Материалы
к дистанционному курсу лекций
по дискретной математике

2020/2021 уч. год

4 курс, I поток

Лектор — проф. В. В. Кочергин

3 сентября 2020 г.

Просьба о замеченных опечатках сообщать автору
по электронной почте на адрес vvkoch@yandex.ru

Глава 1

Булевы функции, замкнутые классы

§ 1.1 Теорема Поста

Лекция № 1

Основной целью данного раздела является доказательство теоремы Поста (1920 г.) о конечной порожденности всякого замкнутого класса булевых функций.

Введем обозначения и напомним известные определения.

Положим $E = \{0, 1\}$. Пусть $f^{(n)} : E^n \to E$ — отображение, тогда $f^{(n)}(x_1, \ldots, x_n)$ — *n-местная булева функция* (или *функция алгебры логики*), $x_1, \ldots, x_n$ — переменные этой функции, а $f^{(n)}$ (или просто f) — функциональный символ. Подчеркнем, что как правило, под булевой функцией понимается не просто отображение, но и упорядоченный набор переменных. Через P_2 обычно обозначается множество всех булевых функций, а через $P_2(n)$ — множество всех булевых функций от n фиксированных переменных (сами переменные указываются только в случае необходимости).

Далее будем считать, что имеется счетное множество X переменных, скажем, множество

$$\{x_1, x_2, \ldots, \quad y_1, y_2, \ldots, \quad z_1, z_2, \ldots, \quad u_1, u_2, \ldots, \quad v_1, v_2, \ldots, \quad w_1, w_2, \ldots\}.$$

Переменная x_i называется *существенной* для функции $f(x_1, \ldots, x_n)$, если существует набор $(\alpha_1, \ldots, \alpha_{i-1}, \alpha_{i+1}, \ldots, \alpha_n)$ значений остальных переменных, при котором изменение значения переменной x_i приводит к

изменению значения функции:

$$f(\alpha_1, \ldots, \alpha_{i-1}, 0, \alpha_{i+1}, \ldots, \alpha_n) \neq f(\alpha_1, \ldots, \alpha_{i-1}, 1, \alpha_{i+1}, \ldots, \alpha_n).$$

Если же набора с таким свойством не существует, то переменная x_i называется *несущественной* или *фиктивной*.

Упражнение 1. Доказать, что из функции, существенно зависящей от n переменных, путем подстановки некоторой константы вместо некоторой переменной можно получить функцию, существенно зависящую от $n-1$ переменной.

Способы задания булевых функций:

1. Перечисление значений (таблица).

Приведем пример задания таблицей следующих функций от двух переменных: $e_1(x, y)$ — селекторной функции для первого аргумента x, $0^{(2)}$ — двухместной константы 0, $x\&y$ — конъюнкции, $x \vee y$ — дизъюнкции, $x \to y$ — импликации, $x \oplus y$ — суммы по модулю 2 (исключающего ИЛИ), $x \sim y$ — эквивалентности (счетчика четности), $x \mid y$ — штриха Шеффера, $x \downarrow y$ — стрелки Пирса.

x	y	$e_1(x,y)$	$0^{(2)}$	$x\&y$	$x \vee y$	$x \to y$	$x \oplus y$	$x \sim y$	$x \mid y$	$x \downarrow y$
0	0	0	0	0	0	1	0	1	1	1
0	1	0	0	0	1	1	1	0	1	0
1	0	1	0	0	1	0	1	0	1	0
1	1	1	0	1	1	1	0	1	0	0

2. Описательный (работает эффективно, но может быть использован только в редких случаях).

Пример — функция от 100 переменных, которая равна единице тогда и только тогда, когда число переменных, принимающих единичное значение, делится на 5.

3. Формула.

Индуктивно определим понятие формулы. Пусть дано множество переменных X и множество функциональных символов (конечное или счетное) $A = \{f_1^{(n_1)}, \ldots, f_k^{(n_k)}, \ldots\}$, а также следующие правила:

1. Переменные из множества X — тривиальные формулы[1]).

2. Если каждая из записей $\Phi_1, \ldots, \Phi_{n_i}$ является либо символом переменной, либо формулой, то формулой является и запись $f_i^{(n_i)}(\Phi_1, \ldots, \Phi_{n_i})$.

[1]) При этом формально остается открытым вопрос о том, является ли тривиальная формула (переменная) формулой.

Теперь можно определить *формулу над множеством переменных* X *и множеством функциональных символов* $A = \{f_1^{(n_1)}, \ldots, f_k^{(n_k)}, \ldots\}$ как слово из функциональных символов, символов переменных, скобок и запятых, которое можно получить по указанным правилам за конечное число шагов.

Формулы будем обозначать, как правило, большими греческими буквами, иногда указывая все символы переменных, которые входят в формулу: $\Phi = \Phi(x_1, \ldots, x_n)$.

Формула вида $f_i^{(n_i)}(\Phi_1, \ldots, \Phi_{n_i})$, в которой $\Phi_1, \ldots, \Phi_{n_i}$ — переменные, называется *простейшей*.

Подформулой данной формулы называется любое подслово этой формулы, являющееся, в свою очередь, формулой.

Индуктивно определим понятие *глубины* $D(\Phi)$ формулы Φ. Для тривиальной формулы Φ полагаем $D(\Phi) = 0$, для простейшей формулы Φ полагаем $D(\Phi) = 1$. Далее, для формулы Φ, имеющий вид $\Phi = f_i^{(n_i)}(\Phi_1, \ldots, \Phi_{n_i})$, полагаем $D(\Phi) = 1 + \max(D(\Phi_1), \ldots, D(\Phi_{n_i}))$.

Произвольной формуле $\Phi(x_1, \ldots, x_n)$ и всякому набору $(\alpha_1, \ldots, \alpha_n)$ значений переменных $x_1, \ldots, x_n$ естественным образом можно сопоставить значение $\Phi(\alpha_1, \ldots, \alpha_n) \in E$. Тем самым формуле $\Phi(x_1, \ldots, x_n)$ сопоставляется некоторая булева функция $f(x_1, \ldots, x_n)$. В этом случае говорят, что *функция* $f(x_1, \ldots, x_n)$ *реализуется формулой* $\Phi(x_1, \ldots, x_n)$.

Упражнение 2. Реализовать функцию $x \downarrow y$ формулой над системой $\{x \mid y\}$ с использованием минимально возможного числа функциональных символов.

Итак, мы практически определили операцию *суперпозиции*. Если функция $f(x_1, \ldots, x_n)$ реализуется нетривиальной формулой $\Phi(x_1, \ldots, x_n)$ над системой функций A, то говорят, что она *получена операцией суперпозиции* из функций системы A.

На множестве булевых функций определим еще одну операцию — *операцию введения фиктивной переменной*, которая произвольную функцию $f(x_1, \ldots, x_n)$ отображает в функцию $g(x_1, \ldots, x_n, x_{n+1})$, задаваемую на произвольном наборе $(\alpha_1, \ldots, \alpha_n, \alpha_{n+1})$ равенством $g(\alpha_1, \ldots, \alpha_n, \alpha_{n+1}) = f(\alpha_1, \ldots, \alpha_n)$.

Пусть $A \subseteq P_2$. *Замыканием* множества A (относительно операций суперпозиции и введения фиктивной переменной) называется множество $[A]$ всех таких булевых функций, которые могут быть получены из функций множества A при помощи операций суперпозиции и введения фиктивной переменной.

Свойства оператора замыкания:

1. $A \subseteq [A]$ (экстенсивность).

2. Если $A \subseteq B$, то $[A] \subseteq [B]$ (монотонность).

3. $[[A]] = [A]$ (идемпотентность).

Если для множества булевых функций A выполняется равенство $A = [A]$, то такое множество называют *замкнутым множеством* (функций) или *замкнутым классом*, а иногда и просто *классом*.

Если для множеств A и F булевых функций выполняется равенство $[A] = F$, то говорят, что система A *полна* в F.

Замкнутый класс F называется *конечнопорожденным* (имеющим конечный базис), если существует система функций A, удовлетворяющая условиям: $[A] = F$, $|A| < \infty$.

Функции *равны*, если у них совпадают множества переменных, от которых они зависят существенно, и на каждом наборе этих переменных функции принимают одинаковые значения.

Формулы *эквивалентны*, если они реализуют равные функции.

Насколько принципиально, скажем, для доказательства теоремы Поста под функцией понимать одну функцию или класс эквивалентности равных функций, получающихся друг из друга путем добавления или изъятия фиктивных переменных?

Обозначим черех $[A]_c$ множество всех таких булевых функций, которые могут быть получены из функций множества A при помощи операций суперпозиции. Если для множества функций F выполняется равенство $[F]_c = F$, то множество F называется классом Поста.

Очевидно, что равенство $[F] = F$ влечет за собой равенство $[F]_c = F$. Обратное, вообще говоря, неверно:

$$[\{1^{(1)}(x)\}]_c = \{1^{(1)}(x)\} \neq \{1^{(1)}(x), 1^{(2)}(x), \ldots\} = [\{1^{(1)}(x)\}].$$

Упражнение 3. Доказать, что существует 17 классов F, удовлетворяющих условиям:

$$[F]_c = F, \quad [F] \neq F.$$

Теперь предъявим некоторые полные в P_2 системы функций. В силу разложения по переменной

$$f(x_1, \ldots, x_n) = f(x_1, \ldots, x_{n-1}, 0)\overline{x_n} \vee f(x_1, \ldots, x_{n-1}, 1)x_n$$

справедливо равенство

$$[\{xy, x \vee y, \overline{x}\}] = P_2,$$

из которого, в свою очередь, легко получаются равенства

$$P_2 = [\{xy, \overline{x}\}] = [\{x \vee y, \overline{x}\}] = [\{x \mid y\}] = [\{x \downarrow y\}] = [\{xy, x \oplus y, 1\}].$$

Последняя полная система называется *базисом Жегалкина*. Из произвольной формулы над базисом Жегалкина путем раскрытия скобок и приведения подобных слагаемых полуается полином над полем $\mathbb{Z}_2$. Напомним, что у каждой булевой функции ровно один такой полином, и это полином называется *полиномом Жегалкина*.

Упражнение 4. Из полинома Жегалкина степени (максимальное число переменных в одном слагаемом) k ($k \geq 3$) путем отождествления переменных можно получить полином Жегалкина степени ровно $k-1$.

1.1.1 Принцип двойственности

Пусть $f(x_1, \ldots, x_n)$ — произвольная булева функция. Функция $f(\overline{x}_1, \ldots, \overline{x}_n)$, обозначаемая обычно $f^*(x_1, \ldots, x_n)$, называется *двойственной (к) функции* $f(x_1, \ldots, x_n)$.

Принцип двойственности: Если в произвольной формуле Φ, реализующей булеву функцию f, заменить все функциональные символы на функциональные символы двойственных функций, то получившаяся формула Φ' будет реализовывать функцию f^*.

В качестве примера применения принципа двойственности рассмотрим задачу о нахождении полных систем в классах T_0 и T_1, где T_i ($i = 0, 1$) — класс всех булевых функций, принимающих на наборе из всех значений i также значение i. Легко понять, что $T_0 = [\{xy, x \oplus y\}]$, так как полином Жегалкина любой функции из класса T_0 имеет нулевой свободный член, а любой такой полином может быть реализован над системой функций $[\{xy, x \oplus y\}]$. Теперь, по принципу двойственности, двойственный к классу T_0 класс T_1 порождается функциями, двойственными к функциям xy и $x \oplus y$, т. е. справедливо равенство $T_1 = [\{x \vee y, x \sim y\}]$.

1.1.2 Классы функций, содержащие константы 0 и 1

Приведем примеры таких классов и их свойств.

I. Класс L линейных функций.

Функция $f(x_1, \ldots, x_n)$ является *линейной* (т. е. $f \in L$) тогда и только тогда, когда существует набор $(a_0, a_1, \ldots, a_n)$ из нулей и единиц, такой что справедливо равенство $f(x_1, \ldots, x_n) = a_0 \oplus a_1 x_1 \oplus \ldots \oplus a_n x_n$.

Свойства класса L:

1. $0, 1, x, \overline{x}, x \oplus y \in L; \quad xy, x \vee y \notin L$.
2. $[L] = L$.
3. $[\{x \oplus y, 1\}] = L$.

4. Любая линейная функция от n переменных, отличная от константы, принимает значение 1 на 2^{n-1} наборе.

5. Если[2]) $f_L \notin L$, то найдется такая функция $g(x,y)$, что выполняются условия: $g(x,y) \notin L$, $g(x,y) \in [\{0, f_L\}]$ (т. е. из нелинейной функции и константы 0 можно получить нелинейную функцию от двух переменных).

$\triangleright$ Выделим в полиноме Жегалкина функции f_L нелинейную часть, а в ней — слагаемое наименьшей длины (степени). Теперь для получения искомой функции $g(x,y)$ достаточно вместо первой переменной из этого слагаемого подставить x, вместо других переменных из этого слагаемого подставить y, а вместо всех остальных переменных — константу 0. $\square$

6. (Лемма о нелинейной функции.) Если $f_L \notin L$, то $xy \in [\{0, \overline{x}, f_L\}]$ (т. е. из нелинейной функции, отрицания и константы 0 можно получить конъюнкцию двух переменных).

$\triangleright$ По предыдущему свойству из нелинейной функции f_L и константы 0 можно получить нелинейную функцию от двух переменных $g(x,y)$, полином Жегалкина которой имеет вид $xy \oplus ax \oplus by \oplus c$, где $(a,b,c) \in E^3$. Если $a = 1$, то подставив вместо y отрицание этой переменной $y \oplus 1$, получим полином $xy \oplus by \oplus b \oplus c$. Теперь, если $b = 1$, подставив $x \oplus 1$ вместо x, получим полином $xy \oplus b \oplus c$. Для получение конъюнкции осталось в случае выполнения равенства $b \oplus c = 1$ взять отрицание получившейся функции. $\square$

7. Выполняется равенство $|L(n)| = 2^{n+1}$, где $L(n) = L \cap P_2(n)$.

8. Функциональное (относительно f) уравнение $[\{f\}] = L$ не имеет решений.

Упражнение 5. Доказать свойство 8 класса линейных функций.

II. Классы K и D (конъюнкций и дизъюнкций соответственно).

Функция $f(x_1, \ldots, x_n)$ является *конъюнкцией* (т. е. $f \in K$) тогда и только тогда, когда существует набор $(a_0, a_1, \ldots, a_n)$ из нулей и единиц, такой что справедливо равенство $f(x_1, \ldots, x_n) = a_0(a_1 \vee x_1) \ldots (a_n \vee x_n)$.

Функция $f(x_1, \ldots, x_n)$ является *дизъюнкцией* (т. е. $f \in D$) тогда и только тогда, когда существует набор $(a_0, a_1, \ldots, a_n)$ из нулей и единиц, такой что справедливо равенство $f(x_1, \ldots, x_n) = a_0 \vee a_1 x_1 \vee \ldots \vee a_n x_n$.

Свойства классов K и D:

1. $0, 1, x, xy, \in K$; $\overline{x}, x \oplus y, x \vee y \notin K$. $0, 1, x, x \vee y, \in D$; $\overline{x}, x \oplus y, xy \notin K$.
2. $[K] = K$; $[D] = D$.

[2]) Здесь и далее будем часто использовать обозначение вида f_F для функции, не принадлежащей классу F.

3. $[\{0,1,xy\}] = K$; $[\{0,1,x \vee y\}] = D$.

4. Если $f \in K$ и $\{x_1, \ldots, x_n\}$ — множество всех переменных, от которых функция f зависит существенно, то $f = x_1 \ldots x_n$; если $f \in D$ и $\{x_1, \ldots, x_n\}$ — множество всех переменных, от которых функция f зависит существенно, то $f = x_1 \vee \ldots \vee x_n$.

III. Класс M монотонных функций.

На множестве всех наборов n-мерного единичного (булева) куба E^n введем частичный порядок «$\preccurlyeq$» следующим образом: $(\alpha_1, \ldots, \alpha_n) \preccurlyeq (\beta_1, \ldots, \beta_n)$ тогда и только когда для всех $i = 1, \ldots, n$ выполняются неравенства $\alpha_i \leq \beta_i$. Наборы $\tilde{\alpha}$ и $\tilde{\beta}$ называются *сравнимыми*, если выполняется хотя бы одно из отношений $\tilde{\alpha} \preccurlyeq \tilde{\beta}$ или $\tilde{\beta} \preccurlyeq \tilde{\alpha}$. Если ни одно из этих двух отношений не выполняется, то такие наборы называются *несравнимыми*.

Функция $f(x_1, \ldots, x_n)$ является *монотонной* (т. е. $f \in M$) тогда и только тогда, когда для любых двух наборов $(\alpha_1, \ldots, \alpha_n)$ и $(\beta_1, \ldots, \beta_n)$ соотношение $(\alpha_1, \ldots, \alpha_n) \preccurlyeq (\beta_1, \ldots, \beta_n)$ влечет неравенство $f(\alpha_1, \ldots, \alpha_n) \leq f(\beta_1, \ldots, \beta_n)$.

Свойства класса M:

1. $0, 1, x, xy, x \vee y \in M$; $\quad \overline{x}, x \oplus y, x \to y \notin M$.

2. $[M] = M$.

3. Если $f(x_1, \ldots, x_n) \in M$, то справедливо разложение

$$f(x_1, \ldots, x_n) = f(x_1, \ldots, x_{n-1}, 0) \vee f(x_1, \ldots, x_{n-1}, 1)x_n.$$

4. $[\{0,1,xy,x \vee y\}] = M$.

5. Если $f \in M$, $f \not\equiv 0, 1$, то $f \in [\{xy, x \vee y\}]$.

$\triangleright$ Это утверждение легко устанавливается индукцией по числу n переменных, от которых функция f зависит существенно. База индукции ($n = 1$) очевидна. Переход индукции обоснован разложением из свойства 3 монотонных функций — достаточно только заметить, что $f(x_1, \ldots, x_{n-1}, 0) \not\equiv 1$ и $f(x_1, \ldots, x_{n-1}, 1) \not\equiv 0$. $\square$

6. $K \subset M$; $D \subset M$.

7. (Лемма о немонотонной функции.) Если $f_M \notin M$, то $\overline{x} \in [\{0, 1, f_M\}]$ (т. е. из немонотонной функции и констант можно получить отрицание).

$\triangleright$ Для немонотонной функции $f_M(x_1, \ldots, x_n)$ найдутся два набора $\tilde{\alpha} = (\alpha_1, \ldots, \alpha_n)$ и $\tilde{\beta} = (\beta_1, \ldots, \beta_n)$ из E^n, удовлетворяющие условиям: $\tilde{\alpha} \preccurlyeq \tilde{\beta}$, $f(\tilde{\alpha}) = 1$, $f(\tilde{\beta}) = 0$. Для $i = 1, \ldots, n$ при выполнении равенства $\alpha_i = \beta_i$ в функцию $f_M(x_1, \ldots, x_n)$ вместо переменной x_i подставим константу α_i, а при выполнении неравенства $\alpha_i < \beta_i$ — переменную x, тем самым получим функцию $\overline{x}$. $\square$

8. Если $f_K, f_D \in M$, $f_K \notin K$, $f_D \notin D$, то $x \vee y \in [\{1, f_K\}]$, $xy \in [\{0, f_D\}]$.

$\triangleright$ Так как $f_D \in M$, $f_D \notin D$, то $f_D \notin \{0, 1, x, \overline{x}\}$. Следовательно функция f_D существенно зависит не менее, чем от двух перемееных. Пусть $x_1, \ldots, x_n$ — все переменные, от которых функция f_D зависит существенно.

Так как $f_D \in M$, $f_D \notin D$, то существует набор $\tilde{\alpha} \in E^n$, в котором ровно одна единица и на котором функция обращается в 0. Без ограничения общности будем считать, что $f_D(1, 0, \ldots, 0) = 0$.

Функция f зависит от переменной x_1 существенно. Поэтому найдется набор $(\beta_2, \ldots, \beta_n)$, для которого выполняется условие $f_D(0, \beta_2, \ldots, \beta_n) \neq f_D(1, \beta_2, \ldots, \beta_n)$. В силу монотонности функции f_D имеем: $f_D(0, \beta_2, \ldots, \beta_n) = 0$, $f_D(1, \beta_2, \ldots, \beta_n) = 1$.

Кроме того, $(\beta_2, \ldots, \beta_n) \neq (0, \ldots, 0)$, так как $f_D(1, 0, \ldots, 0) = 0$. Без ограничения общности будем считать, что

$$(\beta_2, \ldots, \beta_n) = (\underbrace{1, \ldots, 1}_{k \geq 1}, 0, \ldots, 0).$$

Тогда

$$f_D(x, \underbrace{y, \ldots, y}_{k}, 0, \ldots, 0) = g(x, y) = xy.$$

Для того, чтобы установить принадлежность функции $x \vee y$ классу $[\{1, f_K\}]$, достаточно применить принцип двойственности. $\square$

9. Выполняются неравенства

$$2^{C_n^{\lfloor n/2 \rfloor}} \leq |M(n)| \leq n^{C_n^{\lfloor n/2 \rfloor}},$$

где $M(n) = M \cap P_2(n)$.

Упражнение 6. Доказать свойство 9 класса монотонных функций.

Теорема 1. *Пусть $F \subseteq P_2$, $F = [F]$, $\{0, 1\} \subseteq F$. Тогда найдется система функций A, удовлетворяющая условиям: $[A] = F$, $|A| < \infty$.*

Доказательство. Отдельно рассмотрим несколько случаев.

Случай 1°: $F \not\subseteq M$.

Тогда в классе F найдется немонотонная функция f_M. Применяя лемму о немонотонной функции, получаем: $\overline{x} \in [\{0, 1, f_M\}] \subseteq F$.

Случай 1.1°: $F \not\subseteq L$.

Тогда в классе F найдется нелинейная функция f_L. Применяя лемму о нелинейной функции, получаем: $xy \in [\{0, \overline{x}, f_L\}] \subseteq F$.

Таким образом, $\{xy, \overline{x}\} \subseteq F$. Поэтому

$$F = P_2 = [\{0, 1, f_M, f_L\}].$$

Случай 1.2°: $F \subseteq L$.

Случай 1.2.1°: в классе F есть функция, существенно зависящая не менее чем от двух переменных.

Тогда $F = L = [\{0, 1, x \oplus y\}]$.

Случай 1.2.2°: в классе F нет функций, существенно зависящих не менее чем от двух переменных.

Тогда $F = \{0, 1, x, \overline{x}\} = U$.

Случай 2°: $F \subseteq M$.

Случай 2.1°: $F \not\subseteq K$, $F \not\subseteq D$.

Тогда в классе F найдется монотонная функция f_K, не лежащая в классе K и найдется монотонная функция f_D, не лежащая в классе D. Используя свойство 8 монотонных функций, получаем: $\{xy, x \vee y\} \subset [\{0, 1, f_K, f_D\}] \subseteq F$. Поэтому $F = M = [\{0, 1, f_K, f_D\}]$.

Случай 2.2°: $F \subseteq K$, $F \not\subseteq D$.

Тогда $F = K = [\{0, 1, xy\}]$.

Случай 2.3°: $F \not\subseteq K$, $F \subseteq D$.

Тогда $F = D = [\{0, 1, x \vee y\}]$.

Случай 2.4°: $F \subseteq K$, $F \subseteq D$.

Тогда либо $F = C = [\{0, 1\}]$, либо $F = MU = [\{0, 1, x\}]$. □

На рисунке 1 показана диаграмма вложености замкнутых классов булевых функций, содержащих обе константы (если два класса соединены ребром, то нижний содержится в верхнем).

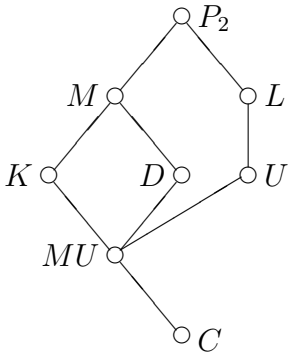

Рис. 1

1.1.3 Классы функций, содержащие константу 1 и не содержащие константу 0

Сначала докажем лемму, которая в дальнейшем будет многократно использоваться. Дадим ей специальное название — лемма Z.

Лемма 1 (лемма Z). *Пусть* $A \subseteq P_2$, $x \vee y \in [A]$, $f \in [A \cup \{0\}]$, $g \leq f$, $g \in [A]$. *Тогда* $f \in [A]$.

Доказательство. Так как $f(x_1, \ldots, x_n) \in [A \cup \{0\}]$, то найдется формула $\Phi(x_1, \ldots, x_n, y)$ над системой A, такая что формула $\Phi(x_1, \ldots, x_n, 0)$ над системой $A \cup \{0\}$ реализует функцию $f(x_1, \ldots, x_n)$.

Пусть формула $\Psi(x_1, \ldots, x_n)$ реализует функцию $g(x_1, \ldots, x_n)$.

Рассмотрим формулу $\Psi(x_1, \ldots, x_n) \vee \Phi(x_1, \ldots, x_n, \Psi(x_1, \ldots, x_n))$ над системой A. Покажем, что реализуемая этой формулой функция $h(x_1, \ldots, x_n)$ есть функция $f(x_1, \ldots, x_n)$.

1°. Если $g(\tilde{\alpha}) = 1$, то $h(\tilde{\alpha}) = 1$ и $f(\tilde{\alpha}) = 1$.

2°. Если $g(\tilde{\alpha}) = 0$, то $h(\tilde{\alpha}) = f(\tilde{\alpha})$. □

Булева функция $f(x_1, \ldots, x_n)$ *удовлетворяет условию* 0^∞, если для некоторого i, $1 \leq i \leq n$, на всех наборах значений переменных выполняется неравенство $f(x_1, \ldots, x_n) \geq x_i$.

Определим класс O^∞ как множество всех булевых функций, удовлетворяющих условию 0^∞.

Свойства класса O^∞:

1. $1, x, x \vee y, x \to y \in O^\infty$; $\quad 0, xy \notin O^\infty$.
2. $[O^\infty] = O^\infty$.
3. $[\{x \to y\}] = O^\infty$.

▷ Проверим выполнение условий леммы Z:

$A = [\{x \to y\}]$;

$x \vee y = (x \to y) \to y \in [A]$;

если $f \in O^\infty$, то $f \geq x$ для некоторой переменной x, и $x = x \vee x \in [A]$;

$f \in [\{0, x \to y\}] = P_2$ (так как $x \vee y \in [\{0, x \to y\}]$, $\overline{x} = x \to 0$).

Следовательно, по лемме Z имеем: $f \in [\{x \to y\}]$. □

Введем важные для дальнейшего изложения функции:

$$\omega(x, y, z) = x \vee yz,$$

$$d_p(x_1, \ldots, x_p) = \bigvee_{1 \leq i < j \leq p} x_i x_j, \quad p = 2, 3, \ldots$$

Свойства функций ω и d_p:

1. $\omega, d_p \in M,\ \omega \in O^\infty,\ d_p \notin O^\infty$.

2. Если $f \in MO^\infty = M \cap O^\infty$, $f \not\equiv 1$, то $f \in [\{\omega\}]$ (т. е. функция ω порождает класс $M \cap O^\infty \cap T_0$).

$\triangleright$ По лемме Z. $\square$

3. Если $f_K \in M \setminus K$, $f_D \in M \setminus D$, то $\omega \in [\{1, f_K, f_D\}]$.

$\triangleright$ По лемме Z. $\square$

4. Если $p \geq 3$, то $\omega \in [\{1, d_p\}]$ (следует из свойства 3); если $p > 3$, то $\omega \in [\{d_p\}]$ (следует из представления $x \vee yz = d_p(y, z, x, \ldots, x)$).

5. Справедливо равенство

$$d_{p+1}(x_1, \ldots, x_{p+1}) = x_{p+1}(x_1 \vee \ldots \vee x_p) \vee d_p(x_1, \ldots, x_p).$$

Поэтому, в частности,

$$d_{p+1}(x_1, \ldots, x_p, 0) = d_p(x_1, \ldots, x_p),$$
$$d_{p+1}(x_1, \ldots, x_p, 1) = x_1 \vee \ldots \vee x_p.$$

6. $d_{p+1} \in [\{\omega, d_p\}]$ (следует из свойства 5 и включения $x \vee y \in [\{x \vee yz\}]$).

7. Имеет место цепочка включений

$$[\omega] \subseteq \ldots \subseteq [\{\omega, d_{p+1}\}] \subseteq [\{\omega, d_p\}] \subseteq \ldots \subseteq [\{\omega, d_2\}].$$

8. Выполняются соотношения

$$d_p \notin [\{\omega, d_{p+1}\}], \quad p = 2, 3, \ldots$$

Упражнение 7. Доказать свойство 8. (*Указание*: доказать, что при $p \geq 2$ функции ω и d_{p+1} принадлежат, а функция d_p не принадлежит замкнутому классу O^p всех булевых функций, удовлетворяющих условию 0^p (функция удовлетворяет условию 0^p тогда и только тогда, когда у любых p наборов, на которых функция обращается в 0, есть общая нулевая компонента).)

Замечание 1. Из свойств 7 и 8 следует, что семейство замкнутых классов булевых функций бесконечно.

Для произвольной монотонной функции введем множество функций, которые получаются из этой функции путем отождествления любых двух переменных. Дадим строгие определения.

Пусть $f(x_1, \ldots, x_n) \in M$, $i \neq j$, $1 \leq i, j \leq n$. Положим

$$f_j^i(x_1, \ldots, x_n) = f(x_1, \ldots, x_{j-1}, x_i, x_{j+1}, \ldots, x_n).$$

Свойства функций f_j^i:

1. $f_j^i \in [\{f\}] \subset M$.
2. Если $f \notin \{0,1\}$, то $f_j^i \notin \{0,1\}$
3. При $x_i = 0$ имеем $f_j^i \le f$, поэтому $f_j^i \le f \vee x_i$.
   При $x_i = 1$ имеем $f_j^i \ge f$, поэтому $f_j^i \ge f x_i$.
   При $x_j = 0$ имеем $f_j^i \ge f$, поэтому $f \le f_j^i \vee x_j$.
   При $x_j = 1$ имеем $f_j^i \le f$, поэтому $f \ge f_j^i x_j$.

Теперь для монотонной функции$f(x_1, \dots, x_n)$ от n переменных, $n \ge 2$, положим

$$A_{n-1}(f(x_1, \dots, x_n)) = \{f_j^i \mid 1 \le i \le n,\ 1 \le j \le n,\ i \ne j\}.$$

Лемма 2 (основная)**.** *Пусть* $f(x_1, \dots, x_n) \in M$, $n \ge 2$. *Тогда*

$$f \in [\{\omega, d_n\} \cup A_{n-1}(f)].$$

Доказательство. Если $f \equiv 0$ или $f \equiv 1$, то утверждение леммы справедливо ввиду равенства $f_j^i = f$. Если $f \in O^\infty \setminus \{1\}$, то $f \in [\{\omega\}]$. Далее будем считать, что функция f отлична от константы и не лежит в классе O^∞. Проведем индукцию по n (числу переменных функции f).

База индукции: $n = 2$. Единственной монотоной функцией не более чем от двух переменных, отличной от константы и не лежащей в классе O^∞, является функция $xy = d_2(x, y)$.

Переход индукции, $n \ge 3$. Так как $f \in M \setminus O^\infty$, то функция f на любом наборе, содержащем ровно одну единицу, принимает значение 0.

В разложении

$$f(x_1, \dots, x_n) = f(x_1, \dots, x_{n-1}, 1)x_n \vee f(x_1, \dots, x_{n-1}, 0)$$

обозначим $f(x_1, \dots, x_{n-1}, 0) = g(x_1, \dots, x_{n-1})$. Так как $f \ge g$ и $f \notin O^\infty$, то $g \notin O^\infty$ (иначе бы функция f не меньше какой-то своей переменной).

По предположению индукции $g \in [\{\omega, d_{n-1}\} \cup A_{n-2}(g)]$, где

$$A_{n-2}(g(x_1, \dots, x_{n-1})) = \{g_j^i = f_j^i(x_1, \dots, x_{n-1}, 0) \mid 1 \le i, j \le n-1,\ i \ne j\}.$$

Следовательно, для функции $g(x_1, \dots, x_{n-1})$ существует реализующая ее формула Γ над системой $\{\omega, d_{n-1}\} \cup \{g_j^i \mid 1 \le i, j \le n-1,\ i \ne j\}$.

Последовательно в соответствии с шагами индуктивного построения формулы преобразуем формулу Γ в формулу H:

— шаги вида $\omega(\Psi_1, \Psi_2, \Psi_3)$ не изменяем;

— шаг вида $d_{n-1}(\Psi_1, \dots, \Psi_{n-1})$ заменяем на шаг $d_n(\Psi_1, \dots, \Psi_{n-1}, x_n)$;

— шаг вида $g_j^i(\Psi_1, \ldots, \Psi_{n-1})$ заменяем на шаг $f_j^i(\Psi_1, \ldots, \Psi_{n-1}, x_n)$.

Пусть построенная формула H реализует функцию $h(x_1, \ldots, x_n)$. Эта функция обладает следующими свойствами:

1) $h \in [\{\omega, d_n\} \cup A_{n-1}(f)]$,

2) $h(x_1, \ldots, x_{n-1}, 0) = g(x_1, \ldots, x_{n-1}) = f(x_1, \ldots, x_{n-1}, 0)$,

3) $h(0, \ldots, 0, 1) = f(0, \ldots, 0, 1) = g(0, \ldots, 0) = 0$.

Равенство $h(0, \ldots, 0, 1) = 0$ следует из того факта, что на наборе $(0, \ldots, 0, 1)$ при построении формулы H на каждом шаге аргументами очередной подформулы являются только нули и не более одной единицы. Функция f на любом наборе с одной единицей равна 0, а функция g не превосходит функцию f.

Положим

$$\varphi(x_1, \ldots, x_n) = h(x_1, \ldots, x_n) \vee (x_1 \vee \ldots \vee x_{n-1})x_n.$$

Покажем, что

$$\varphi(x_1, \ldots, x_n) = g(x_1, \ldots, x_{n-1}) \vee (x_1 \vee \ldots \vee x_{n-1})x_n.$$

Если $x_n = 0$, то справедливость равенства вытекает непосредственно из определения функции h.

На наборе $(0, \ldots, 0, 1)$ установлены равенства $h(0, \ldots, 0, 1) = g(0, \ldots, 0) = 0$.

Если же $x_n = 1$ и $x_1 \vee \ldots \vee x_{n-1} = 1$, то обе части доказываемого соотношения обращаются в единицу.

Наконец, положим

$$\lambda(x_1, \ldots, x_n) = \varphi(f_n^1, \ldots, f_n^{n-1}, x_n).$$

Для завершения доказательства основной леммы осталось применить лемму Z. Действительно,

$$x \vee y \in [\{\omega\}] \subset [\{\omega, d_n\} \cup A_{n-1}(f)], \quad f \in [\{\omega, 0\}], \quad \lambda \in [\{\omega, d_n\} \cup A_{n-1}(f)].$$

Используя неравенства $f_n^i \leq f \vee x_i$ и $f_n^i x_n \leq f$, установим недостающее соотношение $\lambda \leq f$:

$$\lambda(x_1, \ldots, x_n) = g(f_n^1, \ldots, f_n^{n-1}) \vee (f_n^1 \vee \ldots \vee f_n^{n-1})x_n \leq$$
$$\leq g(f \vee x_1, \ldots, f \vee x_{n-1}) \vee f = f.$$

Последнее равенство отдельно проверется для случаев $f = 0$ (используется неравенство $f \geq g$) и $f = 1$. $\square$

Для монотонной функции $f(x_1, \ldots, x_n)$ помимо уже определенного множества $A_{n-1}(f)$ последовательно введем множества $A_{n-2}(f), \ldots, A_1(f)$:

$$A_k(f) = \bigcup_{g \in A_{k+1}(f)} A_k(g), \quad k = n-2, \ldots, 1.$$

Отметим, что множества $A_k(g)$ из правой части равенства определены, так как у функций g из множества $A_{k+1}(f)$ ровно $k+1$ переменная.

Таким образом, $A_k(f)$ — множество всех функций $g(x_{i_1}, \ldots, x_{i_k})$, получающихся из функции f отождествлением переменных и у которых все переменные x_{i_j} различны.

Для полноты картины положим $A_n(f) = \{f\}$.

Очевидно, что выполняются включения

$$[A_1(f)] \subseteq [A_2(f)] \subseteq \ldots \subseteq [A_{n-1}(f)] \subseteq [A_n(f)].$$

Пусть теперь $f(x_1, \ldots, x_n) \in M \setminus O^\infty$, $f \not\equiv 0$. Тогда у функции f не менее двух существенных переменных. Положим

$$F_f = \bigcup_{k=2}^{n} \left\{ g \in A_k(f) \mid g \notin O^\infty,\ g_j^i \in O^\infty,\ 1 \leq i, j \leq k,\ i \neq j \right\}.$$

Таким образом, F_f — множество всех функций из семейства $\bigcup_{k=2}^{n} A_k(f)$, которые сами не принадлежат классу O^∞, но при отождествлении любых двух различных переменных получается функция, лежащая в классе O^∞.

Свойства множества F_f:

1. Выполняется соотношение $F_f \neq \varnothing$.

2. Справедливо включение $F_f \subseteq [\{f\}]$.

3. Верно неравенство $p(f) \geq 2$, где $p(f)$ — наименьшее число существенных переменных у функций из множесва F_f.

4. Если $g(x_1, \ldots, x_p) \in F_f$, то $g(x_1, \ldots, x_p) = d_p(x_1, \ldots, x_p)$.

$\triangleright$ Так как $g \notin O^\infty$, то на наборах с одной единицей функция g принимает значение 0. Покажем, что на произвольном наборе с двумя единицами функция g принимает значение 1. Без ограничения общности рассмотрим набор $(1, 1, 0, \ldots, 0)$. Так как $g_2^1(x_1, x_3, \ldots, x_p) \in O^\infty$, то найдется k, $k \in \{1, 3, \ldots, p\}$, такое что $g_2^1(x_1, x_3, \ldots, x_p) \geq x_k$.

Если $k \neq 1$ (для определенности будем считать, что $k = 3$), то $g(0, 0, 1, 0, \ldots, 0) = g_2^1(0, 1, 0, \ldots, 0) = 1$. Это противоречит тому, что функция g не лежит в классе O^∞. Значит, $k = 1$. Тогда $g(1, 1, 0, \ldots, 0) =$

$g_2^1(1,0,\ldots,0) = 1$. Теперь осталось воспользоваться монотонностью функции g. $\square$

5. Справедливо соотношение $d_{p(f)} \in [\{f\}]$.

6. Верно включение $A_{p(f)-1}(f) \subseteq O^\infty$.

Лемма 3. *Пусть $f \in M \setminus O^\infty$, $f \not\equiv 0$. Тогда $f \in [\{\omega, d_{p(f)}\}]$.*

Доказательство. В силу основной леммы имеем: $f \in [\{\omega, d_n\} \cup A_{n-1}(f)]$. Рассмотрим произвольную функцию $g \in A_{n-1}$. Если $g \in O^\infty$, то $g \in [\{\omega\}]$. Иначе опять применяем основную лемму и т. д. Таким образом, получаем:

$$f \in [\{\omega, d_n\} \cup A_{n-1}(f)] \subseteq [\{\omega, d_n, d_{n-1}\} \cup A_{n-2}(f)] \subseteq \ldots$$
$$\ldots \subseteq [\{\omega, d_n, \ldots, d_{p(f)}\} \cup A_{p(f)-1}(f)] \subseteq [\{\omega, d_{p(f)}\}].$$

Здесь использовалось включение $[A_{p(f)-1}(f)] \subseteq [\{\omega\}]$, справедливое ввиду того, что отождествление переменных у монотонной функции, отличной от константы, не может привести к получению константы. $\square$

Следствие 1. *Пусть $f \in M \setminus O^\infty$, $f \not\equiv 0$, $p(f) > 3$. Тогда $[\{f\}] = [\{d_{p(f)}\}]$.*

Лекция № 3

Теорема 2. *Пусть $F \subseteq P_2$, $F = [F]$, $0 \notin F$, $1 \in F$, $F \subset M$. Тогда найдется система функций A, удовлетворяющая условиям: $[A] = F$, $|A| < \infty$.*

Доказательство. Отдельно рассмотрим несколько случаев.

Случай 1°: $F \not\subseteq K$.

Тогда в классе F найдется функция $f_K \in M \setminus K$.

Случай 1.1°: $F \not\subseteq D$.

Тогда в классе F найдется функция $f_D \in M \setminus D$. Тогда $\omega \in [\{1, f_K, f_D\}] \subseteq F$.

Случай 1.1.1°: $F \subseteq O^\infty$.

Тогда $F = [\{1, \omega\}] = [\{1, f_K, f_D\}]$.

Случай 1.1.2°: $F \not\subseteq O^\infty$.

Если $f \in O^\infty$, то $f \in [\{1, \omega\}]$. Если $f \in F \setminus O^\infty$, то $f \in [\{1, \omega, d_{p(f)}\}]$. Положим

$$p(F) = \min_{f \in F \setminus O^\infty} p(f).$$

Обозначим через $\hat{f}$ некоторую функцию, удовлетворяющую условиям $\hat{f} \in F \setminus O^\infty$ и $p(\hat{f}) = p(F)$. Тогда

$$[\{1, f_K, f_D, \hat{f}\}] \subseteq F \subseteq [\{1, \omega, d_{p(F)}\}] \subseteq [\{1, f_K, f_D, \hat{f}\}].$$

Поэтому

$$F = [\{1, \omega, d_{p(F)}\}] = [\{1, f_K, f_D, \hat{f}\}].$$

Случай 1.2°: $F \subseteq D$.
Тогда $F = [\{1, x \vee y\}]$.
Случай 2°: $F \subseteq K$.
Случай 2.1°: $F \not\subseteq D$.
Тогда $F = [\{1, xy\}]$.
Случай 2.2°: $F \subseteq D$.
Тогда либо $F = \{1\}$, либо $F = \{1, x\}$. □

Следствие 2. *Список всех замкнутых классов монотонных функций, не содержащих* 0 *и содержащих* 1:

$$\{1\}, \{1, x\}, [\{1, xy\}], [\{1, x \vee y\}], [\{1, \omega\}] = MO^\infty, [\{1, \omega, d_p\}] \ (p = 2, 3, \ldots).$$

Лемма 4. *Если $f_L \notin L$, то найдется такая функция $h(x, y)$, что выполняются условия: $h(x, y) \notin L$, $h(x, y) \in [\{1, f_L\}]$.*

Доказательство. Воспользуемся доказанным ранее фактом (свойство 4 класса линейных функций), сформулированным относительно функции f_L^*, двойственной к функции f_L (функция f_L^* также будет нелинейна): если $f_L^* \notin L$, то найдется такая функция $g(x, y)$, что выполняются условия: $g(x, y) \notin L$, $g(x, y) \in [\{0, f_L^*\}]$.

В силу принципа двойственности в качестве искомой фунции $h(x, y)$ можно взять функцию $g^*(x, y)$. □

Обозначим через T_1 класс булевых функций, сохраняющих константу 1, т. е. класс всех функций, принимающих значение 1 в случае, когда все переменные принимают значение 1.

Свойства класса T_1:

1. $1, xy, x \vee y, x \to y \in T_1$; $0, x \oplus y \notin T_1$.
2. $[T_1] = T_1$.
3. $[\{x \to y, xy\}] = T_1$.

▷ По лемме Z. □

Лемма 5 (о порождении импликации)**.** *Если $f_M \in T_1 \setminus M$, $f_L \in T_1 \setminus L$, то $x \to y \in [\{1, f_M, f_L\}]$.*

Доказательство. Для немонотонной функции $f_M(x_1, \ldots, x_n)$ найдутся два набора $\tilde{\alpha} = (\alpha_1, \ldots, \alpha_n)$ и $\tilde{\beta} = (\beta_1, \ldots, \beta_n)$ из E^n, удовлетворяющие условиям: $\tilde{\alpha} \preccurlyeq \tilde{\beta}$, $f(\tilde{\alpha}) = 1$, $f(\tilde{\beta}) = 0$. Для $i = 1, \ldots, n$ при выполнении равенств $\alpha_i = 0$ и $\beta_i = 0$ (в силу включения $f_M \in T_1$ хотя бы для одного значения i эти равенства будут выполняться) в функцию $f_M(x_1, \ldots, x_n)$ вместо переменной x_i подставим переменную y, при выполнении равенств $\alpha_i = 1$ и $\beta_i = 1$ — константу 1, а при выполнении неравенства $\alpha_i < \beta_i$ — переменную x. Тем самым получаем следующую функцию $g(x, y)$:

x	y	$g(x,y)$
0	0	1
0	1	0 или 1
1	0	0
1	1	1

Таким образом, либо $g(x, y) = x \to y$, либо $g(x, y) = x \oplus y \oplus 1$. Во втором случае по предыдущей лемме из функции f_L получаем функцию $xy \oplus ax \oplus by \oplus c$, где $a \oplus b \oplus c = 0$, так как $f_L \in T_1$. Если получена функция $xy \oplus x \oplus 1$ (или $xy \oplus y \oplus 1$), то цель достигнута — это импликация. С другой стороны, и из функции xy, и из функции $xy \oplus x \oplus y$ с помощью функции $x \oplus y \oplus 1$ легко получить функцию $xy \oplus x \oplus 1$. □

Лемма 6 (о монотонной функции)**.** *Для произвольной булевой функции f найдется монотонная функция g, удовлетворяющая условиям $g \leq f$ и $g \in [\{1, x \vee y, f\}]$.*

Доказательство. Рассмотрим произвольную булеву функцию $f(x_1, \ldots, x_n)$.

Если $f \in M$, то в качестве искомой функции g можно взять саму функцию f. Далее считаем, что функция f немонотонна.

Если $n = 1$, то $f(x) = \overline{x}$. Тогда в качестве функции g можно взять константу 0. Далее будем считать, что $n \geq 2$.

Положим $N_f = \{\tilde{\alpha} \in E^n \mid f(\tilde{\alpha}) = 1\}$. Проведем доказательство индукцией по $|N_f|$.

При $|N_f| = 0$ функция f монотонна и поэтому утверждение леммы для нее выполняется.

Пусть утверждение леммы справедливо для всех функций f', удовлетворяющих условию $|N_{f'}| < |N_f|$. Докажем его для функции $f(x_1, \ldots, x_n)$. Можно считать, что $f \notin M$, причем она немонотонна по переменной x_1.

Положим

$$R = \{\tilde{\beta} \in E^{n-1} \mid f(0, \tilde{\beta}) > f(1, \tilde{\beta})\}.$$

Очевидно, что $R \neq \varnothing$.

Обозначим через $\chi_R(x_2, \ldots, x_n)$ характеристическую функцию множества R, т. е. $\chi_R(\tilde{\beta}) = 1$ тогда и только тогда, когда $\tilde{\beta} \in R$.

Положим

$$h(x_1, \ldots, x_n) = f(x_1, \ldots, x_n) \vee \chi_R(x_2, \ldots, x_n).$$

В силу леммы Z справедливо соотношение $h \in [\{1, x \vee y, f\}]$ (так как при добавлении константы 0 из немонотонной функции можно получить отрицание).

Положим $g_1(x_1, \ldots, x_n) = f(h(x_1, \ldots, x_n), x_2, \ldots, x_n)$.

Очевидно, что $g_1 \in [\{1, x \vee y, f\}]$.

Докажем неравенство $g_1 < f$.

Действительно, если в наборе $(\alpha, \tilde{\beta})$ из E^n поднабор $\tilde{\beta}$ принадлежит множеству R, то

$$g_1(0, \tilde{\beta}) = g_1(1, \tilde{\beta}) = f(1, \tilde{\beta}) = 0, \quad f(0, \tilde{\beta}) = 1.$$

Следовательно, $g_1(0, \tilde{\beta}) < f(0, \tilde{\beta})$, $g_1(1, \tilde{\beta}) = f(1, \tilde{\beta})$.

Если $\tilde{\beta} \notin R$, то функция $f(x_1, \tilde{\beta})$ монотонна по своей переменной x_1 и выполняются равенства

$$g_1(\alpha, \tilde{\beta}) = f(f(\alpha, \tilde{\beta}), \tilde{\beta}) = f(\alpha, \tilde{\beta}).$$

Справедливость последнего равенства следует из следующих соображений: если $f(\alpha, \tilde{\beta}) = 0$, то

$$f(f(\alpha, \tilde{\beta}), \tilde{\beta}) = f(0, \tilde{\beta}) \leq f(\alpha, \tilde{\beta})) = 0;$$

если $f(\alpha, \tilde{\beta}) = 1$, то

$$f(f(\alpha, \tilde{\beta}), \tilde{\beta}) = f(1, \tilde{\beta}) \geq f(\alpha, \tilde{\beta})) = 1.$$

Итак, доказано, что $g_1 < f$. Следовательно, верно неравенство $|N_{g_1}| < |N_f|$. Поэтому по предположению индукции найдется монотонная функция g, удовлетворяющая условиям: $g \leq g_1$, $g \in [\{1, x \vee y, g_1\}]$. В силу соотношений $g_1 \leq f$ и $g_1 \in [\{1, x \vee y, f\}]$ функция g удовлетворяет требованиям леммы. □

Теорема 3. *Пусть $F \subseteq P_2$, $F = [F]$, $0 \notin F$, $1 \in F$. Тогда найдется система функций A, удовлетворяющая условиям: $[A] = F$, $|A| < \infty$.*

Доказательство. Покажем, что $F \subseteq T_1$. Действительно, если это не так, то в классе F найдется функция, принимающая на единичном наборе значение 0. Но тогда верно включение $0 \in F$ — противоречие.

Отдельно рассмотрим несколько случаев.

Случай 1°: $F \subseteq M$.

Применяем теорему 2.

Случай 2°: $F \not\subseteq M$.

Случай 2.1°: $F \not\subseteq L$.

Тогда в классе F найдутся функции $f_M \notin M$ и $f_L \notin L$. По лемме о порождении импликации имеем: $x \to y \in [\{1, f_M, f_L\}]$.

Случай 2.1.1°: $F \subseteq O^\infty$.

Тогда $F = O^\infty = [\{x \to y\}] = [\{1, f_M, f_L\}]$.

Случай 2.1.2°: $F \not\subseteq O^\infty$.

Если $f \in O^\infty$, то $f \in [\{x \to y\}]$. Если $f \in F \setminus O^\infty$, то для функции f применяем лемму о монотонной функции: существует монотонная функция g_f, такая что $g_f \leq f$, $g_f \in [\{1, x \vee y, f\}]$. Кроме того, справедливо соотношение $g_f \notin O^\infty$, так как иначе бы функция f была бы не меньше некоторой своей переменной.

Положим

$$G_F = \bigcup_{f \in F \setminus O^\infty} g_f.$$

Очевидно, что $G_F \subseteq M \setminus O^\infty$. Положим

$$p(G_F) = \min_{g \in G_F} p(g).$$

По лемме 3 справедливо включение $G_F \subseteq [\{1, \omega, d_{p(G_F)}\}]$. Обозначим через $\hat{g}$ некоторую функцию, удовлетворяющую условиям $\hat{g} \in G_F$ и $p(\hat{g}) = p(G_F)$. Обозначим через $\hat{f}$ такую функцию из класса F, что $g_{\hat{f}} = \hat{g}$. Тогда

$$G_F \subseteq [\{1, \omega, d_{p(\hat{g})}\}] \subseteq [\{x \to y, d_{p(\hat{g})}\}] \subseteq [\{x \to y, \hat{g}\}] \subseteq [\{x \to y, \hat{f}\}].$$

Применяя лемму Z, имеем:

$$F \subseteq [\{x \to y\} \cup G_F] \subseteq [\{x \to y, d_{p(\hat{g})}\}] \subseteq$$
$$\subseteq [\{x \to y, \hat{f}\}] \subseteq [\{1, f_M, f_L, \hat{f}\}] \subseteq F.$$

Случай 2.2°: $F \subseteq L$.

Так как немонотонная функция, лежащая в классе F, не может быть отрицанием (в силу соотношения $\overline{x} \notin T_1$), то в классе F есть функция, существенно зависящая не менее чем от двух переменных. Поэтому $F = [\{x \oplus y \oplus 1\}]$. □

Следствие 3. *Список всех замкнутых классов функций, не содержащих* 0*, содержащих* 1 *и не лежащих в классе* M :

$$[\{x \oplus y \oplus 1\}], [\{x \to y\}], [\{x \to y, d_p\}] \ (p = 2, 3, \ldots).$$

Следствие 4. *Список всех замкнутых классов функций, не содержащих* 0*, содержащих* 1 *и не лежащих в классе* M :

$$L \cap T_1, O^\infty, T_1, O^\mu \ (\mu = 2, 3, \ldots).$$

Лекция № 4

1.1.4 Классы функций, не содержащие констант

Приведем пример такого класса — класс S самодвойственных функций.

Функция $f(x_1, \ldots, x_n)$ является *самодвойственной* (т. е. $f \in S$) тогда и только тогда, когда функция f совпадает с двойственной к ней функцией f^*: $f(x_1, \ldots, x_n) \equiv \overline{f(\overline{x}_1, \ldots, \overline{x}_n)}$.

Свойства класса S:

1. $x, \overline{x}, x \oplus y \oplus z, d_3 \in S$; $0, 1, xy, x \vee y \notin S$.
2. $[S] = S$.
3. Если $f(x, y) \in S$, то $f(x, y) \in \{x, \overline{x}, y, \overline{y}\}$.
4. Самодвойственная функция от n переменных обращается в единицу на 2^{n-1} наборе.
5. (Лемма о несамодвойственной функции.) Если $f_S \notin S$, то $\{0, 1\} \subseteq [\{f_S, \overline{x}\}]$ (т. е. из несамодвойственной функции и отрицания можно получить обе константы).

$\triangleright$ Для несамодвойственной функции $f(x_1, \ldots, x_n)$ найдется набор $(\alpha_1, \ldots, \alpha_n)$, удовлетворяющей условию $f(\alpha_1, \ldots, \alpha_n) = f(\overline{\alpha}_1, \ldots, \overline{\alpha}_n)$. Теперь для $i = 1, \ldots, n$, подставив в функцию $f(x_1, \ldots, x_n)$ вместо переменной x_i переменную x, если $\alpha_i = 0$, и функцию $\overline{x}$, если $\alpha_i = 1$, получим константу. Имея отрицание, получим и вторую константу. $\square$

Прежде чем перейти к теореме о конечнопорожденности замкнутых классов, не содержащих констант, сформулируем очень простой, но полезный факт.

Лемма 7. *Пусть* $[A] = F$, $|A| < \infty$, $[B] = F$. *Тогда найдется система функций* B_0, *удовлетворяющая условиям* $B_0 \subseteq B$, $[B_0] = F$ *и* $|B_0| < \infty$.

Теорема 4. *Пусть* $F \subseteq P_2$, $0 \notin F$, $1 \notin F$. *Тогда найдется система функций* A, *удовлетворяющая условиям:* $[A] = F$, $|A| < \infty$.

Доказательство. Отдельно рассмотрим два случая.

Случай 1°: $F \subseteq S$.

Положим $F_1 = [F \cup \{1\}]$. Применяя либо теорему 1, либо теорему 3, получаем, что найдется система функций A_1, удовлетворяющая условиям: $[A_1] = F_1$, $|A_1| < \infty$.

По лемме 7 из системы $F \cup \{1\}$ можно извлечь конечную полную в F_1 подсистему (обязательно содержащую константу 1, так как иначе подсистема порождала бы подкласс самодвойственных функций) $A \cup \{1\}$. Докажем, что $[A] = F$.

Пусть $f(x_1, \ldots, x_n) \in F$ и формула $\Phi(x_1, \ldots, x_n)$ реализует функцию f над системой $A \cup \{1\}$. Заменим все вхождения константы 1 в формуле Φ на новую переменную x_{n+1}. Полученная таким образом формула $\mathrm{H}(x_1, \ldots, x_n, x_{n+1})$ над системой A реализует некоторую функцию $h(x_1, \ldots, x_n, x_{n+1})$. Покажем, что $h(x_1, \ldots, x_n, x_{n+1}) = f(x_1, \ldots, x_n)$.

Действительно, при $x_{n+1} = 1$ равенство очевидно, а при $x_{n+1} = 0$ в силу самодвойственности функций f и h имеем:

$$h(x_1, \ldots, x_n, 0) = \overline{h(\overline{x}_1, \ldots, \overline{x}_n, 1)} = \overline{f(\overline{x}_1, \ldots, \overline{x}_n)} = f(x_1, \ldots, x_n).$$

Случай 2°: $F \not\subseteq S$.

Тогда справедливо включение $F \subseteq T_0 \cap T_1$. Действительно, иначе нашлась бы функция $f \in F$, удовлетворяющая условию $f(x, \ldots, x) \in \{0, 1, \overline{x}\}$. Таким образом, учитывая лемму о несамодвойственной функции, получаем, что класс F содержит константу, что противоречит условию теоремы.

Докажем также, что хотя бы одна из функций xy и $x \vee y$ содержится в классе F. Для этого рассмотрим функцию $f_S(x_1, \ldots, x_n) \in F \setminus S$. Найдется набор $(\alpha_1, \ldots, \alpha_n)$, удовлетворяющей условию $f_S(\alpha_1, \ldots, \alpha_n) = f_S(\overline{\alpha}_1, \ldots, \overline{\alpha}_n)$. Этот набор не может состоять только из нулей или только из единиц, так как справедливо включение $f_S \in T_0 \cap T_1$. Теперь для $i = 1, \ldots, n$, подставив в функцию $f_S(x_1, \ldots, x_n)$ вместо переменной x_i переменную x, если $\alpha_i = 0$, и переменную y, если $\alpha_i = 1$, получим одну из искомых функций.

Ввиду двойственности случаев без ограничения общности будем считать, что $x \vee y \in F$.

Положим $F_1 = [F \cup \{1\}]$. Применяя либо теорему 1, либо теорему 3, получаем, что найдется система функций A_1, удовлетворяющая условиям: $[A_1] = F_1$, $|A_1| < \infty$.

По лемме 7 из системы $F \cup \{1\}$ можно извлечь конечную полную в F_1 подсистему (обязательно содержащую константу 1, так как иначе подсистема порождала бы подкласс класса T_0) $B \cup \{1\}$. Докажем, что

конечная система $A = B \cup \{x \vee y\}$ обладает нужным свойством, т. е. $[A] = F$.

Пусть $f(x_1, \ldots, x_n) \in F$ и формула $\Phi(x_1, \ldots, x_n)$ реализует функцию f над системой $B \cup \{1\}$. Заменим все вхождения константы 1 в формуле Φ на переменную y. Полученная таким образом формула $\mathrm{H}(x_1, \ldots, x_n, y)$ над системой B реализует некоторую функцию $h(x_1, \ldots, x_n, y)$. Покажем, что $h(x_1, \ldots, x_n, x_1 \vee \ldots \vee x_n) = f(x_1, \ldots, x_n)$.

Действительно, для ненулевого набора значений переменных $x_1, \ldots, x_n$ равенство очевидно, а на нулевом наборе и левая, и правая части доказываемого равенства обращаются в 0, так как $f, h \in T_0$. $\square$

1.1.5 Заключительные утверждения

Теорема 5 (Э. Пост)**.** *Каждый замкнутый класс булевых функций конечнопорожден.*

Доказательство. Следует из теорем 1, 3, 4 с применением принципа двойственности. $\square$

Следствие 5. *Семейство замкнутых классов булевых функций счетно.*

Наконец, сформулируем условия, при которых константы принадлежат или не принадлежат замкнутому классу.

Утверждение 1. *Пусть $A \subseteq P_2$. Тогда:*

1. *Константы* 0 *и* 1 *содержатся в классе $[A]$ тогда и только тогда, когда существуют функции f_{T_0}, f_{T_1} и f_S, такие что $f_{T_0} \in A \setminus T_0$, $f_{T_1} \in A \setminus T_1$, $f_S \in A \setminus S$.*

2. *Константа* 0 *не содержится, а константа* 1 *содержится в классе $[A]$ тогда и только тогда, когда выполняется условие $A \subseteq T_1$ и существует функция f_{T_0}, такая что $f_{T_0} \in A \setminus T_0$.*

3. *Константа* 0 *содержится, а константа* 1 *не содержится в классе $[A]$ тогда и только тогда, когда выполняется условие $A \subseteq T_0$ и существует функция f_{T_1}, такая что $f_{T_1} \in A \setminus T_1$.*

4. *Константы* 0 *и* 1 *не содержатся в классе $[A]$ тогда и только тогда, когда выполняется хотя бы одно из условий $A \subseteq S$ или $A \subseteq T_0 \cap T_1$.*

Упражнение 8. Доказать утверждение 1.

§ 1.2 Пример Мучника

Теорема Поста и следствие к ней описывают важнейшие свойства булевых функций: каждый замкнутый класс булевых функций конечнопорожден и семейство замкнутых классов булевых функций счетно. Верны ли аналогичные утверждения для семейства замкнутых классов функций k-значной логики[3])? Отрицательный ответ на эти вопросы при $k \geq 3$ был получен в 1959 году в работе Ю. И. Янова и А. А. Мучника.

Теорема 6 (А. А. Мучник)**.** *При $k \geq 3$ семейство замкнутых классов функций k-значной логики континуально.*

Доказательство. Количество замкнутых классов не более числа различных подмножеств счетного множества, т. е. семейство замкнутых классов функций k-значной логики не более чем континуально.

Построим континуальное семейство замкнутых классов. Для $s = 2, 3, \ldots$ определим (симметрическую) функцию $f_s(x_1, \ldots, x_s)$ следующим образом. Функция f_s принимает значение 1 на наборах, состоящих из $s-1$ двойки и одной единицы, и значение 0 на всех остальных наборах. Положим

$$F = \bigcup_{i=2}^{\infty} \{f_i\}.$$

Покажем, что для любого $s \geq 2$ верно соотношение $f_s \notin [F \setminus \{f_s\}]$.

Пусть это не так, т. е. найдется реализующая функцию f_s формула:

$$f_s(x_1, \ldots, x_s) = f_n(A_1, \ldots, A_n).$$

Рассмотрим три случая.

Случай 1°: среди подформул $A_1, \ldots, A_n$ не менее двух нетривиальных подформул. Тогда в формуле $f_n(A_1, \ldots, A_n)$ не менее двух аргументов всегда отличны от 2, поэтому вся формула обращается в 0 на любом наборе переменных — противоречие.

Случай 2°: среди подформул $A_1, \ldots, A_n$ в точности одна является нетривиальной подформулой. Тогда в формуле $f_n(A_1, \ldots, A_n)$ в силу неравенства $n \geq 2$ среди подформул $A_1, \ldots, A_n$ есть хотя бы одна тривиальная подформула. Без ограничения общности будем считать, что это переменная x_1. Но в этом случае на наборе $(1, 2, \ldots, 2)$ формула $f_n(A_1, \ldots, A_n)$ обращается в 0 — противоречие.

[3]) Все переменные функций k-значной логики, а также сами функции, принимают значение из множества $\{0, 1, \ldots, k-1\}$. Для функций k-значной логики точно так же как и для булевых функций можно ввести понятия формулы, суперпозиции, замыкания, замкнутого класса и т. д., подробнее на этом останавливаться не имеет смысла.

Случай 3°: все подформулы $A_1, \ldots, A_n$ являются тривиальными подформулами, т. е. переменными. Тогда выполняется неравенство $n > s$ и поэтому найдется переменная, скажем x_1, которая встречается среди подформул $A_1, \ldots, A_n$ не менее двух раз. Следовательно на наборе $(1, 2, \ldots, 2)$ формула $f_n(A_1, \ldots, A_n)$ обращается в 0 — противоречие.

Обозначим множество бесконечных ненулевых двоичных последовательностей через R. Для произвольной последовательности $\tilde{\alpha} = (\alpha_1, \ldots, \alpha_s, \ldots)$ положим

$$F_{\tilde{\alpha}} = \bigcup_{i\colon \alpha_i = 1} \{f_{i+1}\}.$$

Если $\tilde{\alpha}, \tilde{\beta} \in R$, $\tilde{\alpha} \neq \tilde{\beta}$, то $[F_{\tilde{\alpha}}] \neq [F_{\tilde{\beta}}]$. Поэтому семейство замкнутых классов $\{[F_{\tilde{\alpha}}] \mid \tilde{\alpha} \in R\}$ континуально. □

Следствие 6. *При $k \geq 3$ семейство замкнутых классов функций k-значной логики содержит классы, не являющиеся конечнопорожденными.*

§ 1.3 Инвариантные классы Яблонского

Итак, доказано, что семейство замкнутых классов булевых функций счетно, а семейство замкнутых классов функций k-значной логики при $k \geq 3$ континуально. Однако, если изменить определения замыкания и замкнутого класса, то можно получить и в случае булевых функций континуальное семейство замкнутых (в новом смысле) классов.

Вспомним общематематическое (алгебраическое) определение оператора замыкания.

Пусть $(P, \leq)$ — частично упорядоченное множество. Оператор $Z\colon P \to P$ называется *оператором замыкания*, если выполняются три условия (аксиомы):

1) для любого $x \in P$ выполняется соотношение $x \leq Z(x)$ (экстенсивность);

2) для любых $x, y \in P$ из соотношения $x \leq y$ вытекает соотношение $Z(x) \leq Z(y)$ (монотонность);

3) для любого $x \in P$ справедливо равенство $Z(Z(x)) = Z(x)$ (идемпотентность).

Очевидно, что рассматриваемый выше оператор $A \to [A]$ (оператор замыкания относительно операций суперпозиции и добавления фиктивной переменной, действующий из 2^{P_2} в 2^{P_2}) действительно является оператором замыкания.

Теперь рассмотрим следующий набор операций над булевыми функциями:

1) добавление и изъятие фиктивных переменных;

2) переименование переменных без отождествления;

3) подстановка вместо переменных констант.

Нетрудно понять, что оператор, действующий из 2^{P_2} в 2^{P_2} и добавляющий к исходному множеству булевых функций все функции, которые получаются из них применением трех описанных операций, удовлетворяет всем условиям оператора замыкания. Изучим классы булевых функций, замкнутые (инвариантные) относительно этого оператора.

Напомним, что булевы функции называются *равными*, если у них совпадают множества переменных, от которых они зависят существенно, и на каждом наборе этих переменных функции принимают одинаковые значения. Таким образом, равные функции получаются друг из друга путем добавления или изъятия фиктивных переменных.

Булевы функции называются *конгруэнтными* (или *подобными*), если они получаются одна из другой переименованием переменных без отождествления.

Подфункцией булевой функции называется любая функция, получающаяся из исходной функции подстановкой вместо каких-либо переменных произвольных констант.

Множество булевых функций Q называется *инвариантным классом Яблонского*, если для любой функции из класса Q классу Q также принадлежат:

1) все равные ей функции;

2) все конгруэнтные (подобные) ей функции;

3) все ее подфункции.

Примеры инвариантных классов: L, M, $KSym$ (класс квазисимметрических, т. е. симметрических по всем своим существенным переменным, булевых функций), P^n (класс всех булевых функций, существенно зависящих не более чем от n переменных). Отметим, что пустой класс также формально является инвариантным классом. В дальнейшем будем рассматривать только непустые инвариантные классы, не всегда оговаривая это отдельно.

Свойства инвариантных классов Яблонского:

1. Пусть Q — инвариантный класс и класс Q содержит функцию, отличную от константы. Тогда $\{0,1\} \subseteq Q$.

2. Пусть $Q = [Q]$, $Q \neq \varnothing, \{0\}, \{1\}$. Тогда класс Q является инвариантным в том и только том случае, когда выполняется включение $\{0,1\} \subseteq Q$.

3. Замкнутый класс не всегда является инвариантным (примеры: S, T_1, O^∞).

4. Инвариантный класс не всегда является замкнутым (пример: класс, состоящий из функции $x \vee y$ и функций, получающийся из нее добавлением и изъятием фиктивных переменных, переименованием переменных без отождествления и подстановкой вместо переменных констант).

Теорема 7. *Семейство инвариантных классов Яблонского континуально.*

Доказательство. Количество инвариантных классов не более числа различных подмножеств счетного множества, т. е. семейство инвариантных классов булевых функций не более чем континуально.

Построим континуальное семейство инвариантных классов. Для $k = 2, 3, \ldots$ положим

$$f_k(x_1, \ldots, x_k) = x_1 \ldots x_k \vee \overline{x}_1 \ldots \overline{x}_k.$$

Обозначим множество бесконечных ненулевых двоичных последовательностей через R. Для произвольной последовательности $\tilde{\alpha} = (\alpha_1, \ldots, \alpha_s, \ldots)$ положим

$$F_{\tilde{\alpha}} = \bigcup_{i:\, \alpha_i = 1} \{f_{i+1}\}.$$

Через $Q_{\tilde{\alpha}}$ обозначим инвариантный класс, который получается из функций множества $F_{\tilde{\alpha}}$ добавлением и изъятием фиктивных переменных, переименованием переменных без отождествления и подстановкой вместо переменных констант.

В классе $Q_{\tilde{\alpha}}$ найдется функция, существенно зависящая от всех своих $k+1$ переменной и принимающая значение 1 ровно на двух наборах этих переменных, тогда и только тогда, когда $\alpha_k = 1$. Поэтому если $\tilde{\alpha}, \tilde{\beta} \in R$, $\tilde{\alpha} \neq \tilde{\beta}$, то $Q_{\tilde{\alpha}} \neq Q_{\tilde{\beta}}$. Следовательно семейство инвариантных классов $\{Q_{\tilde{\alpha}} \mid \tilde{\alpha} \in R\}$ континуально. □

Положим $Q(n) = Q \cap P_2(n)$.

Лемма 8. *Пусть Q — непустой инвариантный класс Яблонского. Тогда последовательность*

$$\sqrt[2^n]{|Q(n)|}, \quad n = 1, 2, \ldots$$

не возрастает и для любого n выполняются неравенства

$$1 \leq \sqrt[2^n]{|Q(n)|} \leq 2.$$

Доказательство. Для функции $f(x_1, \ldots, x_n, x_{n+1})$ из класса Q справедливо разложение

$$f(x_1, \ldots, x_n, x_{n+1}) = f(x_1, \ldots, x_n, 0)\overline{x}_{n+1} \vee f(x_1, \ldots, x_n, 1)x_{n+1},$$

где функции $f(x_1, \ldots, x_n, 0)$ и $f(x_1, \ldots, x_n, 1)$ также принадлежат классу Q. Поэтому $|Q(n+1)| \leq |Q(n)|^2$. Следовательно,

$$\sqrt[2^{n+1}]{|Q(n+1)|} \leq \sqrt[2^{n+1}]{|Q(n)|^2} = \sqrt[2^n]{|Q(n)|},$$

т. е. последовательность $\sqrt[2^n]{|Q(n)|}$ не возрастает.

Справедливость неравенств $1 \leq \sqrt[2^n]{|Q(n)|} \leq 2$ следует из очевидных соотношений $1 \leq |Q(n)| \leq 2^{2^n}$. □

В силу леммы 8 у последовательности $\sqrt[2^n]{|Q(n)|}$ существует предел, не превосходящий никакого члена последовательности. Число

$$\sigma_Q = \log_2 \left(\lim_{n\to\infty} \sqrt[2^n]{|Q(n)|} \right)$$

называется *параметром инвариантного класса* Q.

Свойства параметра σ_Q:

1. Верно равенство

$$\sigma_Q = \lim_{n\to\infty} \frac{\log_2 |Q(n)|}{2^n}.$$

2. Для любого инвариантного класса Q справедливы неравенства $0 \leq \sigma_Q \leq 1$.

3. Если $\sigma_Q \neq 0$, то

$$|Q(n)| = 2^{\sigma_Q 2^n (1+\varepsilon_Q(n))},$$

где $\varepsilon_Q(n) \to 0$ при $n \to \infty$; если $\sigma_Q = 0$, то

$$|Q(n)| = 2^{2^n \delta_Q(n)},$$

где $\delta_Q(n) \to 0$ при $n \to \infty$

4. Существует единственный инвариантный класс Q с параметром $\sigma_Q = 1$ — это класс P_2.

▷ Пусть инвариантный класс Q отличен от класса P_2. Тогда найдется $f(x_1, \ldots, x_n) \notin Q$ и следовательно выполняются соотношения $2^{\sigma_Q} \leq \sqrt[2^n]{|Q(n)|} < \sqrt[2^n]{2^{2^n}} = 2$. □

Теперь покажем, что инвариантные классы L, M, $KSym$ — нулевые (т. е. имеют параметр, равный нулю). Действительно, с использованием свойства 9 монотонных функций при $n \to \infty$ имеем:

$$\log_2 |L(n)| = n + 1 = o\left(2^n\right),$$

$$\log_2 |M(n)| \le \log_2 \left(n^{C_n^{\lfloor n/2 \rfloor}}\right) = \frac{n! \log_2 n}{\lceil n/2 \rceil! \lfloor n/2 \rfloor!} = O\left(\frac{2^n \log_2 n}{n^{1/2}}\right) = o\left(2^n\right),$$

$$\log_2 |KSym(n)| \le \log_2 \left(\sum_{k=0}^{n} 2^{k+1} C_n^k\right) = \log_2 \left(2(1+2)^n\right) = o\left(2^n\right).$$

Лекция № 5

Построим инвариантный класс Яблонского, у которого значение параметра равно $1/2$.

Обозначим через $Q_{1/2}$ класс, состоящий из функций $f(x_1, \ldots, x_n)$, $n = 1, 2, \ldots$, представимых в виде

$$f(x_1, \ldots, x_n) = f_1(x_{i_1}, \ldots, x_{i_k}) f_2(x_{i_1}, \ldots, x_{i_k}),$$

где $f_1(x_{i_1}, \ldots, x_{i_k}) = x_{i_1} \oplus \ldots \oplus x_{i_k} \oplus c$, а $f_2(x_{i_1}, \ldots, x_{i_k})$ — произвольная булева функция, все существенные переменные которой содержатся среди переменных $x_{i_1}, \ldots, x_{i_k}$.

Очевидно, что $Q_{1/2}$ — инвариантный класс. Оценим величину $|Q_{1/2}(n)|$.

С одной стороны, в классе $Q_{1/2}$ содержатся все функции вида $(x_1 \oplus \ldots \oplus x_n) g(x_1, \ldots, x_n)$, число которых $2^{2^{n-1}}$, так как функцию $g(x_1, \ldots, x_n)$ можно задавать произвольным образом на всех наборах, на которых функция $(x_1 \oplus \ldots \oplus x_n)$ обращается в единицу, т. е. на 2^{n-1} наборах. Поэтому

$$|Q_{1/2}(n)| \ge 2^{2^{n-1}}.$$

С другой стороны, справедливо неравенство

$$|Q_{1/2}(n)| \le 2^{n+1} 2^{2^{n-1}},$$

так как количество возможных вариантов для выбора функции f_1, т. е. число линейных функций от n переменных, равно 2^{n+1}, а число возможных вариантов для выбора функции f_2 при фиксированной функции f_1 равно $2^{2^{k-1}}$, где k — число существенных переменных у функции f_1, и $k \le n$.

Используя свойство 1, имеем:

$$\frac{1}{2} = \lim_{n\to\infty} \frac{\log_2\left(2^{2^{n-1}}\right)}{2^n} \leq \sigma_{Q_{1/2}} \leq \lim_{n\to\infty} \frac{\log_2\left(2^{n+1}2^{2^{n-1}}\right)}{2^n} = \frac{1}{2}.$$

В заключение отметим, что для любого σ, $0 \leq \sigma < 1$, семейство инвариантных классов Яблонского с параметром σ континуально. Доказательство этого нетривиального факта содержится, например, в работе С. В. Яблонского «Об алгоритмических трудностях синтеза минимальных контактных схем» (сб. «Проблемы кибернетики», 1959, вып. 2, с. 75–121).

Глава 2

Элементы комбинаторики

§ 2.1 Определения и простейшие свойства основных комбинаторных объектов

С чего начинается комбинаторика? В каком-то смысле она начинается с ответа на вопрос: сколькими способами можно среди n элементов выбрать k элементов? На самом деле это не один вопрос, а четыре — выборка может быть упорядоченной и неупорядоченной, повторный выбор одного и того же элемента может допускаться, а может не допускаться. Упорядоченные выборки называются *размещениями* или *наборами*, а неупорядоченные — *сочетаниями*. Как правило, по умолчанию под выборкой понимается выборка без повторений, а если речь идет о выборке с повторениями, то это оговаривается явным образом.

Итак, пусть есть n-элементное множество, например, $M_n = \{1, 2, \ldots, n\}$. Дадим ответ на вопрос о числе выборок k элементов из этого множества.

1. Упорядоченные выборки с повторениями (размещения с повторениями).

В этом случае число выборок равно числу k-значных чисел в системе счисления по основанию n, т. е. равно

$$n^k.$$

2. Упорядоченные выборки без повторений (размещения).

Первый элемент можно выбрать n способами, второй элемент — $n-1$ способами, ..., k-й элемент — $n-k+1$ способами. Таким образом общее число способов выбора равно

$$n(n-1)\ldots(n-k+1) = \frac{n!}{(n-k)!}.$$

3. Неупорядоченные выборки без повторений (сочетания).

Так как одной неупорядоченной выборке k элементов без повторений соответствует $k!$ упорядоченных выборок без повторений, то число сочетаний из n элементов по k элементов, обозначаемое C_n^k, при $n \geq k \geq 0$ равно

$$\frac{n!}{(n-k)!\,k!}.$$

4. Неупорядоченные выборки с повторениями (сочетания с повторениями).

Между множеством всех сочетаний с повторениями из n элементов по k элементов и множеством всех двоичных наборов длины $n+k-1$ с k нулями следующим образом можно установить взаимно однозначное соответствие: выборке, в которой k_1 единиц, k_2 двоек, ..., k_n чисел n, $k_i \geq 0$ $(i = 1, \ldots, n)$, $k_1 + \ldots + k_n = k$, соответствует набор, в котором сначала расположены k_1 нулей, а затем последовательно для $i = 2, \ldots, n$ расположены наборы, состоящие из единицы и стоящих вслед за ней k_i нулей:

$$\underbrace{0\ldots0}_{k_1}1\underbrace{0\ldots0}_{k_2}1\ldots1\underbrace{0\ldots0}_{k_n}.$$

Таким образом, число искомых выборок равно числу C_{n+k-1}^k сочетаний из $n+k-1$ элементов по k элементов, т. е. числу

$$\frac{(n+k-1)!}{(n-1)!\,k!}.$$

Свойства сочетаний:

1. Если $k > n$, то $C_n^k = 0$.

2. Для любого целого k, $0 \leq k \leq n$, справедливо равенство $C_n^k = C_n^{n-k}$.

3. Функция $f(k) = C_n^k$ целочисленной переменной k возрастает на множестве $\{0, \ldots, \lfloor n/2 \rfloor\}$ и убывает на множестве $\{\lceil n/2 \rceil, \ldots, n\}$.

4. Выполняется равенство

$$C_n^k = C_{n-1}^{k-1} + C_{n-1}^k.$$

5. Для любого ненулевого x и произвольного целого неотрицательного n справедливо равенство (бином Ньютона):

$$(1+x)^n = \sum_{k=0}^{n} C_n^k x^k.$$

6. Верны равенства

$$\sum_{k=0}^{n} C_n^k = 2^n,$$

$$\sum_{k=0}^{n} (-1)^k C_n^k = \begin{cases} 1, & \text{если } n = 0; \\ 0, & \text{если } n \geq 1, \end{cases},$$

$$\sum_{k=0}^{t} (-1)^k C_n^k = (-1)^t C_{n-1}^t \quad (n \geq 1,\ t \geq 0).$$

Свойства сочетаний с повторениям:

1. Число неотрицательных целочисленных решений уравнения

$$z_1 + z_2 + \ldots + z_n = k$$

равно числу сочетаний с повторениями из n элементов по k элементов, т. е. равно C_{n+k-1}^k.

2. Число монотонных отображений[1]) из множества $\{1, 2, \ldots, k\}$ в множество $\{1, 2, \ldots, n\}$ равно числу сочетаний с повторениями из n элементов по k элементов, т. е. равно C_{n+k-1}^k.

$\triangleright$ Между множеством из всех C_{n+k-1}^k двоичных наборов длины $n+k-1$ с k единицами и множеством монотонных отображений из множества $\{1, 2, \ldots, k\}$ в множество $\{1, 2, \ldots, n\}$ следующим образом установим взаимно однозначное соответствие. Пусть в наборе длины $n+k-1$ с k единицами число нулей до первой единицы равно r_1, число нулей, расположенных между единицами с номерами $i-1$ и i, $i = 2, \ldots, k$, равно r_i:

$$\underbrace{0\ldots0}_{r_1}1\underbrace{0\ldots0}_{r_2}1\ldots1\underbrace{0\ldots0}_{r_k}10\ldots0.$$

Соответствующее этому набору монотонное отображение f из множества $\{1, 2, \ldots, k\}$ в множество $\{1, 2, \ldots, n\}$ задается таким образом:

$$f(s) = 1 + \sum_{i=1}^{s} r_i, \quad s = 1, 2, \ldots, k.$$

Легко понять, что и по монотонному отображению исходный набор восстанавливается однозначно. $\square$

[1]) Отображение f называется *монотонным*, если для любых a и b из области определения отображения f из неравенства $a \leq b$ следует неравенство $f(a) \leq f(b)$.

3. Для любого натурального n при $0 < |x| < 1$ справедливо равенство

$$(1-x)^{-n} = \sum_{k=0}^{\infty} C_{n+k-1}^{k} x^k.$$

▷ Установим коэффициент при слагаемом x^k после раскрывания скобок и приведения подобных слагаемых в выражении

$$\underbrace{(1+x+x^2+\ldots)\ldots(1+x+x^2+\ldots)}_{n \text{ раз}}.$$

Если какое-либо выражение в скобках не дает «вклад» в конкретное слагаемое x^k, то считаем, что это выражение не входит в выборку, а если дает вклад x^s, $1 \leq s \leq k$, то — входит с кратностью s. Тогда коэффициент при x^k совпадает с числом сочетаний с повторениями из n элементов по k элементов. □

Теперь преобразуем выражение из свойства 3 сочетаний с повторениями следующим образом:

$$(1-x)^{-n} = 1 + \sum_{k=1}^{\infty} \frac{(n+k-1)(n+k-2)\ldots n}{k!} x^k = \\ = 1 + \sum_{k=1}^{\infty} \frac{(-n)(-n-1)\ldots(-n-k+1)}{k!} (-x)^k.$$

Таким образом, формулы из свойства 5 сочетаний и из свойства 3 сочетаний с повторениями являются частными случаями известного из курса математического анализа и справедливого для любого действительного α при $0 < |x| < 1$ разложения

$$(1+x)^{\alpha} = \sum_{k=0}^{\infty} \binom{\alpha}{k} x^k,$$

где

$$\binom{\alpha}{0} = 1; \qquad \binom{\alpha}{k} = \frac{\alpha(\alpha-1)\ldots(\alpha-k+1)}{k!} \quad (k = 1, 2, \ldots).$$

Очевидно, что для всех целых неотрицательных n и k верно равенство

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

§ 2.2 Формула включений-исключений

Пусть есть N предметов и свойства $p_1, \ldots, p_n$. Каждый предмет может одними свойствами обладать, а другими не обладать. Обозначим через $N_{i_1,\ldots,i_k}$ количество предметов, которые обладают свойствами $p_{i_1}, \ldots, p_{i_k}$ (обладание остальными свойствами — произвольное).

Через $N(r)$ обозначим число предметов, обладающих ровно r свойствами. Положим

$$S_0 = N, \quad S_k = \sum_{1 \le i_1 < \ldots < i_k \le n} N_{i_1,\ldots,i_k} \quad (k = 1, \ldots, n).$$

Здесь особо отметим, что S_k — просто удобные обозначения, способствующие уменьшению громоздкости выкладок, не стоит за ними усматривать какой-то сакраментальный смысл.

Теорема 8. *Справедливо равенство*

$$N(0) = \sum_{k=0}^{n} (-1)^k S_k.$$

Доказательство. Покажем, что каждый предмет дает одинаковый вклад при подсчете левой и правой частей устанавливаемого равенства.

Пусть предмет не обладает ни одним свойством. Тогда вклад в левую часть будет равен 1, вклад в правую часть будет ненулевым только в слагаемое $S_0 = N$, соответствующее мощности множества всех предметов, и этот вклад тоже равен 1.

Пусть предмет обладает s свойствами, $1 \le s \le n$, и это свойства $p_{j_1}, \ldots, p_{j_s}$. Тогда данный предмет дает ненулевой (единичный) вклад в слагаемое $N_{i_1,\ldots,i_t}$ тогда и только тогда, когда верно включение

$$\{i_1, \ldots, i_t\} \subseteq \{j_1, \ldots j_s\}.$$

Вклад этого предмета в левую часть равен нулю, а в правую —

$$\sum_{t=1}^{s} (-1)^t C_s^t,$$

т. е. вклад тоже нулевой.

Суммируя по всем предметам доказанные равенства вкладов в левую и правую часть, получаем справедливость исходного равенства. □

Переформулируем формулу включений-исключений в терминах множеств.

Следствие 7. *Справедливо равенство*

$$|A_1 \cup \ldots \cup A_n| = \sum_{i=1}^{n} |A_i| - \sum_{1 \le i_1 < i_2 \le n} |A_{i_1} \cap A_{i_2}| + \ldots$$
$$\ldots + (-1)^{k+1} \sum_{1 \le i_1 < \ldots < i_k \le n} |A_{i_1} \cap \ldots \cap A_{i_k}| + \ldots + (-1)^{n+1} |A_{i_1} \cap \ldots \cap A_{i_n}|.$$

Упражнение 9 (задача о беспорядках). Найти точное значение числа подстановок σ симметрической группы S_n, удовлетворяющих условию $\sigma(i) \neq i$ для всех i, $1 \le i \le n$.

Что произойдет, если сумму из правой части формулы включений-исключений оборвать на каком-либо слагаемом? Оказывается, что если последнее выписанное слагаемое положительное (точнее, оно соответствует четному числу свойств), то получается оценка величины $N(0)$ сверху, а в противном случае — снизу.

Теорема 9 (Неравенства Бонферрони)**.** *Для любого* l, $0 \le l \le \lfloor (n-1)/2 \rfloor$

$$\sum_{k=0}^{2l+1} (-1)^k S_k \le N(0) \le \sum_{k=0}^{2l} (-1)^k S_k.$$

Доказательство. Для каждого из неравенств подобно доказательству формулы включений-исключений достаточно установить соответствующее неравенство для вкладов в левую и правую часть доказываемого соотношения каждого из предметов. Нужное неравенство легко следует из равенства

$$\sum_{k=0}^{t} (-1)^k C_n^k = (-1)^t C_{n-1}^t$$

(третье равенство свойства 6 сочетаний). □

Также аналогично формуле включений-исключений можно установить справедливость следующих утверждений.

Теорема 10. *Справедливы равенства*

$$N(r) = \sum_{k=r}^{n} (-1)^{k-r} \binom{k}{r} S_k,$$

$$N_{\ge r} = \sum_{k=r}^{n} (-1)^{k-r} \binom{k-1}{r-1} S_k,$$

где $N_{\ge r}$ *— количество предметов, обладающих не менее чем* r *свойствами.*

Упражнение 10. Доказать теорему 10.

Обозначим через $\varphi(m)$ функцию Эйлера, численно равную количеству натуральных чисел, не превосходящих m и взаимнопростых с m.

Теорема 11. *Пусть* $m = p_1^{k_1} \dots p_n^{k_n}$ *— разложение числа* m *на простые множители. Тогда*

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

Доказательство. Применим формулу включений-исключений, положив $N = m$ и считая i-м свойством делимость на p_i, $i = 1, \dots, n$. Тогда имеем:

$$\varphi(m) = N(0) = m - \sum_{1\le i\le n}\frac{m}{p_i} + \dots + (-1)^k \sum_{1\le i_1<\dots i_k\le n}\frac{m}{p_{i_1}\dots p_{i_k}} + \dots$$
$$\dots + (-1)^n\frac{m}{p_1\dots p_n} = m\left(1 - \frac{1}{p_1}\right)\dots\left(1 - \frac{1}{p_n}\right).$$

□

§ 2.3 Числа Стирлинга

Произведение

$$x(x-1)\dots(x-k+1), \quad k \ge 1,$$

будем называть k-*й нижней факториальной степенью* числа x и будем обозначать через $x^{\underline{k}}$.

Произведение

$$x(x+1)\dots(x+k-1), \quad k \ge 1,$$

будем называть k-*й верхней факториальной степенью* числа x и будем обозначать через $x^{\overline{k}}$.

При $k = 0$ полагаем $x^{\underline{0}} = x^{\overline{0}} = 1$. Отметим, что такое определение значения $x^{\underline{0}}$ согласуется с тем, что $\binom{\alpha}{0} = 1$.

Коэффициенты $c(n,k)$ и $s(n,k)$ в представлениях

$$x^{\overline{n}} = \sum_{k=0}^{n} c(n,k)x^k, \quad x^n = \sum_{k=0}^{n} s(n,k)x^{\underline{k}}$$

называются *числами Стирлинга первого и второго рода* соответственно.

Свойства чисел Стирлинга:

1. Верны представления

$$(x)^{\underline{n}} = \sum_{k=0}^{n} (-1)^{n-k} c(n,k) x^k, \quad (x)^n = \sum_{k=0}^{n} (-1)^{n-k} s(n,k) x^{\overline{k}}.$$

▷ В силу очевидных равенств $(-x)^{\overline{n}} = (-1)^n x^{\underline{n}}$ и $(-1)^m = (-1)^{-m}$ имеем:

$$(x)^{\underline{n}} = (-1)^n (-x)^{\overline{n}} = (-1)^n \sum_{k=0}^{n} c(n,k)(-1)^k x^k = \sum_{k=0}^{n} (-1)^{n-k} c(n,k) x^k,$$

$$(x)^n = (-1)^n (-x)^n = (-1)^n \sum_{k=0}^{n} s(n,k)(-x)^{\underline{k}} = \sum_{k=0}^{n} (-1)^{n-k} s(n,k) x^{\overline{k}}.$$

Свойство 1 доказано. □

Коэффициенты $(-1)^{n-k} c(n,k)$ в первом представлении из свойства 1 еще иногда называются *числами Стирлинга первого рода со знаком.*

2. Для любого целого неотрицательного n верны равенства

$$c(n,n) = s(n,n) = 1.$$

3. Для любого натурального n верны равенства

$$c(n,0) = s(n,0) = 0.$$

4. При $n < k$ верны равенства

$$c(n,k) = s(n,k) = 0.$$

5. При $n \geq k \geq 1$ справедливы рекуррентные соотношения

$$c(n,k) = c(n-1,k-1) + (n-1)c(n-1,k),$$
$$s(n,k) = s(n-1,k-1) + k\, s(n-1,k).$$

▷ Действительно, при $n \geq 1$ имеем:

$$\sum_{k=1}^{n} c(n,k) x^k = \sum_{k=0}^{n} c(n,k) x^k = x^{\overline{n}} = x^{\overline{n-1}}(x+n-1) =$$
$$= x \sum_{k=0}^{n-1} c(n-1,k) x^k + (n-1) \sum_{k=0}^{n-1} c(n-1,k) x^k =$$
$$= \sum_{k=1}^{n} c(n-1,k-1) x^k + (n-1) \sum_{k=1}^{n} c(n-1,k) x^k =$$
$$= \sum_{k=1}^{n} \big(c(n-1,k-1) + (n-1)c(n-1,k)\big) x^k;$$

$$\sum_{k=1}^{n} s(n,k)x^{\underline{k}} = \sum_{k=0}^{n} s(n,k)x^{\underline{k}} = x^n = x^{n-1}x =$$

$$= x\sum_{k=0}^{n-1} s(n-1,k)x^{\underline{k}} = \sum_{k=0}^{n-1} s(n-1,k)x^{\underline{k}}\big((x-k)+k\big) =$$

$$= \sum_{k=1}^{n} s(n-1,k-1)x^{\underline{k}} + \sum_{k=0}^{n-1} k\,s(n-1,k)x^{\underline{k}} =$$

$$= \sum_{k=1}^{n} s(n-1,k-1)x^{\underline{k}} + \sum_{k=1}^{n} k\,s(n-1,k)x^{\underline{k}} =$$

$$= \sum_{k=1}^{n} \big(s(n-1,k-1) + k\,s(n-1,k)\big)x^{\underline{k}}.$$

Свойство 5 доказано. □

6. Числа Стирлинга $c(n,k)$ и $s(n,k)$ при малых значениях k и n указаны в двух следующих таблицах.

Лекция № 6

Теперь покажем, что у чисел Стирлинга первого и второго рода есть эквивалентные комбинаторные определения.

Теорема 12. *Число Стирлинга первого рода $c(n,k)$ равно количеству подстановок симметрической группы S_n, раскладывающихся в произведение ровно k циклов (включая циклы длины 1).*

Доказательство. Обозначим через $c'(n,k)$ число подстановок симметрической группы S_n, раскладывающихся в произведение ровно k циклов (при этом по определению положим $c'(0,0) = 1$). Очевидно, что $c'(n,n) = 1$, $c'(n,0) = 0$ при $n \neq 0$, $c'(n,k) = 0$ при $n < k$. Кроме того, при $n \geq k \geq 1$ справедливо рекуррентное соотношение

$$c'(n,k) = c'(n-1,k-1) + (n-1)c'(n-1,k).$$

Поэтому $c'(n,k) = c(n,k)$. □

Следствие 8. *Справедливо равенство*

$$\sum_{k=0}^{n} c(n,k) = n!$$

Упражнение 11. Доказать равенства

$$c(n,1) = (n-1)!, \quad c(n,n-1) = C_n^2.$$

Теорема 13. *Число Стирлинга второго рода $s(n,k)$ равно количеству способов разбиения n-элементного множества на k непустых непересекающихся подмножеств.*

Доказательство. Обозначим через $s'(n,k)$ число способов разбиения n-элементного множества на k непустых непересекающихся подмножеств (при этом по определению положим $s'(0,0) = 1$). Очевидно, что $s'(n,n) = 1$, $s'(n,0) = 0$ при $n \neq 0$, $s'(n,k) = 0$ при $n < k$. Кроме того, при $n \geq k \geq 1$ справедливо рекуррентное соотношение

$$s'(n,k) = s'(n-1,k-1) + k\, s'(n-1,k).$$

Поэтому $s'(n,k) = s(n,k)$. □

Упражнение 12. Доказать равенства

$$s(n,1) = 1, \quad s(n,2) = 2^{n-1} - 1,$$
$$s(n,3) = \frac{1}{2}\left(3^{n-1} - 2^n + 1\right), \quad s(n,n-1) = C_n^2.$$

Для всех целых неотрицательных n положим

$$B_n = \sum_{k=0}^{n} s(n,k).$$

Число B_n будем называть *n-м числом Белла*. Очевидно, что B_n равно числу разбиений n-элементного множества. Для чисел Белла есть простая рекуррентная формула.

Теорема 14. *Для любого целого неотрицательного n справедливо равенство*

$$B_{n+1} = \sum_{k=0}^{n} \binom{n}{k} B_k.$$

Доказательство. Существует ровно $\binom{n}{k}B_k$ разбиений $(n+1)$-элементного множества, в которых $(n+1)$-й элемент входит в множество мощности $n-k+1$, так как в это множество мощности $n-k+1$ не входят ровно k элементов и эти k элементов нужно выбрать, а затем разбить это множество мощности k. □

Теперь найдем явную формулу для чисел Стирлинга второго рода.

Теорема 15. *Для любых целых неотрицательных n и k справедливо равенство*

$$s(n,k) = \frac{1}{k!}\sum_{t=0}^{k}(-1)^{k-t}\binom{k}{t}t^n.$$

Доказательство. Сначала решим вспомогательную задачу — найдем количество способов разложить n разных предметов в k пронумерованных ящиков, чтобы при этом все ящики были непустыми.

Будем решать эту задачу по формуле включений-исключений. Положим $N = k^n$. Под i-м свойством, $i = 1, \ldots, k$, будем понимать свойство «i-й ящик пуст». Тогда

$$N(0) = \sum_{s=0}^{k}(-1)^s\binom{k}{s}(k-s)^n = \sum_{t=0}^{k}(-1)^{k-t}\binom{k}{t}t^n.$$

Использование очевидного соотношения

$$s(n,k) = \frac{N(0)}{k!}$$

дает нужное равенство. □

§ 2.4 Метод производящих функций

Сопоставим последовательности $\{a_n\}$ или бесконечному вектору

$$(a_0, a_1, \ldots, a_n \ldots)$$

формальный степенной ряд

$$A(x) = \sum_{n=0}^{\infty} a_n x^n,$$

который называется *производящей функцией* последовательности $\{a_n\}$.

Иногда для последовательности $\{a_n\}$ также используется понятие *экспоненциальной производящей функции*, которая определяется как такой формальный степенной ряд:

$$A_e(x) = \sum_{n=0}^{\infty} \frac{a_n}{n!} x^n.$$

В общих чертах метод производящих функций заключается в переходе от последовательности к производящей функции, ее изучение с дальнейшим использованием свойств производящей функции для получения информации об исходной последовательности. Наиболее типичная ситуация для метода производящих функций такая. Имеется рекуррентное соотношение, которому удовлетворяет искомая последовательность. По этому соотношению на элементы последовательности находится уравнение, которому удовлетворяет производящая функция последовательности. Решая уравнение (здесь, вообще говоря, должна быть немалая доля везения), находим производящую функцию, а следовательно, так или иначе, находим элементы последовательности.

К использованию метода производящих функций можно условно выделить три подхода.

1. Развитие и использование теории формальных степенных рядов, обобщающих теорию обычных степенных рядов.

2. Переход к функциям, к которым сходятся (абсолютно) в некоторой окрестности нуля формальные степенные ряды; использование возможностей классического анализа.

3. Работа с формальными степенными рядами как с обычными функциями (без уточнения вопросов сходимости и других вопросов корректности) с последующей проверкой ответа.

Отметим, что третий подход, даже если его рассматривать только как способ «угадывания» ответа, зачастую бывает достаточно эффективен. Например, задача нахождения суммы кубов первых n натуральных чисел труднее задачи проверить равенство

$$\sum_{k=1}^{n} k^3 = \frac{n^2(n+1)^2}{4},$$

которое очень легко устанавливается по индукции: база индукции очевидна, а переход следует из соотношения

$$\frac{n^2(n+1)^2}{4} - \frac{(n-1)^2 n^2}{4} = n^3.$$

Вообще, проверить, удовлетворяет ли исходному рекуррентному соотношению вида $a_n = f(n, a_{n-1}, a_{n-2}, \ldots, a_{n-k})$ найденная последовательность — задача существенно более простая нежели задача найти эту последовательность.

Однако, мы будем, в основном, придерживаться второго подхода. В отдельных случаях, связанных, как правило, с производящими функциями для нечисловых последовательностей, будем использовать элементы теории формальных рядов. Для этого дадим некоторые определения.

Пусть непустое множество K замкнуто относительно операции «$+$» (сложение) и «$\times$» (умножение), является абелевой группой по сложению и для любых элементов $a, b, c \in K$ выполняются законы дистрибутивности $a(b+c) = ab+ac$ и $(a+b)c = ac+bc$. Тогда множество K (или тройка $(K, +, \times)$) называется *кольцом*.

Пусть K — кольцо. Обозначим через $K[[x]]$ множество формальных степенных рядов с коэффициентами из кольца K. На множестве $K[[x]]$ введем операции сложения и умножения. Пусть

$$A(x) = \sum_{n=0}^{\infty} a_n x^n, \quad B(x) = \sum_{n=0}^{\infty} b_n x^n.$$

Для элементов $A(x)$ и $B(x)$ множества $K[[x]]$ положим:

$$A(x) + B(x) = \sum_{n=0}^{\infty} (a_n + b_n) x^n,$$

$$A(x) \times B(x) = \sum_{n=0}^{\infty} c_n x^n, \quad \text{где} \quad c_n = \sum_{k=0}^{n} a_k b_{n-k}.$$

Последовательность $\{c_n\}$, задаваемая равенствами

$$c_n = \sum_{k=0}^{n} a_k b_{n-k}, \quad n = 0, 1, \ldots,$$

называется *сверткой* последовательностей $\{a_n\}$ и $\{b_n\}$. Производящая функция свертки двух последовательностей равна произведению производящих функций этих последовательностей.

Свойства множества $K[[x]]$ с введенными операциями:

1. $K[[x]] = (K[[x]], +, \times)$ — кольцо.

2. Нулем (нейтральным элементом по сложению) кольца $K[[x]]$ является элемент

$$\sum_{n=0}^{\infty} 0 \times x^n,$$

где 0 — нейтральный элемент по сложению кольца K.

3. Если кольцо K ассоциативно, то кольцо $K[[x]]$ тоже ассоциативно.

4. Если кольцо K коммутативно, то кольцо $K[[x]]$ тоже коммутативно.

5. Если кольцо K с единицей (нейтральным элементом по умножению), то кольцо $K[[x]]$ тоже с единицей. В этом случае единицей кольца $K[[x]]$ является элемент

$$1 \times x^0 + \sum_{n=1}^{\infty} 0 \times x^n,$$

где 0 — нейтральный элемент по сложению кольца K, 1 — нейтральный элемент по умножению кольца K.

6. Если кольцо K является областью целостности (отсутствуют делители нуля), то кольцо $K[[x]]$ тоже является областью целостности.

7. Кольцо $K[[x]]$ не является полем, даже если K — поле. А именно, элемент $A(x) \in K[[x]]$ обратим тогда и только тогда, когда элемент $a_0 = A(0) \in K$ обратим (здесь к записи $A(0)$ формально нужно относиться только как к обозначению для a_0).

$\triangleright$ Действительно, для существования для элемента $A(x) = \sum_{n=0}^{\infty} a_n x^n$ обратного элемента $B(x) = \sum_{n=0}^{\infty} b_n x^n$ необходимым и достаточным условием является справедливость равенств

$$1 = a_0 b_0, \qquad 0 = \sum_{k=0}^{n} a_k b_{n-k}, \quad n = 1, 2, \ldots$$

Отсюда необходимость обратимости элемента a_0 очевидна. С другой стороны, если элемент a_0 обратим, то равенства

$$b_0 = a_0^{-1}, \qquad b_n = -a_0^{-1} \sum_{k=1}^{n} a_k b_{n-k}, \quad n = 1, 2, \ldots$$

последовательно определяют коэффициенты обратного к $A(x)$ элемента $B(x) \in K[[x]]$. $\square$

Теперь временно оставим формальный подход к изучению производящих функций. Рассмотрим примеры, когда формальным степеным рядам соответствуют функции, к которым в некоторой окрестности нуля сходятся эти степенные ряды.

1. Если $a_n = \binom{m}{n}$, $n = 1, 2, \ldots$, то

$$A(x) = \sum_{n=0}^{\infty} \binom{m}{n} x^n = (1 + x)^m.$$

2. Если $a_n = \binom{m+n-1}{n}\lambda^n$, $n = 1, 2, \ldots$, то

$$A(x) = \sum_{n=0}^{\infty} \binom{m+n-1}{n} \lambda^n x^n = (1-\lambda x)^{-m}.$$

Пусть производящая функция $A(x)$ последовательности $\{a_n\}$ абсолютно сходится в некоторой окрестности нуля (к функции, которую естественно также обозначать $A(x)$), а последовательность $\{b_n\}$ тесно связана с последовательностью $\{a_n\}$. Выразим производящую функцию $B(x)$ последовательности $\{b_n\}$ через функцию $A(x)$.

Если

$$b_n = \begin{cases} 0, & \text{если } 0 \le n \le k-1; \\ a_{n-k}, & \text{если } n \ge k, \end{cases}$$

то

$$B(x) = x^k A(x).$$

Если

$$b_n = a_{n+k}, \quad n = 0, 1, \ldots,$$

то

$$B(x) = \frac{A(x) - \sum\limits_{i=0}^{k-1} a_i x^i}{x^k} = \frac{A(x) - \sum\limits_{i=0}^{k-1} \frac{A^{(i)}(0)}{i!} x^i}{x^k}.$$

Если

$$b_n = n a_n, \quad n = 0, 1, \ldots,$$

то

$$B(x) = xA'(x).$$

Теперь применим метод производящих функций для решения одной известной задачи.

Определим *правильную скобочную структуру* как последовательность левых и правых скобок, которую можно получить за конечное число применений следующих правил:

— последовательность ( ) является правильной скобочной структурой;

— если П — правильная скобочная структура, то (П) — тоже правильная скобочная структура;

— если Π_1 и Π_2 — правильные скобочные структуры, то $\Pi_1\Pi_2$ — тоже правильная скобочная структура.

Задача заключается в нахождении числа правильных скобочных структур, содержащих ровно по n левых и правых скобок. Обозначим это число через a_n.

Очевидно, что в каждом начальном отрезке любой правильной скобочной структуры число левых скобок не менее числа правых скобок. Разобьем множество $\mathcal{A}_{n+1}$ всех правильных скобочных структур, содержащих ровно по $n+1$ левых и правых скобок, на подмножества $\mathcal{A}_{n+1}^{(1)}, \ldots, \mathcal{A}_{n+1}^{(n+1)}$, где $\mathcal{A}_{n+1}^{(i)}$, $i = 1, \ldots, n+1$, — подмножество правильных скобочных структур из множества $\mathcal{A}_{n+1}$, обладающих следующим свойством: впервые равенство числа левых и правых скобок достигается на элементе последовательности скобок с номером $2i$. Тогда справедливы равенства

$$a_{n+1} = |\mathcal{A}_{n+1}| = \left|\mathcal{A}_{n+1}^{(1)}\right| + \ldots + \left|\mathcal{A}_{n+1}^{(n+1)}\right| = a_n + a_1 a_{n-1} + \ldots + a_{n-1} a_1 + a_n.$$

Теперь, полагая $a_0 = 1$, получаем рекуррентную формулу

$$a_{n+1} = \sum_{k=0}^{n} a_k a_{n-k}.$$

Переходя к производящей функции $A(x)$ последовательности $\{a_n\}$, заметим, что для всех n верно неравенство $a_n \leq 4^n$. Поэтому в некоторой окрестности нуля соответствующий этой последовательности степенной ряд абсолютно сходится.

Введем последовательности $\{b_n\}$ и $\{c_n\}$ равенствами

$$b_n = a_{n+1}, \qquad c_n = \sum_{k=0}^{n} a_k a_{n-k}, \qquad n = 0, 1, \ldots$$

Для производящих функций $B(x)$ и $C(x)$ последовательностей $\{b_n\}$ и $\{c_n\}$ справедливы равенства:

$$B(x) = C(x), \quad B(x) = \frac{A(x) - a_0}{x}, \quad C(x) = A^2(x).$$

Следовательно,

$$\frac{A(x) - 1}{x} = A^2(x).$$

Таким образом, имеем квадратное уравнение

$$xA^2(x) - A(x) + 1 = 0$$

относительно $A(x)$. Решая его, получаем:

$$A(x) = \frac{1 \pm \sqrt{1 - 4x}}{2x}.$$

Подходит только один корень:

$$A(x) = \frac{1 - \sqrt{1-4x}}{2x}.$$

Разложим в ряд Маклорена функцию $1 - \sqrt{1-4x}$:

$$\begin{aligned}
1 - \sqrt{1-4x} &= 1 - \sum_{k=0}^{\infty} \binom{1/2}{k} (-4x)^k = \\
&= -\sum_{k=1}^{\infty} \frac{\frac{1}{2}\left(\frac{1}{2}-1\right)\ldots\left(\frac{1}{2}-k+1\right)}{k!}(-4x)^k = \\
&= \sum_{k=1}^{\infty} \frac{(-1)(1-2)(1-4)\ldots(1-2(k-1))}{2^k k!}(-4x)^k = \\
&= \sum_{k=1}^{\infty} \frac{(2k-3)!!}{2^k k!} 4^k x^k = \sum_{k=1}^{\infty} \frac{(2k-2)!!(2k-3)!!}{2^{k-1}(k-1)!2^k k!} 4^k x^k = \\
&= \sum_{k=1}^{\infty} \frac{(2k-2)!}{(k-1)!k!} 2x^k = \sum_{k=1}^{\infty} \frac{2}{k} \binom{2k-2}{k-1} x^k.
\end{aligned}$$

Таким образом,

$$A(x) = \frac{1}{2x} \sum_{k=1}^{\infty} \frac{2}{k} \binom{2k-2}{k-1} x^k = \sum_{n=0}^{\infty} \frac{1}{n+1} \binom{2n}{n} x^n.$$

Следовательно,

$$a_n = \frac{1}{n+1} \binom{2n}{n}, \quad n = 0, 1, \ldots$$

Число $\frac{1}{n+1}\binom{2n}{n}$ называется *n-м числом Каталана*. Числа Каталана возникают во многих комбинаторных задачах, в частности, n-му числу Каталана равны:

а) количество наборов, состоящих из n нулей и n единиц, в которых в любом начальном отрезке число единиц не превосходит числа нулей;

б) количество способов разбиения (путем проведения непересекающихся диагоналей) выпуклого $(n+2)$-угольника на треугольники;

в) количество монотонных отображений f из множества $\{1, 2, \ldots, n\}$ в множество $\{1, 2, \ldots, n\}$, удовлетворяющих условию $f(i) \leq i$ для всех i, $1 \leq i \leq n$;

г) количество всех $(2 \times n)$ матриц, множество элементов которых совпадает со множством $\{1, 2, \ldots, 2n\}$, причем в каждой строке и в каждом столбце элементы расположены в возрастающем порядке.

Упражнение 13. Найти экспоненциальную производящую функцию последовательности $\{a_n\}$, где a_n — количество возрастающе-убывающих подстановок симметрической группы S_n (подстановка $\sigma \in S_n$ возрастающе-убывающая, если в последовательности $\sigma(1), \sigma(2), \ldots, \sigma(n)$ все элементы, стоящие на четных местах, больше своих соседей).

Ответ: $A_e(x) = \frac{1+\sin x}{\cos x}$ (функция $A_e(x)$ удовлетворяет уравнению $2A_e'(x) - A_e^2(x) - 1 = 0$).

Здесь вставить метод траекторий и замечание про важность метода производящих функций несмотря на наличие у задач более простых и элегантных решений

Лекция № 7

§ 2.5 Линейные рекуррентные последовательности

Рассмотрим следующую задачу. Нужно найти число f_n наборов из нулей и единиц длины n, обладающие свойствами:

1) первый разряд набора равен единице;

2) последний разряд набора равен единице;

3) в наборе нет двух стоящих рядом нулей.

Последовательность $\{f_n\}$ удовлетворяет рекуррентному соотношению $f_{n+2} = f_{n+1} + f_n$ и начальным условиям $f_1 = f_2 = 1$. Такая последовательность называется *последовательностью Фибоначчи* (в исходной задаче специально добавлены первые два условия, чтобы получилось именно эта последовательность, а не «сдвинутая»). Последовательность Фибоначчи возникает очень часто в самых разных областях математики. В частности, f_n равно количеству таких подстановок σ симметрической группы S_{n-1}, что для любого i, $1 \le i \le n-1$, выполняются неравенства $|i - \sigma(i)| \le 1$.

Последовательность Фибоначчи является частным случаем возвратных последовательностей, задаваемых линейными однородными соотношениями с постоянными коэффициентами. Найдем общее решение такого

соотношения.

Теорема 16. *Пусть последовательность* $\{a_n\}$ *с элементами из* $\mathbb{C}$ *удовлетворяет рекуррентному соотношению*

$$a_{n+k} = u_1 a_{n+k-1} + \ldots + u_{k-1} a_{n+1} + u_k a_n,$$

где $u_i \in \mathbb{C}$, $i = 1, \ldots, k$, $u_k \neq 0$. *Тогда*

$$a_n = \sum_{i=1}^{s} \lambda_i^n P_i(n), \qquad (*)$$

где λ_i *— корень кратности* r_i *характеристического многочлена*

$$x^k - u_1 x^{k-1} - \ldots - u_{k-1} x - u_k,$$

$P_i(n)$ *— многочлен от переменной* n *степени* $r_i - 1$, $i = 1, \ldots, s$. *Коэффициенты многочленов* P_i *в количестве* $r_1 + \ldots + r_s = k$ *штук определяются из справедливости формулы* $(*)$, *например, для первых* k *членов последовательности.*

Доказательство. Найдется такая константа c, что для любого n, $n \geq 1$, справедливо неравенство

$$|a_{n+1}| \leq c \max_{0 \leq i \leq n} |a_i|.$$

Поэтому рост элементов последовательности по абсолютной величине не более чем экспоненциальный. Следовательно, переходя от последовательности $\{a_n\}$ к производящей функции $A(x)$ этой последовательности, получаем степенной ряд, абсолютно сходящийся в некоторой окрестности нуля.

Производящая функция $B(x)$ последовательности $\{b_n\}$, задаваемой при всех n равенством $b_n = a_{n+i}$, имеет такой вид:

$$B(x) = x^{-i}(A(x) - Q_{i-1}(x)),$$

где $Q_{i-1}(x)$ — многочлен степени $i-1$, определяемый первыми i членами последовательности $\{a_n\}$.

Поэтому исходное рекуррентное соотношение дает следующее уравнение относительно производящей функции $A(x)$:

$$x^{-k}(A(x) - Q_{k-1}(x)) = \sum_{i=1}^{k-1} u_i x^{-(k-i)}(A(x) - Q_{k-i-1}(x)) + u_k A(x).$$

Домножим левую и правую часть этого равенства на x^k и соберем слагаемые, содержащие функцию $A(x)$ в левой части равенства:

$$A(x)\left(1-\sum_{i=1}^{k}u_ix^i\right)=\tilde{Q}_{k-1}(x),$$

где $\tilde{Q}_{k-1}(x)$ — некоторый многочлен степени $k-1$.

Пусть $\lambda_1,\ldots,\lambda_s$ — корни характеристического многочлена

$$x^k-u_1x^{k-1}-\ldots-u_{k-1}x-u_k,$$

кратности $r_1,\ldots,r_s$, соответственно. Так как $u_k\neq 0$, то $\lambda_i\neq 0$, $i=1,\ldots,s$. Тогда, делая замену $x\to\frac{1}{y}$, получаем, что многочлен

$$1-u_1x-\ldots-u_{k-1}x^{k-1}-u_kx^k$$

имеет корни $\frac{1}{\lambda_1},\ldots,\frac{1}{\lambda_s}$ кратности $r_1,\ldots,r_s$, соответственно. Следовательно,

$$A(x)=\frac{\tilde{Q}_{k-1}(x)}{(-u_k)\left(x-\frac{1}{\lambda_1}\right)^{r_1}\ldots\left(x-\frac{1}{\lambda_s}\right)^{r_s}}=\frac{(-\lambda_1)^{r_1}\ldots(-\lambda_s)^{r_s}\tilde{Q}_{k-1}(x)}{(-u_k)(1-\lambda_1x)^{r_1}\ldots(1-\lambda_sx)^{r_s}}.$$

Используя разложение рациональной функции в сумму простейших дробей, получаем:

$$A(x)=\sum_{i=1}^{s}\sum_{j=1}^{r_i}\frac{\alpha_{ij}}{(1-\lambda_ix)^j},$$

где α_{ij}, $i=1,\ldots,s$, $j=1,\ldots,r_i$, — некоторые константы (вообще говоря, комплексные). Далее, для $i=1,\ldots,s$ имеем:

$$\sum_{j=1}^{r_i}\frac{\alpha_{ij}}{(1-\lambda_ix)^j}=\sum_{j=1}^{r_i}\alpha_{ij}\sum_{n=0}^{\infty}\binom{n+j-1}{n}\lambda_i^nx^n=$$

$$=\sum_{n=0}^{\infty}\lambda_i^nx^n\sum_{j=1}^{r_i}\alpha_{ij}\binom{n+j-1}{j-1}=\sum_{n=0}^{\infty}\lambda_i^nP_i(n)x^n,$$

где $P_i(n)$ — некоторый многочлен от переменной n степени r_i-1.

Таким образом,

$$A(x)=\sum_{i=1}^{s}\sum_{n=0}^{\infty}\lambda_i^nP_i(n)x^n=\sum_{n=0}^{\infty}\left(\sum_{i=1}^{s}\lambda_i^nP_i(n)\right)x^n.$$

Приравнивая коэффициенты при степенях переменной x, получаем утверждение теоремы. □

Замечание 2. Линейное неоднородное соотношение с постоянными коэффициентами вида

$$a_{n+k} - u_1 a_{n+k-1} - \ldots - u_{k-1} a_{n+1} - u_k a_n = f(n),$$

в случае, когда функция $f(n)$ является квазимногочленом (т. е. $f(n) = \lambda^n P(n)$, где $P(n)$ — многочлен от переменной n), решается методом производящих функций практически так же, как и однородное.

Возвращаясь к последовательности Фибоначчи $\{f_n\}$, удовлетворяющей рекуррентному соотношению $f_{n+2} = f_{n+1} + f_n$ и начальным условиям $f_1 = f_2 = 1$ (или, что то же самое, начальным условиям $f_0 = 0$, $f_1 = 1$), отметим, что характеристический многочлен $x^2 - x - 1 = 0$ последовательности Фибоначчи имеет корни

$$\lambda_{1,2} = \frac{1 \pm \sqrt{5}}{2}.$$

Поэтому, в силу доказанной теоремы, общим решением соотношения $f_{n+2} = f_{n+1} + f_n$ будет такая совокупность последовательностей:

$$a_n = c_1 \left(\frac{1+\sqrt{5}}{2}\right)^n + c_2 \left(\frac{1-\sqrt{5}}{2}\right)^n.$$

Из начальных условий находим, что $c_1 = 1/\sqrt{5}$, $c_2 = -1/\sqrt{5}$. Таким образом,

$$f_n = \frac{1}{\sqrt{5}} \left(\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n \right).$$

Свойства последовательности Фибоначчи:

1. Пусть φ — «золотое сечение», т. е. $\varphi = (1+\sqrt{5})/2$. Тогда для n-го члена последовательности Фибоначчи справедливы формулы:

а) $f_n = \frac{1}{\sqrt{5}} (\varphi^n - (-\varphi)^{-n})$ (формула Бине);

б) $f_n = \left\lfloor \frac{\varphi^n}{\sqrt{5}} \right\rceil$.

2. Элементы последовательности Фибоначчи удовлетворяют следующим соотношениям:

а) $f_1 + f_2 + \ldots + f_n = f_{n+2} - 1$;

б) $f_1 + f_3 + \ldots + f_{2n-1} = f_{2n}$;

в) $f_2 + f_4 + \ldots + f_{2n} = f_{2n+1} - 1$;

г) $f_1^2 + f_2^2 + \ldots + f_n^2 = f_n f_{n+1}$;

д) $f_n^2 + f_{n+1}^2 = f_{2n+1}$;

е) $f_{n+m} = f_{n-1}f_m + f_n f_{m+1} = f_{n+1}f_{m+1} - f_{n-1}f_{m-1}$;

ж) $f_{3n} = f_{n+1}^3 + f_n^3 - f_{n-1}^3$;

з) $\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}^n = \begin{pmatrix} f_{n+1} & f_n \\ f_n & f_{n-1} \end{pmatrix}$.

3. Любые два соседних члена последовательности Фибоначчи взаимнопросты. Более того,

$$(f_n, f_m) = f_{(n,m)},$$

где (a, b) — наибольший общий делитель чисел a и b.

4. Любое натуральное число n может быть представлено, причем единственным образом, в виде

$$n = \sum_{k=2}^{\infty} \alpha_k f_k,$$

где $\alpha_i \in \{0, 1\}$, $\alpha_i + \alpha_{i+1} \leq 1$, $i = 2, 3, \ldots$

5. Для производящей функции $F(x)$ последовательности Фибоначчи справедливо равенство

$$F(x) = \frac{x}{1 - x - x^2}$$

(так как степенной ряд абсолютно сходится в некоторой окрестности нуля, то под производящей функцией можно понимать функцию, к которой сходится соответствующий степенной ряд).

Упражнение 14. Доказать свойства последовательности Фибоначчи.

§ 2.6 Формула обращения Мебиуса

Пусть $p_1^{m_1} \ldots p_k^{m_k}$ — разложение числа n на простые множители. Введем *функцию Мебиуса* $\mu(n)$ натурального аргумента n равенством

$$\mu(n) = \begin{cases} 1, & \text{если } n = 1; \\ (-1)^k, & \text{если } m_1 = \ldots = m_k = 1; \\ 0, & \text{иначе.} \end{cases}$$

Лемма 9. *Справедливо равенство*

$$\sum_{d:\, d|n} \mu(d) = \begin{cases} 1, & \textit{если } n = 1; \\ 0, & \textit{если } n \geq 2. \end{cases}$$

Доказательство. При $n = 1$ равенство очевидно.

При $n \geq 2$ рассмотрим разложение $n = p_1^{m_1} \dots p_k^{m_k}$ на простые множители. Положим $\tilde{n} = p_1 \dots p_k$. Тогда

$$\sum_{d\colon d|n} \mu(d) = \sum_{d\colon d|\tilde{n}} \mu(d) + \sum_{d\colon d|n,\, d\nmid\tilde{n}} \mu(d) = \\ = \sum_{d\colon d|\tilde{n}} \mu(d) = 1 - C_k^1 + C_k^2 - \ldots + (-1)^k C_k^k = 0.$$

Тем самым равенство доказано в обоих случаях. □

Теорема 17 (формула обращения Мебиуса)**.** *Пусть* $f\colon \mathbb{N} \to \mathbb{R}$, $g\colon \mathbb{N} \to \mathbb{R}$ *и для всех натуральных* n *верно равенство*

$$f(n) = \sum_{d\colon d|n} g(d).$$

Тогда для всех натуральных n *справедливо равенство*

$$g(n) = \sum_{d\colon d|n} \mu(d) f\left(\frac{n}{d}\right).$$

Доказательство. Верна следующая цепочка равенств

$$\sum_{d\colon d|n} \mu(d) f\left(\frac{n}{d}\right) = \sum_{d\colon d|n} \mu(d) \sum_{b\colon b|\frac{n}{d}} g(b) = \sum_{d\colon d|n} \sum_{b\colon b|\frac{n}{d}} \mu(d) g(b) = \\ = \sum_{(d,b)\colon d|n,\, b|\frac{n}{d}} \mu(d) g(b) = \sum_{(d,b)\colon db|n} \mu(d) g(b) = \\ = \sum_{(d,b)\colon b|n,\, d|\frac{n}{b}} g(b)\mu(d) = \sum_{b\colon b|n} g(b) \sum_{d\colon d|\frac{n}{b}} \mu(d) = g(n),$$

где последнее равенство следует из того, что сумма

$$\sum_{d\colon d|\frac{n}{b}} \mu(d)$$

по лемме 9 при $b = n$ равна 1, а в остальных случаях равна 0. □

Теперь применим формулу обращения Мебиуса для перечисления циклических последовательностей.

Циклическая последовательность

$$(\ldots, a_1, a_2, \ldots, a_n, a_1, a_2, \ldots, a_n, a_1, a_2, \ldots)$$

длины (периода) n порождается линейными последовательностями

$$(a_1, a_2, \ldots, a_n),\ (a_2, \ldots, a_n, a_1),\ \ldots,\ (a_n, a_1, \ldots, a_{n-1})$$

длины n и, в свою очередь, порождает эти n линейных последовательностей, среди которых, вообще говоря, могут быть совпадающие.

Для циклической последовательности, порождаемой линейной последовательностью $(a_1, a_2, \ldots, a_n)$, наименьшее число d, удовлетворяющее условию

$$(a_1, a_2, \ldots, a_n) = (a_{d+1}, \ldots, a_n, a_1, \ldots, a_d),$$

называется *наименьшим периодом* этой циклической последовательности. Очевидно, что всегда d делит n.

Пусть $|A| = r$. Тогда число линейных последовательностей длины n с элементами из множества A равно r^n. Обозначим через $M(d)$ число циклических последовательностей наименьшего периода d. Тогда

$$\sum_{d:\ d|n} dM(d) = r^n.$$

Применяя формулу обращения Мебиуса для функций $f(n) = r^n$ и $g(n) = nM(n)$, имеем:

$$nM(n) = \sum_{d:\ d|n} \mu(d) r^{n/d},$$

и следовательно,

$$M(n) = \frac{1}{n} \sum_{d:\ d|n} \mu(d) r^{n/d}.$$

Таким образом, мы нашли число циклических последовательностей с минимальным периодом n. Для нахождения числа $T(n)$ всех циклических последовательностей периода n достаточно использовать простое равенство

$$T(n) = \sum_{d:\ d|n} M(d).$$

Лекция № 8

§ 2.7 Количество неприводимых многочленов над $\mathbb{Z}_p$

Рассмотрим поле $\mathbb{Z}_p$ и кольцо многочленов $\mathbb{Z}_p[x]$ над этим полем. Будем рассматривать только многочлены, у которых коэффициент при стар-

шем члене равен 1. Такие многочлены будем называть *нормированными*[2]. Отметим, что произведение нормированных многочленов также является нормированным многочленом.

Количество нормированных многочленов степени n над полем $\mathbb{Z}_p$ равно p^n (всего у многочлена $n+1$ коэффициентов, один из которых фиксирован, а остальные n коэффициентов могут принимать любые из p значений). Следовательно, переходя от последовательности числа нормированных многочленов степени n к производящей функции этой последовательности, получаем степенной ряд, абсолютно сходящийся в некоторой окрестности нуля.

С одной стороны, производящая функция числа нормированных многочленов имеет вид

$$\sum_{n=0}^{\infty} p^n x^n = \frac{1}{1-px}.$$

С другой стороны, обозначив число неприводимых нормированных многочленов над полем $\mathbb{Z}_p$ степени m через I_m, имеем:

$$\sum_{n=0}^{\infty} p^n x^n = \prod_{m=1}^{\infty} \underbrace{\left(1 + x^m + x^{2m} + \ldots\right) \ldots \left(1 + x^m + x^{2m} + \ldots\right)}_{I_m \text{ раз}}.$$

Поясним это равенство. Для каждого m, $m \geq 1$, занумеруем I_m соответствующих скобок из правой части равенства неприводимыми нормированными многочленами $g_{m,1}, \ldots, g_{m,I_m}$ степени m. Теперь произвольному нормированному многочлену f степени n, представленному как произведение неприводимых многочленов

$$f = (g_{s_1,i_1})^{t_1} \ldots (g_{s_k,i_k})^{t_k}$$

(здесь $t_1 s_1 + \ldots + t_k s_k = n$), сопоставим следующее произведение, получающееся в правой части после раскрытия скобок. Для $j = 1, \ldots, k$ из скобки, которой сопоставлен неприводимый многочлен g_{s_j,i_j}, берем слагаемое $x^{t_j s_j}$, а из остальных скобок берем слагаемое 1. Тем самым многочлену f будет сопоставлено произведение $x^{t_1 s_1} \ldots x^{t_k s_k}$, равное x^n. Верно и обратное — любому получающемуся после раскрытия скобок в правой части произведению, равному x^n сопоставляется произведение степеней неприводимых многочленов, дающее нормированный многочлен степени n. Поэтому при раскрытии скобок в правой части при x^n будет коэффициент, равный числу всех нормированных многочленов степени n, т. е. равный p^n.

[2]) Обычно такие многочлены называют приведенными, но так как текущая цель — подсчитать число неприводимых многочленов, то во избежание словосочетания «приведенные неприводимые многочлены», будем использовать термин «нормированные».

Объединяя два равенства, получаем, что

$$1 - px = \prod_{m=1}^{\infty} (1 - x^m)^{I_m}.$$

Приравниваем логарифмические производные (производные логарифма) левой и правой частей предыдущего равенства:

$$\frac{-p}{1 - px} = \sum_{m=1}^{\infty} m I_m \frac{-x^{m-1}}{1 - x^m}.$$

Умножаем обе части последнего равенства на x, а также добавляем и вычитаем единицу в числителях всех дробей:

$$1 - \frac{1}{1 - px} = \sum_{m=1}^{\infty} m I_m \left(1 - \frac{1}{1 - x^m}\right).$$

Раскладывая в ряд, получаем:

$$\sum_{n=1}^{\infty} p^n x^n = \sum_{m=1}^{\infty} m I_m \left(x^m + x^{2m} + x^{3m} + \ldots\right).$$

Приравнивая коэффициенты при x^n, получаем, что

$$p^n = \sum_{d:\ d|n} d I_d.$$

Свойства чисел I_n:

1. Выполняется равенство $I_1 = p$.
2. Если n — простое, то

$$I_n = \frac{p^n - p}{n}.$$

Тем самым, получены точные значения I_n для $n = 1, 2, 3$.

3. При $n \geq 4$ справедливы неравенства

$$I_n \leq \frac{p^n - p}{n} < \frac{p^n}{n}.$$

▷ В сумме для представления p^n выделим первое и последнее слагаемое: $p^n \geq n I_n + I_1 = n I_n + p$. □

4. При $n \geq 4$ верна следующая оценка снизу на I_n:

$$I_n > \frac{p^n - p^{n/2+1}}{n}.$$

▷ Действительно, с использованием свойства 3 получаем соотношения

$$p^n = \sum_{d\colon d|n} dI_d \leq nI_n + \sum_{d=1}^{\lfloor n/2 \rfloor} dI_d \leq nI_n + \sum_{d=1}^{\lfloor n/2 \rfloor} d\frac{p^d}{d} < nI_n + p^{n/2+1},$$

из которых следует нужное неравенство. □

5. Для любого натурального n над полем $\mathbb{Z}_p$ существует нормированный неприводимый многочлен степени n.

▷ Неравенство $I_n > 0$, а следовательно и неравенство $I_n \geq 1$, непосредственно следует из свойств 1, 2 и 4. □

6. При $n \to \infty$ справедливо асимптотическое равенство

$$I_n \sim \frac{p^n}{n}.$$

▷ Асимптотика роста величины I_n с ростом n легко устанавливается с использованием свойств 3 и 4. □

7. Для любого натурального n справедливо равенство

$$I_n = \frac{1}{n} \sum_{d\colon d|n} \mu(d) p^{n/d}.$$

▷ Для доказательства этого равенства достаточно применить формулу обращения Мебиуса, положив $f(n) = p^n$, $g(n) = nI_n$ для всех натуральных n. □

§ 2.8 Тождества Ньютона

Здесь нам потребуется один фрагмент теории формальных степенных рядов.

Пусть K — коммутативное ассоциативное кольцо с единицей. Тогда в кольце $K[[x]]$ введем операцию D формального дифференцирования, положив для произвольного элемента $A(x) = \sum_{n=0}^{\infty} a_n x^n$ кольца $K[[x]]$

$$D(A(x)) = \sum_{n=1}^{\infty} n a_n x^{n-1},$$

где для произвольного элемента a кольца K под записью na понимается элемент

$$\underbrace{a + \ldots + a}_{n \text{ слагаемых}}$$

кольца K. Отметим, что $D(A(x)) \in K[[x]]$.

Свойства операции формального дифференцирования:

1. Пусть $A(x), B(x) \in K[[x]]$. Тогда

$$D(A(x) + B(x)) = D(A(x)) + D(B(x)).$$

2. Пусть $A(x), B(x) \in K[[x]]$. Тогда

$$D(A(x)B(x)) = D(A(x))B(x) + A(x)D(B(x)).$$

3. Пусть $A(x) \in K[[x]]$. Тогда

$$D(A^N(x)) = NA^{N-1}(x)D(A(x)).$$

4. Пусть $A(x)$ и $B(x)$ — обратимые элементы кольца $K[[x]]$. Тогда

$$\frac{D(A(x)B(x))}{A(x)B(x)} = \frac{D(A(x))}{A(x)} + \frac{D(B(x))}{B(x)}.$$

Теперь перейдем непосредственно к тождествам Ньютона.

Пусть F — некоторое поле, $F[\alpha_1, \ldots, \alpha_n]$ — кольцо многочленов от переменных $\alpha_1, \ldots, \alpha_n$ с коэффициентами из поля F, а $F[\alpha_1, \ldots, \alpha_n][[x]]$ — кольцо формальных степенных рядов над кольцом $F[\alpha_1, \ldots, \alpha_n]$.

Выделим в кольце $F[\alpha_1, \ldots, \alpha_n]$ множество *элементарных симметрических многочленов*:

$$\begin{aligned}
&\sigma_0 = 1,\\
&\sigma_1 = -(\alpha_1 + \ldots + \alpha_n),\\
&\sigma_2 = \alpha_1\alpha_2 + \alpha_1\alpha_3 + \ldots + \alpha_{n-1}\alpha_n,\\
&\quad \ldots \qquad \ldots \qquad \ldots\\
&\sigma_k = (-1)^k \sum_{1 \le i_1 < \ldots < i_k \le n} \alpha_{i_1}\alpha_{i_2}\ldots\alpha_{i_k},\\
&\quad \ldots \qquad \ldots \qquad \ldots\\
&\sigma_n = (-1)^n \alpha_1 \ldots \alpha_n,
\end{aligned}$$

а также множество *степенных сумм*:

$$\begin{aligned}
&S_0 = 0,\\
&S_k = \alpha_1^k + \alpha_2^k + \ldots + \alpha_n^k, \quad k = 1, 2, \ldots
\end{aligned}$$

Последовательностям $\{\sigma_k\}$ (считаем, что $\sigma_k = 0$ при $k > n$) и $\{S_k\}$ сопоставим формальные степенные ряды (производящие функции) из кольца $F[\alpha_1, \ldots, \alpha_n][[x]]$:

$$S(x) = \sum_{k=0}^{\infty} S_k x^k,$$

$$\sigma(x) = \sum_{k=0}^{\infty} \sigma_k x^k = \sum_{k=0}^{n} \sigma_k x^k = (1 - \alpha_1 x)(1 - \alpha_2 x) \ldots (1 - \alpha_n x).$$

Следующая цепочка равенств устанавливает связь между введенными формальными степенными рядами:

$$x \frac{D(\sigma(x))}{\sigma(x)} = x \sum_{k=1}^{n} \frac{D((1 - \alpha_k x))}{1 - \alpha_k x} = x \sum_{k=1}^{n} \frac{-\alpha_k}{1 - \alpha_k x} =$$

$$= \sum_{k=1}^{n} \left(1 - \frac{1}{1 - \alpha_k x}\right) = -\sum_{k=1}^{n} \left(\alpha_k x + \alpha_k^2 x^2 + \ldots\right) =$$

$$= -\sum_{i=1}^{\infty} S_i x^i = -S(x).$$

Таким образом,

$$xD(\sigma(x)) + S(x)\sigma(x) = 0.$$

С другой стороны,

$$xD(\sigma(x)) = \sum_{k=1}^{n} k\sigma_k x^k,$$

а $S(x)\sigma(x)$ — производящая функция последовательности, являющейся сверткой последовательностей $\{\sigma_k\}$ и $\{S_k\}$. Приравнивая коэффициенты при степенях формальной переменной x, получаем:

$$\left.\begin{array}{l} \sigma_1 + S_1 = 0, \\ 2\sigma_2 + S_2 + S_1\sigma_1 = 0, \\ \quad \ldots \qquad \ldots \qquad \ldots \\ k\sigma_k + S_k + S_{k-1}\sigma_1 + \ldots + S_1\sigma_{k-1} = 0, \\ \quad \ldots \qquad \ldots \qquad \ldots \\ n\sigma_n + S_n + S_{n-1}\sigma_1 + \ldots + S_1\sigma_{n-1} = 0, \end{array}\right\} \text{ формулы Ньютона}$$

$$\left.\begin{array}{l} S_{n+1} + S_n\sigma_1 + \ldots + S_1\sigma_n = 0, \\ \ldots \quad \ldots \quad \ldots \\ S_{n+k} + S_{n+k-1}\sigma_1 + \ldots + S_k\sigma_n = 0, \\ \ldots \quad \ldots \quad \ldots \end{array}\right\} \begin{array}{l} \text{обобщенные} \\ \text{формулы} \\ \text{Ньютона} \end{array}$$

Таким образом, степенные суммы однозначно выражаются через элементарные симметрические многочлены. Кроме того, если исходное поле F имеет нулевую характеристику, то и элементарные симметрические многочлены однозначно выражаются через степенные суммы.

В заключение первые n равенств, входящих в обобщенные формулы Ньютона, запишем в удобном для последующего использования матричном виде:

$$\begin{pmatrix} S_n & S_{n-1} & \ldots & S_1 \\ S_{n+1} & S_n & \ldots & S_2 \\ \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots \\ S_{2n-1} & S_{2n-2} & \ldots & S_n \end{pmatrix} \begin{pmatrix} \sigma_1 \\ \sigma_2 \\ \cdot \\ \cdot \\ \cdot \\ \sigma_n \end{pmatrix} = - \begin{pmatrix} S_{n+1} \\ S_{n+2} \\ \cdot \\ \cdot \\ \cdot \\ S_{2n} \end{pmatrix}.$$

Глава 3

Кодирование

Вопросы кодирования в математике возникали давно. Изначально они имели важное, но вспомогательное значение — например, изображение чисел в десятичной системе счисления, введение координат как алгебраических образов геометрических объектов. Толчком к формированию теории кодирования как самостоятельного раздела дискретной математики стало стремительное развитие таких направлений как связь и криптология, а также управляющих систем (простейшие примеры — диагностика двигателя или программирование сигнализации).

Процессы, так или иначе связанные с кодированием, схематично изображены на рис. 1.

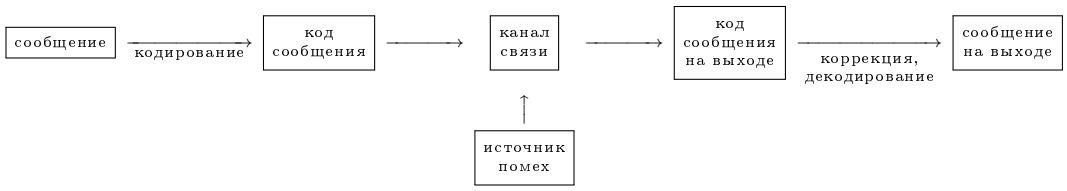

Рис. 1.

Введем некоторые обозначения.

Пусть $A = \{a_1, \ldots, a_r\}$ — исходный алфавит, $r \geq 2$. Тогда $\alpha = a_{i_1} a_{i_2} \ldots a_{i_k}$, где $a_{i_j} \in A$, — *слово* над алфавитом A, *длина* $l(\alpha)$ слова α равна k.

Обозначим через Λ пустое слово. Отметим, что для произвольного слова α справедливы равенства $\Lambda\alpha = \alpha\Lambda = \alpha$; кроме того, $l(\Lambda) = 0$, $\{\Lambda\} \neq \varnothing$.

Множество всех слов длины k над алфавитом A обозначим через A^k. Положим

$$A^+ = \bigcup_{k \geq 1} A^k, \quad A^* = \bigcup_{k \geq 0} A^k.$$

Пусть $B = \{b_1, \ldots, b_q\}$ — второй алфавит, $S \subset A^*$. Отображение

$$F: \ S \to B^*$$

называется *кодированием*, множество S — *множеством сообщений*, произвольное слово α из множества S — *сообщением*, слово $F(\alpha)$ из множества B^* — *кодом сообщения* α, а множество $F(S)$ — *кодом*.

Кодирование обладает свойством *взаимной однозначности*, если для любых двух различных сообщений α_1 и α_2 из множества сообщений S выполняется соотношение $F(\alpha_1) \neq F(\alpha_2)$.

Пример 1. Равномерное кодирование.

Пусть $\{\alpha_1, \ldots, \alpha_m\} \subset A^*$. Определим множество сообщений S, $S \subset A^*$, как множество всех слов $\{\alpha_1, \ldots, \alpha_m\}^*$, рассматриваемое как подмножество множества A^*, т. е как множество слов над алфавитом A. Зададим *схему кодирования* Σ следующим образом:

$$\alpha_i \to \beta_i, \quad l(\beta_i) = n; \qquad i = 1, \ldots, m.$$

Тогда схема Σ определяет кодирование F, заданное на множестве сообщений S:

$$\alpha_{i_1} \ldots \alpha_{i_k} \xrightarrow{F} \beta_{i_1} \ldots \beta_{i_k}.$$

Другим примером кодирования является алфавитное (побуквенное) кодирование, к более подробному изучению которого и переходим.

§ 3.1 Алфавитное кодирование

Пусть $A = \{a_1, \ldots, a_r\}$, $B = \{b_1, \ldots, b_q\}$. Схема Σ, определяемая следующим образом:

$$\begin{aligned} a_1 &\to v_1 = b_{11} b_{12} \ldots b_{1l_1}, \\ &\vdots \\ a_r &\to v_r = b_{r1} b_{r2} \ldots b_{rl_r}, \end{aligned} \tag{3.1}$$

порождает *алфавитное* (*побуквенное*) *кодирование* F:

$$a_{i_1} \ldots a_{i_k} \xrightarrow{F} v_{i_1} \ldots v_{i_k}.$$

Слова $v_1, \ldots, v_r$ называются *кодовыми словами*, множество кодовых слов $V = \{v_1, \ldots, v_r\}$ — *кодом алфавита* A. Для множества слов V будем допускать также название *алфавитный код* и даже просто код, хотя, формально говоря, в данной ситуации кодом будет множество слов $F(A^*)$.

Алфавитный код V называется *разделимым*, если из справедливости равенства

$$v_{i_1} v_{i_2} \dots v_{i_s} = v_{j_1} v_{j_2} \dots v_{j_t}$$

двух слов из множества V^* следуют равенства:

1) $s = t$;

2) $i_1 = j_1,\ i_2 = j_2,\ \dots,\ i_s = j_s$.

Свойства разделимых кодов:

1. $v_i \neq v_j$ при $i \neq j$.

2. $v_i \neq \Lambda$, $i = 1, \dots, r$.

3. Кодирование, осуществляемое с помощью разделимого кода, обладает свойством взаимной однозначности.

Примеры разделимых кодов:

1. Азбука Морзе.

2. Кодирование, задаваемое схемой:

$$a_1 \to 010,$$
$$a_2 \to 0.$$

3. Кодирование, задаваемое схемой:

$$a_1 \to 00,$$
$$a_2 \to 01,$$
$$a_3 \to 10,$$
$$a_4 \to 11.$$

Если слово $\alpha \in A^*$ представимо в виде $\alpha = \alpha_1 \alpha_2$, где $\alpha_1, \alpha_2 \in A^*$, то говорят, что α_1 — *префикс* слова α, а α_2 — *суффикс* слова α.

Отметим, что слова Λ и α являются и префиксами, и суффиксами слова α.

Алфавитный код V называется *префиксным*, если никакое кодовое слово v_i не является префиксом другого кодового слова v_j $(i \neq j)$.

Алфавитный код V называется *суффиксным*, если никакое кодовое слово v_i не является суффиксом другого кодового слова v_j $(i \neq j)$.

Алфавитное кодирование из примера 3 является и префиксным, и суффиксным, а кодирование из примера 2 не является ни префиксным, ни суффиксным.

Очевидно, что и префиксность, и суффиксность алфавитного кодирования являются достаточными условиями разделимости, но как показывает тот же пример 2, не являются необходимыми условиями разделимости.

Префиксный код удобно представлять в виде корневого помеченного дерева. Пусть $A = \{a_1, \ldots, a_r\}$, $B = \{b_1, \ldots, b_q\}$ и префиксный код из алфавита A в алфавит B задается схемой

$$a_i \to b_{i1} b_{i2} \ldots b_{il_i}, \quad i = 1, \ldots, r.$$

Такому коду естественным образом сопоставляется дерево с корнем: из каждой вершины исходит не более q ребер, которым приписываются буквы из алфавита B, концевым вершинам (т. е. вершинам степени 1) соответствуют буквы алфавита A, причем если концевой вершине приписана буква a_i, то ребрам на пути от корня до этой концевой вершины последовательно приписаны буквы $b_{i1}, b_{i2}, \ldots, b_{il_i}$ (ребру, инцидентному корню, приписана буква b_{i1}).

Построенное таким образом корневое помеченное дерево будем называть *кодовым деревом*.

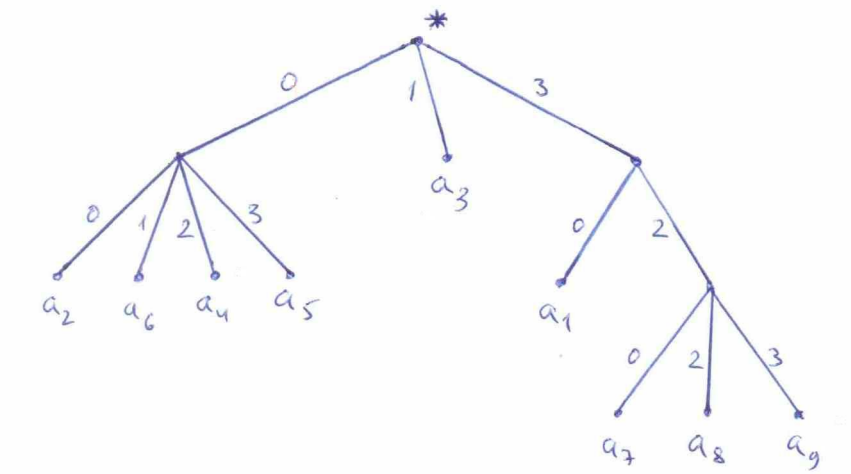

Рис. 3.1:

Пример 2. На рис. 3.1 представлено кодовое дерево для префиксного кода из алфавита $A = \{a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9\}$ в алфавит $B =$

$\{0, 1, 2, 3\}$, задаваемого схемой

$$
\begin{aligned}
a_1 &\to 30,\\
a_2 &\to 00,\\
a_3 &\to 1,\\
a_4 &\to 02,\\
a_5 &\to 03,\\
a_6 &\to 01,\\
a_7 &\to 320,\\
a_8 &\to 322,\\
a_9 &\to 323.
\end{aligned}
$$

Корень дерева помечен звездочкой.

Лекция № 9

Как проверить алфавитный код на разделимость? Проблема в том, что нужно проверить на однозначность декодирования бесконечное множество слов.

По схеме Σ, задаваемой соотношениями (3.1), определим следующие параметры

$$r = r(\Sigma) = |A|,$$

$$L = L(\Sigma) = l_1 + l_2 + \ldots + l_r,$$

$$W = W(\Sigma) = \max\{w \mid v_i = \beta_0 v_{i_1} v_{i_2} \ldots v_{i_w} \beta_1,\ i = 1, \ldots, r\},$$

при этом в последней формуле тривиальные разложения $v_i = v_i$ не учитываются.

Теорема 18 (Ал. А. Марков)**.** *Для любого алфавитного кодирования, задаваемого схемой Σ, существует такое число*

$$N \le \left\lfloor \frac{(W+1)(L-r+2)}{2} \right\rfloor,$$

что проблема взаимной однозначности кодирования всех слов сводится к аналогичной проблеме для слов из множества $\bigcup\limits_{i=0}^{N} A^i$.

Доказательство. Пусть v — слово из B^* минимальной длины, допускающее не менее двух способов расшифровки (декодирования).

Берем два разбиения слова v на кодовые слова (две расшифровки) и множество точек разбиения двух разбиений объединяем в одно множество. Это новое разбиение порождает множество отрезков (слов) двух типов: первый тип — отрезки, сответствующие кодовым словам, второй тип — все остальные отрезки (на рис. 3.2 одно разбиение слова v схематично изображено сверху, другое — снизу, а отрезки второго рода выделены жирным).

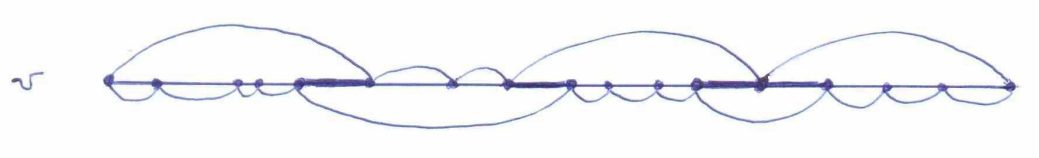

Рис. 3.2:

В силу минимальности исходного слова v все отрезки второго типа различны. Кроме того, любой отрезок второго типа является нетривиальным префиксом некоторого кодового слова v_i (т. е. отличным от Λ и v_i). Поэтому всего отрезков второго типа не более $(l_1-1)+(l_2-1)+\ldots+(l_r-1)=L-r$.

Отрезки второго типа разбивают слово v не более, чем на $L-r+1$ кусков. Между отрезками второго типа с номерами $i-1$ и i и между отрезками второго типа с номерами i и $i+1$ в одном и том же разбиении находится один раз ровно одно кодовое слово (включающее и граничные отрезки второго типа) и один раз не более W кодовых слов. Поэтому в каждом из изначальных разбиений (расшифровок) не более

$$\left\lceil \frac{L-r+1}{2} \right\rceil + W \left\lceil \frac{L-r+1}{2} \right\rceil \leq$$
$$\leq (W+1) \left\lfloor \frac{L-r+2}{2} \right\rfloor \leq \left\lfloor \frac{(W+1)(L-r+2)}{2} \right\rfloor$$

кодовых слов. □

Теорема 19 (неравенство Крафта — МакМиллана). *Пусть разделимый код задается схемой* $a_i \to v_i,\ v_i \in B^*,\ i=1,\ldots,r$. *Тогда*

$$\sum_{i=1}^{r} \frac{1}{q^{l(v_i)}} \leq 1,$$

где $q=|B|$.

Доказательство. Зафиксируем $n \in \mathbb{N}$. Обозначим через l максимальную длину кодового слова, т. е. $l = \max l(v_i)$. Положим

$$U = \{u \in B^* \mid u = v_{i_1} v_{i_2} \dots v_{i_n}\}, \quad U_k = \{u \in U \mid l(u) = k\}.$$

Рассмотрим производящую функцию последовательности $\{|U_k|\}$. С одной стороны,

$$\sum_{k=0}^{\infty} |U_k| x^k = \sum_{k=0}^{nl} |U_k| x^k,$$

с другой стороны,

$$\sum_{k=0}^{\infty} |U_k| x^k = \left(x^{l(v_1)} + x^{l(v_2)} + \ldots + x^{l(v_r)}\right)^n.$$

Поэтому, полагая $x = \frac{1}{q}$, имеем:

$$\left(\sum_{i=1}^{r} \frac{1}{q^{l(v_i)}}\right)^n = \sum_{k=0}^{nl} \frac{|U_k|}{q^k} \leq nl.$$

Таким образом для любого натурального n выполняется неравенство

$$\sum_{i=1}^{r} \frac{1}{q^{l(v_i)}} \leq \sqrt[n]{nl}.$$

Для завершения доказательства достаточно в последнем неравенстве перейти к пределу при $n \to \infty$. □

Теорема 20 (о построении префиксного кода с заданным набором длин)**.** *Пусть натуральные $l_1, l_2, \dots, l_r$ удовлетворяют неравенству*

$$\sum_{i=1}^{r} \frac{1}{q^{l_i}} \leq 1.$$

Тогда существует префиксный код, задаваемый схемой $a_i \to v_i$, $v_i \in \{0, 1, \dots, q-1\}^*$, $i = 1, \dots, r$, такой что $l(v_i) = l_i$, $i = 1, \dots, r$.

Доказательство. Без ограничения общности будем считать, что

$$l_1 \leq l_2 \leq \ldots \leq l_r.$$

Положим

$$n_1 = 0, \quad n_k = \sum_{i=1}^{k-1} q^{-l_i}, \ k = 2, \dots, r.$$

Тогда $0 = n_1 < n_2 < \ldots < n_r < 1$. При этом каждое число n_k ($k = 1, 2, \ldots, r$) является суммой чисел, у которых в записи по основанию q не более l_k разрядов после запятой. Кроме того все числа n_k меньше единицы. Поэтому число n_k в q-ичной системе счисления может быть записано следующим образом:

$$n_k = 0, b_{1k} b_{2k} \ldots b_{l_k k}, \quad k = 1, 2, \ldots, r,$$

где $b_{ij} \in \{0, 1, \ldots, q-1\}$.

Теперь рассмотрим код из алфавита $\{a_1, \ldots, a_r\}$ в алфавит $\{0, 1, \ldots, q-1\}$, задаваемый схемой:

$$a_k \to b_{1k} b_{2k} \ldots b_{l_k k} = v_k, \quad k = 1, 2, \ldots, r.$$

Условие $l(v_k) = l_k$ выполняется. Покажем что указанная схема задает префиксный код. Пусть это не так, т. е. существуют такие s и k, что

$$v_s = v_k b_{l_k+1,s} \ldots b_{l_s s}.$$

Но тогда $s > k$ (так как $l_s > l_k$) и справедливы соотношения

$$n_s = n_k + 0, \underbrace{00\ldots00}_{l_k \text{ разрядов}} b_{l_k+1,s} \ldots b_{l_s s} \le n_k + 0, \underbrace{00\ldots00}_{l_k \text{ разрядов}} \underbrace{(q-1)\ldots(q-1)}_{l_s - l_k \text{ разрядов}} <$$
$$< n_k + 0, \underbrace{00\ldots01}_{l_k \text{ разрядов}} = n_k + q^{-l_k} = n_{k+1} \le n_s.$$

Получили противоречие ($n_s < n_s$), которое завершает доказательство. □

Следствие 9. *Для любого разделимого кода существует префиксный код с таким же набором длин кодовых слов.*

Следствие 10. *Для существования разделимого кода с заданным набором $l_1, l_2, \ldots, l_r$ длин кодовых слов необходимо и достаточно выполнение неравенства*

$$\sum_{i=1}^{r} \frac{1}{q^{l_i}} \le 1.$$

3.1.1 Полные коды

Вспомним два примера разделимых кодов со стр. 63, задаваемых двумя разными схемами. В одном случае есть последовательности букв второго алфавита, которые никогда не могут встретиться (в примере 2 такой

последовательностью, в частности, является последовательность из двух единиц), а в другом случае (пример 3) таких последовательностей нет.

Разделимый алфавитный код, задаваемый схемой

$$a_i \to v_i,\ v_i \in B^*,\ i = 1, \ldots, r,$$

называется *полным*, если любое слово $v \in B^*$ представимо в виде

$$v = v_{i_1} \ldots v_{i_s} \beta,$$

где v_{i_j}, $j = 1, \ldots, s$, — кодовые слова, β — префикс некоторого кодового слова.

Свойство полных кодов:

Для любого слова $v \in B^*$ либо v — префикс некоторого кодового слова (соответствует случаю $s = 0$ в определении полного кода), либо некоторое кодовое слово — префикс слова v (соответствует случаю $s > 0$).

Теорема 21 (критерий полноты разделимого кода)**.** *Пусть алфавитный код V, задаваемый схемой*

$$a_i \to v_i,\ v_i \in B^*,\ i = 1, \ldots, r,$$

является разделимым. Тогда для того, чтобы код V был полным, необходимо и достаточно, чтобы выполнялось следующее условие: для любого слова $u \in B^*$, удовлетворяющего неравенству $l(u) \geq \max l(v_i)$, найдется кодовое слово v_j, являющееся префиксом слова u.

Доказательство. Необходимость следует непосредственно из определения полного кода.

Достаточность. Установим существование требуемого определением полного кода представления произвольного слова $u \in B^*$.

Будем «отрезать» от слова u слева кодовое слово, пока это сделать можно (это делать точно можно пока длина слова не менее чем $\max l(v_i)$). В какой-то момент получаем слово $u' \in B^*$, никакой префикс которого не является кодовым словом. Очевидно, что $l(u') < \max l(v_i)$.

Если $u' = \Lambda$, то искомое представление получено.

Если $u' \neq \Lambda$, то к слову u' допишем справа произвольным образом $\max l(v_i) - l(u')$ символов из множества B. Полученное слово u'' имеет длину $\max l(v_i)$ и, следовательно, по условию имеет в качестве префикса некоторое кодовое слово v_j. С другой стороны, по построению слово u'' имеет префикс u'. Таким образом слова v_j и u' являются префиксами одного и того же слова и следовательно одно из них является префиксом другого. Но слово u' по построению не имеет префикса в виде кодового слова. Поэтому u' — префикс кодового слова v_j. Следовательно искомое представление получено. □

Теорема 22 (критерий полноты кода)**.** *Пусть алфавитный код V, задан схемой*

$$a_i \to v_i,\ v_i \in B^*,\ i = 1, \ldots, r.$$

Тогда для того, чтобы код V был полным, необходимо и достаточно, чтобы код V был префиксным и выполнялось условие

$$\sum_{i=1}^{r} \frac{1}{q^{l(v_i)}} = 1,$$

где $q = |B|$.

Доказательство. Необходимость. Код V полный, и следовательно разделимый. Поэтому выполняется неравенство Крафта — МакМиллана

$$\sum_{i=1}^{r} \frac{1}{q^{l(v_i)}} \le 1.$$

Пусть $l = \max l(v_i)$. Обозначим через U_i множество слов из B^l (длины l) с префиксом v_i:

$$U_i = \{u \in B^l \mid u = v_i v,\ v \in B^*\}, \quad i = 1, \ldots, r.$$

В силу полноты кода выполняется равенство $B^l = \bigcup\limits_{i=1}^{r} U_i$. Поэтому

$$q^l = |B_l| = \left|\bigcup_{i=1}^{r} U_i\right| \le \sum_{i=1}^{r} |U_i| = \sum_{i=1}^{r} q^{l-l(v_i)} = q^l \sum_{i=1}^{r} q^{-l(v_i)} \le q^l.$$

Следовательно, содержащиеся в этой цепочке два нестрогих неравенства на самом деле являются равенствами. Второе из них устанавливает равенство в неравенстве Крафта — МакМиллана. А равенство

$$\left|\bigcup_{i=1}^{r} U_i\right| = \sum_{i=1}^{r} |U_i|$$

означает, что множества U_i, $i = 1, \ldots, r$, не пересекаются. Следовательно, алфавитный код V — префиксный.

Достаточность. Код V префиксный и поэтому разделимый. Покажем, что для любого слова $u \in B^*$, удовлетворяющего неравенству $l(u) \ge l$, найдется кодовое слово v_j, являющееся префиксом слова u.

Пусть $l(u) = n$. Положим

$$U_i = \{u \in B^n \mid u = v_i v,\ v \in B^*\}, \quad i = 1, \ldots, r.$$

Код V — префиксный, поэтому $U_i \cap U_j =$ при $i \neq j$. Тогда

$$\left|\bigcup_{i=1}^{r} U_i\right| = \sum_{i=1}^{r} |U_i| = \sum_{i=1}^{r} q^{n-l(v_i)} = q^n \sum_{i=1}^{r} q^{-l(v_i)} = q^n = |B^n|.$$

Следовательно существует такое i, что $u \in U_i$. По предыдущей теореме код V — полный. □

Теорема 23. *Пусть натуральные $l_1, \ldots, l_r$ удовлетворяют условию*

$$\sum_{i=1}^{r} 2^{-l_i} \leq 1.$$

Тогда существует полный код со схемой

$$a_i \to v_i,\ v_i \in \{0,1\}^*,\ i = 1, \ldots, r,$$

такой что выполняются неравенства $l(v_i) \leq l_i$, $i = 1, \ldots, r$.

Доказательство. Строим префиксный двоичный код с длинами кодовых слов $l_1, \ldots, l_r$ — согласно теореме 20 такой код существует. Теперь будем переделывать этот код сохраняя свойство префиксности.

Если $\sum\limits_{i=1}^{r} 2^{-l_i} = 1$, то по теореме 22 код полный.

Пусть $\sum\limits_{i=1}^{r} 2^{-l_i} < 1$. Без ограничения общности будем считать, что $l_1 \leq l_2 \leq \ldots \leq l_r$. Тогда

$$\sum_{i=1}^{r} 2^{-l_i} + 2^{-l_r} \leq 1.$$

Теперь, полагая $l'_1 = l_1, \ldots, l'_{r-1} = l_{r-1}$, $l'_r = l_r - 1$, получаем:

$$\sum_{i=1}^{r} 2^{-l'_i} = \sum_{i=1}^{r-1} 2^{-l_i} + 2^{-l'_r} = \sum_{i=1}^{r-1} 2^{-l_i} + 2^{-l_r} + 2^{-l_r} = \sum_{i=1}^{r} 2^{-l_i} + 2^{-l_r} \leq 1.$$

Последовательно уменьшая в наборе длин наибольшее значение на единицу, за конечное число шагов получим набор, для которого соответствующая сумма обращается в единицу. □

3.1.2 Оптимальное кодирование

Пусть $A = \{a_1, \ldots, a_r\}$, $B = \{b_1, \ldots, b_q\}$ — алфавиты, алфавитное кодирование задано схемой Σ:

$$a_1 \to v_1, \ldots, a_r \to v_r; \quad v_i \in B^*,\ i = 1, \ldots, r.$$

Пусть $V = \{v_1, \ldots, v_r\}$ — алфавитный код, заданный схемой Σ.

Будем считать, что заданы частоты появления символов из алфавита A и как следствие — их вероятности $p_i = p(a_i)$, удовлетворяющие условиям $p_i > 0,\ i = 1, \ldots, r$, $p_1 + \ldots + p_r = 1$.

Набор $P = (p_1, \ldots, p_r)$ будем называть *распределением*.

Далее везде считаем, что алфавитный код V — разделимый.

Определим *стоимость* $L_V(P)$ алфавитного кода V при распределении P равенством

$$L_V(P) = \sum_{i=1}^{r} p_i l(v_i).$$

Положим

$$L(P) = \inf L_V(P),$$

где инфинум берется по всем разделимым кодам V. Если справедливо равенство $L_P(V) = L(P)$, то код V называется *оптимальным кодом* (или *кодом с минимальной избыточностью*).

Свойства оптимальных кодов:

1. Для любого распределения существует оптимальный код.

$\triangleright$ Построим равномерный код $V = \{v_1, \ldots, v_r\}$. Тогда $l(v_i) = \lceil \log_q r \rceil$, $i = 1, \ldots, r$. Обозначим $p_{\min} = \min\{p_1, \ldots, p_r\}$. Любой код $V' = \{v'_1, \ldots, v'_r\}$, в котором найдется кодовое слово v'_i, удовлетворяющее неравенству

$$l(v'_i) > \frac{\lceil \log_q r \rceil}{p_{\min}},$$

не является оптимальным:

$$L_{V'}(P) > p_i \frac{\lceil \log_q r \rceil}{p_{\min}} \geq \lceil \log_q r \rceil = \sum_{i=1}^{r} p_i l(v_i) = L_V(P).$$

Таким образом, достаточно взять инфинум по конечному множеству кодов. $\square$

2. Для любого распределения существует оптимальный префиксный код.

3. В оптимальном коде:

а) если $p_i < p_j$, то $l(v_i) \geq l(v_j)$;

б) если $l(v_i) < l(v_j)$, то $p_i \geq p_j$.

$\triangleright$ Пусть это не так, т. е. $p_i < p_j$ и $l(v_i) < l(v_j)$. Тогда ввиду неравенства

$$(p_j - p_i)(l(v_j) - l(v_i)) > 0$$

выполняется соотношение

$$p_i l(v_i) + p_j l(v_j) > p_i l(v_j) + p_j l(v_i).$$

Поэтому, поменяв местами кодовые слова v_i и v_j, получим код меньшей стоимости — противоречие. $\square$

4. Справедлива следующая нижняя оценка стоимости оптимального кода:

$$L(P) \geq \sum_{i=1}^{r} p_i \log_q \frac{1}{p_i}.$$

$\triangleright$ Действительно,

$$\begin{aligned} -L(P) + \sum_{i=1}^{r} p_i \log_q \frac{1}{p_i} &= \sum_{i=1}^{r} \left(-l(v_i)p_i + p_i \log_q \frac{1}{p_i} \right) = \\ &= \sum_{i=1}^{r} p_i \log_q \frac{q^{-l(v_i)}}{p_i} = \sum_{i=1}^{r} p_i \frac{1}{\ln q} \ln \frac{q^{-l(v_i)}}{p_i} \leq \\ \leq \sum_{i=1}^{r} p_i \frac{1}{\ln q} \left(\frac{q^{-l(v_i)}}{p_i} - 1 \right) &= \frac{1}{\ln q} \left(\sum_{i=1}^{r} q^{-l(v_i)} - \sum_{i=1}^{r} p_i \right) \leq 0. \quad \square \end{aligned}$$

5. Построение кода, близкого к оптимальному, методом Шеннона.

Пусть $P = (p_1, \ldots, p_r)$, $p_i > 0$, $p_1 + \ldots + p_r = 1$, $q \geq 2$. Полагаем $l_i = \lceil \log_q \frac{1}{p_i} \rceil$. Тогда $-l_i \leq \log_q p_i$ и следовательно $q^{-l_i} \leq p_i$. Значит,

$$\sum_{i=1}^{r} q^{-l_i} \leq \sum_{i=1}^{r} p_i = 1.$$

При доказательстве теоремы 20 предъявлен алгоритм построения префиксного кода V с длинами кодовых слов, удовлетворяющих неравенству Крафта — МакМиллана. При этом

$$L_V(P) = \sum_{i=1}^{r} p_i \left\lceil \log_q \frac{1}{p_i} \right\rceil \leq \sum_{i=1}^{r} p_i \log_q \frac{1}{p_i} + \sum_{i=1}^{r} p_i = \sum_{i=1}^{r} p_i \log_q \frac{1}{p_i} + 1.$$

6. Величину $L(P)$ можно установить с точностью до единицы:

$$\sum_{i=1}^{r} p_i \log_q \frac{1}{p_i} \le L(P) \le \sum_{i=1}^{r} p_i \log_q \frac{1}{p_i} + 1.$$

7. Если все числа $\log_q \frac{1}{p_i}$ — натуральные, то код, построенный методом Шеннона, — оптимальный.

8. Пусть $V = \{v_1, \ldots, v_r\}$ — полный код. Тогда найдется распределение $P = \{p_1, \ldots, p_r\}$, для которого выполняется равенство $L_V(P) = L(P)$, т. е. при этом распределении код V — оптимальный.

$\triangleright$ Положим $p_i = q^{-l(v_i)}$, $i = 1, \ldots, r$. Тогда $p_1 + \ldots + p_r = 1$. Кроме того,

$$L_V(P) = \sum_{i=1}^{r} p_i l(v_i) = \sum_{i=1}^{r} p_i \log_q q^{l(v_i)} = \sum_{i=1}^{r} p_i \log_q \frac{1}{p_i} \le_{\text{(по свойству 4)}} L(P).$$

Следовательно, код V оптимальный. $\square$

Упражнение 15. Доказать, что двоичный ($q = 2$) префиксный оптимальный (для некоторого распределения) код является полным.

`Указание`. Если в неравенстве Крафта — МакМиллана строгое неравенство, то можно построить код с меньшим набором длин — противоречие с оптимальностью.

3.1.3 Построение оптимального кода

1. Можно искать оптимальный код в классе префиксных кодов.

2. Префиксный код задается кодовым деревом — корневым помеченным деревом, ребрам которого приписаны буквы из алфавита B, а концевым вершинам соответствуют буквы алфавита A. Припишем концевым вершинам вероятности соответствующих им букв алфавита A.

3. Остальным вершинам также припишем вероятности как суммы вероятностей всех концевых вершин поддерева с корнем в этой вершине.

4. Все вершины разобъем на ярусы, в зависимости от расстояния до корня. Корень — вершина нулевого уровня.

Свойства кодовых деревьев оптимальных префиксных кодов:

1. Если вероятность, приписанная вершине w_1, меньше вероятности, приписанной вершине w_2, то номер яруса вершины w_1 не меньше номера яруса вершины w_2.

2. Все «пучки» ребер, исходящие не из предпоследнего яруса, либо «насыщенные» (т. е. состоят из q ребер), либо пустые.

Без ограничения общности можно рассматривать приведенное кодовое дерево — такое дерево, в котором все «пучки» ребер, кроме, быть может, одного (исходящего из некоторой вершины предпоследнего яруса), — насыщенные, при этом число q_0 ребер в этом особом пучке определяется равенством

$$q_0 = \begin{cases} 2, & \text{если } q = 2; \\ q-1, & \text{если } r \equiv 0 \pmod{q-1}; \\ q, & \text{если } r \equiv 1 \pmod{q-1}; \\ r-(q-1)\lfloor \frac{r}{q-1} \rfloor, & \text{иначе,} \end{cases}$$

и эти ребра ведут в концевые вершины, которым приписаны q_0 самых маленьких вероятностей из исходного распределения.

Теорема 24 (о редукции)**.** *Пусть есть два кодовых дерева, причем второе дерево получается из первого путем замены концевой вершины на пучок из s ребер (при этом распределение P_2 получается из распределения P_1 соответствующей заменой p на $p_{i_1}, \ldots, p_{i_s}$). Тогда:*

1. *Если второй код — оптимальный, то первый — тоже оптимальный.*

2. *Если выполняются следующие условия: а) первый код — оптимальный; б) вероятности $p_{i_1}, \ldots, p_{i_s}$ — это s самых маленьких вероятностей в распределении P_2; в) $s = q_0$; то второй код — тоже оптимальный.*

Доказательство. Очевидно, что справедливо равенство $L_{V_1}(P_1) + p = L_{V_2}(P_2)$.

1. Предположим, что первый код не будет оптимальным. В этом случае найдется такой код V_1', что $L_{V_1'}(P_1) < L_{V_1}(P_1)$. Делая аналогичную замену вершины на пучок ребер в кодовом дереве кода V_1', получаем кодовое дерево некоторого кода V_2', для которого

$$L_{V_2'}(P_2) = L_{V_1'}(P_1) + p < L_{V_1}(P_1) + p = L_{V_2}(P_2),$$

что противоречит оптимальности кода V_2.

2. Предположим, что второй код не будет оптимальным. Тогда существует оптимальный код V_2'' с приведенным кодовым деревом, в котором есть соответствующий пучок из s ребер. Заменяя этот пучок на концевую вершину, получаем кодовое дерево некоторого кода V_1'', для которого

$$L_{V_1''}(P_1) = L_{V_2''}(P_2) - p < L_{V_2}(P_2) - p = L_{V_1}(P_1),$$

что противоречит оптимальности кода V_1. □

На использовании теоремы о редукции основан **метод Хаффмана** построения оптимального префиксного кода, заключающегося в последовательной замене q (на первом шаге q_0) букв, имеющих наименьшие вероятности, на одну новую, имеющую вероятность, равную сумме вероятностей заменяемых букв.

Пример 3. Построение оптимального кода методом Хаффмана в случае, когда $|A| = 11$, $|B| = 4$,

$$P = (0,25; 0,24; 0,15; 0,08; 0,07; 0,06; 0,05; 0,04; 0,03; 0,02; 0,01).$$

Сначала определяем, что $q_0 = 2$. Затем четырежды проделываем следующие процедуры: выписываем все вероятности в порядке убывания, заменяем 4 (в первый раз 2) самые маленькие вероятности на их сумму и переходим к новому распределению. Проделав это, последовательно в обратном порядке строим корневое дерево, заменяя вершину на пучок из 4 ребер (в последний раз — на пучок из 2 ребер), как показано на рис. 3.3.

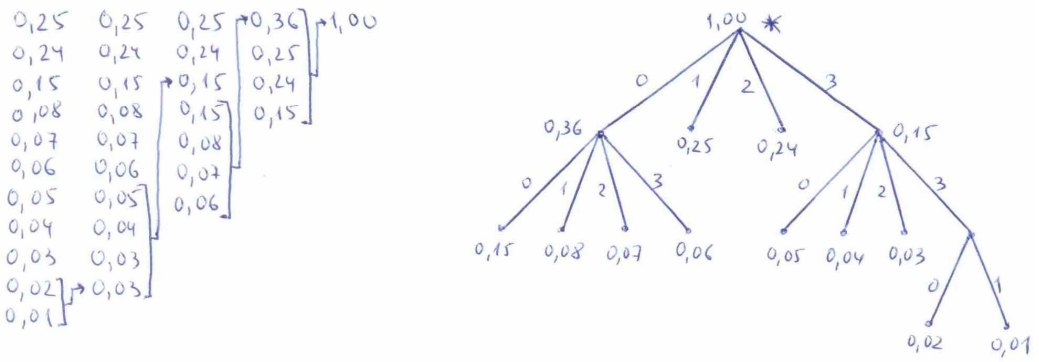

Рис. 3.3:

Пусть $p(a_1) = 0,25$, $p(a_2) = 0,24$, $p(a_3) = 0,15$, $p(a_4) = 0,08$, $p(a_5) = 0,07$, $p(a_6) = 0,06$, $p(a_7) = 0,05$, $p(a_8) = 0,04$, $p(a_9) = 0,03$, $p(a_{10}) = 0,02$, $p(a_{11}) = 0,01$. Тогда построенное кодовое дерево определяет такую схему кодирования:

$$a_1 \to 1, \quad a_2 \to 2, \quad a_3 \to 00, \quad a_4 \to 01,$$
$$a_5 \to 02, \quad a_6 \to 03, \quad a_7 \to 30, \quad a_8 \to 31,$$
$$a_9 \to 32, \quad a_{10} \to 330, \quad a_{11} \to 331.$$

Отметим, что если при третьем упорядочивании вероятностей поменять местами две одинаковые вероятности, равные 0,15, то мы построим оптимальный код, в котором будет три кодовых слова длины 1 и два кодовых слова длины 4.

Упражнение 16. При алфавитном кодировании четырехбуквенного алфавита с распределением $P = (p_1, p_2, p_3, p_4)$, $p_1 \geq p_2 \geq p_3 \geq p_4 > 0$, в двухбуквенный существует оптимальный код, содержащий кодовое слово длины 3, и существует оптимальный код, в котором нет кодовых слов длины 3. Найти все возможные значения, которые может принимать величина

а) p_1,
б) p_2,
в) p_3,
г) p_4.

Для каждого такого значения привести пример набора вероятностей, на котором это значение достигается.

Лекция № 11

§ 3.2 Коды, исправляющие ошибки

Пусть требуется по зашумленному каналу связи передать некоторое сообщение, т. е. конечный набор символов фиксированного алфавита (на самом деле, как правило, передается не само сообщение, а код этого сообщения, что соответствует схеме на рис. 1). Зашумленность подразумевает возможность искажения передаваемой информации:

$$x_1 \dots x_n \ \to \ \boxed{\text{КАНАЛ СВЯЗИ}} \ \to \ y_1 \dots y_m.$$

Обозначим через $R(\tilde{\alpha})$ множество слов, в которое может перейти сообщение $\tilde{\alpha}$ под воздействием «шума».

Если для любых сообщений $\tilde{\alpha}$ и $\tilde{\beta}$ справедливо соотношение $R(\tilde{\alpha}) \cap R(\tilde{\beta}) = \varnothing$, то код называется *самокорректирующимся* относительно заданного источника ошибок.

Далее будем предполагать, что возможны только ошибки типа замещения, а ошибки типа вставки или выпадения символа отсутсвуют. Отметим, что ошибки типа замещения являются наиболее характерными ошибками, особенно при хранении («передаче во времени») информации.

Теперь под кодом будем понимать множество слов одинаковой длины.

Пусть: p — простое, $B = \mathbb{Z}_p$ — поле вычетов по модулю p, $B^n = \{(\alpha_1, \ldots, \alpha_n) \mid \alpha_i \in B\}$.

Введем обозначения: $\rho(\tilde{\alpha}, \tilde{\beta})$ — *расстояние Хэмминга между наборами* $\tilde{\alpha}$ и $\tilde{\beta}$, т. е. число несовпадающих разрядов; $S_n^t(\tilde{\alpha}) = \{\tilde{\beta} \in B^n \mid \rho(\tilde{\alpha}, \tilde{\beta}) \leq t\}$ — n-мерный шар радиуса t с центром в $\tilde{\alpha}$.

Если в канале происходит не более t ошибок, то $R(\tilde{\alpha}) = S_n^t(\tilde{\alpha})$.

Код $V = \{\tilde{\alpha}^1, \ldots, \tilde{\alpha}^N\}$ *исправляет* t ошибок типа замещения, если для любых различных i и j выполняется неравенство $\rho(\tilde{\alpha}^i, \tilde{\alpha}^j) \geq 2t + 1$.

Код $V = \{\tilde{\alpha}^1, \ldots, \tilde{\alpha}^N\}$ *обнаруживает* t ошибок типа замещения, если для любых различных i и j выполняется неравенство $\rho(\tilde{\alpha}^i, \tilde{\alpha}^j) \geq t + 1$.

Положим $d(V) = \min \rho(\tilde{\alpha}, \tilde{\beta})$, где минимум берется по всем парам $(\tilde{\alpha}, \tilde{\beta})$ различных наборов из кода V. Величина $d(V)$ называется *кодовым расстоянием кода* V.

Код V обнаруживает $d(V) - 1$ ошибку и исправляет $\lfloor (d(V) - 1)/2 \rfloor$ ошибок.

Положим $N_n^t = |S_n^t(\tilde{\alpha})|$.

Теорема 25 (Граница Хэмминга или граница сферической упаковки)**.** *Пусть* V_n^t *— код из* B^n*, исправляющий* t *ошибок (типа замещения). Тогда*

$$|V_n^t| \leq \frac{p^n}{N_n^t}.$$

Доказательство. Из неравенства

$$|V_n^t| \; N_n^t \leq p^n$$

непосредственно следует утверждение теоремы. □

Если в границе Хэмминга равенство, то такой код называется *совершенным*.

Пример 4. $p = 2$, $n = 2k + 1$, $t = k$, $V = \{(0, \ldots, 0), (1, \ldots, 1)\}$.

Метод построения кода, исправляющего t ошибок

В качестве набора $\tilde{\alpha}^1$ берем произвольный набор, например, $\tilde{0}$.

Если наборы $\tilde{\alpha}^1, \ldots, \tilde{\alpha}^{k-1}$ уже построены, то набор $\tilde{\alpha}^k$ определяем из условия

$$\tilde{\alpha}^k \notin \bigcup_{i=1}^{k-1} S_n^{2t}(\tilde{\alpha}^i)$$

пока это можно сделать.

Получаем код V, исправляющий t ошибок. По построению

$$B^n \subseteq \bigcup_{i=1}^{|V|} S_n^{2t}(\tilde{\alpha}^i).$$

Поэтому

$$|V| \geq \frac{p^n}{N_n^{2t}}.$$

Положим $M_n^t(p) = \max |V_n^t|$, где максимум берется по всем кодам V_n^t, $V_n^t \subset B^n$, исправляющим t ошибок. Тогда

$$\frac{p^n}{N_n^{2t}} \leq M_n^t(p) \leq \frac{p^n}{N_n^t}.$$

Оценим число наборов в шаре радиуса t:

$$N_n^t = \sum_{k=0}^{t} C_n^k (p-1)^k,$$

$$n^k \frac{1}{k^k} \leq C_n^k = \frac{n(n-1)\dots(n-k+1)}{k(k-1)\dots(k-k+1)} \leq \frac{n^k}{k!} \leq n^k \frac{3^k}{k^k}.$$

Таким образом,

$$c'(t) n^t \leq N_n^t \leq c''(t) n^t,$$

и следовательно

$$c_1(t) \frac{p^n}{n^{2t}} \leq M_n^t(p) \leq c_2(t) \frac{p^n}{n^t}.$$

Содержательно это означает, что на исправление одной ошибки требуется «заложить» порядка $\log n$ разрядов.

Отметим, что

$$M_n^1(p) \leq \frac{p^n}{1+(p-1)n}, \quad M_n^1(2) \leq \frac{2^n}{1+n}.$$

§ 3.3 Линейные коды

Пусть $n \in \mathbb{N}$, $V \subset B^n$. Множество V будем называть *линейным* (*или групповым*) *кодом*, если для любых наборов $\tilde{\alpha}$ и $\tilde{\beta}$ из V набор $\tilde{\alpha} + \tilde{\beta}$ (сложение поразрядное по модулю p) тоже лежит в V.

Таким образом, V — подпространство пространства B^n.

Пусть вектора $\tilde{\alpha}^1, \ldots, \tilde{\alpha}^k$ — базис пространства V. Выписывая их в виде строк матрицы, получаем *порождающую матрицу*:

$$G = \begin{pmatrix} \alpha_1^1 & \ldots & \alpha_n^1 \\ \ldots & \ldots & \ldots \\ \alpha_1^k & \ldots & \alpha_n^k \end{pmatrix}_{k \times n}$$

Тогда

$$V = \{x \in B^n \mid x = (u_1, \ldots, u_k)G,\ u_1, \ldots, u_k \in B\}.$$

Параметр n — *длина* кода V, параметр k — *размерность* кода V.

Пусть $x, y \in B^n$. Положим

$$(x, y) = x_1 y_1 + \ldots + x_n y_n \pmod{p}.$$

Если справедливо равенство $(x, y) = 0$, то векторы x и y *дуальны*. Иногда такие векторы называют ортогональными, однако употребление этого термина не совсем корректно, так как введеная операция (x, y) над векторами x и y, вообще говоря, не является скалярным произведением, так как из равенства $(x, x) = 0$ не всегда следует равенство $x = 0$ — например, при $p = 2$ имеем: $((1, 1), (1, 1)) = 1 + 1 = 0$.

Положим

$$V^\perp = \{y \in B^n \mid \forall x\ (x \in V) \Rightarrow ((x, y) = 0)\};$$

$V^\perp$ — код, *двойственный* или *дуальный* к коду V.

Для двойственного кода выполняется равенство

$$\dim V^\perp = n - \dim V.$$

Действительно, размерность двойственного кода равна размерности пространства решений матричного уравнения $Gy^T = (0, \ldots, 0)^T$ (или уравнения $yG^T = (0, \ldots, 0)$) и, следовательно,

$$\dim V^\perp = n - \operatorname{rank} G = n - \dim V = n - k.$$

Отметим, что утверждение $V \cap V^\perp = \varnothing$, вообще говоря, неверно — например, при $p = 2$ имеем: $V = \{(0, 0), (1, 1)\} = V^\perp$.

Очевидно, что код, двойственный к двойственному, совпадает с исходным кодом:

$$(V^\perp)^\perp = V.$$

Возьмем $n-k$ линейно независимых векторов $\tilde{\beta}^1, \ldots, \tilde{\beta}^{n-k}$ из $V^\perp$ и запишем их в виде матрицы:

$$H = \begin{pmatrix} \beta_1^1 & \ldots & \beta_n^1 \\ \ldots & \ldots & \ldots \\ \beta_1^{n-k} & \ldots & \beta_n^{n-k} \end{pmatrix}_{(n-k)\times n}.$$

Тогда

$$V = (V^\perp)^\perp = \{y \in B^n \mid yH^T = (0, \ldots, 0)\} = \{y \in B^n \mid Hy^T = (0, \ldots, 0)^T\}.$$

Матрица H, являющаяся порождающей для кода $V^\perp$, называется *проверочной матрицей* для кода V.

Для линейного кода V определим еще один параметр — кодовое расстояние $d(V)$:

$$d(V) = \min_{x,y\in V,\ x\neq y} \rho(x,y) = \min_{x\in V,\ x\neq 0} \|x\|,$$

где $\|x\|$ — число ненулевых разрядов вектора x.

Код с кодовым расстоянием d исправляет $\left\lfloor \frac{d-1}{2} \right\rfloor$ ошибок.

Линейный код длины n, размерности k с кодовым расстоянием d называется $[n,k,d]$-кодом.

Если коды V и V' отличаются только в порядке разрядов, то такие коды называются *эквивалентными*.

Среди кодов, эквивалентных коду V, найдется код, порождающую матрицу которого элементарными операциями над строками можно привести (с сохранением свойства «порождаемости») к виду

$$G_{k\times n} = \left(I_k \ A_{k\times(n-k)}\right),$$

где I_k — единичная матрица порядка k. Без ограничения общности пусть таким кодом является сам код V. Тогда при порождении элемента $(x_1, \ldots, x_n)$ кода V произвольным набором $(u_1, \ldots, u_k)$ по правилу $(x_1, \ldots, x_n) = (u_1, \ldots, u_k)G$ выполняются равенства $x_1 = u_1, \ldots, x_k = u_k$. В этом случае эти первые k разрядов вектора x называются *информационными*, а остальные $n-k$ разрядов — *проверочными*.

Матрица H, определяемая равенством

$$H_{(n-k)\times n} = \left(-A^T \ I_{n-k}\right),$$

будет проверочной для кода V, так как справедливы равенства $GH^T = 0_{k\times(n-k)}$ и $HG^T = 0_{(n-k)\times k}$.

Под словами «проверочная матрица» иногда будем понимать любую матрицу, строками которой являются вектора системы (не обязательно линейно независимой), порождающей пространство $V^{\perp}$.

Следующее утверждение справедливо для проверочной матрицы в таком более широком смысле.

Теорема 26. *Пусть H — матрица над B, V — линейный код с проверочной матрицей H. Тогда для того, чтобы кодовое расстояние кода V было равно d, необходимо и достаточно, чтобы в матрице H любые $d-1$ столбцов были линейно независимыми над B и существовали d линейно зависимых столцов над B.*

Доказательство. Следующие утверждения равносильны.

1. Справедливо равенство $d(V) = d$.

2. Для любого ненулевого вектора $x \in B^n$ если выполняется условие $\|x\| \leq d-1$, то $x \notin V$; при этом найдется вектор $x \in V$, такой что $\|x\| = d$.

3. Никакая линейная комбинация из $d-1$ столбцов матрицы H не равна 0; существует линейная комбинация из d столбцов, равная 0. $\square$

Следствие 11. *Пусть H — матрица над B размера $(n-k) \times n$, причем любые $d-1$ столбцов матрицы H линейно независимы, V — линейный код с проверочной (в широком смысле) матрицей H. Тогда*

$$n(V) = n, \quad k(V) \geq k, \quad d(V) \geq d.$$

Следствие 12 (Граница Синглтона для линейных кодов)**.** *Для любого линейного кода V справедливо неравенство*

$$k(V) \leq n(V) - d(V) + 1.$$

Это утверждение вытекает из того факта, что число $d(V) - 1$ линейно независимых столбцов в проверочной (в узком смысле) матрице не превосходит числа строк в этой матрице, т. е. величины $n(V) - k(V)$.

Пример 5. Двоичный код Хэмминга (исправляющий одну ошибку).

Код V задается проверочной матрицей H размера $m \times (2^m - 1)$, в которой i-й столбец есть двоичная запись числа i:

$$H = \begin{pmatrix} 0 & 0 & 0 & \dots & 1 \\ . & . & . & \dots & . \\ . & . & . & \dots & . \\ . & . & . & \dots & . \\ 0 & 1 & 1 & \dots & 1 \\ 1 & 0 & 1 & \dots & 1 \end{pmatrix}.$$

Параметры кода V:

$$n(V) = 2^m - 1, \quad k(V) = n(V) - m = 2^m - m - 1, \quad d(V) = 3.$$

Декодирование. Пусть $y = x + e$, где e — вектор ошибки (нулевой, если нет ошибок, и единичный с единицей в i-м разряде, если произошла одна ошибка в i-м разряде). Найдем *синдром*, т. е. вектор Hy^T.

Если $Hy^T = (0, \ldots, 0)^T$, то ошибок не было.

Если произошла одна ошибка в i-м разряде, то

$$Hy^T = H\left(x^T + e_i^T\right) = Hx^T + He_i^T = (0, \ldots, 0)^T + h_i,$$

где h_i — i-й столбец матрицы H, который, в свою очередь, является двоичной записью числа i — разряда, в котором произошла ошибка.

Проверим на совершенность:

$$2^{n(V)} = 2^{2^m - 1}, \quad |V| = 2^{k(V)} = 2^{2^m - m - 1}, \quad |S_n^1| = n + 1 = 2^m.$$

Поэтому $|V| \times |S_n^1| = 2^{n(V)}$, и, следовательно, двоичный код Хэмминга совершенный.

Пример 6. Расширенный код Хэмминга.

«Расширение» заключается в том, что к проверочной матрице двоичного кода Хэмминга дописывается слева нулевой столбец высоты m, а затем сверху дописывается единичная строка длины 2^m. Выпишем параметры расширенного кода Хэмминга V, задаваемого описанной выше проверочной матрицей:

$$n(V) = 2^m, \quad k(V) = n(V) - (m + 1) = 2^m - m - 1, \quad d(V) = 4.$$

Кодовое расстояние равно 4, так как любые три столбца линейно независимы в силу того, что в линейной комбинации первый разряд отличен от нуля.

Код обнаруживает три ошибки (синдром отличен от нуля в случае, если произошло 1–3 ошибки).

Если произошла одиночная ошибка, то первый разряд синдрома равен единице, а остальные m разрядов указывают на номер разряда с ошибкой (разряды имеют номера $0, 1, \ldots, 2^m - 1$).

Если произошла двойная ошибка, то первый разряд синдрома равен нулю, а остальные m разрядов — не все нулевые.

Расширенный код Хэмминга не является совершенным, так как $|S_n^1| = n + 1 = 2^m + 1$ — не степень двойки. Но:

Упражнение 17. Расширенный код Хэмминга является максимальным кодом с кодовым расстоянием 4.

Лекция № 12

Пример 7. Обобщенный код Хэмминга (исправляющий одну ошибку) над $\mathbb{Z}_p$.

Код V задается проверочной матрицей H, состоящей из m строк. Столбцы матрицы разбиваются на m групп. В i-й группе, $i = 0, 1, \ldots, m-1$, первые $m - i - 1$ разрядов — нулевые, в разряде с номером $m - i$ во всех столбцах группы стоит единица, а в остальных i разрядах столбцов этой группы — все возможные наборы из $\{0, 1, \ldots, p-1\}^i$ (тем самым, в i-й группе p^i столбцов). Таким образом, матрица H имеет вид:

$$\begin{pmatrix}
0 & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & \ldots & 1 & 1 & \ldots & 1 \\
0 & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & \ldots & 0 & 0 & \ldots & p-1 \\
\cdot & \cdot & \cdot & \ldots & \cdot & \cdot & \cdot & \ldots & \cdot & \ldots & \cdot & \cdot & \ldots & \cdot \\
\cdot & \cdot & \cdot & \ldots & \cdot & \cdot & \cdot & \ldots & \cdot & \ldots & \cdot & \cdot & \ldots & \cdot \\
\cdot & \cdot & \cdot & \ldots & \cdot & \cdot & \cdot & \ldots & \cdot & \ldots & \cdot & \cdot & \ldots & \cdot \\
0 & 0 & 0 & \ldots & 0 & 1 & 1 & \ldots & 1 & \ldots & 0 & 0 & \ldots & p-1 \\
0 & 1 & 1 & \ldots & 1 & 0 & 0 & \ldots & p-1 & \ldots & 0 & 0 & \ldots & p-1 \\
1 & 0 & 1 & \ldots & p-1 & 0 & 1 & \ldots & p-1 & \ldots & 0 & 1 & \ldots & p-1
\end{pmatrix}.$$

Параметры кода V:

$$n(V) = 1 + p + p^2 + \ldots + p^{m-1} = \frac{p^m - 1}{p - 1},$$
$$k(V) = n(V) - \operatorname{rank} H = n - m = \frac{p^m - 1}{p - 1} - m,$$
$$d(V) = 3.$$

Оценим сверху число проверочных разрядов $m = n(V) - k(V)$ (число всех разрядов минус число информационных разрядов). Так как $n(V) = \frac{p^m - 1}{p-1}$, то $p^m = n(p-1) + 1$ и, следовательно, $m \leq \log_p n + 1$.

Декодирование. Находим синдром

$$S = Hy^T = \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_m \end{pmatrix}.$$

Если $S = (0, \ldots, 0)^T$, то ошибок нет, ничего не делаем.

Если $S \neq (0,\ldots,0)^T$, то выделив первый ненулевой разряд s_i, $i \in \{1,\ldots,m\}$, покоординатно разделим синдром на s_i:

$$S/s_i = (0,\ldots,0,1,s_{i+1}/s_i,\ldots,s_m/s_i)^T.$$

Далее, находим в матрице H столбец $(0,\ldots,0,1,s_{i+1}/s_i,\ldots,s_m/s_i)^T$. Пусть этот столбец имеет номер l. Тогда $x = y - e$, где вектор e определяется следующим образом: $e_l = s_i$ и $e_j = 0$ при $j \neq l$.

Проверка корректности:

$$Hx^T = Hy^T - He^T = (0,\ldots,0,s_i,s_{i+1},\ldots,s_m)^T - s_i h_l = (0,\ldots,0)^T,$$

следовательно $x \in V$. Кроме того, $\rho(x,y) = 1$.

Проверим обобщенный код Хэмминга на совершенность:

$$|V| = p^{k(V)} = p^{n-m}, \quad |S_n^1| = (p-1)n + 1 = \frac{p^m - 1}{p-1}(p-1) + 1 = p^m.$$

Поэтому $|V| \times |S_n^1| = p^{n(V)}$, и, следовательно, обобщенный код Хэмминга совершенный.

3.3.1 Алгоритм построения матриц, в которых любые $d-1$ столбцов линейно независимы

Строим над B матрицы из m строк, добавляя на каждом шаге очередной столбец так, чтобы в получающейся матрице любые $d-1$ столбцов были линейно независимы.

В качестве первого столбца h_1 берем произвольный ненулевой столбец высоты m.

Пусть столбцы $h_1,\ldots,h_{i-1}$ построены. Любые $d-1$ из них линейно независимы. Нужно добавить столбец h_i таким образом, чтобы он вместе с любыми $d-2$ столбцами из множества $h_1,\ldots,h_{i-1}$ образовывал линейно независимую систему, т. е. чтобы он не выражался ни через какие $d-2$ из них. Сколько линейных комбинаций столбцов $h_1,\ldots,h_{i-1}$ с не более чем $d-2$ слагаемыми? Обозначим это число через N и оценим его сверху:

$$N \leq 1 + C_{i-1}^1(p-1) + C_{i-1}^2(p-1)^2 + \ldots + C_{i-1}^{d-2}(p-1)^{d-2}.$$

Если выполняется неравенство $1 + C_{i-1}^1(p-1) + C_{i-1}^2(p-1)^2 + \ldots + C_{i-1}^{d-2}(p-1)^{d-2} < p^m$, то $N < p^m$ и, следовательно, найдется столбец h_i, такой что любые $d-1$ столбцов из множества $h_1,\ldots,h_i$ линейно независимы.

Таким образом, справедливо следующее утверждение.

Теорема 27 (Граница Варшамова — Гилберта)**.** *Пусть* $n \geq 2$, $d \geq 2$, $m \geq 1$. *Если выполняется неравенство*

$$1 + C_{n-1}^1(p-1) + C_{n-1}^2(p-1)^2 + \ldots + C_{n-1}^{d-2}(p-1)^{d-2} < p^m,$$

то существует линейный код с параметрами

$$n(V) = n, \quad k(V) \geq n - m, \quad d(V) \geq d.$$

Отметим, что в настоящее время неизвестно, лежат ли на границе какие-либо коды.

3.3.2 Коды Рида — Маллера

Для произвольной булевой функции f обозначим через π_f ее полином Жегалкина, а через $\deg \pi_f$ — степень этого полинома.

Пусть натуральные m и r удовлетворяют неравенствам $0 \leq r \leq m$. Зафиксируем некоторый (например, лексикографический) порядок наборов из множества E^m.

Кодом Рида — Маллера $R(r,m)$ *длины* 2^m *порядка* r называется множество векторов значений всех m-местных булевых функций, у каждой из которых степень полинома Жегалкина не превосходит r. Для краткости будем говорить не о векторах значений функций, а просто о функциях. Тогда можно условно записать так:

$$R(r,m) = \{f(x_1, \ldots, x_m) \mid \deg \pi_f \leq r\}.$$

Приведем простейшие примеры кодов Рида — Маллера:

$$R(0,m) = \Big\{(\underbrace{0\ldots0}_{2^m}), (\underbrace{1\ldots1}_{2^m})\Big\},$$
$$R(1,m) = L(m), \quad R(m,m) = P_2(m).$$

Свойства кодов Рида — Маллера:

1. Коды Рида — Маллера являются линейными кодами.

▷ Следует из неравенства $\deg(\pi_{f+g}) \leq \max\{\deg(\pi_f), \deg(\pi_g)\}$. □

2. Пусть $f \in R(r,m)$, $f \not\equiv 0$. Тогда

$$\|f\| \geq 2^{m-r},$$

где $\|f\|$ — вес функции f, т. е. число наборов из E^m, на которых функция f обращается в единицу.

$\triangleright$ Приведем два доказательства этого факта.

Первое доказательство проведем индукцией по r, а при фиксированном r — по m $(m \geq r)$.

База индукции.

При $r = 0$ единственной отличной от константы 0 функцией является константа 1, для которой требуемое неравенство выполняется.

При $m = r$ доказываемое неравенство превращается в соотношение $\|f\| \geq 1$, которое справедливо для любой отличной от константы 0 функции.

Шаг индукции. Пусть для всех пар (r', m'), удовлетворяющих условиям $0 \leq r' \leq m'$, $r' \leq r$, $m' \leq m$, $r' + m' < r + m$, требуемое неравенство установлено.

Представим произвольную отличную от константы 0 функцию f из кода $R(r, m)$ в следующем виде:

$$f(x_1, x_2, \ldots, x_m) = (x_1 \oplus 1) f(0, x_2, \ldots, x_m) \oplus x_1 f(1, x_2, \ldots, x_m).$$

Обозначим $f_0 = f(0, x_2, \ldots, x_m)$, $f_1 = f(1, x_2, \ldots, x_m)$. Отдельно рассмотрим три случая.

Случай 1°: $f_0 \not\equiv 0$, $f_1 \not\equiv 0$.

Тогда в силу включений $f_0 \in R(r, m-1)$, $f_1 \in R(r, m-1)$ к функциям f_0 и f_1 можно применить предположение индукции. Поэтому

$$\|f\| = \|f_0\| + \|f_1\| \geq 2^{m-1-r} + 2^{m-1-r} = 2^{m-r}.$$

Случай 2°: $f_0 \equiv 0$.

В условиях этого случая справедливы соотношения $f_1 \not\equiv 0$, $f_1 \in R(r-1, m-1)$. С использованием предположения индукции получаем, что

$$\|f\| = \|f_1\| \geq 2^{(m-1)-(r-1)} = 2^{m-r}.$$

Случай 3°: $f_1 \equiv 0$.

В условиях этого случая справедливы соотношения $f_0 \not\equiv 0$, $f_0 \in R(r-1, m-1)$. С использованием предположения индукции получаем, что

$$\|f\| = \|f_0\| \geq 2^{(m-1)-(r-1)} = 2^{m-r}.$$

Таким образом, во всех случаях верно неравенство $\|f\| \geq 2^{m-r}$.

Второе доказательство. Выделим в полиноме Жегалкина π_f какую-либо конъюнкцию максимальной длины. Без ограничения общности будем считать, что это конъюнкция $x_1 x_2 \ldots x_s$, $s \leq r$. Разложим функцию $f(x_1, \ldots, x_m)$ по последним $m - s$ переменным:

$$f(x_1, \ldots, x_m) = \bigvee_{(\sigma_{s+1}, \ldots, \sigma_m)} f(x_1, \ldots, x_s, \sigma_{s+1}, \ldots, \sigma_m) x_{s+1}^{\sigma_{s+1}} \ldots x_m^{\sigma_m}.$$

Для любого из 2^{m-s} наборов $(\sigma_{s+1}, \ldots, \sigma_m)$ в полиноме Жегалкина функции $f(x_1, \ldots, x_s, \sigma_{s+1}, \ldots, \sigma_m)$ будет присутствовать конъюнкция $x_1 x_2 \ldots x_s$ и, следовательно, эта функция не является константой 0. Поэтому для любого из 2^{m-s} наборов $(\sigma_{s+1}, \ldots, \sigma_m)$ найдется набор $(\alpha_1, \ldots, \alpha_s, \sigma_{s+1}, \ldots, \sigma_m)$, на котором функция $f(x_1, \ldots, x_s, \sigma_{s+1}, \ldots, \sigma_m) x_{s+1}^{\sigma_{s+1}} \ldots x_m^{\sigma_m}$ обращается в единицу. Следовательно, $\|f\| \geq 2^{m-s} \geq 2^{m-r}$. □

3. Для параметров (длина, размерность и кодовое расстояние) кода Рида — Маллера $R(r, m)$ справедливы равенства

$$\begin{aligned} n(R(r,m)) &= 2^m, \\ k(R(r,m)) &= 1 + C_m^1 + \ldots + C_m^r, \\ d(R(r,m)) &= 2^{m-r}. \end{aligned}$$

4. Порождающей матрицей размера $(1 + C_m^1 + \ldots + C_m^r) \times 2^m$ кода Рида — Маллера $R(r, m)$ является матрица G, состоящая из единичной строки и строк, являющихся векторами значений всевозможных функций вида $x_{i_1} \ldots x_{i_l}$, $1 \leq l \leq r$, как функций переменных $x_1, \ldots, x_m$.

3.3.3 Декодирование кодов Рида — Маллера

Сначала подробно изложим рассуждения, лежащие в основе эффективного метода декодирования кодов Рида — Маллера, не использующий вычисления синдрома декодируемого набора.

Пусть функция $f(x_1, \ldots, x_m)$ удовлетворяет условию $\deg(f) \leq r$. Опишем, как можно установить, входит ли моном степени r, скажем моном $x_1 \ldots x_r$, в полином π_f, если известен не вектор значений функции f, а вектор, отличающийся от него не более, чем в $2^{m-r-1} - 1$ разряде.

Для каждого набора $\tilde{\sigma} = (\sigma_1, \ldots, \sigma_{m-r})$ из $(m-r)$-мерного единичного куба E^{m-r} определим следующее подмножество m-мерного единичного куба E^m:

$$E_{\tilde{\sigma}}^r = \{(\alpha_1, \ldots, \alpha_m) \in E^m \mid (\alpha_{r+1}, \ldots, \alpha_m) = (\sigma_1, \ldots, \sigma_{m-r})\}.$$

Свойства множеств $E_{\tilde{\sigma}}^r$:

1. Для любого набора $\tilde{\sigma} \in E^{m-r}$ верно равенство $|E_{\tilde{\sigma}}^r| = 2^r$.
2. Пусть $\tilde{\sigma}, \tilde{\tau} \in E^{m-r}$, $\tilde{\sigma} \neq \tilde{\tau}$. Тогда $E_{\tilde{\sigma}}^r \cap E_{\tilde{\tau}}^r = \varnothing$.
3. Справедливо равенство

$$E^n = \bigcup_{\tilde{\sigma} \in E^{m-r}} E_{\tilde{\sigma}}^r.$$

4. Для любого набора $\tilde{\sigma} \in E^{m-r}$ для конъюнкции $K_0(x_1, \ldots, x_m) = x_1 \ldots x_r$ выполняется равенство

$$\bigoplus_{(\alpha_1,\ldots,\alpha_m)\in E^r_{\tilde{\sigma}}} K_0(\alpha_1, \ldots, \alpha_m) = 1.$$

5. Для любого набора $\tilde{\sigma} \in E^{m-r}$ и для любой отличной от $x_1 \ldots x_r$ конъюнкции $K(x_1, \ldots, x_m)$ вида $ax_{i_1} \ldots x_{i_s}$, где $0 \leq s \leq r$, $a \in \{0, 1\}$, выполняется равенство

$$\bigoplus_{(\alpha_1,\ldots,\alpha_m)\in E^r_{\tilde{\sigma}}} K(\alpha_1, \ldots, \alpha_m) = 0.$$

$\triangleright$ Любая конъюнкция $K(x_1, \ldots, x_m)$ вида $ax_{i_1} \ldots x_{i_s}$, где $0 \leq s \leq r$, $a \in \{0, 1\}$, отличная от конъюнкции $x_1 \ldots x_r$, имеет хотя бы одну фиктивную переменную из множества $\{x_1, \ldots, x_r\}$. Поэтому для любого набора $\tilde{\sigma} \in E^{m-r}$функция $K(x_1, \ldots, x_m)$ принимает единичное значение на четном числе наборов из множества $E^r_{\tilde{\sigma}}$. $\square$

Переходя непосредственно к решению задачи о вхождении монома $x_1 \ldots x_r$ в полином π_f, отметим, что в силу свойств множеств $E^r_{\tilde{\sigma}}$ для любого набора $\tilde{\sigma}$ из множества E^{m-r} сумма

$$\bigoplus_{(\alpha_1,\ldots,\alpha_m)\in E^r_{\tilde{\sigma}}} f(\alpha_1, \ldots, \alpha_m)$$

равна 1, если моном $x_1 \ldots x_r$ входит в полином π_f, и равна 0 в противном случае. Тем самым, в случае, когда известен вектор значений функции f, получаем 2^{m-r} (такова мощность множества E^{m-r}) независимых контрольных (для монома $x_1 \ldots x_r$) сумм, четность каждой из которых определяет наличие или отсутствие монома $x_1 \ldots x_r$ в полиноме π_f.

Далее, если изменить значение функции f не более чем на $2^{m-r-1} - 1$ наборах, то изменится значение не более $2^{m-r-1} - 1$ контрольных сумм, а значение не менее чем $2^{m-r-1} + 1$ контрольных сумм не изменятся. Таким образом, достаточно установить четность большинства контрольных сумм. Если большинство контрольных сумм нечетно, то моном $x_1 \ldots x_r$ входит в полином π_f, а если четно, то не входит.

Мажоритарный алгоритм декодирования (алгоритм Рида)

Пусть вектор значений функции $f(x_1, \ldots, x_m)$ является элементом кода Рида — Маллера $R(r, m)$, а нам известен вектор, отличающийся от него не более чем в $2^{m-r-1} - 1$ разрядах и, соответственно, задающий некоторую функцию $f(x_1, \ldots, x_m) \oplus g(x_1, \ldots, x_m)$, где $\|g\| \leq 2^{m-r-1} - 1$.

Так как выполняется неравенство $\deg \pi_f \le r$, то в полиноме Жегалкина π_f отсутствуют конъюнкции длины более r.

На первом этапе для каждой конъюнкции $x_{i_1} \dots x_{i_r}$ длины r последовательно выясняем входит ли эта конъюнкция в полином π_f.

Наличие конъюнкции $x_1 \dots x_r$ в полиноме π_f определяется большинством из 2^{m-r} значений (для всех наборов $\tilde{\sigma}$ из E^{m-r}) выражений

$$\bigoplus_{(\alpha_1,\dots,\alpha_m)\in E^r_{\tilde{\sigma}}} \big(f(\alpha_1,\dots,\alpha_m) \oplus g(\alpha_1,\dots,\alpha_m)\big).$$

Если большинство значений равно 1, то конъюнкция $x_1 \dots x_r$ входит в полином π_f, в противном случае не входит.

Аналогично проверяем наличие в полиноме π_f остальных конъюнкций длины r, естественные отличия связаны только с изменениями в определении множеств $E^{m-r}_{\tilde{\sigma}}$. Тем самым на первом этапе в полиноме π_f устанавливаются все слагаемые степени r. Обозначим через $h_r(x_1,\dots,x_m)$ функцию, полином Жегалкина которой состоит именно из этих слагаемых (положим $h_r = 0$ если таких слагаемых нет).

На втором этапе вместо функции $f(x_1,\dots,x_m)$ исследуется функция $f_1(x_1,\dots,x_m) = f(x_1,\dots,x_m) \oplus h_r(x_1,\dots,x_m)$, удовлетворяющая условию $\deg(\pi_{f_1}) \le r-1$. Вектор значений этой функции неизвестен, но известен вектор значений функции $f_1 \oplus g$, отличающийся от вектора значений функции f_1 не более чем в $2^{m-r-1}-1$ разрядах. Так как $2^{m-r-1}-1 < 2^{m-(r-1)-1}-1$, то для выявления слагаемых степени $r-1$ в полиноме π_{f_1} (а следовательно, и в полиноме π_f) достаточно проделать такие же процедуры, как и на предыдущем этапе.

Поступая аналогичным образом на этапах с номерами $3,\dots,r+1$ в конце концов установим все слагаемые полинома π_f.

Упражнение 18. Доказать, что код Рида — Маллера $R(m-2,m)$ совпадает с расширенным кодом Хэмминга.

Лекция № 13

3.3.4 Повторение сведений из алгебры. Конечные поля

Пусть $F = (F,+,\times)$ — поле, т. е. коммутативное ассоциативное кольцо с единицей, в котором каждый ненулевой элемент обратим.

Обозначим через $|F|$ *порядок* поля, т. е. число элементов поля. Если $|F| < \infty$, то поле F называется конечным. Пример: $\mathbb{Z}_p$.

Характеристикой поля называется минимальное значение p, такое что сумма p копий единицы дает ноль. Если такого числа не существует, то полагают, что характеристика поля равна 0.

Упражнение 19. Привести пример бесконечного поля ненулевой характеристики.

Свойства конечных полей

1. Характеристика поля — простое число (иначе есть делители нуля).

2. Пусть $|F| = q$. Тогда существует n, такое что $q = p^n$, где p — характеристика поля F.

$\triangleright$ Положим $F_1 = \{0, 1, \ldots, p-1\}$. Тогда F_1 — подполе (достаточно проверить замкнутость).

Случай 1°. $F_1 = F$. Тогда $q = p^1$, $F \cong \mathbb{Z}_p$.

Случай 2°. $F_1 \neq F$. Тогда возьмем элемент $\alpha \in F \setminus F_1$. Положим

$$F_2 = \{x_1\alpha + x_2 \mid x_1, x_2 \in F_1\}.$$

В множестве F_2 (уже не являющимся, вообще говоря, подполем) все элементы разные. Пусть это не так, т. е. $(x_1, x_2) \neq (y_1, y_2)$, но $(x_1 - y_1)\alpha + (x_2 - y_2) = 0$. Тогда если $x_1 = y_1$, то $x_2 = y_2$ — противоречие; если $x_1 \neq y_1$, то $\alpha = \frac{y_2 - x_2}{x_1 - y_1} \in F_1$ — противоречие.

Случай 2.1°. $F_2 = F$. Тогда $q = p^2$.

Случай 2.2°. $F_2 \neq F$. Тогда возьмем элемент $\beta \in F \setminus F_2$. Положим

$$F_3 = \{x_1\beta + x_2 \mid x_1 \in F_1, x_2 \in F_2\}.$$

В множестве F_3 все элементы разные (как следствие, $|F_3| = p^3$). Пусть это не так, т. е. $(x_1, x_2) \neq (y_1, y_2)$, но $(x_1 - y_1)\beta + (x_2 - y_2) = 0$. Тогда если $x_1 = y_1$, то $x_2 = y_2$ — противоречие; если $x_1 \neq y_1$, то $\beta = \frac{y_2 - x_2}{x_1 - y_1} \in F_2$ (так как $y_2 - x_2 \in F_2$, $\frac{1}{x_1 - y_1} \in F_1$) — противоречие.

Продолжая рассуждения аналогичным образом, за конечное число шагов установим, что существует n, такое что $q = p^n$. $\square$

3. В каждом конечном поле есть примитивный элемент (элемент α поля называется *примитивным*, если для всякий ненулевой элемент поля есть некоторая степень элемента α). Или иными словами — мультипликативная группа конечного поля циклическая.

$\triangleright$ Сначала докажем следующий факт.

Пусть m — максимальный порядок элемента в конечной абелевой группе G. Тогда порядок любого элемента группы G делит m.

Разложим группу G в прямое произведение примарных циклических подгрупп. Для каждого простого p, являющегося делителем $|G|$, берем

примарную компоненту максимального порядка, соответствующего этому p. Пусть M — произведение (по всем простым делителям $|G|$) порядков таких примарных компонент. Тогда для любого элемента g группы G справедливо равенство $g^M = e$ и, следовательно, порядок элемента g делит M. С другой стороны, произведение порождающих элементов выбранных примарных компонент имеет порядок M, поэтому $m = M$.

Рассмотрим в исходном конечном поле F уравнение $x^m = 1$, где m — максимальный порядок элементов мультипликативной группы поля. С одной стороны, каждый ненулевой элемент поля F является корнем этого уравнения, а с другой стороны, у многочлена степени m корней не более m. Следовательно, $m = |F| - 1$, т. е. мультипликативная группа конечного поля циклическая. □

4. Если d делит $|F| - 1$, то в поле F есть элемент порядка d (следует из цикличности).

5. Для любого элемента $\alpha \in F$ выполняется равенство $p\alpha = 0$ (так как $p = \underbrace{1 + \ldots + 1}_{p}$).

6. $(x + y)^p = x^p + y^p$, так как C_p^k делится на p при $k \neq 0, p$.

7. $(x_1 + \ldots + x_m)^p = x_1^p + \ldots + x_m^p$.

8. $(x_1 + \ldots + x_m)^{p^u} = x_1^{p^u} + \ldots + x_m^{p^u}$.

9. Для любого $n \geq 1$ существует поле из $p^n = q$ элементов.

▷ При $n = 1$ искомое поле $\mathbb{F}_p = \mathbb{Z}_p$ — поле вычетов по простому модулю p. Пусть $n \geq 1$ (случай $n = 1$ тоже вкладывается в общий случай). Для любого натурального n существует неприводимый нормированный многочлен степени n над $\mathbb{Z}_p$. Выберем какой-либо неприводимый нормированный многочлен $\pi(x)$ степени n. На множестве M многочленов из $\mathbb{Z}_p[x]$ степени не выше $n - 1$ введем операции сложения и умножения следующим образом.

Сложение: пусть $a(x), b(x) \in M$, тогда $a(x) + b(x) \in M$.

Умножение: пусть $a(x), b(x) \in M$, тогда $a(x) * b(x) = c(x)$, где $c(x)$ — остаток от деления $a(x)b(x)$ на $\pi(x)$, $c(x) \in M$.

Докажем обратимость всех ненулевых элементов (очевидно, что остальные аксиомы поля выполняются). Для этого сначала установим отсутствие делителей нуля.

Пусть $a(x) * b(x) = 0$. Тогда $a(x)b(x) = \pi(x)f(x)$ (здесь имеется ввиду обычное умножение многочленов над $\mathbb{Z}_p$), где степени многочленов $a(x)$, $b(x)$ и $f(x)$ не превосходят $n - 1$, а степень многочлена $\pi(x)$ равна n, но это противоречит неприводимости многочлена $\pi(x)$.

Установим теперь существование обратного элемента для любого ненулевого многочлена $a(x)$ из M. Очевидно, что $|M| = q$. Пусть

$b_1(x), \ldots, b_{q-1}(x)$ — все ненулевые многочлены из M. Тогда все многочлены $a(x) * b_1(x), \ldots, a(x) * b_{q-1}(x)$ — ненулевые и различные (в силу отсутствия делителей нуля). Значит, среди них есть и единичный многочлен. $\square$

Построенное поле будем обозначать $GF(p^n)$. Корректность такого обозначения доставляет следующее свойство.

10. Любые два конечных поля, состоящих из равного числа элементов, изоморфны (без доказательства).

3.3.5 Код Боуза — Чоудхури — Хоквингема

Будем рассматривать двоичный случай: $F = GF(2^m)$ (хотя построения можно вести над произвольным конечным полем). Полагаем, что поле $GF(2^m)$ построено по некоторому неприводимому многочлену $\pi(x)$ степени m в соответствии с тем, как это сделано в п. 9 свойств конечных полей.

Элементами поля $GF(2^m)$ считаем многочлены $a_0 + a_1 x + \ldots + a_{m-1}x^{m-1}$, где $a_i \in GF(2)$, $i = 0, 1, \ldots, m-1$. Тем самым элементу $\alpha \in GF(2^m)$ естественным образом можно поставить в соответствие столбец коэффициентов γ:

$$\alpha \longleftrightarrow \gamma = \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{m-1} \end{pmatrix}.$$

Чтобы подчеркнуть тот факт, что элемент поля $GF(2^m)$ представлен в виде столбца из m элементов поля $GF(2)$, будем при обозначении добавлять сверху символ «$\widehat{\ }$»: так, например, $\widehat{\alpha} = \gamma$.

Пусть $n = 2^m - 1$, $\alpha_1, \ldots, \alpha_n$ — все ненулевые элементы поля $GF(2^m)$, выписанные в произвольном порядке. Пусть также параметр t удовлетворяет условию $t < 2^m/m$. Введем матрицу A размера $t \times n$ над полем $GF(2^m)$ следующим образом:

$$A = \begin{pmatrix} \alpha_1 & \alpha_2 & \ldots & \alpha_n \\ \alpha_1^3 & \alpha_2^3 & \ldots & \alpha_n^3 \\ . & . & \ldots & . \\ \alpha_1^{2t-1} & \alpha_2^{2t-1} & \ldots & \alpha_n^{2t-1} \end{pmatrix}.$$

Матрица A задает матрицу H размера $tm \times n$ над полем $GF(2)$:

$$H = \begin{pmatrix} \widehat{\alpha_1} & \widehat{\alpha_2} & \dots & \widehat{\alpha_n} \\ \widehat{\alpha_1^3} & \widehat{\alpha_2^3} & \dots & \widehat{\alpha_n^3} \\ . & . & \dots & . \\ \widehat{\alpha_1^{2t-1}} & \widehat{\alpha_2^{2t-1}} & \dots & \widehat{\alpha_n^{2t-1}} \end{pmatrix}, \quad \text{где } \widehat{\alpha_i} = \begin{pmatrix} a_{0i} \\ a_{1i} \\ \vdots \\ a_{m-1,i} \end{pmatrix}, \ i = 1, \dots, n.$$

Отметим, что матрицы A и H построены над разными полями, но оба эти поля имеют характеристику 2.

Линейный код с проверочной матрицей H называется *кодом Боуза — Чоудхури — Хоквингема* (*кодом БЧХ*).

Теорема 28. *Пусть $tm < 2^m$. Тогда любые $2t$ столбцов в матрице H линейно независимы.*

Доказательство. Пусть это не так, т. е. при некотором k, $1 \le k \le 2t$, имеется k столбцов $h_{i_1}, \dots, h_{i_k}$, удовлетворяющих равенству

$$h_{i_1} + \dots + h_{i_k} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Тогда для полосы высоты m матрицы H с номером j, $j = 1, \dots, t$, выполняется соотношение

$$\widehat{\alpha_{i_1}^{2j-1}} + \dots + \widehat{\alpha_{i_k}^{2j-1}} = \widehat{0},$$

из которого в силу справедливости равенства $\widehat{\alpha} + \widehat{\beta} = \widehat{\alpha + \beta}$ следует, что

$$\alpha_{i_1}^{2j-1} + \dots + \alpha_{i_k}^{2j-1} = 0,$$

т. е. степенная сумма $S_{2j-1} = S_{2j-1}(\alpha_{i_1}, \dots, \alpha_{i_k})$ равна нулю (нулевому элементу поля $GF(2^m)$). Таким образом,

$$S_1 = S_3 = \dots = S_{2t-1} = 0.$$

Покажем, что и при четных значениях r, $2 \le r \le 2t$, степенная сумма S_r также равна нулю. Представим число r в виде: $r = 2^u v$, $u \ge 1$, $1 \le v \le t$, v — нечетное. Тогда справедливы равенства

$$S_r = S_{2^u v} = (\alpha_{i_1})^{2^u v} + \dots + (\alpha_{i_1})^{2^u v} = ((\alpha_{i_1})^v)^{2^u} + \dots + ((\alpha_{i_k})^v)^{2^u} = \\ = (\alpha_{i_1}^v + \dots + \alpha_{i_k}^v)^{2^u} = S_v^{2^u} = 0.$$

Итак, все степенные суммы $S_1, S_2, \ldots, S_{2t}$ от переменных $\alpha_{i_1}, \ldots, \alpha_{i_k}$ равны нулю. Тогда в матрице

$$\begin{pmatrix} \alpha_{i_1} & \alpha_{i_2} & \ldots & \alpha_{i_k} \\ \alpha_{i_1}^2 & \alpha_{i_2}^2 & \ldots & \alpha_{i_k}^2 \\ . & . & \ldots & . \\ \alpha_{i_1}^k & \alpha_{i_2}^k & \ldots & \alpha_{i_k}^k \end{pmatrix}$$

столбцы линейно зависимы, и, следовательно, $\det M = 0$. С другой стороны, в силу правила вычисления определителя Вандермонда над полем характеристики 2 получаем:

$$\det M = \alpha_{i_1} \ldots \alpha_{i_k} \prod_{1 \le s < t \le k} (\alpha_{i_s} + \alpha_{i_t}) \neq 0,$$

что противоречит предыдущему равенству. □

Следствие 13. *Пусть V — код БЧХ с проверочной матрицей (в широком смысле) H размера $mt \times n$. Тогда*

$$n(V) = 2^m - 1, \quad k(V) \ge n - mt = 2^m - mt - 1, \quad d(V) \ge 2t + 1.$$

Величина $2t + 1$ называется *конструктивным расстоянием* кода БЧХ V.

3.3.6 Алгоритм декодирования кода БЧХ при $t = 2$

В случае, когда $t = 2$, матрицы A и H имеют следующий вид:

$$A = \begin{pmatrix} \alpha_1 & \alpha_2 & \ldots & \alpha_n \\ \alpha_1^3 & \alpha_2^3 & \ldots & \alpha_n^3 \end{pmatrix}, \quad H = \begin{pmatrix} \widehat{\alpha_1} & \widehat{\alpha_2} & \ldots & \widehat{\alpha_n} \\ \widehat{\alpha_1^3} & \widehat{\alpha_2^3} & \ldots & \widehat{\alpha_n^3} \end{pmatrix}.$$

Пусть произошло две ошибки в разрядах с номерами i и j:

$$x = (x_1, \ldots, x_n) \longrightarrow y = (y_1, \ldots, y_n) = (x_1, \ldots, x_n) + (e_1, \ldots, e_n),$$

где $e_i = e_j = 1$, $e_r = 0$ при $r \neq i, j$.

Вычислим синдром S:

$$S = Hy^T = Hx^T + He^T = 0 + He^T = h_i + h_j.$$

В силу того, что

$$h_i = \begin{pmatrix} \widehat{\alpha_i} \\ \widehat{\alpha_i^3} \end{pmatrix}, \quad h_j = \begin{pmatrix} \widehat{\alpha_j} \\ \widehat{\alpha_j^3} \end{pmatrix}, \quad S = \begin{pmatrix} \widehat{S_1} \\ \widehat{S_3} \end{pmatrix},$$

выполняется система равенств

$$\begin{cases} \alpha_i + \alpha_j = S_1 \\ \alpha_i^3 + \alpha_j^3 = S_3. \end{cases}$$

Теперь отметим, что аналогичная система равенств справедлива и в случае, когда произошла одна ошибка или ошибок не было вовсе — в этих случаях одно или оба значения α_i и α_j равны нулю. В частности, если $S_1 = 0$, то при $S_3 = 0$ можно сделать вывод, что ошибок нет, а при $S_3 \neq 0$ — что все же произошло более двух ошибок.

Возвращаясь к решению системы, имеем:

$$\alpha_i^3 + \alpha_j^3 = (\alpha_i + \alpha_j)(\alpha_i^2 + \alpha_i\alpha_j + \alpha_j^2) = S_1(S_1^2 + \alpha_i\alpha_j).$$

Поэтому

$$\begin{cases} \alpha_i + \alpha_j = S_1 \\ \alpha_i\alpha_j = \dfrac{S_3}{S_1} + S_1^2, \end{cases}$$

т. е. известны сумма и произведение корней. Поэтому α_i и α_j — корни квадратного трехчлена

$$x^2 + S_1 x + \frac{S_3}{S_1} + S_1^2$$

или уравнения

$$S_1 x^2 + S_1^2 x + S_3 + S_1^3 = 0. \qquad (*)$$

Теперь опишем порядок действий.

1. Вычисляем S, находим $\widehat{S_1}$ и $\widehat{S_3}$, затем S_1 и S_3.

2. Если $S_1 = S_3 = 0$, то ошибок нет и ничего исправлять не нужно.

3. Если $S_1 \neq 0$, $S_3 = S_1^3$, то у уравнения $(*)$ ровно одно ненулевое решение $\alpha = S_1$. В этом случае произошла ошибка в том разряде, в котором в первых m строках матрицы H находится столбец $\widehat{S_1}$.

4. Если $S_1 \neq 0$, $S_3 \neq S_1^3$, то составляем уравнение $(*)$ и ищем его корни:

а) если есть два корня α_i и α_j, то определяем, в каких столбцах i' и j' матрицы H в первых m строках стоят наборы $\widehat{\alpha_i}$ и $\widehat{\alpha_j}$, а затем исправляем значения $y_{i'}$ и $y_{j'}$;

б) если нет двух корней, то произошло более двух ошибок.

5. Если $S_1 = 0$, $S_3 \neq 0$, произошло более двух ошибок.

3.3.7 Общая схема декодирования кода БЧХ

Пусть выполняется условие

$$t < \frac{2^m}{m}$$

и код БЧХ, исправляющий t ошибок, задается проверочной матрицей

$$H = \begin{pmatrix} \widehat{\alpha_1} & \widehat{\alpha_2} & \dots & \widehat{\alpha_n} \\ \widehat{\alpha_1^3} & \widehat{\alpha_2^3} & \dots & \widehat{\alpha_n^3} \\ . & . & \dots & . \\ \widehat{\alpha_1^{2t-1}} & \widehat{\alpha_2^{2t-1}} & \dots & \widehat{\alpha_n^{2t-1}} \end{pmatrix}$$

размера $tm \times n$ с элементами из поля $GF(2)$.

Пусть произошло τ, $\tau \le t$, ошибок в разрядах с номерами $i_1, \dots, i_\tau$:

$$x = (x_1, \dots, x_n) \longrightarrow y = (y_1, \dots, y_n) = (x_1, \dots, x_n) + (e_1, \dots, e_n),$$

где $e_{i_1} = \dots = e_{i_\tau} = 1$, а остальные координаты e_j равны нулю.

Вычислим синдром S:

$$S = Hy^T = Hx^T + He^T = 0 + He^T = h_{i_1} + \dots + h_{i_\tau}.$$

В силу того, что

$$h_{i_r} = \begin{pmatrix} \widehat{\alpha_{i_r}} \\ \widehat{\alpha_{i_r}^3} \\ \vdots \\ \widehat{\alpha_{i_r}^{2t-1}} \end{pmatrix}, \quad r = 1, \dots, \tau; \quad S = \begin{pmatrix} \widehat{S_1} \\ \widehat{S_3} \\ \vdots \\ \widehat{S_{2t-1}} \end{pmatrix},$$

выполняется система равенств

$$\begin{cases} \alpha_{i_1} + \dots + \alpha_{i_\tau} = S_1 \\ \alpha_{i_1}^3 + \dots + \alpha_{i_\tau}^3 = S_3 \\ \dots \\ \alpha_{i_1}^{2t-1} + \dots + \alpha_{i_\tau}^{2t-1} = S_{2t-1}. \end{cases}$$

Решив эту систему (над полем $GF(2^m)$), найдем $\alpha_{i_1}, \dots, \alpha_{i_\tau}$. Далее, определим номера столбцов $i'_1, \dots, i'_\tau$ матрицы H, в которых первые m строк совпадают со столбцами $\widehat{\alpha_{i_1}}, \dots, \widehat{\alpha_{i_\tau}}$. Изменяя в векторе y на противоположные значения $y_{i'_1}, \dots, y_{i'_\tau}$, получаем исходный вектор x.

Тем самым задача решена, но есть две проблемы:

1. Решение системы. При решении методом перебора (что само по себе при работе в конечном поле нормально) при вполне «рабочих» параметрах $m = 15$, $t = 100$ будет порядка $C_{2^{15}}^{100}$ различных вариантов, что является слишком большой для перебора величиной.

2. Нам неизвестно значение τ.

Вторую проблему формально можно решать так: значение τ неизвестно, но известно значение t. Будем решать такую систему:

$$\begin{cases} \alpha_{i_1} + \ldots + \alpha_{i_t} = S_1 \\ \alpha_{i_1}^3 + \ldots + \alpha_{i_t}^3 = S_3 \\ \ldots \\ \alpha_{i_1}^{2t-1} + \ldots + \alpha_{i_t}^{2t-1} = S_{2t-1}. \end{cases}$$

Среди решений $(\alpha_{i_1}, \ldots, \alpha_{i_t})$ есть и решения вида $(\alpha_{i_1}, \ldots, \alpha_{i_\tau}, 0, \ldots, 0)$.

Общий алгоритм

1. Вычисляем S, находим $\widehat{S_1}, \ldots, \widehat{S_{2t-1}}$, затем $S_1, \ldots, S_{2t-1}$. Для вычисления степенной суммы S_r при четных значениях r, $2 \le r \le 2t$ представим число r в виде: $r = 2^u v$, где $u \ge 1$, $1 \le v \le t$, v — нечетное, и воспользуемся справедливостью равенств

$$S_r = S_{2^u v} = (\alpha_{i_1})^{2^u v} + \ldots + (\alpha_{i_t})^{2^u v} = ((\alpha_{i_1})^v)^{2^u} + \ldots + ((\alpha_{i_t})^v)^{2^u} = \\ = (\alpha_{i_1}^v + \ldots + \alpha_{i_t}^v)^{2^u} = S_v^{2^u}.$$

Тем самым вычислены все степенные суммы $S_j(\alpha_{i_1}, \ldots, \alpha_{i_t})$, $j = 1, \ldots, 2t$.

2. По степенным суммам $S_j(\alpha_{i_1}, \ldots, \alpha_{i_t})$, $j = 1, \ldots, 2t$, находим элементарные симметрические многочлены $\sigma_j(\alpha_{i_1}, \ldots, \alpha_{i_t})$, $j = 1, \ldots, 2t$.

3. Составляем многочлен $\Lambda(x)$:

$$\Lambda(x) = (x - \alpha_{i_1}) \ldots (x - \alpha_{i_t}) = x^t + \sigma_1 x^{t-1} + \ldots + \sigma_{t-1} x + \sigma_t,$$

который называется *локатором ошибок*. Ищем перебором корни уравнения $\Lambda(x) = 0$, при этом перебор уже порядка 2^m.

4. Находим номера столбцов матрицы A, у которых первые элементы совпадают с ненулевыми решениями $\alpha_{i_1}, \ldots, \alpha_{i_\tau}$.

5. Исправляем соответствующие разряды в векторе y путем инвертирования (путем прибавления единицы).

Осталась одна проблема — по степенным суммам $S_1, \ldots, S_{2t}$ найти элементарные симметрические многочлены $\sigma_1, \ldots, \sigma_t$.

Рассмотрим два способа ее решения, причем оба используют тождества Ньютона для степенных сумм и элементарных симметрических многочленов из кольца $F[\alpha_1, \ldots, \alpha_t]$, где F — произвольное поле (правда в первом случае используются только обобщенные формулы Ньютона, а во втором — и просто формулы Ньютона, и обобщенные).

3.3.8 Алгоритм Питерсона — Горенстейна — Цирлера

Итак, пусть

$$S(x) = \sum_{k=0}^{\infty} S_k x^k, \quad \text{где } S_0 = 0,\ S_k = \alpha_1^k + \ldots + \alpha_t^k,\ k \geq 1;$$

$$\sigma(x) = \sum_{k=0}^{t} \sigma_k x^k, \quad \text{где } \ \sigma_0 = 1,\ \sigma_k = (-1)^k \sum_{1 \leq i_1 < \ldots < i_k \leq t} \alpha_1 \ldots \alpha_k,\ 1 \leq k \leq t.$$

Тогда справедливо равенство $D(\sigma(x))x + S(x)\sigma(x) = 0$ и, следовательно,

$$\sum_{k=0}^{t} k\sigma_k x^k + \sum_{k=0}^{\infty} \left(\sum_{i=0}^{k} S_{k-i}\sigma_i \right) x^k.$$

Поэтому

$$\left.\begin{array}{l} \sigma_1 + S_1 = 0, \\ 2\sigma_2 + S_2 + S_1\sigma_1 = 0, \\ \ldots \quad \ldots \quad \ldots \\ k\sigma_k + S_k + S_{k-1}\sigma_1 + \ldots + S_1\sigma_{k-1} = 0, \\ \ldots \quad \ldots \quad \ldots \\ t\sigma_t + S_t + S_{t-1}\sigma_1 + \ldots + S_1\sigma_{t-1} = 0, \end{array}\right\} \text{ формулы Ньютона}$$

$$\left.\begin{array}{l} S_{t+1} + S_t\sigma_1 + \ldots + S_1\sigma_t = 0, \\ \ldots \quad \ldots \quad \ldots \\ S_{t+i} + S_{t+i-1}\sigma_1 + \ldots + S_i\sigma_t = 0, \\ \ldots \quad \ldots \quad \ldots \end{array}\right\} \begin{array}{l} \text{обобщенные} \\ \text{формулы} \\ \text{Ньютона} \end{array}$$

Теперь рассмотрим способ нахождения элементарных симметрических многочленов через степенные суммы, использующий первые t обобщенных формул Ньютона и лежащий в основе алгоритма Питерсона — Горенстейна — Цирлера.

Запишем первые t обобщенных формул Ньютона в матричном виде:

$$\begin{pmatrix} S_t & S_{t-1} & \ldots & S_1 \\ S_{t+1} & S_t & \ldots & S_2 \\ \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots \\ S_{2t-1} & S_{2t-2} & \ldots & S_t \end{pmatrix} \begin{pmatrix} \sigma_1 \\ \sigma_2 \\ \cdot \\ \cdot \\ \cdot \\ \sigma_t \end{pmatrix} = - \begin{pmatrix} S_{t+1} \\ S_{t+2} \\ \cdot \\ \cdot \\ \cdot \\ S_{2t} \end{pmatrix}.$$

Положим

$$\mathbf{S}_t = \begin{pmatrix} S_t & S_{t-1} & \dots & S_1 \\ S_{t+1} & S_t & \dots & S_2 \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ S_{2t-1} & S_{2t-2} & \dots & S_t \end{pmatrix}.$$

Очевидно, что $\mathbf{S}_t = A_t \times B_t$, где

$$A_t = \begin{pmatrix} \alpha_1 & \alpha_2 & \dots & \alpha_t \\ \alpha_1^2 & \alpha_2^2 & \dots & \alpha_t^2 \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \alpha_1^t & \alpha_2^t & \dots & \alpha_t^t \end{pmatrix}, \qquad B_t = \begin{pmatrix} \alpha_1^{t-1} & \alpha_1^{t-2} & \dots & 1 \\ \alpha_2^{t-1} & \alpha_2^{t-2} & \dots & 1 \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \alpha_t^{t-1} & \alpha_t^{t-2} & \dots & 1 \end{pmatrix}.$$

Определители матриц A_t и B_t легко выражаются через определитель Вандермонда:

$$\det A_t = (\alpha_1 \dots \alpha_t) \prod_{1 \le i < j \le t} (\alpha_i - \alpha_j),$$

$$\det B_t = (-1)^{C_t^2} \prod_{1 \le i < j \le t} (\alpha_i - \alpha_j).$$

Таким образом, если произошло ровно t ошибок, то все величины $\alpha_1, \dots, \alpha_t$ ненулевые и попарно различные. Тогда матрицы A_t и B_t невырождены, а следовательно, невырождена и матрица $\mathbf{S}_t$. Поэтому, элементарные симметрические многочлены $\sigma_1, \dots, \sigma_t$ находятся как решение системы линейных уравнений с невырожденной матрицей.

Если произошло менее t ошибок, то среди величин $\alpha_1, \dots, \alpha_t$ хотя бы одна равна нулю. Поэтому $\det A_t = 0$, и следовательно, матрица $\mathbf{S}_t$ вырождена.

Алгоритм Питерсона — Горенстейна — Цирлера

1. Вычисляем синдром S, находим $\widehat{S}_1, \dots, \widehat{S_{2t-1}}$, затем $S_1, \dots, S_{2t-1}$, после чего вычисляем $S_2, S_4, \dots, S_{2t}$.

2. Составляем матрицу $\mathbf{S}_t$, находим $\det \mathbf{S}_t$.

Если матрица $\mathbf{S}_t$ вырождена, то уменьшаем значение t на единицу и заново составляем матрицу $\mathbf{S}_t$ и находим ее определитель. Если для всех $k = t, t-1, \dots, 1$ выполняется равенство $\det \mathbf{S}_t = 0$, то ошибок не было.

3. Если на некотором шаге выполняется неравенство $\det \mathbf{S}_t \neq 0$, то делаем вывод, что произошло ровно t ошибок и решаем систему линейных уравнений относительно $\sigma_1, \dots, \sigma_t$ (это можно сделать в силу того,

что определитель основной матрицы системы отличен от нуля). Далее составляем многочлен локаторов ошибок и подбором ищем его корни. Должно получиться t корней (если не так, то на самом деле произошло более t ошибок). Смотрим, в каких столбцах матрицы H находятся эти элементы и исправляем соответствующие разряды.

$*$ Лекция № 13′

3.3.9 Алгоритм, основанный на теореме Питерсона

Рассмотрим более экономную (по количеству вычислений) модификацию предыдущего алгоритма, основанную на теореме Питерсона.

Большинство рассуждений, на которых основан алгоритм Питерсона — Горенстейна — Цирлера, с незначительными изменениями справедливы для произвольного конечного поля. Описываемый ниже вариант изложенного алгоритма декодирования кода БЧХ существенно использует специфику полей характеристики 2.

Оставив только первые $2t-1$ равенств в формулах Ньютона (простых и обобщенных) и вычеркнув четные, получаем следующую систему из t линейных уравнений относительно $\sigma_1, \ldots, \sigma_t$:

$$\begin{cases} \sigma_1 = S_1 \\ S_2\sigma_1 + S_1\sigma_2 + \sigma_3 = S_3 \\ S_4\sigma_1 + S_3\sigma_2 + S_2\sigma_3 + S_1\sigma_4 + \sigma_5 = S_5 \\ \qquad \ldots \\ S_{2t-2}\sigma_1 + S_{2t-3}\sigma_2 + \ldots + S_{t-1}\sigma_t = S_{2t-1}. \end{cases}$$

Основную матрицу этой системы обозначим через M_t:

$$M_t = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ S_2 & S_1 & 1 & 0 & 0 & 0 & \cdots & 0 \\ S_4 & S_3 & S_2 & S_1 & 1 & 0 & \cdots & 0 \\ . & . & . & . & . & . & \cdots & . \\ S_{2t-2} & S_{2t-3} & S_{2t-4} & S_{2t-5} & S_{2t-6} & S_{2t-7} & \cdots & S_{t-1} \end{pmatrix}.$$

Теорема 29 (Питерсона). *Матрица M_t размера $t \times t$ невырождена, если степенные симметрические суммы S_j зависят от t или $t-1$ различных элементов поля (характеристики 2), и вырождена, если суммы S_j зависят от меньшего чем $t-1$ числа различных элементов поля.*

Доказательство теоремы опирается на две леммы.

Лемма 10. *Если степенные симметрические суммы S_j являются суммами меньшего чем $t-1$ числа степеней различных элементов поля характеристики 2, то матрица M_t вырождена.*

Доказательство. В соответствии с тождествами Ньютона справедливо равенство

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ S_2 & S_1 & 1 & 0 & 0 & 0 & \cdots & 0 \\ S_4 & S_3 & S_2 & S_1 & 1 & 0 & \cdots & 0 \\ . & . & . & . & . & . & \cdots & . \\ S_{2t-2} & S_{2t-3} & S_{2t-4} & S_{2t-5} & S_{2t-6} & S_{2t-7} & \cdots & S_{t-1} \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ \sigma_1 \\ \vdots \\ \sigma_{t-2} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

при этом условие леммы (из которого, в частности, следуют равенства $\sigma_{t-1} = \sigma_t = 0$) используется только для умножения последних $\lfloor n/2 \rfloor$ строк матрицы M_t. Например при умножении последней строки выкладки будут такими:

$$\begin{aligned} &S_{2t-3} + S_{2t-4}\sigma_1 + S_{2t-5}\sigma_2 + \ldots + S_{t-1}\sigma_{t-2} = \\ &\quad = S_{2t-3} + S_{2t-4}\sigma_1 + S_{2t-5}\sigma_2 + \ldots + S_{t-1}\sigma_{t-2} + S_{t-2}\sigma_{t-1} + S_{t-3}\sigma_t = 0. \end{aligned}$$

Таким образом, столбцы матрицы M_t линейно зависимы. □

Лемма 11. *Если S_j — степенные суммы t переменных $x_1, \ldots, x_t$ (над полем характеристики 2), то справедливо равенство:*

$$\det M_t = \prod_{1 \le i < j \le t} (x_i + x_j).$$

Доказательство. Рассуждения будем проводить в три этапа.

1. В силу предыдущей леммы если $x_i = x_j$, то степенные суммы S_j зависят от $t-2$ переменных и поэтому $\det M_t = 0$. Следовательно выражение $x_i + x_j$ делит $\det M_t$ при любых различных i и j.

2. Оценим сверху степень многочлена $\det M_t$ от переменных $x_1, \ldots, x_t$. Домножим столбцы матрицы M_t соответственно на $1, S_1, S_2, \ldots, S_{t-1}$, а строки — соответственно на $S_{2t-2}, S_{2t-4}, \ldots, S_2, 1$:

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ S_2 & S_1 & 1 & 0 & 0 & 0 & \cdots & 0 \\ S_4 & S_3 & S_2 & S_1 & 1 & 0 & \cdots & 0 \\ . & . & . & . & . & . & \cdots & . \\ S_{2t-2} & S_{2t-3} & S_{2t-4} & S_{2t-5} & S_{2t-6} & S_{2t-7} & \cdots & S_{t-1} \end{pmatrix} \begin{matrix} S_{2t-2} \\ S_{2t-4} \\ S_{2t-6} \\ . \\ 1 \end{matrix}$$

$$1 \qquad S_1 \qquad S_2 \qquad S_3 \qquad S_4 \qquad S_5 \quad \ldots \quad S_{t-1}$$

В полученной матрице каждый элемент будет либо нулем, либо многочленом степени $2t-2$. Таким образом,

$$\deg\left(\det M_t(1 \cdot S_1 \ldots S_{t-1})(S_{2t-2}S_{2t-4}\ldots S_2 \cdot 1)\right) \leq 2t(t-1),$$

при этом

$$\deg\left(1 \cdot S_1 \ldots S_{t-1}\right) = 0 + 1 + \ldots + (t-1) = \frac{t(t-1)}{2};$$

$$\deg\left(S_{2t-2}S_{2t-4}\ldots S_2 \cdot 1\right) = (2t-2) + (2t-4) + \ldots + 2 + 0 = t(t-1).$$

Поэтому

$$\deg M_t \leq \frac{t(t-1)}{2}.$$

Так как

$$\deg \prod_{1\leq i<j\leq t} (x_i + x_j) \leq \frac{t(t-1)}{2},$$

то либо $\det M_t \equiv 0$, либо

$$\det M_t = \prod_{1\leq i<j\leq t} (x_i + x_j).$$

3. Значение $\det M_t$ не зависит от того, какое именно поле характеристики 2 рассматривается. Подберем такое поле, в котором легко показать, что многочлен $\det M_t$ не является тождественно нулевым. Отдельно разберем случаи, когда t нечетно и четно.

Пусть t нечетное. Тогда найдется такое m_0, что $m_0 < t$ и t делит $2^{m_0} - 1$. Действительно, среди остатков от деления чисел $2^0, 2^1, \ldots, 2^{t-1}$ найдутся два одинаковых, так как их t штук и все они ненулевые. Следовательно, найдутся такие i и j, что $j < i \leq t-1$ и $2^i - 2^j$ делится на t. Тогда $2^j(2^{i-j} - 1)$ делится на t, и в силу нечетности t $2^{i-j} - 1$ делится на t. Полагаем $m_0 = i - j$.

В поле $GF(2^{m_0})$ найдется элемент α порядка t. Тогда элементы

$$\alpha_1 = 1,\ \alpha_2 = \alpha,\ \alpha_3 = \alpha^2,\ \ldots,\ \alpha_t = \alpha^{t-1}$$

являются попарно различными корнями уравнения $x^t = 1$. Следовательно

$$x^t - 1 = (x - \alpha_1)\ldots(x - \alpha_t) = x^t + \sigma_1 x^{t-1} + \ldots + \sigma_{t-1}x + \sigma_t.$$

Поэтому

$$\sigma_1 = \sigma_2 = \ldots = \sigma_{t-1} = 0, \quad \sigma_t = 1.$$

Тогда в силу тождеств Ньютона имеем:

$$S_1 = S_2 = \ldots = S_{t-1} = 0, \quad S_t = 1, \quad S_{t+1} = S_{t+2} = \ldots = S_{2t-2} = 0.$$

Учитывая нечетность t, получаем:

$$M_t(\alpha_1, \ldots, \alpha_t) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \ldots & 0 & 0 \\ . & . & . & . & . & \ldots & . & . \\ 0 & 0 & 0 & 0 & 0 & \ldots & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & \ldots & 0 & 0 \\ . & . & . & . & . & \ldots & . & . \\ 0 & 0 & 0 & 0 & 0 & \ldots & 1 & 0 \end{pmatrix}.$$

В этой матрице в каждом столбце и в каждой строке ровно по одной единице и поэтому $\det M_t(\alpha_1, \ldots, \alpha_t) \neq 0$.

Пусть t четное. Тогда $t' = t - 1$ — нечетное. Для этого t' находим m_0. В поле $GF(2^{m_0})$ найдется элемент α порядка $t' = t - 1$. Тогда элементы

$$\alpha_1 = 1, \ \alpha_2 = \alpha, \ \alpha_3 = \alpha^2, \ \ldots, \ \alpha_{t-1} = \alpha^{t-2}, \ \alpha_t = 0$$

являются корнями уравнения $x^t - x = 0$. Следовательно

$$x^t - x = (x - \alpha_1) \ldots (x - \alpha_t) = x^t + \sigma_1 x^{t-1} + \ldots + \sigma_{t-1} x + \sigma_t.$$

Поэтому

$$\sigma_1 = \sigma_2 = \ldots = \sigma_{t-2} = 0, \quad \sigma_{t-1} = 1, \quad \sigma_t = 0.$$

Тогда в силу тождеств Ньютона имеем:

$$S_1 = S_2 = \ldots = S_{t-2} = 0, \ S_{t-1} = 1, \ S_t = S_{t+1} = \ldots = S_{2t-1} = 0, \ S_{2t-2} = 1.$$

Следовательно,

$$M_t(\alpha_1, \ldots, \alpha_t) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \ldots & 0 & 0 \\ . & . & . & . & . & \ldots & . & . \\ 0 & 0 & 0 & 0 & 0 & \ldots & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & \ldots & 0 & 0 \\ . & . & . & . & . & \ldots & . & . \\ 1 & 0 & 0 & 0 & 0 & \ldots & 0 & 1 \end{pmatrix}.$$

После разложения определителя этой матрицы по последнему столбцу получаем матрицу, в каждом столбце и в каждой строке которой ровно по одной единице, и поэтому $\det M_t(\alpha_1, \ldots, \alpha_t) \neq 0$.

Тем самым установлено, что

$$\det M_t = \prod_{1 \leq i < j \leq t} (x_i + x_j). \qquad \square$$

Теперь перейдем непосредственно к доказательству теоремы Питерсона.

Доказательство. Если $\det M_t \neq 0$, то по лемме 10 степенные суммы S_j зависят не менее, чем от $t-1$ различных элементов поля.

Если $\det M_t = 0$, то по лемме 11 найдутся такие i и j, $i \neq j$, что $x_i + x_j = 0$. Тогда $x_i = x_j$, $x_i^k + x_j^k = 0$ и следовательно степенные суммы S_j зависят не более, чем от $t-2$ различных элементов поля. $\square$

Алгоритм, основанный на теореме Питерсона

1. Вычисляем синдром S, находим $\widehat{S}_1, \ldots, \widehat{S_{2t-1}}$, затем $S_1, \ldots, S_{2t-1}$, после чего вычисляем $S_2, S_4, \ldots, S_{2t-2}$.

2. Составляем матрицу M_t, находим $\det M_t$.

Если $\det M_t \neq 0$, то делаем вывод, что произошло либо $t-1$, либо t ошибок, и решаем систему линейных уравнений относительно $\sigma_1, \ldots, \sigma_t$ (это можно сделать в силу того, что определитель основной матрицы системы отличен от нуля). Далее составляем многочлен локаторов ошибок и подбором ищем его корни. Должно получиться t или $t-1$ корней (если не так, то на самом деле произошло более t ошибок). Смотрим, в каких столбцах матрицы H находятся эти элементы и исправляем соответствующие разряды.

Если $\det M_t = 0$, то делаем вывод, что произошло не более $t-2$ ошибок. В качестве нового значения t берем $t-2$ и начинаем снова.

Тем самым за конечное число итераций либо решим задачу, либо придем к случаям $t = 1$ или $t = 2$, а они уже разобраны.

Глава 4

Сложность схемных вычислений

Лекция № 14

Вычисление — это последовательность элементарных действий (из заданного набора) над исходными (входными) данными и величинами, полученными в процессе вычисления. При этом последовательность действий может быть фиксированна, а может зависеть от результатов промежуточных вычислений. В первом случае вычисление называется неветвящимся, а во втором — ветвящимся. Таким образом, неветвящиеся вычисления — это фиксированные последовательности элементарных вычислительных команд. Важнейшим модельным классом неветвящихся вычислений являются схемы из функциональных элементов[1] (СФЭ).

В случае вычисления булевых функций СФЭ являются довольно точной математической моделью электронных логических схем без обратной связи.

Чтобы избежать излишней громоздкости и неоправданного формализма, дадим одно из нескольких эквивалентных определений для случая вычисления булевых функций в конкретном базисе[2] $B_0 = \{x \vee y,\ x\&y,\ \overline{x}\}$. В общем случае определение дается аналогично.

Итак, пусть есть:

1) множество «исходных данных» X (как правило это переменные и,

[1] В зарубежной (да и не только зарубежной) литературе используются также другие названия — «прямолинейный алгоритм», «комбинационная машина», «логическая схема», «булева схема», «схема» и др.

[2] Строго говоря, фраза «базис $\{x \vee y,\ x\&y,\ \overline{x}\}$» математически безграмотна — множество булевых функций $\{x \vee y,\ x\&y,\ \overline{x}\}$ не является минимальной по включению полной системой, т. е. базисом. Однако в данном контексте такое словосочетание стало устойчивым и это связано с тем, что слово «базис» здесь несет другой смысловой оттенок — это набор элементарных средств, «кирпичиков», из которых строится СФЭ.

быть может, константы; в нашем случае $X = \{x_1, \ldots, x_n\}$);

2) множество «базисных операций» B (в нашем случае $B_0 = \{x \vee y,\ x\&y,\ \overline{x}\}$).

Схемой из функциональных элементов в базисе B_0 называется ориентированный граф (кратные ребра допускаются) без ориентированных циклов, в котором входные степени вершин могут быть равны только 0, 1 или 2, при этом если входная степень вершины равна 0, то вершине приписывается символ переменной из множества X (такие вершины называются *входами*), если входная степень вершины равна 1, то вершине приписывается функциональный символ, соответствующий операции отрицания, а если входная степень вершины равна 2, то вершине приписывается функциональный символ, соответствующий либо двухместной конъюнкции, либо двухместной дизъюнкции. Вершины с ненулевой входной степенью (т. е. вершины, которым приписаны символы операций), будем называть *функциональными элементами*. Кроме того, одна или несколько вершин помечены дополнительно «звездочкой» — эти вершины называются *выходами* (с них считывается информация).

На рис. 4.1 слева приведен пример СФЭ в базисе B_0 с двумя входами и одним выходом. Справа приведена та же схема, но в этой схеме функциональные элементы изображены в виде треугольников, внутри которых изображен приписанный данной вершине-элементу функциональный символ. Такой подход позволяет избавиться от ориентации ребер в СФЭ — ребра выходят из вершин треугольников и входят в основания. Кроме того, функциональные элементы на схеме справа специальным образом пронумерованы.

Чтобы корректно определить функционирование СФЭ, сформулируем две леммы, касающиеся такой нумерации.

Лемма 12. *В любом конечном ориентированном графе без ориентированных циклов найдется вершина, из которой не выходит ни одно ребро.*

Нумерация функциональных элементов СФЭ, содержащей n невходовых вершин, числами $1, 2, \ldots, n$ называется *правильной* (или *монотонной*), если каждое ребро СФЭ либо исходит из входа, либо направлено от функционального элемента, имеющего меньший номер, к функциональному элементу, имеющему бóльший номер.

Лемма 13. *У любой СФЭ существует правильная нумерация невходовых вершин (функциональных элементов).*

Правильная нумерация вершин СФЭ, вообще говоря, не единственна — в схеме справа на рис. 4.1 свойство правильности нумерации со-

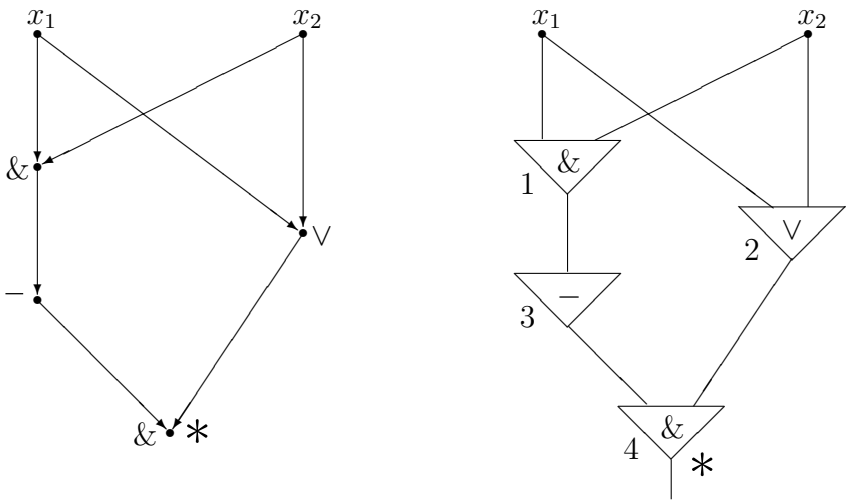

Рис. 4.1:

хранится, если поменять номера у первого и второго функциональных элементов.

Теперь зафиксируем какую-либо правильную нумерацию вершин схемы. Далее в порядке увеличения номера естественным образом приписываем вершине вычисляемую функцию. Тем самым каждой вершине будет приписана своя функция. Будем говорить, что СФЭ *реализует* (*вычисляет*) булеву функцию $f(x_1, \ldots, x_n)$ (систему функций $f_1(x_1, \ldots, x_n), \ldots, f_m(x_1, \ldots, x_n)$), если выходу (выходам) приписана эта функция (эта система функций). Удобно считать, что выходы СФЭ пронумерованы (упорядочены), и тем самым СФЭ вычисляет *булеву* (n, m)-*функцию* — вектор (набор) из m булевых функций от n переменных.

Функцию, вычисляемую i-м функциональным элементом схемы, представленной на рис. 4.1, обозначим через h_i. Тогда

$$
\begin{aligned}
&h_1 = x_1 x_2;\\
&h_2 = x_1 \vee x_2;\\
&h_3 = \overline{h}_1 = \overline{x}_1 \vee \overline{x}_2;\\
&h_4 = h_3 h_2 = (\overline{x}_1 \vee \overline{x}_2)(x_1 \vee x_2) = x_1 \oplus x_2.
\end{aligned}
$$

Таким образом, результатом работы этой схемы будет линейная функция $x_1 \oplus x_2$.

В качестве еще одного примера СФЭ отметим, что при всем формальном различии в определениях СФЭ и формулы, любую формулу можно интерпретировать как СФЭ, в которой выходы функциональных эле-

ментов не ветвятся, т. е. полустепень исхода любой вершины, отличной от входов и выхода, равна единице.

Теперь перейдем к понятию схемы вычислений. Пусть, по-прежнему, есть множество «исходных данных» X и множество «базисных операций» B. *Схема вычислений (над X) в базисе B* — это последовательность равенств

$$\begin{gathered} z_1 = \varphi_1(y_{11}, \ldots, y_{1r_1}); \\ \ldots \quad \ldots \quad \ldots \\ z_i = \varphi_i(y_{i1}, \ldots, y_{ir_i}); \\ \ldots \quad \ldots \quad \ldots \\ z_l = \varphi_l(y_{l1}, \ldots, y_{lr_l}), \end{gathered}$$

где каждая переменная y_{ij} $(i = 1, \ldots, l;\ j = 1, \ldots, r_i)$ — это либо одна из входных независимых переменных из множества X, либо одна из внутренних переменных $z_1, \ldots, z_{i-1}$, вычисленных на предыдущих шагах; $\varphi_1, \ldots, \varphi_l \in B$. Кроме того, одна или несколько внутренних переменных из множества $z_1, \ldots, z_l$ дополнительно помечены «звездочкой» — эти переменные называются *выходными* (с них считывается информация).

Отметим, что иногда под схемой вычислений понимают просто последовательность

$$z_1, \ldots, z_l,$$

удовлетворяющую такому условию: для каждого i, $1 \le i \le l$, найдется такая функция $\varphi_i \in B$, что $z_i = \varphi_i(y_{i1}, \ldots, y_{ir_i})$, где $y_{ij} \in X \cup \{z_1, \ldots, z_{i-1}\}$, $j = 1, \ldots, r_i$.

Понятия СФЭ и схемы вычислений эквивалентны — если в схеме вычислений исходить из того, что любая функция φ_i из B реализуется функциональным элементом, то естественным образом приходим к понятию СФЭ. В дальнейшем, как правило, будем говорить просто «схема».

Сложностью схемы S называется число функциональных элементов (или, что то же самое, число равенств в последовательности) схемы S. Сложность схемы S будем обозначать через $L(S)$.

Если функция f (произвольной природы) реализуется какой-либо схемой S в базисе B, то, очевидно, найдутся еще схемы, реализующие функцию f в базисе B. Определим величину $L_B(f)$ — *сложность реализации функции f схемами в базисе B* — равенством

$$L_B(f) = \min L(S),$$

где минимум берется по всем схемам S, реализующим функцию f в базисе B.

Аналогично определяется *сложность* $L_B(\{f_1, \ldots, f_m\})$ *реализации системы функций* $\{f_1, \ldots, f_m\}$ *схемами в базисе* B:

$$L_B(\{f_1, \ldots, f_m\}) = \min L(S),$$

где минимум берется по всем схемам S, реализующим систему функций $\{f_1, \ldots, f_m\}$ в базисе B.

Минимальной схемой называется такая схема, что не существует другой схемы в том же базисе, реализующей ту же самую функцию или систему функций, и имеющую меньшую сложность. Таким образом, если схема S, реализующая функцию f в базисе B, минимальная, то выполняется равенство $L(S) = L_B(f)$.

К сожалению, задача нахождения точного значения сложности функции или системы функций (и построения минимальной схемы), как правило, очень трудна и может быть решена только в отдельных простых случаях. Это связано с проблемами в доказательстве оценок сложности снизу. Действительно, для получения верхней оценки сложности функции достаточно предъявить конкретную схему, вычисляющую эту функцию, а вот для получения нижней оценки нужно доказать, что никакая схема меньшей сложности не реализует эту функцию, а это, вообще говоря, переборная задача огромного размера. Поэтому часто исследуют следующую упрощенную задачу.

Пусть F — класс (множество) функций, а $F(n)$ $(n = 1, 2, \ldots)$ — некоторое подмножество этого класса, например, множество всех функций от n фиксированных переменных из класса F. *Функцией Шеннона для класса* F, или более точно — функцией Шеннона сложности реализации функций из класса F схемами в базисе B, будем называть функцию $L_B^F(n)$, определяемую равенством

$$L_B^F(n) = \max L_B(f),$$

где максимум берется по всем функциям f из множества $F(n)$.

Нас будет интересовать асимптотическое поведение функции Шеннона в первую очередь для классической задачи, когда $F = P_2$. В этом случае вместо обозначения $L_B^{P_2}(n)$ будем использовать более простое $L_B(n)$. Кроме того, иногда будем у функционалов сложности опускать информацию о базисе (нижний индекс), если эта информация однозначно определяется из контекста.

Однако, сначала остановимся на двух задачах, не относящихся к сложности реализации булевых функций.

§ 4.1 Вычисление степеней

Итак, пусть множество «исходных данных» X состоит из одной переменной x, а множество «базисных операций» B состоит только из ассоциативной[3] операции умножения, применимой к степеням переменной x. Функциями, реализуемыми схемами в рамках такой модели, являются степени переменной x.

В соответствии с введенными определениями сложность $L(x^n)$ вычисления степени x^n численно равна минимально возможному числу операций умножения, достаточному для вычисления x^n.

Тривиальный способ последовательного умножения на x дает грубую оценку $L(x^n) \leq n - 1$.

Существенно более экономный способ вычисления x^n, основан на представлении числа n в двоичной записи и заключается[4] в $\lfloor \log n \rfloor$ последовательных возведениях в квадрат с последующим перемножением тех степеней, которые соответствуют единицам в двоичном разложении числа n. Пусть s — количество единиц в двоичном разложении числа n, тогда предложенная схема дает такую оценку сложности:

$$L(x^n) \leq \lfloor \log n \rfloor + s \leq 2\lfloor \log n \rfloor.$$

Однако, вообще говоря, и эта оценка может быть улучшена.

Пример 8. Схема

$$\begin{aligned} & z_1 = x \times x = x^2; \\ & z_2 = z_1 \times z_1 = x^4; \\ & z_3 = z_2 \times z_2 = x^8; \\ & z_4 = z_3 \times z_2 = x^{12}; \\ & z_5 = z_4 \times z_1 = x^{14}; \\ * \quad & z_6 = z_5 \times x = x^{15}, \end{aligned}$$

вычисляющая x^{15} в соответствии с двоичным разложением, не является минимальной. Действительно, сложность этой схемы равна 6, в то время

[3] Требование ассоциативности операции умножения нужно для корректного определения степени: действительно, если двухместную операцию "$*$" определить равенством $x * y = \overline{x}$, то, с одной стороны, $(x * x) * x = \overline{x} * x = x$, а с другой стороны, $x * (x * x) = x * \overline{x} = \overline{x}$.

[4] Здесь и далее, если не оговорено противное, под записью $\log x$ понимается $\log_2 x$.

как сложность следующей схемы

$$\begin{aligned} & z_1 = x \times x = x^2; \\ & z_2 = z_1 \times z_1 = x^4; \\ & z_3 = z_2 \times x = x^5; \\ & z_4 = z_3 \times z_3 = x^{10}; \\ * \quad & z_5 = z_4 \times z_3 = x^{15}, \end{aligned}$$

вычисляющей x^{15}, равна 5.

Теперь установим нижнюю оценку величины $L(x^n)$, которая основана на следующем простом утверждении.

Лемма 14. *Пусть схема S вычисляет x^n. Тогда выполняется неравенство*

$$n \leq 2^{L(S)}.$$

Доказательство. Проведем индукцию по величине $L(S)$.

При $L(S) = 0$ схема вычисляет саму переменную x и неравенство $1 \leq 2^0$ выполняется.

Теперь в схеме S рассмотрим последний элемент. Этот элемент вычисляет x^n, перемножая какие-то степени x^a и x^b, вычисляемые в свою очередь какими-то схемами сложности не более $L(S) - 1$. Дважды применяя предположение индукции, получаем

$$n = a + b \leq 2^{L(S)-1} + 2^{L(S)-1} = 2^{L(S)}.$$

□

Оформим в виде теоремы следующую нижнюю оценку, непосредственно вытекающую из леммы.

Теорема 30. *Для любого натурального n справедливо неравенство*

$$L(x^n) \geq \log n.$$

Таким образом, установлены верхняя и нижняя оценки величины $L(x^n)$, отличающиеся асимптотически вдвое.

Теперь установим асимптотически неулучшаемую верхнюю оценку.

Теорема 31. *При $n \to \infty$ справедливо соотношение*

$$L(x^n) \leq \log n + \frac{\log n}{\log \log n} \left(1 + O\left(\frac{\log \log \log n}{\log \log n}\right)\right).$$

Доказательство. Пусть d — натуральный параметр, значение которого выберем позже. Представим n в системе счисления по основанию 2^d:

$$n = a_0 \left(2^d\right)^0 + a_1 \left(2^d\right)^1 + \ldots + a_s \left(2^d\right)^s,$$

где $0 \le a_i \le 2^d - 1$, $i = 0, 1, \ldots s$; $a_s \ne 0$. Отметим справедливость неравенств

$$2^{sd} \le n < 2^{(s+1)d}.$$

На первом этапе, использовав sd операций умножения, путем последовательного возведения в квадрат вычисляем степени

$$x^2,\ x^4,\ \ldots, x^{2^d},\ \ldots,\ x^{2^{2d}},\ \ldots,\ x^{2^{sd}}.$$

Положим

$$u_0 = x^{2^{0d}} = x,\ u_1 = x^{2^d},\ \ldots,\ u_s = x^{2^{sd}}.$$

Отметим, что все степени u_i, $i = 0, 1, \ldots, s$, вычислены на первом этапе.

Для $k = 1, \ldots, 2^d - 1$ положим

$$I_k = \{i \mid a_i = k\}, \quad J_k = \{j \mid a_j \ge k\}.$$

Справедливы такие представления вычисляемой степени:

$$x^n = u_0^{a_0} u_1^{a_1} \ldots u_s^{a_s} =$$

$$= \left(\prod_{i \in I_{2^d-1}} u_i\right)^{2^d-1} \left(\prod_{i \in I_{2^d-2}} u_i\right)^{2^d-2} \ldots \left(\prod_{i \in I_1} u_i\right)^1 =$$

$$= \left(\prod_{i \in J_{2^d-1}} u_i\right) \left(\prod_{i \in J_{2^d-2}} u_i\right) \ldots \left(\prod_{i \in J_1} u_i\right).$$

По уже вычисленным степеням $u_0, u_1, \ldots, u_s$ ввиду вложений

$$J_{2^d-1} \subseteq J_{2^d-2} \subseteq \ldots \subseteq J_1$$

произведения

$$\prod_{i \in J_{2^d-1}} u_i,\quad \prod_{i \in J_{2^d-2}} u_i,\ \ldots,\ \prod_{i \in J_1} u_i$$

можно последовательно вычислить, используя не более s операций умножения (по одной операции для «присоединения» каждой новой переменной u_i). Для перемножения этих произведений требуется не более $2^d - 2$ операций умножения.

Таким образом, окончательно имеем:

$$L(x^n) \le sd + s + 2^d - 2 \le \log n + \frac{\log n}{d} + 2^d.$$

Теперь, полагая

$$d = \lfloor \log\log n - 2\log\log\log n \rfloor,$$

из предыдущего неравенства при $n \to \infty$ получаем

$$L(x^n) \le \log n + \frac{\log n}{\log\log n \left(1 - \frac{2\log\log\log n + 1}{\log\log n}\right)} + \frac{\log n}{(\log\log n)^2} =$$
$$= \log n + \frac{\log n}{\log\log n}\left(1 + \frac{2\log\log\log n}{\log\log n} + \frac{2}{\log\log n} + o\left(\frac{1}{\log\log n}\right)\right).$$

Требуемая верхняя оценка установлена. □

Следствие 14. *При $n \to \infty$ справедливо соотношение*[5]

$$L(x^n) \sim \log n.$$

Упражнение 20. Доказать, что

$$L(\{x^{n_1}, x^{n_2}\}) \sim \log\max(n_1, n_2)$$

при $n_1 + n_2 \to \infty$.

Упражнение 21. Пусть выполняются условия $n_i \le \frac{m^2}{\log m}$, $i = 1, \ldots, m$. Доказать, что при $m \to \infty$ справедливо соотношение

$$L(\{x^{n_1}, \ldots, x^{n_m}\}) \sim m.$$

Упражнение 22. Доказать, что

$$\max_{k:\, k \le n} L(x^k) - \log n \to \infty$$

при $n \to \infty$.

[5] Для положительных при всех достаточно больших значениях n последовательностей $\{a_n\}$ и $\{b_n\}$ запись $a_n \sim b_n$ при $n \to \infty$ означает, что $\lim\limits_{n\to\infty} \frac{a_n}{b_n} = 1$.

§ 4.2 Реализация двоичных наборов схемами конкатенации

Пусть $E = \{0, 1\}$. На множестве $E^+ = \bigcup\limits_{k=1}^{\infty} E^k$ всех двоичных наборов (слов) определим операцию *конкатенации* следующим образом. Для наборов $\tilde{\alpha} \in E^n$ и $\tilde{\beta} \in E^m$ результатом $\tilde{\alpha} \circ \tilde{\beta}$ применения операции конкатенации к этим наборам будет набор из множества E^{n+m}, получающийся в результате приписывания к набору $\tilde{\alpha}$ справа набора $\tilde{\beta}$:

$$(\alpha_1, \ldots, \alpha_n) \circ (\beta_1, \ldots, \beta_m) = (\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_m).$$

Будем рассматривать схемы, у которых множество «исходных данных» X состоит не из переменных, а из символов «0» и «1», а множество «базисных операций» B состоит из операции конкатенации, применимой к любой паре двоичных наборов. Функциями, реализуемыми схемами в рамках такой модели, являются наборы из множества E^+. Такие схемы будем называть схемами конкатенации.

Упражнение 23. Доказать, что схема конкатенации

$$\begin{aligned} & z_1 = 0 \circ 0 = 00; \\ & z_2 = 0 \circ 1 = 01; \\ & z_3 = z_1 \circ z_1 = 0000; \\ & z_4 = z_2 \circ z_2 = 0101; \\ & z_5 = z_1 \circ z_4 = 000101; \\ * \quad & z_6 = z_5 \circ z_3 = 0001010000 \end{aligned}$$

является минимальной.

В соответствии с изначальными определениями для любого набора $\tilde{\alpha} \in E^+$ сложность $L_c(\tilde{\alpha})$ численно равна минимальному числу операций конкатенации, достаточному для получения набора $\tilde{\alpha}$ исходя только из нуля и единицы (у функционала сложности нижний индекс «c» означает, что базисная операция — конкатенация; «c» — от англ. «*concatenation*»). Также стандартным образом вводится функция Шеннона:

$$L_c(n) = L_c^{E^+}(n) = \max_{\tilde{\alpha}\colon\, \tilde{\alpha} \in E^n} L_c(\tilde{\alpha}).$$

Тривиальный способ последовательного приписывания очередного символа дает такую верхнюю оценку функции Шеннона: $L_c(n) \leq n - 1$.

Добавление к произвольной схеме конкатенации еще одной операции конкатенации может увеличить максимум длин вычисляемых наборов не более чем в два раза. Поэтому $L_c(n) \geq \log n$.

Перейдем к уточнению и верхней и нижней оценок функции Шеннона.

Теорема 32. *При $n \to \infty$ справедливо асимптотическое неравенство*[6]

$$L_c(n) \lesssim \frac{n}{\log n}.$$

Доказательство. Пусть t — натуральный параметр, значение которого выберем позже, а $\tilde{\alpha}$ — набор из множества E^n, удовлетворяющий условию $L_c(\tilde{\alpha}) = L_c(n)$.

Опишем некоторую схему конкатенации, реализующую набор $\tilde{\alpha}$. Эта схема сначала тривиальным образом вычисляет все наборы длины t. Для этого требуется $(t-1)2^t$ операций конкатенации. После этого набор $\tilde{\alpha}$ собирается из $\lfloor n/t \rfloor$ кусков длины t и, быть может, куска длины $t' < t$. Для этого требуется менее $n/t + t$ операций. Следовательно,

$$L_c(n) = L_c(\tilde{\alpha}) < (t-1)2^t + \frac{n}{t} + t \leq t2^t + \frac{n}{t}.$$

Полагая

$$t = \lfloor \log n - 3 \log\log n \rfloor,$$

получаем, что

$$L_c(n) \leq \frac{n}{\log n} + O\left(\frac{n \log\log n}{(\log n)^2}\right).$$

Требуемая оценка получена. □

Отметим, что при доказательстве теоремы сложность реализации системы всех наборов заданной длины оценивалась очень грубо, причем это не помешало получить оценку функции Шеннона, которая, как будет показано ниже, является асимптотически неулучшаемой. На самом деле можно построить такую схему конкатенации, реализующую множество всех наборов заданной длины, в которой на каждый набор будет тратиться асимптотически всего одна (!) операция конкатенации. Аккуратно сформулируем этот факт в виде задачи.

Упражнение 24. Доказать, что при $n \to \infty$ выполняется равенство

$$L_c(E^n) \sim 2^n.$$

[6] Для положительных при всех достаточно больших значениях n последовательностей $\{a_n\}$ и $\{b_n\}$ запись $a_n \lesssim b_n$ при $n \to \infty$ означает, что $\overline{\lim\limits_{n\to\infty}} \frac{a_n}{b_n} \leq 1$.

Для получения нижней оценки будем использовать комбинаторно-графовый аппарат.

Рассмотрим класс конечных ориентированных графов (допускающих петли). Такой граф назовем *эйлеровым*, если в этом графе существует ориентированный эйлеров цикл, т. е. ориентированный цикл, содержащий по одному разу все ребра (дуги) графа.

Лемма 15. *Для того, чтобы конечный ориентированный связный*[7] *граф был эйлеровым, необходимо и достаточно, чтобы для любой вершины графа полустепень захода совпадала с полустепенью исхода.*

Доказательство. Необходимость очевидна. Докажем достаточность индукцией по числу ребер в графе. База индукции проверяются непосредственно.

Теперь в конечном ориентированном связном графе G, в котором для любой вершины полустепень захода совпадает с полустепенью исхода, выберем произвольную вершину. Она неизолированная, поэтому ее полустепени захода и исхода ненулевые. Следовательно, существует выходящее из этой вершины ребро. Переходя по этому ребру, попадаем в вершину, из которой выходит хотя бы одно ребро. В силу конечности графа через некоторое число таких переходов попадем в вершину, в которой уже были. Обозначим получившийся ориентированный цикл через C. Выделим на этом цикле произвольную вершину v_1. В ориентированном графе G', получающемся из графа G путем удаления всех ребер, входящих в цикл C, также полустепени захода всех вершин совпадают с полустепенями исхода.

Построим эйлеров цикл C_E в графе G следующим образом. В компоненте связности графа G', в которой лежит вершина v_1, по предположению индукции есть ориентированный эйлеров цикл. Все ребра этого цикла, начиная с некоторого ребра, исходящего из вершины v_1, последовательно включим в строящийся цикл C_E. Далее идем по ребрам цикла C, добавляя их в цикл C_E, пока не попадем в вершину, лежащую в другой относительно вершины v_1 компоненте связности графа G'. Обозначим эту вершину v_2. В компоненте связности графа G', в которой лежит вершина v_2, по предположению индукции есть ориентированный эйлеров цикл. Все ребра этого цикла, начиная с некоторого ребра, исходящего из вершины v_2, также последовательно включим в строящийся цикл C_E. Далее идем по ребрам цикла C, добавляя их в цикл C_E, пока не попадем в вершину, лежащую компоненте связности графа G', отличной

[7] для данного утверждения не имеет значения, какая связность имеется ввиду — обычная или сильная.

от компонент, в которых лежат вершины v_1 и v_2. Так поступаем пока не вернемся в вершину v_1. Тем самым будет построен ориентированный цикл C_E, содержащий все ребра графа G ровно по одному разу. □

Упражнение 25. Сформулировать и доказать аналогичный критерий эйлеровости для неориентированных связных графов.

Построим на 2^{k-1} вершинах, каждой из которых приписан уникальный двоичный набор длины $k-1$, ориентированный граф, проведя для каждого двоичного набора $(\alpha_1, \ldots, \alpha_k)$ помеченное этим набором ориентированное ребро из вершины, которой приписан набор $(\alpha_1, \ldots, \alpha_{k-1})$, в вершину, которой приписан набор $(\alpha_2, \ldots, \alpha_k)$. Назовем этот граф *графом де Брёйна порядка* k. На рисунке 4.2 изображен граф де Брёйна порядка 3.

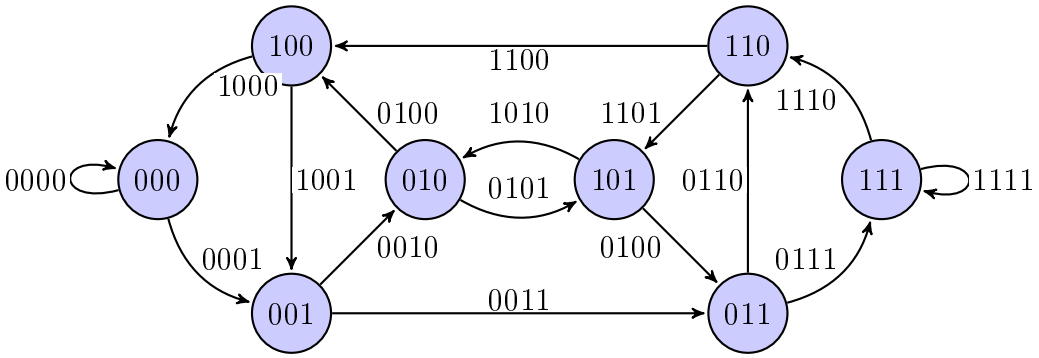

Рис. 4.2:

Свойства графа де Брёйна порядка k:

1. Число вершин равно 2^{k-1}, каждая из них помечена уникальным двоичным набором длины $k-1$.

2. Число ребер (включая две петли) равно 2^k, каждое из них помечено уникальным двоичным набором длины k.

3. В каждую вершину входит ровно два ребра.

4. Из каждой вершины исходит ровно два ребра.

5. Последние $k-1$ разрядов набора, помечающего входящее в вершину ребро, совпадают с первыми $k-1$ разрядами набора, помечающего исходящее из этой вершины ребро.

В силу леммы 15 граф де Брёйна является эйлеровым. Эйлеров цикл графа де Брёйна порядка k естественным образом задает двоичную последовательность длины $2^k + k - 1$: пометка первого ребра цикла определяет первые k символов, а каждое из остальных $2^k - 1$ ребер в силу

свойства 5 определяет ровно один символ. Определенный таким образом двоичный набор длины $2^k + k - 1$ называется *последовательностью де Брёйна порядка* k.

Для произвольного набора назовем *подсловом* любую последовательность идущих подряд разрядов этого набора.

Сформулируем вытекающее из построения основное свойство последовательности де Брёйна.

Лемма 16. *Все* 2^k *подслов длины* k *последовательности де Брёйна порядка* k *различны.*

Теперь можно переходить к доказательству асимптотически совпадающей с верхней оценкой из теоремы 32 нижней оценки функции Шеннона.

Лекция № 15

Теорема 33. *Для любого* $n \geq 3$ *выполняется неравенство*

$$L_c(n) \geq \frac{n}{\log n}.$$

Доказательство. При $n = 3, 4, 5$ утверждение теоремы проверяется непосредственно.

Пусть $n \geq 6$. Определим k как минимальное натуральное число r, удовлетворяющее условию

$$n \leq 2^r + r - 1.$$

Отметим, что при $n \geq 6$ верно неравенство $k \geq 3$. Кроме того, из минимальности k следует, что $2^{k-1} + k - 2 < n$, и следовательно, $k - 1 < \log n$. Обозначим через $\tilde{\beta}_n$ первые n элементов некоторой последовательности де Брёйна порядка k.

Рассмотрим как схему из функциональных элементов минимальную схему конкатенации S, реализующую набор $\tilde{\beta}_n$. Множество всех вершин схемы S разобьем на два подмножества в зависимости от длины вычисляемого в этой вершине набора: если длина вычисляемого в вершине набора не превосходит $k - 1$, то вершину относим к первому множеству, а если длина вычисляемого набора k или больше, то — ко второму.

Вершины второго множества образуют дерево, листья которого (возможно, склеенные) находятся в первом множестве. Пусть во втором множестве t вершин. Тогда

$$(t + 1)(k - 1) \geq n.$$

Отсюда, учитывая, что в первом множестве при $k \geq 3$ есть хотя бы одна невходовая вершина, получаем

$$L_c(n) \geq L_c\left(\tilde{\beta}_n\right) \geq t + 1 \geq \frac{n}{k-1}.$$

Наконец, используя неравенство $k - 1 < \log n$, устанавливаем справедливость доказываемой оценки. □

Упражнение 26. Пусть $\tilde{\alpha}_n^m$ — произвольный двоичный набор длины n с m единицами. Тогда если m — фиксировано, а n неограниченно растет, то

$$L_c\left(\tilde{\alpha}_n^m\right) \sim \log n.$$

§ 4.3 Точная по порядку верхняя оценка сложности булевых функций

Возвращаемся к задаче оптимальной реализации булевых функций схемами. В случае, когда набором элементарных операций является классический базис $B_0 = \{x \vee y,\ x\&y,\ \overline{x}\}$, индекс B_0 у функционалов сложности будем опускать.

Лемма 17. *Для любого натурального n выполняется неравенство*

$$L(n) \leq n2^n.$$

Доказательство. Константы 0 и 1 можно реализовать в соответствии с формулами $x_1\overline{x}_1$ и $x_1 \vee \overline{x}_1$ схемами сложности 2. Для любой функции $f(x_1, \ldots, x_n)$, отличной от констант, промоделируем схемой представление этой функции в виде совершенной дизъюнктивной нормальной формы:

$$f(x_1, \ldots, x_n) = \bigvee_{\tilde{\sigma}:\ f(\tilde{\sigma})=1} x_1^{\sigma_1} \ldots x_n^{\sigma_n}.$$

Для этого потребуется не более n отрицаний, чтобы реализовать отрицания переменных, по $n - 1$ операций конъюнкции на каждую из не более чем $2^n - 1$ элементарных конъюнкций из совершенной дизъюнктивной нормальной формы, а затем не более $2^n - 2$ операций дизъюнкции для реализации функции f. Таким образом,

$$L(n) \leq n + (n-1)(2^n - 1) + 2^n - 2 < n2^n.$$

□

Обозначим через $\mathcal{K}_n(x_1, \ldots, x_n)$ множество всех 2^n элементарных конъюнкций от переменных $x_1, \ldots, x_n$:

$$\mathcal{K}_n(x_1, \ldots, x_n) = \left\{ x_1^{\sigma_1} \ldots x_n^{\sigma_n} \mid (\sigma_1, \ldots, \sigma_n) \in E^n \right\}.$$

Лемма 18. *При $n \to \infty$ справедливо асимптотическое соотношение*

$$L(\mathcal{K}_n) \sim 2^n.$$

Доказательство. Построим схему, реализующую систему функций $\mathcal{K}_n(x_1, \ldots, x_n)$, как показано на рис. 4.3. Схема состоит из трех блоков (подсхем). Первая подсхема по переменным $x_1, \ldots, x_{\lfloor n/2 \rfloor}$ реализует систему конъюнкций $\mathcal{K}_{\lfloor n/2 \rfloor}(x_1, \ldots, x_{\lfloor n/2 \rfloor})$, вторая подсхема по переменным $x_{\lfloor n/2 \rfloor + 1}, \ldots, x_n$ реализует систему конъюнкций $\mathcal{K}_{\lceil n/2 \rceil}(x_{\lfloor n/2 \rfloor + 1}, \ldots, x_n)$, а третья — каждую элементарную конъюнкцию из множества $\mathcal{K}_n(x_1, \ldots, x_n)$ «собирает» (с помощью одной операции конъюнкции) из двух ее «половинок», реализованных на некоторых выходах первой и второй подсхемы соответственно.

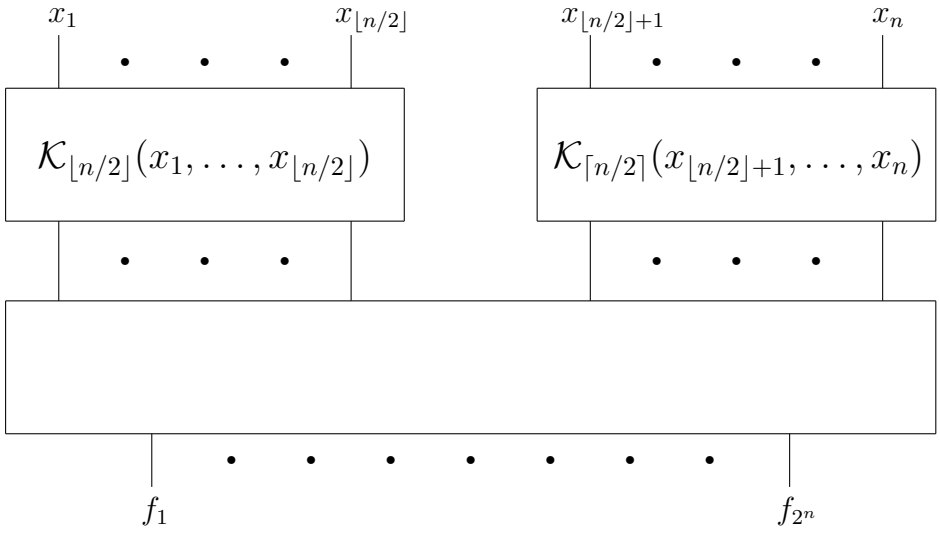

Рис. 4.3:

Используя тривиальные верхние оценки сложности первых двух подсхем, получаем:

$$\begin{aligned} L(\mathcal{K}_n) &\le L(\mathcal{K}_{\lfloor n/2 \rfloor}) + L(\mathcal{K}_{\lceil n/2 \rceil}) + 2^n \le \\ &\le \left\lfloor \frac{n}{2} \right\rfloor + \left(\left\lfloor \frac{n}{2} \right\rfloor - 1 \right) 2^{\lfloor n/2 \rfloor} + \left\lceil \frac{n}{2} \right\rceil + \left(\left\lceil \frac{n}{2} \right\rceil - 1 \right) 2^{\lceil n/2 \rceil} + 2^n \le \\ &\le 2^n + O\left(n\sqrt{2^n} \right). \end{aligned}$$

□

Теорема 34. *При $n \to \infty$ верна следующая верхняя оценка функции Шеннона*

$$L(n) \lesssim 6\frac{2^n}{n}.$$

Доказательство. Пусть $f(x_1, \ldots, x_n)$ — самая сложная функция от n переменных. Опишем метод построения схем, позволяющий получить указанную оценку сложности. Этот метод принято называть *методом Шеннона*.

Введем натуральный параметр $k = k(n)$, значение которого выберем позже. Пока потребуем только, чтобы выполнялись условия $k < n$ и $n - k \to \infty$ при $n \to \infty$.

Разложим функцию $f(x_1, \ldots, x_n)$ по первым $n - k$ переменным:

$$f(x_1, \ldots, x_n) = \bigvee_{(\sigma_1, \ldots, \sigma_{n-k})} x_1^{\sigma_1} \ldots x_{n-k}^{\sigma_{n-k}} f(\sigma_1, \ldots, \sigma_{n-k}, x_{n-k+1}, \ldots, x_n).$$

Построим схему, реализующую функцию f как показано на рис. 4.4. Схема состоит из трех блоков (подсхем). Первая подсхема по переменным $x_1, \ldots, x_{n-k}$ реализует систему конъюнкций $\mathcal{K}_{n-k}(x_1, \ldots, x_{n-k})$, вторая по переменным $x_{n-k+1}, \ldots, x_n$ реализует систему всех 2^{2^k} функций от этих k переменных, а третья в соответствии с указанным разложением функции f реализует саму функцию f.

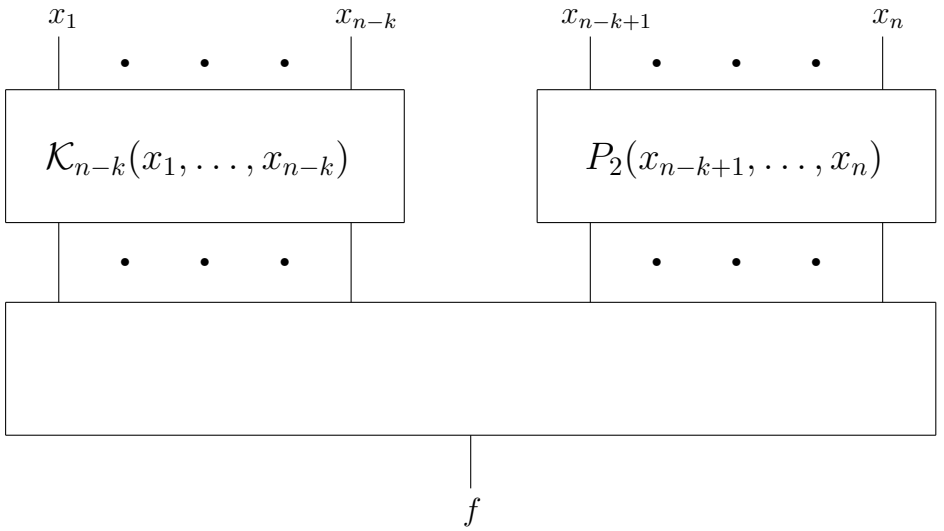

Рис. 4.4:

Применяя леммы 17 и 18, получаем:

$$L(n) = L(f) \leq 2^{n-k} + o\left(2^{n-k}\right) + k2^k 2^{2^k} + 2 \cdot 2^{n-k} = \frac{3 \cdot 2^n}{2^k} + o\left(\frac{2^n}{2^k}\right) + k2^k 2^{2^k}.$$

Теперь положим $k = \lfloor \log(n - 3\log n) \rfloor$. Тогда

$$\frac{n - 3\log n}{2} < 2^k \leq n - 3\log n.$$

Поэтому

$$L(n) \leq 6\frac{2^n}{n} + o\left(\frac{2^n}{n}\right).$$

□

Отметим, что оценка сложности реализации системы всех булевых функций от фиксированных переменных, использованная при доказательстве теоремы 34, очень груба. На самом деле нетрудно найти точное значение этой величины, причем при реализации в любом полном базисе.

Упражнение 27. Пусть $[B] = P_2$. Доказать, что для любого натурального n верно равенство

$$L_B(P_2(x_1, \ldots, x_n)) = 2^{2^n} - n.$$

Теперь сформулируем еще два утверждения о сложности реализации булевых функций в произвольном базисе.

Теорема 35. *Пусть B_1 и B_2 — конечные множества булевых функций, причем $[B_1] = [B_2] = P_2$. Тогда найдутся такие положительные константы c_1 и c_2, что для любой булевой функции f выполняются неравенства*

$$c_1 L_{B_1}(f) \leq L_{B_2}(f) \leq c_2 L_{B_1}(f).$$

Доказательство. Положим

$$c_2 = \max_{\varphi \in B_1} L_{B_2}(\varphi).$$

Теперь, если в какой-либо минимальной схеме S, реализующей произвольную функцию f в базисе B_1, заменить все функциональные элементы, соответствующие функциям из базиса B_1, минимальными схемами в базисе B_2, реализующие те же самые функции, получим схему S', реализующую функцию f в базисе B_2 и имеющую сложность не более $c_2 L(S)$. Следовательно,

$$L_{B_2}(f) \leq L(S') \leq c_2 L(S) = c_2 L_{B_1}(f),$$

и второе неравенство доказано.

Полагая

$$c_1 = \left(\max_{\varphi \in B_2} L_{B_1}(\varphi) \right)^{-1},$$

можно аналогично установить первое неравенство. □

Теоремы 34 и 35 устанавливают верхнюю оценку сложности реализации булевых функций в произвольном базисе. Сформулируем эту оценку как отдельную теорему.

Теорема 36. *Пусть $[B] = P_2$. Тогда при $n \to \infty$ выполняется соотношение*

$$L_B(n) \leq O\left(\frac{2^n}{n}\right).$$

Ниже будет показано, что эта оценка по порядку неулучшаема.

§ 4.4 Нижние оценки сложности булевых функций

Пусть B — конечное множество булевых функций. Обозначим через $r = r(B)$ максимальное число существенных переменных у функций из множества B.

Сначала установим простую линейную нижнюю оценку.

Теорема 37. *Пусть B — конечное множество булевых функций, удовлетворяющее условию $[B] = P_2$. Тогда для произвольной булевой функции f, существенно зависящей от n переменных, выполняется неравенство*

$$L_B(f) \geq \left\lceil \frac{n-1}{r(B)-1} \right\rceil.$$

Доказательство. В минимальной схеме, реализующей функцию f в базисе B, ровно $L_B(f)$ элементов. Оценим число ребер в этой схеме. С одной стороны, число ребер, исходящих из вершин схемы, не менее $n+L_B(f)-1$. С другой стороны, число ребер, входящих в вершины схемы, не более $r(B)L_B(f)$. Поэтому справедливо неравенство $n+L_B(f)-1 \leq r(B)L_B(f)$. Из этого неравенства и целочисленности величины $L_B(f)$ следует нужная оценка. □

Оценку из теоремы 37 в отдельных случаях можно усиливать, но для конструктивно заданных[8] последовательностей булевых функций известны лишь не более чем линейные по числу переменных нижние оценки сложности.

Однако, мы обещали доказать, что оценка из теоремы 36 по порядку неулучшаема, а для этого нужно уметь доказывать экспоненциальные нижние оценки. Эти оценки будут неконструктивными и будут получаться из мощностных соображений, идею которых проиллюстрируем на примере одной задачи из теории чисел.

Как доказать, что есть трансцендентные действительные числа (не являющиеся корнями ненулевых многочленов с рациональными коэффициентами)? Да, можно доказать, что $e, \pi, \sin 1, 2^{\sqrt{2}}, \log 3$ трансцендентны. Однако эти доказательства отнюдь не просты. В то же время элементарные мощностные рассуждения — множество действительных чисел континуально, а множество корней ненулевых многочленов с рациональными коэффициентами счетно — дают не только простейшее (правда, некоструктивное) доказательство этого факта, но и устанавливают, что почти все действительные числа трансцендентны.

Аналогичные рассуждения лежат и в основе высоких нижних оценок сложности функций Шеннона. Действительно, если число минимальных схем в базисе B с n входами и одним выходом сложности, не превосходящей k, менее 2^{2^n}, то и число булевых функций от n фиксированных переменных сложности, не превосходящей k, менее числа всех булевых функций от n переменных. Поэтому выполняется неравенство $L_B(n) > k$. В связи с этим при доказательстве нижних оценок сложности функций Шеннона на первый план выходит задача о получении приемлемых верхних оценок числа указанных схем, к которой мы и переходим.

Схему будем называть *приведенной*, если в ней нет двух разных элементов, реализующих одну и ту же функцию. Очевидно, что из любой схемы путем удаления некоторых элементов и «переподключения» выходящих из этих элементов ребер можно получить приведенную схему. Отметим также, что любая минимальная схема является приведенной.

Обозначим через $N_B^{=}(l, n, m)$ и $N_B^{\leqslant}(l, n, m)$ число приведенных схем в базисе B с n входами и m выходами сложности в точности l и, соответственно, не более l.

Сначала оценим сверху величину $N_B^{=}(l, n, m)$.

[8]Под конструктивным заданием последовательности булевых функций можно понимать, например, такое задание, при котором ответ на вопрос, равно ли значение функции на полученном наборе единице, может быть получен за полиномиальное время.

Пусть S — приведенная схема в базисе B сложности l с n входами и m выходами. Занумеруем в произвольном порядке числами $1, 2, \ldots, l$ элементы схемы S. Схеме S с выбранной нумерацией элементов сопоставим таблицу T высоты l и ширины $r(B)+1$ следующим образом. В i-й строке таблицы T в первом столбце указывается функция из базиса B, приписанная элементу E с номером i в выбранной нумерации, а в остальных столбцах этой строки — символы из множества $\{x_1, \ldots, x_n\} \cup \{1, \ldots, l\}$: в $(j+1)$-й столбец, $j = 1, \ldots, r$, этой строки помещается информация о том, из какой вершины ведет ребро, соответствующее j-му входу элемента E. Если ребро ведет из входа, помеченного переменной x_i, то в соответствующую клетку помещается символ x_i, а если ребро ведет из элемента с номером k, то — число k. Если у элемента E менее r входов, то все оставшиеся пустыми клетки i-й строки заполним для определенности так же, как и вторую клетку этой строки. Кроме того, если элемент E является выходом схемы с номером t, то i-я строка помечается дополнительно символом $*_t$ (если выходом является переменная, то помечается эта переменная). На рис. 4.5 представлена таблица для схемы, изображенной на рис. 4.1, при следующей нумерации элементов: первый элемент — дизъюнктор (элемент, которому приписана операция дизъюнкции), второй элемент — инвертор (элемент, которому приписана операция отрицания), третий элемент — конъюнктор (элемент, которому приписана операции конъюнкции), являющийся выходом схемы, четвертый элемент — конъюнктор, не являющийся выходом схемы.

$\vee$	x_1	x_2	
$-$	4	4	
&	2	1	$*_1$
&	x_1	x_2	

Рис. 4.5:

Лемма 19. *Приведенной схеме при разных нумерациях элементов соответствуют разные таблицы.*

Доказательство. Пусть это не так: двум разным нумерациям некоторой приведенной схемы S соответствуют одинаковые таблицы T_1 и T_2. Будем последовательно рассматривать элементы схемы S в порядке, соответствующем некоторой монотонной нумерации. В какой-то момент впервые элемент будет иметь разные номера в двух рассматриваемых нумерациях, и следовательно, этому элементу будут соответствовать разные

строки в таблицах T_1 и T_2. Но эти строки будут совпадать, так как до этого все элементы схемы имели одинаковые номера в рассматриваемых нумерациях. В силу совпадения таблиц T_1 и T_2 получаем, что в этой таблице есть две одинаковые строки, что противоречит неприводимости схемы S. □

Лемма 20. *Найдется $c_0 > 0$, такое что при всех значениях n выполняется неравенство*

$$N_B^{=}(l,n,m) \le \frac{c_0^l (l+n)^{r(B)l+m}}{l^l}.$$

Доказательство. В силу леммы 19 приведенной схеме сложности l соответствуют $l!$ различных таблиц. Кроме того, таблицы, соответствующие разным приведенным схемам, тоже разные. Оценивая сверху число таблиц, имеем:

$$N_B^{=}(l,n,m) \le \frac{|B|^l (l+n)^{r(B)l} (l+n)^m}{l!}.$$

Используя соотношение $l! \ge (l/3)^l$ и полагая $c_0(B) = 3|B|$, получаем нужное неравенство. □

В силу очевидного равенства

$$N_B^{\leqslant}(l,n,m) = \sum_{l'=0}^{\lfloor l \rfloor} N_B^{=}(l',n,m)$$

из леммы 20 непосредственно следует такое утверждение.

Лемма 21. *Найдется $c > 0$, такое что при всех значениях n выполняется неравенство*

$$N_B^{\leqslant}(l,n,m) \le \frac{c^l (l+n)^{r(B)l+m}}{l^l}.$$

Пусть F_{nm} — класс булевых $(n, m(n))$-функций; $\mathcal{F} = \cup F_{nm}$. Положим

$$L_B^{\mathcal{F}}(n,m) = \max_{\{f_1,\ldots,f_m\} \in F_{nm}} L_B(\{f_1,\ldots,f_m\}).$$

При $m = 1$ получаем стандартное определение функции Шеннона для класса $\mathcal{F}$.

Теперь в достаточно общей форме сформулируем мощностную нижнюю оценку.

Теорема 38 (мощностная нижняя оценка). *Пусть при $n \to \infty$ выполняется условие*

$$n + m(n) = o\left(\frac{\log|F_{nm}|}{\log\log|F_{nm}|}\right).$$

Тогда для любого полного конечного базиса B справедлива асимптотическая оценка

$$L_B^{\mathcal{F}}(n,m) \gtrsim \frac{1}{r(B)-1}\frac{\log|F_{nm}|}{\log\log|F_{nm}|},$$

причем для любого $\varepsilon > 0$ доля систем функций $\{f_1,\ldots,f_m\}$ из класса F_{nm}, удовлетворяющих неравенству

$$L_B(\{f_1,\ldots,f_m\}) \geq \frac{1-\varepsilon}{r(B)-1}\frac{\log|F_{nm}|}{\log\log|F_{nm}|},$$

стремится к 1 при $n \to \infty$.

Доказательство. Положим

$$l_\varepsilon = \frac{1-\varepsilon}{r(B)-1}\frac{\log|F_{nm}|}{\log\log|F_{nm}|}.$$

Обозначим через $\hat{N}_B(l,n,m)$ число (n,m)-функций, сложность реализации которых в базисе B не превосходит l. Для доказательства теоремы достаточно установить, что для любого $\varepsilon > 0$ отношение

$$\frac{\hat{N}_B(l_\varepsilon,n,m)}{|F_{nm}|}$$

стремится к 0 при $n \to \infty$. Используя лемму 21 и условие теоремы, оценим сверху логарифм этого отношения:

$$\begin{aligned}
\log\left(\frac{\hat{N}_B(l_\varepsilon,n,m)}{|F_{nm}|}\right) \leq \log\left(\frac{N_B^{\leqslant}(l_\varepsilon,n,m)}{|F_{nm}|}\right) \leq \\
\leq (r(B)l_\varepsilon + m)\log(2l_\varepsilon) + l_\varepsilon\log c - l_\varepsilon\log l_\varepsilon - \log|F_{nm}| = \\
= (r(B)-1)l_\varepsilon\log l_\varepsilon - \log|F_{nm}| + (r(B)l_\varepsilon + m + l_\varepsilon\log c) \leq \\
(r(B)-1)\frac{1-\varepsilon}{r(B)-1}\frac{\log|F_{nm}|}{\log\log|F_{nm}|}\log\log|F_{nm}| - \log|F_{nm}| + O\left(\frac{\log|F_{nm}|}{\log\log|F_{nm}|}\right) = \\
= -\varepsilon\log|F_{nm}| + O\left(\frac{\log|F_{nm}|}{\log\log|F_{nm}|}\right).
\end{aligned}$$

Последнее выражение стремится к $-\infty$ при $n \to \infty$. □

Следствие 15. *При $n \to \infty$ выполняется асимптотическое неравенство*

$$L(n) \gtrsim \frac{2^n}{n}.$$

Упражнение 28. Усилить теорему 38: доказать, что в условиях теоремы 38 для любого $\varepsilon > 0$ доля систем функций $\{f_1, \ldots, f_m\}$ из класса F_{nm}, удовлетворяющих неравенству

$$L_B(\{f_1, \ldots, f_m\}) \geq \frac{1}{r(B)-1} \frac{\log |F_{nm}|}{\log\log |F_{nm}|} \left(1 + (1-\varepsilon)\frac{\log\log\log |F_{nm}|}{\log\log |F_{nm}|}\right),$$

стремится к 1 при $n \to \infty$.

`Указание`. Подправить доказательство теоремы 38, положив

$$l_\varepsilon = \frac{1}{r(B)-1} \frac{\log |F_{nm}|}{\log\log |F_{nm}|} \left(1 + (1-\varepsilon)\frac{\log\log\log |F_{nm}|}{\log\log |F_{nm}|}\right).$$

Упражнение 29. Пусть $\mathcal{L}_{nn}$ — множество всех булевых линейных (n,n)-операторов (т. е. наборов из n линейных булевых функций от n переменных). Для функции Шеннона сложности класса $\mathcal{L} = \cup\mathcal{L}_{nn}$ доказать соотношения:

а) $\frac{n^2}{2\log n} \lesssim L^{\mathcal{L}}_{\{\oplus,\&,\neg\}}(n) \lesssim \frac{n^2}{\log n}$;

б)* $L^{\mathcal{L}}_{\{\oplus,\&,\neg\}}(n) \sim \frac{n^2}{2\log n}$.

Итак, известные нижние оценки сложности реализации булевых функций можно содержательно отнести к одному из двух типов — либо это линейные относительно числа полюсов (входов и выходов) оценки для конкретных конструктивно задаваемых последовательностей функций, либо мощностые (энтропийные) оценки для почти всех функций из некоторого класса, причем для «широких» классов последние оценки экспоненциальные.

С этой точки зрения особняком стоит рассмотренная ранее «неклассическая» задача о сложности реализации двоичных наборов схемами конкатенации. Для функции Шеннона $L_c(n)$ посредством применения оценки из леммы 21 на число схем с двумя входами и одним выходом и сравнивнением числа схем конкатенации сложности не более $(1-\varepsilon)\frac{n}{\log n}$ с числом двоичных наборов длины n легко устанавливается мощностная нижняя оценка

$$L_c(n) \gtrsim \frac{n}{\log n},$$

являющаяся, как мы уже знаем, асимптотически неулучшаемой. Однако такую же нижнюю оценку мы ранее получили для конструктивно задаваемого набора — первых n разрядов последовательности де Брёйна соответствующего порядка.

§ 4.5 Асимптотически наилучший метод О. Б. Лупанова

При реализации не системы, а одной булевой функции теоремы 36 и 38 устанавливают порядок роста функции Шеннона $L_B(n)$ в произвольном базисе B. Следующая теорема утверждает, что устанавливаемая теоремой 38 нижняя оценка сложности реализации одной булевой функции асимптотически неулучшаема. Этот факт справедлив для произвольного полного конечного базиса B, однако ввиду технических трудностей при доказательстве общего случая, он будет доказан только для базиса $B_0 = \{\vee, \&, {}^-\}$.

Теорема 39 (О. Б. Лупанов)**.** *Для произвольного конечного полного базиса B при $n \to \infty$ верно соотношение*

$$L_B(n) \leq \frac{1}{r(B)-1}\frac{2^n}{n}\left(1 + O\left(\frac{\log n}{n}\right)\right).$$

Доказательство. Установим утверждение теоремы для базиса $B_0 = \{\vee, \&, {}^-\}$. Опишем метод, который позволяет для произвольной (в том числе и для самой сложной) функции от n переменных построить схему, состоящую не более чем из $\frac{2^n}{n}\left(1 + O\left(\frac{\log n}{n}\right)\right)$ элементов.

Пусть $k = k(n)$ — натуральный параметр, удовлетворяющий при $n \to \infty$ условиям $k \to \infty$ и $n - k \to \infty$. Точное значение этого параметра укажем позже.

Таблицу из 2^n значений произвольной функции $f(x_1, \ldots, x_n)$ представим в виде прямоугольной таблицы высоты 2^k и ширины 2^{n-k} как показано на рис. 4.6.

Пусть $s = s(n)$ — также некоторый натуральный параметр, удовлетворяющий при $n \to \infty$ условиям $s \to \infty$ и $\frac{2^k}{s} \to \infty$. Точное значение этого параметра укажем позже. Таблицу разобъем на горизонтальные полосы $A_1, \ldots, A_p$ высоты s (полоса A_p имеет высоту $s' \leq s$), $p = \left\lceil \frac{2^k}{s} \right\rceil$. Для $i = 1, \ldots, p$ через $f_i(x_1, \ldots, x_n)$ обозначим функцию, значения которой совпадают со значениями функции $f(x_1, \ldots, x_n)$ на полосе A_i, и равны 0 на остальных полосах. Тогда

$$f(x_1, \ldots, x_n) = \bigvee_{i=1}^{p} f_i(x_1, \ldots, x_n).$$

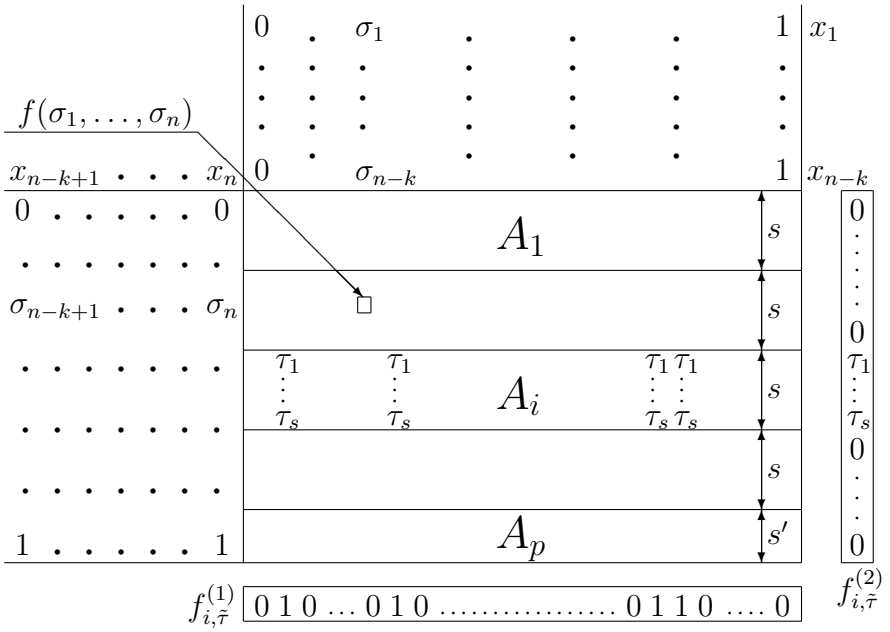

Рис. 4.6:

Теперь для каждой пары $(i,\tilde\tau)$, $i=1,\ldots,p$, $\tilde\tau\in E^s$ (или $\tilde\tau\in E^{s'}$ при $i=p$), обозначим через $f_{i,\tilde\tau}(x_1,\ldots,x_n)$ функцию, таблица которой получается из таблицы функции $f(x_1,\ldots,x_n)$ путем обнуления всех полос, кроме полосы A_i, а также обнуления всех столбцов полосы A_i, значения в которых не совпадают с набором $\tilde\tau$. Тогда

$$f(x_1,\ldots,x_n)=\bigvee_{i=1}^{p}\bigvee_{\tilde\tau} f_{i,\tilde\tau}(x_1,\ldots,x_n).$$

Наконец, у каждой функции $f_{i,\tilde\tau}(x_1,\ldots,x_n)$ можно разделить переменные, точнее представить эту функцию в виде

$$f_{i,\tilde\tau}(x_1,\ldots,x_n)=f^{(1)}_{i,\tilde\tau}(x_1,\ldots,x_{n-k})f^{(2)}_{i,\tilde\tau}(x_{n-k+1},\ldots,x_n),$$

где функция $f^{(1)}_{i,\tilde\tau}(x_1,\ldots,x_{n-k})$ обращается в единицу только на таких наборах $(\sigma_1,\ldots,\sigma_{n-k})$, что в соответствующим этим наборам столбцах полосы A_i находится набор $\tilde\tau$, а столбец значений функции $f^{(2)}_{i,\tilde\tau}(x_{n-k+1},\ldots,x_n)$ совпадает с набором $\tilde\tau$ на полосе A_i, а на наборах, соответствующих остальным полосам, функция $f^{(2)}_{i,\tilde\tau}(x_{n-k+1},\ldots,x_n)$ равна 0.

Возвращаясь к представлению функции f, окончательно получаем:

$$f(x_1,\dots,x_n)=\bigvee_{i=1}^{p}\bigvee_{\tilde\tau} f_{i,\tilde\tau}^{(1)}(x_1,\dots,x_{n-k})f_{i,\tilde\tau}^{(2)}(x_{n-k+1},\dots,x_n).$$

Схема S, реализующая функцию $f(x_1,\dots,x_n)$, будет состоять из подсхем S_i, $i=1,\dots,6$, — см рис. 4.7.

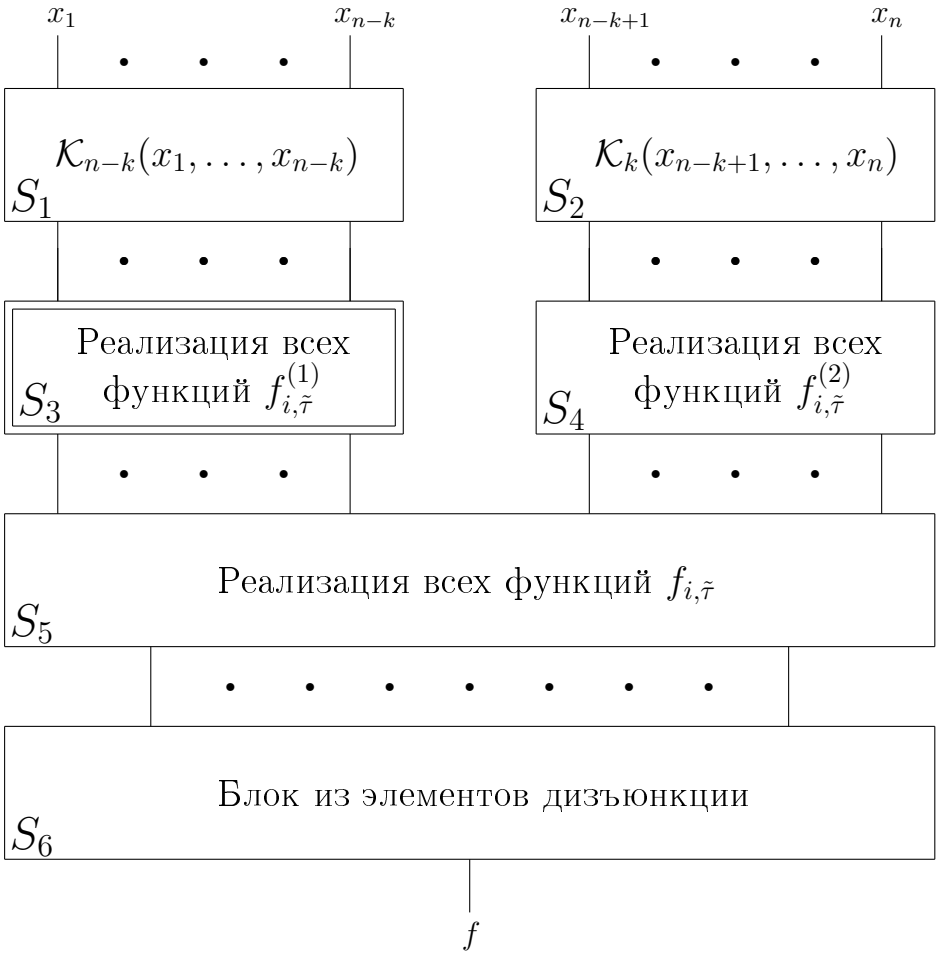

Рис. 4.7:

Подсхема S_1, на входы которой подаются переменные $x_1,\dots,x_{n-k}$, реализует систему функций $\mathcal{K}_{n-k}(x_1,\dots,x_{n-k})$. В силу леммы 18 можно считать, что при $n\to\infty$

$$L(S_1)=2^{n-k}+o\left(2^{n-k}\right)\le 2\times 2^{n-k}.$$

Подсхема S_2, на входы которой подаются переменные $x_{n-k+1}, \ldots, x_n$, реализует систему функций $\mathcal{K}_k(x_{n-k+1}, \ldots, x_n)$. Также в силу леммы 18 можно считать, что при $n \to \infty$

$$L(S_2) = 2^k + o\left(2^k\right) \leq 2 \times 2^k.$$

Подсхема S_3, на входы которой подаются функции системы $\mathcal{K}_{n-k}(x_1, \ldots, x_{n-k})$, состоит только из дизъюнкторов и реализует функции $f^{(1)}_{i,\tilde{\tau}}(x_1, \ldots, x_{n-k})$ для всех i и $\tilde{\tau}$ в соответствии с представлением этих функций в виде совершенной дизъюнктивной нормальной формы. Учитывая равенство[9]

$$\sum_{\tilde{\tau}} \left| N_{f^{(1)}_{i,\tilde{\tau}}} \right| = 2^{n-k},$$

справедливое для всех $i = 1, \ldots, p$, получаем такую оценку:

$$L(S_3) \leq p2^{n-k}.$$

Подсхема S_4, на входы которой подаются функции системы $\mathcal{K}_k(x_{n-k+1}, \ldots, x_n)$, состоит только из дизъюнкторов и реализует функции $f^{(2)}_{i,\tilde{\tau}}(x_{n-k+1}, \ldots, x_n)$ для всех i и $\tilde{\tau}$ также в соответствии с представлением этих функций в виде совершенной дизъюнктивной нормальной формы. Используя очевидное равенство

$$\left| N_{f^{(2)}_{i,\tilde{\tau}}} \right| \leq s,$$

выполняющееся для всех допустимых значений i и $\tilde{\tau}$, получаем:

$$L(S_4) \leq p2^s s.$$

Подсхема S_5, на входы которой подаются выходы подсхем S_3 и S_4, состоит только из конъюнкторов и реализует функции $f_{i,\tilde{\tau}}(x_1, \ldots, x_n)$ для всех i и $\tilde{\tau}$. Сложность подсхемы S_5 можно оценить числом различных пар $(i, \tilde{\tau})$:

$$L(S_5) \leq p2^s.$$

Подсхема S_6, на входы которой подаются все функции$f_{i,\tilde{\tau}}(x_1, \ldots, x_n)$ для всех i и $\tilde{\tau}$, состоит только из дизъюнкторов и реализует функцию $f(x_1, \ldots, x_n)$.Сложность подсхемы S_6 также можно оценить числом различных пар $(i, \tilde{\tau})$:

$$L(S_6) < p2^s.$$

[9]Здесь и далее через N_g обозначается множество наборов переменных функции g, на которых эта функция равна 1.

Таким образом,

$$L(f) \le L(S) = \sum_{i=1}^{6} L(S_i) \le$$
$$\le 2 \times 2^{n-k} + 2 \times 2^k + p2^{n-k} + p2^s s + 2p2^s \le$$
$$\le \left(\frac{2^k}{s} + 1\right) 2^{n-k} + \left(\frac{2^k}{s} + 1\right) 2^s(s+2) + 2^{n-k+1} + 2^{k+1} \le$$
$$\le \frac{2^n}{s} + 2^{s+k} + 2^s(s+2) + 3 \times 2^{n-k} + 2 \times 2^k.$$

Теперь полагая

$$k = \lfloor 3 \log n \rfloor, \quad s = \lfloor n - 5 \log n \rfloor,$$

удостоверяемся, что условия $k \to \infty$, $n - k \to \infty$, $s \to \infty$, $\frac{2^k}{s} \to \infty$ выполнены. Подставляя значения k и s в полученную оценку, имеем:

$$L(f) \le \frac{2^n}{n}\left(1 + O\left(\frac{\log n}{n}\right)\right).$$

Утверждение теоремы в базисе $B_0 = \{\vee, \&, {}^-\}$ следует из справедливости этой оценки для функции $f(x_1, \ldots, x_n)$, удовлетворяющей условию $L(f) = L(n)$. □

Упражнение 30. Доказать, что при $n \to \infty$ для любого $r \ge 2$ справедливо асимптотическое равенство

$$L_{\{x_1 \& \ldots \& x_r, \overline{x}\}}(n) \sim \frac{1}{r-1}\frac{2^n}{n}.$$

Отметим важный факт, который вытекает из теорем 38 и 39 — почти все функции от n переменных имеют сложность, асимптотически совпадающую со сложностью самой сложной функции. Такой эффект называется *эффектом Шеннона.*

§ 4.6 Принцип локального кодирования и его применения

Разработанный О. Б. Лупановым *принцип локального кодирования* имеет весьма общий характер и заключается в сведении путем кодирования реализации функций из заданного класса к реализации функций от меньшего числа переменных. При этом кодирование должно обладать

некоторыми свойствами. Во-первых, оно должно быть локальным (это значит, что для вычисления значения функции на конкретном наборе достаточно знать только часть кода относительно небольшой длины), во-вторых, вычисление функции по куску кода осуществляется со сложностью, существенно меньшей общей сложности схемы, а в-третьих, само кодирование должно быть достаточно экономным (например, асимптотически наилучшим). Не будем давать точных формулировок, связанных с использованием принципа локального кодирования, а рассмотрим отдельные простые примеры его применения.

4.6.1 Реализация симметрических функций

Прежде чем непосредственно заняться реализацией симметрических функций, рассмотрим две задачи, вспомогательные для данного раздела, но имеющие серьезное самостоятельное значение.

Обозначим через Σ_n булев оператор суммирования n-разрядных чисел, т. е. булеву $(2n, n+1)$-функцию, которая по двум n-разрядным двоичным числам вычисляет $(n+1)$-разрядое двоичное представление их суммы, а через N_n — булев оператор подсчета числа единиц в наборе длины n, т. е. булеву $(n, \lceil\log(n+1)\rceil)$-функцию, которая по n-разрядному двоичному набору вычисляет $\lceil\log(n+1)\rceil$-разрядое двоичное представление количества единиц в этом наборе.

Лемма 22. *Для любого конечного полного базиса B при $n \to \infty$ верно равенство*

$$L_B(\Sigma_n) = O(n).$$

Доказательство. Утверждение леммы в силу теоремы 35 достаточно доказать для какого-нибудь конкретного базиса. Пусть базис B содержит функции $x\&y$, $x \vee y$ и $x \oplus y$.

Построим схему S, которая по двум группам входов — $(x_1, \ldots, x_n)$ и $(y_1, \ldots, y_n)$, на которые подаются двоичные n-разрядные числа (младшие разряды x_1 и y_1), вычисляет набор $(z_1, \ldots, z_{n+1})$, представляющий двоичную запись их суммы. Тогда, обозначив через u_i, $i = 2, \ldots, n+1$, значение переноса в i-й разряд, получаем:

$$\begin{gathered} z_1 = x_1 \oplus y_1, \quad u_2 = x_1 y_1 \\ z_i = x_i \oplus y_i \oplus u_i, \quad u_{i+1} = x_i y_i \vee x_i u_i \vee y_i u_i, \qquad i = 2, \ldots, n-1; \\ z_n = x_n \oplus y_n \oplus u_n, \quad z_{n+1} = u_{n+1} = x_n y_n \vee x_n u_n \vee y_n u_n. \end{gathered}$$

Следовательно, $L_B(\Sigma_n) \leq 7n - 5$. □

Лемма 23. *Для любого конечного полного базиса B при $n \to \infty$ верно равенство*

$$L_B(N_n) = O(n).$$

Доказательство. Очевидно, что результат применения оператора N_n равен двоичной записи суммы n одноразрядных двоичных чисел. Опишем способ вычисления этой суммы схемой линейной сложности.

Сначала будем считать, что $n = 2^k$ для некоторого k. Построим схему S, имеющую k ярусов. Ярус с номером t, $t = 1, \ldots, k$, будет состоять из 2^{k-t} подсхем, каждая из которых реализует оператор Σ_t и, следовательно, имеет две группы по t входов, а также $t+1$ выходов. Таким образом, считая в силу леммы 22, что $L_B(\Sigma_t) \leq ct$, в случае, когда $n = 2^k$, имеем:

$$L_B(N_n) \leq L_B(S) \leq \sum_{t=1}^{k} 2^{k-t} ct = c2^k \sum_{t=1}^{k} \frac{t}{2^t} < 2c2^k = 2cn.$$

Переходя к общему случаю, полагаем $n' = 2^{\lceil \log n \rceil}$. Очевидно, что $n \leq n' < 2n$. Схему, реализующую оператор N_n, можно получить из схемы S', реализующую оператор $N_{n'}$, подав на $n' - n$ входов схемы S' константу 0. Поэтому

$$L_B(N_n) \leq L_B(0) + L_B(N_{n'}) \leq L_B(0) + 2cn' \leq L_B(0) + 4cn = O(n).$$

□

Теперь рассмотрим реализацию симметрических булевых функций. Напомним, что функция $f(x_1, \ldots, x_n)$ называется *симметрической*, если для любой перестановки σ из симметрической группы S_n выполняется равенство $f(x_{\sigma(1)}, \ldots, x_{\sigma(n)}) = f(x_1, \ldots, x_n)$.

Теорема 40. *Для произольного конечного полного базиса B найдутся такие положительные константы c_1 и c_2, что для любой симметрической булевой функции $f(x_1, \ldots, x_n)$, отличной от константы, выполняются неравенства*

$$c_1 n \leq L_B(f(x_1, \ldots, x_n)) \leq c_2 n.$$

Доказательство. Нижняя оценка непосредственно следует из теоремы 37 в силу существенной зависимости от всех своих переменных любой отличной от константы симметрической функции.

Переходя к доказательству верхней оценки отметим, что произвольная симметрическая функция f от n переменных может быть задана двоичной последовательностью $\tilde{\pi}(f) = (\pi_0(f), \pi_1(f), \ldots, \pi_n(f))$, где $\pi_k(f)$ —

значение функции f на наборах, состоящих из k единиц и $n-k$ нулей. Значение симметрической функции $f(x_1,\ldots,x_n)$ однозначно определяется по числу единиц в наборе $x_1,\ldots,x_n$, а следовательно, по двоичной записи этого числа. На этом и основан метод построения схемы S, вычисляющей функцию $f(x_1,\ldots,x_n)$.

Схема S состоит из подсхем S_1 и S_2. Подсхема S_1 реализует оператор N_n, на выходах подсхемы S_1 вычисляется двоичная запись длины $\lceil\log(n+1)\rceil$ числа единиц во входном наборе. Подсхема S_2 по двоичной записи числа единиц во входном наборе вычисляет значение функции f на этом наборе. В силу леммы 23 и теоремы 36 получаем:

$$L_B(f) \le L_B(N_n) + L_B(\lceil\log(n+1)\rceil) = O(n) + O\left(\frac{n}{\log n}\right) = O(n).$$

□

4.6.2 Реализация самодвойственных функций

В соответствии с введенными ранее обозначениями функция Шеннона сложности реализации самодвойственных функций в базисе B_0 определяется равенством

$$L^S(n) = \max_{f(x_1,\ldots,x_n)\in S} L(f).$$

Теорема 41. *При $n\to\infty$ справедливо асимптотическое равенство*

$$L^S(n) \sim \frac{2^{n-1}}{n}.$$

Доказательство. В силу равенства $|S(n)| = 2^{2^{n-1}}$ теорема 38 дает такую нижнюю оценку:

$$L^S(n) \gtrsim \frac{2^{n-1}}{n-1} \sim \frac{2^{n-1}}{n}.$$

Построим схему в базисе $B_0 = \{\vee, \&, {}^-\}$ для произвольной самодвойственной функции $f(x_1,\ldots,x_n)$. Положим

$$g(x_1,\ldots,x_{n-1}) = f(x_1,\ldots,x_{n-1},0).$$

Тогда

$$f(x_1,\ldots,x_{n-1},1) = \overline{f(\overline{x}_1,\ldots,\overline{x}_{n-1},0)} = \overline{g(\overline{x}_1,\ldots,\overline{x}_{n-1})}.$$

Таким образом, для реализации функции $f(x_1,\ldots,x_n)$ достаточно при $x_n = 0$ реализовать функцию $g(x_1,\ldots,x_{n-1})$, а при $x_n = 1$ — функцию $\overline{g(\overline{x}_1,\ldots,\overline{x}_{n-1})}$.

Обозначим схему, изображенную на рис. 4.1, через $S_\oplus$. Преобразуем минимальную схему S_g, реализующую в базисе B_0 функцию $g(x_1, \ldots, x_{n-1})$, в схему S_f, реализующую функцию $f(x_1, \ldots, x_n)$. Для этого следующим образом используем n экземпляров схемы $S_\oplus$: во-первых, на i-й вход, $i = 1, \ldots, n-1$, схемы S_g вместо переменной x_i подадим выход схемы $S_\oplus$, на входы которой, в свою очередь, подадим переменные x_i и x_n; во-вторых, выход схемы S_g подадим на вход еще одной схемы $S_\oplus$, на второй вход которой подается переменная x_n. Выход последней подсхемы $S_\oplus$ объявим выходом схемы S_f. Тогда схема S_f реализует функцию f, причем

$$L(S_f) = L(S_g) + nL(S_\oplus) = L(S_g) + 4n.$$

Применяя для оценки сложности минимальной схемы, реализующей функцию g от $n-1$ переменной, теорему 39, получаем требуемую верхнюю оценку. □

4.6.3 Реализация функции на r последовательных наборах.

Упорядочим все двоичные наборы длины n лексикографически. Кроме того, будем считать, что за набором $(1, \ldots, 1)$ следует набор $(0, \ldots, 0)$. Тем самым для любого набора $(\sigma_1, \ldots, \sigma_n)$ однозначно определен набор, следующий за ним.

Пусть r — фиксированное натуральное число, а f — произвольная булева функция от n переменных. Обозначим через $f^{(r)}(x_1, \ldots, x_n)$ булеву (n, r)-функцию, которая по произвольному набору $(\sigma_1, \ldots, \sigma_n)$ вычисляет значение $f(\sigma_1, \ldots, \sigma_n)$, а также значение функции f на $r-1$ наборах, следующих за набором $(\sigma_1, \ldots, \sigma_n)$. Положим

$$L^{(r)}(n) = \max_{f \in P_2(n)} L\left(f^{(r)}\right).$$

Теорема 42. *При $n \to \infty$ справедливо асимптотическое равенство*

$$L^{(r)}(n) \sim \frac{2^n}{n}.$$

Доказательство. Нижняя оценка непосредственно следует из нижней оценки функции Шеннона для случая реализации одной функции (следствие из теоремы 38).

Пусть $f(x_1, \ldots, x_n)$ — произвольная булева функция. Для доказательства верхней оценки построим схему S в базисе $B_0 = \{\vee, \&, \bar{}\}$, на r

выходах которой реализуется значение функции f на входном наборе переменных, а также на $r-1$ наборах, следующих за этим набором. Схема S будет состоять из двух подсхем, которые обозначим S_1 и S_2. Эти подсхемы будут зависеть от натурального параметра $k = k(n)$, удовлетворяющего условиям $k \to \infty$ и $n - k \to \infty$ при $n \to \infty$.

Подсхема S_1, на входы которой подаются переменные $x_1, \ldots, x_{n-k}$, имеет $2^k + r - 1$ выходов, которые на входном наборе $(\sigma_1, \ldots, \sigma_{n-k})$ вычисляют значения $f(\sigma_1, \ldots, \sigma_{n-k}, 0, \ldots, 0, 0)$, $f(\sigma_1, \ldots, \sigma_{n-k}, 0, \ldots, 0, 1)$, $\ldots$, $f(\sigma_1, \ldots, \sigma_{n-k}, 1, \ldots, 1)$, а также значения функции f на $r-1$ наборах, следующих за набором $(\sigma_1, \ldots, \sigma_{n-k}, 1, \ldots, 1)$.

Подсхема S_2, на входы которой подаются выходы подсхемы S_1, а также переменные $x_{n-k+1}, \ldots, x_n$, по конкретному набору $(\sigma_{n-k+1}, \ldots, \sigma_n)$ значений переменных $x_{n-k+1}, \ldots, x_n$ выбирает нужный кусок длины r значений, поступающих на первые $2^k + r - 1$ входов, т. е. значение $f(\sigma_1, \ldots, \sigma_n)$, а также значение функции f на $r-1$ наборах, следующих за набором $(\sigma_1, \ldots, \sigma_n)$.

Построение подсхем S_1 и S_2 методом Лупанова (см. теорему 39) отдельно для каждого выхода дает следующие оценки:

$$L\left(f^{(r)}\right) \leq L(S) = L(S_1) + L(S_2) \lesssim$$
$$\lesssim \left(2^k + r - 1\right) \frac{2^{n-k}}{n-k} + r \frac{2^{2^k + r - 1 + k}}{2^k + r - 1 + k} =$$
$$= \frac{2^n}{n-k} + O\left(\frac{2^{n-k}}{n-k}\right) + O\left(2^{2^k}\right).$$

Полагая $k(n) = \lfloor \log n \rfloor - 1$ (в качестве $k(n)$ можно взять любую неограниченно возрастающую функцию, не превосходящую указанную), получаем оценку

$$L\left(f^{(r)}\right) \lesssim \frac{2^n}{n},$$

которая в силу произвольности выбора функции $f(x_1, \ldots, x_n)$ завершает доказательство верхней оценки. □

4.6.4 Сложность инвариантных классов Яблонского

Напомним, что множество булевых функций Q является инвариантным классом Яблонского, если для любой функции из класса Q классу Q также принадлежат все равные ей функции, все конгруэнтные (подобные) ей функции, а также все ее подфункции. Количество функций от n фиксированных переменных в инвариантном классе Q выражается через

параметр инвариантного класса, определяемого равенством

$$\sigma_Q = \log_2 \left(\lim_{n\to\infty} \sqrt[2^n]{|Q(n)|} \right),$$

следующим образом:

$$|Q(n)| = 2^{(\sigma_Q + \varepsilon(n))2^n},$$

где $\varepsilon(n) \to 0$ при $n \to \infty$.

В соответствии с общим определением функция Шеннона сложности инвариантного класса Q в базисе B_0 определяется равенством

$$L^Q(n) = \max_{f \in Q(n)} L(f).$$

Теорема 43. *Пусть Q — инвариантный класс Яблонского с параметром σ_Q. Тогда*

$$\lim_{n\to\infty} \frac{L^Q(n)}{L(n)} = \sigma_Q.$$

Доказательство. Достаточно установить, что при $n \to \infty$ в случае выполнения условия $\sigma_Q \neq 0$ справедливо асимптотическое соотношение

$$L^Q(n) \sim \sigma_Q \frac{2^n}{n},$$

а для любого нулевого инвариантного класса выполняется равенство

$$L^Q(n) = o\left(\frac{2^n}{n}\right).$$

Нижняя оценка в случае ненулевого инвариантного класса непосредственно следует из теоремы 38.

Пусть $f(x_1, \ldots, x_n)$ — произвольная функция из $Q(n)$. Введем натуральный параметр $k = k(n)$, удовлетворяющий условиям $k \to \infty$ и $n - k \to \infty$ при $n \to \infty$. Положим

$$l = \lceil \log |Q(k)| \rceil.$$

Все функции из множества $Q(k)$ закодируем двоичными наборами длины l. Построим схему S, реализующую функцию f, как показано на рис. 4.8. Схема S состоит из трех подсхем, которые обозначим S_1, S_2 и S_3.

Подсхема S_1, на входы которой подаются переменные $x_1, \ldots, x_{n-k}$, имеет l выходов. Эта подсхема по набору $(\alpha_1, \ldots, \alpha_{n-k})$ значений своих

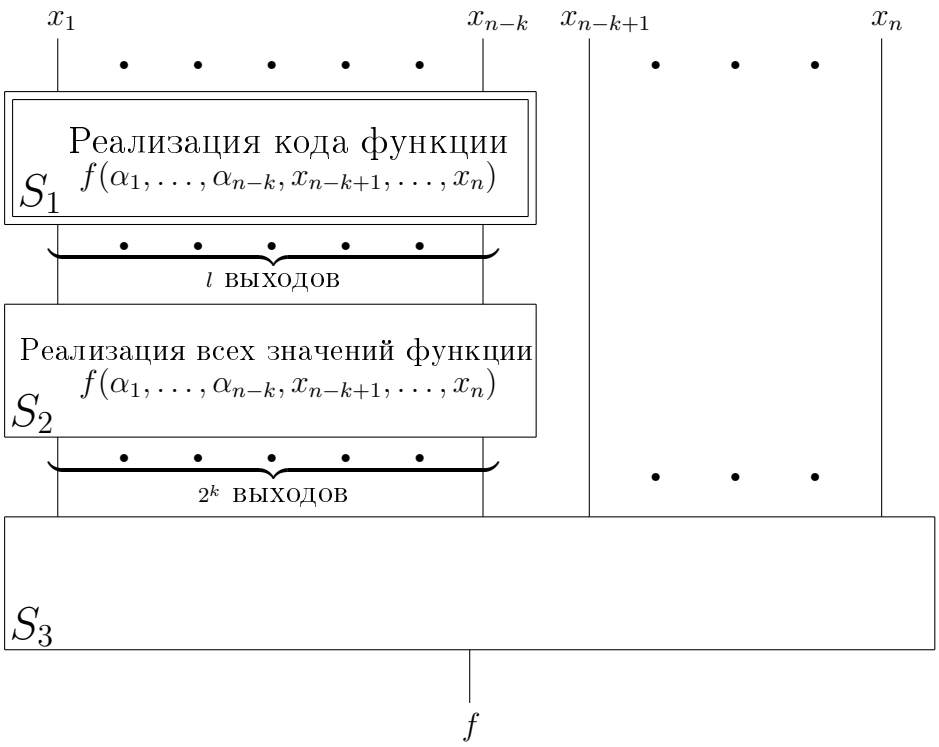

Рис. 4.8:

входов вычисляет код функции $f((\alpha_1, \ldots, \alpha_{n-k}, x_{n-k+1}, \ldots, x_n)$, принадлежащей инвариантному классу Q.

Подсхема S_2, на входы которой подаются выходы подсхемы S_1, вычисляет все 2^k значений функции $f((\alpha_1, \ldots, \alpha_{n-k}, x_{n-k+1}, \ldots, x_n)$, т. е. значения $f((\alpha_1, \ldots, \alpha_{n-k}, 0, \ldots, 0), \ldots, f((\alpha_1, \ldots, \alpha_{n-k}, 1, \ldots, 1)$.

На входы подсхемы S_3 подаются выходы подсхемы S_2 и переменные $x_{n-k+1}, \ldots, x_n$. Эта подсхема по набору $(\alpha_{n-k+1}, \ldots, \alpha_n)$ переменных $x_{n-k+1}, \ldots, x_n$ среди подаваемых на вход подсхемы значений $f((\alpha_1, \ldots, \alpha_{n-k}, 0, \ldots, 0), \ldots, f((\alpha_1, \ldots, \alpha_{n-k}, 1, \ldots, 1)$ выбирает нужное значение $f(\alpha_1, \ldots, \alpha_n)$.

Учитывая, что $l = (\sigma_Q + o(1))2^k$ при $n \to \infty$, построение подсхем S_1, S_2 и S_3 методом Лупанова (см. теорему 39) отдельно для каждого выхода дает следующие оценки:

$$L(f) \leq L(S) = L(S_1) + L(S_2) + L(S_3) \lesssim$$
$$\lesssim l\frac{2^{n-k}}{n-k} + 2^k\frac{2^l}{l} + \frac{2^{2^k+k}}{2^k+k} \leq (\sigma_Q + o(1))\frac{2^n}{n-k} + o\left(2^{2^k+k}\right).$$

Полагая $k(n) = \lfloor (\log n)/2 \rfloor$, получаем оценку

$$L(f) \leq \sigma_Q \frac{2^n}{n} + o\left(\frac{2^n}{n}\right),$$

которая в силу произвольности выбора функции $f(x_1, \ldots, x_n)$ завершает доказательство верхней оценки и для нулевого и для ненулевого значения параметра σ_Q. □

§ 4.7 Теорема Яблонского

В соответствии с теоремой 35 величина сложности реализации булевой функции схемами качественно не меняется при переходе от одного конечного полного базиса к другому. Поэтому для определенности будем рассматривать сложность реализации булевых функций схемами в базисе $B_0 = \{\vee, \&, \bar{}\}$.

С одной стороны, теорема 38 утверждает, что для любого $\varepsilon > 0$ почти все булевы функции от n переменных имеют сложность больше величины $(1-\varepsilon)\frac{2^n}{n}$ (а в силу задачи 28 — даже больше величины $\frac{2^n}{n}\left(1 + (1-\varepsilon)\frac{\log n}{n}\right)$).

С другой стороны, не известно ни одного конструктивного примера булевой функции (а точнее говоря, последовательности функций n переменных $\{f_n = f_n(x_1, \ldots, x_n)\}$), для которой была бы получена нелинейно растущая по n нижняя оценка сложности. Принципиальные сложности получения высоких нижних оценок сложности индивидульных последовательностей булевых функций подтверждаются тем фактом, что эффективное (в разумном смысле) построение последовательности функций, сложность которых растет быстрее любого полинома от n, решило бы в отрицательном смысле известную проблему равенства классов P и NP (т. е. было бы доказано, что P≠NP).

Конечно, для каждого значения n можно определить функцию $f_n(x_1, \ldots, x_n)$ равенством $L(f_n) = L(n)$. Такая последовательность $\{f_n\}$, безусловно, имеет максимально возможную нижнюю оценку сложности, однако, эту последовательность трудно назвать «конкретной». Кроме того, при определении этой последовательности используется понятие сложности и доказательство нижней оценки для последовательности «самых сложных» функций, по существу, является тавтологичным. Последний недостаток можно обойти и построить последовательности функций экспоненциальной сложности, используя другие «сильные» средства, например, нумерации примитивно-рекурсивных функций или

формальные теории большой выразительной силы. При этом такие примеры в каком-то смысле сходны с последовательностью самых сложных функций своей «неосязаемостью». Еще в 1954 г. С. В. Яблонский высказал гипотезу о том, что этот недостаток является непреодолимым. Для формализации такого подхода он ввел понятие правильного алгоритма.

Алгоритм называется *правильным*, если при построении любой бесконечной последовательности функций $\{f_n(x_1, \ldots, x_n)\}$ он одновременно строит некоторый инвариантный класс, содержащий эту функцию. Определение правильного алгоритма хорошо согласуется с интуитивным пониманием конструктивности задаваемой последовательности. Действительно, под конструктивным заданием функции естественно понимать такое описание этой функции свойствами, что по этому описанию достаточно просто определяется значение функции на любом наборе. Но тогда также просто определяются значения на произвольном наборе любой подфункции этой функции, а также функций, получающихся из исходной путем переименования переменных или добавления фиктивных переменных.

Алгоритмические трудности построения последовательности самых сложных функций подтверждаются следующим утверждением, известным как теорема Яблонского о невозможности элиминации[10] перебора при построении последовательности самых сложных функций.

Теорема 44 (С. В. Яблонский)**.** *Любой правильный алгоритм, строящий последовательность функций $\{f_n(x_1, \ldots, x_n)\}$, удовлетворяющих условию $L(f_n) = L(n)$, строит множество всех функций P_2.*

Доказательство. Предположим, что правильный алгоритм, строящий последовательность функций $\{f_n(x_1, \ldots, x_n)\}$, строит инвариантный класс Q, отличный от класса всех булевых функций. Тогда выполняется неравенство $\sigma_Q < 1$, так как параметр любого инвариантного класса не превосходит единицы, причем существует только один инвариантный класс с единичным параметром — это класс P_2. Поэтому, применяя теорему 43, получаем, что при $n \to \infty$ выполняется неравенство

$$L(f_n) \leq (\sigma_Q + o(1))L(n),$$

которое противоречит условию теоремы. □

Стоить отметить, что теорема 44 является одним из уточнений весьма общей *гипотезы С. В. Яблонского*, которая может быть сформулирована следующим образом: *любой алгоритм, строящий последовательность*

[10] Элиминация — здесь: исключение, избежание.

самых сложных функций, требует числа операций не меньшего, чем полный перебор всех булевых функций.