1)\}. \end{aligned}$$

Задача 3.37. Дано универсальное множество

$$M = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}$$

и нечеткие множества $A \subseteq M$ и $B \subseteq M$:

$$\begin{aligned} A &= \{(x_1, 1.0); (x_2, 0.1); (x_3, 0.2); (x_4, 0.3); (x_5, 0.4); (x_6, 0.5); (x_7, 0.2)\}, \\ B &= \{(x_1, 0.7); (x_2, 0.5); (x_3, 0.4); (x_4, 0.8); (x_5, 0.2); (x_6, 0.5); (x_7, 1.0)\}. \end{aligned}$$

Выполнить операции объединения, пересечения, дополнения, разности и симметрической разности.

Задача 3.38. Дано универсальное множество $M = \{x_1, x_2, x_3, x_4\}$ и нечеткие множества $A \subseteq M$ и $B \subseteq M$:

$$A = 0.3/x_1 + 0.1/x_2 + 1.0/x_3 + 0/x_4, \quad B = 0.6/x_1 + 0.8/x_2 + 1.0/x_3 + 0.9/x_4.$$

Найти $A \oplus B$.

Задача 3.39. Пусть $A = \{(a, 0.7); (b, 0.24); (c, 0.8); (d, 0.25)\}$. Найти $\text{CON}(A)$ и $\text{DIL}(A)$

Задача 3.40. Пусть

$$\begin{aligned} M &= \{x_1, x_2, x_3, x_4\}, \quad A = 0.8/x_1 + 0.6/x_2 + 0.0/x_3 + 0.0/x_4, \\ K(x_1) &= 1.0/x_1 + 0.4/x_2, \quad K(x_2) = 1.0/x_2 + 0.4/x_3 + 0.4/x_4, \\ K(x_4) &= 1.0/x_3 + 0.5/x_4, \quad K(x_4) = 1.0/x_4. \end{aligned}$$

Найти $\Phi(A, K)$.

3.5. Нечеткие бинарные отношения

3.5.1. Основные понятия

Определение 3.87. Перенося определение нечетких множеств на отношения, определим бинарное нечеткое отношение как нечеткое подмножество $R \subseteq M \times M$. Таким образом, под нечетким отношением R будем понимать функцию принадлежности $\mu_R(x, y)$ такую, что

$$\mu_R : M \times M \rightarrow [0, 1].$$

Значение функции принадлежности понимается как степень выполнения отношения xRy .

Существуют различные способы, которыми в общем случае формально могут быть заданы те или иные нечеткие бинарные отношения. Наибольшее распространение из них получили следующие.

1. В форме списка с явным перечислением всех пар нечеткого бинарного отношения и соответствующих им значений функции принадлежности:

$$\{(\langle x_1, y_1 \rangle, \mu(x_1, y_1)), \dots, (\langle x_k, y_k \rangle, \mu(x_k, y_k))\}, \quad \mu(x_i, y_i) > 0, \forall i \in 1 : k.$$

При этом для сокращения подобной записи пары с нулевыми значениями функции принадлежности не указываются в данном списке.

2. Аналитически в форме некоторого математического выражения для соответствующей функции принадлежности этого нечеткого отношения. Этот способ может быть использован для задания произвольных нечетких отношений как с конечным, так и с бесконечным числом элементов. Рассмотрим пример.

Пример 3.117. Пусть X – множество всех вещественных чисел. Отношение $x >> y$ (x многое больше y) можно задать функцией принадлежности:

$$\mu_R(x, y) = \begin{cases} 0, & x \leqslant y, \\ \frac{1}{1 + \frac{1}{(x - y)^2}}, & x > y. \end{cases}$$

3. В форме матрицы нечеткого отношения. Этот способ основан на представлении нечеткого бинарного отношения с конечным числом элементов в форме матрицы M_R , строки которой соответствуют первым элементам пары, а столбцы – вторым рассматриваемого нечеткого отношения.

При этом элементами матрицы являются соответствующие значения функции принадлежности данного отношения.

Пример 3.118. Рассмотрим нечеткое бинарное отношение R , заданное на множестве $\{a, b, c, d\}$. Матрица отношения M_R имеет вид (табл. 3.49)

Таблица 3.49

	a	b	c	d
a	0.1	0.9	0.7	0.8
b	0.3	0.5	0.0	0.2
c	0.0	0.3	0.4	0.1
d	0.0	0.0	0.9	0.2

4. В форме ориентированного нечеткого графа.

Определение 3.88. *Ориентированный нечеткий граф* есть тройка объектов $\mathbf{G} = (\mathbf{V}, \mathbf{E}, \mu_G)$, где $\mathbf{V} = \{v_1, \dots, v_n\}$ – множество вершин нечеткого графа, $\mathbf{E} = \{e_1, \dots, e_m\}$ – множество дуг нечеткого графа, μ_G – функция принадлежности дуг данному нечеткому графу, т. е. $\mu_G : \mathbf{E} \rightarrow [0, 1]$.

Замечание 3.55. Таким образом, ориентированный нечеткий граф представляет собой нагруженный, или взвешенный ориентированный граф (см. определение 2.44).

Пример 3.119. Граф нечеткого бинарного отношения из предыдущего примера 3.118 представлен на рис. 3.28.

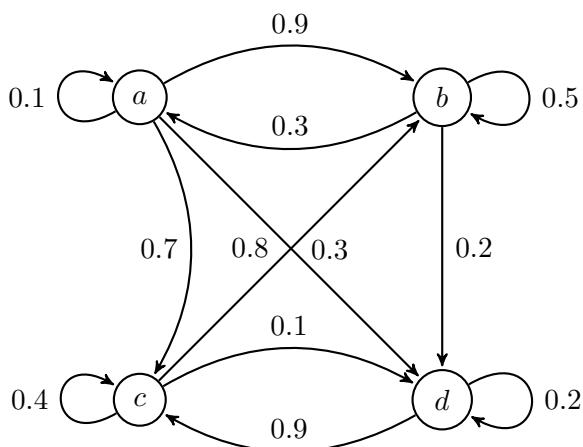


Рис. 3.28

Пример 3.120. Пусть $X = \{1, 2, 3\}$. Нечеткое отношение *приблизительно равняться*, заданное на множестве X , определим с помощью спис-

ка:

$$\begin{aligned}\mu_R(1, 1) &= \mu_R(2, 2) = \mu_R(3, 3) = 1, \\ \mu_R(1, 2) &= \mu_R(2, 1) = \mu_R(2, 3) = \mu_R(3, 2) = 0.8, \\ \mu_R(1, 3) &= \mu_R(3, 1) = 0.3.\end{aligned}$$

Функция принадлежности в аналитическом виде может быть задана следующим образом:

$$\mu_R(u, v) = \begin{cases} 1.0, & u = v, \\ 0.8, & |u - v| = 1, \\ 0.3 & |u - v| = 2. \end{cases}$$

Матрица отношения имеет вид (табл. 3.50)

Таблица 3.50

	1	2	3
1	1.0	0.8	0.3
2	0.8	1.0	0.8
3	0.3	0.8	1.0

Граф данного нечеткого бинарного отношения представлен на рис. 3.29.

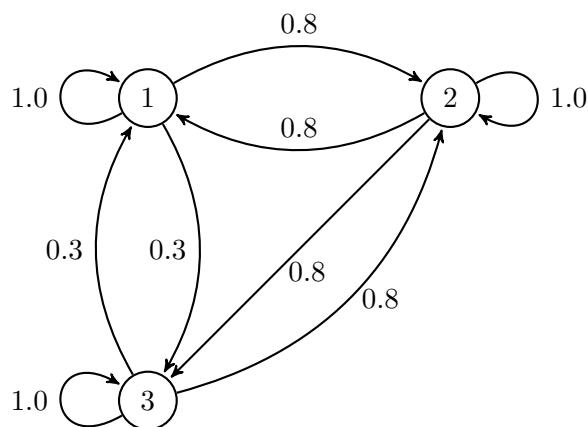


Рис. 3.29

3.5.2. Свойства нечетких бинарных отношений

Различные типы нечетких бинарных отношений определяются с помощью свойств, аналогичных свойствам обычных отношений, причем для

нечетких отношений можно указать различные способы обобщения этих свойств. В качестве основных свойств рассмотрим свойства, имеющие такую же алгебраическую запись, что и для обычных бинарных отношений.

Перейдем к изучению основных свойств нечетких отношений. Рассмотрим бинарное нечеткое отношение R , заданное на декартовом произведении $X \times X$.

Рефлексивность. Отношение R , называется *рефлексивным*, если выполняется равенство:

$$\mu_R(x, x) = 1, \forall x \in X. \quad (3.56)$$

Слабая рефлексивность. Отношение R , называется *слабо рефлексивным*, если выполняется неравенство:

$$\mu_R(x, y) \leq \mu_R(x, x), \forall x, y \in X. \quad (3.57)$$

Сильная рефлексивность. Отношение R называется *сильно рефлексивным*, если выполняется неравенство:

$$\mu_R(x, y) < 1, \forall x, y \in X. \quad (3.58)$$

Антирефлексивность. Отношение R называется *антирефлексивным*, если выполняется равенство:

$$\mu_R(x, x) = 0, \forall x \in X. \quad (3.59)$$

Слабая антирефлексивность. Отношение R называется *слабо антирефлексивным*, если выполняется неравенство:

$$\mu_R(x, x) \leq \mu_R(x, y), \forall x, y \in X. \quad (3.60)$$

Сильная антирефлексивность. Отношение R называется *сильно антирефлексивным*, если выполняется неравенство:

$$\mu_R(x, y) > 0, \forall x, y \in X. \quad (3.61)$$

Симметричность. Отношение R называется *симметричным*, если выполняется равенство:

$$\mu_R(x, y) = \mu_R(y, x), \forall x, y \in X. \quad (3.62)$$

Асимметричность. Отношение R называется *асимметричным*, если выполняется равенство:

$$\min \{\mu_R(x, y), \mu_R(y, x)\} = 0, \forall x, y \in X. \quad (3.63)$$

Антисимметричность. Отношение R называется *антисимметричным*, если выполняется равенство:

$$\min \{\mu_R(x, y), \mu_R(y, x)\} = 0, \quad \forall x, y \in X, \quad x \neq y. \quad (3.64)$$

Сильная полнота. Отношение R называется *сильно полным*, если выполняется равенство:

$$\max \{\mu_R(x, y), \mu_R(y, x)\} = 1, \quad \forall x, y \in X. \quad (3.65)$$

Слабая полнота. Отношение R называется *слабо полным*¹, если выполняется неравенство:

$$\max \{\mu_R(x, y), \mu_R(y, x)\} > 0, \quad \forall x, y \in X, \quad x \neq y. \quad (3.66)$$

Транзитивность. Отношение R называется *транзитивным*, если выполняется неравенство:

$$\mu_R(x, z) \geq \max_{y \in X} \{\min \{\mu_R(x, y), \mu_R(y, z)\}\}, \quad \forall x, y, z \in X. \quad (3.67)$$

Котранзитивность. Отношение R называется *котранзитивным*, если выполняется неравенство:

$$\mu_R(x, z) \leq \min_{y \in X} \{\max \{\mu_R(x, y), \mu_R(y, z)\}\}, \quad \forall x, y, z \in X. \quad (3.68)$$

Теорема 3.38. Для матрицы M_R бинарного нечеткого отношения R с конечным универсумом X справедливы следующие свойства.

1. Все элементы главной диагонали матрицы рефлексивного бинарного нечеткого отношения равны 1.
2. Все элементы главной диагонали матрицы антирефлексивного бинарного нечеткого отношения равны 0.
3. Матрица симметричного бинарного нечеткого отношения симметрична относительно главной диагонали.
4. Все элементы главной диагонали матрицы асимметричного бинарного нечеткого отношения равны 0.

Доказательство. Перечисленные свойства очевидным образом следуют из данных выше определений (3.56) – (3.68). ■

¹линейным или связным

3.5.3. Задачи для самостоятельного решения

Задача 3.41. Показать, что бинарное нечеткое отношение R , заданное на $N \times N$ с функцией принадлежности

$$\mu_R(x, y) = e^{-k(x-y)^2}, k > 1$$

нетранзитивно.

Задача 3.42. Для нечеткого бинарного отношения R , заданного на множестве $\{a, b, c, d\}$ матрицей отношения M_R (табл. 3.51) построить бли-

Таблица 3.51

	a	b	c	d
a	0.3	1.0	0.2	0.1
b	0.9	0.2	0.0	0.5
c	0.9	0.5	1.0	0.5
d	0.5	0.0	0.7	0.7

жайшее четкое отношение.

3.6. Нечеткая логика

3.6.1. Нечеткие высказывания

Определение 3.89. Нечетким высказыванием называется высказывание \tilde{F} , степень истинности которого $\mu(\tilde{F})$ можно оценить числом из интервала $[0, 1]$, $\mu(\tilde{F}) \in [0, 1]$. Если $\mu(\tilde{F}) = 0.5$, то высказывание называется индифферентным.

Нечеткой высказывательной переменной \tilde{X} называется нечеткое высказывание \tilde{X} , степень истинности которого может меняться в интервале $[0, 1]$.

Аналогично обычным (четким) высказываниям будем отождествлять нечеткое высказывание с его степенью истинности¹. Нечеткие высказывания и степень их истинности будем обозначать большими буквами с тильдой: $\tilde{F}, \tilde{G}, \tilde{X}, \tilde{Y}$ и т. д.

Введем на множестве нечетких высказываний логические операции, аналогичные операциям логики высказываний (соглашения о приоритете и записи операций такие же как и для обычных высказываний).

¹Для четких высказываний это были два значения *true* и *false*.

Отрицание нечеткого высказывания:

$$\neg \tilde{F} = 1 - \tilde{F}. \quad (3.69)$$

Помимо приведенного выше основного определения нечеткого логического отрицания, введенного Заде (3.69), могут использоваться следующие альтернативные формулы:

$$\begin{aligned}\neg \tilde{F} &= \frac{1 - \tilde{F}}{1 + \lambda \tilde{F}}, \quad \lambda > -1, \quad \text{нечеткое } \lambda - \text{дополнение по Сугено}, \\ \neg \tilde{F} &= \sqrt[p]{(1 - \tilde{F})}, \quad p > 0, \quad \text{нечеткое } p - \text{дополнение по Ягеру}.\end{aligned}$$

Конъюнкция нечетких высказываний:

$$\tilde{F}\tilde{G} = \min \left\{ \tilde{F}, \tilde{G} \right\}. \quad (3.70)$$

Помимо приведенного выше основного определения логической конъюнкции, введенного Заде (3.70), могут использоваться следующие альтернативные формулы:

$$\tilde{F}\tilde{G} = \max \left\{ \tilde{F} + \tilde{G} - 1, 0 \right\}, \quad \text{в базисе Лукашевича - Гилеса};$$

$$\tilde{F}\tilde{G} = \begin{cases} \tilde{G}, & \text{при } \tilde{F} = 1; \\ \tilde{F}, & \text{при } \tilde{G} = 1; \\ 0, & \text{в остальных случаях.} \end{cases} \quad \text{в базисе Вебера}$$

Дизъюнкция нечетких высказываний:

$$\tilde{F} \vee \tilde{G} = \max \left\{ \tilde{F}, \tilde{G} \right\}. \quad (3.71)$$

Помимо приведенного выше основного определения логической дизъюнкции, введенного Заде (3.71), могут использоваться следующие альтернативные формулы:

$$\tilde{F} \vee \tilde{G} = \tilde{F} + \tilde{G} - \tilde{F}\tilde{G}, \quad \text{в базисе Бандлера-Кохоута};$$

$$\tilde{F} \vee \tilde{G} = \min \left\{ \tilde{F} + \tilde{G}, 1 \right\}, \quad \text{в базисе Лукашевича-Гилеса};$$

$$\tilde{F} \vee \tilde{G} = \begin{cases} \tilde{G}, & \text{при } \tilde{F} = 0; \\ \tilde{F}, & \text{при } \tilde{G} = 0; \\ 1, & \text{в остальных случаях.} \end{cases} \quad \text{в базисе Вебера}$$

Импликация нечетких высказываний:

$$\tilde{F} \rightarrow \tilde{G} = \neg \tilde{F} \vee \tilde{G} = \max \left\{ 1 - \tilde{F}, \tilde{G} \right\}. \quad (3.72)$$

Помимо приведенного выше основного определения нечеткой импликации, введенного Заде (3.72), могут использоваться следующие альтернативные определения нечеткой импликации, предложенные различными исследователями в области теории нечетких множеств:

$$\begin{aligned} \tilde{F} \rightarrow \tilde{G} &= \max \left\{ 1 - \tilde{F}, \tilde{G} \right\}, && \text{Гедель;} \\ \tilde{F} \rightarrow \tilde{G} &= \min \left\{ \tilde{F}, \tilde{G} \right\}, && \text{Мамдани;} \\ \tilde{F} \rightarrow \tilde{G} &= \min \left\{ 1, 1 - \tilde{F} + \tilde{G} \right\}, && \text{Лукашевич;} \\ \tilde{F} \rightarrow \tilde{G} &= \min \left\{ \tilde{F} + \tilde{G}, 1 \right\}, && \text{Лукашевич-Гилес;} \\ \tilde{F} \rightarrow \tilde{G} &= \tilde{F}\tilde{G}, && \text{Бандлер-Кохоут;} \\ \tilde{F} \rightarrow \tilde{G} &= \max \left\{ \tilde{F}\tilde{G}, 1 - \tilde{F} \right\} && \text{Вади.} \end{aligned}$$

Замечание 3.56. Общее число введенных определений нечеткой импликации не ограничивается приведенными выше. Большое количество работ по изучению различных вариантов нечеткой импликации обусловлено тем, что понятие нечеткой импликации является ключевым при нечетких выводах и принятии решений в нечетких условиях. Наибольшее применение при решении прикладных задач нечеткого управления находит нечеткая импликация Заде.

Эквивалентность нечетких высказываний:

$$\tilde{F} \leftrightarrow \tilde{G} = \tilde{F}\tilde{G} \vee \neg \tilde{F} \neg \tilde{G} = \max \left\{ \min \left\{ 1 - \tilde{F}, 1 - \tilde{G} \right\}, \min \left\{ \tilde{F}, \tilde{G} \right\} \right\} \quad (3.73)$$

Формулу (3.73) можно записать в другом виде:

$$\tilde{F} \leftrightarrow \tilde{G} = (\tilde{F} \rightarrow \tilde{G}) (\tilde{G} \rightarrow \tilde{F}) = \min \left\{ \max \left\{ 1 - \tilde{F}, \tilde{G} \right\}, \max \left\{ \tilde{F}, 1 - \tilde{G} \right\} \right\}. \quad (3.74)$$

Замечание 3.57. Для нечетких логических операций выполняются все законы логики высказываний, кроме законов противоречия и исключенного третьего, т. е.

$$\neg \tilde{F} \wedge \tilde{F} \neq 0, \quad \neg \tilde{F} \vee \tilde{F} \neq 1.$$

Пример 3.121. Найдем степень истинности высказывания \tilde{H} :

$$\tilde{H} = (\tilde{F} \vee \tilde{G}) \leftrightarrow (\tilde{F} \rightarrow \tilde{F}\tilde{G}),$$

при $\tilde{F} = 0.75$, $\tilde{G} = 0.2$.

Обозначим $\widetilde{F}_1 = (\widetilde{F} \vee \widetilde{G})$, а $\widetilde{F}_2 = (\widetilde{F} \rightarrow \widetilde{F}\widetilde{G})$. Согласно (3.69) – (3.72) получаем:

$$\begin{aligned}\widetilde{F}_1 &= \max \left\{ \widetilde{F}, \widetilde{G} \right\} = 0.75, \\ \widetilde{F}_2 &= \max \left\{ 1 - \widetilde{F}, \widetilde{F}\widetilde{G} \right\} = \max \left\{ 0.25, \min \left\{ \widetilde{F}, \widetilde{G} \right\} \right\} = \\ &= \max \left\{ 0.25, \min \left\{ 0.75, 0.2 \right\} \right\} = 0.25.\end{aligned}$$

Отсюда, используя (3.73) имеем

$$\begin{aligned}\widetilde{H} = \widetilde{F}_1 \leftrightarrow \widetilde{F}_2 &= \max \left\{ \min \left\{ 1 - \widetilde{F}_1, 1 - \widetilde{F}_2 \right\}, \min \left\{ \widetilde{F}_1, \widetilde{F}_2 \right\} \right\} = \\ &= \max \left\{ \min \left\{ 0.25, 0.75 \right\}, \min \left\{ 0.75, 0.25 \right\} \right\} = 0.25.\end{aligned}$$

3.6.2. Нечеткие формулы логики высказываний

Аналогично определению для формул логики высказываний введем индуктивное определение нечеткой логической формулы.

Определение 3.90. Нечеткой логической формулой называется:

- a) любая нечеткая высказывательная переменная;
- b) если \widetilde{F} и \widetilde{G} – нечеткие логические формулы, то

$$\neg \widetilde{F}, \quad \widetilde{F}\widetilde{G}, \quad \widetilde{F} \vee \widetilde{G}, \quad \widetilde{F} \rightarrow \widetilde{G}, \quad \widetilde{F} \leftrightarrow \widetilde{G}$$

тоже нечеткие логические формулы.

Определение 3.91. Пусть $\widetilde{F}(\widetilde{X}_1, \dots, \widetilde{X}_n)$ и $\widetilde{G}(\widetilde{X}_1, \dots, \widetilde{X}_n)$ – две нечеткие логические формулы. Степенью равносильности формул \widetilde{F} и \widetilde{G} называется величина

$$\mu(\widetilde{F}, \widetilde{G}) = \bigwedge_{(\alpha_1, \dots, \alpha_n)} \left\{ \widetilde{F}(\alpha_1, \dots, \alpha_n) \leftrightarrow \widetilde{G}(\alpha_1, \dots, \alpha_n) \right\}, \quad (3.75)$$

где конъюнкция берется по всем наборам степеней истинности $(\alpha_1, \dots, \alpha_n)$ нечетких переменных $(\widetilde{X}_1, \dots, \widetilde{X}_n)$.

Очевидно, что множество всех наборов степеней истинности имеет мощность континuum¹ в отличие от двузначной логики высказываний, где число всех наборов переменных конечно и равно 2^n .

Если $\mu(\widetilde{F}, \widetilde{G}) = 0.5$, то нечеткие формулы \widetilde{F} и \widetilde{G} называются *индиффирентными*.

¹Мощность множества всех вещественных чисел.

Если $\mu(\tilde{F}, \tilde{G}) > 0.5$, то нечеткие формулы \tilde{F} и \tilde{G} называются *нечетко равносильными*.

Если $\mu(\tilde{F}, \tilde{G}) < 0.5$, то нечеткие формулы \tilde{F} и \tilde{G} называются *нечетко неравносильными*.

Степенью неравносильности формул \tilde{F} и \tilde{G} называется величина

$$\bar{\mu}(\tilde{F}, \tilde{G}) = 1 - \mu(\tilde{F}, \tilde{G}).$$

Пример 3.122. Определим степень равносильности нечетких формул $\tilde{F} = \tilde{X} \rightarrow \tilde{Y}$ и $\tilde{G} = \neg(\tilde{X}\tilde{Y})$ при условии, что нечеткие переменные \tilde{X} и \tilde{Y} принимают значения степеней истинности из множества $\{0.1, 0.2\}$.

Рассмотрим все возможные наборы значений \tilde{X} и \tilde{Y} :

$$(0.1, 0.1), (0.1, 0.2), (0.2, 0.1), (0.2, 0.2). \quad (3.76)$$

С учетом (3.69) – (3.73) имеем

$$\begin{cases} \tilde{F} = \tilde{X} \rightarrow \tilde{Y} = \max \left\{ 1 - \tilde{X}, \tilde{Y} \right\}, \\ \tilde{G} = \neg(\tilde{X}\tilde{Y}) = 1 - (\tilde{X}\tilde{Y}) = 1 - \min \left\{ \tilde{X}, \tilde{Y} \right\}. \end{cases}$$

Вычислим значения формул \tilde{F} и \tilde{G} на каждом из наборов (3.76):

$$\begin{aligned} \tilde{F}(0.1, 0.1) &= \max \{1 - 0.1, 0.1\} = 0.9, \\ \tilde{F}(0.1, 0.2) &= \max \{1 - 0.1, 0.2\} = 0.9, \\ \tilde{F}(0.2, 0.1) &= \max \{1 - 0.2, 0.1\} = 0.8, \\ \tilde{F}(0.2, 0.2) &= \max \{1 - 0.2, 0.2\} = 0.8 \\ \tilde{G}(0.1, 0.1) &= 1 - \min \{0.1, 0.1\} = 0.9, \\ \tilde{G}(0.1, 0.2) &= 1 - \min \{0.1, 0.2\} = 0.9, \\ \tilde{G}(0.2, 0.1) &= 1 - \min \{0.2, 0.1\} = 0.9, \\ \tilde{G}(0.2, 0.2) &= 1 - \min \{0.2, 0.2\} = 0.8 \end{aligned}$$

Вычислим теперь степень равносильности формул \tilde{F} и \tilde{G} . Согласно (3.74) и (3.75) имеем:

$$\begin{aligned} \tilde{F}(0.1, 0.1) \leftrightarrow \tilde{G}(0.1, 0.1) &= \min \{ \max \{1 - 0.9, 0.9\}, \max \{0.9, 1 - 0.9\} \} = 0.9, \\ \tilde{F}(0.1, 0.2) \leftrightarrow \tilde{G}(0.1, 0.2) &= \min \{ \max \{1 - 0.9, 0.9\}, \max \{0.9, 1 - 0.9\} \} = 0.9, \\ \tilde{F}(0.2, 0.1) \leftrightarrow \tilde{G}(0.2, 0.1) &= \min \{ \max \{1 - 0.8, 0.9\}, \max \{0.8, 1 - 0.9\} \} = 0.8, \\ \tilde{F}(0.2, 0.2) \leftrightarrow \tilde{G}(0.2, 0.2) &= \min \{ \max \{1 - 0.8, 0.8\}, \max \{0.8, 1 - 0.8\} \} = 0.8. \end{aligned}$$

Окончательно, согласно (3.75) получим:

$$\mu(\tilde{F}, \tilde{G}) = \min(0.9, 0.9, 0.8, 0.8) = 0.8,$$

таким образом, формулы \tilde{F} и \tilde{G} нечетко равносильны на заданных наборах степеней истинности.

Замечание 3.58. На других наборах степеней истинности нечетких переменных \tilde{X} и \tilde{Y} формулы \tilde{F} и \tilde{G} из примера 3.122 могут быть нечетко неравносильны.

3.7. Теория алгоритмов

Историческая справка. Понятие *алгоритма*, являющееся одним из основных понятий математики, возникло задолго до появления вычислительных машин. Арабский математик IX в. Аль-Хорезми (из города Хорезма на реке Аму-Дарья)¹ в своей книге подробно описал индийскую арифметику. В ней он впервые выдвинул идею о том, что решение любой поставленной математической задачи может быть оформлено в виде последовательности выполняемых инструкций, т. е. может быть алгоритмизировано. Триста лет спустя эту книгу перевели на латинский язык, и она стала первым учебником «индийской» (то есть нашей современной) арифметики для всей Европы. Переводчик, имя которого не сохранилось в истории, дал ей название *Algoritmi de numero Indorum*. Кроме того, Аль-Хорезми написал книгу об общих правилах решения арифметических задач при помощи уравнений. Она называлась «Китаб ал-Джебр». Эта книга дала имя науке алгебре.

Формальные определения алгоритма появились в первой половине XX в. Можно выделить три основные алгоритмические модели:

- 1) алгоритм представляется как некоторое детерминированное устройство, способного выполнять в каждый отдельный момент некоторые примитивные операции;
- 2) понятие алгоритма связывают с вычислениями и числовыми функциями и рассматривают алгоритм с точки зрения того, что можно вычислить с его помощью;
- 3) алгоритм рассматривается как преобразование слов в некотором алфавите, при этом элементарными операциями являются подстановки, т. е. замены части слова другим словом.

3.7.1. Машина Тьюринга и функции, вычислимые по Тьюрингу

Историческая справка. Алан Матисон Тьюринг (1912–1954), английский математик, получил докторскую степень в Принстонском университете (США), где его

¹Полное имя Аль-Хорезми следующее: Аль-Хорезми Абу Абдалла Мухаммед бен Муса аль-Маджуси.

научным руководителем был А. Черч. После получения докторской степени Тьюринг отклонил предложение Дж. фон Неймана остатся в США и вернулся в Кембридж.

В 1936 г. в трудах Лондонского математического общества появилась знаменитая работа Тьюринга «О вычислимых числах, с приложением к проблеме разрешимости», в которой он описал универсальную вычислительную машину, предназначенную для решения любых математических или логических задач. Устройство, впоследствии названное «машиной Тьюринга», обладало основными свойствами современного компьютера. Кстати, именно Тьюринг впервые употребил термин «компьютер» применительно к машине (раньше так называли людей, проводивших расчеты на арифмометре).

Идея машины Тьюринга легла в основу современной теории алгоритмов

С началом Второй мировой войны Тьюринг перешел на работу в правительственный Школу кодов и шифров (Government Code and Cypher School). Используя более ранние польские наработки, совместно с У. Уэлчманом раскрыл шифры германских BBC, создав дешифровочную машину «Бомба», а затем взломал гораздо более сложный шифр, использовавшийся в шифровальных машинах «Энigma», которыми были оснащены германские подводные лодки. В этот период Тьюринг занимался также разработкой шифров для переписки У. Черчилля и Ф. Рузвельта.

Определение 3.92. Машина Тьюринга состоит из следующих элементов:

1. *Ленты*, разбитой на ячейки и бесконечной в обе стороны . В каждой ячейке ленты может быть записан один из символов конечного алфавита $\mathcal{A} = \{\alpha_0, \alpha_1, \dots, \alpha_t\}$, называемого *внешним алфавитом*. Условимся считать, что символ α_0 является пустым символом. Иногда для обозначения пустого символа алфавита используют символы Λ или ϵ .

2. *Управляющего устройства*, которое может находиться в одном из конечного числа *внутренних состояний* $\mathcal{Q} = \{q_0, q_1, \dots, q_n\}$. Число элементов в \mathcal{Q} характеризует объем внутренней памяти машины. В множестве \mathcal{Q} выделены два специальных состояния: q_0 и q_1 , называемых *заключительным* и *начальным* состояниями соответственно. Машина всегда начинает работу в состоянии q_1 и всегда останавливается, попав в состояние q_0 (часто это состояние называют «стоп-сигналом»).

3. *Считывающей/пишущей головки*, которая может перемещаться вдоль ленты и в каждый момент времени обозревает (читывает) одну из ее ячеек. Машина Тьюринга функционирует в дискретные моменты времени $t = 0, 1, 2, \dots$ и в зависимости от внутреннего состояния машины и считываемого символа на ленте:

- 1) записывает в эту ячейку символ внешнего алфавита;
- 2) сдвигает считающую головку на один шаг влево (вправо) или оставляет ее на месте;
- 3) переходит в новое внутреннее состояние.

Схематично машины Тьюринга представлена на рис. 3.30.

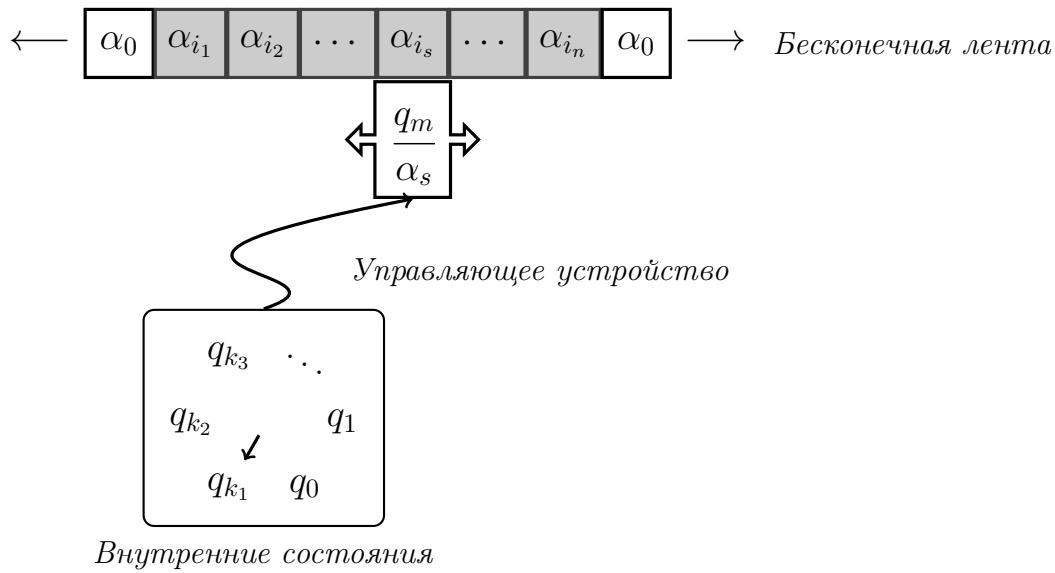


Рис. 3.30

Таким образом, работа машины Тьюринга определяется системой *команд* вида

$$q_i \alpha_j \rightarrow q_k \alpha_l M_s, \quad (3.77)$$

где q_i — текущее внутреннее состояние машины; α_j — считываемый символ; q_k — новое внутреннее состояние; α_l — записываемый символ; M_s — направление движения головки, обозначаемое одним из символов: $M_1 = L$ (влево), $M_2 = R$ (вправо), $M_3 = C$ (на месте).

Определение 3.93. Считаем, что для каждой пары

$$\{q_i, \alpha_j\}, \quad q_i \in \mathcal{Q}, \quad \alpha_j \in \mathcal{A}, \quad i \in 1 : n, \quad j \in 0 : m$$

существует ровно одна команда вида (3.77). Множество этих команд называют *программой* машины. Таким образом, в программе всего $n(m + 1)$ команд.

Замечание 3.59. Из определения 3.92 следует, что у машины Тьюринга заключительное состояние q_0 может стоять только в правой части (3.77).

Определение 3.94. Совокупность внутреннего состояния, состояния ленты (т. е. размещения символов внешнего алфавита по ячейкам) и положения головки на ленте называют *конфигурацией*.

Работа машины заключается в изменении конфигураций согласно программе.

Состояние ленты для конфигурации с заключительным внутренним состоянием g_0 называют *результатом работы* машины Тьюринга.

Часто программу для машины Тьюринга записывают в виде таблицы команд, называемой *Тьюринговой функциональной схемой*.

Пример 3.123. Рассмотрим машину Тьюринга с внешним алфавитом $\mathcal{A} = \{\alpha_0, \alpha_1, \alpha_2\}$ и множеством состояний $\mathcal{Q} = \{q_0, q_1, q_2, q_3, q_4\}$. Ее функциональная схема представлена в табл. 3.52. Проанализируем работу данной машины Тьюринга.

Таблица 3.52

	α_0	α_1	α_2		α_0	α_1	α_2
q_1	$q_3\alpha_2L$	$q_2\alpha_1R$	$q_1\alpha_2L$	q_3	$q_0\alpha_0R$	$q_4\alpha_1R$	$q_1\alpha_2C$
q_2	$q_2\alpha_0C$	$q_1\alpha_2C$	$q_2\alpha_1C$	q_4	$q_3\alpha_1C$	$q_4\alpha_0R$	$q_4\alpha_2R$

Пусть начальная конфигурация имеет вид $\mathcal{K}_1 = \alpha_0 \alpha_2 \frac{\alpha_2}{q_1} \alpha_0$, т. е. управляющая головка обозревает символ α_2 , а машина находится в состоянии q_1 . Тогда

$$\alpha_0 \alpha_2 \frac{\alpha_2}{q_1} \alpha_0 \rightarrow \alpha_0 \frac{\alpha_2}{q_1} \alpha_2 \alpha_0 \rightarrow \frac{\alpha_0}{q_1} \alpha_2 \alpha_2 \alpha_0 \rightarrow \frac{\alpha_0}{q_3} \alpha_2 \alpha_2 \alpha_2 \alpha_0 \rightarrow \alpha_0 \frac{\alpha_2}{q_0} \alpha_2 \alpha_2 \alpha_0.$$

Так как при последней конфигурации машина находится в состоянии q_0 , то слово $\alpha_2 \alpha_2 \alpha_2$ является результатом.

Рассмотрим другую начальную конфигурацию: $\mathcal{K}_2 = \alpha_0 \alpha_1 \alpha_1 \alpha_2 \frac{\alpha_2}{q_1} \alpha_0$.

Тогда

$$\begin{aligned} \alpha_0 \alpha_1 \alpha_1 \alpha_2 \frac{\alpha_2}{q_1} \alpha_0 &\rightarrow \alpha_0 \alpha_1 \alpha_1 \frac{\alpha_2}{q_1} \alpha_2 \alpha_0 \rightarrow \alpha_0 \alpha_1 \frac{\alpha_1}{q_1} \alpha_2 \alpha_2 \alpha_0 \rightarrow \\ &\rightarrow \alpha_0 \alpha_1 \alpha_1 \frac{\alpha_2}{q_2} \alpha_2 \alpha_0 \rightarrow \alpha_0 \alpha_1 \alpha_1 \frac{\alpha_1}{q_2} \alpha_2 \alpha_0 \rightarrow \alpha_0 \alpha_1 \alpha_1 \frac{\alpha_2}{q_1} \alpha_2 \alpha_0. \end{aligned}$$

Видно, что последняя конфигурация повторяет вторую, следовательно, машина никогда не придет в заключительное состояние q_0 .

Определение 3.95. Если машина Тьюринга начиная работу в конфигурации \mathcal{K} , за конечное число шагов приходит в состояние q_0 (в конфигурации \mathcal{K}_0), то ее называют *применимой* к исходной конфигурации \mathcal{K} , в противном случае — *неприменимой*.

Конфигурацию \mathcal{K} называют *начальной*, а конфигурацию \mathcal{K}_0 — *заключительной*.

Пример 3.124. В примере 3.123 машина применима к конфигурации \mathcal{K}_1 и неприменима к конфигурации \mathcal{K}_2 .

Определение 3.96. Активной зоной конфигурации называют минимальную связную часть ленты, содержащую обозреваемую ячейку, а также все ячейки, в которых записаны непустые символы.

Замечание 3.60. Будем считать, что машина Тьюринга всегда начинает и заканчивает работу в начале активной зоны. Такие начальную и конечную конфигурации называют *стандартными*.

Определение 3.97. Две машины Тьюринга *эквивалентны*, если они применимы к одним и тем же начальным конфигурациям и результаты применения обеих машин совпадают.

Теорема 3.39. Для любой машины Тьюринга \mathcal{T} можно построить эквивалентную ей машину \mathcal{T}_S , у которой все заключительные конфигурации находятся в стандартной форме.

Доказательство. Машину \mathcal{T}_S построим, добавив к машине \mathcal{T} два новых состояния q' и q'' и команды

$$q_0\alpha_i \rightarrow q'\alpha_i L, \quad i \in 0 : m, \quad q'\alpha_i \rightarrow q''\alpha_i L, \quad i \in 1 : m, \quad q'\alpha_0 \rightarrow q''\alpha_0 R.$$

При этом состояние q'' будем считать заключительным для машины \mathcal{T}_S . Очевидно, что построенная машина \mathcal{T}_S эквивалентна машине \mathcal{T} . ■

Замечание 3.61. Условимся далее *по умолчанию* рассматривать машины Тьюринга с внешним алфавитом $\mathcal{A} = \{\alpha_0, 1\}$, в который при необходимости будем включать символ-разделитель *.

Произвольное число $x \in N_0 = N \cup \{0\}$ будем представлять на ленте в виде $x + 1$ единицы

$$1^{x+1} = \overbrace{1 \ 1 \ \dots \ 1 \ 1}^{x+1}, \quad (3.78)$$

чтобы запись нуля была непустой. Представление числа в виде (3.78) называют *унарным*.

Определение 3.98. Говорят, что машина \mathcal{T} *правильно вычисляет функцию* $f(x_1, \dots, x_n) : N_0^n \rightarrow N_0$, если начальную конфигурацию

$$\mathcal{K}_{start} = 1^{x_1+1} * \dots * 1^{x_n+1}$$

она переводит в заключительную конфигурацию

$$\mathcal{K}_{finish} = 1^{f(x_1, \dots, x_n)+1}$$

при условии, что значение функции $f(x_1, \dots, x_n)$ определено, и машина \mathcal{T} неприменима к начальной конфигурации \mathcal{K}_{start} , если значение функции не определено.

Рассмотрим несколько примеров на построение машин Тьюринга (табл. 3.53–3.55).

Таблица 3.53

	α_0	1
q_1	$q_2 1L$	$q_1 1R$
q_2	$q_0 \alpha_0 R$	$q_2 1L$

Таблица 3.54

	α_0	1
q_1		$q_2 \alpha_0 R$
q_2	$q_0 1C$	$q_2 \alpha_0 R$

Таблица 3.55

	α_0	1	*
q_1		$q_2 \alpha_0 R$	
q_2	$q_3 \alpha_0 L$	$q_2 1R$	$q_2 1R$
q_3		$q_4 \alpha_0 L$	
q_4	$q_0 \alpha_0 R$	$q_4 1L$	

Пример 3.125. Пусть \mathcal{T}_1 — машина Тьюринга с программой в табл. 3.53. Машина \mathcal{T}_1 , начиная работать в конфигурации $\mathcal{K}_0 = q_1 1^{x+1}$, останавливается в конфигурации $\mathcal{K}_1 = q_0 1^{x+2}$, таким образом, правильно вычисляет функцию $f(x) = x + 1$, $x \in N_0$.

Пусть машина \mathcal{T}_2 имеет программу в табл. 3.54. Если $\mathcal{K}_0 = q_1 1^{x+1}$ — начальная конфигурация, то заключительная конфигурация выглядит так: $\mathcal{K}_1 = q_0 1$, т. е. машина \mathcal{T}_2 правильно вычисляет функцию $f(x) = 0$, $x \in N_0$.

Рассмотрим машину \mathcal{T}_3 заданную табл. 3.55. Тогда при начальной конфигурации $\mathcal{K}_0 = q_1 1^{x+1} * 1^{y+1}$ заключительная конфигурация имеет вид $\mathcal{K}_1 = q_0 1^{x+y+1}$. Следовательно, машина \mathcal{T}_3 правильно вычисляет функцию $f(x) = x + y$, $x, y \in N_0$.

Машину Тьюринга можно представить с помощью ориентированного графа. Каждому состоянию $q_i \in \mathcal{Q}$ ставится в соответствие вершина графа, а каждой команде — помеченная дуга.

Пример 3.126. Машина Тьюринга \mathcal{T}_1 из примера 3.125 с программой в табл. 3.53, будет представлена в виде графа следующим образом (рис. 3.31).

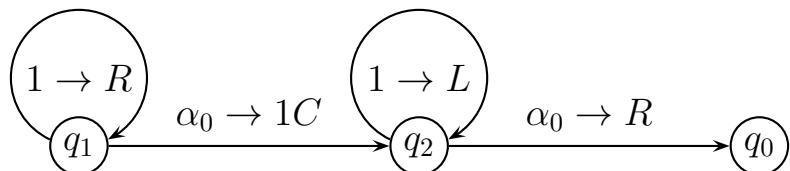


Рис. 3.31

Если на очередном такте работы машины Тьюринга символ на ленте не изменяется, то в правой части команды его можно не писать, например, α_0 на ребре $q_2 \rightarrow q_0$.

Замечание 3.62. Для простоты изложения, конфигурацию в унарном алфавите $q_i 1^{\alpha+1}$ будем записывать, как $q_i \alpha$.

В определении 3.92 для машины Тьюринга лента предполагалась бесконечной в обе стороны. Ограничим ленту с одной стороны и покажем, что

машина Тьюринга с правой или левой полулентой эквивалентна машине Тьюринга с бесконечной лентой.

Теорема 3.40. Функция, правильно вычислимая на машине Тьюринга с обычной лентой, правильно вычислима и на машине Тьюринга с правой (левой) полулентой, т. е. для любой машины Тьюринга \mathcal{T} существует эквивалентная ей машина с правой (левой) полулентой \mathcal{T}_R (\mathcal{T}_L).

Доказательство. Пусть лента ограничена слева неподвижным маркером. Машина \mathcal{T}_R работает как обычная машина Тьюринга \mathcal{T} до тех пор, пока активная зона текущей конфигурации не выходит на неподвижный маркер. Тогда \mathcal{T}_R производит сдвиг активной зоны вправо на одну ячейку и снова работает, как машина \mathcal{T} .

Такие же рассуждения можно провести для машины с левой полулентой. ■

3.7.2. Примеры и задачи на построение машин Тьюринга

Пример 3.127. Построим машину Тьюринга (как и в примере 3.125) правильно вычисляющую функцию $f(x) = x + 1$, но не для унарного алфавита, а в алфавите $\mathcal{A} = \{\alpha_0, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$.

Идея построения машины в следующем. Нужно „передвинуть“ управляющую головку на последнюю цифру числа. Если это цифра от 0 до 8, то заменить ее цифрой на 1 больше, вернуться в начало и остановиться. Если же это цифра 9, тогда заменить ее на 0 и сдвинуть управляющую головку к предыдущей цифре, после чего таким же способом увеличить на 1 эту предпоследнюю цифру.

Пусть на ленте только девятки (например, 9...9). Тогда управляющую головку будет сдвигаться влево, заменяя девятки на нули, и в конце концов окажется на пустой ячейке. В эту пустую ячейку надо записать 1 и остановиться.

$$\begin{aligned}
 q_1 \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\} &\rightarrow q_1 \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\} R, \\
 q_1 \alpha_0 &\rightarrow q_2 \alpha_0 L, \\
 q_2 \{0, 1, 2, 3, 4, 5, 6, 7, 8\} &\rightarrow q_3 \{1, 2, 3, 4, 5, 6, 7, 8, 9\} L, \\
 q_2 9 &\rightarrow q_2 0 L, \\
 q_3 \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\} &\rightarrow q_3 \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\} L, \\
 q_3 \alpha_0 &\rightarrow q_0 \alpha_0 R, \\
 q_2 \alpha_0 &\rightarrow q_0 1 E
 \end{aligned}$$

Пример 3.128. Построим машину Тьюринга, которую назовем „переводчиком“. Алфавит этой машины $\mathcal{A} = \{\alpha_0, s_1, s_2, t_1, t_2, *\}$. Машина должна осуществлять *посимвольный перевод* цепочки s_{i_1}, \dots, s_{i_k} в цепочку t_{i_1}, \dots, t_{i_k} , сохраняя при этом *оригинальный текст*. То есть, машина Тьюринга начиная работать в конфигурации $\mathcal{K}_0 = q_1 s_{i_1}, \dots, s_{i_k} *$ должна закончить работу в конфигурации $\mathcal{K}_1 = q_1 s_{i_1}, \dots, s_{i_k} * t_{i_1}, \dots, t_{i_k} \alpha_0$ (рис. 3.32).

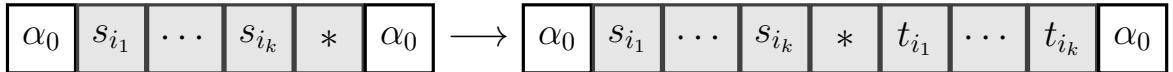


Рис. 3.32

Как должна работать такая машина? Она должна запомнить обозреваемый символ s_i , то место, где он считан, найти первую пустую ячейку и записать туда символ t_i . Будем запоминать символ s_1 состоянием q_{s_1} , символ s_2 – состоянием q_{s_2} , а текущую копируемую ячейку символом пустой ячейки α_0 . Имеем:

$$q_1 s_i \rightarrow q_{s_i} \alpha_0 R, \quad i = 1, 2.$$

Далее, машина в состоянии q_{s_1} или q_{s_2} должна дойти, двигаясь вправо, до первой пустой ячейки и записать в нее символы t_1 или t_2 (перейдя, соответственно, в состояния q_{t_1} и q_{t_2}). Это реализуется следующими командами:

$$\begin{aligned} q_{s_i} \{s_1, s_2, t_1, t_2, *\} &\rightarrow q_{s_i} \{s_1, s_2, t_1, t_2, *\} R, \\ q_{s_i} \alpha_0 &\rightarrow q_{t_i} t_i L, \quad i = 1, 2. \end{aligned}$$

Теперь нам нужно вернуться в текущую копирующую ячейку и символ α_0 заменить на s_1 или s_2 . На этом один *цикл* работы машины закончится. Добавим следующие команды:

$$\begin{aligned} q_{t_i} \{s_1, s_2, t_1, t_2, *\} &\rightarrow q_{t_i} \{s_1, s_2, t_1, t_2, *\} L, \\ q_{t_i} \alpha_0 &\rightarrow q_1 t_i R, \quad i = 1, 2. \end{aligned}$$

Возникает вопрос: когда машина должна закончить свою работу – тогда, когда после завершения очередного цикла встретит символ-разделитель *. Добавляем последнюю команду:

$$q_1 * \rightarrow q_0 * C.$$

Замечание 3.63. Как и в задаче 3.127, легко добавить команды, чтобы машина завершила работу в стандартной конечной конфигурации (в самой левой непустой ячейке ленты).

Пример 3.129. Построим машину Тьюринга, которая обращает непустое входное слово алфавите $\{a, b, \alpha_0\}$, где α_0 – пустой символ. Более точно, машина должна начальную конфигурацию $\mathcal{K}_S = q_1 W$ переводить в конфигурацию $\mathcal{K}_F = q_0 W^r$, где W – непустое слово в алфавите $\{a, b\}$, а W^r –

слово W , записанное в обратном порядке. Назовем такую машину „конвертором“.

Основная идея работы машины состоит в том, что обращение слова W надо начинать с конца этого слова. Машина будет считывать очередную букву $x \in \{a, b\}$, ставить на ее месте временный маркер x_1 , двигаться вправо и в первую пустую ячейку записывать x .

В состоянии q_1 машина находит первый с конца слова W еще „не обращенный“ символ $x \in \{a, b\}$, заменяет его временным маркером $x_1 \in \{a_1, b_1\}$ и переходит в состояние q_a или q_b в зависимости от считанного символа.

$$q_1 \{a, b\} \rightarrow q_1 \{a, b\} R, q_1 \alpha_0 \rightarrow q_2 * L, q_2 a \rightarrow q_a a_1 R, q_2 b \rightarrow q_b b_1 R$$

В состоянии q_a или q_b она находит первую пустую ячейку, записывает в нее соответственно символ a или b и переходит в состояние q_3 .

$$\begin{aligned} q_a \{a, b, a_1, b_1, *\} &\rightarrow q_a \{a, b, a_1, b_1, *\} R, \\ q_b \{a, b, a_1, b_1, *\} &\rightarrow q_b \{a, b, a_1, b_1, *\} R, \\ q_a \alpha_0 &\rightarrow q_3 a L, q_b \alpha_0 \rightarrow q_3 b L \end{aligned}$$

В этом состоянии машина идет влево по входной ленте и при переходе через символ $*$ приходит в состояние q_2 и цикл повторяется.

$$q_3 \{a, b\} \rightarrow q_3 \{a, b\} L, q_3 * \rightarrow q_2 * L$$

Если в состоянии q_2 она читает пустой символ α_0 , то это означает, что обращение слова W завершено. Тогда машина переходит в состояние q_4 , удаляет временные маркеры $\{a_1, b_1\}$ и останавливается в состоянии q_0 .

$$q_2 \alpha_0 \rightarrow q_4 \alpha_0 R, q_4 \{a_1, b_1\} \rightarrow q_4 \alpha_0 R, q_4 * \rightarrow q_0 \alpha_0 R$$

Диаграмма данной машины Тьюринга приведена на рис.3.33.

Задача 3.43. Постройте машины Тьюринга, правильно вычисляющие следующие функции:

$$1) f(x) = x + k, x, k \in N_0;$$

$$2) f(x) = \text{sign}(x) = \begin{cases} 0 & \text{при } x = 0, \\ 1 & \text{при } x > 0, \end{cases} \quad x \in N_0;$$

$$3) f(x) = \overline{\text{sign}(x)} = \begin{cases} 1 & \text{при } x = 0, \\ 0 & \text{при } x > 0, \end{cases} \quad x \in N_0;$$

$$4) \text{Функцию выбора аргумента } f(x_1, x_2, x_3) = x_2, \quad x_1, x_2, x_3 \in N_0;$$

$$5) f(x, y) = xy, x, y \in N_0.$$

Задача 3.44. Используя машину Тьюринга из примера 3.128 постройте машину, правильно вычисляющую функцию $f(x) = 2x$.

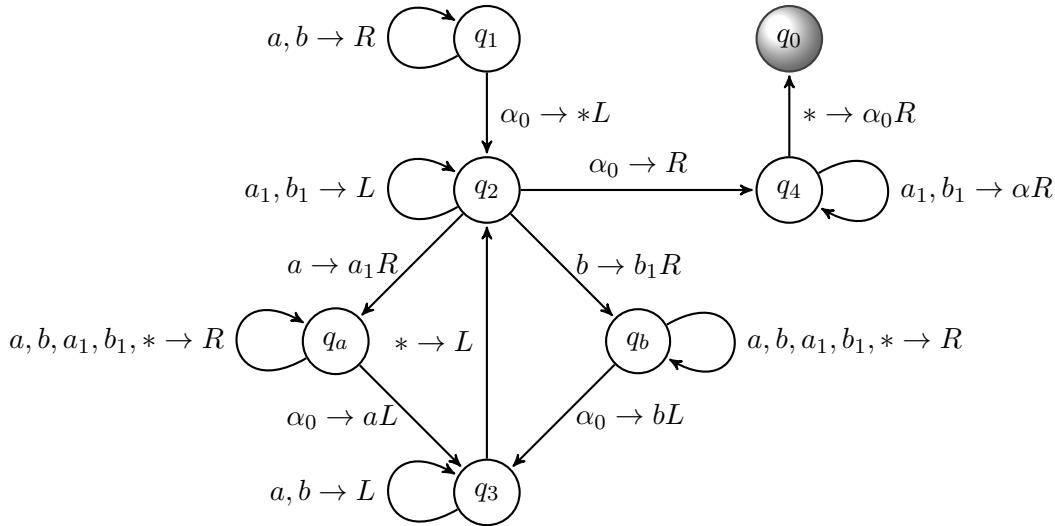


Рис. 3.33

3.7.3. Операции над машинами Тьюринга

Прямое построение машин Тьюринга для решения даже простых задач может оказаться затруднительным. Однако можно облегчить работу, если использовать сочетания программ нескольких машин в результирующую программу.

Напомним, что согласно определению 2.22 словом, или цепочкой в алфавите \mathcal{A} называют последовательность символов из этого алфавита. Множество слов алфавита \mathcal{A} обозначают \mathcal{A}^* .

► **Композиция машин Тьюринга.** Пусть даны две машины Тьюринга \mathcal{T}_1 и \mathcal{T}_2 , вычисляющие соответственно функции $f_1(\alpha)$ и $f_2(\alpha)$, заданные на множестве \mathcal{A}^* . Тогда существует машина Тьюринга \mathcal{T} , которая вычисляет функцию суперпозиции $f(\alpha) = f_2(f_1(\alpha))$. При этом для любого слова α функция $f(\alpha)$ определена в том и только в том случае, когда $f_1(\alpha)$ определена и $f_2(f_1(\alpha))$ также определена.

Состояния машин \mathcal{T}_1 и \mathcal{T}_2 обозначим дополнительно верхними индексами с номерами машин. Не умаляя общности, считаем, что множества внутренних состояний машин не пересекаются. Программа машины \mathcal{T} строится следующим образом. Полагаем

$$q_1 = q_1^{\mathcal{T}_1}, \quad q_0 = q_0^{\mathcal{T}_2}, \quad q_0^{\mathcal{T}_1} = q_1^{\mathcal{T}_2}. \quad (3.79)$$

Модифицированные с учетом (3.79) команды для обеих машин \mathcal{T}_1 и \mathcal{T}_2 объединяем в одну программу.

Рассмотрим начальную конфигурацию $\mathcal{K}_1 = q_1\alpha$. В силу (3.79) машина \mathcal{T} начинает работать, как машина \mathcal{T}_1 , и, если последняя применима к \mathcal{K}_1 ,

то на некотором шаге будет получена конфигурация

$$\mathcal{K}_2 = q_0^{\mathcal{T}_1} f_1(\alpha) = q_1^{\mathcal{T}_2} f_1(\alpha),$$

являющаяся стартовой для \mathcal{T}_2 . Теперь машина \mathcal{T} действует как машина \mathcal{T}_2 .

Если \mathcal{T}_2 применима к конфигурации \mathcal{K}_2 , то за конечное число шагов будет получена конфигурация $\mathcal{K}_3 = q_0^{\mathcal{T}_2} f_2(f_1(\alpha))$, которая, согласно (3.79) является заключительной для машины \mathcal{T} .

Если машина \mathcal{T}_1 неприменима к конфигурации \mathcal{K}_1 или машина \mathcal{T}_2 неприменима к конфигурации \mathcal{K}_2 , то машина \mathcal{T} неприменима к конфигурации \mathcal{K}_1 .

Машину \mathcal{T} называют *композицией* (*суперпозицией*) машин \mathcal{T}_1 и \mathcal{T}_2 и обозначают $\mathcal{T} = \mathcal{T}_2 \circ \mathcal{T}_1$ или $\mathcal{T} = \mathcal{T}_2(\mathcal{T}_1)$. Схематически работу машины \mathcal{T} можно изобразить следующим образом:

$$q_1 \alpha = q_1^{\mathcal{T}_1} \alpha \xrightarrow{\mathcal{T}_1} q_0^{\mathcal{T}_1} f_1(\alpha) = q_1^{\mathcal{T}_2} f_1(\alpha) \xrightarrow{\mathcal{T}_2} q_0^{\mathcal{T}_2} f_2(f_1(\alpha)) = q_0 f_2(f_1(\alpha)).$$

► **Разветвление машин Тьюринга.** Пусть машины Тьюринга \mathcal{T}_1 и \mathcal{T}_2 , вычисляют функции $f_1(\alpha)$ и $f_2(\alpha)$ соответственно. Тогда существует машина Тьюринга \mathcal{T} , которая начальную конфигурацию $\mathcal{K}_1 = q_1 \lambda * \alpha$, где $\lambda \in \{0, 1\}$ переводит в заключительную $\mathcal{K}_0 = q_0 f_1(\alpha)$, если $\lambda = 0$ и в $\mathcal{K}_0 = q_0 f_2(\alpha)$, если $\lambda = 1$. Машину \mathcal{T} называют *разветвлением* машин и обозначают $\mathcal{T}_1 \vee \mathcal{T}_2$.

Как и ранее, состояния машин \mathcal{T}_1 и \mathcal{T}_2 будем обозначать верхними индексами, с номерами машин. Полагаем также, что множества внутренних состояний машин не пересекаются. Программа машины \mathcal{T} строится следующим образом. Объединим программы машин \mathcal{T}_1 и \mathcal{T}_2 , добавим новые начальное и заключительное состояния q_1, q_0 и следующие команды:

$$\begin{aligned} q_1 0 &\rightarrow q_1^{\mathcal{T}_1} \alpha_0 R, & q_1^{\mathcal{T}_1} * &\rightarrow q_1^{\mathcal{T}_1} \alpha_0 R, & q_1 1 &\rightarrow q_1^{\mathcal{T}_2} \alpha_0 R, & q_1^{\mathcal{T}_2} * &\rightarrow q_1^{\mathcal{T}_2} \alpha_0 R, \\ q_0 &= q_0^{\mathcal{T}_1} = q_0^{\mathcal{T}_2}. \end{aligned} \tag{3.80}$$

Если $\mathcal{K}_1 = q_1 \lambda * \alpha$ — начальная конфигурация, то \mathcal{T} , согласно (3.80), через два шага перейдет в конфигурацию $\mathcal{K}_2 = q_1^{\mathcal{T}_1} \alpha$, если $\lambda = 0$, и в конфигурацию $\mathcal{K}_2 = q_1^{\mathcal{T}_2} \alpha$ при $\lambda = 1$, а затем будет работать как \mathcal{T}_1 или \mathcal{T}_2 соответственно. Схематически работу машины можно представить так:

$$q_1 \lambda * \alpha \xrightarrow{\mathcal{T}_1 \vee \mathcal{T}_2} \begin{cases} q_1 0 * \alpha \rightarrow q_1^{\mathcal{T}_1} * \alpha \rightarrow q_1^{\mathcal{T}_1} \alpha \rightarrow q_0^{\mathcal{T}_1} f_1(\alpha) = q_0 f_1(\alpha), & \text{при } \lambda = 0, \\ q_1 1 * \alpha \rightarrow q_1^{\mathcal{T}_2} * \alpha \rightarrow q_1^{\mathcal{T}_2} \alpha \rightarrow q_0^{\mathcal{T}_2} f_2(\alpha) = q_0 f_2(\alpha), & \text{при } \lambda = 1. \end{cases}$$

► **Машина Тьюринга, реализующая цикл.** Представим работу некоторого цикла следующим образом. Пусть на множестве \mathcal{A}^* заданы

функции $f_1(\alpha)$, $f_2(\alpha)$ и предикат $P(\alpha)$. Для произвольного слова α проверяется условие $P(\alpha) = 1$. Если оно выполнено, то цикл завершает работу и результат его работы есть $f_1(\alpha)$. Если же $P(\alpha) = 0$, то полагаем $\alpha = f_2(\alpha)$ и вся описанная процедура повторяется.

Построим машину \mathcal{T} , реализующую данную процедуру. Обозначим машины Тьюринга для вычисления функций f_1 , f_2 и предиката P соответственно \mathcal{T}_1 , \mathcal{T}_2 и \mathcal{T}_P . Пусть \mathcal{T}_0 — машина, которая добавляет к текущей конфигурации справа разделитель и слово α . Машина \mathcal{T} строится в соответствии со схемой (рис. 3.34).

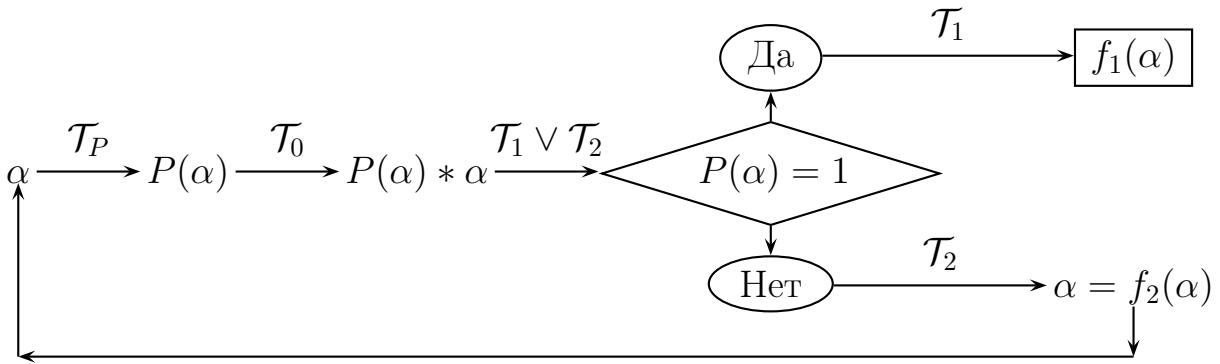


Рис. 3.34

Положим

$$q_0 = q_0^{\mathcal{T}_1}, \quad q_1 = q_1^{\mathcal{T}_P}, \quad q_1^{\mathcal{T}_0} = q_0^{\mathcal{T}_P}, \quad q_1^{\mathcal{T}_1 \vee \mathcal{T}_2} = q_0^{\mathcal{T}_0}.$$

Тогда, если $\mathcal{T}_1 \vee \mathcal{T}_2$ работает как \mathcal{T}_1 , то полученное ею значение является выходом машины \mathcal{T} , если же $\mathcal{T}_1 \vee \mathcal{T}_2$ работает как \mathcal{T}_2 , то полученное ею значение снова подается на вход машины \mathcal{T} .

Упражнение 3.23. Используя цикл, постройте машину Тьюринга, реализующую перевод произвольного числа $n \geq 1$, заданного в унарной записи, т. е. в виде 1^{n+1} , в его двоичную запись, начинаяющуюся с 1.

Замечание 3.64. Таким образом, машины Тьюринга позволяют осуществлять последовательное выполнение программ (композиция), использовать условные переходы (разветвление), реализовывать циклы, т. е. содержат основные операторы программирования.

Это дает основания предположить, что для любого алгоритма существует реализующая его машина Тьюринга. Данное предположение известно как *тезис Тьюринга*. Строго доказать его нельзя, поскольку для формулировки используется интуитивное понятие алгоритма. Однако принятие тезиса Тьюринга позволяет ставить знак равенства между утверждением о несуществовании машины Тьюринга для решения некоторой задачи и утверждением об отсутствии алгоритма ее решения вообще.

3.7.4. Проблема останова машины Тьюринга

Зададимся вопросом: существует ли алгоритм (машина Тьюринга), позволяющий по описанию произвольного алгоритма и его исходных дан-

ных (алгоритм и данные заданы символами на ленте машины Тьюринга) определить, останавливается ли этот алгоритм на этих данных или работает бесконечно. Данная задача получила название *проблемы останова*.

Введем необходимые определения.

Предположим, что существуют общие внешний и внутренний алфавиты \mathcal{A}_0 и \mathcal{Q}_0 , которые включают символы для всех машин Тьюринга. Тогда для произвольной команды машины Тьюринга в виде (3.77) символы $q_i, \alpha_j, q_k, \alpha_l$ есть символы из \mathcal{A}_0 и \mathcal{Q}_0 . Укажем способ нумерации всех машин Тьюринга (*геделева нумерация*¹).

Рассмотрим последовательность $\{p_i\}, i = 1, 2, \dots$, — всех простых чисел в порядке возрастания, т. е. последовательность 2, 3, 5, 7, 11, 13, ...

Определение 3.99. Пусть машина Тьюринга \mathcal{T} определена набором команд вида (3.77):

$$q_{i_1} \alpha_{j_1} \rightarrow q_{k_1} \alpha_{l_1} M_{s_1}, q_{i_2} \alpha_{j_2} \rightarrow q_{k_2} \alpha_{l_2} M_{s_2}, \dots, q_{i_m} \alpha_{j_m} \rightarrow q_{k_m} \alpha_{l_m} M_{s_m}. \quad (3.81)$$

Ее номером будем называть число

$$n(\mathcal{T}) = \prod_{t=1}^m p_{5t-4}^{i_t} p_{5t-3}^{j_t} p_{5t-2}^{k_t} p_{5t-1}^{l_t} p_{5t}^{s_t}, \text{ где } i_t, k_t \in \mathcal{Q}_0, j_t, l_t \in \mathcal{A}_0, s_t \in \{1, 2, 3\}.$$

Замечание 3.65. Очевидно, что если $n(\mathcal{T})$ — номер машины Тьюринга, то по нему можно однозначно восстановить набор команд (3.81). Таким образом, есть взаимно-однозначное соответствие между номером и программой машины Тьюринга.

Определение 3.100. Машину \mathcal{T} применимую к слову $n(\mathcal{T})$, называют *самоприменимой*, в противном случае — *несамоприменимой*.

Пример 3.130. Очевидно, что машины \mathcal{T}_1 и \mathcal{T}_2 из примера 3.125 являются самоприменимыми.

Примером несамоприменимой машины является машина, в правых частях команд которой не встречается заключительное состояние q_0 . Такая машина неприменима ни к одной конфигурации.

Существует ли алгоритм, который по любой машине Тьюринга устанавливает, самоприменима она или нет? Эту задачу называют *проблемой самоприменимости*.

Согласно тезису Тьюринга (см. замечание 3.64), такой алгоритм следует искать в виде машины Тьюринга, т. е. требуется построить машину Тьюринга \mathcal{T}_A , которая была бы применима к номерам всех машин Тьюринга и в зависимости от самоприменимости машины с данным номером имела бы различные заключительные конфигурации. Пусть, например, в случае самоприменимой машины заключительная конфигурация имеет вид q_01 , а в случае несамоприменимой машины — q_00 .

¹Название связано с именем К. Геделя, впервые в 1931 г. предложившего идею кодирования произвольных объектов натуральными числами.

Теорема 3.41. Машины Тьюринга \mathcal{T}_A , решающей проблему самоприменимости, не существует.

Доказательство. Предположим противное. Используя \mathcal{T}_A , построим машину \mathcal{T}_S со следующими свойствами:

- а) \mathcal{T}_S применима ко всем номерам несамоприменимых машин;
- б) \mathcal{T}_S неприменима ко всем номерам самоприменимых машин.

Машину \mathcal{T}_S получается из \mathcal{T}_A следующим образом: заключительное состояние $q_0^{\mathcal{T}_A}$ машины \mathcal{T}_A считается незаключительным состоянием машины \mathcal{T}_S , ее заключительное состояние обозначим $q_0^{\mathcal{T}_S}$. К программе машины \mathcal{T}_A добавляются две команды:

$$q_0^{\mathcal{T}_A} 1 \rightarrow q_0^{\mathcal{T}_A} 1 E, \quad q_0^{\mathcal{T}_A} 0 \rightarrow q_0^{\mathcal{T}_S} 0 E. \quad (3.82)$$

Очевидно, что в силу (3.82) машина \mathcal{T}_S удовлетворяет условиям (а) и (б).

Сама машина \mathcal{T}_S либо самоприменима, либо несамоприменима. Если она самоприменима, то тем самым применима к номеру некоторой самоприменимой машины, но тогда нарушается требование (б). Если же машина \mathcal{T}_S несамоприменима, тогда не выполняется требование (а).

Полученное противоречие доказывает теорему. ■

Используя теорему 3.41 докажем неразрешимость *проблемы останова*.

Будем считать, что описание произвольного алгоритма (машины Тьюринга) задается ее номером. Сформулируем задачу: существует ли машина Тьюринга \mathcal{T}_S , которая была бы применима ко всем словам вида $n(\mathcal{T}) * \alpha$, где $n(\mathcal{T})$ — номер произвольной машины, α — произвольное слово, и в случае, если машина \mathcal{T} применима к слову α , приходила бы к заключительной конфигурации $q_0 1$, а в случае неприменимости — к конфигурации $q_0 0$.

Теорема 3.42. Машины Тьюринга \mathcal{T}_S , решающей проблему останова, не существует.

Доказательство. Предположим, что машина \mathcal{T}_S существует. Рассмотрим композицию машин $\mathcal{T}_D = \mathcal{T}_S \circ \mathcal{T}_{copy}$, где \mathcal{T}_{copy} — машина, которая выполняет копирование аргумента на ленте из задачи 3.128. Запишем на ленте номер произвольной машины Тьюринга $n(\mathcal{T})$ и начнем работу машины \mathcal{T}_D с конфигурации $q_1^{\mathcal{T}_D} n(\mathcal{T})$.

После того, как машина \mathcal{T}_{copy} завершит работу, на ленте будет слово $n(\mathcal{T}) * n(\mathcal{T})$. По предположению машина \mathcal{T}_S применима ко всем таким словам, следовательно, машина \mathcal{T}_D остановится в конфигурации $q_0^{\mathcal{T}_D} 1$, если машина \mathcal{T} применима к слову $n(\mathcal{T})$ и в конфигурации $q_0^{\mathcal{T}_D} 0$, если неприменима. Таким образом, построенная машина \mathcal{T}_D решает проблему самоприменимости, что противоречит теореме 3.41. ■

3.7.5. Частично рекурсивные функции

Рассмотрим второй подход (исторически он был первым) для формализации понятия алгоритма — класс частично рекурсивных функций. Функции данного класса строятся из базисных функций с помощью вычислимых операторов. Как и в предыдущем подразделе будем рассматривать функции $f : N_0^n \rightarrow N_0$.

Множество базисных функций таково:

- 1) *нуль-функция*: $0(x) = 0$;
- 2) *одноместная функция следования*: $s(x) = x + 1$;
- 3) *проектирующая функция*: $I_m^n(x_1, \dots, x_n) = x_m$, $1 \leq m \leq n$.

Вычислимые операторами являются операторы *суперпозиции*, *примитивной рекурсии* и *минимизации*.

1. *Оператор суперпозиции*. Будем говорить, что функция

$$h(x_1, \dots, x_m) = g(f_1(x_1, \dots, x_m), \dots, f_n(x_1, \dots, x_m))$$

получается с помощью оператора суперпозиции из функций g, f_1, \dots, f_n ($h = S(g, f_1, \dots, f_n)$).

Пример 3.131. Функция $f(x_1, x_2) = x_1 + 1$ может быть получена как $s(I_1^2(x_1, x_2))$.

2. *Оператор примитивной рекурсии*. Функция $f(x_1, \dots, x_n, y)$ получается из функций $g(x_1, \dots, x_n)$ и $h(x_1, \dots, x_n, y, z)$ с помощью оператора примитивной рекурсии ($f = R(g, h)$), если она может быть определена следующей схемой:

$$\begin{cases} f(x_1, \dots, x_n, 0) = g(x_1, \dots, x_n); \\ f(x_1, \dots, x_n, y + 1) = h(x_1, \dots, x_n, y, f(x_1, \dots, x_n, y)). \end{cases}$$

Пример 3.132. Функция $f(x_1, x_2) = x_1 + x_2$ задается схемой примитивной рекурсии

$$\begin{cases} x_1 + 0 = I_1^1(x_1); \\ x_1 + (x_2 + 1) = s(I_3^3(x_1, x_2, x_1 + x_2)). \end{cases}$$

Таким образом, $f = R(I_1^1, s(I_3^3))$.

3. *Оператор минимизации.* Функция $f(x_1, \dots, x_n)$ получается из функции $g(x_1, \dots, x_n, y)$ с помощью оператора минимизации ($f = M_y g$)

$$f(x_1, \dots, x_n) = \mu y [g(x_1, \dots, x_n, y) = 0],$$

если значение $f(x_1, \dots, x_n)$ определено и равно y тогда и только тогда, когда значения $g(x_1, \dots, x_n, 0), \dots, g(x_1, \dots, x_n, y-1)$ определены и отличны от 0, а $g(x_1, \dots, x_n, y) = 0$.

Оператор минимизации иногда называют μ -оператором.

Замечание 3.66. Очевидно, что операторы S и R , применяемые ко всюду определенным функциям, дают всюду определенные функции, однако μ -оператор может давать частичные функции даже при применении ко всюду определенным функциям.

Дадим теперь основное определение.

Определение 3.101. Функцию $f(x_1, \dots, x_n)$ называют *примитивно рекурсивной*, если она может быть получена из базисных функций с помощью конечного числа применений операторов суперпозиции и примитивной рекурсии.

Функцию $f(x_1, \dots, x_n)$ называют *частично рекурсивной*, если она может быть получена из базисных функций с помощью конечного числа применений операторов суперпозиции, примитивной рекурсии и минимизации.

Функцию $f(x_1, \dots, x_n)$ называют *общерекурсивной*, если она частично рекурсивна и всюду определена.

Замечание 3.67. Из замечания 3.66 следует, что всякая примитивно-рекурсивная функция общерекурсивна.

По аналогии с тезисом Тьюринга приведем *тезис Чёрча*: класс алгоритмически вычислимых функций совпадает с классом всех частично рекурсивных функций [56].

Принятие данного тезиса позволяет считать доказательство того, что некоторая функция не является частично рекурсивной, доказательством отсутствия алгоритма вычисления ее значений.

Пример 3.133. Установим примитивную рекурсивность некоторых числовых функций.

1. Функция $f(x) = m = s(s(\dots s(0(x)) \dots))$.
2. Функция $f(x_1, x_2) = x_1 x_2$.

В примере 3.132 было показано, что $h(x_1, x_2, x_3) = x_1 + x_3$ частично рекурсивная функция. Тогда $f(x_1, x_2)$ задается схемой примитивной рекурсии

$$\begin{cases} x_1 \cdot 0 = 0(x_1); \\ x_1(x_2 + 1) = h(x_1, x_2, x_1 x_2) = x_1 + x_1 x_2. \end{cases}$$

3. Функция $f(x_1, x_2) = x_1^{x_2}$.

Из 2 следует, что $h(x_1, x_2, x_3) = x_1 x_3$ частично рекурсивная функция. Тогда $f(x_1, x_2)$ задается схемой примитивной рекурсии

$$\begin{cases} x_1^0 = s(0(x)); \\ x_1^{x_2+1} = h(x_1, x_2, x_1^{x_2}). \end{cases}$$

Упражнение 3.24. Установите частичную рекурсивность функции $f(x) = \text{sign}(x)$ из примера 3.43.

3.7.6. Нормальные алгорифмы Маркова

Рассмотрим еще одну алгоритмическую модель, предложенную А. А. Марковым [30] и называемую *нормальные алгорифмы*¹. Данный подход представляет интуитивное понятие алгоритма, как переработку слов в некотором конечном алфавите по определенным правилам. В качестве примитивных операций используются подстановки одного слова вместо другого. Множество таких подстановок определяет схему алгоритма.

Дадим необходимые определения. Напомним, что, согласно определению 2.22, словом (цепочкой) в алфавите \mathcal{A} называют последовательность символов из этого алфавита. Пустое слово обозначают символом \wedge или ε . Множество слов алфавита \mathcal{A} обозначают \mathcal{A}^* .

Определение 3.102. Рассмотрим алфавит $\mathcal{A} = \{a_1, \dots, a_n\}$. Если $P, Q \in \mathcal{A}^*$, то выражения

$$P \rightarrow Q, \quad P \rightarrow \cdot Q$$

называют *формулами подстановки* в алфавите \mathcal{A} . При этом формулу $P \rightarrow Q$ называют *простой*, а формулу $P \rightarrow \cdot Q$ — *заключительной*. В общем виде будем писать $P \rightarrow (\cdot)Q$. Конечную последовательность таких формул называют *схемой* $\mathcal{S}_{\mathcal{N}}$ *нормального алгорифма* \mathcal{N} . Таким образом, схема $\mathcal{S}_{\mathcal{N}}$ имеет вид:

$$P_1 \rightarrow (\cdot)Q_1, \quad P_2 \rightarrow (\cdot)Q_2, \quad \dots, \quad P_m \rightarrow (\cdot)Q_m. \quad (3.83)$$

Замечание 3.68. В определении 3.102 предполагается, что знаки \rightarrow и \cdot не входят в алфавит \mathcal{A} .

Определение 3.103. Слово Q называют *подсловом* слова R , если справедливо представление

$$R = V_1 Q V_2, \quad Q, R, V_1, V_2 \in \mathcal{A}^*.$$

¹При транскрипции греческих слов латинскими буквами, звук «ф» передается сочетанием «th», так что русское прочтение слова **алгорифм**, как алгорифм имеет некоторое основание.

Определение 3.104. Рассмотрим формулу подстановки $P_i \rightarrow (\cdot) Q_i$ схемы $\mathcal{S}_{\mathcal{N}}$ нормального алгорифма \mathcal{N} . Пусть слово $R \in \mathcal{A}^*$ содержит подслово P_i (таких подслов в слове R может быть несколько).

Подстановкой называют замену первого по порядку подслова P_i исходного слова R на слово Q_i . В этом случае будем говорить, что формула подстановки действует на слово R .

Схема $\mathcal{S}_{\mathcal{N}}$ действует на слово R , если в схеме существует формула подстановки, действующая на это слово.

Замечание 3.69. В формулах подстановок $P \rightarrow (\cdot) Q$ слова P и(или) Q могут быть пустыми. Если слово Q пустое, то результатом действия такой подстановки на слово $R = V_1 Q V_2$ будет слово $V_1 V_2$. Если пустым является слово P , то считается, что результатом подстановки является слово QR , т. е. правая часть формулы подстановки приписывается слева к слову R .

Работа схемы $\mathcal{S}_{\mathcal{N}}$ нормального алгорифма \mathcal{N} применительно к слову R может быть описана следующим образом.

Если в схеме $\mathcal{S}_{\mathcal{N}}$ имеются формулы подстановок, действующие на слово R , то первая из них применяется к R , в результате чего оно переходит в другое слово R_1 . К нему вновь применяется схема подстановок и т. д. Процесс прекращается в двух случаях:

- схема $\mathcal{S}_{\mathcal{N}}$ не действует на очередное слово R_n ;
- при получении R_n была применена заключительная подстановка.

Если процесс не заканчивается за конечное число шагов, то говорят, что алгорифм \mathcal{N} неприменим к слову R , обозначая это следующим образом: $\mathcal{S}_{\mathcal{N}} : R \dashv$

Если же для слова R процесс заканчивается некоторым результатом $\tilde{R} = R_n$, то говорят, что алгорифм \mathcal{N} применим к слову R и обозначают этот факт так: $\mathcal{S}_{\mathcal{N}} : R \models \tilde{R}$.

Определение 3.105. Словарную функцию $f : \mathcal{A}^* \rightarrow \mathcal{A}^*$ называют вычислимой по Маркову, если существует схема $\mathcal{S}_{\mathcal{N}}$ нормального алгорифма \mathcal{N} , такая, что

$$f(P) = Q \iff \mathcal{S}_{\mathcal{N}} : P \models Q, \forall P, Q \in \mathcal{A}^*.$$

Аналогично тезисам Тьюринга и Чёрча, А. А. Марков предложил принцип нормализации, который утверждает, что для любого алгоритма в произвольном конечном алфавите можно построить эквивалентный ему нормальный алгорифм в том же алфавите.

Принятие данного принципа позволяет принимать доказательство non-existence нормального алгоритма как утверждение о non-existence алгоритма вообще.

Пример 3.134. Рассмотрим алгорифм \mathcal{N} со схемой $\mathcal{S}_{\mathcal{N}} : \wedge \rightarrow \cdot \wedge$ в произвольном алфавите \mathcal{A} . Данный алгорифм вычисляет тождественную функцию $f(P) = P$ для любого слова P .

Если же взять схему $\mathcal{S}_{\mathcal{N}} : \wedge \rightarrow \wedge$, то соответствующий алгорифм вычисляет нигде не определенную функцию.

Пример 3.135. Пусть алфавит $\mathcal{A} = \{*, 1\}$. Рассмотрим алгорифм \mathcal{N} со схемой $\mathcal{S}_{\mathcal{N}} : *1 \rightarrow \wedge, \wedge \rightarrow \cdot \wedge$.

Найдем результат работы алгорифма для слова $R = 111 * 1111 * 11$. Имеем

$$\overbrace{111 * 1111 * 11}^{2+3+1} = 111 \frac{*1}{\wedge} 111 * 11 \rightarrow 111111 \frac{*1}{\wedge} 1 \rightarrow 111111 \rightarrow \cdot \overbrace{111111}^6,$$

т. е. алгорифм \mathcal{N} суммирует числа в унарной системе счисления.

Пример 3.136. Рассмотрим алфавит $\mathcal{A} = \{\alpha_1, \alpha_2\}$ и алгорифм \mathcal{N} со схемой $\mathcal{S}_{\mathcal{N}} : \alpha_1 \rightarrow \cdot \wedge, \alpha_2 \rightarrow \alpha_2$.

Данный алгорифм слово P переводит в слово Q , полученное из P путем вычеркивания первого вхождения буквы α_1 . Алгорифм \mathcal{N} неприменим к словам, не содержащим вхождений буквы α_1 .

Пример 3.137. Пусть алфавит \mathcal{A} содержит символы русского языка: $\mathcal{A} = \{а, б, \dots, я\}$. Рассмотрим алгорифм \mathcal{N} со схемой $\mathcal{S}_{\mathcal{N}} : \pi \rightarrow м, \wedge \rightarrow \cdot \wedge$.

Применим алгорифм к слову **п а п а**. Получим:

$$\text{п а п а} = \frac{\pi}{м} \text{ а п а} \rightarrow \frac{\pi}{м} \text{ а} = \text{м а м а} \rightarrow \cdot \text{м а м а}.$$

Пример 3.138. Построим алгорифм \mathcal{N} обращения слова P в алфавите $\mathcal{A} = \{\alpha_1, \dots, \alpha_n\}$, т. е. слова из тех же букв, но записанных в обратном порядке (обозначаем как P^{-1}).

Добавим к алфавиту \mathcal{A} два служебных символа $\mathcal{B} = \mathcal{A} \cup \{\beta, \gamma\}$ и рассмотрим алгорифм \mathcal{N} со схемой $\mathcal{S}_{\mathcal{N}}$:

- 1) $\beta\beta \rightarrow \gamma$;
- 2) $\gamma\alpha_i \rightarrow \alpha_i\gamma$, $\forall \alpha_i \in \mathcal{A}$;
- 3) $\gamma\beta \rightarrow \gamma$;
- 4) $\gamma \rightarrow \cdot \wedge$;
- 5) $\beta\alpha_i\alpha_j \rightarrow \alpha_j\beta\alpha_i$, $\forall \alpha_i, \alpha_j \in \mathcal{A}$;
- 6) $\wedge \rightarrow \beta$.

Пусть $P = \alpha_{j_0} \dots \alpha_{j_k} \in \mathcal{A}^*$. Как и прежде над стрелкой \rightarrow будем обозначать номер примененной формулы. Тогда

$$\begin{aligned} P &\xrightarrow{6} \beta P \xrightarrow{5} \alpha_{j_1}\beta\alpha_{j_0}\alpha_{j_2} \dots \alpha_{j_k} \xrightarrow{5} \alpha_{j_1}\alpha_{j_2}\beta\alpha_{j_0}\alpha_{j_3} \dots \alpha_{j_k} \xrightarrow{5} \dots \xrightarrow{5} \alpha_{j_1}\alpha_{j_2} \dots \alpha_{j_k}\beta\alpha_{j_0} \\ &\xrightarrow{6} \beta\alpha_{j_1}\alpha_{j_2} \dots \alpha_{j_k}\beta\alpha_{j_0} \xrightarrow{5} \dots \xrightarrow{5} \alpha_{j_2}\alpha_{j_3} \dots \alpha_{j_k}\beta\alpha_{j_1}\beta\alpha_{j_0}. \end{aligned}$$

Повторяя этот процесс, получим

$$\begin{aligned} \beta\alpha_{j_k}\beta\alpha_{j_{k-1}}\dots\beta\alpha_{j_1}\beta\alpha_{j_0} &\xrightarrow{6} \beta\beta\alpha_{j_k}\beta\alpha_{j_{k-1}}\dots\beta\alpha_{j_1}\beta\alpha_{j_0} \xrightarrow{4} \gamma\alpha_{j_k}\beta\alpha_{j_{k-1}}\dots\beta\alpha_{j_1}\beta\alpha_{j_0} \\ &\xrightarrow{2} \alpha_{j_k}\gamma\beta\alpha_{j_{k-1}}\dots\beta\alpha_{j_1}\beta\alpha_{j_0} \xrightarrow{3} \alpha_{j_k}\gamma\alpha_{j_{k-1}}\dots\beta\alpha_{j_1}\beta\alpha_{j_0} \\ &\xrightarrow{234} \dots \xrightarrow{234} \alpha_{j_k}\alpha_{j_{k-1}}\dots\alpha_{j_1}\alpha_{j_0} = P^{-1}. \end{aligned}$$

Пример 3.139. Построим алгоритм инвертирования булевых наборов, т. е. перевод строки $\{\alpha_1, \dots, \alpha_n\}$, где $\alpha_i \in \{0, 1\}, 1 \leq i \leq n$ в двойственную строку $\{\overline{\alpha_1}, \dots, \overline{\alpha_n}\}$

Добавим к алфавиту служебный символ * и рассмотрим алгорифм \mathcal{N} со схемой $\mathcal{S}_{\mathcal{N}}$:

- 1) $*0 \rightarrow 1*$
- 2) $*1 \rightarrow 0*$ *индикатор позиции ** движется вдоль строки меняя α_i на $\overline{\alpha_i}$.
- 3) $* \rightarrow \cdot \wedge$ *удаляем служебный символ ** индикатора позиции и заканчиваем работу алгоритма.
- 4) $\wedge \rightarrow *$ *ставим служебный символ ** индикатора позиции в начало слова.

Применяя описанный алгоритм для слова $P = \{1101\}$, получаем:

$$\{1101\} \rightarrow \{*1101\} \rightarrow \{0*101\} \rightarrow \dots \rightarrow \{0010*\} \rightarrow \{\cdot 0010\}$$

Пример 3.140. Построим алгоритм сортировки числовых строк в порядке возрастания компонент, для простоты рассмотрим строки, содержащие три символа 0, 1, 2.

Рассмотрим алгорифм \mathcal{N} со схемой $\mathcal{S}_{\mathcal{N}}$:

$$20 \rightarrow 02, \quad 10 \rightarrow 01, \quad 21 \rightarrow 12$$

Применим данный алгоритм к слову $P = \{2110\}$, имеем:

$$\{2110\} \rightarrow \{2101\} \rightarrow \{2011\} \rightarrow \{0211\} \rightarrow \{0121\} \rightarrow \{0112\}$$

3.7.7. Примеры и задачи на нормальные алгорифмы Маркова

Пример 3.141. Рассмотрим алфавит скобочных выражений $\mathcal{A} = \{(,)\}$. Построить нормальный алгоритм Маркова, определяющий *правильные скобочные выражения*: строки правильных скобочных выражений (и только они) должны преобразовываться в пустое слово.

Решение представляет собой схему из одной команды:

$$() \rightarrow \Lambda$$

Пример 3.142. Рассмотрим алфавит $\mathcal{A} = \{0, 1, 2, 3\}$. Пусть P – непустое слово. Считая его записью неотрицательного целого числа в четверичной системе счисления, требуется получить запись этого же числа, но в двоичной системе. Например: $0123 \rightarrow 00011011$.

Такая замена, казалось бы, реализуется следующим нормальным алгорифмом Маркова:

$$0 \rightarrow 00, 1 \rightarrow 01, 2 \rightarrow 10, 3 \rightarrow 11 \quad (3.84)$$

Но алгоритм (3.84) работает неправильно, в чем можно убедиться на входном слове 0123. Действительно,

$$0123 \rightarrow 00123 \rightarrow 000123 \rightarrow \dots$$

Ошибка здесь в том, что после замены четверичной цифры на пару двоичных цифр уже никак нельзя отличить двоичные цифры от четверичных, поэтому наш алгоритм начинает заменять и двоичные цифры. Значит, надо как-то отделить ту часть числа, в которой уже была произведена замена, от той части, где замены ещё не было. Для этого предлагается пометить слева дополнительным символом * ту четверичную цифру, которая сейчас должна быть заменена на пару соответствующих двоичных цифр, а после того как такая замена будет выполнена, дополнительный символ нужно поместить перед следующей четверичной цифрой:

$$0123 \rightarrow 00 * 0123 \rightarrow 00 * 123 \rightarrow 0001 * 23 \rightarrow 000110 * 3 \rightarrow 00011011 *$$

Как видно, слева от * всегда находится та часть числа, которая уже переведена в двоичный вид, а справа – часть, которую ещё предстоит заменить. Поэтому никакой путаницы между четверичными и двоичными цифрами уже не будет. Итак, * сначала нужно разместить слева от первой цифры четверичного числа, а затем он должен *перепрыгивать* через очередную четверичную цифру, оставляя слева от себя соответствующие ей двоичные цифры. В конце, когда справа от * уже не окажется никакой цифры, этот символ надо удалить и остановиться.

Получаем следующий нормальный алгорифм Маркова:

$$*0 \rightarrow 00*, *1 \rightarrow 01*, *2 \rightarrow 10*, *3 \rightarrow 11*, * \rightarrow \cdot\Lambda, \Lambda \rightarrow * \quad (3.85)$$

Проверим алгоритм (3.85) на входном слове 0123:

$$0123 \rightarrow *0123 \rightarrow 00*123 \rightarrow 0001*23 \rightarrow 000110*3 \rightarrow 00011011* \rightarrow 00011011$$

Пример 3.143. Пусть $\mathcal{A} = \{a, b\}$. Требуется приписать символ a к концу слова P .

Чтобы приписать символ a справа, надо сначала пометить конец слова. Для этого, как и в предыдущем примере воспользуемся дополнительным

знаком $*$, который сначала поместим в конец слова P , а затем заменим его на символ q .

Но как поместить $*$ в конец слова? Сделаем это следующим образом: сначала $*$ припишем слева к слову P , а затем *перегоним* ее через все буквы слова:

$$bbab \rightarrow *bbab \rightarrow b * bab \rightarrow bb * ab \rightarrow bba * b \rightarrow bbab *$$

Заметим, что *перегон* символа $*$ через какой-то символ α – это замена пары $*\alpha$ на пару $\alpha*$, которая реализуется заменой $*\alpha \rightarrow *\alpha$. таким образом, следующий алгоритм:

$$*a \rightarrow a*, *b \rightarrow b*, * \rightarrow \cdot a, \Lambda \rightarrow *$$

Задача 3.45. Пусть $\mathcal{A} = \{a, b, c\}$. Построить нормальный алгоритм Маркова, заменяющий все пары cb на a .

Задача 3.46. Пусть $\mathcal{A} = \{a, b, c\}$. Построить нормальный алгоритм Маркова, заменяющий любое входное слово на слово a .

Задача 3.47. Пусть $\mathcal{A} = \{0, 1, 2, 3\}$. Построить нормальный алгоритм Маркова, преобразующий слово так, чтобы сначала шли все четные цифры 0 и 2, а затем – все нечетные.

Задача 3.48. Пусть $\mathcal{A} = \{a, b, c\}$. Построить нормальный алгоритм Маркова, который за первым символом непустого слова вставляет символ c .

Задача 3.49. Пусть $\mathcal{A} = \{a, b, c\}$. Построить нормальный алгоритм Маркова, который удаляет из слова второй символ, если такой есть.

Задача 3.50. Пусть $\mathcal{A} = \{a, b, c\}$. Построить нормальный алгоритм Маркова, который удаляет из непустого слова его последний символ.

Задача 3.51. Пусть $\mathcal{A} = \{a, b, c\}$. Построить нормальный алгоритм Маркова, который оставляет в слове только первое вхождение символа a , если такое есть.

Упражнение 3.25. Построить нормальный алгоритм Маркова, который применим ко всем словам в алфавите $\mathcal{A} = \{a, b, c\}$, кроме двух слов – abc и $baac$.

Упражнение 3.26. Построить нормальный алгоритм Маркова, который из всех слов в алфавите $\mathcal{A} = \{a, b, c\}$ применим только к двум словам – пустому слову и слову $abccba$.

3.8. Элементы теории сложности алгоритмов

Если бы строители строили здания так же, как программисты пишут программы, первый залетевший дятел разрушил бы цивилизацию.

— Второй закон Вейнберга

3.8.1. Вычислительная сложность алгоритма

Определение 3.106. Назовем *размером входа* общее количество двоичных разрядов (n), необходимых для представления входов с помощью выбранного способа кодирования.

Наихудшим временем работы алгоритма будем считать верхнюю границу времени работы для произвольного входа, выраженную в виде функции размера входа $T(n)$.

Часто невозможно определить точное время работы алгоритма. В такой ситуации приходится приближенно оценивать время работы и обычно удается получить лишь асимптотические оценки при безграничном увеличении размера задачи.

Далее будем рассматривать функции, определенные на множестве натуральных чисел и принимающие натуральные значения, начиная с некоторых значений аргумента.

Определение 3.107. Ведем следующие обозначения (O -символика)

a) $f(n) = O(g(n))$ (*верхняя оценка*), если существует константа $\alpha > 0$ и $n_0 \in \mathbb{N}$, такие что

$$0 \leq f(n) \leq \alpha g(n), \quad n \geq n_0$$

b) $f(n) = \Omega(g(n))$ (*нижняя оценка*), если существует константа $\alpha > 0$ и $n_0 \in \mathbb{N}$, такие что

$$0 \leq \alpha g(n) \leq f(n), \quad n \geq n_0$$

c) $f(n) = \Theta(g(n))$ (*плотная оценка*), если существуют положительные константы α, β и $n_0 \in \mathbb{N}$, такие что

$$\alpha g(n) \leq f(n) \leq \beta g(n), \quad n \geq n_0$$

Замечание 3.70. Таким образом, $f(n) = O(g(n))$ означает, что $f(n)$ растет асимптотически не быстрее, чем $g(n)$ с точностью до мультипликативной константы, тогда как $f(n) = \Omega(g(n))$ означает, что $f(n)$ растет асимптотически по крайней мере так же быстро, как $g(n)$ с точностью до мультипликативной константы.

Определение 3.108. Будем говорить, что $T(n)$ алгоритма имеет *порядок роста* или *теоретическую сложность* $f(n)$, если $T(n) = O(f(n))$.

В порядке возрастания $T(n)$ выделяют следующие асимптотические классы сложности алгоритмов:

Константный $T(n) = O(1)$. Любой алгоритм, всегда требующий независимо от размера данных одного и того же времени, имеет константную сложность. Если не считать эффективность в наилучшем случае, в этот класс попадает очень небольшое количество алгоритмов.

Логарифмический $T(n) = O(\log(n))$. Эффективность алгоритма логарифмическая. Алгоритм начинает работать намного медленнее с увеличением размера входных данных (n). Такое время работы характерно для алгоритмов, в которых большая задача делится на маленькие, каждая из которых решается по отдельности.

Линейный $T(n) = O(n)$. Эффективность алгоритма меняется линейно в зависимости от размера входных данных (обычно такое наблюдается, когда каждый элемент входных данных требуется обработать линейное число раз). К этому классу относят алгоритмы, выполняющие сканирование списка, состоящего из n элементов (например, алгоритм поиска методом последовательного перебора).

Линейно-логарифмический $T(n) = O(n \log(n))$. Такая эффективность характерна для алгоритмов, которые делят большую проблему на маленькие, а затем, решив их, соединяют их решения. К этому классу относятся большое количество алгоритмов декомпозиции, таких как алгоритмы сортировки, быстрого преобразования Фурье.

Квадратичный $T(n) = O(n^2)$. Обычно, подобная зависимость характеризует эффективность алгоритмов, содержащих два встроенных цикла (например, ряд операций, выполняемых над матрицами размера $n \times n$).

Кубический $T(n) = O(n^3)$. Подобная зависимость характеризует эффективность алгоритмов, содержащих три встроенных цикла.

Экспоненциальный $T(n) = O(2^n)$. Данная зависимость типична для алгоритмов, выполняющих обработку всех подмножеств некоторого n -элементного множества. Часто термин «экспоненциальный» используется в широком смысле и означает очень высокий порядок роста.

Факториальный $T(n) = O(2!)$. Характеризует алгоритмы, выполняющие обработку всех перестановок некоторого n -элементного множества.

Замечание 3.71. Иногда квадратичный и кубический классы объединяют в *полиномиальный* класс: $T(n) = O(n^k)$, $k \geq 2$.

Пример 3.144. Алгоритм 1.11 вычисления a^n имеет логарифмическую сложность $T(n) = O(\log_2(n))$.

Вычисление чисел Фибоначчи по рекуррентному соотношению (1.46) имеет линейную сложность алгоритма $T(n) = O(n)$.

Замечание 3.72. Представим, что у нас есть алгоритм сложности 1.7^n для некоторой задачи, который за „разумное“ время позволяет решать примеры этой задачи размера не более n_0 .

- The „hardware“ approach: возьмем в 10 раз более быстрый компьютер. Теперь мы можем решать примеры размера $n_0 + 4$. Действительно, $1.7^4 \cong 10$.
- The „brainware“ approach: придумаем алгоритм сложности 1.3^n . Это позволит нам решать примеры размера $2n_0$. Действительно, $1.3^2 \cong 1.7$.

Таким образом, уменьшение основания экспоненты времени работы алгоритма увеличивает размер решаемых за данное время примеров **в константное число раз**, в то время как использование более быстрого компьютера способно увеличить размер **лишь на константу**.

3.8.2. Сложностные классы задач

Установление точных оценок сложности алгоритма, о которых речь шла в предыдущем разделе удается далеко не всегда. В работах Кобмена (Alan Cobham, 1964), и Эдмундса (Jack Edmonds, 1965) получил развитие другой подход — **сложностные классы задач**.

□ **Класс P (Polynomial)** — задачи с **полиномиальной сложностью**.

Определение 3.109. Задача называется *полиномиальной*, если существует константа k и алгоритм A , решающий задачу со сложностью $T_A(n) = O(n^k)$ (для большинства таких задач $k < 6$, это „практически разрешимые задачи“)

□ **Класс NP (Non-deterministic Polynomial)** —**полиномиально проверяемые задачи**.

Определение 3.110. Задача называется *полиномиальной проверяемой* если ее решение может быть проверено с помощью алгоритма полиномиальной сложности.

□ **Класс NP (NP-complete) – NP-полные задачи.**

Определение 3.111. Задача называется NP -полной, если:

- 1) она принадлежит к классу NP ;
- 2) к ней полиномиально сводятся все задачи из класса NP .

Полиномиальная сводимость одной задачи к другой означает существование алгоритма полиномиальной сложности, способного выполнить трансляцию описания исходной задачи на одном языке в эквивалентное ему описание на другом языке.

Пример 3.145. Значение класса NP -полных проблем состоит еще и в том, что ему принадлежат очень многие известные и важные в прикладном отношении задачи. Перечислим некоторые из них.

- *Гамильтонов цикл.* Существует ли в графе гамильтонов цикл – последовательность вершин и дуг (ребер) графа, содержащая все вершины графа по одному разу, но, может быть, содержащая не все дуги.

- *Задача коммивояжера.* Задан нагруженный график и некоторое натуральное число γ (иногда называемое *бюджетом*). Требуется ответить на вопрос: найдется ли в графике маршрут, проходящий через все вершины графа такой, что его длина не превышает γ ?

- *Задача пропозициональной выполнимости (Boolean satisfiability problem, SAT).* Определить, выполнима ли данная формула в конъюнктивной нормальной форме, то есть существует ли набор булевых значений переменным формулы, выполняющий формулу. Такой набор называют выполняющим (*satisfying assignment*), а формулу, для которой такой набор существует, — выполнимой (*satisfiable*).

После введения в теорию алгоритмов понятий сложностных классов была поставлена основная проблема теории сложности – $P = NP$?

Словесная формулировка проблемы имеет вид: можно ли все задачи, решение которых проверяется с полиномиальной сложностью, решить за полиномиальное время? Очевидно, что любая задача, принадлежащая классу P , принадлежит и классу NP , т. к. она может быть полиномиально проверена – задача проверки решения может состоять просто в повторном решении задачи.

На сегодня отсутствуют теоретические доказательства как совпадения этих классов ($P = NP$), так и их несовпадения. Однако тот факт, что ни для одной NP – полной задачи не найдено полиномиального алгоритма, косвенно подтверждает гипотезу строгого включения $P \subset NP$, $P \neq NP$. Математический институт Клэя включил эту проблему в список проблем тысячелетия.

3.9. Формальные языки и грамматики

Впервые понятие грамматики было формализовано лингвистами при изучении естественных языков. Предполагалось, что это может помочь при их автоматической трансляции. На интуитивном уровне язык можно определить как некоторое множество предложений допустимой структуры. Правила, определяющие допустимость той или иной конструкции языка, составляют синтаксис языка. Эти рассуждения можно отнести и к искусственным языкам программирования. Именно применительно к ним были достигнуты наилучшие результаты в этом направлении.

Если бы все языки состояли из небольшого конечного числа предложений, то проблемы синтаксического анализа не существовало. Можно было бы просто перечислить все предложения данного языка. Но так как большинство языков содержат неограниченное (или достаточно большое) число правильно построенных предложений, то возникает проблема разработки формальных средств описания и анализа языка (формальных грамматик).

Для формального описания языка необходимо задать набор символов (алфавит) языка и правила, по которым из символов строятся их последовательности (предложения), принадлежащие данному языку.

Существуют два вида описания языков:

- 1) *порождающая грамматика*, которая дает алгоритм, порождающий все правильные предложения языка;
- 2) *распознающая грамматика*, которая предполагает наличие алгоритма, определяющего принадлежность любого предложения данному языку.

Основные идеи формальных грамматик были заложены в работах Н. Хомского.

Историческая справка. Ноам Хомский, американский лингвист и общественный деятель, родился в Филадельфии 7 декабря 1928 г. Его отец эмигрировал из России незадолго до начала Первой мировой войны. С 1945 г. Хомский изучал в Пенсильванском университете лингвистику, математику и философию. Там же, в 1955 г. получил докторскую степень, однако большая часть исследований, положенных в основу его первой монографии [53], была выполнена в Гарвардском университете в 1951–1955 г.

Хомский – создатель системы грамматического описания, известной как *генеративная (порождающая) грамматика*; соответствующее направление лингвистики часто называют *генеративизмом*. Уже несколько десятилетий он является самым знаменитым лингвистом мира; ему и его теории посвящено множество статей и монографий, а развитие лингвистики в последней трети XX в. иногда называют «хомскианской революцией». Работы Хомского оказали влияние не только на лингвистику, но и инициировали развитие новых идей в философии, психологии, музыке. Например, Леонард Бернстайн использовал теорию Хомского для анализа музыкального ряда. Хомский также входит в первую десятку самых цитируемых в научном мире авторов. За резкость и безкомпромиссность суждений его часто называют «Че Гевара в лингвистике».

3.9.1. Основные определения порождающих грамматик

Напомним, что согласно определению 2.22, цепочкой в алфавите \mathcal{A} называют последовательность символов из этого алфавита. Сохраняя ранее введенные обозначения, пустое слово будем обозначать символом ϵ , а множество слов алфавита \mathcal{A} как \mathcal{A}^* .

Введем формальные определения.

Определение 3.112. *Длиной* цепочки называют число составляющих ее символов.

Например, если $\alpha = abcdabcd$, то длина α равна 8. Длину цепочки α будем обозначать $|\alpha|$. Длина пустой цепочки ϵ по определению равна 0.

Через $\mathcal{A}^+ \subset \mathcal{A}^*$ будем обозначать множество, содержащее все *непустые* цепочки конечной длины в алфавите \mathcal{A} .

Определение 3.113. *Порождающей грамматикой* G называют четверку объектов (V_T, V_N, P, S) , где

- 1) V_T — алфавит (словарь) терминальных символов (терминалов или примитивов);
- 2) V_N — алфавит (словарь) нетерминальных символов (нетерминалов), предполагается, что $V_T \cap V_N = \emptyset$;
- 3) множество P — конечное подмножество декартова произведения $(V_T \cup V_N)^+ \times (V_T \cup V_N)^*$. Элементы (α, β) множества P называют правилами вывода (продукционными правилами или продукциями) и записываются в виде $\alpha \rightarrow \beta$;
- 4) S — начальный символ (цель или аксиома) грамматики, $S \in V_N$.

Замечание 3.73. Для того чтобы различать терминальные и нетерминальные символы, принято обозначать терминальные символы строчными, а нетерминальные символы прописными буквами латинского алфавита.

Для записи n правил с одинаковыми левыми частями

$$\alpha \rightarrow \beta_1, \dots, \alpha \rightarrow \beta_n$$

часто используют сокращенную запись

$$\alpha \rightarrow \beta_1 | \dots | \beta_n.$$

Иногда при описании грамматики словари терминальных и нетерминальных символов не указывают. В таком случае обычно предполагается, что грамматика содержит только те терминальные и нетерминальные символы, которые встречаются в правилах вывода

Определение 3.114. Цепочка $\beta \in (V_T \cup V_N)^*$ *непосредственно выводима* из цепочки $\alpha \in (V_T \cup V_N)^+$ в грамматике $G = (V_T, V_N, P, S)$ (обозначают $\alpha \rightarrow \beta$), если

$$\alpha = \xi_1 \gamma \xi_2, \quad \beta = \xi_1 \delta \xi_2, \quad \xi_1, \xi_2, \delta \in (V_T \cup V_N)^*, \quad \gamma \in (V_T \cup V_N)^+$$

и правило вывода $\gamma \rightarrow \delta$ содержится в P .

Пример 3.146. Рассмотрим грамматику $G_1 = (\{a, b\}, \{A, S\}, P_1, S)$, где множество P_1 состоит из правил

$$S \rightarrow aAb, \quad aA \rightarrow aaAb, \quad A \rightarrow \epsilon.$$

Цепочка $aaAbb$ непосредственно выводима из цепочки aAb .

Определение 3.115. Цепочка $\beta \in (V_T \cup V_N)^*$ выводима из цепочки $\alpha \in (V_T \cup V_N)^+$ в грамматике $G = (V_T, V_N, P, S)$ (обозначают $\alpha \Rightarrow \beta$), если существуют последовательность цепочек $\gamma_0, \dots, \gamma_n$, такая, что

$$\alpha = \gamma_0 \rightarrow \dots \rightarrow \gamma_n = \beta.$$

Последовательность $\gamma_0, \dots, \gamma_n$ называют *выводом* длины n .

Пример 3.147. В грамматике G_1 из примера 3.146 $S \Rightarrow aaaAbbb$, т. к. существует вывод

$$S \rightarrow aAb \rightarrow aaAbb \rightarrow aaaAbbb.$$

При этом длина вывода равна 3.

Замечание 3.74. Бинарное отношение \Rightarrow является транзитивным замыканием отношения \rightarrow , определенного на множестве $(V_T \cup V_N)^*$.

Определение 3.116. Языком, порождаемым грамматикой G , называют множество $L(G) = \{\alpha \in V_T^* | S \Rightarrow \alpha\}$, т. е. $L(G)$ — это все цепочки в алфавите V_T , которые выводимы из начального символа S с помощью правил вывода P .

Упражнение 3.27. Покажите, что для грамматики G_1 из примера 3.146 $L(G_1) = \{a^n b^n | n > 0\}$.

Определение 3.117. Грамматики G_1 и G_2 называют эквивалентными, если $L(G_1) = L(G_2)$.

Пример 3.148. Рассмотрим грамматику $G_2 = (\{a, b\}, \{S\}, P_2, S)$, где множество P_2 состоит из правил $S \rightarrow aSb | ab$. Эта грамматика эквивалентна грамматике G_1 из примера 3.146, так как обе порождают один и тот же язык $L(G_1) = L(G_2) = \{a^n b^n | n > 0\}$.

3.9.2. Классификация грамматик

Накладывая различные ограничения на правила вывода можно выделить классы грамматик. Рассмотрим классификацию, предложенную Н. Хомским [53].

► **Тип G_0 (свободные или неограниченные грамматики).**

Грамматику G называют грамматикой типа 0, если на ее правила вывода не накладывается никаких ограничений.

► **Тип G_1 (контекстно-зависимые, контекстные или НС-грамматики).**

Грамматику $G = (V_T, V_N, P, S)$ называют грамматикой типа 1, если каждое правило из P имеет вид

$$\xi_1 A \xi_2 \rightarrow \xi_1 \gamma \xi_2, \quad A \in V_N, \gamma \in (V_T \cup V_N)^+, \xi_1, \xi_2 \in (V_T \cup V_N)^*.$$

Цепочку ξ_1 называют левым контекстом, цепочку ξ_2 называют правым контекстом.

Пример 3.149. Грамматика $G_1 = (\{a, b, c\}, \{S, B, C, W, Z\}, P, S)$, где множество P состоит из правил

$$\begin{aligned} S \rightarrow aSBC \mid aBC, \quad CB \rightarrow CZ, \quad CZ \rightarrow WZ, \quad WZ \rightarrow WC, \quad WC \rightarrow BC, \\ aB \rightarrow ab, \quad bB \rightarrow bb, \quad bC \rightarrow bc, \quad cC \rightarrow cc \end{aligned}$$

является контекстно-зависимой грамматикой. Грамматика G_1 порождает язык $L(G_1) = \{a^n b^n c^n \mid n \geq 1\}$.

► **Тип G_2 (контекстно-свободные или КС-грамматики).**

Грамматику $G = (V_T, V_N, P, S)$ называют грамматикой типа 2, если каждое правило из P имеет вид

$$A \rightarrow \beta, \quad A \in V_N, \beta \in (V_T \cup V_N)^+.$$

Пример 3.150. Грамматика $G_2 = (\{a, b, c\}, \{S, A\}, P, S)$, где множество правил P имеет вид

$$S \rightarrow aAb \mid accb, \quad A \rightarrow cSc$$

является контекстно-свободной грамматикой. Грамматика G_2 порождает язык $L(G_2) = \{(ac)^n(cb)^n \mid n > 0\}$.

Пример 3.151. Рассмотрим контекстно-свободную грамматику G_D , описывающую язык *Дика* (последовательность правильных расстановок скобок)

$$G_D = (\{a_1, b_1, \dots, a_n, b_n\}, \{S\}, P, S),$$

где множество правил P имеет вид

$$S \rightarrow \epsilon, \quad S \rightarrow a_1 S b_1, \quad \dots \quad S \rightarrow a_n S b_n$$

Замечание 3.75. Если интерпретировать символ a_i как открывающую скобку i -го типа, а b_i – как закрывающую скобку i -го типа, то язык Дика описывает правильную расстановку скобок n типов. А именно:

1. Для любой начальной последовательности число открывающих скобок i -го типа не меньше числа закрывающих скобок этого же типа
2. Для всей последовательности число вхождений открывающей скобки i -го типа равно числу вхождений закрывающей скобки i -го типа.

► **Тип G_3 (автоматные или регулярные грамматики).** Грамматики типа 3 делятся на две группы:

- a) леволинейные (леворекурсивные), правила вывода для которых имеют вид: $A \rightarrow a|Ba$, где $A, B \in V_N$, $a \in V_T$;
- б) праволинейные (праворекурсивные), правила вывода для которых имеют вид: $A \rightarrow a|aB$, где $A, B \in V_N$, $a \in V_T$.

Пример 3.152. Грамматика $G_3 = (\{a, b\}, \{S, A, B\}, P, S)$, где множество P содержит правила

$$S \rightarrow aA \mid bB, \quad A \rightarrow aA \mid a, \quad B \rightarrow b$$

является регулярной (праволинейной) грамматикой. Грамматика G_3 порождает язык $L(G_3) = \{a^n, b^2 \mid n > 1\}$.

Пример 3.153. Грамматика $G_4 = (\{a, b, c\}, \{S, A, B, C\}, P, S)$, где множество P содержит правила

$$(1) S \rightarrow aA, \quad (2) A \rightarrow aA \mid bC, \quad (3) C \rightarrow cC \mid c \quad (3.86)$$

является регулярной (праволинейной) грамматикой. Грамматика G_4 порождает язык $L(G_4) = \{a^nbc^m \mid n, m > 0\}$.

Рассмотрим вывод слова $aaabcc \in L(G_4)$ в грамматике G_4 . При применении правила сверху будем ставить его номер согласно (3.86).

$$S \xrightarrow{1} aA \xrightarrow{2} aaA \xrightarrow{2} aaaA \xrightarrow{2} aaabC \xrightarrow{3} aaabcC \xrightarrow{3} aaabcc$$

Любая автоматная грамматика может быть представлена ориентированным графом, который строится по следующим правилам:

- 1) вершины графа соответствуют нетерминальным элементам из V_N ; вводится одна дополнительная терминалная вершина, обозначаемая T ;
- 2) каждому правилу вывода $A \rightarrow aB$ (или $A \rightarrow Ba$ для леволинейной грамматики) соответствует дуга из вершины A в вершину B , которой сопоставляется терминальный символ a ;
- 3) каждому правилу вывода $A \rightarrow a$ соответствует дуга из вершины A в вершину T , которой сопоставляется терминальный символ a .

Таким образом из вершины T никакая дуга не выходит, а в вершину S никакая дуга не входит.

Регулярной грамматике из примера 3.152 соответствует граф на рис. 3.35.

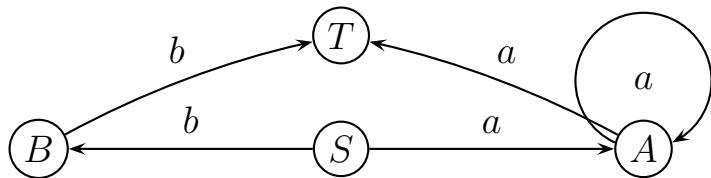


Рис. 3.35

Классы грамматик типа G_0, G_1, G_2 и G_3 образуют *иерархию Хомского*.

Определение 3.118. Язык называют *контекстным* (*контекстно-свободным*, *регулярным*), если он порождается некоторой контекстной (*контекстно-свободной*, *регулярной*) грамматикой. Контекстно-свободные языки называют также *алгебраическими* языками.

Замечание 3.76. Двигаясь от грамматики G_0 к грамматике G_3 можно заметить, что в то время как ограничения на правила вывода усиливаются, описательные возможности языков уменьшаются.

Справедливо следующее очевидное утверждение.

Теорема 3.43. $L(G_3) \subset L(G_2) \subset L(G_1) \subset L(G_0)$.

Упражнение 3.28. Постройте грамматику, порождающую язык булевых формул от трех переменных x, y, z .

3.9.3. Контекстно-свободные грамматики

Наиболее важный класс грамматик в иерархии Хомского представляют собой контекстно-свободные грамматики. Во-первых, на них основаны почти все используемые языки программирования, во-вторых для КС грамматик разработаны эффективные грамматические анализаторы.

Замечание 3.77. Часто для контекстно-свободных грамматик правила вывода $A \rightarrow \beta_1 \mid \dots \mid \beta_n$ записывают в виде $A ::= \beta_1 \mid \dots \mid \beta_n$. Такое описание правил называют записью в *форме Бэкуса–Наура*.

Историческая справка. В 1958 году Джон Бэкус решил принять активное участие в обсуждении нового языка (он впоследствии и получил название Algol) в Цюрихе. Однако возникла проблема — английский язык, на котором изъяснялся Бэкус, был мало понятен швейцарским программистам. В связи с этим для описания конструкций языка были применены специальные диаграммы, которые Бэкус разработал совместно с датским астрономом и программистом Питером Науром. С тех пор форма Бэкуса–Наура (Backus-Naur Form — BNF) стала неким «эсперанто» мирового программирования.

Пример 3.154. Рассмотрим бесскобочные алгебраические выражения, т. е. множество формул вида: $a+b/c*d-a*b$. Попробуем это неформально заданное множество формул превратить в язык с контекстно-свободной грамматикой.

Положим $V_T = \{a, \dots, z\}$. Далее определим множество нетерминалов

$$V_N = \{\langle \text{бесскобочное выражение} \rangle, \langle \text{буква} \rangle, \langle \text{знак операции} \rangle\}$$

(нетерминальному символу будем давать смысловое название, которое будет заключаться в угловые скобки). В качестве начального символа грамматики возьмем

$$S = \langle \text{бесскобочное выражение} \rangle .$$

Множество P состоит из трех правил:

- 1) $\langle \text{бесскобочное выражение} \rangle ::=$
 $\langle \text{бесскобочное выражение} \rangle \langle \text{знак операции} \rangle \langle \text{бесскобочное выражение} \rangle \mid \langle \text{буква} \rangle,$
 - 2) $\langle \text{буква} \rangle ::= a \mid \dots \mid z,$
 - 3) $\langle \text{знак операции} \rangle ::= + \mid - \mid * \mid /.$
- (3.87)

Грамматик, описывающих данный язык, может быть много. Например, первое правило в (3.87) можно заменить на:

$$\begin{aligned} \langle \text{бесскобочное выражение} \rangle ::= \\ \langle \text{бесскобочное выражение} \rangle \langle \text{знак операции} \rangle \langle \text{буква} \rangle \mid \langle \text{буква} \rangle \end{aligned}$$

или

$$\begin{aligned} \langle \text{бесскобочное выражение} \rangle ::= \\ \langle \text{буква} \rangle \langle \text{знак операции} \rangle \langle \text{бесскобочное выражение} \rangle \mid \langle \text{буква} \rangle . \end{aligned}$$

Можно расширить множество нетерминальных символов пустым символом ε и получить такую грамматику:

$$\begin{aligned} \langle \text{бесскобочное выражение} \rangle ::= & \langle \text{буква} \rangle \langle \text{окончание} \rangle, \\ \langle \text{окончание} \rangle ::= & \langle \text{знак операции} \rangle \langle \text{бесскобочное выражение} \rangle \mid \varepsilon. \end{aligned} \quad (3.88)$$

Задача 3.52. Постройте несколько вариантов грамматик, описывающих полную скобочную запись арифметических выражений. Под полными скобочными записями понимаются записи вида $((a+b)*c) - ((a+d)/e)$, в которых каждая бинарная операция вместе с ее аргументами заключается в скобки.

Понятие грамматики как средства описания языка неотделимо от понятия грамматического (синтаксического) разбора.

Определение 3.119. Последовательность цепочек нетерминальных и терминальных символов β_1, \dots, β_n называют *грамматическим разбором* для цепочки терминальных символов α , если $\beta_1 = S$, $\beta_n = \alpha$, а β_i отличается от β_{i+1} тем, что один из нетерминальных символов цепочки β_i заменен по одному из правил грамматики.

В дальнейшем грамматический разбор будем записывать в столбик.

Пример 3.155. Проведем грамматический разбор выражения $a + b * d$ языка бесскобочных арифметических выражений, заданного грамматикой (3.87) из предыдущего примера 3.154:

```

<бесскобочное выражение>
<бесскобочное выражение><знак операции><бесскобочное выражение>
<буква><знак операции><бесскобочное выражение>
a <знак операции><бесскобочное выражение>
a + <бесскобочное выражение>
a + <бесскобочное выражение><знак операции><бесскобочное выражение>
a + <буква><знак операции><бесскобочное выражение>
a + b <знак операции><бесскобочное выражение>
a + b * <бесскобочное выражение>
a + b * <буква>
a + b * d

```

Замечание 3.78. Нетрудно видеть, что, как правило, существует несколько грамматических разборов одной и той же формулы.

Представляет интерес построение программы — *синтаксического анализатора*, которая бы автоматически выясняла принадлежность цепочки терминальных символов (слова) данному языку. Однако, вследствие неоднозначности выбора расшифровки нетерминального символа, в общем случае грамматический разбор требует перебора вариантов.

Рассмотрим один важный частный случай, когда выбор цепочки при расшифровке нетерминального символа определяется однозначно. Назовем это свойство грамматики *однозначностью ветвления по первому символу*.

Формально это свойство можно описать следующим образом.

- Каждое правило грамматики $\alpha ::= \beta_1 | \dots | \beta_n$ должно удовлетворять условию: $L(\beta_i) \cap L(\beta_j) = \emptyset$, где $L(\beta_i)$ обозначает множество терминальных символов, с которых может начинаться цепочка β_i (после применения к ней правил вывода).

- Каждое правило грамматики

$$\alpha ::= \dots | \beta\gamma | \dots, \text{ где } \beta ::= \xi_1 | \dots | \xi_n | \varepsilon$$

должно удовлетворять условию: $L(\gamma) \cap L(\xi_i) = \emptyset$.

Замечание 3.79. Второй случай запрещает скрытую неоднозначность, когда за счет использования пустого символа ее проверка «сдвигается» на другое правило.

Пример 3.156. В примере 3.154 для грамматики (3.87) неоднозначность имеет место, так как

$$L(<\text{бесскобочное выражение}>) = \{a, \dots, z\} = L(<\text{буква}>)$$

и, очевидно, их пересечение не пусто.

Для грамматики (3.88) того же языка неоднозначности уже не будет (так как $\alpha = <\text{бесскобочное выражение}>$, $\beta = <\text{окончание}>$, γ отсутствует).

Для грамматик, удовлетворяющих свойству однозначности ветвления по первому символу, существует алгоритм, который, получая на вход правила грамматики, на выходе порождает алгоритм синтаксического анализа слов языка, задаваемого этой грамматикой. Дадим неформальное описание этого алгоритма, демонстрируя изложение на примере.

Пример 3.157. Рассмотрим технологию построения синтаксического анализатора для языка арифметических выражений в полной скобочной записи (для краткости будем называть их формулами).

Грамматика, определяющая этот язык, устроена очень просто:

$$\begin{aligned} <\text{формула}> ::= & <\text{переменная}> | (<\text{формула}> <\text{знак операции}> <\text{формула}>) \\ <\text{переменная}> ::= & a | \dots | z, \quad <\text{знак операции}> ::= + | - | * | /. \end{aligned} \quad (3.89)$$

Нетрудно видеть, что грамматика (3.89) удовлетворяет условию ветвления по первому символу.

Для построения синтаксического анализа сделаем преобразование грамматики из формы Бэкуса–Наура в форму графа. Такое представление грамматики называют *синтаксической диаграммой*.

Синтаксическая диаграмма строится по следующим правилам.

Правило 1. Каждый терминальный символ t изображается кружочком (вершиной графа), помеченным самим символом t (рис. 3.36,*a*).

Правило 2. Каждый нетерминальный символ N изображается прямоугольником (вершиной графа), помеченным тем же символом N (рис. 3.36,*b*).

Правило 3. Последовательность терминальных и нетерминальных символов изображается соединением соответствующих вершин дугами в порядке их следования. Например, изображение последовательности $N_1 N_2$ синтаксической диаграммой представлено на рис. 3.36,*c*.

Правило 4. Значку $|$ сопоставляется ветвление. Например, $N_1 | N_2$ на синтаксической диаграмме изображено на рис. 3.36,*d*.

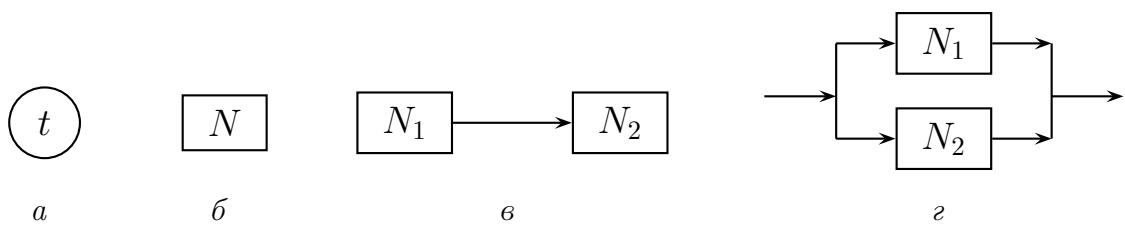


Рис. 3.36

Правило 5. Пустой цепочке символов ставится в соответствие дуга графа.

Синтаксическая диаграмма для языка полных арифметических скобочных записей, построенная на основе этих правил, показана на рис. 3.37.

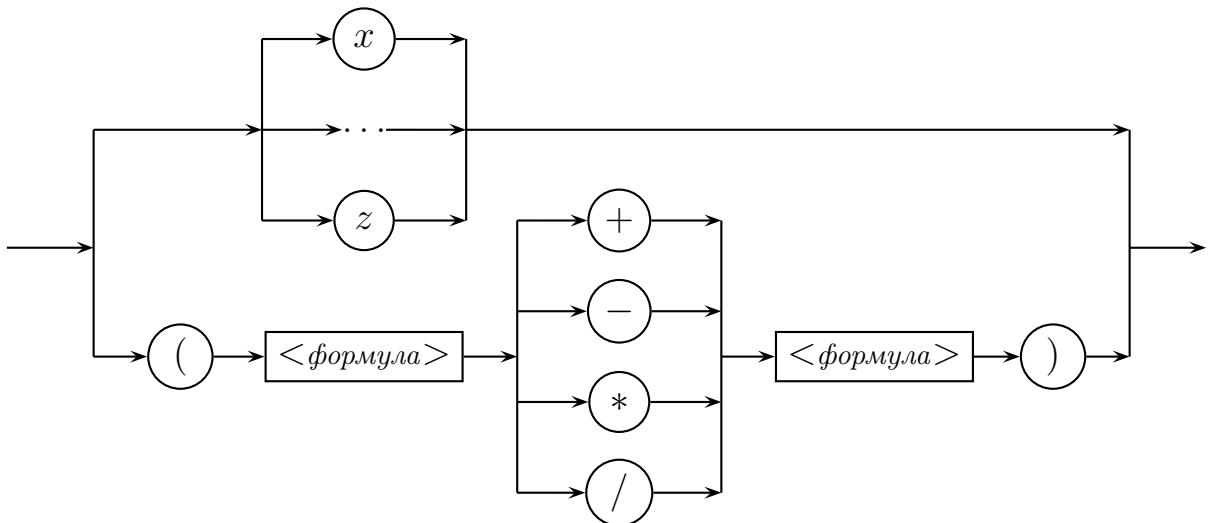


Рис. 3.37

Далее на основе следующих правил по этой диаграмме можно написать программу синтаксического анализа. Заметим, что для нашей грамматики условие однозначности ветвления выполнено, так как единственное ветвление осуществляется на основе проверки, является ли очередной символ выражения открывающей скобкой или латинской буквой.

Правило 1. Терминалному символу t синтаксической диаграммы в программе синтаксического анализа сопоставляется элементарный оператор распознавания:

ЕСЛИ $ch = t$ ТО $\text{read}(ch)$ ИНАЧЕ error

Здесь переменная ch обозначает текущий символ записи выражения, процедура $\text{read}(ch)$ считывает следующий символ записи, процедура error обрабатывает ошибку (в простейшем варианте просто прерывает дальнейший анализ выражения).

Правило 2. Нетерминальному символу N синтаксической диаграммы сопоставляется процедура $P(N)$ распознавания этого символа (при выполнении программы анализа вызывается соответствующая процедура).

Правило 3. Последовательности терминальных или нетерминальных символов соответствует линейная программа, в которой последовательно перечислены соответствующие символам операторы и процедуры распознавания. Последовательность заключается в операторные скобки. Например, для последовательности $N_1 N_2$ программа выглядит так:

```
НАЧАЛО
  P(N1)
  P(N2)
КОНЕЦ
```

Правило 4. Ветвлению соответствует оператор условного перехода. Так, диаграмме с ветвлением $N_1 | N_2$ на рис. 3.36,г соответствует такая программа распознавания:

```
ЕСЛИ ch ∈ L(N1) ТО P(N1) ИНАЧЕ
  ЕСЛИ ch ∈ L(N2) ТО P(N2) ИНАЧЕ error
```

Здесь $L(N_1)$ обозначает множество начальных символов, с которых могут начинаться правила вывода для нетерминального символа N_1 . (Заметим, что множества начальных символов N_1 и N_2 не имеют общих элементов в силу однозначности ветвления. Если это условие не выполняется, то строить синтаксический анализатор по предложенным правилам нельзя).

После применения к диаграмме языка полных скобочных записей арифметических выражений (см. рис. 3.37) перечисленных правил, получается следующая программа синтаксического анализа этого языка.

```
ПРОГРАММА FORMULA
НАЧАЛО
  ЕСЛИ ch ∈ {a, …, z} ТО
    ЕСЛИ ch ∈ {a, …, z} ТО read(ch) ИНАЧЕ error
  ИНАЧЕ
    ЕСЛИ ch = ( ТО
      НАЧАЛО
        ЕСЛИ ch = ( ТО read(ch) ИНАЧЕ error
        FORMULA
        ЕСЛИ ch ∈ {+, −, *, /} ТО read(ch) ИНАЧЕ error
        FORMULA
        ЕСЛИ ch = ) ТО read(ch) ИНАЧЕ error
```

```

КОНЕЦ
ИНАЧЕ error
КОНЕЦ

```

Замечание 3.80. Простота построения алгоритма синтаксического анализа основана на использовании рекурсии (в нашем примере программа обращается к себе дважды).

Алгоритм синтаксического анализа представлен в псевдокоде, который легко перевести на любой алгоритмический язык, допускающий рекурсивный вызов процедур.

Нетрудно видеть, что программа иногда осуществляет лишние проверки и поэтому не является идеальной. Однако такая запись в точности соответствует последовательному применению перечисленных правил.

3.9.4. Преобразования контекстно-свободных грамматик

Определение 3.120. Нетерминальный символ A называется *непродуктивным* (*непроизводящим*), если он не порождает ни одной терминальной цепочки, т. е. не существует вывода $A \Rightarrow \alpha | \alpha \in V_T^*$.

Представляет интерес задача удаления из КС-грамматики всех непродуктивных нетерминальных символов. Рассмотрим алгоритм построения КС-грамматики эквивалентной исходной, все нетерминальные символы которой продуктивны.

Алгоритм 3.18 (удаление непродуктивных терминалов).

```

 $W_0 \leftarrow \emptyset$ 
 $W_1 \leftarrow \{A | A \rightarrow \alpha \in P, \alpha \in V_T^*\}$ 
 $k \leftarrow 1$ 
while  $W_k \neq W_{k-1}$  do
     $W_{k+1} \leftarrow W_k \cup \{B | B \rightarrow \beta \in P, \beta \in (V_T \cup W_k)^*\}$ 
     $k \leftarrow k + 1$ 
end while

```

Пример 3.158. Пусть задана грамматика G со следующими правилами вывода:

$$S \rightarrow SA | BSb | bAb, A \rightarrow aSa | bb, B \rightarrow bBb | BaA.$$

Применим алгоритм 3.18 для построения множества продуктивных нетерминалов. Имеем:

1. $W_1 = \{A\}$, т. к. в множестве P есть правило $A \rightarrow bb$.
2. $W_2 = \{A, S\}$, т. к. имеется правило $S \rightarrow bAb$ и $A \in W_1$.
3. $W_3 = W_2$.

Все символы множества $V_N \setminus W_2$ являются непродуктивными, не используются в выводе никакой терминальной цепочки и их можно удалить

из грамматики вместе со всеми правилами, в которые они входят. Грамматика G_1 , эквивалентная исходной грамматике, будет включать следующее множество правил:

$$S \rightarrow SA \mid bAb, A \rightarrow aSa \mid bb.$$

В грамматике G_1 все нетерминалы продуктивны.

Существует еще один тип нетерминальных символов, которые можно удалять из грамматики вместе с правилами, в которые они входят. Например, в грамматике G , заданной множеством правил

$$S \rightarrow aSb \mid ba, A \rightarrow aAa \mid bba,$$

нетерминал A не участвует ни в каком выводе, т. к. из аксиомы нельзя вывести цепочку, содержащую этот нетерминал. Поэтому из заданной грамматики можно удалить нетерминал A .

Рассмотрим, как для произвольной КС-грамматики можно построить эквивалентную КС-грамматику, аксиома которой зависит от всех нетерминальных символов.

Алгоритм 3.19.

```

 $W_0 \leftarrow \emptyset$ 
 $W_1 \leftarrow \{S\}$ 
 $k \leftarrow 1$ 
while  $W_k \neq W_{k-1}$  do
     $W_{k+1} \leftarrow W_k \cup \{B \mid A \rightarrow \alpha B \beta \in P, A \in W_k\}$ 
     $k \leftarrow k + 1$ 
end while

```

Замечание 3.81. Для последнего множества W_n нетерминал $B \in W_n$ тогда и только тогда, когда аксиома S зависит от B . Все нетерминалы, не содержащиеся в W_n , можно удалить из грамматики вместе с правилами, в которые они входят.

Пример 3.159. Пусть дана КС-грамматика G :

$$S \rightarrow abS \mid ASa \mid ab, A \rightarrow abAa \mid ab, B \rightarrow bAab \mid bB.$$

Применим алгоритм 3.19 для построения грамматики, аксиома которой зависит от всех нетерминалов. Имеем:

1. $W_1 = \{S\}$.
2. $W_2 = \{S, A\}$, т. к. имеется правило $S \rightarrow ASa$ и $S \in W_1$.
3. $W_3 = W_2$.

Получаем эквивалентную грамматику G_1 , аксиома которой зависит от всех нетерминальных символов:

$$S \rightarrow abS \mid ASa \mid ab, A \rightarrow abAa \mid ab.$$

3.9.5. Примеры и задачи на построение грамматик

Пример 3.160. Построить грамматику G_1 , порождающую язык

$$L(G_1) = \{a^n b c^m \mid n, m > 0\}.$$

Структура любой цепочки языка имеет вид $\alpha b \beta$, где часть α состоит только из символов a , а часть β – только из символов c . Поэтому обе части можно порождать независимо друг от друга. Получаем следующую грамматику:

$$S \rightarrow AbC; \quad A \rightarrow aA \mid a; \quad C \rightarrow cC \mid c.$$

Построенная грамматика G_1 является контекстно-свободной грамматикой.

Пример 3.161. Построить грамматику G_2 , порождающую язык *правильно расставленных скобок круглых скобок* (частный случай грамматики Дика, пример 3.151). Имеем

$$S \rightarrow (S) \mid ()S \mid \epsilon$$

Построенная грамматика G_2 является контекстно-свободной грамматикой.

Пример 3.162. Построить грамматику G_3 , порождающую язык

$$L(G_3) = \{\alpha \mid \alpha \in \{a, b\}^*\},$$

где слово α начинается и заканчивается одним и тем же символом. Имеем:

$$S \rightarrow aAa \mid bAb; \quad A \rightarrow aA \mid bA \mid \epsilon.$$

Построенная грамматика G_3 является контекстно-свободной грамматикой.

Пример 3.163. Построить грамматику G_4 , порождающую язык

$$L(G_4) = \{a_1 a_2 \dots a_n a_n a_{n-1} \dots a_1 \mid a_i \in \{0, 1\}, 1 \leq i \leq n\}$$

Имеем:

$$S \rightarrow 0S0 \mid 1S1 \mid \epsilon$$

Построенная грамматика G_4 является контекстно-свободной грамматикой.

Пример 3.164. Построить грамматику G_5 , порождающую язык

$$L(G_5) = \{a^{2n+1} \mid n \geq 0\}.$$

Имеем:

$$S \rightarrow aB \mid a; \quad B \rightarrow aS.$$

Построенная грамматика G_5 является автоматной грамматикой.

Пример 3.165. Построить грамматику G_6 , порождающую язык

$$L(G_6) = \{\lambda \mid \lambda \in \{a, b\}^+\},$$

где цепочка λ не содержит двух рядом стоящих символов a . Имеем:

$$S \rightarrow a \mid aB \mid b \mid bB \mid bA; \quad A \rightarrow a \mid aB; \quad B \rightarrow b \mid bB \mid bA$$

Построенная грамматика G_6 является автоматной грамматикой.

3.9.6. Префиксная и постфиксная записи формул

Привычная запись арифметического выражения в виде инфиксной записи, когда знак операции ставится между operandами, обладает очевидным недостатком: невозможно записать любое арифметическое выражение без использования скобок. В то же время, есть другие записи формулы строкой символов, при которых последовательность действий в формуле определяется однозначно без использования дополнительных символов (скобок).

Определим *постфиксную запись* арифметического выражения.

$$\langle \text{постфикс} \rangle ::= \langle \text{постфикс} \rangle \langle \text{постфикс} \rangle \langle \text{знак операции} \rangle \mid \langle \text{буква} \rangle,$$

где $\langle \text{постфикс} \rangle$ означает постфиксную запись арифметического выражения, а $\langle \text{знак операции} \rangle$ и $\langle \text{буква} \rangle$ определяются как и раньше в грамматике (3.87)

$$\langle \text{буква} \rangle ::= a \mid \dots \mid z, \quad \langle \text{знак операции} \rangle ::= + \mid - \mid * \mid /.$$

Пример 3.166. Написать постфиксную запись арифметического выражения $(a + b) * c - d$.

Найдем в нашем выражении последнюю операцию (минус) и перепишем выражение в виде, соответствующем определению постфиксной записи:

$$\langle (a + b) * c \rangle \langle d \rangle -,$$

где выражение, взятое в угловые скобки ($\langle f \rangle$), следует читать как *постфиксная запись формулы* f . Далее аналогично раскроем выражения в угловых скобках

$$\langle a + b \rangle \langle c \rangle * \langle d \rangle -$$

и, наконец

$$\langle a \rangle \langle b \rangle + \langle c \rangle * \langle d \rangle -$$

или

$$ab + c * d -$$

поскольку постфиксная запись выражения из одной буквы — это сама эта буква.

Аналогично определяется *префиксная запись*.

$$<\text{префикс}> ::= <\text{знак операции}> <\text{префикс}> <\text{префикс}> | <\text{буква}> \quad (3.90)$$

Пример 3.167. Найти обычную запись арифметического выражения по его префиксной записи $/a + b * -c d f$.

Заметим, что грамматика (3.90) удовлетворяет условию однозначности ветвления по первому символу. Поэтому грамматический анализ выражения выполняется однозначно:

```

<префикс>
<знак операции><префикс><префикс>
| <префикс><префикс>
/a <префикс>
/a <знак операции><префикс><префикс>
/a+ <префикс><префикс>
/a + b <префикс>
/a + b <знак операции><префикс><префикс>
/a + b * <префикс><префикс>
/a + b * <знак операции><префикс><префикс><префикс>
/a + b * - <префикс><префикс><префикс>
/a + b * -c <префикс><префикс>
/a + b * -c d <префикс>
/a + b * -c d f

```

Если просмотреть полученный разбор повторно, переставив операнды (для удобства знаки операций и операнды можно пронумеровать) и взяв при необходимости выражения в скобки (это можно делать всегда, либо, если использовать приоритет операции, когда предыдущий знак имеет приоритет больше данного), то получим исковую формулу (см. табл. 3.56).

Таблица 3.56

$<\text{префикс}>_1$	$<\text{инфикс}>_1$
$<\text{знак операции}>_1 <\text{префикс}>_{2a}$	$<\text{инфикс}>_{2a} <\text{знак операции}>_1$
$<\text{префикс}>_{2b}$	$<\text{инфикс}>_{2b}$
$/ <\text{префикс}>_{2a} <\text{префикс}>_{2b}$	$<\text{инфикс}>_{2a} / <\text{инфикс}>_{2b}$
$/a <\text{префикс}>_{2b}$	$a / <\text{инфикс}>_{2b}$
$/a <\text{знак операции}>_2 <\text{префикс}>_{3a}$	$a / <\text{инфикс}>_{3a} <\text{знак операции}>_2$
$<\text{префикс}>_{3b}$	$<\text{инфикс}>_3 b$

Продолжение табл. 3.56

$/a + <\text{префикс}>_{3a} <\text{префикс}>_{3b}$	$a / (<\text{инфикс}>_{3a} + <\text{инфикс}>_{3b})$
$/a + b <\text{префикс}>_{3b}$	$a / (b + <\text{инфикс}>_{3b})$
$/a + b <\text{знак операции}>_3 <\text{префикс}>_{4a} <\text{префикс}>_{4b}$	$a / (b + <\text{инфикс}>_{4a} <\text{знак операции}>_3 <\text{инфикс}>_{4b})$
$/a + b * <\text{префикс}>_{4a} <\text{префикс}>_{4b}$	$a / (b + <\text{инфикс}>_{4a} * <\text{инфикс}>_{4b})$
$/a + b * <\text{знак операции}>_4 <\text{префикс}>_{5a} <\text{префикс}>_{5b} <\text{префикс}>_{4b}$	$a / (b + <\text{инфикс}>_{5a} <\text{знак операции}>_4 <\text{инфикс}>_{5b} * <\text{инфикс}>_{4b})$
$/a + b * - <\text{префикс}>_{5a} <\text{префикс}>_{5b} <\text{префикс}>_{5b}$	$a / (b + (<\text{инфикс}>_{4a} - <\text{инфикс}>_{5b}) * <\text{инфикс}>_{4b})$
$/a + b * - c <\text{префикс}>_{5b} <\text{префикс}>_{4b}$	$a / (b + (c - <\text{инфикс}>_{5b}) * <\text{инфикс}>_{4b})$
$/a + b * - c d <\text{префикс}>_{4b}$	$a / (b + (c - d) * <\text{инфикс}>_{4b})$
$/a + b * - c d f$	$a / (b + (c - d) * f)$

Докажем, что постфиксная и префиксная записи однозначно определяют арифметическое выражение.

Теорема 3.44. Каждому арифметическому выражению соответствует единственная постфиксная запись.

Доказательство. Сопоставим каждой постфиксной записи упорядоченный набор целых чисел следующим образом: двигаясь слева направо (и начиная с нуля) будем добавлять к предыдущему числу 1, если очередной символ буква, и -1 , если это знак операции.

Пример 3.168. Выражению $a b + c d * f - /$ соответствует такой набор

a	b	$+$	c	d	$*$	f	$-$	$/$
1	2	1	2	3	2	3	2	1

Лемма 3.21. Набор, соответствующий постфиксной записи арифметического выражения, обязательно заканчивается на 1.

Доказательство. Докажем индукцией по n — длине постфиксной записи. Если запись имеет длину 1, то она состоит из одной буквы и по правилу ей будет сопоставлена 1, что дает базу индукции. Предположим теперь, что утверждение доказано для всех постфиксных записей арифметических выражений, длина которых меньше n , и докажем его справедливость для n . По определению постфиксная запись арифметического выражения для $n > 1$ имеет вид:

$$<\text{постфикс}> ::= <\text{постфикс}> <\text{постфикс}> <\text{знак операции}>,$$

где длина постфиксных записей в правой части меньше n , и поэтому к ним применимо индукционное предположение. Таким образом первая постфиксная запись будет оканчиваться на 1. Тогда числа, сопоставляемые второй постфиксной записи, будут увеличены на 1, а значит, (по индукционному предположению) вторая запись будет оканчиваться на 2.

$<\text{постфикс}>$	$<\text{постфикс}>$	$<\text{знак операции}>$
1 1 2	1

Наконец, по правилу сопоставления знаку операции будет сопоставлено $2 - 1 = 1$, что и доказывает лемму. ■

Лемма 3.22. Все числа набора, соответствующего постфиксной записи арифметического выражения, больше либо равны 1 (т. е., натуральные).

Упражнение 3.29. Докажите лемму 3.22 самостоятельно по аналогии с леммой 3.21.

Рассмотрим последнюю выполняемую операцию арифметического выражения и два операнда относительно нее. Заметим, что по определению постфиксной записи все символы одного операнда идут подряд. Как выделить из записи первый и второй операнды? Очевидно, знак операции стоит последним. Где граница между первым и простым операндом? Оказывается, она определяется предпоследней единицей записи.

Лемма 3.23. Все числа, соответствующие символам второго операнда, больше либо равны 2.

Доказательство. По лемме 3.21 все числа набора соответствующего постфиксной записи арифметического выражения, больше либо равны 1. При нумерации символов второго операнда (который сам по себе является постфиксной записью арифметического выражения и при нумерации с нуля дает числа больше либо равные 1) происходит увеличение всех чисел на 1 (так как первый операнд заканчивается на 1), что и доказывает лемму. ■

Для окончания доказательства теоремы сформулируем алгоритм разбора, основанный на леммах 3.21–3.23.

Алгоритм 3.20.

Шаг 1. Сопоставляем символам постфиксной записи числа, как указано выше.

Шаг 2. Находим предпоследнюю 1 в полученном наборе, которая укажет конец первого операнда (последняя единица ограничивает справа второй операнд).

Шаг 3. Применяем *шаги 1 и 2* к полученным операндам.

Теорема 3.44 доказана. ■

Пример 3.169. Построить синтаксическое дерево арифметического выражения, заданного постфиксной записью $a b + c d * f - /$.

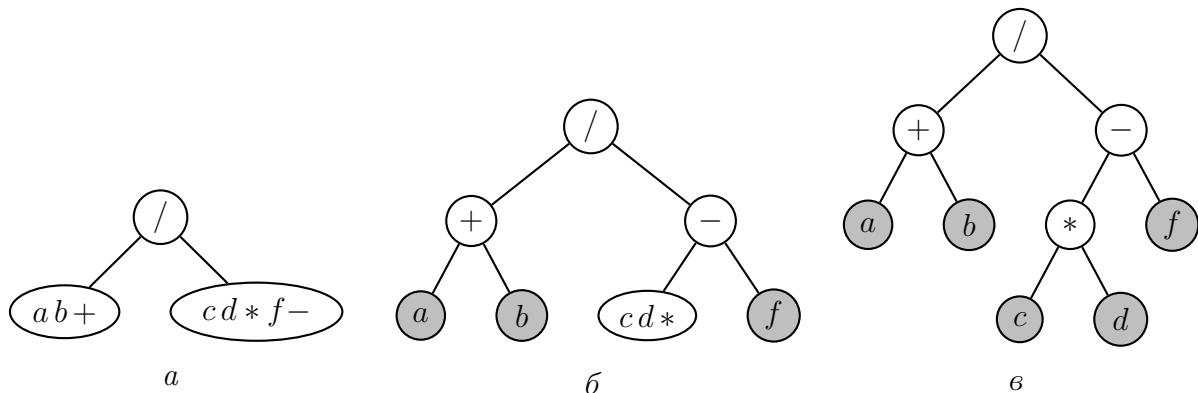


Рис. 3.38

Разделив операнды, поместим знак в корень дерева, а операнды рассмотрим как соответственно левое и правое поддеревья (рис. 3.38,*а*).

<i>a</i>	<i>b</i>	<i>+</i>	<i>c</i>	<i>d</i>	<i>*</i>	<i>f</i>	<i>-</i>	<i>/</i>
1	2	1	2	3	2	3	2	1

Далее применим ту же процедуру к поддеревьям (рис. 3.38,*б*).

<i>a</i>	<i>b</i>	<i>+</i>	<i>c</i>	<i>d</i>	<i>*</i>	<i>f</i>	<i>-</i>
1	2	1	1	2	1	2	1

Процедура закончится, когда все поддеревья будут состоять из отдельных символов (рис. 3.38,*в*).

<i>c</i>	<i>d</i>	<i>*</i>	<i>f</i>
1	2	1	1

Историческая справка. Способ записи арифметических формул, не использующий скобок предложил в 1920 г. польский математик Ян Лукасевич (1878 – 1956). В честь него постфиксную и префиксную формы записи называют прямой и обратной польской записью. На практике наиболее часто используется постфиксная форма. Поэтому под польской обычно понимают именно постфиксную форму записи.

3.9.7. Алгоритмы обработки скобочных записей

Рассмотрим сначала алгоритм вычисления значений в постфиксной форме. Отчасти он отвечает на вопрос, зачем нужна такая запись арифметического выражения. Алгоритм крайне прост. Он использует стековую

структурой организации памяти для хранения промежуточных вычислений (напомним, что стек определяется операциями *очистить стек*, *поместить в стек*, *извлечь из стека*, причем последовательность извлечений является обратной по сравнению с последовательностью добавлений).

Алгоритм 3.21 (вычисление значений в постфиксной форме).

В качестве входной строки рассматриваем выражение, записанное в постфиксной записи.

Шаг 1. Очищаем стек.

Шаг 2. Если очередной символ входной строки — константа, то помещаем ее в стек.

Шаг 3. Если очередной символ — знак операции, то извлекаем последовательно из стека две верхние константы, используем их в качестве второго и соответственно первого операндов для этой операции, затем помещаем результат обратно в стек.

Когда вся входная строка будет разобрана, в стеке должно остаться одно число, которое и будет результатом данного выражения.

Пример 3.170. Вычислим значение выражения $a b + c d * f - /$ при $a = 7$, $b = 8$, $c = 4$, $d = 3$, $f = 9$. Тогда исходная строка примет вид $78 + 43 * 9 - /$. Протокол работы алгоритма запишем в табл. 3.57.

Таблица 3.57

Текущий символ	Действие	Состояние стека
7	Помещаем константу 7 в стек.	7
8	Помещаем константу 8 в стек.	8 7
+	Извлекаем из стека две верхних константы (8 и 7), совершаем операцию $8 + 7 = 15$, результат помещаем в стек	15
4	Помещаем константу 4 в стек.	4 15
3	Помещаем константу 3 в стек.	3 4 15
*	Извлекаем из стека две верхних константы (3 и 4), совершаем операцию $4 * 3 = 12$, результат помещаем в стек	12 15
9	Помещаем константу 9 в стек.	9 12 15
-	Извлекаем из стека две верхних константы (9 и 12), совершаем операцию $12 - 9 = 3$, результат помещаем в стек	3 15
/	Извлекаем из стека две верхних константы (3 и 15), совершаем операцию $15 / 3 = 5$, результат помещаем в стек	5

Теперь строка разобрана и в стеке находится одна константа 5, которая является результатом исходного выражения.

Теорема 3.45. Приведенный алгоритм 3.21 вычисляет значение постфиксной записи арифметического выражения (если оно определено).

Доказательство. Докажем индукцией по n — длине постфиксной записи. Если $n = 1$, то правильность алгоритма очевидна (значение переменной заносится в стек, и оно является значением выражения).

Пусть утверждение верно для всех постфиксных записей арифметических выражений с числом символов меньше n . Докажем, что тогда оно будет верно и для записи с n символами.

Воспользуемся определением и рассмотрим постфиксную запись как последовательность трех частей. Протокол вычислений запишем в табл. 3.58.

Таблица 3.58

$\langle \text{постфикс} \rangle$	$\langle \text{постфикс} \rangle$	$\langle \text{знак операции} \rangle$
Значение первого операнда	Значение второго операнда	Результат выполнения данной операции над первым и вторым операндом
	Значение первого операнда	

Здесь использован тот факт, что правильность вычисления, как первого, так и второго операнда обеспечивается индукционным предположением. При вычислении второго операнда значение первого останется *на дне* стека, поэтому перед вычислением операции в стеке будут два числа: значение первого операнда — внизу, и второго операнда — наверху. В соответствии с алгоритмом вычислений над ними будет проведена операция, которая, очевидно, и даст значение выражения. ■

Сформулируем теперь алгоритм перевода скобочной инфиксной записи арифметического выражения в постфиксную форму. Этот алгоритм также будет использовать стек, но только для символьных переменных. Кроме того, для каждой операции будет определен приоритет (табл. 3.59). (Поскольку запись не является полной скобочной, это необходимо для определения точного порядка действий.)

Таблица 3.59

Операция	Приоритет	Операция	Приоритет
(0	-	1
)	0	*	2
+	1	/	2

Алгоритм 3.22 (перевод инфиксной формы в постфиксную). В качестве входной строки рассматриваем выражение в инфиксной записи.

Шаг 1. Очищаем стек.

Шаг 2. Если текущий символ — константа или переменная, то помещаем его в выходную строку.

Шаг 3. Если текущий символ — знак операции или скобка, то определяем приоритет операции согласно табл. 3.59. Далее проверяем стек:

а) если стек пуст, или находящиеся в нем символы (находиться в нем могут только знаки операций и открывающая скобка) имеют меньший приоритет, чем приоритет текущего символа, то помещаем текущий символ в стек;

б) если символ, находящийся на вершине стека имеет приоритет, больший или равный приоритету текущего символа, то извлекаем символы из стека в выходную строку до тех пор, пока выполняется это условие; затем переходим к п. а).

Шаг 4. Если текущий символ — открывающая скобка, то помещаем ее в стек.

Шаг 5. Если текущий символ — закрывающая скобка, то извлекаем символы из стека в выходную строку до тех пор, пока не встретим в стеке открывающую скобку, которую уничтожаем. Закрывающая скобка также уничтожается.

Шаг 6. Если вся входная строка разобрана, а в стеке еще остаются знаки операций, извлекаем их из стека в выходную строку.

Пример 3.171. Переведем арифметическое выражение $(a + b) * c - d$ в постфиксную форму по алгоритму 3.59. Протокол его работы запишем в табл. 3.60.

Таблица 3.60

Текущий символ	Действие	Выходная строка	Состояние стека
(Помещаем (в стек		(
a	Помещаем a в выходную строку	a	(
+	Проверяем стек: на вершине находится символ (, приоритет которого меньше, чем приоритет текущего символа +. Помещаем в стек +	a	+ (
b	Помещаем b в выходную строку	a b	+ (
)	Извлекаем из стека + и (. Помещаем + в выходную строку. Скобки уничтожаем.	a b +	
*	Помещаем * в стек	a b +	*
c	Помещаем c в выходную строку	a b + c	*

Продолжение табл. 3.60

Текущий символ	Действие	Выходная строка	Состояние стека
–	Проверяем стек: на вершине находится символ *, приоритет которого больше, чем приоритет текущего символа –. Извлекаем из стека в выходную строку *. Помещаем в стек –	$a b + c *$	–
d	Помещаем d в выходную строку	$a b + c * d$	–
	Входная строка пуста. Извлекаем из стека – и помещаем выходную строку	$a b + c * d -$	

3.10. Конечные автоматы распознаватели

3.10.1. Введение

Теория автоматов относится к числу ключевых разделов современной математики. Она непосредственно связана с математической логикой, теорией алгоритмов, теорией формальных грамматик. Универсальность и сравнительная простота автоматов обусловили широкое использование автоматных моделей на практике. Теория автоматов, например, успешно используется при построении узлов цифровых вычислительных машин, применяется при разработке парсеров для формальных языков (в том числе языков программирования), а также при построении компиляторов и разработке самих языков программирования..

Конечный автомат – это один из самых простых механизмов, используемых при работе с языками. Основная часть конечного автомата – это функция перехода, определяющая возможные переходы для любого текущего состояния и любого входного символа. Подразумевается, что допускается возможность перехода сразу в несколько различных состояний автомата, т.е. управляющее устройство распознавателя может быть и недетерминированным.

Историческая справка. Основы теории автоматов были заложены выдающимся математиком венгерского происхождения¹ Джоном фон Нейманом (1903 - 1957). В 1952г. в работе „Вероятностная логика и синтез надежных организмов из ненадеж-

¹Джоном фон Нейман родился в Будапеште в семье состоятельного венгерского банкира. Родители назвали его Яношем, в годы учебы и работы в Швейцарии и Германии он именовал себя Иоганном, а после переезда в США (в 1930г.) - Джоном.

ных элементов² он показал путь создания надежных ЭВМ и других автоматов. Это даже не книга, а 5 лекций, написанных Нейманом для своего друга Ричарда Пирса.

Джоном фон Нейман — один из создателей атомной бомбы (математически доказал осуществимость взрывного способа детонации атомной бомбы), сформулировал основную концепцию логической организации ЭВМ (принцип совместного хранения команд и данных в памяти компьютера, называемый *архитектурой фон Неймана, моделью фон Неймана или принстонской архитектурой*), что послужило огромным толчком к развитию электронно-вычислительной техники.

Под автоматом в общем смысле обычно понимают дискретный преобразователь информации, который принимает входные сигналы в дискретные моменты времени, изменяет свое состояние (с учетом прежнего состояния) и возможно формирует выходные сигналы.

3.10.2. Автоматы Мили и Мура

Дадим теперь формальные определения. Начнем с детерминированных автоматов.

Определение 3.121. Конечный детерминированный автомат с выходом (*автомат Мили*¹) — это система $A = (Q, \Sigma, \Omega, \delta, \lambda, q_0)$ где

- Q — конечное *непустое* множество состояний;
- $q_0 \in Q$ — начальное состояние;
- Σ — конечное *непустое* множество входных символов (входной алфавит);
- Ω — конечное *непустое* множество выходных символов (выходной алфавит);
- $\delta : Q \times \Sigma \rightarrow Q$ — *всюду определенное* отображение множества $Q \times \Sigma$ в множество Q , определяющее поведение автомата (эту функцию часто называют *функцией переходов*);
- $\lambda : Q \times \Sigma \rightarrow \Omega$ — *всюду определенное* отображение множества $Q \times \Sigma$ в множество Ω , определяющее выходную последовательность автомата (эту функцию часто называют *функцией выходов*).

Автомат начинает работу в состоянии q_0 , считывая по одному символу входной строки. Считанный символ переводит автомат в новое состояние из Q в соответствии с функцией переходов δ и возвращает соответствующий выходной символ согласно функции выходов λ .

Существуют два основных способа описания конечного автомата.

²На русском языке работа была опубликована в 1960 году в сборнике „Автоматы“

¹Иногда его называют автоматом I рода.

1. *Диаграмма состояний* (или иногда граф переходов) – графическое представление множества состояний и функции переходов. Представляет собой нагруженный ориентированный граф, вершины которого – состояния автомата, дуги – переходы из одного состояния в другое, а нагрузка – входные/выходные символы, при которых осуществляется данный переход. Если переход из состояния q_i в q_j может быть осуществлен при появлении одного из нескольких символов, то над дугой должны быть написаны все они.

2. *Таблица переходов* (табличное представление) функции δ . Обычно в такой таблице каждой строке соответствует одно состояние, а столбцу – один допустимый входной символ. В ячейке на пересечении строки и столбца записывается состояние, в которое должен перейти автомат, если в ситуации, когда он находился в данном состоянии на входе он получил данный символ и соответствующий выходной символ.

Пример 3.172. Рассмотрим автомат Мили по продаже шоколадок стоимостью 20 рублей, принимающий монеты номиналом в 5 и 10 рублей и возвращающий сдачу, если это необходимо².

Состояний автомата четыре: q_0, q_5, q_{10}, q_{15} – 0, 5, 10 и 15 рублей.

Входных сигналов два: σ_5 – 5 рублей и σ_{10} – 10 рублей.

Выходных сигналов три: ω_n – ничего не выдавать, ω_0 – выдать шоколадку и 0 рублей сдачи, ω_5 – выдать шоколадку и 5 рублей сдачи.

На рис. 3.39 *a* приведена диаграмма состояний, а на рис. 3.39 *b* – таблица переходов рассматриваемого автомата.

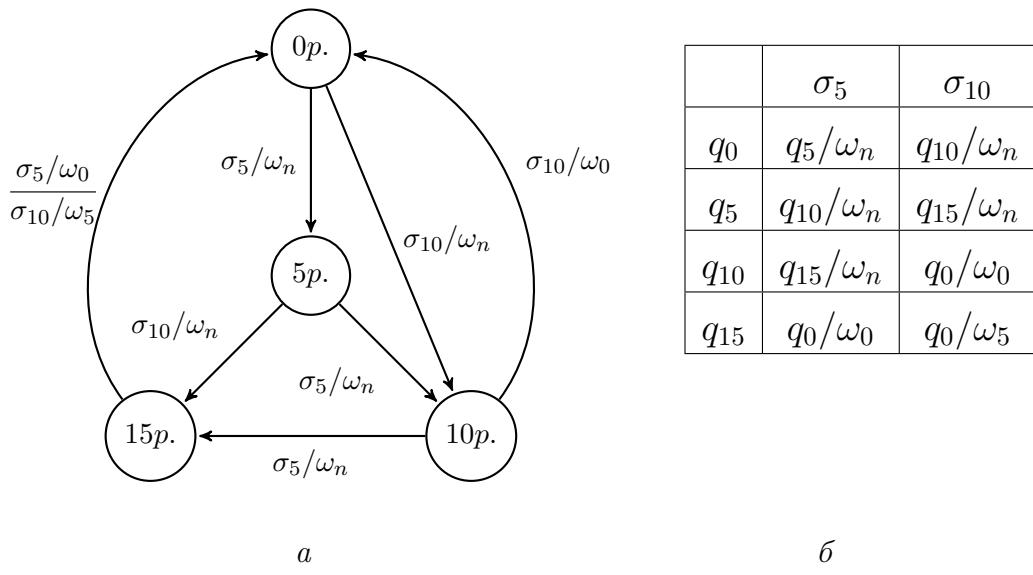


Рис. 3.39

²Пример взят из Викиконспекты: <https://neerc.ifmo.ru/wiki/....>

Определение 3.122. Пусть

$$A = (Q_A, \Sigma, \Omega, \delta_A, \lambda_A, q_0), \quad B = (Q_B, \Sigma, \Omega, \delta_B, \lambda_B, p_0)$$

два автомата с одинаковыми входными и выходными алфавитами. Состояние q автомата A и состояние p автомата B называются *неотличимыми*, если для любого слова $\alpha \in \Sigma^*$ выходные слова при достижении q и p совпадают.

Автоматы A и B называются *эквивалентными*, если для любого состояния автомата A найдется неотличимое от него состояние автомата B и наоборот.

Иногда удобным бывает более простое определение автомата, в котором функция переходов задает следующее состояние, а выход автомата зависит лишь от его текущего состояния.

Определение 3.123. Конечный автомат называется *автоматом Мура* или *автоматом II рода*, если для любого $q \in Q$ значения функции выходов на дугах, входящих в q , совпадают. То есть функция выходов не зависит от входного символа, а только от состояния: $\lambda : Q \rightarrow \Omega$.

В графическом представлении автоматов Мура часто вершину состояния $q_i \in Q$ помечают также соответствующим выходным символом $w_j \in \Omega$.

Пример 3.173. Пусть на вход автомата подается последовательность из нулей и единиц. Построим автомат Мура, который возвращает накопленное число единиц по модулю 3.

Как и в предыдущем примере 3.176 начальное состояние обозначим q_0 . Диаграмма состояний рассматриваемого автомата приведена на рис. 3.40. Выходной сигнал однозначно определяются состоянием, в которое автомат перешел, поэтому выходные сигналы приписаны прямо к состояниям.

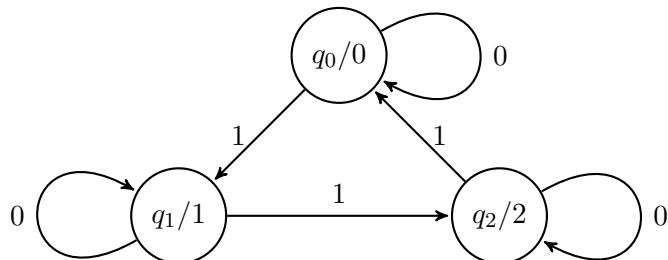


Рис. 3.40

3.10.3. Примеры автоматов Мили и Мура

Пример 3.174. Реализуем с помощью автомата Мили двоичный последовательный сумматор для одного разряда входного кода.

Автомат имеет два состояния: q_0 — нет переноса единицы, начальное состояние, q_1 — есть перенос. Входной алфавит автомата это пары битов обоих операндов, последовательно поступающих на сумматор: $(0, 0)$, $(0, 1)$, $(1, 0)$ и $(1, 1)$. Выходной алфавит, результат работы сумматора: $\{0, 1\}$.

Диаграмма состояний рассматриваемого автомата приведена на рис. 3.41.

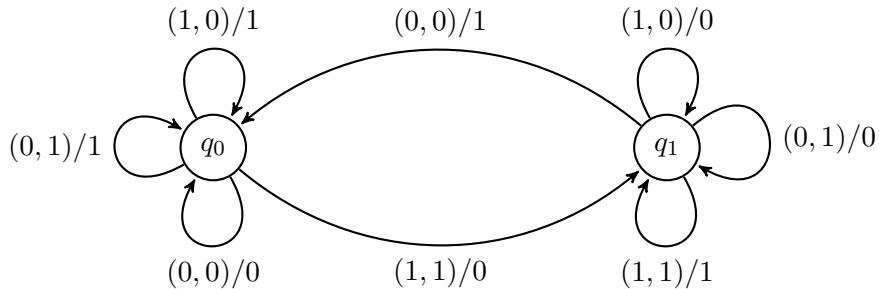


Рис. 3.41

Пример 3.175. Построим автомат Мили, управляющий лифтом в трехэтажном доме.

Входной алфавит автомата состоит из нажатия кнопки вызова соответствующего этажа: $\{C_1, C_2, C_3\}$. Выходной алфавит состоит из перемещений на один или два этажа вверх или вниз, а также остановки лифта: $\{U_1, U_2, D_1, D_2, S\}$; состояние соответствует этажу, на котором находится автомат: $\{q_1, q_2, q_3\}$; для определенности выберем начальное состояние q_1 .

Диаграмма состояний рассматриваемого автомата приведена на рис. 3.42.

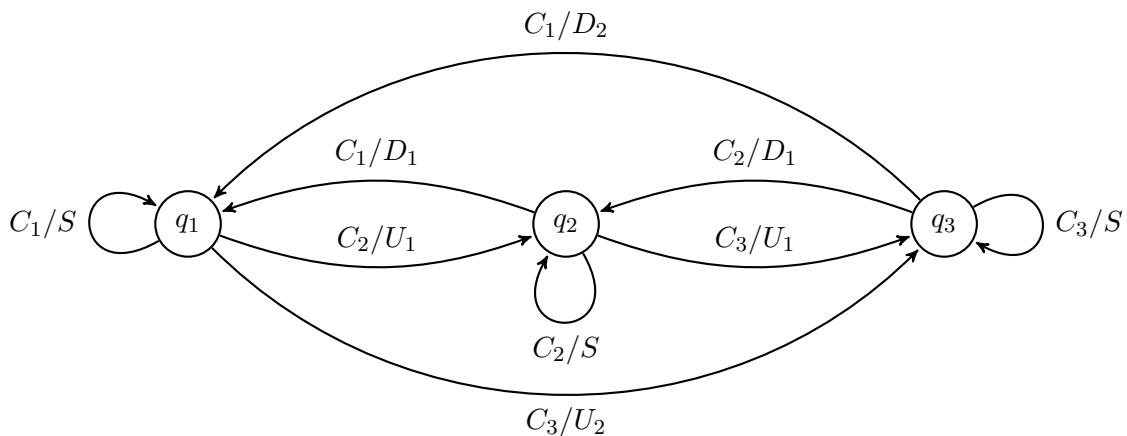


Рис. 3.42

Пример 3.176. Построим автомат Мили, который удаляет повторяющиеся подряд символы в словах в алфавите $\{a, b, c\}$ (оставляя только один). Например, слово $abbbaccaa$ должно преобразоваться в слово $abca$.

В выходной алфавит добавим пустой символ ϵ . Пустой символ будем возвращать при удалении символов входного слова. Начальное состояние обозначим q_0 . Диаграмма состояний рассматриваемого автомата приведена на рис. 3.43.

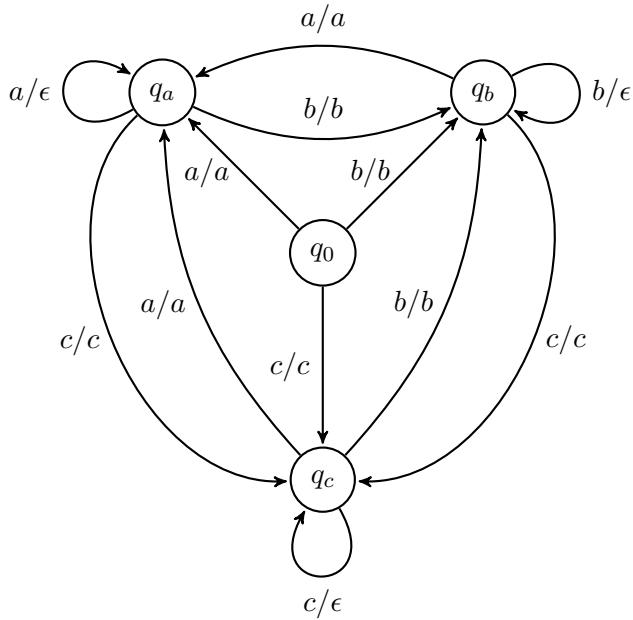


Рис. 3.43

3.10.4. Эквивалентность автоматов Мили и Мура

Согласно определениям 3.121 и 3.123 автомат Мура есть частный случай автомата Мили. Однако оба автомата являются эквивалентными.

Теорема 3.46 (об эквивалентности автоматов Мили и Мура). Для любого автомата Мили существует эквивалентный ему автомат Мура.

Доказательство. Преобразование автомата Мура в эквивалентный автомат Мили легко провести в их графическом представлении. Для этого, для каждой вершины-состояния, символ выходного алфавита Ω , которым помечена данная вершина, переносим на все дуги, входящие в нее. Очевидно, что получившийся автомат Мили эквивалентен исходному.

На рис. 3.44 показано данное преобразование для автомата Мура из примера 3.173.

Рассмотрим алгоритм обратного преобразования автомата Мили в автомат Мура. Для этого выполним действия, обратные действиям при преобразовании автомата Мура в автомат Мили, т. е. выходной символ, которым помечена дуга графа автомата, перенесем в вершину, в которую эта дуга входит. Однако такое преобразование невозможно для случая несколько

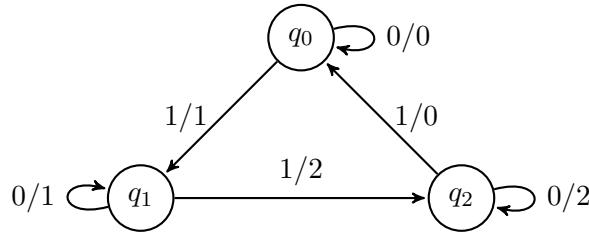


Рис. 3.44

входящих в вершину дуг с разными выходными символами. Тогда это состояние необходимо «расщепить» на такое количество состояний, сколько различных входных символов имеется на всех входных ребрах этого состояния, и сопоставить каждому из них свой выходной символ.

Схематически идея «расщепления» состояний показана на рис. 3.45.

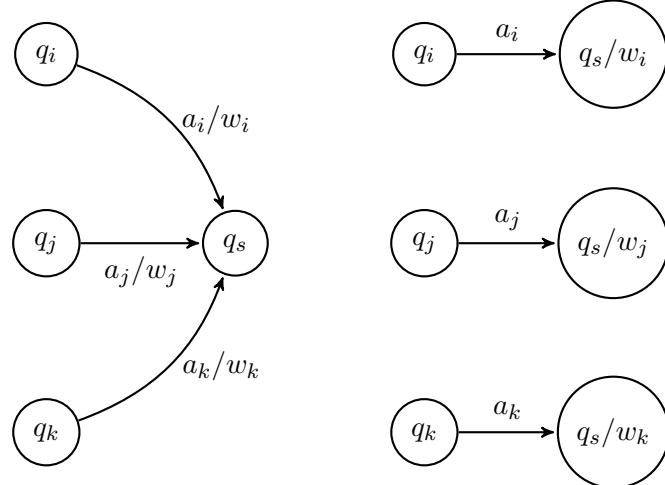


Рис. 3.45

Рассмотрим эту процедуру на примере автомата Мили на рис. 3.46.

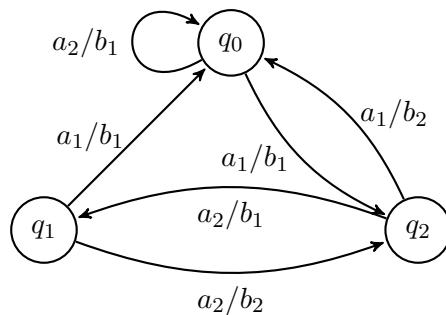


Рис. 3.46

Каждому состоянию автомата Мили $q_i \in Q$ поставим в соответствие множество пар A_i :

$$A_i = \{(q_i, \omega_j) \mid q_i = \delta(q_m, \alpha_s), \omega_j = \lambda(q_m, \alpha_s), q_m \in Q, \alpha_s \in \Sigma, \omega_j \in \Omega\}. \quad (3.91)$$

Согласно (3.91) для автомата на рис. 3.46 имеем:

$$A_0 = \begin{cases} (q_0, b_1) = p_0, \\ (q_0, b_2) = p_1, \end{cases} \quad A_1 = \{(q_1, b_1) = p_2\}, \quad A_2 = \begin{cases} (q_2, b_1) = p_3, \\ (q_2, b_2) = p_4 \end{cases} \quad (3.92)$$

Таким образом, получаем состояния автомата Мура: $\{p_0, p_1, p_2, p_3, p_4\}$. Обозначим функции переходов и выходов для автомата Мура как $\hat{\delta}$ и $\hat{\lambda}$. Функция выхода определяется вторым элементом соответствующей пары в (3.92).

$$\hat{\lambda}(p_0) = b_1, \hat{\lambda}(p_1) = b_2, \hat{\lambda}(p_2) = b_1, \hat{\lambda}(p_3) = b_1, \hat{\lambda}(p_4) = b_2.$$

Определим функцию переходов для автомата Мура. Рассмотрим произвольное состояние $q_i \in Q$ и произвольный символ входного алфавита $\alpha \in \Sigma$. Пусть $\delta(q_i, \alpha) = q_s$ и $\lambda(q_i, \alpha) = \omega_m \in \Omega$. Тогда в автомате Мура должны быть переходы из всех состояний A_i в состояние $(q_s, \omega_m) \in A_s$, определенное согласно (3.91), под воздействием символа α . Например, в автомате Мили $\delta(q_0, a_1) = q_2, \lambda(q_0, a_1) = b_1$. Тогда для автомата Мура для состояний $p_0, p_1 \in A_0$ имеем $\hat{\delta}(p_0, a_1) = (q_2, b_1) = p_3$ и $\hat{\delta}(p_1, a_1) = p_3$.

В общем виде можно записать таблицу переходов (таблица 3.61).

Таблица 3.61

$\delta(q_0, a_1) = q_2, \lambda(q_0, a_1) = b_1$	$(q_2, b_1) = p_3$	$\hat{\delta}(p_0, a_1) = \hat{\delta}(p_1, a_1) = p_3$
$\delta(q_0, a_2) = q_0, \lambda(q_0, a_2) = b_1$	$(q_0, b_1) = p_0$	$\hat{\delta}(p_0, a_2) = \hat{\delta}(p_1, a_2) = p_0$
$\delta(q_1, a_1) = q_0, \lambda(q_1, a_1) = b_1$	$(q_0, b_1) = p_0$	$\hat{\delta}(p_2, a_1) = p_0$
$\delta(q_1, a_2) = q_2, \lambda(q_1, a_2) = b_2$	$(q_2, b_2) = p_4$	$\hat{\delta}(p_2, a_2) = p_4$
$\delta(q_2, a_1) = q_0, \lambda(q_2, a_1) = b_2$	$(q_0, b_2) = p_1$	$\hat{\delta}(p_3, a_1) = \hat{\delta}(p_4, a_1) = p_1$
$\delta(q_2, a_2) = q_1, \lambda(q_2, a_2) = b_1$	$(q_1, b_1) = p_2$	$\hat{\delta}(p_3, a_2) = \hat{\delta}(p_4, a_2) = p_2$

В итоге получаем эквивалентный автомат Мура (рис. 3.47).

В качестве начального состояния автомата Мура можно взять любое из состояний p_0 или p_1 , так как оба порождены состоянием q_0 автомата Мили.

Таким образом, можно рассматривать только автоматы Мура. Для задачи определения принадлежности слова языку достаточно ограничиться случаем $|\Omega| = 2$. Тогда каждому состоянию может соответствовать один из двух выходных символов, будем считать, что тем самым состояния разбиваются на два класса, назовем их *заключительными* и *незаключительными*. Приходим к понятию детерминированного конечного автомата распознавателя (раздел 3.10.5). ■

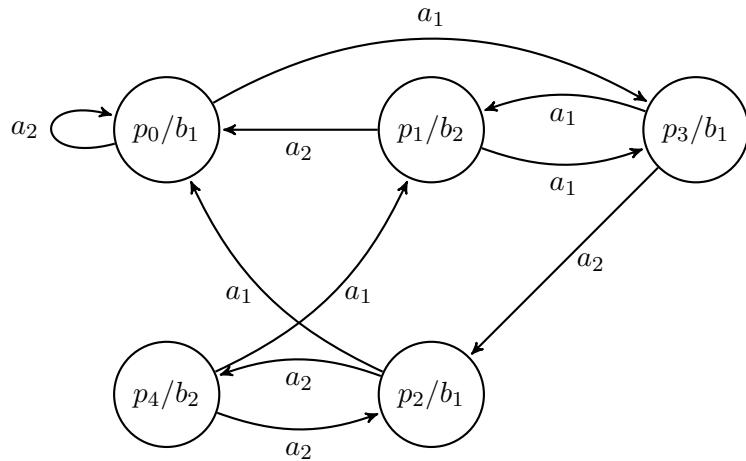


Рис. 3.47

3.10.5. Основные понятия автоматов распознавателей

Определение 3.124. Детерминированный конечный автомат распознаватель A – это пятерка объектов $A = (Q, \Sigma, \delta, q_0, F)$, где

- Q – конечное *непустое* множество состояний;
- $q_0 \in Q$ – начальное состояние;
- Σ – конечное *непустое* множество входных символов (входной алфавит);
- $\delta : Q \times \Sigma \rightarrow Q$ – *всюду определенное* отображение множества $Q \times \Sigma$ в множество Q , определяющее поведение автомата. Как и ранее (см. определение 3.121) эту функцию будем называть *функцией переходов*;
- $F \subseteq Q$ – множество заключительных (финальных) состояний.

Автомат начинает работу в состоянии q_0 , считывая по одному символу входной строки. Считанный символ переводит автомат в новое состояние из Q в соответствии с функцией переходов δ .

Замечание 3.82. Автомат, описанный в определении 3.124, называется *детерминированным*, так как для любой пары $q \in Q, a \in \Sigma$ существует единственное состояние $p \in Q : p = \delta(q, a)$.

В принципе, для определения последующих действий конечного автомата достаточно знать его текущее состояние и последовательность еще необработанных символов на входной ленте. Этот набор данных называется *конфигурацией автомата*.

Определение 3.125. Слово $w = a_1 \dots a_k$ над алфавитом Σ *допускается* конечным автоматом $M = (Q, \Sigma, \delta, q_0, F)$, если существует последовательность состояний q_1, q_2, \dots, q_n такая, что

$$q_1 = q_0, q_n \in F, \delta(q_i, a_j) = q_{i+1}, 1 \leq i < n, 1 \leq j < k.$$

Определение 3.126. Язык L распознается конечным автоматом A , если каждое слово языка L допускается этим конечным автоматом. При этом язык называется *автоматным* или *регулярным* и обозначается L_A .

Определение 3.127. Два состояния автомата называются *эквивалентными*, если для любой строки $\alpha \in \Sigma^*$ результаты ее обработки (допуск или недопуск автоматом), начиная с указанных состояний, совпадают.

Определение 3.128. Автоматы A и B называются эквивалентными, если совпадают распознаваемые ими языки, т.е. $L_A = L_B$.

3.10.6. Примеры автоматов распознавателей

Рассмотрим диаграммы состояний простейших автоматов, (конечные состояния — заштрихованы).

Пример 3.177. На рис. 3.48 *a* приведена диаграмма состояний, а на рис. 3.48 *б* — таблица переходов детерминированного конечного автомата допускающего цепочки из 0 и 1, заканчивающиеся символом 0.

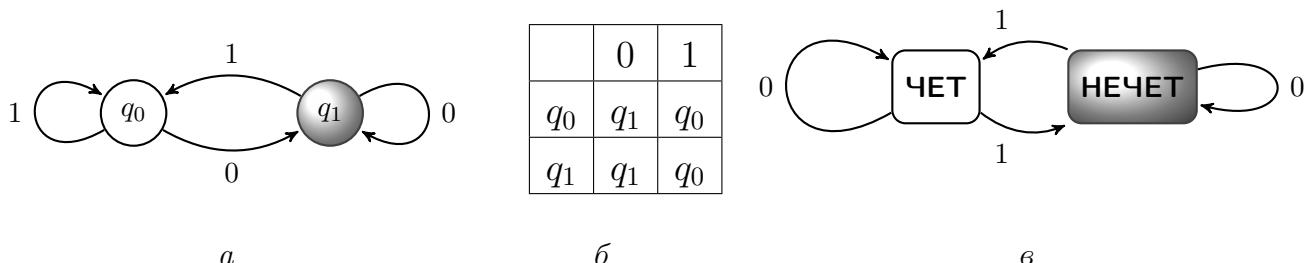


Рис. 3.48

Пример 3.178. Рассмотрим диаграмму состояний автомата, контролирующего нечетность единиц, то есть распознающий цепочки, состоящих из 0 и 1 и имеющие нечетное число единиц (рис. 3.48 *в*). Начальное состояние: $g_0 = \text{ЧЕТ}$.

Пример 3.179. Рассмотрим автомат распознающий слова, содержащие только парные вхождения букв a и b , например, $aa, aabbaaaa, bbaa$ и т.д. Диаграмма состояний автомата приведена на рис. 3.49.

Замечание 3.83. Заметим, что автомат из примера 3.179 попав в незаключительное состояние q_3 под воздействием сигналов a, b не может выйти из него под воздействием тех же входных сигналов (состояние и переходы отмечены на диаграмме 3.49 пунктиром). Будем считать, что переходы автомата в это состояние под воздействием сигналов a, b запрещены. То есть, запрещены: сигнал b в состоянии q_1 и сигнал a в состоянии q_2 . Далее запрещенные состояния и переходы в диаграмме состояний не отображаем, считая, что если символ входного слова привел к запрещенному переходу данной слово не принимается автоматом. Такие состояния (в которое переходят при обработке любого недопустимого символа) в теории часто называют *невозратными*.

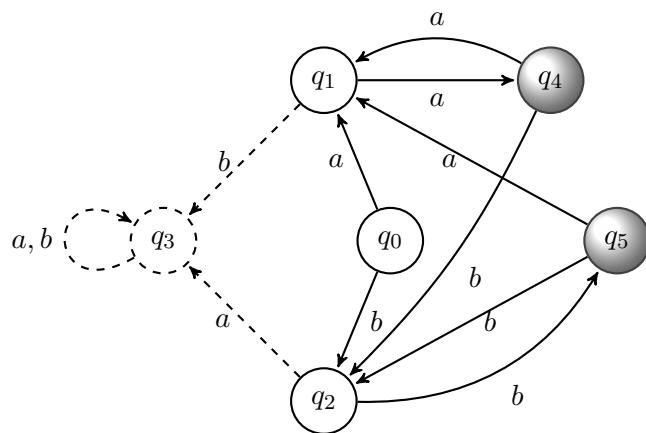


Рис. 3.49

Пример 3.180. С учетом замечания 3.83 построим автомат распознавающий слова над алфавитом $\{\alpha, \beta\}$, начинающиеся с символов $\alpha\alpha$ и содержащих нечетное число вхождений символа β . Его диаграмма представлена на рис. 3.50.

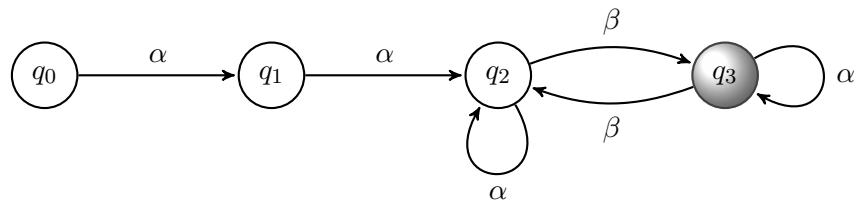


Рис. 3.50

Пример 3.181. Построим автомат распознавающий слова над алфавитом $\{0, 1\}$, содержащие подслово 00, например $\alpha = 01001$. Его диаграмма представлена на рис. 3.51.

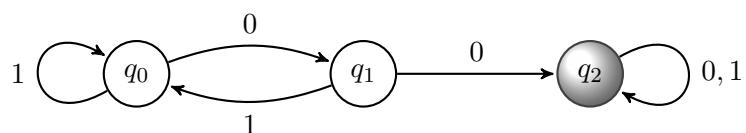


Рис. 3.51

Пример 3.182. Рассмотрим автомат распознавающий множество вещественных констант ($V_T = \{+, -, 0, \dots, 9, .\}$) (рис. 3.52). При этом в константе может быть опущен знак и отсутствовать дробная или целая часть (но не обе сразу). Десятичная точка, как признак вещественного числа — обязательна. При построении диаграммы учтем замечание 3.83.

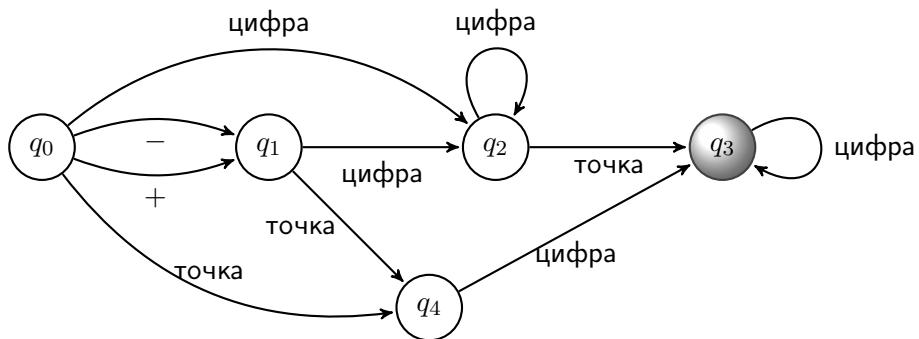


Рис. 3.52

3.10.7. Недетерминированные автоматы распознаватели. Детерминизация

Перейдем теперь к недетерминированным автоматам распознавателям, которые являются обобщением детерминированных.

Определение 3.129. Недетерминированный конечный автомат распознаватель A – это пятерка объектов $A = (Q, \Sigma, \delta, H, F)$, где

- Q – конечное *непустое* множество состояний;
- Σ – конечное *непустое* множество входных символов;
- $\delta : Q \times \Sigma \rightarrow 2^Q$ – функция перехода: всюду определенное отображение множества $Q \times \Sigma$ в множество подмножеств Q ;

$$\delta(q_i, \alpha_j) = \{q_{i_1}, \dots, q_{i_n}\} \subseteq Q,$$

означает, что из состояния q_i по входному символу α_j можно осуществить переход в любое из состояний q_{i_1}, \dots, q_{i_n} .

- $H \subseteq Q$ – множество начальных состояний;
- $F \subseteq Q$ – множество заключительных (финальных) состояний.

Замечание 3.84. Отличия от детерминированного автомата (см. определение 3.124) состоят в задании множества начальных состояний H вместо одного q_0 и выходному значению функции перехода δ , которая возвращает подмножество выходных состояний вместо одного.

Для дальнейшего изучения свойств конечных автоматов важное значение имеет следующая теорема.

Теорема 3.47 (о детерминизации). Для любого конечного автомата может быть построен эквивалентный ему детерминированный конечный автомат.

Доказательство. В качестве доказательства приведем два алгоритма (3.24 (детерминизация объединением) и 3.25 (детерминизация вытягиванием) для построения детерминированного конечного автомата по недетерминированному. ■

Введем необходимые определения.

Определение 3.130. Состояние $q_i \in Q$ автомата A называется *недостижимым*, если под воздействием любого слова автомата A не переходит в состояние q_i .

Состояния, не достижимые из начального, можно найти поиском в ширину, например, используя, следующий алгоритм.

Алгоритм 3.23 (Нахождение недостижимых состояний).

Начало работы. Заносим в список L начальное состояние $q_0 \rightarrow L$.

Текущий шаг. Для каждого состояния, уже включенного в список, добавляем все еще не занесенные в него состояния, которые могут быть достигнуты из этого нового состояния под действием каждого входного символа.

Повторяем этот шаг до тех пор пока удается добавлять новые состояния в список L .

Состояния не включенные в список L удаляем как недостижимые со всеми инцидентными им дугами.

Рассмотрим теперь алгоритм построения детерминированного конечного автомата по недетерминированному.

На вход алгоритма подается недетерминированный конечный автомат $A = (Q, \Sigma, \delta, H, F)$. В результате работы получаем детерминированный конечный автомат $\tilde{A} = (\tilde{Q}, \tilde{\Sigma}, \tilde{\delta}, \tilde{q}_0, \tilde{F})$, допускающий тот же язык, что и автомат A .

Алгоритм 3.24 (Детерминизация объединением).

Шаг 1. Множество состояний \tilde{Q} состоит из всех подмножеств множества Q . Каждое состояние из \tilde{Q} будем обозначать

$$\tilde{q}_i = [q_{i_1} \dots q_{i_k}], q_{i_j} \in Q, 1 \leq j \leq k.$$

Таким образом, если $|Q| = n$, то после первого шага $|\tilde{Q}| = 2^n$.

Шаг 2. Отображение $\tilde{\delta}$ определим как

$$\tilde{\delta}([q_{i_1} \dots q_{i_n}], \alpha) = [q_{j_1} \dots q_{j_m}], \quad (3.93)$$

где

$$\forall q_{j_s} \in \{q_{j_1}, \dots, q_{j_m}\}, \exists q_{i_p} \in \{q_{i_1}, \dots, q_{i_n}\} : \delta(q_{i_p}, \alpha) = q_{j_s}.$$

Объединение состояний согласно (3.93) показано на рис. 3.53.

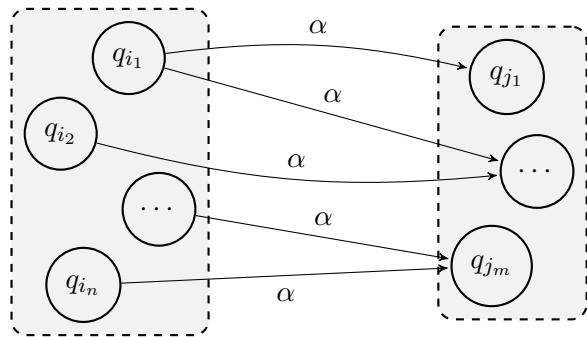


Рис. 3.53

Шаг 3. Начальное состояние \tilde{q}_0 определим как объединение *всех* начальных состояний H :

$$\tilde{q}_0 = [q_{0_1} \dots q_{0_k}], q_{0_i} \in H, 1 \leq i \leq k, k = |H|. \quad (3.94)$$

Шаг 4. Определим множество заключительных состояний \tilde{F} . Пусть $F = \{q_{l_1}, \dots, q_{l_k}\}$. Тогда \tilde{F} это все состояния из \tilde{Q} имеющие вид $[\dots q_l \dots]$, где $q_l \in F$.

Шаг 5. Используя алгоритм 3.23 исключаем возможные недостижимые состояния в \tilde{Q} .

Пример 3.183. На рис. 3.54, б изображен детерминированный автомат \tilde{A} , построенный по алгоритму 3.24 из недетерминированного автомата A на рис. 3.54, а.

Всего, согласно алгоритму 3.24, у нового автомата \tilde{A} должно быть $2^4 = 16$ состояний, но только 4 из них оказываются достижимыми.

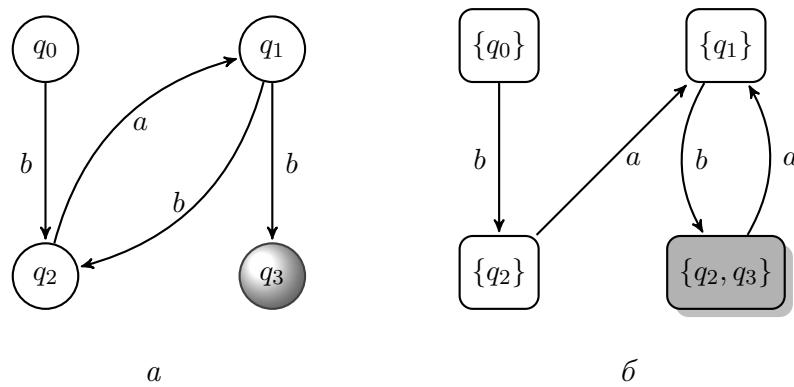


Рис. 3.54

Замечание 3.85. В алгоритме 3.24 на шаге 2 необходимо перебрать 2^n возможных состояний нового детерминированного автомата, а отсчетение недостижимых состояний происходит на шаге 4. Рассмотрим другой подход, где достижимость состояния нового детерминированного автомата определяется сразу при построении.

Алгоритм 3.25 (Детерминизация вытягиванием).

Шаг 1. Как и в предыдущем алгоритме 3.24 начальное состояние \tilde{q}_0 определим как объединение всех начальных состояний H (см. (3.94)). Далее определяем все множества состояний, достижимые из начального, т.е. для каждого входного символа $\alpha \in \Sigma$ находим множество $\delta(\tilde{q}_0, \alpha)$. Каждое такое множество в новом автомате является состоянием, непосредственно достижимым из начального.

Шаг 2. Для каждого из определенных состояний-множеств S и каждого входного символа $\alpha \in \Sigma$ находим множество $\bigcup_{q \in S} \delta(q, \alpha)$. Все полученные на этом шаге состояния будут состояниями нового (детерминированного) автомата, достижимыми из начальной вершины по пути длины 2.

Шаг k . Повторяем предыдущий шаг до тех пор, пока не перестанут появляться новые состояния-множества. Легко показать, что при этом получаются все такие состояния конечного автомата, которые достижимы из начального состояния $\{\tilde{q}_0\}$.

Пример 3.184. Рассмотрим работу алгоритма 3.25 на примере недетерминированного автомата (рис. 3.54,*a*). Имеем

$$\begin{aligned}\delta_1(\{q_0\}, b) &= \{q_2\}; \\ \delta_1(\{q_2\}, a) &= \{q_1\}; \\ \delta_1(\{q_1\}, b) &= \{q_2, q_3\}; \\ \delta_1(\{q_2, q_3\}, a) &= \delta(q_2, a) \cup \delta(q_3, a) = \{q_1\} \cup \emptyset = \{q_1\}.\end{aligned}$$

Так как новых состояний-множеств больше не появилось, процедура «вытягивания» на этом заканчивается, и мы получаем граф, изображенный на рис. 3.54,*b*.

Пример 3.185. Рассмотрим работу алгоритма 3.25 на примере недетерминированного автомата (рис. 3.55). Имеем

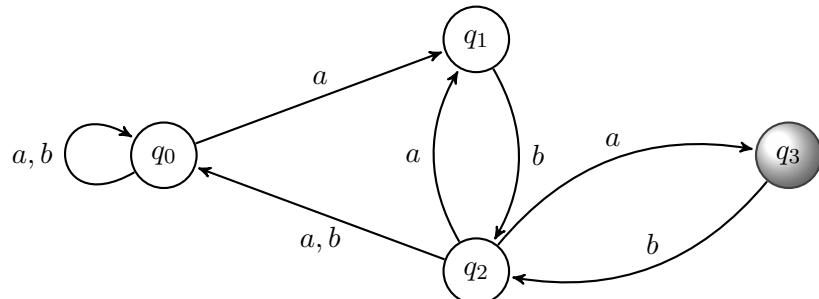


Рис. 3.55

$$\begin{aligned}
 \delta_1(\{q_0\}, a) &= \{q_0, q_1\}; \\
 \delta_1(\{q_0, q_1\}, b) &= \{q_2\}; \\
 \delta_1(\{q_2\}, a) &= \{q_0, q_1, q_3\}; \\
 \delta_1(\{q_0, q_1, q_3\}, b) &= \{q_0, q_2\}; \\
 \delta_1(\{q_0, q_2\}, a) &= \{q_0, q_3\}.
 \end{aligned}$$

Результат работы алгоритма 3.25 показан на рис. 3.56.

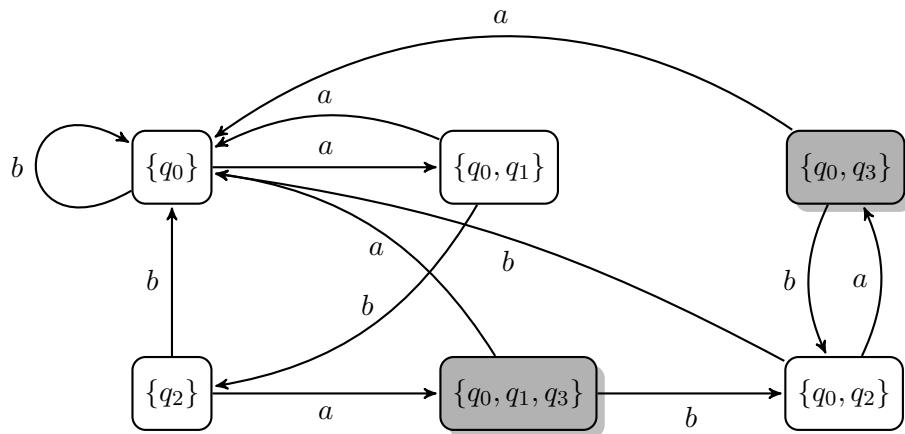


Рис. 3.56

3.10.8. Теорема о разрастании для автоматных языков

Очевидно, что любой конечный язык может быть распознан конечным автоматом, поэтому *все конечные языки автоматные*. Один и тот же язык может распознаваться разными автоматами. Однако, не для всех языков существуют распознающие их конечные автоматы. Иными словами *существуют и неавтоматные языки*.

Установим некоторое необходимое условие, которому удовлетворяют все автоматные языки. После этого, проверив, что некоторый язык этому условию не удовлетворяет, можно заключить, что он не является автоматным. Очевидно, что это условие имеет смысл формулировать только для бесконечных языков.

Теорема 3.48 (о разрастании для автоматных языков). Пусть L – бесконечный автоматный язык. Тогда существует такая константа n , что любое слово $w \in L$ длины $|w| > n$ можно разбить на три части $\alpha_1, \alpha_2, \alpha_3$ так, что $w = \alpha_1\alpha_2\alpha_3$ и

1. $|\alpha_1\alpha_2| \leq n$;
2. $|\alpha_2| > 0$;

$$3. \forall m \geq 0 \quad w_m = \alpha_1 \alpha_2^m \alpha_3 \in L.$$

В последнем условии $\alpha_2^0 = \varepsilon$, $\alpha_2^1 = \alpha_2$, $\alpha_2^{i+1} = \alpha_2^i \alpha_2$.

Доказательство. Так как язык L автоматный, то существует детерминированный конечный автомат $A = (Q, \Sigma, \delta, q_0, F)$, распознающий L . Пусть $|Q| = n$ и слово $w = w_1 w_2 \dots w_k \in L$ имеет длину $k > n$. Рассмотрим путь

$$p = (q_0 = q_{i_0}, q_{i_1}, \dots, q_{i_k})$$

в диаграмме автомата A , который принимает слово w . Очевидно, что среди первых $(n + 1)$ состояний этого пути хотя бы одно встречается дважды. Выберем первое из таких состояний $q \in Q$. Тогда для некоторой пары чисел $l < j \leq n$ имеем $q_{i_l} = q_{i_j} = q$.

Пусть $\alpha_1 = w_1 w_2 \dots w_l$ – это префикс w , который переводит q_0 в $q_{i_l} = q$, $\alpha_2 = w_{i+1} \dots w_j$ – это подслово w , которое переводит $q_{i_l} = q$ в $q_{i_j} = q$, и $\alpha_3 = w_{j+1} \dots w_k$ – это суффикс w , который переводит $q_{i_j} = q$ в $q_{i_k} \in F$.

Части α_1 и α_3 могут быть пусты, но $|\alpha_2| = j - l \geq 1$. Длина $|\alpha_1 \alpha_2| = j \leq n$. Таким образом, условия (1) и (2) теоремы выполнены. Нетрудно убедиться и в выполнении условия (3).

Действительно, выбросив из пути p цикл $q_{i_l} = q, \dots, q_{i_j}$, получим путь p_0 из q_0 в $q_{i_k} \in F$, который принимает слово $\alpha_1 \alpha_3$, а повторив этот цикл m раз, получим путь p_0 из q_0 в $q_{i_k} \in F$, который принимает слово $\alpha_1 \alpha_2^m \alpha_3$. Следовательно,

$$\forall m \geq 0 \quad w_m = \alpha_1 \alpha_2^m \alpha_3 \in L.$$

■

Замечание 3.86. Содержательно, теорема 3.48 утверждает, что у всякого достаточно длинного слова из автоматного языка имеется непустое подслово, которое можно вырезать или повторить сколько угодно раз, оставаясь *внутри* языка.

Пример 3.186. Рассмотрим автомат с диаграммой, представленной на рис. 3.57. Проиллюстрируем теорему 3.48 для слова $w = 1101111$. Имеем $n = 4$, $|w| = 7 > 4$. Путь p для слова w :

$$p = q_0 \xrightarrow{1} q_2 \xrightarrow{1} \underbrace{q_1 \xrightarrow{0} q_2 \xrightarrow{1} q_1}_{\text{цикл}} \xrightarrow{1} q_3 \xrightarrow{1} q_3 \xrightarrow{1} q_3.$$

Повторение состояний $q_1 \xrightarrow{0} q_2 \xrightarrow{1} q_1$. Тогда

$$w = \alpha_1 \alpha_2 \alpha_3, \quad \alpha_1 = 11, \quad \alpha_2 = 01, \quad \alpha_3 = 111;$$

При этом

$$|\alpha_1 \alpha_2| = n = 4, \quad |\alpha_2| = 2.$$

Очевидно, что слово

$$\alpha_1 \alpha_2^m \alpha_3 \in L, \quad \forall m \geq 0.$$

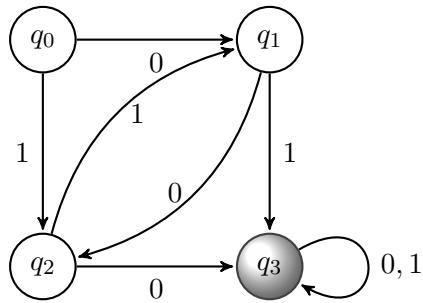


Рис. 3.57

Замечание 3.87. Теорема 3.48 имеет другие названия в литературе: лемма о накачке, лемма Огдена, pumping lemma.

Как, используя теорему 3.48, доказать, что некоторый язык L не является автоматным? Это можно сделать, используя схему доказательства „от противного“.

1. Предположим, что L – автоматный язык. Тогда для него имеется константа n из утверждения теоремы 3.48.

2. Определим по n „специальное“ слово $w \in L$, $|w| > n$ и докажем, что для любого разбиения $w = \alpha_1\alpha_2\alpha_3$, удовлетворяющего условиям (1) и (2) теоремы, найдется такое $k \geq 0$, что слово $w_k = \alpha_1\alpha_2^k\alpha_3 \notin L$.

3. На основании полученного противоречия делаем вывод, что L – не автоматный язык.

Замечание 3.88. В этой схеме самым сложным является выбор „специального“ слова w в пункте (2). Что касается, подбора такого $k \geq 0$, для которого $w_k \notin L$, то, как правило, достаточно рассмотреть $k = 0$ или $k = 2$.

Рассмотрим несколько примеров применения данной схемы.

Пример 3.187. Покажем, что язык $L_1 = \{w = 0^k 1^k \mid k \geq 1\}$ не является автоматным.

Предположим, что он автоматный. Тогда для него имеется n из утверждения теоремы 3.48. Рассмотрим „специальное“ слово $w = 0^n 1^n$. Очевидно, что $w \in L_1$. Предположим, что существует разбиение $w = \alpha_1\alpha_2\alpha_3$, удовлетворяющее условиям (1) и (2) теоремы.

Так как по условию (2) $|\alpha_1\alpha_2| \leq n$, то $\alpha_2 = 0^i$ для некоторого $i > 0$. Но тогда слово $w_0 = \alpha_1\alpha_3 = 0^{n-i}1^n \notin L_1$, что противоречит условию (3) теоремы. Следовательно язык L_1 не автоматный.

Пример 3.188. Покажем, что язык правильных скобочных последовательностей L_2 в алфавите $\{(,)\}$ не является автоматным.

В качестве „специального“ слова выберем слово $w = (^n)^n \in L_2$. Тогда для всякого разбиения $w = \alpha_1\alpha_2\alpha_3$ такого, что $|\alpha_1\alpha_2| \leq n$ слово

$\alpha_2 = (^i \text{ для некоторого } i > 0)$. И, как и в предыдущем примере 3.187, слово $w_0 = \alpha_1\alpha_3 = (^{n-i})^n \notin L_2$, что противоречит условию (3) теоремы. Следовательно, язык L_2 не автоматный.

Упражнение 3.30. С помощью теоремы 3.48 покажите, что язык $L_3 = \{w = 0^k 1 0^k \mid k \geq 1\}$ не является автоматным.

3.10.9. Автоматы и автоматные грамматики

Установим связь между автоматными грамматиками G_3 и недерминированными автоматами-распознавателями.

Определение 3.131. Порождающая грамматика G и конечный автомат-распознаватель A называются эквивалентными, если $L(G) = L(A)$.

Теорема 3.49. Для каждой праволинейной грамматики существует эквивалентный ей конечный автомат.

Доказательство. Каждому нетерминальному символу произвольной праволинейной грамматики G поставим в соответствие одно состояние конечного автомата A . Добавим еще одно состояние T – единственное конечное состояние. Состояние, соответствующее аксиоме грамматики S , будем считать начальным состоянием. Каждому правилу $A \rightarrow aB$ поставим в соответствие команду $Aa \rightarrow B$, а каждому терминальному правилу $A \rightarrow a$ поставим в соответствие команду $Aa \rightarrow T$. Таким образом, выводу цепочки в грамматике

$$S \Rightarrow a_1A_1 \Rightarrow a_1a_2A_2 \Rightarrow \dots \Rightarrow a_1a_2 \dots a_{k-1}A_{k-1} \Rightarrow a_1a_2 \dots a_k$$

взаимно однозначно соответствует последовательность команд построенного автомата A :

$$Aa_1 \rightarrow A_1; A_1a_2 \rightarrow A_2; \dots; A_{k-1}a_k \rightarrow T.$$

Таким образом, язык $L(G) = L(A)$. ■

Пример 3.189. Для грамматики G , заданной следующими правилами вывода

$$S \rightarrow aS|bB; \quad A \rightarrow aA|bS; \quad B \rightarrow bB|c|cA$$

построим эквивалентный ей конечный автомат A .

Граф автомата (рис. (3.58)) будет иметь четыре вершины, три из них помечены нетерминальными символами грамматики S, A, B , четвертая вершина, помеченная символом T , является единственным заключительным состоянием. Начальным состоянием является вершина, соответствующая аксиоме грамматики S .

Согласно теореме 3.49 каждому правилу грамматики поставим в соответствие команду конечного автомата:

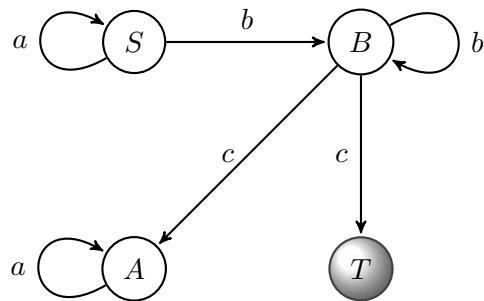


Рис. 3.58

- 1) $Sa \rightarrow S$ – в начальном состоянии при поступлении на вход терминального символа a автомат остается в том же состоянии S ;
- 2) $Sb \rightarrow B$ – из начального состояния при поступлении на вход терминального символа b автомат переходит в состояние B ;
- 3) $Bb \rightarrow B$ – в состоянии B при поступлении на вход терминального символа b автомат остается в том же состоянии B ;
- 4) $Bc \rightarrow T$ – из состояния B при поступлении на вход терминального символа c автомат переходит в заключительное состояние T ;
- 5) $Bc \rightarrow A$ – из состояния B при поступлении на вход терминального символа c автомат переходит в состояние A ;
- 6) $Aa \rightarrow A$ – в состоянии A при поступлении на вход терминального символа a автомат остается в этом же состоянии A ;
- 7) $Ab \rightarrow S$ – из состояния A при поступлении на вход терминального символа b автомат переходит в состояние S .

Полученный *недетерминированный* конечный автомат распознает цепочки языка, порождаемые праворекурсивной грамматикой G .

Пример 3.190. Для грамматики G_4 из примера 3.153 построим эквивалентный ей конечный автомат A_4 . Правила вывода для этой грамматики согласно (3.86) имеют вид:

$$S \rightarrow aA, \quad A \rightarrow aA \mid bC, \quad C \rightarrow cC \mid c$$

Грамматика G_4 является регулярной (праволинейной) грамматикой и порождает язык $L(G_4) = \{a^nbc^m \mid n, m > 0\}$.

Согласно теореме 3.49 каждому правилу грамматики G_4 поставим в соответствие команду конечного автомата. Рассуждая аналогично примеру 3.189 получаем следующий автомат (рис. 3.59).

Полученный *недетерминированный* конечный автомат распознает цепочки языка, порождаемые праворекурсивной грамматикой G_4 , например слово $aaabcc$ из примера 3.153.

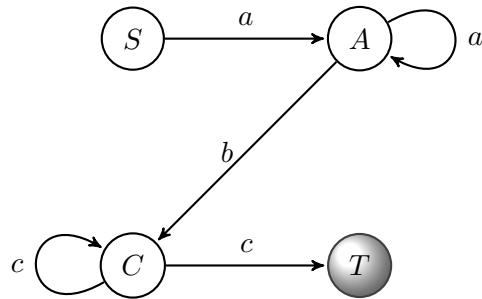


Рис. 3.59

Пример 3.191. Построим конечный автомат, распознающий язык $L(A) = \{(ab)^n \mid n \in N\}$.

Сначала построим саму грамматику G , которая порождает язык $L(A)$:

$$S \rightarrow aA; \quad A \rightarrow bS|b.$$

Проверим, действительно ли эта грамматика порождает язык $L(A)$. Для этого построим несколько выводов возможных вариантов цепочек:

- 1) $S \Rightarrow aA \Rightarrow ab$
- 2) $S \Rightarrow aA \Rightarrow abS \Rightarrow abaA \Rightarrow abab$
- 3) $S \Rightarrow aA \Rightarrow abS \Rightarrow abaA \Rightarrow ababS \Rightarrow ababaA \Rightarrow ababab$

И.т.д.

Таким образом, грамматика G действительно порождает язык $L(A)$, следовательно, можно построить соответствующий этой грамматике конечный автомат (рис. (3.60)). Для этого введем заключительное состояние T , начальное состояние соответствует аксиоме грамматики S .

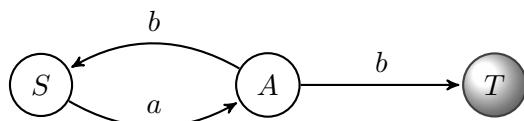


Рис. 3.60

Запишем преобразование правил вывода грамматики G в команды автомата A :

- 1) $Sa \rightarrow A$ – из состояния S при поступлении на вход терминального символа a автомат переходит в состояние A ;
- 2) $Ab \rightarrow S$ – из состояния A при поступлении на вход терминального символа b автомат переходит в состояние S ;
- 3) $Ab \rightarrow F$ – из состояния A при поступлении на вход терминального символа b автомат переходит в заключительное состояние T .

Таким образом, построенный недетерминированный конечный автомат A , распознает заданный язык $L(G)$.

Легко установить и утверждение, обратное к теореме 3.49.

Теорема 3.50. Для произвольного конечного автомата-распознавателя существует эквивалентная порождающая автоматная праволинейная грамматика.

Доказательство. Каждому состоянию произвольного автомата поставим в соответствие нетерминальный символ грамматики, причем начальному состоянию будет соответствовать аксиома грамматики. Тогда для каждой команды $Ac \rightarrow B$ в множество правил грамматики включим правило $A \rightarrow cB$, причем в случае, если B – заключительное состояние, добавим правило $A \rightarrow c$. Эквивалентность исходного конечного автомата и построенной грамматики очевидна. ■

Пример 3.192. Построим порождающую автоматную праволинейную грамматику по недетерминированному конечному автомatu-распознавателю на рис 3.61.

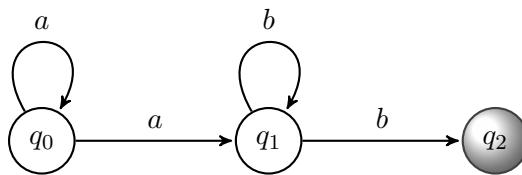


Рис. 3.61

Введя согласно теореме 3.50 обозначения для порождающей грамматики G , получаем набор правил:

$$q_0 \rightarrow S, q_1 \rightarrow A, \quad G : S \rightarrow aS \mid aA, \quad A \rightarrow bA \mid b$$

Построенная грамматика G порождает язык

$$L(G) = \{a^n b^m \mid n \geq 0, m > 0\}.$$

3.10.10. Минимизация автоматов распознавателей

Поставим задачу нахождения наименьшего (по количеству состояний) конечного автомата, распознающего данный язык. Введем необходимые определения.

Определение 3.132. На множестве состояний Q автомата A зададим семейство отношений эквивалентности следующим образом:

a) 0-эквивалентность: для произвольных состояний $q_1, q_2 \in Q$ полагаем $q_1 \stackrel{0}{\equiv} q_2$ тогда и только тогда, когда они оба являются заключительными или оба не являются заключительными;

b) n -эквивалентность: при $n \geq 1$ полагаем $q_1 \stackrel{n}{\equiv} q_2$ тогда и только тогда, когда

$$q_1 \stackrel{n-1}{\equiv} q_2, \quad \delta(q_1, \alpha) \stackrel{n-1}{\equiv} \delta(q_2, \alpha), \quad \forall \alpha \in \Sigma \quad (3.95)$$

Замечание 3.89. Предлагаемый ниже алгоритм 3.26 нахождения минимального автомата был предложен Ауфенкампом и Хоном в [60].

Алгоритм 3.26 (Построение минимального автомата).

Начало работы. Удаляем все недостижимые состояния используя алгоритм 3.23.

Шаг 0. Разбиваем все множество состояний Q на два подмножества 0-эквивалентных состояний: $S_1 = F$ и $S_2 = Q - F$.

Шаг $n + 1$. Разбиваем каждое из подмножеств n -эквивалентных состояний на подмножества $(n + 1)$ -эквивалентных состояний.

Повторяем этот шаг до тех пор пока удается разбить хотя бы одно подмножество.

Завершение работы. Получаем набор подмножеств эквивалентных состояний S_1, \dots, S_k . Внесем в множество состояний минимизированного автомата по одному представителю каждого из множеств S_i .

Замечание 3.90. При работе алгоритма 3.26 возможны два крайних случая:

- 1) все состояния исходного автомата окажутся эквивалентными и тогда в минимальном автомате останется только одно состояние;
- 2) итоговое разбиение будет состоять из одноэлементных классов эквивалентности – это означает, что исходный автомат нельзя минимизировать, он уже минимален.

Первый случай замечания 3.90 проиллюстрирован на рис 3.62. Это автомат, допускающий произвольные цепочки символов a, b .

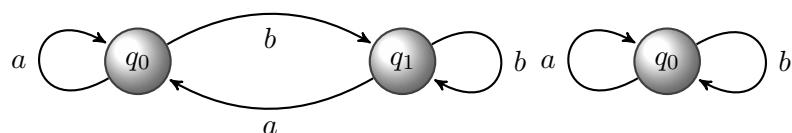


Рис. 3.62

Замечание 3.91. Алгоритм 3.26 часто называют „методом разбиения“, а сам процесс приведения автомата к минимальному – *редукцией конечного автомата*.

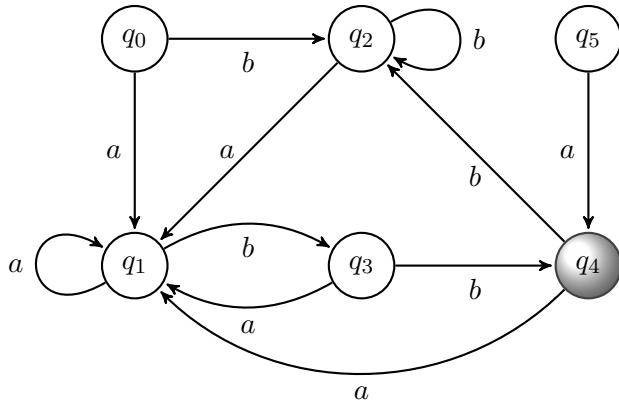


Рис. 3.63

Пример 3.193. Рассмотрим процесс минимизации автомата, диаграмма состояний которого, представлена на рис. 3.63.

Начало работы. Согласно алгоритму 3.23, вначале произведем поиск и удаление недостижимых состояний. Получаем список достижимых состояний:

$$\{q_0\} \rightarrow \{q_0, q_2\} \rightarrow \{q_0, q_2, q_1\} \rightarrow \{q_0, q_2, q_1, q_3\} \rightarrow \{q_0, q_2, q_1, q_3, q_4\}$$

Состояние q_5 – недостижимо, удаляем его вместе с дугой $[q_5, q_4]$ и производим разбиение множества достижимых состояний автомата на классы эквивалентности.

Шаг 0. Запишем начальное разбиение множества состояний автомата для отношения $\stackrel{0}{\equiv}$:

$$\{q_4\}, \{q_0, q_1, q_2, q_3\}.$$

Шаг 1. Заметим, что для состояний q_0, q_1, q_2 автомат переходит в одно из состояний класса эквивалентности предыдущего уровня – $\{q_0, q_1, q_2, q_3\}$, поэтому все они 1-эквивалентны.

Но состояние $\delta(q_3, b) = q_4 \notin \{q_0, q_1, q_2, q_3\}$, поэтому, согласно (3.95), состояние q_3 не 1-эквивалентно состояниям q_0, q_1, q_2 . В разбиении для 1-эквивалентности они «разойдутся» по разным классам; разбиение, определяемое отношением $\stackrel{1}{\equiv}$, будет иметь вид:

$$\{q_4\}, \{q_0, q_1, q_2\}, \{q_3\}.$$

Шаг 2. Далее, для состояний q_0 и q_2 автомат переходит в одно из состояний класса эквивалентности предыдущего уровня $\{q_0, q_1, q_2\}$, поэтому они 2-эквивалентны.

Однако состояние $\delta(q_1, b) = q_3 \notin \{q_0, q_1, q_2\}$, поэтому q_1 не 2-эквивалентно q_0 и q_2 . Произойдет разделение; разбиение, определяемое отношением $\stackrel{2}{\equiv}$, будет иметь вид:

$$\{q_4\}, \{q_0, q_2\}, \{q_1\}, \{q_3\}.$$

Завершение работы. Заметим, что согласно (3.95), $q_0 \stackrel{3}{\equiv} q_2$. Действительно,

$$q_0, q_2, \delta(q_0, b), \delta(q_2, b) \in \{q_0, q_2\}, \quad \delta(q_0, a), \delta(q_2, a) \in \{q_1\}.$$

Поэтому разбиение на классы 2-эквивалентности есть искомое разбиение. Таким образом, состояния q_0 и q_2 неразличимы. Полученный минимизированный автомат представлен на рис. 3.64.

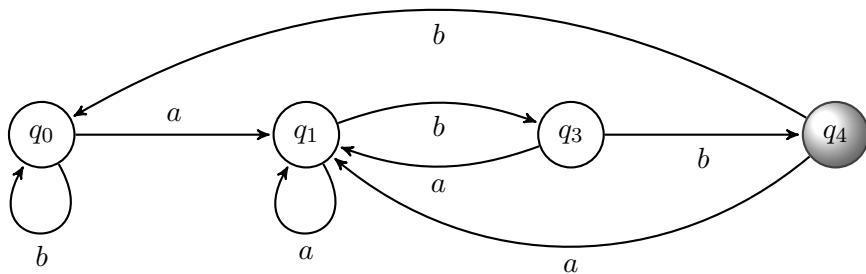


Рис. 3.64

Пример 3.194. Проведем минимизацию автомата, диаграмма состояний которого, представлена на рис. 3.65.

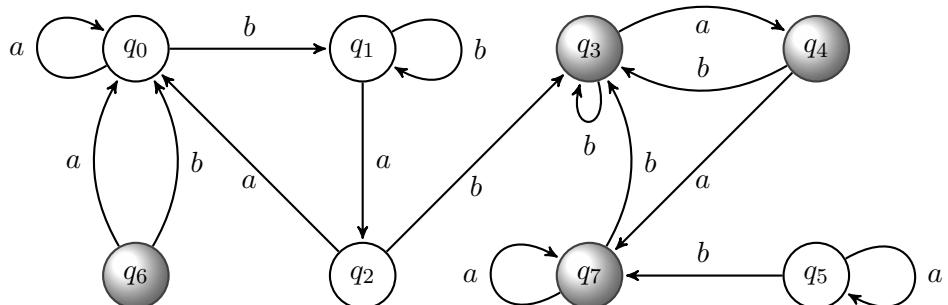


Рис. 3.65

Начало работы. Используя алгоритм 3.23 удаляем недостижимые состояния: q_6, q_5 .

Шаг 0. Начальное разбиение множества состояний автомата для отношения $\stackrel{0}{\equiv}$:

$$\{q_0, q_1, q_2\}, \{q_3, q_4, q_7\}.$$

Шаг 1. Для состояний q_0 и q_1 автомат переходит в одно из состояний класса эквивалентности предыдущего уровня — $\{q_0, q_1, q_2\}$, поэтому они 1-эквивалентны. Но состояние $\delta(q_2, b) = q_3 \notin \{q_0, q_1, q_2\}$, следовательно, q_3 не 1-эквивалентно q_0 и q_1 . Для отношения $\stackrel{1}{\equiv}$ получаем новое разбиение:

$$\{q_0, q_1\}, \{q_2\}, \{q_3, q_4, q_7\}.$$

Шаг 2. Далее, состояния $\delta(q_0, a) = q_0$ и $\delta(q_1, a) = q_2$ не 1-эквивалентны, для отношения $\stackrel{2}{\equiv}$ получаем разбиение:

$$\{q_0\}, \{q_1\}, \{q_2\}, \{q_3, q_4, q_7\}.$$

Завершение работы. Для всех состояний из множества $\{q_3, q_4, q_7\}$ автомат переходит в одно из этих же состояний, то разбиение на классы 2-эквивалентности и есть искомое разбиение. Таким образом, состояния q_3, q_4, q_7 неразличимы. Минимизированный автомат представлен на рис. 3.66.

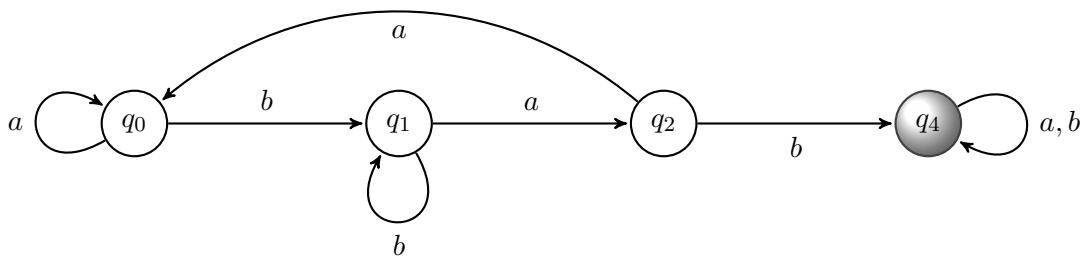


Рис. 3.66

Замечание 3.92. Неизвестно, существует ли быстрый (полиномиальный) алгоритм, позволяющий по произвольному конечному автомату находить минимальный автомат среди всех (не обязательно детерминированных) конечных автоматов, эквивалентных исходному автомату.

3.10.11. Конечные автоматы с эпсилон переходами

Определение 3.133. Эпсилон переходом называется переход между состояниями, который может быть выполнен автоматом «просто так», без входного символа. На графах и в таблицах такие переходы обычно помечаются символом ϵ .

Одно из наиболее полезных свойств ϵ -переходов – возможность простого объединения нескольких автоматов в один. При этом если на вход подаются детерминированные автоматы, то при объединении может получиться недетерминированный автомат. Рассмотрим этот прием на простом примере:

Пример 3.195. На рис. 3.67 изображены два исходных конечных автомата.

На рис. 3.68 показан результат объединения двух входных автоматов.

Никаких дополнительных ограничений на ϵ переходы при работе конечного автомата не накладывается.



Рис. 3.67

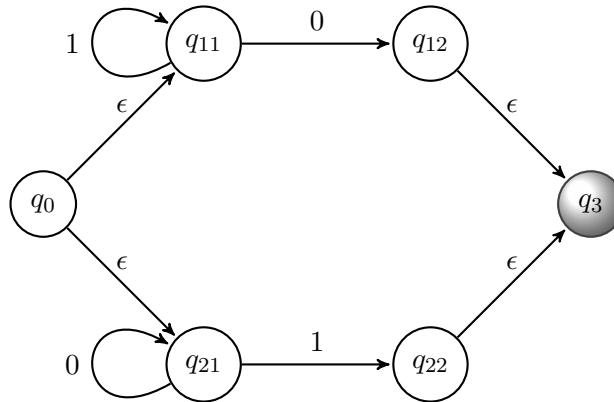


Рис. 3.68

1. Из одного состояния недерминированного конечного автомата может выходить сколько угодно как обычных, так и ϵ переходов.
2. Может существовать цепочка из нескольких последовательных ϵ переходов.
3. Автомат может проходить цепочку целиком, частично или не проходить ее вообще (если у автомата есть другие варианты поведения).

В общем, график/таблица переходов автомата выглядят так, как будто во входном алфавите появился еще один символ – ϵ , а его работа так, будто в любом месте обрабатываемой цепочки (в начале, в конце, между символами) находится произвольное количество этих символов. При этом для каждого состояния есть как минимум один переход по ϵ – в себя.

Замечание 3.93. Нетрудно показать, что по недерминированному автомату с ϵ переходами можно построить эквивалентный ему конечный детерминированный автомат, но проектировать ϵ автомат гораздо проще и удобнее.

Действительно, уберем ϵ -переходы следующим образом. Найдем все пары состояний (q_k, q_l) , такие, что q_l достижимо из q_k по ϵ -переходам:

$$q_k \xrightarrow{\epsilon} \dots q_i \xrightarrow{\epsilon} \dots \xrightarrow{\epsilon} q_l$$

Для каждого не- ϵ -перехода $\delta(q_l, \alpha) = q_s$ добавим переход $\delta(q_k, \alpha) = q_s$. Кроме того, если $q_l \in F$, добавим и q_k в F . Далее, при необходимости, нужно применить один из рассмотренных выше алгоритмов детерминизации (3.24 или 3.25).

Пример 3.196. Уберем ϵ -переходы для объединенного автомата из примера 3.195 (рис. 3.68). На рис. 3.69 представлен результат. Детерминизация не потребовалась.

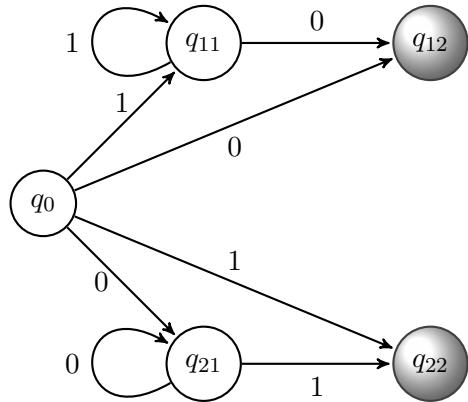


Рис. 3.69

3.10.12. Конечные автоматы и регулярные выражения

Регулярные выражения являются достаточно удобным средством для построения «алгебраических» описаний языков. Они строятся из элементарных выражений с помощью операций *объединения* ($+$, $|$, \cup)¹, *конкатенации (цепления)* (\cdot) и *итерации* ($*$). Каждому такому выражению r соответствует представляемый им язык L_r .

Замечание 3.94. Другие общеупотребительны названия регулярных выражений: *Regular Expressions*, *RegExpr*. Активно используются в языках программирования *Python*, *Perl*, *Ruby*, *Java*, *.Net* и в текстовых редакторах *Vim*, *EmEdit*.

Введем необходимые определения.

Определение 3.134. *Объединение* языков $L_1 + L_2$ содержит все слова, содержащиеся или в L_1 , или в L_2 .

Конкатенация (цепление) $L_1 \hat{\cdot} L_2$ содержит все слова, удовлетворяющие форме vw , где $v \in L_1$, а $w \in L_2$.

*Итерацию*² (L) * языка L образуют все слова которые можно разбить на несколько подряд идущих слов из L :

$$(L)^* = \{\varepsilon\} \cup \{w \mid (\exists n \geq 0)(w = w_1 w_2 \dots w_n), w_i \in L, i \in 1 : n\}$$

Ее можно представить с помощью степеней:

$$(L)^* = \bigcup_{i=0}^{\infty} L^i$$

¹ В различных описаниях возможно использование одного из данных символов

² Иногда ее называют замыканием Клини

Заметим, что $(L)^*$ включает и пустое слово ϵ , так как $n = 0$ допустимо по условию.

Введем некоторые соглашения. При записи регулярных выражений будем опускать знак конкатенации \cdot . Расставим приоритеты операций по уменьшению: итерация, конкатенация, объединение. Это позволит опустить многие скобки.

Пример 3.197. Выражение $((1^0)^\wedge((1)^* + 0))$ можно записать как $10(1^* + 0)$.

Дадим строгое определение регулярного выражения. Регулярное выражение над алфавитом Σ определяется рекурсивно следующим образом.

Определение 3.135.

- 1) Пустой символ ϵ является регулярным выражением;
- 2) для любого $a \in \Sigma$, a является регулярным выражением;
- 3) если r и p являются регулярными выражениями, то $(r + p)$, (rp) и r^* тоже являются регулярными выражениями.

Определение 3.136. Два регулярных выражения r и p называются *эквивалентными*, если совпадают представляемые ими языки, т.е. $L_r = L_p$. В этом случае будем писать $r = p$.

Теорема 3.51. Справедливы следующие свойства регулярных операций:

- 1) $r + p = p + r$ (коммутативность объединения),
- 2) $(r + p) + q = r + (p + q)$ (ассоциативность объединения),
- 3) $(rp)q = r(pq)$ (ассоциативность конкатенации),
- 4) $(r^*)^* = r^*$ (идемпотентность итерации),
- 5) $(r + p)q = rq + pq$ (дистрибутивность).

Доказательство. Перечисленные свойства очевидным образом следуют из введенных выше определений. ■

Следствие 3.10. Класс регулярных языков замкнут относительно операций конкатенации и итерации.

Рассмотрим несколько примеров регулярных выражений и представляемых ими языков.

Пример 3.198. Регулярное выражение $(0 + 1)^*$ представляет множество всех слов в алфавите $\{0, 1\}$ включая пустое слово.

Пример 3.199. Множество всех слов в алфавите $\{0, 1\}$, содержащих ровно одно вхождение 1 соответствует регулярному выражению 0^*10^* .

Пример 3.200. Регулярное выражение $11(0 + 1)^*001$ представляет язык, состоящий из всех слов в алфавите $\{0, 1\}$, которые начинаются на 11, а заканчиваются на 001.

Теорема 3.52. Для каждого регулярного выражения r можно эффективно построить такой недетерминированный конечный автомат A , который распознает язык, задаваемый r , т.е. $L_A = L_r$.

Доказательство этой теоремы достаточно техническое, построение автомата A по выражению r проводится индукцией по длине регулярного выражения r .

Следствие 3.11. Для каждого регулярного выражения можно эффективно построить детерминированный конечный автомат, который распознает язык, представляемый этим выражением.

Рассмотрим алгоритм построения недетерминированного конечного автомата на основе регулярного выражения. Используем автоматы с ϵ переходами, рассмотренными в разделе 3.10.11.

Будем строить недетерминированный конечный автомат (НКА) как композицию таких автоматов для более простых выражений исходя из определения 3.135. Комбинирование выполняется по правилам, представленным ниже.

Алгоритм 3.27.

Шаг 1. Для ϵ строим НКА, i – новое начальное состояние, f – новое заключительное состояние. Этот НКА (рис. 3.70) распознает язык $L = \{\epsilon\}$.

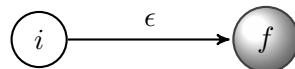


Рис. 3.70

Шаг 2. Для $a \in \Sigma$ строим НКА, i – новое начальное состояние, f – новое заключительное состояние. Этот НКА (рис. 3.71) распознает язык $L = \{a\}$

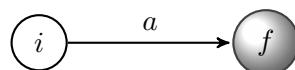


Рис. 3.71

Шаг 3. Пусть A_s и A_t представляют собой НКА для регулярных выражений s и t . Для регулярного выражения $s + t$ строим НКА A_{s+t} , i – новое начальное состояние, f – новое заключительное состояние. Этот НКА (рис. 3.72) распознает язык $L_s \cup L_t$.

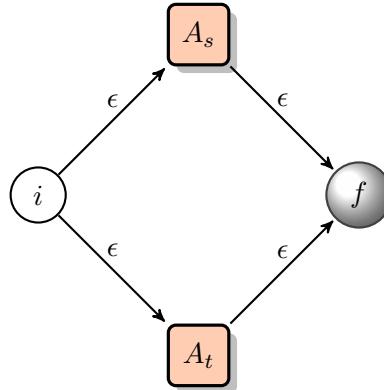


Рис. 3.72

Шаг 4. Для регулярного выражения st строим НКА A_{st} . Здесь начальное состояние НКА A_s становится новым начальным состоянием A_{st} , заключительное состояние A_t становится заключительным состоянием A_{st} . А конечное состояние A_s и начальное состояние A_t объединяются. Этот НКА (рис. 3.73) распознает язык $L_s L_t$.

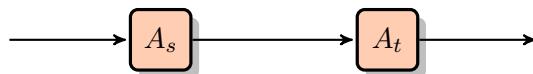


Рис. 3.73

Шаг 5. Для регулярного выражения s^* строим НКА A_{s^*} , i – новое начальное состояние, f – новое заключительное состояние. Этот НКА (рис. 3.74) распознает язык $L(s^*)$.

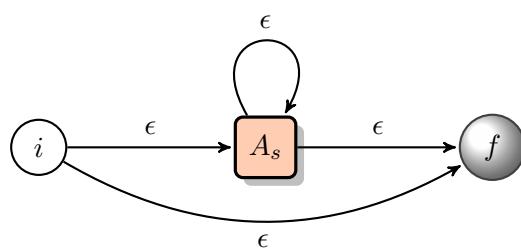


Рис. 3.74

Пример 3.201. Используя алгоритм 3.27 построим НКА по регулярному выражению $r = (ab)^*a^*ba$.

Используя шаги 2,4 и 5 строим НКА A_r (рис. 3.75) для регулярного выражения $r = (ab)^*$.

Используя шаги 2 и 5 строим НКА A_s (рис. 3.76) для регулярного

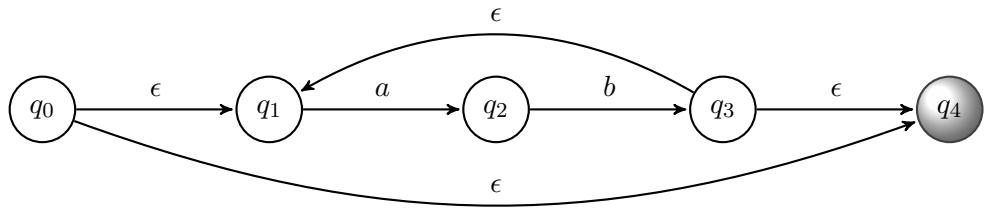


Рис. 3.75

выражения $s = a^*$.

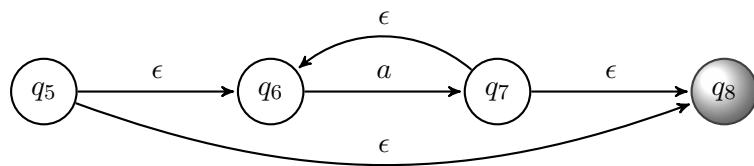


Рис. 3.76

Используя шаги 2 и 4 строим НКА A_t (рис. 3.77) для регулярного выражения $t = ba$.

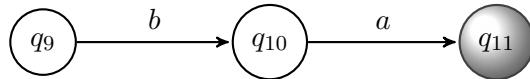


Рис. 3.77

Соединяем построенные автоматы A_r, A_s и A_t (рис. 3.78).

Пример 3.202. Построим детерминированный конечный автомат для регулярного выражения $(a + b)^*a(b + a)$.

Построим сначала недетерминированный автомат (рис. 3.79 а), затем детерминизируем его. В данном случае при построении можно обойтись без эпсилон переходов.

Применяя алгоритм 3.25 детерминизации вытягиванием, получаем:

$$\begin{array}{ll} \delta_1(\{q_0\}, a) = \{q_0, q_1\}; & \delta_1(\{q_0\}, b) = \{q_0\}; \\ \delta_1(\{q_0, q_1\}, a) = \{q_0, q_1, q_2\}; & \delta_1(\{q_0, q_1\}, b) = \{q_0, q_2\}; \\ \delta_1(\{q_0, q_2\}, a) = \{q_0, q_1\}; & \delta_1(\{q_0, q_2\}, b) = \{q_0\}; \\ \delta_1(\{q_0, q_1, q_2\}, a) = \{q_0, q_1, q_2\}; & \delta_1(\{q_0, q_1, q_2\}, b) = \{q_0, q_3\}; \\ \delta_1(\{q_0, q_3\}, a) = \{q_0, q_1\}; & \delta_1(\{q_0, q_3\}, b) = \{q_0\}; \end{array}$$

Полученный детеминированный автомат представлен на рис. 3.79 б.

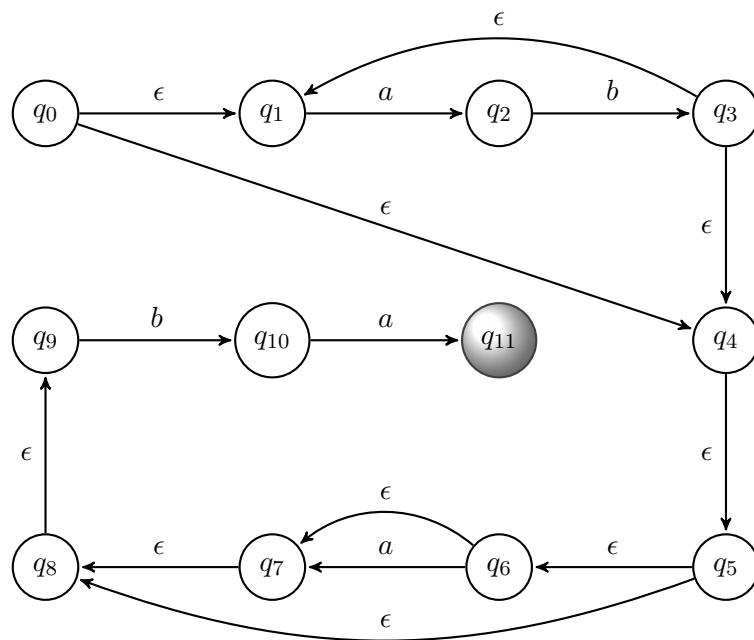


Рис. 3.78

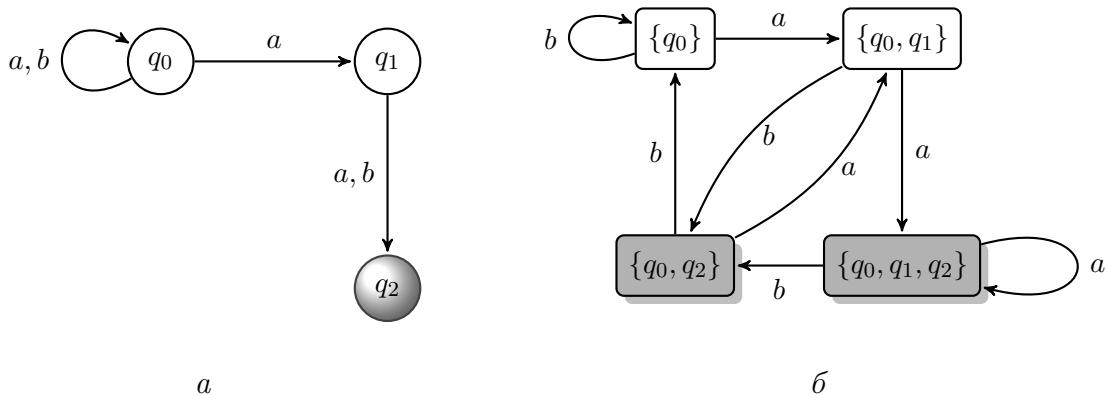


Рис. 3.79

3.11. Тесты и задачи для экзамена

3.11.1. Целочисленные алгоритмы

Тест 3.1. Как провести быструю факторизацию чисел при наличии „компьютера бога“?

Тест 3.2. Как провести факторизацию „без делений“?

Тест 3.3. Для какой пары чисел алгоритм Евклида работает медленнее всего?

Тест 3.4. Критерий разрешимости диофантова уравнения: $ax+by = c$?

Тест 3.5. Возможно ли такое представление числа в фибоначчиевой системе счисления: $(1, 0, 0, 1, 1, 0, 1, 1, 0, 1)$?

Тест 3.6. Приведите пример технологии распараллеливания арифметических операций: работа без переноса в следующий разряд.

3.11.2. Комбинаторика, кодирование

Тест 3.7. Обладает ли данный код свойством префикса:

$$ac, c, bb, ca, abc, bac, cab, abb, bcab, abcb, abbc$$

Задача 3.53. Расставьте в лексикографическом (антилексикографическом) порядке слова:

$$(dcda), (badc), (cabd), (abdc), (bbaa)$$

Тест 3.8. Улучшите двоичный код чтобы не было резких изменений между соседними элементами, например: $(0, 1, 1, 1), (1, 0, 0, 0)$.

Задача 3.54. Есть открытая часть ключа RSA: (e, m) . Зашифруйте число N .

Тест 3.9. На чем основана криптостойкость системы шифрования RSA?

Задача 3.55. Есть открытая часть ключа RSA: (e, m) и зашифрованное сообщение M . Предложите способ дешифровки сообщения M .

3.11.3. Бинарные отношения

Тест 3.10. Определим бинарное отношение на множестве статей P следующим образом: пара статей α и β принадлежат P тогда и только тогда, когда они имеют общую ссылку на одну и ту же статью. Является ли это отношение отношением эквивалентности?

Тест 3.11. Является ли отношение перпендикулярности прямых отношением эквивалентности?

Тест 3.12. Отношение M на множестве точек плоскости: $(P_1, P_2) \in M$, если ординаты точек P_1 и P_2 равны. Является ли отношение M отношением эквивалентности?

Тест 3.13. Может ли отношение эквивалентности быть одновременно и отношением порядка?

Задача 3.56. Какими свойствами обладает отношение

$$xRy \leftrightarrow |x - y| = 1, \quad x, y \in \{1, 2, 3, 4, 5\}.$$

Задача 3.57. Отношение R на некотором множестве слов определено следующим образом: пара слов W_1 и W_2 принадлежит M , если и только если они начинаются с одного и того же символа. Является ли R отношением эквивалентности? Отношением толерантности?

Задача 3.58. Отношение R на некотором множестве слов определено следующим образом: пара слов W_1 и W_2 принадлежит M , тогда и только тогда, когда они отличаются не более чем на одну букву. Является ли R отношением эквивалентности? Отношением толерантности?

Задача 3.59. Нарисуйте граф с пятью вершинами, соответствующий рефлексивному симметричному и нетранзитивному отношению.

Задача 3.60. Как построить классы эквивалентности для отношения эквивалентности?

Задача 3.61. Как сделать отношение частичного порядка полным?

Задача 3.62. Является ли отношение, заданное на рисунке, транзитивным?



3.11.4. Графы

Тест 3.14. Справедливо ли утверждение: „каждое дерево является двудольным графом“?

Тест 3.15. В чем отличие понятий слабой и сильной связности для ориентированных графов?

Тест 3.16. Что такое транспонированный граф?

Тест 3.17. Что такое мост, точка сочленения?

Тест 3.18. Что такое граф конденсации (граф Герца)?

Тест 3.19. Пусть все степени вершин неориентированного графа равны 2. В каком случае этот граф будет эйлеровым?

Тест 3.20. Может ли эйлеров граф содержать мост?

Тест 3.21. Существует ли граф с набором степеней: $(2, 2, 3, 3, 3)$?

Задача 3.63. В графе 100 вершин и 800 ребер. Есть ли в этом графе есть хотя бы одна вершина степени не меньше 16?

Тест 3.22. Сколько ребер в графе $K_{m,n}$?

Задача 3.64. В полном двудольном графе 143 ребра Определить число вершин в каждой доле V_1 и V_2 , если $|V_1| > 1$ и $|V_2| > 1$.

Тест 3.23. Сколько мостов имеет дерево с p ребрами?

Тест 3.24. Можно ли представить неориентированное дерево в виде двудольного графа?

Тест 3.25. Что собой представляет операция „излом ребра“? Для чего используется?

Тест 3.26. Являются ли графы K_4 и K_5 планарными?

Тест 3.27. Существует ли полный граф с семью ребрами?

Задача 3.65. Найти хроматическое число связного графа, у которого p вершин и p ребер.

Задача 3.66. Какое наибольшее число ребер может быть в несвязном графе с p вершинами?

Задача 3.67. Какое наименьшее число ребер может быть в связном графе с p вершинами?

Задача 3.68. Найти число компонент связности леса, имеющего 20 вершин и 10 ребер.

Задача 3.69. В группе 24 студента. Может ли быть так, что 9 из них имеют по 5 друзей среди студентов этой группы, 5 – по 4 друга и 10 – по 3 друга?

Тест 3.28. Можно ли построить 6-вершинный неориентированный граф, имеющий такой набор степеней вершин: $(2, 2, 2, 4, 5, 5)$.

Тест 3.29. Сколько ребер в связном графе с p вершинами, если в нем имеется единственный цикл?

Тест 3.30. Какое минимальное число вершин связного графа с p ребрами?.

Задача 3.70. Каково наибольшее число ребер в двудольном графе с p вершинами?

Тест 3.31. Сколько ребер может быть у регулярного графа с 9 вершинами?

Тест 3.32. Сколько ребер имеет регулярный граф с p вершинами степени r ?

Тест 3.33. Связный граф с p вершинами содержит единственный цикл. Сколько у него остовов?

Тест 3.34. Может ли связный граф иметь ровно два остова?

Тест 3.35. Пусть все ребра связного графа с p вершинами имеют один и тот же вес α . Каков наименьший из весов его остовов?

Тест 3.36. Степени валентности вершин связного графа: $(2, 3, 6, 4, 3, 2, 2, 8)$. Существует ли в нем замкнутый (открытый) эйлеров путь?

Тест 3.37. Как по коду Прюфера вычислить степень каждой вершины, не восстанавливая само дерево?

Задача 3.71. Связный планарный граф, имеющий 7 ребер, изображен на плоскости так, что они не имеют пересечений. Граф разбивает плоскость на три части. Сколько в нем вершин?

Задача 3.72. Связный планарный граф, имеющий 7 вершин, изображен на плоскости так, что его ребра не имеют пересечений. Он разбивает плоскость на пять частей. Сколько в нем ребер?

Тест 3.38. Три поссорившихся соседа имеют три общих колодца. Можно ли провести непересекающиеся дорожки от каждого дома к каждому колодцу?

Задача 3.73. Опишите граф с хроматическим числом $\chi(\mathbf{G}) = 1$ или $\chi(\mathbf{G}) = 2$.

Тест 3.39. Что является инвариантом при построении максимального паросочетания?

Тест 3.40. В чем разница между эйлеровым и гамильтоновым графом?

Тест 3.41. Что такое „правильная“ раскраска графа?

Тест 3.42. Что сохраняется в качестве инварианта при построении минимального каркаса алгоритмом Прима или Краскала?

Тест 3.43. Как построить кратчайший путь в графе с „препятствиями“?

3.11.5. Логика высказываний, булевы функции

Тест 3.44. Какой дизъюнкт выводим по правилу резолюций в логике высказываний из дизъюнктов $X \vee F$ и $\neg X \vee G$.

Тест 3.45. Для какой формулы логики высказываний нет равносильной ей, имеющей СДНФ.

Тест 3.46. Для какой формулы логики высказываний нет равносильной ей, имеющей СКНФ.

Задача 3.74. Существует ли формула F такая, что формула G будет тождественно истинна:

$$G = (F \wedge Y \rightarrow \neg Z) \rightarrow ((Z \rightarrow \neg Y) \rightarrow F).$$

Задача 3.75. Является ли формула G логическим следствием формул F_1, F_2, F_3, F_4 :

$$F_1 = X \vee Y \vee Z, F_2 = X \rightarrow X_1, F_3 = Y \rightarrow X_1 \vee Y_1, F_4 = \neg Y_1, G = Z \rightarrow X_1.$$

Задача 3.76. Методом минимизирующих карт или Куайна – МакКласски найдите минимальную ДНФ для формулы логики высказываний $F(X, Y, Z)$, принимающую значение 1 на наборах $(1, 1, 1), (1, 0, 1), (0, 0, 1)$ и $(0, 0, 0)$.

Задача 3.77. Методом минимизирующих карт или Куайна – МакКласски найдите минимальную ДНФ для формулы логики высказываний $F(X, Y, Z)$, принимающую значение 1 на наборах $(1, 1, 0), (1, 0, 0), (0, 1, 1)$ и $(0, 0, 0)$.

Задача 3.78. Построить булеву функцию от трех переменных, для которой

$$f(1, 0, 0) = 1, f(0, 1, 1) = 0$$

Задача 3.79. Найти все функционально полные системы, состоящие из одной 2-местной булевой функции.

Тест 3.47. Опишите классы Поста T_0, T_1, M, S, L .

Задача 3.80. Булева функция $f(x, y, z)$ обладает следующими свойствами:

$$f \in M, f(1, 1, 1) = 0.$$

Опишите эту функцию.

Задача 3.81. Булева функция $f(x, y, z)$ обладает следующими свойствами:

$$f \in M, f(0, 0, 0) = 1.$$

Опишите эту функцию.

Тест 3.48. Про булеву функцию f известно, что она монотонная и непостоянная. Справедливо ли утверждение: f сохраняет ноль.

Тест 3.49. Сколько различных булевых функций от n переменных.

Тест 3.50. Сколько различных самодвойственных булевых функций от n переменных.

Тест 3.51. Сколько различных булевых функций от n переменных, сохраняющих ноль.

Тест 3.52. Сколько различных линейных булевых функций от n переменных.

Тест 3.53. Как построить разложение Шеннона для заданной булевой функции?

Задача 3.82. Покажите, что $\{\leftrightarrow, \oplus\}$ не составляют функционально полной системы функций. Предложите способ сделать эту систему полной добавлением одной, не более чем 2-местной функции.

Задача 3.83. Найти все функционально полные системы, состоящие из одной 2-местной булевой функции.

Задача 3.84. Выяснить функциональную полноту следующей системы функций: $\{xy \oplus x, x \oplus y, 1\}$.

Задача 3.85. Выяснить функциональную полноту следующей системы функций: $\{xy, x \oplus y, x \leftrightarrow (xy)\}$.

Задача 3.86. Выяснить функциональную полноту следующей системы функций: $\{\bar{x}, xy \vee xz \vee yz\}$.

Задача 3.87. Выяснить функциональную полноту следующей системы функций: $\{x \oplus y, x \vee y \vee \bar{z}\}$.

Тест 3.54. Можно ли получить с помощью операций суперпозиции над множеством $F = \{\rightarrow\}$ функцию $f(x, y) = x \oplus y$

Тест 3.55. Можно ли получить с помощью операций суперпозиции над множеством $F = \{\wedge, \vee\}$ функцию $f(x, y) = x \rightarrow y$

Тест 3.56. Можно ли получить отрицание с помощью операций суперпозиции над множеством $F = \{\vee, \wedge, \rightarrow, \equiv\}$

Тест 3.57. Можно ли получить импликацию с помощью операций суперпозиции над множеством $F = \{\vee, \wedge\}$

Тест 3.58. Можно ли получить конъюнкцию с помощью операций суперпозиции над множеством $F = \{\vee, \rightarrow\}$

Тест 3.59. Может ли суперпозиция нелинейных функций дать функцию линейную?

3.11.6. Логика предикатов

Задача 3.88. Запишите бинарную резольвенту дизъюнктов:

$$D_1 = \neg Q(\alpha, f(x)) \vee R(x), D_2 = Q(u, z) \vee P(z).$$

Задача 3.89. Примените правило склейки к дизъюнкту

$$Q(x, y, \gamma) \vee Q(\alpha, y, z) \vee Q(x, \beta, z) \vee \neg R(x, y, z).$$

Тест 3.60. Что такое *терм* в исчислении предикатов?

Тест 3.61. Что такое область действия квантора?

Тест 3.62. Объясните понятие связного и свободного вхождения переменной в формулу.

Тест 3.63. Сколько свободных вхождений переменной y в формулу логики предикатов

$$F = T(x) \wedge (\forall x) [S(x, y) \rightarrow (\exists x) (R(x, y) \vee T(y))]$$

Тест 3.64. Что такое *интерпретация* в логике предикатов?

Тест 3.65. Что такое *выполнимая* формула в исчислении предикатов

Тест 3.66. Можно ли в формулах логики предикатов переставлять рядом стоящие *одноименные* кванторы?

Тест 3.67. Можно ли в формулах логики предикатов переставлять рядом стоящие *разноименные* кванторы?

Тест 3.68. Привести пример формулы логики предикатов в *предваренной нормальной форме*.

Тест 3.69. Привести пример формулы логики предикатов в *сколемской нормальной форме*.

Тест 3.70. Какое свойство формулы логики предикатов сохраняется при ее сколемизации?

3.11.7. Нечеткие множества

Тест 3.71. Что такое нечеткое множество, заданное на универсуме M ?

Тест 3.72. Пусть имеются два нечетких множества A и B , заданные функциями принадлежности $\mu_A(x)$, $\mu_B(x)$. Определите $A \cup B$.

Тест 3.73. Пусть имеются два нечетких множества A и B , заданные функциями принадлежности $\mu_A(x), \mu_B(x)$. Определите $A \cap B$.

Тест 3.74. Приведите пример функции принадлежности для нечеткого множества.

Тест 3.75. Что такое мера нечеткости для нечеткого множества.

Задача 3.90. Приведите пример нечеткого бинарного отношения, заданного графом или таблицей.

Задача 3.91. Дано универсальное множество

$$M = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}$$

и нечеткие множества $A \subseteq M$ и $B \subseteq M$:

$$\begin{aligned} A &= \{(x_1, 0.6); (x_2, 0.3); (x_3, 0.7); (x_4, 0.4); (x_5, 0.4); (x_6, 0.5); (x_7, 1.0)\}, \\ B &= \{(x_1, 0.4); (x_2, 0.6); (x_3, 0.5); (x_4, 0.4); (x_5, 0.5); (x_6, 0.5); (x_7, 0.7)\}. \end{aligned}$$

Найти $A \cup B$, $A \cap B$, меры нечеткости A и B по Егеру.

3.11.8. Машина Тьюринга, алгоритмы Маркова

Тест 3.76. Что означает, что машина Тьюринга *не применима* к данной конфигурации?

Тест 3.77. Что означает, что машина Тьюринга *применима* к данной конфигурации?

Задача 3.92. Постройте машину Тьюринга не применимую ни к одной конфигурации.

Тест 3.78. Что означает, что машина Тьюринга правильно вычисляет функцию $f(x_1, \dots, x_n)$?

Задача 3.93. Постройте машину Тьюринга, правильно вычисляющую $f(x) = x$.

Тест 3.79. Что такое *геделева нумерация* для машины Тьюринга?

Задача 3.94. Построить машину Тьюринга в алфавите $\mathcal{A} = \{a, b, c\}$, которая заменяет на a каждый второй символ конфигурации.

Задача 3.95. Построить машину Тьюринга в алфавите $\mathcal{A} = \{a, b, c\}$, которая определяет, входит ли в конфигурацию символ a . Ответ: конфигурация из одного символа a (да, входит) или пустая конфигурация (нет).

Задача 3.96. Построить машину Тьюринга в алфавите $\mathcal{A} = \{0, 1\}$, которая считая конфигурацию записью двоичного числа, удалить из нее незначащие нули, если такие есть.

Тест 3.80. Что такое Р и NP классы сложности?

Тест 3.81. Когда нормальный алгоритм Маркова заканчивает работу?

Задача 3.97. Постройте схему нормального алгоритма Маркова для вычисления функции $F(P) = P$, для любого слова P .

Задача 3.98. Постройте схему нормального алгоритма Маркова, неприменимого ни к одному входному слову.

3.11.9. Языки и грамматики

Тест 3.82. Дайте определение формальной грамматики.

Задача 3.99. Цепочки 01 и 000111 являются правильными словами языка. Постройте грамматику для этого языка.

Тест 3.83. Что такое *непосредственная выводимость* в порождающей грамматике?

Тест 3.84. Что такое *выводимость* в порождающей грамматике?

Тест 3.85. Что такое язык, порожденный грамматикой?

Тест 3.86. Какие ограничения накладываются на правила вывода контексто-зависимых грамматик?

Тест 3.87. Какие ограничения накладываются на правила вывода контексто-свободных грамматик?

Тест 3.88. Какие ограничения накладываются на правила вывода автоматных грамматик?

Тест 3.89. Что означает условие *однозначности ветвления по первому символу* для КС-грамматики?

Тест 3.90. Что такое *непродуктивный (непроизводящий)* нетерминальный символ грамматики? Как устраниТЬ непродуктивные символы из грамматики?

Тест 3.91. К какому типу по Хомскому относится данная грамматика? Какой язык она порождает?

1. $S \rightarrow bSS \mid a$. Ответ: $L(G) = \{\alpha \mid \alpha \in \{a, b\}^*\}, |a| = |b| + 1$, причем цепочки начинаются с терминала b и заканчиваются терминалом a .

2. $S \rightarrow Ab \mid a, A \rightarrow Ba, B \rightarrow cS$.
3. $S \rightarrow Ab, A \rightarrow Aa \mid ba$.
4. $S \rightarrow aSb, S \rightarrow \epsilon$.

Задача 3.100. Определить язык, порождаемый грамматикой G со следующими правилами:

$$S \rightarrow AaA \mid AaB, \quad A \rightarrow aA \mid \epsilon \quad B \rightarrow b$$

Задача 3.101. Построить КС-грамматику, порождающую язык

$$L = \{a^n b^n c^m \mid n, m \geq 1\}.$$

Задача 3.102. Построить КС-грамматику, порождающую слова из нулей и единиц с одинаковым числом тех и других.

Задача 3.103. Построить грамматику, порождающую язык регулярных выражений $0^*1(0+1)^*$.

3.11.10. Автоматы

Тест 3.92. Дайте определение автоматов Мили и Мура. В чем различие?

Тест 3.93. Как преобразовать автомат Мура в эквивалентный ему автомат Мили.

Тест 3.94. Дайте определение недетерминированного конечного автомата-распознавателя.

Тест 3.95. Опишите алгоритм построения недетерминированного конечного автомата по автоматной грамматике.

Задача 3.104. Постройте конечный автомат Мили, который для заданных двух целых двоичных чисел N_1 и N_2 одинаковой длины (меньшее может быть дополнено слева нулями) выдает $\max(X_1, X_2)$. Числа вводятся со старших разрядов, вход автомата – пары разрядов исходных чисел.

Задача 3.105. Приведите пример конечного автомата с ϵ -переходами. Для чего нужны такие автоматы?

Задача 3.106. Построить конечный автомат, распознающий язык L в алфавите $\{a, b, c\}$, где слово принадлежит языку L , если в нем не встречается буква a .

Задача 3.107. Построить конечный автомат по регулярному выражению $R = a^*b^*c^*$.

Задача 3.108. Построить конечные распознаватели для описанных ниже множеств цепочек из нулей и единиц:

- 1) число единиц четное, а нулей - нечетное;
- 2) между вхождениями единиц четное число нулей;
- 3) за каждым вхождением пары 11 следует 0;
- 4) каждый третий символ - единица;
- 5) имеется по крайней мере одна единица.

Задача 3.109. Построить конечный автомат с входным алфавитом $\{0, 1\}$, который допускает в точности такое множество цепочек:

- 1) все входные цепочки, начинающиеся с 0 и заканчивающиеся на 1;
- 2) все цепочки, не содержащие ни одной единицы;
- 3) все цепочки, содержащие в точности три единицы;
- 4) все цепочки, в которых перед и после каждой единицы стоит 0;
- 5) подцепочки 01, например 01, 11010, 1000111;
- 6) подцепочку 11;
- 7) подцепочку 100;
- 8) множество цепочек, которые начинаются или оканчиваются (или и то и другое) последовательностью 01;
- 9) множество всех цепочек, в которых всякая подцепочка из пяти последовательных символов содержит хотя бы два 0;
- 10) множество всех цепочек, которые начинаются с 11 и в цепочке нули стоят только по два;
- 11) множество всех цепочек, у которых между вхождениями единиц четное число нулей;
- 12) множество всех цепочек, у которых за каждым вхождением пары единиц следует нуль;
- 13) множество всех цепочек, у которых каждый пятый символ единица;
- 14) множество всех цепочек, у которых перед каждой единицей стоит нуль.

Задача 3.110. Построить конечный автомат с входным алфавитом $\Sigma = \{a, b, c\}$, допускающий:

- 1) все цепочки, в которых две последние буквы не совпадают;

- 2) все цепочки, начинающиеся и заканчивающиеся различными символами;
- 3) все цепочки, у которых последний символ цепочки уже появлялся раньше.

Задача 3.111. Построить конечный детеминированный автомат с входным алфавитом $\Sigma = \{a, b\}$, распознающий язык L :

- 1) в каждом слове которого сочетание ab встречается не более двух раз;
- 2) в каждом слове которого содержится подслово $bbcc$;
- 3) каждое слово которого имеет длину не более 8 и содержит одинаковое число букв a и b .

Задача 3.112. Построить конечный детеминированный автомат, распознающий числа, кратные 4.

Задача 3.113. Построить конечный детеминированный автомат для распознания цепочек в алфавите $\Sigma = \{a, b, c\}$:

- 1) которые содержат символ b только парами и заканчиваются на c ;
- 2) которые начинаются с b и в цепочке встречается только один раз ac .

Задача 3.114. Построить конечный детеминированный автомат для распознания цепочек в алфавите $\Sigma = \{0, 1, 2\}$, которые начинаются на 0 и в цепочке встречается ровно два символа 2.

Задача 3.115. Построить конечный автомат, распознающий язык

$$L = \{ucv \mid u \in \{a, b\}^*, v \in \{a, b\}^*\}.$$

Задача 3.116. Построить конечный автомат, который будет допускать только те цепочки, которые можно построить из подцепочек *go*, *goto*, *too*, *on*. Возможны повторения, но не пересечения.

Так, например, одна из допустимых цепочек *goongotoongotooon*. Можно сначала построить недетеминированный автомат, а затем преобразовать его в детеминированный.

Задача 3.117. Построить конечный автомат, распознающий зарезервированные слова языка C++:

- 1) *inline*, *int*, *if*;
- 2) *continue*, *class*, *const*;
- 3) *private*, *protected*, *public*;

- 4) *union, using, union;*
- 5) *double, delete, default;*
- 6) *virtual, void, volatile.*

Задача 3.118. Построить конечный недетерминированный автомат для языка над $\{0, 1, \dots, 9\}$ из слов, в которых крайняя справа цифра больше нигде в них не встречается (чтение происходит, как обычно, слева направо).

Задача 3.119. Для грамматик, заданных следующими правилами вывода построить эквивалентные им конечные автоматы:

- 1) $S \rightarrow 0A; A \rightarrow 0A|1S|0$
- 2) $S \rightarrow 0C|0D; C \rightarrow 0D|0S|1; D \rightarrow 1C|1S|0$
- 3) $S \rightarrow bS|aA|a|b; A \rightarrow aA|bS|a$
- 4) $S \rightarrow aA|bB|aC; A \rightarrow bA|bB|c; B \rightarrow aA|cC|b; C \rightarrow bB|bC|a$

Задача 3.120. Для недетерминированного конечного автомата A :

$$A = (Q, \Sigma, \delta, q_0, F), Q = \{q_0, q_1, q_2, q_3\}, \Sigma = \{7, 8, 9\}, F = \{q_3\},$$

$$\begin{cases} \delta(q_0, 7) = \{q_0, q_1\}, \\ \delta(q_1, 8) = \{q_1, q_2\}, \\ \delta(q_2, 9) = \{q_3\}, \\ \delta(q_3, 9) = \{q_3\} \end{cases}$$

построить эквивалентный ему детерминированный конечный автомат.

Задача 3.121. Дан детерминированный конечный автомат A :

$$A = (Q, \Sigma, \delta, q_0, F), Q = \{q_0, q_1, q_2, q_3\}, \Sigma = \{0, 1\}, F = \{q_3\},$$

$$\begin{cases} \delta(q_0, 1) = q_0, \\ \delta(q_0, 0) = q_1, \\ \delta(q_1, 1) = q_0, \\ \delta(q_1, 0) = q_2, \\ \delta(q_2, 1) = q_0, \\ \delta(q_2, 0) = q_3, \\ \delta(q_3, 1) = q_3, \\ \delta(q_3, 0) = q_3. \end{cases}$$

Применив теорему 3.48 о разрастании для автоматных языков построить представление слова $w = 0010001$ в виде $z = \alpha_1\alpha_2\alpha_3$.

Задача 3.122. Применив теорему 3.48 о разрастании для автоматных языков доказать нерегулярность следующих языков:

- 1) $\{0^n1^m2^n \mid n, m \in N\}$;
- 2) $\{0^n1^m \mid n \leq m, n, m \in N\}$;
- 3) $\{0^n1^{2n} \mid n \in N\}$;
- 4) множество цепочек из нулей и единиц, длины которых – полные квадраты;
- 5) множество цепочек из нулей и единиц вида ww , где некоторая цепочка w повторяется дважды;
- 6) множество цепочек из нулей и единиц вида $w1^n$, где w – цепочка из нулей и единиц длиной n ;
- 7) $\{0^n \mid n – полный квадрат\}$;
- 8) $\{0^n \mid n – полный куб\}$;
- 9) $\{0^n \mid n – степень числа 2\}$;
- 10) множество цепочек из нулей и единиц вида $w\bar{w}$, где цепочка \bar{w} образована из w путем замены всех нулей единицами и наоборот. Например, $\overline{011} = 100$, так что цепочка 011100 принадлежит данному языку;
- 11) $\{ww^R \mid w \in \{a, b\}\}$, где w^R – обратная цепочка к w (например, для $w = aba$, $w^R = aab$).

Задача 3.123. Доказать, что язык над алфавитом $\Sigma = \{a, b\}$ содержащий только слова, в котором букв a больше, чем букв b невозможно задать никаким конечным автоматом.

Задача 3.124.

1. По заданному регулярному выражению построить диаграмму переходов конечного автомата.
2. Если конечный автомат является недетерминированным, то преобразовать его в детерминированный.
3. Если конечный автомат не является минимальным, то минимизировать его.

1. $a^2(ab)^*c \mid d$
2. $d \mid a^2(ab)^*c$
3. $d \mid a^2(ab)^*c$

4. $(d | a)^2(ab)^*c$
5. $c^+((d | a)^2ab)^*$
6. $(a | b | c)^+abc(a | b | c)^*$
7. $(d | a)^*(ab)^*c^*$
8. $(d | a)^* | (ab)^* | c^*$
9. $(d | a)^*(ab)^* | c^*$
10. $(d | a^+)^*(ab)^* | c^*$
11. $(d | a^*)^*(ab)^* | c^*$
12. $a^*b^*c^* | a^+b^+c^+$

Задача 3.125. Минимизировать конечный детерминированный автомат, представленный на рис. 3.80.

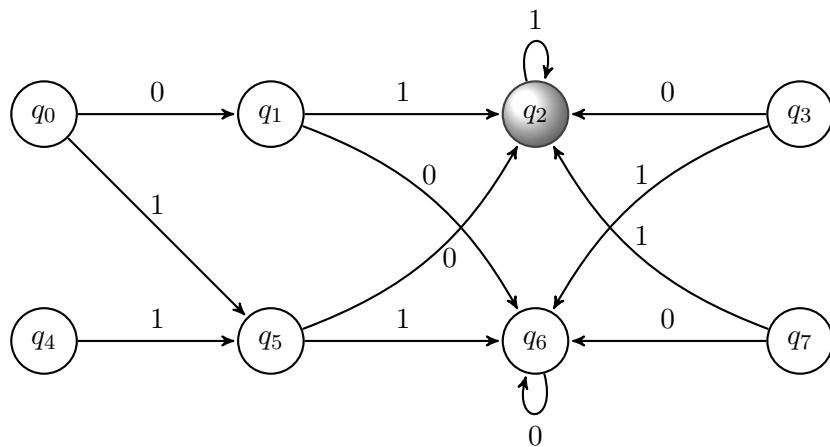


Рис. 3.80

Задачи для самостоятельного решения

1. Определите сколькими способами можно расставить скобки в выражении

$$X \rightarrow \neg Y \vee Z \vee Y \wedge Z,$$

чтобы получилась формула.

2. Формула логики высказываний F не содержит никаких логических связок, кроме \leftrightarrow . Докажите, что формула F является тавтологией тогда и только тогда, когда каждая переменная входит в F четное число раз.

3. Докажите, что если формулы $F \vee G$ и $\neg F \vee H$ тавтологии, то и формула $G \vee H$ — тавтология.

4. Является ли формула G логическим следствием формул F_1, F_2, F_3 :

$$F_1 = X \neg Y \vee Z, \quad F_2 = Y \rightarrow W, \quad F_3 = Z \rightarrow X, \quad G = X \rightarrow W.$$

5. Будет ли логичным следующее рассуждение:

Намеченная атака удастся, если захватить противника врасплох или его позиции плохо защищены. Захватить противника врасплох можно только, если он беспечен. Он не будет беспечен, если его позиции плохо защищены. Следовательно, намеченная атака не удастся.

6. Методом минимизирующих карт найдите минимальную ДНФ для формулы логики высказываний $F(X, Y, Z)$, принимающую значение 1 на наборах $(1, 1, 1), (1, 0, 1), (0, 0, 1)$ и $(0, 0, 0)$.

7. Найдите все существенные переменные следующих булевых функций:

- а) $f(x, y, z) = xy \vee \bar{y}z$;
- б) $f(x, y, z) = (x \rightarrow (y \rightarrow z)) \rightarrow ((x \rightarrow y) \rightarrow (x \rightarrow z))$.

8. Докажите, что класс булевых функций, сохраняющих ноль, но не сохраняющих единицу не является замкнутым.

9. Докажите, что множество булевых функций, удовлетворяющих условию

$$f(x_1, \dots, x_n) \geq x_n$$

образует замкнутый класс.

10. Докажите, что множество булевых функций вида $a_0(a_1 \vee x_1) \dots (a_n \vee x_n)$ образует замкнутый класс

11. Для булевой функции $f(x_1, x_2, x_3) = x_1x_2x_3 \vee x_2(\bar{x}_1 \vee x_3)$ постройте таблицу истинности, СДНФ и СКНФ.

12. Булеву функцию $f(x_1, x_2, x_3) = x_1\bar{x}_2 \vee \bar{x}_1x_3$ представьте в виде полиномом Жегалкина.

13. Сформулируйте алгоритм, проверяющий булеву функцию на монотонность, используя метод перебора с возвратом.

14. Выразите с помощью суперпозиции:

- а) \rightarrow через 1, \oplus и \wedge ;
- б) \wedge и \vee через \rightarrow , \neg ;
- в) \wedge, \vee и \neg через стрелку Пирса $f(x, y) = x \downarrow y = \overline{x \vee y}$.

15. Докажите, что нельзя выразить с помощью суперпозиции \rightarrow через \wedge и \vee .

16. Докажите полноту систем булевых функций:

- а) $\{x \rightarrow y, \overline{y \rightarrow x}\}$;
- б) $\{x \rightarrow y, 0\}$.

17. Приведите пример полной системы булевых функций, состоящий из одной функции n переменных ($n \geq 2$).

18. Запишите утверждение, состоящее в том, что d есть НОД чисел a и b в терминах предиката $D(x, y)$, означающего « x делит y » и предиката $G(x, y)$, означающего « $x \geq y$ ».

19. Определите выполнимы ли следующие формулы логики предикатов:
- $(\exists x)(\forall y) \left(G(x, y) \wedge \overline{H(x, y)} \right);$
 - $(\exists x)(F(x) \rightarrow F(y)).$
20. Определите общезначимы ли следующие формулы логики предикатов:
- $Q(x) \rightarrow (\forall y)Q(y);$
 - $(\forall x)Q(x) \rightarrow Q(y).$
21. Приведите к предваренной нормальной форме формулу логики предикатов:
- $$(\exists x)(\forall y)S(x, y) \rightarrow (\exists x)(\forall y)S(x, y).$$
22. Приведите к сколемовской нормальной форме формулу логики предикатов:
- $$(\exists x)(\forall y)R(x, y) \rightarrow (\forall x)(\exists y)R(x, y).$$
23. Докажите примитивную рекурсивность следующих функций:
- $f(x, y) = \max(x, y);$
 - $f(x, y) = \min(x, y);$
 - $f(x) = x!$
24. Докажите, что если функция $f(x_1, \dots, x_n)$ частично рекурсивна, то функция
- $$g(x_1, x_2, x_3, \dots, x_n) = f(x_2, x_1, x_3, \dots, x_n)$$
- также частично рекурсивна.
25. Постройте машины Тьюринга осуществляющие
- транспозицию: $q_1 1^{x+1} * 1^{y+1} \rightarrow q_0 1^{y+1} * 1^{x+1};$
 - циклический сдвиг: $q_1 1^{x_1+1} * \dots * 1^{x_n+1} \rightarrow q_0 1^{x_2+1} * \dots * 1^{x_n+1} * 1^{x_1+1}.$
26. Постройте машины Тьюринга правильно вычисляющие:
- проектирующую функцию $I_m^n(x_1, \dots, x_n) = x_m$, $1 \leq m \leq n;$
 - $f(x) = \text{sign}(x);$
 - $f(x_1, x_2) = x_1 x_2.$
27. Постройте нормальный алгорифм Маркова:
- для сложения двух чисел, записанных в унарной системе счисления;
 - для возведения числа в унарной записи в квадрат.
28. Сформулируйте грамматику постфиксной записи арифметических выражений содержащих, кроме бинарных операций (сложение, умножение и т. д.), также и унарные операции (не умоляя общности, достаточно рассмотреть одну — извлечение квадратного корня).
29. Опишите язык, который определяет КС-грамматика

$$S ::= S101 \mid 010.$$

Удовлетворяет ли она условию однозначности ветвлений?

30. Сформулируйте алгоритм вычисления значений арифметического выражения в постфиксной записи, если оно содержит как бинарные, так и унарные операции.
31. Сформулируйте алгоритм перевода арифметических выражений в скобочной инфиксной записи, содержащих как бинарные, так и унарные операции, в постфиксную форму.
32. Сформулируйте алгоритм перевода инфиксной записи ДНФ, не содержащей операций отрицания, в префиксную форму и докажите его корректность.
33. Сформулируйте алгоритм перевода инфиксной записи многочлена Жегалкина в постфиксную форму и докажите его корректность.
34. Сформулируйте алгоритм перевода выражения из скобочной инфиксной записи в префиксную.
35. Сформулируйте алгоритм вычисления значений арифметического выражения в префиксной записи.
36. Сколько различных алгебраически эквивалентных постфиксных записей соответствует выражению $a_1 + \dots + a_n$?
37. Докажите теорему о преобразовании арифметического выражения из инфиксной скобочной записи в постфиксную¹.
38. Сравните механизм сопоставления натуральных чисел символам постфиксной записи (который использовался в доказательстве теоремы об однозначности постфиксной записи) с изменением загрузки стека при вычислении значений выражения в постфиксной записи. Дайте содержательную трактовку сопоставляемым числам.
39. Найдите максимальный и минимальный размеры стека, необходимого для вычисления значения выражения с n знаками арифметических операций в постфиксной записи.
40. Найдите максимальный размер стека, необходимый для перевода выражения с n знаками арифметических операций из инфиксной скобочной записи в постфиксную.

¹Задача повышенного уровня сложности.

Список литературы

1. *J. H. van Lint.* A Course in Combinatorics. /J. H. van Lint, R. M. Wilson. – Cambridge. University Press, 2001.
2. *M. A. Айзerman, Л. А. Гусев, Л. И. Розонэр, И. М. Смирнова, А. А Таль.* Логика. Автоматы. Алгоритмы. Гос. изд. физ.-мат. литературы 1963.
3. *K. Айерлэнд, М. Роузен.* Классическое введение в современную теорию чисел. – М.: Мир, 1987.
4. *Г. Биркгоф.* Теория решеток. – М.: Наука, 1984.
5. *P. Блейхут.* Теория и практика кодов, исправляющих ошибки. – М.: Мир, 1986.
6. *B. Брауэр.* Введение в теорию конечных автоматов. – М.: Радио и связь, 1987.
7. *Дж. Бунос, Р. Джесеффри.* Вычислимость и логика. – М.: Мир, 1994.
8. *A. A. Бухштаб.* Теория чисел. – М.: Просвещение, 1966.
9. Введение в криптографию. /Под ред. В. В. Ященко. – М.: МЦНМО, 1998.
10. *И. М. Виноградов.* Основы теории чисел. – М.: Наука, 1972.
11. *A. Гилл.* Введение в теорию конечных автоматов. – М.: Наука, 1966.
12. *B. A. Горбатов, A. B. Горбатов, M. B Горбатова.* Теория автоматов. – М.: Высшая школа, 2008.
13. *P. Грэхем, Д. Кнут, О. Паташиник.* Конкретная математика. – М.: Мир, 1998.
14. *Л. А. Заде.* Понятие лингвистической переменной и его применение к принятию приближенных решений. – М.: Мир, 1976.

15. Ю. Г. Карпов. Теория автоматов. – СПб.: Питер, 2003.
16. В. Н. Касьянов. Лекции по теории формальных языков, автоматов и сложности вычислений. – М.: Вильямс, 2002.
17. С. Клини. Математическая логика. – М.: Мир, 1978.
18. У. Клоксин. Программирование на языке Пролог. /У. Клоксин, К. Меллиш. – М.: Мир, 1987.
19. Д. Кнут. Искусство программирования для ЭВМ. Получисленные алгоритмы.: в 3 т. Т.2 – М.: Мир, 1977.
20. Л. К. Конышева, Д. М. Назаров. Основы теории нечетких множеств. – СПб: Питер, 2011.
21. А Кофман. Введение в теорию нечетких множеств. – М.: Радио и связь, 1982.
22. Н. Кристофидес. Теория графов. Алгоритмический подход. – М.: Мир, 1997.
23. В. Б. Кудрявцев, С. В. Алешин, А. С. Подколзин. Введение в теорию автоматов. – М.: Наука, 1985.
24. О. П. Кузнецов, Г. М. Адельсон-Вельский. Дискретная математика для инженера. – М.: Энергоатомиздат, 1988.
25. И. А. Лавров, Л. Л. Максимова. Задачи по теории множеств, математической логике и теории алгоритмов. – М.: Наука, 1984.
26. А. В. Леоненков. Нечеткое моделирование в среде MATLAB и fuzzyTECH. – СПб.: БХВ-Петербург, 2005.
27. В. Липский. Комбинаторика для программистов. – М.: Мир, 1988.
28. Л. М. Лихтарников, Т. Г. Сукачева. Математическая логика. – М.: Лань, 1999
29. А. И. Мальцев. Алгоритмы и рекурсивные функции. – М.: Наука, 1965.
30. А. А. Марков, Н. М. Нагорный. Теория алгорифмов. – М.: Наука, 1984.
31. А. Н. Мелихов. Ориентированные графы и конечные автоматы. – М.: Наука, 1971.

32. Э. Мендельсон. Введение в математическую логику. – М.: Наука, 1984.
33. М. Минский. Вычисления и автоматы. - М.: Мир, 1971.
34. А. В. Могиленко, А. В. Балуев. Элементарные понятия теории нечетких множеств. – Новосибирск, 2003.
35. М. Мозговой. Классика программирования: алгоритмы, языки, автоматы, компиляторы. Практический подход. – М.: Наука и техника, 2006.
36. В. Н. Нефедов, В. А. Осипова. Курс дискретной математики. – М.: Изд-во МАИ, 1992.
37. В. Новак, И. Перфильева, И. Мочкорж. Математические принципы нечеткой логики. – М.: ФИЗМАТЛИТ, 2006.
38. Ф. А. Новиков. Дискретная математика для программистов. – СПб.: Питер, 2005.
39. А. Е. Пентус, М. Р. Пентус. Математическая теория формальных языков. – М.: БИНОМ, 2006
40. Н. И. Поликарпова, А. А. Шалыто. Автоматное программирование. СПб.: Питер, 2008.
41. И. В. Романовский. Дискретный анализ. – СПб.: Невский диалект, 1999.
42. А. П. Рыжов. Элементы теории нечетких множеств и ее приложений. – М.: Диалог, 2003.
43. S. Martelo, P. Toth. Knapsack?problems. – Great Britain: Wiley, 1990.
44. С. Д. Штөвба. Введение в теорию нечетких множеств и нечеткую логику. – Винница: Континент-Прим. – 2003.
45. Р. М. Смаллиан. «Принцесса или тигр?» – М.: Мир, 1985.
46. В. П. Супрун. Табличный метод полиномиального разложения булевых функций // Кибернетика. – 1987. – № 1. – С.116-117.
47. В. П. Супрун. Основы теории булевых функций. – М.: Ленанд / URSS. – 2017. – 208 с.
48. Р. Стенли. Перечислительная комбинаторика. – М.: Мир, 1990.

49. *Б. А. Трахтенброт, Я. М. Барздинъ.* Конечные автоматы (поведение и синтез). – М.: Наука, 1970.
50. *P. Уилсон.* Введение в теорию графов. – М.: Мир, 1977.
51. *T. Фудзисава, T. Касами.* Математика для радиоинженеров. Теория дискретных структур. – М.: Радио и связь, 1984.
52. *M. Холл.* Комбинаторика. – М.: Мир, 1970.
53. *H. Хомский.* Синтаксические структуры. //В кн.: Новое в лингвистике. Вып. II. М., 1962.
54. *Джон Хопкрофт, Раджисв Мотвани, Джесеффри Ульман.* Введение в теорию автоматов, языков и вычислений. Второе издание. – М.: Вильямс, 2015.
55. *Ч. Ченъ, Р. Ли.* Математическая логика и автоматическое доказательство теорем. – М.: Мир, 1983.
56. *A. Черч.* Введение в математическую логику. – М.: изд-во ин.лит., 1960.
57. *Г. Эндрюс.* Теория разбиений. – М.: Наука, 1982.
58. *C. В. Яблонский.* Введение в дискретную математику. – М.: Наука, 1997.
59. *Amos Omondi, Benjamin Premkumar.* Residue Number Systems: Theory and Implementation, 2007.
60. *D.D. Aufenkamp, F.E. Hohn.* Analysis of Sequential Machines. IRE Transactions on Electronic Computers. Volume: EC-6, Issue: 4, Dec. 1957, 276–285.

Оглавление

Предисловие	3
1. Целочисленные алгоритмы	5
1.1. Арифметика целых чисел	5
1.1.1. Деление с остатком	5
1.1.2. Наибольший общий делитель, наименьшее общее кратное и их свойства	6
1.1.3. Простые числа	9
1.1.4. Решето Эратосфена. Разложение числа на простые множители	10
1.1.5. Позиционная запись натуральных чисел	18
1.1.6. Алгоритмы арифметических действий с p -ичными записями натуральных чисел	19
1.1.7. Алгоритмы перевода p -ичной записи натурального числа в q -ичную	21
1.1.8. Алгоритм эффективного возведения числа в натуральную степень	22
1.2. Алгоритм Евклида и цепные дроби	24
1.2.1. Классический алгоритм Евклида	24
1.2.2. Бинарный алгоритм Евклида	25
1.2.3. Линейное представление наибольшего общего делителя	27
1.2.4. Обобщенный алгоритм Евклида	27
1.2.5. Разложение числа в цепную дробь	30
1.2.6. Свойства подходящих дробей и их вычисление	32
1.2.7. Цепные дроби и анализ алгоритма Евклида	33
1.2.8. Бесконечная цепная дробь и ее вычисление	37
1.2.9. Наилучшие приближения	44
1.2.10. Некоторые применения цепных дробей	45
1.2.11. Диофантовы уравнения	47
1.2.12. Числа Фибоначчи	51
1.3. Арифметика остатков	54
1.3.1. Алгебраические структуры	54
1.3.2. Арифметика и свойства сравнений	56

1.3.3.	Функция Эйлера и ее свойства	61
1.3.4.	Линейные сравнения	67
1.3.5.	Китайская теорема об остатках	71
1.3.6.	Система остаточных классов	75
1.3.7.	Применение теоремы Эйлера в криптографии. Система шифрования RSA	76
1.3.8.	Простейшие атаки на систему RSA	81
1.3.9.	Другие примеры использования RSA	83
1.4.	Арифметика многочленов	86
1.4.1.	Основные операции и свойства	86
1.4.2.	Схема Горнера	93
1.4.3.	Алгоритм Евклида для многочленов. Линейное представление НОД	95
1.4.4.	Интерполяционная формула Лагранжа	97
1.4.5.	Разложение многочлена на свободные от квадратов множители	99
1.4.6.	Разложение многочленов на неприводимые множители	100
1.5.	Базисы Гребнера	104
1.5.1.	Задачи на доказательство и многочлены	104
1.5.2.	Случай многочленов одной переменной	106
1.5.3.	Понятие идеала	107
1.5.4.	Линейные многочлены многих переменных. Понятие базиса Гребнера	109
1.5.5.	Алгоритм Бухбергера	111
1.6.	Теория множеств и комбинаторика	120
1.6.1.	Вводные задачи	121
1.6.2.	Размещения и сочетания	130
1.6.3.	Биномиальные коэффициенты	132
1.6.4.	Кодирование с исправлением ошибок. Граница Хемминга	135
1.6.5.	Полиномиальное кодирование	137
1.6.6.	Код Шеннона–Фэно и алгоритм Хаффмена	143
1.6.7.	Лексикографический порядок. Генерирование k -элементных подмножеств	145
1.6.8.	Перечислительная комбинаторика	147
1.6.9.	Бинарный код Грэя	158
1.6.10.	Числа Стирлинга первого и второго родов	160
1.6.11.	Числа Каталана	163
1.6.12.	Разбиения чисел	168
1.6.13.	Введение в динамическое программирование. Задача о рюкзаке	172
1.6.14.	Принцип включения–исключения	174

1.6.15. Обращение Мёбиуса	179
1.6.16. Понятие о группе: теорема Бернсайда	183
1.7. Производящие функции и рекуррентные уравнения	195
1.7.1. Производящие функции	195
1.7.2. Решение однородного линейного рекуррентного уравнения	201
1.7.3. Решение неоднородного линейного рекуррентного уравнения	207
1.7.4. Производящая функция для чисел Каталана	209
1.7.5. Число многочленов заданной степени, не приводимых над полем вычетов	211
1.8. Элементы дискретной теории вероятностей	215
1.8.1. Основные определения	215
1.8.2. Условные вероятности и формула Байеса	217
1.8.3. Дискретные случайные величины. Математическое ожидание и дисперсия	222
Задачи для самостоятельного решения	225
2. Графы и бинарные отношения	229
2.1. Определения графов	229
2.1.1. Простейшие определения и свойства	229
2.1.2. Машинаное представление графов	234
2.1.3. Поиск в глубину и ширину в графе	237
2.1.4. Связность	241
2.1.5. Эйлеров путь в графе	248
2.1.6. Графы де Брёюна	256
2.1.7. Деревья	260
2.1.8. Корневые деревья	263
2.1.9. Код Прюфера	266
2.1.10. Главные циклы и коциклы	268
2.1.11. Двудольные графы	274
2.1.12. Планарность	276
2.1.13. Раскраска графов	280
2.2. Алгоритмы на графах	286
2.2.1. Построение наибольшего паросочетания	286
2.2.2. Построение минимальных стягивающих деревьев	291
2.2.3. Задача нахождения кратчайших путей в графе	297
2.2.4. Алгоритм Форда–Беллмана	299
2.2.5. Алгоритм Форда–Беллмана для нахождения дерева кратчайших путей	302
2.2.6. Алгоритм Дейкстры	307

2.2.7. Сравнительный анализ трудоемкости алгоритмов Форда–Беллмана и Дейкстры	310
2.2.8. Алгоритм Флойда	311
2.2.9. Волновой алгоритм	316
2.2.10. Алгоритм Форда–Фалкерсона	318
2.2.11. Перебор с возвратом	326
2.2.12. Метод ветвей и границ	330
2.3. Бинарные отношения	333
2.3.1. Введение. Постановка задачи	333
2.3.2. Свойства бинарных отношений	334
2.3.3. Способы задания отношений	337
2.3.4. Отношения эквивалентности и толерантности	339
2.3.5. Отношения предпорядка и порядка	341
2.3.6. Алгоритм топологической сортировки	343
2.3.7. Частично упорядоченные множества. Диаграмма Хассе	345
2.3.8. Транзитивное замыкание бинарного отношения. Алго- ритм Уоршелла	346
2.3.9. Примеры и задачи на бинарные отношения	351
Задачи для самостоятельного решения	353
3. Математическая логика и теория алгоритмов	358
3.1. Логика высказываний	358
3.1.1. Основные понятия логики высказываний	358
3.1.2. Равносильность формул логики высказываний	364
3.1.3. Принцип двойственности	367
3.1.4. Логическое следствие	369
3.1.5. Нормальные формы в логике высказываний	372
3.1.6. Контактные схемы	380
3.1.7. Метод минимизирующих карт	383
3.1.8. Метод Куайна – Мак-Класки	387
3.1.9. Метод резолюций в логике высказываний	391
3.1.10. Стратегии метода резолюций	399
3.1.11. Линейная резолюция	401
3.1.12. Решетки, булевы алгебры, кольца	403
3.2. Булевы функции	408
3.2.1. Основные понятия	408
3.2.2. Замкнутость и полнота	412
3.2.3. Полиномы Жегалкина и линейные функции	415
3.2.4. Быстрое вычисление полинома Жегалкина	421
3.2.5. Булевы функции в криптографии	423
3.2.6. Самодвойственные функции	424
3.2.7. Монотонные функции	428

3.2.8. Теорема Поста	432
3.2.9. Разложение Шеннона	438
3.2.10. Бинарные диаграммы решений	440
3.3. Логика предикатов	445
3.3.1. Основные определения	446
3.3.2. Семантика логики предикатов	449
3.3.3. Примеры и задачи на выразительность логики предикатов	451
3.3.4. Равносильность формул логики предикатов	452
3.3.5. Логическое следствие	456
3.3.6. Нормальные формы	457
3.3.7. Примеры и задачи на приведение к ПНФ и СНФ	460
3.3.8. Метод резолюций в логике предикатов	462
3.3.9. Метод резолюций и ПРОЛОГ	471
3.4. Нечеткие множества	474
3.4.1. Введение	474
3.4.2. Основные понятия	477
3.4.3. Операции над нечеткими множествами	481
3.4.4. Мера нечеткости множеств	486
3.4.5. Функции принадлежности	488
3.4.6. Методы построения функций принадлежности	492
3.4.7. Задачи для самостоятельного решения	493
3.5. Нечеткие бинарные отношения	495
3.5.1. Основные понятия	495
3.5.2. Свойства нечетких бинарных отношений	497
3.5.3. Задачи для самостоятельного решения	500
3.6. Нечеткая логика	500
3.6.1. Нечеткие высказывания	500
3.6.2. Нечеткие формулы логики высказываний	503
3.7. Теория алгоритмов	505
3.7.1. Машина Тьюринга и функции, вычислимые по Тьюрингу	505
3.7.2. Примеры и задачи на построение машин Тьюринга	511
3.7.3. Операции над машинами Тьюринга	514
3.7.4. Проблема останова машины Тьюринга	516
3.7.5. Частично рекурсивные функции	519
3.7.6. Нормальные алгорифмы Маркова	521
3.7.7. Примеры и задачи на нормальные алгорифмы Маркова	524
3.8. Элементы теории сложности алгоритмов	527
3.8.1. Вычислительная сложность алгоритма	527
3.8.2. Сложностные классы задач	529
3.9. Формальные языки и грамматики	531

3.9.1. Основные определения порождающих грамматик	532
3.9.2. Классификация грамматик	533
3.9.3. Контекстно-свободные грамматики	536
3.9.4. Преобразования контекстно-свободных грамматик	542
3.9.5. Примеры и задачи на построение грамматик	544
3.9.6. Префиксная и постфиксная записи формул	545
3.9.7. Алгоритмы обработки скобочных записей	549
3.10. Конечные автоматы распознаватели	553
3.10.1. Введение	553
3.10.2. Автоматы Мили и Мура	554
3.10.3. Примеры автоматов Мили и Мура	557
3.10.4. Эквивалентность автоматов Мили и Мура	558
3.10.5. Основные понятия автоматов распознавателей	561
3.10.6. Примеры автоматов распознавателей	562
3.10.7. Недетерминированные автоматы распознаватели. Детерминизация	564
3.10.8. Теорема о разрастании для автоматных языков	568
3.10.9. Автоматы и автоматные грамматики	571
3.10.10. Минимизация автоматов распознавателей	574
3.10.11. Конечные автоматы с эпилон переходами	578
3.10.12. Конечные автоматы и регулярные выражения	580
3.11. Тесты и задачи для экзамена	585
3.11.1. Целочисленные алгоритмы	585
3.11.2. Комбинаторика, кодирование	586
3.11.3. Бинарные отношения	586
3.11.4. Графы	587
3.11.5. Логика высказываний, булевы функции	589
3.11.6. Логика предикатов	592
3.11.7. Нечеткие множества	592
3.11.8. Машина Тьюринга, алгоритмы Маркова	593
3.11.9. Языки и грамматики	594
3.11.10. Автоматы	595
Задачи для самостоятельного решения	600
Список литературы	604
Предметный указатель	614

Предметный указатель

А

Автомат

- Детерминизация 564
- Мили
 - определение 554
 - примеры 556
 - способы задания 554
- Мура
 - определение 556
 - примеры 556
 - способы задания 556
- Регулярные выражения 580
- автоматные грамматики 571
- алгоритм построения для регулярного выражения 582
- детеминированный
 - определение 561
 - детерминизация вытягиванием 566
 - детерминизация объединением 565
 - лемма о накачке 568
 - минимизация 574
 - невозвратное состояние 562
 - распознаватель
 - определение 561
 - примеры 562
 - редукция 574
 - с эпсилон переходами 578
- Алгебраическая система 449
- Алгоритм
 - Бухбергера 116
 - Дейкстры 307

— Евклида

- — анализ 33–37
- — бинарный 26
- — для многочленов 95
- — классический 25
- — обобщенный 29
- Задача о неограниченном рюкзаке 172
- Задача о рюкзаке 0 – 1 173
- Косарайю и Шарира 246
- Краскала 294
- — модифицированный 296
- Куайна – Мак-Класки 387
- Ли 316
- Прима 292
- Уоршелла 348
- Ферма 14
- Флери 252
- Флойда 312
- Форда–Беллмана 300
- — нахождения дерева кратчайших путей 303
- Форда–Фалкерсона 324
- Хаффмена 143
- Эратосфена 10
- ближайшего соседа 292
- возвведения числа в степень 22
- волновой 316
- вычисления значений в постфиксной форме 550
- вычисления квадратного корня 17
- генерации

- — подмножеств 146
- — разбиений числа 171
- деления многочленов 88, 92
- детерминизации вытягиванием 566
- детерминизации объединением 565
- для метода ветвей и границ 330
- жадный 145
- метода резолюций 395, 467
- минимизирующих карт 384
- нахождения максимального потока в плоской транспортной сети 325
- нахождения мостов графа 243
- нахождения недостижимых состояний 565
- нахождения эйлеровой цепи 249
- перебора с возвратом 329
- перевода из p -ичной записи в q -ичную в p -ичной арифметике 21
- перевода из p -ичной записи в q -ичную в q -ичной арифметике 22
- перевода инфиксной в постфиксную форму 551
- перемножения чисел в p -ичной системе счисления 21
- поиска
 - в глубину 237
 - в ширину 239
- поиска компонент связности в графе 241
- поиска компонент сильной связности в графе 246
- полиномиального декодирования 141
- последовательной раскраски 283
- построение кода Прюфера 266
- построения
 - каркаса 262
- — максимального паросочетания 289
- построения автомата для регулярного выражения 582
- построения графа де Брюина 257
- построения дерева по коду Прюфера 267
- построения корневого дерева 264
- построения минимального автомата 575
- построения полинома Жегалкина
- — метод неопределенных коэффициентов 417
- — метод равносильных преобразований 416
- — метод треугольника Паскаля 418
- приведения
 - к дизъюнктивной нормальной форме 373
 - к предваренной нормальной форме 458
 - к совершенной дизъюнктивной нормальной форме 374, 376
 - к совершенной конъюнктивной нормальной форме 378, 379
- пробных делителей 12
- разбиения на классы эквивалентности 340
- разложения многочлена на свободные от квадратов множители 100
- распознавания двудольности графа 275
- решения
 - диофантова уравнения 48
 - линейного сравнения 68

- сдвига числа в p -ичной системе счисления 20
- сколемизации 459
- сложения в p -ичной системе счисления 19
- сложения многочленов 87
- топологической сортировки 343
- удаления непродуктивных терминалов 542
- умножения многочленов 87
- умножения на цифру в p -ичной системе счисления 20
- унификации 464
- шифрования RSA 77
- Алгоритм RSA**
 - затемняющий множитель 81, 85
 - ослепление 81
 - схема с нотариусом 81
 - циклическая атака 82
- Алгорифм Маркова** 521
 - вычислимая функция 522
 - подстановка 522
 - схема 521
 - формула подстановки 521
- Алфавит** 256

- Б**
- Базис**
 - Бандлера-Кохоута 501
 - Вебера 501
 - Гребнера 115
 - Лукашевича –Гилеса 501
 - замкнутого класса 435
 - идеала 107
- Байеса**
 - Формула 219
- Биграф** 274
- Бинарное отношение** 333
 - Парето 341
 - антирефлексивное 335
 - антисимметричное 335
 - асимметричное 335
 - достижимости 339
 - замыкание 347
 - иррефлексивное 335
 - порядка 341
 - в решетке 403
 - — линейного 341
 - — нестрогого 341
 - — строгого 341
 - — частичного 341
 - — предпорядка 341
 - — рефлексивное 334
 - — симметричное 335
 - — согласованное 342
 - — толерантности 340
 - — транзитивное 336
 - — транзитивное замыкание 347
 - — эквивалентности 339, 365
 - Бинарные диаграммы решений 440
 - упорядоченные 440
 - Бинарные отношения
 - нечеткие 495
 - — антирефлексивность 498
 - — антисимметричность 498
 - — асимметричность 498
 - — граф 496
 - — котранзитивность 499
 - — матрица 495
 - — рефлексивность 497
 - — свойства 497
 - — сильная антирефлексивность 498
 - — сильная полнота 499
 - — сильная рефлексивность 498
 - — симметричность 498
 - — слабая антирефлексивность 498
 - — слабая полнота 499
 - — слабая рефлексивность 498
 - — способы задания 495
 - — транзитивность 499
 - Бином Ньютона 127, 132

Божественная пропорция 53, 200
Буква 256
Булеан 404
Булева алгебра 405
— Кантора 405
— Линденбаума-Тарского 405
— двойственная 407
Бэктрекинг (перебор с возвратом) 328

В

Валентность 230
Вероятность
— независимые события 216
— определение 215
— полная система событий 216
— условная 217
Вершина 229
— висячая 230
— изолированная 230
— источник 319
— концевая 229
— корень дерева 263
— лист дерева 263
— множество смежности 229
— насыщенная в паросочетании 286
— разделяющая 242
— свободная в паросочетании 286
— сток 319
— точка сочленения 242
— шарнир 242
Ветвь 269
Вхождение переменной
— свободное 449
— связное 449
Выход 467
— резолютивный 393
Высказывание 358, 446
Высказывания
— нечеткие 500
— — логические операции 500

— — логические формулы 503
— — степень равносильности формул 503

Г

Гипотеза
— Кармайкла 64
Гомеоморфизм графов 280
Грамматика 532
— автоматная 535
— алфавит нетерминальных символов 532
— алфавит терминальных символов 532
— вывод 533
— грамматический разбор 538
— классификация Хомского 533
— контекстно-зависимая 534
— контекстно-свободная 534, 536
— — непродуктивные нетерминалы 542
— начальный символ 532
— правила вывода 532
— регулярная 535
— свободная 534
— синтаксический анализатор 538
— — однозначность ветвления 538
— — синтаксическая диаграмма 539
— язык 533
— язык Дика 534
Граница 272
— Хемминга 136
Грань 278
Граф 233
— k -раскрашиваемый 280
— Герца 244
— Дирака 256
— Ферре 168
— ациклический 260
— бикомпонента 244
— бинарного отношения 337

- взвешенный 291
 - гамильтонов 255
 - двудольный 274
 - де Брюина 256
 - дерево 260
 - код Прюфера 266
 - корневое 263
 - каркас 261
 - компоненты реберной двусвязности 242
 - компоненты связности 241
 - конденсации 244
 - коранг 268
 - кубический 230
 - лес 260
 - нагруженный 291
 - неориентированный 229
 - ориентированный 232
 - планарный 277
 - плоский 277
 - полный 229
 - полугамильтонов 255
 - полуэйлеров 248
 - реберно отделимый 242
 - реберно связный 242
 - регулярный 230
 - связности булевой функции 412
 - связный 241
 - сильная компонента 244
 - сильносвязный 244
 - слабосвязный 244
 - транспонированный 245
 - цикломатическое число 268
 - четный 274
 - эйлеров 248
 - Группа 54
 - абелева 55
 - классы смежности 186
 - коммутативная 55
 - подгруппа 186
 - подстановок 185
 - симметрий тетраэдра 190
 - система образующих 188
- Д**
- Делитель
 - наибольший общий 6
 - — линейное представление 27
 - нуля 55
 - Деньги электронные 83
 - Дерево 260
 - каркас 261
 - корень 263
 - корневое 263
 - алгоритм построения 264
 - — потомок 263
 - — предок 263
 - минимальное оствовное 291
 - ориентированное 263
 - оствов 261
 - оствовое 261
 - покрывающее граф 261
 - стягивающе 261
 - Джон фон Нейман 554
 - Диаграмма
 - Ферре 168
 - Хассе 345
 - Юнга 168
 - Дивергенция потока 320
 - Дизъюнкт 377
 - хорновский 471
 - Дизъюнкция 359
 - элементарная 377
 - конституента нуля 379
 - Динамическое программирование
 - Задача о рюкзаке 172
 - — Неограниченый рюкзак 172
 - — Рюкзак 0 – 1 173
 - Дискретная случайная величина
 - дисперсия 223
 - закон распределения 222
 - математическое ожидание 222
 - независимость 222
 - определение 222

- Дополнение 405
- по Сугено 501
- по Ягеру 501
- Дробь
 - непрерывная 30
 - подходящая 32
 - цепная 30
 - бесконечная 40
 - — наилучшие приближения 44
 - — применения 45–46
- Дуга 232

З

- Задача
 - Кислера 358, 361, 363
 - Монмора 178
 - коммивояжера 332
 - о Ханойских башнях 159
 - о встречах 178
 - о домино 254
 - о заполнении домино 195
 - о кенигсбергских мостах 233, 252
 - о путешественнике 364
 - о разбиении числа 168
 - о размене монет 122
 - о рыцарях и лжецах 372
 - о футбольной команде 125
 - о числе беспорядков 178
 - о числе перестановок 124
 - построения генома 259
 - расстановки ферзей 326
- Закон логики высказываний 366
- ассоциативности 366
- двойного отрицания 367
- де Моргана 367
- идемпотентности 366
- исключенного третьего 367
- коммутативности 366
- контрапозиции 367
- противоречия 367
- Замыкание

- булевых функций 413
- транзитивное 347, 533
- Запись
 - инфиксная 409
 - польская
 - — обратная 549
 - — прямая 549
 - постфиксная 545
 - префиксная 409, 546
- Звезда 230, 232
- Золотое сечение 53, 200

И

- Идеал 108
- мономиальный 116
- Изоморфизм
 - булевых алгебр 406
 - графов 276
- Импликация 359
- Бандлера-Кохоута 502
- Вади 502
- Геделя 502
- Лукашевича 502
- Лукашевича-Гилеса 502
- Мамдани 502
- Интерпретация формулы
 - логики высказываний 362
 - логики предикатов 449
- Иrrациональность квадратичная 41

К

- Квадрат Дюрфи 168
- Квантификация переменной 447
- Класс
 - булевых функций
 - — Поста 432
 - — замкнутый 413
 - — линейных 419
 - — монотонных 428
 - — полный 414
 - — предполный 434
 - — самодвойственных 426

- — сохраняющих ноль 413
- — сохраняющих единицу 414
- вычетов 56, 58, 78
- грамматик 533, 536
- частично рекурсивных функций 519
- эквивалентности 339
- Кограница** 272
- Код**
 - Грея 158
 - Шеннона-Фэно 143
 - беспрефиксный 143
 - линейный 137
 - циклический 138
 - плотно-упакованный 137
 - совершенный 137
 - с проверкой на четность 142
- Кодовое расстояние**
 - Хемминга 135
 - минимальное 137
- Кольцо** 55
 - ассоциативное 55
 - булево 407
 - вычетов 58
 - коммутативное 55
 - многочленов 88
 - с единицей 55
- Конечный автомат**
 - Детерминизация 564
 - Мили
 - определение 554
 - примеры 556
 - способы задания 554
 - Мура
 - определение 556
 - примеры 556
 - способы задания 556
 - Регулярные выражения 580
 - автоматные грамматики 571
 - детеминированный
 - определение 561
 - с эпсилон переходами 578
- Константа**
 - предметная 448
- Контакт** 380
- Контактная схема** 380
 - замкнутая 380
 - разомкнутая 380
- Континуанта** 33
- Конъюнкция** 359
 - элементарная 372
 - — конституента единицы 376
- Коцикл** 271
 - главный 272
- Коэффициент**
 - биномиальный 66, 132
- Коэффициенты Безу** 27
- Кратное наименьшее общее 6**
- Критерий**
 - Поста 432
 - Эйзенштейна 102
- Л**
- Лемма**
 - Огдена 568
 - Эйлера 34
 - о моделировании 414
 - о накачке 568
 - о нелинейной функции 420
 - о немонотонной функции 431
 - о несамодвойственной функции 426
 - о переименовании связанных переменных 455
 - о подстановке 367
 - о рукопожатиях 230
- Лес** 260
- Литерал** 372, 458
 - контрарный 392
- М**
- Матрица**
 - весов 291
 - инциденций 234
 - отношения 337

- смежности 235
- Машина Тьюринга 506
 - активная зона 508
 - геделева нумерация 517
 - команда 506
 - композиция 514
 - конфигурация 507
 - правильно вычисляющая функцию 509
 - проблема останова 516
 - программа 507
 - разветвление 515
 - реализация цикла 515
 - функциональная схема 507
- Мера
 - Егера 486
 - Коско 486
- Метод
 - гарвадский 383
 - минимизирующих карт 383
 - неопределенных коэффициентов 383
 - разбиения 575
 - раскраски 340
- Мили
 - автомат
 - — определение 554
 - — примеры 556
 - — способы задания 554
- Минимизация автоматов распознавателей 574
- Многочлен
 - Лорана 107
 - Чебышева 203
 - зацепление 115
 - многих переменных 107
 - — линейный 109
 - мультииндекс 107
 - над полем 86
 - неприводимый 99
 - одночлен 107
 - ошибок 139
- порождающий 138
- примитивный 99
- разложение на свободные от квадратов множители 99
- редукция 112
- свободный от квадратов 99
- синдромный 139
- степень 86
- факториальный 162
- характеристический 201
- Множества
 - логические операции
 - — Дизъюнктивная сумма 484
 - — контрастная интенсификация 482
 - — концентрирование 482
 - — размывание 482
 - — функция принадлежности
 - — *S*-образная функция 490
 - — Метод прямого оценивания 492
 - — Методы построения 492
 - — гауссова 489
 - — колоколообразная 489
 - — трапецеидальная 488
 - Множество
 - линейно упорядоченное 345
 - предметное 446
 - унифицируемое 464
 - частично упорядоченное 345
- Модель 449
- Модулярная арифметика 75
- Моном 114
- Мост 242
- Мультиграф 229
- Мура
 - автомат
 - — определение 556
 - — примеры 556
 - — способы задания 556

Н

Нечеткие

- бинарные отношения 495
- — антирефлексивность 498
- — антисимметричность 498
- — асимметричность 498
- — граф 496
- — котранзитивность 499
- — матрица 495
- — рефлексивность 497
- — свойства 497
- — сильная антирефлексивность 498
- — сильная полнота 499
- — сильная рефлексивность 498
- — симметричность 498
- — слабая антирефлексивность 498
- — слабая полнота 499
- — слабая рефлексивность 498
- — способы задания 495
- — транзитивность 499
- — высказывания 500
- — логические операции 500
- — логические формулы 503
- — степень равносильности формул 503

Нечеткие множества 477

- α -уровня 478
- Ближайшее четкое 478
- Выпуклая комбинация 485
- Кардинальное число 478
- Мощность 478
- Ядро 478
- высота 478
- декомпозиция 480
- логические операции 481
- — Дизъюнктивная сумма 484
- — контрастная интенсификация 482
- — концентрирование 482
- — размывание 482

— мера Егера 486

- мера Коско 486
- мера нечеткости 486
- нормальное множество 478
- носитель 478
- оператор увеличения нечеткости 485
- операции 481
- отношения включения 481
- субнормальное множество 478
- точка перехода 478
- умножение на число 485
- функция принадлежности 478, 488
- — S -образная функция 490
- — Метод попарных сравнений 492
- — Метод прямого оценивания 492
- — Методы построения 492
- — Частотный метод 492
- — Шкала Саати 493
- — гауссова 489
- — колоколообразная 489
- — сплайн-функция 490
- — трапецидальная 488

О

Область

- декодирования 136
- Область действия квантора 448
- Оператор
- минимизации 520
- примитивной рекурсии 519
- суперпозиции 519
- Оправдание 393
- Остов 261
- Отрицание 359

П

- Пакет символьных вычислений
- Maple 119–120, 213–214
- Паросочетание 286

- максимальное 286
 - совершенное 286
 - Переменная
 - предметная 446
 - существенная 409
 - фиктивная 409
 - Перестановка 130
 - Перечисление
 - декартова произведения 151–152
 - перестановок 153–158
 - подмножеств 147–151
 - Перешеек 242
 - Подпись электронная 83
 - Подформула 362
 - Позиционные системы счисления
 - алгоритмы арифметических действий 19–21
 - алгоритмы перевода 21–22
 - факториальная запись 19
 - Поле 56
 - Полная система вычетов 57
 - Пометка вершины 298, 300, 301, 305
 - инициализация 312
 - модификация 299, 307, 312
 - Порядок
 - антилексикографический 145
 - лексикографический 145, 410
 - обратный лексикографическому 145
 - Поток в сети 319
 - Правило резолюций
 - в логике высказываний 392
 - в логике предикатов 466
 - Предикат 446
 - Принцип
 - Дирихле 128
 - включения–исключения 175
 - двойственности 369, 377–379, 424
 - нормализации 522
 - Производящая функция 128, 198
 - для чисел Каталана 209
 - для чисел Фибоначчи 199
 - числа разбиений 199
 - Пролог
 - база знаний 471
 - запрос 471
 - правило 471
 - факт 471
 - Псевдограф 229
 - Путь 230
 - гамильтонов 310
 - кратчайший 297–299
 - направленный 232
 - эйлеров 248, 252
- ## Р
- Разбиение ребра 280
 - Разложение Шеннона 438
 - Размещение 130
 - без повторений 130
 - с повторениями 131
 - Разрез 270
 - $s - t$ -разрез 321
 - величина потока через разрез 321
 - минимальный 270
 - пропускная способность 321
 - простой 270
 - Ребро 229
 - ветвь 269
 - кратное 229
 - мост 242
 - отделимости 242
 - перешеек 242
 - петля 229
 - разделяющее 242
 - хорда 269
 - Регулярные выражения 580
 - замыкание Клини 580
 - итерация языков 580
 - конкатенация языков 580
 - объединение языков 580

- свойства 581
- эквивалентность 581
- Редукция конечного автомата 574
- Резольвента 392
 - бинарная 466
- Решетка 403
 - дистрибутивная 404
 - ограниченная 404
 - с дополнением 405
- С**
 - Связки логические 359
 - Сеть 319
 - транспортная 319
 - Сигнатура 448
 - Символ функциональный 448
 - Система остаточных классов 75
 - Склейка дизьюнкта 467
 - Сколемизация 459
 - Слово 256
 - длина 256
 - префикс 256
 - суффикс 256
 - Соединение контактов
 - параллельное 380
 - последовательное 380
 - Соотношение Безу 27
 - Сочетание 130
 - без повторений 130
 - с повторениями 131
 - Список инцидентности 236
 - Сравнение по модулю 56
 - линейное 67
 - Степень вершины 230
 - Стиль календарный 46
 - Стратегия метода резолюций
 - вычеркивания 400
 - насыщения уровней 399
 - Субфакториал 178
 - Схема Горнера 93
- Т**
 - Таблица
- Поста 435
- истинности 359, 363
- Тезис
- Тьюринга 516
- Черча 520
- Теорема
 - Безу 93
 - Бернсайда 192
 - Вандермонда 163
 - Визинга 281
 - Вильсона 71
 - Евклида 11
 - Кенига 274
 - Лагранжа 42
 - Ламе 36
 - Лежандра 176
 - Понtryгина–Куратовского 280
 - Поста 432
 - Ферма (малая) 66
 - Форда–Фалкерсона 323
 - Хивуда 281
 - Цекендорфа 52
 - Чебышева 11
 - Эйлера 230, 279
 - Эйлера–Ферма 66
 - китайская об остатках
 - для многочленов 97
 - для чисел 71
 - об эквивалентности автоматов
 - Мили и Мура 558
 - о декомпозиции 480
 - о детерминизации 564
 - о полноте метода резолюций 393
 - о разрастании для автоматных языков 568
- Терм 448
- Тождество Коши 134
- Треугольник Паскаля 122, 133
- У**
 - Узел 232
 - Укладка графа

- плоская 277
- трехмерная 277

Унификатор 464

- наиболее общий 464

Уравнение

- диофантово 47
- рекуррентное 197
- — неоднородное 207
- — однородное 201
- — порядок 197

Ф

Форма Бэкуса–Наура 536

Формула

- Байеса 219
- Бине 200, 203
- Гаусса 63
- Лагранжа 97
- логики высказываний 360
- — выполнимость 371
- — дизъюнктивная нормальная форма 373, 381
- — конъюнктивная нормальная форма 377, 381
- — логическое следствие 369
- — минимальная дизъюнктивная нормальная форма 383
- — противоречие 365
- — равносильность 364
- — совершенная дизъюнктивная нормальная форма 374
- — совершенная конъюнктивная нормальная форма 377
- — с тесными отрицаниями 373
- — тавтология 365
- логики предикатов 448
- — замкнутая 449
- — логическое следствие 456
- — общезначимая 453
- — предваренная нормальная форма 457
- — предложение 449

- — пренексная нормальная форма 457
- — равносильность 452
- — сколемовская нормальная форма 458
- обращения Мёбиуса 180
- обращения Мебиуса 213
- Функция
- Мёбиуса 65, 179
- Эйлера 61
- булева 408
- — Шеффера обобщенная 435
- — вес 411
- — геометрическое представление 411
- — двойственная 424
- — замыкание 413
- — линейная 419
- — мажоритарная 411
- — монотонная 428
- — неравнозначности 409
- — область истинности 411
- — полином Жегалкина 415
- — полная 435
- — самодвойственная 425
- — сохраняющая единицу 414
- — сохраняющая ноль 413
- — стрелка Пирса 409
- — суперпозиция 412
- — четности 428
- — штрих Шеффера 409
- мультипликативная 64
- односторонняя 76
- оценочная для целевой функции 330
- расстояния 135
- рекурсивная
- — общерекурсивная 520
- — примитивно рекурсивная 520
- — частично рекурсивная 520
- с секретом 77

X

- Хемминга
— вес кодового слова 137
— граница 136
— кодовое расстояние 135
Хорда 269

Ц

- Цепочка 256
Цепь 230
Цикл 230
— главный 269
Цикломатическое число 268

Ч

- Числа
— Белла 161
— Каталана 168

- Стирлинга
— — второго рода 160
— — первого рода 162
— Фибоначчи 51, 203
— — соотношение Кассини 52
— — система счисления 53
— бесквадратные 10
— взаимно простые 7
— каноническое представление 9
— простые 9
— свободные от квадратов 10
— хроматические 280
Число
— Фидия 53, 200

Э

- Эквиваленция 359