Кибернетический сборник

НОВАЯ СЕРИЯ

ВЫПУСК

7

Сборник переводов

Под редакцией

А. А. ЛЯПУНОВА и О. Б. ЛУПАНОВА

ИЗДАТЕЛЬСТВО «МИР»

Москва 1970

УДК 519.95

Научный совет по кибернетике
Академии наук СССР

Седьмой выпуск новой серии кибериетнческих сборииков состоит из двух разделов: математические вопросы и прикладиые вопросы. В первом разделе представлены работы по теории кодирования, сиитезу логических устройств, теории графов, а также статьи по вопросам машиного доказательства теорем. Во втором разделе помещеиа статья Р. Якубовски об алгоритме моделироваиия динамических систем.

Сборник рассчитан иа научиых работиков, инженеров, аспираитов и студеитов различиых специальностей, занимающихся и иитересующихся кибериетикой.

Редакция литературы по математическим наукам

Инд 2-2-3
5-70

Предельное распределение для минимального расстояния в случайном линейном коде[1])

Дж. Н. Пирс

Распределение отношения минимального расстояния к длине кода в случайном линейном коде приближается к скачкообразному распределению, когда длина кода становится достаточно большой при фиксированной скорости. Скачок возникает при наименьшем значении p, удовлетворяющем соотношению

$$1 + p \log_2 p + (1-p) \log_2 (1-p) = k/n.$$

ВВЕДЕНИЕ

Рассматривается следующее семейство случайных кодов. Пусть $G_1, \ldots, G_k$ есть n-разрядные двоичные последовательности, не обязательно различные. Линейный код из 2^k слов $A_0, \ldots, A_{2^k-1}$ формируется из этих последовательностей с помощью соотношений

$$A_i = \sum_{j=1}^{k} B_{ij} G_j, \tag{1a}$$

где суммирование осуществляется поразрядно по модулю 2, а вектор

$$B_i = (B_{i1}, \ldots, B_{ik}) \tag{1b}$$

есть двоичное представление индекса i.

Минимальное кодовое расстояние определяется как

$$D = \min_{i \geqslant 1} D_i, \tag{2a}$$

---

[1]) Pierce J. N., Limit distribution of the minimum distance of random linear codes, *IEEE Transactions on Information Theory*, IT-13, № 4 (1967), 595 – 599.

где

D_i равно числу ненулевых координат в A_i. (2b)

Постулируется вероятностное пространство Ω, содержащее 2^{nk} точек, выбираемых с равными вероятностями 2^{-nk} и соответствующих числу способов выбора последовательностей $G_1, \ldots, G_k$.

Рассмотрим теперь бесконечную последовательность таких семейств случайных кодов с фиксированным отношением k/n и обозначим через p наименьшее решение уравнения

$$1 + p \log_2 p + (1-p)\log_2(1-p) = k/n. \tag{3}$$

Ниже будет показано, что

$$P(D/n \leqslant p) \to 0 \text{ при } n \to \infty \tag{4a}$$

и что

$$P(D/n > p + \varepsilon) \to 0 \text{ при } n \to \infty \text{ для любого } \varepsilon > 0. \tag{4b}$$

Граница Варшамова – Гилберта [1]) позволяет установить, что имеется по крайней мере один код, для которого $D/n \geqslant p$; соотношение (4a) показывает, что существует очень мало длинных кодов, для которых $D/n \leqslant p$, в то время как из (4b) следует, что существует очень мало длинных кодов, для которых D/n превышает p на сколь угодно малую величину.

НЕКОТОРЫЕ ПРЕДВАРИТЕЛЬНЫЕ СВЕДЕНИЯ

Для упрощения различных выражений повсюду будет использоваться обозначение

$$K = 2^k - 1. \tag{5}$$

Применение формулы Стирлинга приводит к следующему асимптотическому выражению для биномиальных коэффициентов:

$$\binom{n}{na} \sim (2\pi n a(1-a))^{-1/2} 2^{nH(a)}$$

при $n \to \infty$ равномерно по $\delta \leqslant a \leqslant 1 - \delta$. (6a)

Здесь символ $\sim$ означает, что отношение левой и правой частей стремится к единице; δ – произвольное положительное число; $H(a)$ – функция энтропии,

$$H(a) = -a \log_2 a - (1-a)\log_2(1-a). \tag{6b}$$

[1]) См. Питерсон У., Коды, исправляющие ошибки, „Мир", М., 1964, 68.

Используя (6) и простые границы для отношения соответствующих членов, получаем следующее выражение для суммы биномиальных коэффициентов:

$$\sum_{m=0}^{na}\binom{n}{m}\sim(1-a)^{1/2}(1-2a)^{-1}(2\pi na)^{-1/2}2^{nH(a)}$$

при $n\to\infty$ равномерно по $\delta\leqslant a\leqslant\frac{1}{2}-\delta$. **(7)**

Из (7) **легко** следует, что

$$K\sum_{m=0}^{M}\binom{n}{m}2^{-n}\to 0 \text{ при } n\to\infty, \text{ если } M\leqslant np, \quad (8a)$$

и что

$$K\sum_{m=0}^{M}\binom{n}{m}2^{-n}\to\infty \quad \text{при } n\to\infty,$$

если $M\geqslant n(p+\varepsilon)$ для любого фиксированного $\varepsilon>0$. (8b)

В дальнейшем понадобится аппроксимировать некоторое число с помощью суммы, аналогичной сумме в левых частях соотношений (8a) и (8b). Легко показать, что для заданных положительных чисел x, ε_1 и ε_2 существуют число N_0, две последовательности M_n и I_n и число α, такие, что

$$np\leqslant M_n\leqslant n(p+\varepsilon_1), \quad (9a)$$

$$\alpha K\leqslant I_n\leqslant K, \quad (9b)$$

$$\left|x-I_n\sum_{m=0}^{M_n}\binom{n}{m}2^{-n}\right|\leqslant\varepsilon_2, \quad \text{если } n\geqslant N_0. \quad (9c)$$

ДОКАЗАТЕЛЬСТВО СООТНОШЕНИЯ (4a)

Запишем верхнюю границу для $P(D\leqslant d)$,

$$P(D\leqslant d)=P\left(\bigcup_{i=1}^{K}(D_i\leqslant d)\right)\leqslant\sum_{i=1}^{K}P(D_i\leqslant d). \quad (10)$$

Однако легко убедиться, что символы в каждой последовательности A_i независимы и принимают с равной вероятностью значения 1 и 0, так что

$$P(D_i\leqslant d)\leqslant\sum_{m=0}^{d}\binom{n}{m}2^{-n} \quad (11)$$

и, следовательно,

$$P(D\leqslant d)\leqslant K\sum_{m=0}^{d}\binom{n}{m}2^{-n}. \quad (12)$$

Используя (8a), немедленно получаем

$$P(D \leqslant d) \to 0 \quad \text{при } n \to \infty, \text{ если } d \leqslant np, \tag{13}$$

что и завершает доказательство соотношения (4a).

ДОКАЗАТЕЛЬСТВО СООТНОШЕНИЯ (4b)

Некоторые пояснения. Для того чтобы пояснить использованный метод доказательства, полезно доказать сначала более слабое утверждение:

$$\lim_{n\to\infty} \inf P(D \leqslant n(p+\varepsilon)) \geqslant \frac{1}{2}. \tag{14}$$

Для упрощения обозначений положим

$$S = (D \leqslant d), \tag{15a}$$

$$S_i = (D_i \leqslant d). \tag{15b}$$

Следует иметь в виду, что события S_i и S **зависят от** d. Далее, при $I \leqslant K$ имеем

$$S = \bigcup_{i=1}^{K} S_i \supset \bigcup_{i=1}^{I} S_i, \tag{16a}$$

так что

$$P(S) \geqslant P\left(\bigcup_{i=1}^{I} S_i\right). \tag{16b}$$

Кроме того,

$$\bigcup_i S_i = \bigcup_{i_1} \left(S_{i_1} - S_{i_1} \cap \left(\bigcup_{i_2<i_1} S_{i_2}\right)\right), \tag{17a}$$

поэтому

$$P\left(\bigcup_i S_i\right) = \sum_{i_1} P\left(S_{i_1} - S_{i_1} \cap \left(\bigcup_{i_2<i_1} S_{i_2}\right)\right) = \sum_{i_1}\left(P(S_{i_1}) - P\left(S_{i_1} \cap \right.\right.$$

$$\left.\left. \cap \left(\bigcup_{i_2<i_1} S_{i_2}\right)\right)\right) = \sum_{i_1} P(S_{i_1}) - \sum_{i_1} P\left(\bigcup_{i_2<i_1} (S_{i_1} \cap S_{i_2})\right)$$

и, следовательно,

$$P\left(\bigcup_{i=1}^{I} S_i\right) \geqslant \sum_{i_1=1}^{I} P(S_{i_1}) - \sum_{i_1=1}^{I} \sum_{i_2=1}^{i_1-1} P(S_{i_1} \cap S_{i_2}). \tag{17b}$$

(Написанные равенства следуют из того, что объединение множеств можно представить, как видно из (17a), в виде объединения непересекающихся множеств, каждое из которых есть разность между множеством и подмножеством; в последнем

неравенстве использована обычная верхняя граница для вероятности объединения.)

Следует заметить, что для любой последовательности $i_1, \ldots, i_M$ имеет место

$$P(S_{i_1} \cap \ldots \cap S_{i_M}) = \prod_{m=1}^{M} P(S_{i_m}) = [P(S_i)]^M, \tag{18}$$

если $(B_{i_1}, \ldots, B_{i_M})$ линейно независимы, где векторы B_i суть векторы, определенные в (1b). В частности, так как B_{i_1} и B_{i_2} всегда линейно независимы при $i_1 \neq i_2$, то из (16b) и (17b) получаем

$$P(S) \geqslant Iy - I(I-1)y^2/2, \tag{19a}$$

где

$$y = P(S_i) = \sum_{m=1}^{d} \binom{n}{m} 2^{-n}. \tag{19b}$$

Но, согласно (9), при достаточно больших n, некотором d из интервала $(np,\ n(p+\varepsilon))$ и подходящем выборе I величину Iy можно сделать как угодно близкой к единице, и, следовательно, нижнюю границу в (19a) можно сделать как угодно близкой к 1/2, что и доказывает соотношение (14).

Естественно попытаться распространить этот подход для доказательства (4b), используя границу в виде знакопеременного ряда

$$P(\bigcup S_i) \geqslant \sigma_1 - \sigma_2 + \ldots - \sigma_{2J},$$

где

$$\sigma_J = \sum_{i_1 > \ldots > i_J} P(S_{i_1} \cap \ldots \cap S_{i_J}).$$

Всего в многократной сумме содержится $\binom{I}{j}$ слагаемых; если бы события под знаком пересечения были независимы, то была бы применима следующая оценка снизу:

$$P(S) \geqslant \sum_{j=1}^{2J} (-1)^{j+1} \binom{I}{j} y^j.$$

Если в соответствии с (9) аппроксимировать Iy некоторым значением x, то эта нижняя граница приближается к величине

$$x - \frac{x^2}{2!} + \ldots - \frac{x^{2J}}{(2J)!},$$

которая в свою очередь равна приближенно $1 - \exp(-x)$, если J выбрано с самого начала достаточно большим.

Чтобы этот способ доказательства был эффективным и в том случае, когда некоторые события в пересечениях не являются независимыми, необходимо привести знакопеременную границу к такому виду, для которого можно было бы получить достаточно точные оценки вероятностей зависимых событий. Желаемый результат будет получен в следующих леммах.

Лемма 1. *Пусть* θ_j *есть множество* j*-наборов целых чисел, удовлетворяющих условиям*

$$(i_1, \ldots, i_j) \in \theta_j \Rightarrow I \geqslant i_1 > \ldots > i_j \geqslant 1, \tag{20a}$$

$$\left.\begin{array}{l}(i_1, \ldots, i_{2J-1}) \in \theta_{2J-1} \\ (i_1, \ldots, i_{2J}) \in \theta_{2J}\end{array}\right\} \Rightarrow (B_{i_1}, \ldots, B_{i_{2J-1}}) \text{ линейно независимы.} \tag{20b}$$

Пусть

$$\sigma_J = \sum_{\theta_J} P(S_{i_1} \cap \ldots \cap S_{i_J}), \tag{21}$$

и пусть

$$\nu_J = \sum_{\theta_J} P\left(S_{i_1} \cap \ldots \cap S_{i_J} \cap \left(\bigcup_{i_{J+1} < i_J} S_{i_{J+1}}\right)\right). \tag{22}$$

Тогда

$$P\left(\bigcup S_i\right) \geqslant \sum_{j=1}^{2J-1} (-1)^{j+1} \sigma_j - \nu_{2J-1} \text{ для любого } J. \tag{23}$$

Доказательство леммы 1 проводится индукцией по J. Первые два шага, согласно (17a), можно записать в виде

$$P\left(\bigcup S_i\right) = \sigma_1 - \nu_1, \tag{24}$$

что показывает справедливость соотношения (23) при $J = 1$ Далее,

$$-\nu_{2J-1} = -\sum_{\theta_{2J-1}} P\left(S_{i_1} \cap \ldots \cap S_{i_{2J-1}} \cap \left(\bigcup_{i_{2J} < i_{2J-1}} S_{i_{2J}}\right)\right). \tag{25}$$

Используя разложения того же типа, которые привели к неравенству (17b), выражение (25) можно переписать в виде

$$-\nu_{2J-1} = -\sum P(S_{i_1} \cap \ldots \cap S_{i_{2J}}) + \sum P\left[S_{i_1} \cap \ldots \cap S_{i_{2J}} \cap \left(\bigcup_{i_{2J+1} < i_{2J}} S_{i_{2J+1}}\right)\right]. \tag{26}$$

В обеих суммах суммирование ведется по $(i_1, \ldots, i_{2J-1}) \in \theta_{2J-1}$, $i_{2J} < i_{2J-1}$. Ограничения на числа i в суммах эквивалентны определению множества θ_{2J}, так что первая сумма есть просто

σ_{2J}. Вторую сумму можно ограничить снизу, отбросив достаточно много значений i_{2J} под знаком суммы и i_{2J+1} под знаком объединения так, чтобы набор из $2J+1$ чисел i удовлетворял ограничению $(i_1, \ldots, i_{2J+1}) \in \theta_{2J+1}$. Таким образом, получаем, что

$$-\nu_{2J-1} \geqslant -\sigma_{2J} + \sum_{\theta_{2J+1}} P\left(S_{i_1} \cap \ldots \cap S_{i_{2J}} \cap \left(\bigcup S_{i_{2J+1}}\right)\right). \quad (27)$$

Каждую вероятность в сумме снова можно переписать в другом виде, используя разложения того же типа, что и раньше. В результате получим

$$\sum_{\theta_{2J+1}} P\left(S_{i_1} \cap \ldots \cap S_{i_{2J}} \cap \left(\bigcup S_{i_{2J+1}}\right)\right) = \sigma_{2J+1} - \nu_{2J+1}.$$

Отсюда следует, что

$$-\nu_{2J-1} \geqslant -\sigma_{2J} + \sigma_{2J+1} - \nu_{2J+1}, \quad (28)$$

т. е. индукция завершена.

Лемма 2.

$$P\left(\bigcup_{i=1}^{I} S_i\right) \geqslant \sum_{j=1}^{2J} (-1)^{j+1} \sigma_j \text{ для любого } J. \quad (29)$$

Доказательство. Внося в выражении (22) все пересечения под знак объединения и применяя обычную оценку для вероятности объединения событий, получаем неравенство

$$\nu_{2J-1} \leqslant \sigma_{2J}.$$

После подстановки этого неравенства в (23) получаем требуемый результат.

Лемма 3. *Для каждого J существует положительная константа c_J, такая, что*

$$P\left(\bigcup_{i=1}^{I} S_i\right) \geqslant \sum_{j=1}^{2J} (-1)^{j+1} (Iy)^j/j! - c_J \exp(Iy)/I - c_J \exp(Iy) \max_{\theta_{2J}} P\left(S_{i_{2J}} \mid S_{i_1}, \ldots, S_{i_{2J-1}}\right). \quad (30)$$

Доказательство. Число j-наборов, удовлетворяющих условию (20а), равно

$$I(I-1)(I-2)\ldots(I+1-j)/j!.$$

Число j-наборов, удовлетворяющих условию (20а) и дополнительному ограничению, что соответствующие последовательности B_i линейно независимы, больше чем

$$I(I-1)(I-3)\ldots(I+1-2^{j-1})/j!.$$

Представляя каждое из этих выражений как полином от I, можно убедиться, что для каждого j существует такое число d_j, не зависящее от I, что полное число j-наборов в θ_j ограничено сверху величиной

$$I^j/j!$$

и снизу величиной

$$I^j/j! - d_j I^{j-1}.$$

Кроме того, число j-наборов, которые удовлетворяют условию (20а) и для которых соответствующее множество последовательностей B_i имеет ранг $(j-1)$, меньше, чем разность между полным числом j-наборов и числом j-наборов, для которых соответствующие последовательности B_i линейно независимы, а потому ограничено сверху величиной

$$d_j I^{j-1}.$$

В обеих оценках число d_j можно заменить любым числом c_J, большим, чем все числа d_j при $1 \leqslant j \leqslant 2J$.

Члены, соответствующие в неравенстве (29) нечетным j, положительны, а вероятности, фигурирующие в определении величин σ_j в (21), суть вероятности пересечения независимых событий, так что можно сразу написать оценку

$$(2m-1)!\,\sigma_{2m-1} \geqslant (I^{2m-1} - c_J I^{2m-2})\,y^{2m-1} \quad \text{при} \quad m \leqslant J. \tag{31}$$

Члены, соответствующие в (29) четным j, отрицательны. Те вероятности в (21), которые, как и ранее, определяются независимыми событиями, равны y^j, а число таких членов по указанным выше причинам не превосходит $I^j/j!$. Оставшиеся члены соответствуют случаю, когда первые $(j-1)$ событий в пересечении независимы, а последнее событие зависит от них; выше было показано, что число таких членов не превосходит $c_J I^{j-1}$. Следовательно,

$$(2m)!\,\sigma_{2m} \leqslant (Iy)^{2m} + c_J (Iy)^{2m-1} \max_{\theta_{2m}} P(S_{t_{2m}} \mid S_{t_1}, \ldots, S_{t_{2m-1}}) \tag{32}$$

$$\text{при } m \leqslant J.$$

Вероятность в (32) зависит только от числа тех из предыдущих последовательностей B_t, сумма которых дает последовательность $B_{t_{2m}}$; поэтому наиболее слабые ограничения при

максимизации имеют место для случая $m = J$:

$$\max_{\theta_{2m}} P(S_{i_{2m}} | S_{i_1}, \ldots, S_{i_{2m-1}}) \leqslant \max_{\theta_{2J}} P(S_{i_{2J}} | S_{i_1}, \ldots, S_{i_{2J-1}}) \quad (33)$$

при $m \leqslant J$.

Утверждение леммы будет получено, если подставить соотношения (31) – (33) в (29) и использовать для упрощения результата слабые неравенства

$$\sum_{m=1}^{J} I^{2m-2} y^{2m-1}/(2m-1)! \leqslant I^{-1} \sum_{j=0}^{\infty} (Iy)^j/j! = I^{-1} \exp(Iy),$$

$$\sum_{m=1}^{J} (Iy)^{2m-1}/(2m)! \leqslant \sum_{j=0}^{\infty} (Iy)^j/j! = \exp(Iy).$$

Лемма 4. *Для заданного* $\delta > 0$ *существуют положительные константы* a_1 *и* a_2, *такие, что*

$$P(S_{i_{2J}} | S_{i_1}, \ldots, S_{i_{2J-1}}) \leqslant a_1 \exp(-a_2 n), \quad (34)$$

если $(i_1, \ldots, i_{2J}) \in \theta_{2J}$ *и если* $\frac{d}{n} < \frac{1}{2} - \delta$.

Доказательство. Согласно условию леммы, последовательности $A_{i_1}, \ldots, A_{i_{2J-1}}$ линейно независимы, а последовательность $A_{i_{2J}}$ равна сумме по модулю 2 некоторого числа (скажем, L) этих последовательностей. Для удобства предположим, что ими являются первые L последовательностей, и введем обозначения

$$V_j = A_{i_j}, \quad 1 \leqslant j \leqslant L, \quad (35a)$$

$$V_0 = \sum_{j>0}^{L} V_j, \quad (35b)$$

$w(V_j)$ – число единиц в последовательности V_j. (35c)

Теперь оценим вероятность ρ, определяемую следующим выражением:

$$\rho = P(w(V_0) \leqslant d | w(V_1) \leqslant d, \ldots, w(V_L) \leqslant d). \quad (36)$$

Пусть

Ψ – множество $(v_1, \ldots, v_L)$, такое, что если

$$V_j = v_j \text{ для } 1 \leqslant j \leqslant L, \text{ то } w(V_0) \leqslant d. \quad (37)$$

Тогда имеем

$$\rho = \sum_{\Psi} \prod_{j=1}^{L} P(V_j = v_j | w(V_j) \leqslant d). \quad (38)$$

Так как все возможные последовательности V_j, удовлетворяющие ограничению на полное число единиц, равновероятны,

условные вероятности в (38) определятся следующим образом:

$$P(V_j = v_j | w(V_j) \leqslant d) = \begin{cases} 0, & \text{если } w(v_j) > d, \\ p_0, & \text{если } w(v_j) \leqslant d, \end{cases} \tag{39a}$$

где

$$p_0 = \left[\sum_{m=0}^{d} \binom{n}{m} \right]^{-1}. \tag{39b}$$

Теперь рассмотрим функцию

$$g(t) = \left(\frac{d}{n}\right)^t \left(1 - \frac{d}{n}\right)^{n-t}. \tag{40}$$

Так как функция $g(t)$ монотонно убывает по t при $\frac{d}{n} < \frac{1}{2}$, имеем $g(t) \geqslant g(d)$, если $t \leqslant d$, а потому

$$\log_2 g(t) \geqslant -nH\left(\frac{d}{n}\right), \quad \text{если} \quad t \leqslant d. \tag{41}$$

Но в таком случае из (7), (39b) и (41) находим

$$\log_2 p_0 \leqslant \text{const} + \frac{1}{2}\log_2 n + \log_2 g(t) \quad \text{для} \quad t \leqslant d, \tag{42}$$

и, следовательно, ρ можно оценить с помощью неравенств

$$\rho \leqslant a n^{L/2} \rho^* \leqslant a n^l \rho^*, \tag{43a}$$

где a — некоторая константа и

$$\rho^* = \sum_{\Psi} \prod_{j=1}^{L} g[w(v_j)]. \tag{43b}$$

Но величину ρ^* можно сразу же отождествить с вероятностью того, что $w(V_0) \leqslant d$ при условии, что *все* координаты *всех* других последовательностей V_j независимы и имеют одинаковое распределенне, причем

$$P(V_{jm} = 1) = \frac{d}{n}.$$

Однако при этих условиях координаты последовательности V_0 также являются независимыми случайнымн величинами и

$$P(V_{0m} = 1) = \sum_{j \text{ нечетно}} \binom{L}{j} \left(\frac{d}{n}\right)^j \left(1 - \frac{d}{n}\right)^{L-j}.$$

Так как

$$\sum_{j \text{ нечетно}} \binom{L}{j} s^j (1-s)^{L-j} = \frac{1}{2} - \frac{1}{2}(1-2s)^L > s + \varepsilon$$

для некоторого положительного ε при всех $L \geqslant 2$ и $s < 1/2$, отсюда находим (с помощью легко получаемых оценок бино-

миального распределения), что для заданного $\delta > 0$ существует положительное число a', такое, что

$$\rho^* \leqslant \exp(-a'n), \quad \text{если} \quad \frac{d}{n} \leqslant \frac{1}{2} - \delta. \tag{44}$$

Из (43) и (44) следует утверждение леммы для некоторых a_1 и a_2 при $a_2 < a'$.

Лемма 5. *Для последовательности* (I_n, d_n), *удовлетворяющей условиям* $I_n P(D_i \leqslant d_n) \to x$ *и* $I_n \to \infty$, *имеет место неравенство*

$$\liminf_{n\to\infty} P(D \leqslant d_n) \geqslant \sum_{j=1}^{2J} (-1)^{j+1} x^j / j! \tag{45}$$

для любого фиксированного J.

Доказательство. Подставим в соотношение (30) определения S_i и y; первый из остаточных членов в оценке стремится к нулю, так как $I_n \to \infty$, второй член стремится к нулю, согласно лемме 4.

Лемма 6. *Соотношение* (4b) *справедливо.*

Доказательство. Нужно показать, что для заданных ε и ε_1 и достаточно больших n имеет место неравенство

$$P\left(\frac{D}{n} \leqslant p + \varepsilon\right) \geqslant 1 - \varepsilon_1.$$

Выберем x так, чтобы

$$\exp(-x) \leqslant \frac{\varepsilon_1}{4}; \tag{46a}$$

выберем J так, чтобы

$$1 - \exp(-x) - \sum_{j=0}^{2J} (-1)^{j+1} x^j / j! \leqslant \frac{\varepsilon_1}{4}, \tag{46b}$$

и, следовательно, так, чтобы

$$\sum_{j=0}^{2J} (-1)^{j+1} x^j / j! \geqslant 1 - \frac{\varepsilon_1}{2}. \tag{46c}$$

В силу непрерывности можно подобрать такое ε_2, что

$$\sum_{j=0}^{2J} (-1)^{j+1} (x')^j / j! \geqslant 1 - \frac{3\varepsilon_1}{4}, \quad \text{если} \quad |x' - x| \leqslant \varepsilon_2. \tag{46d}$$

Используя (9), выберем последовательность (I_n, d_n), причем $I_n \to \infty$, $d_n < n(p+\varepsilon)$, и число N_0, такое, что

$$|I_n P(D_i \leqslant d_n) - x| < \varepsilon_2, \quad \text{если } n \geqslant N_0. \tag{46e}$$

Из леммы 5 следует, что

$$\lim_{n\to\infty} \inf P(D \leqslant d_n) \geqslant 1 - \frac{3\varepsilon_1}{4}. \tag{46f}$$

Наконец, выберем $N_1 (N_1 \geqslant N_0)$ так, чтобы

$$P(D \leqslant d_n) \geqslant 1 - \varepsilon_1, \quad \text{если } n \geqslant N_1. \tag{46g}$$

Этим заканчивается доказательство леммы и соотношения (4b).

РАСПРОСТРАНЕНИЕ РЕЗУЛЬТАТОВ НА КОДЫ СО СЛУЧАЙНЫМИ ПРОВЕРКАМИ

Путем несложных преобразований результаты (4a) и (4b) можно распространить на коды со случайными проверками, в которых первые k разрядов в кодовых словах предназначены для информационных символов. В частности, пусть на элементы порождающих последовательностей наложены следующие ограничения:

$$G_{jm} = \begin{cases} 1, & \text{если } m = j, \\ 0, & \text{если } m \neq j \text{ и } m \leqslant k, \\ 0 \text{ или } 1 \text{ с одинаковой вероятностью, если } m > k. \end{cases} \tag{47}$$

Число кодов в этом подмножестве пространства Ω равно $2^{k(n-k)}$; обозначим это подмножество Ω'.

Из каждого кода в Ω' можно построить семейство кодов, используя в качестве порождающих последовательностей кодов семейства некоторые линейные комбинации последовательностей G:

$$\Gamma_i = \sum_{j=1}^{k} \gamma_{ij} G_j, \quad 1 \leqslant i \leqslant k.$$

Если матрица γ, образованная величинами γ_{ij}, не сингулярна, то упорядоченный набор кодовых слов $(A_0, \ldots, A_K)$, связанный с $(\Gamma_1, \ldots, \Gamma_k)$, есть просто перестановка кодовых слов в наборе, связанном с $(G_1, \ldots, G_k)$, и, конечно, имеет то же самое минимальное расстояние D. Если обозначить число различных несингулярных матриц γ через η, то легко проверить, что

$$\eta = \prod_{m=1}^{k} (2^k - 2^{m-1}) \tag{48}$$

и что все η кодов, образованных из заданной точки множества Ω', различны.

Пусть $\varphi(\omega)$ — набор кодов, связанных с точкой $\omega \in \Omega'$.

Пусть Ω'' — подмножество Ω, образованное объединением всех таких наборов,

$$\Omega'' = \bigcup_{\omega \in \Omega'} \varphi(\omega). \tag{49}$$

Можно доказать, что множества в объединении не пересекаются. Отсюда сразу следует, что если T — любое событие, не затрагивающее порядка в списке слов, то условная вероятность события T одна и та же на Ω' и Ω'':

$$P(T|\Omega') = P(T|\Omega''). \tag{50}$$

Далее, вероятность Ω'' есть просто отношение числа точек в Ω'' к 2^{nk}, или

$$P(\Omega'') = 2^{k(n-k)}\eta/2^{nk} = 2^{-k^2}\eta = \prod_{m=1}^{k}(1-2^{m-1-k}) \geqslant \prod_{m=1}^{\infty}(1-2^{-m}).$$

Бесконечное произведение сходится к некоторой положительной константе, скажем b, поэтому получается оценка

$$P(T|\Omega') = P(T \cap \Omega'')/P(\Omega'') \leqslant P(T)/b. \tag{51}$$

В частности, события с исчезающе малой вероятностью во всем пространстве имеют исчезающе малую вероятность и в подпространстве:

$$P(T) \to 0 \text{ при } n \to \infty \Rightarrow P(T|\Omega') \to 0 \text{ при } n \to \infty. \tag{52}$$

Таким образом, соотношения (4a) и (4b) остаются справедливыми, если ограничиться рассмотрением кодов со случайными проверками.

Класс оптимальных нелинейных кодов с исправлением двойных ошибок[1])

Ф. П. Препарата

В этой статье представлены нелинейные блочные коды длины $(2^n - 1)$ (n четно) с исправлением двойных ошибок. Они имеют максимально возможное число кодовых слов для их длины и минимального расстояния и образованы добавлением к определенному линейному коду (называемому ядром) специального подмножества из его смежных классов. Ядро получается из соединения и суперпозиции кодов Боуза – Чоудхури – Хоквингема (БЧХ кодов) длины $(2^{n-1} - 1)$. Представленные коды являются систематическими и сравнимы с соответствующими линейными кодами относительно сложности операций кодирования и декодирования.

1. ВВЕДЕНИЕ

В течение последних лет в литературе появились сообщения о некоторых примерах нелинейных двоичных кодов (Васильев [2], Надлер [8], Грин [6]). Особенно интересным в силу своей структуры и общности был класс, открытый Васильевым [2], т. е. класс совершенных групповых и негрупповых кодов с исправлением одиночных ошибок, содержащий коды Хемминга.

Недавно некоторый интерес к линейным кодам был возрожден открытием Нордстромом и Робинсоном [9] нелинейного (15,8) кода с исправлением двойных ошибок. Из этого кода путем укорачивания получаются опубликованные ранее негрупповой (12,5) код Надлера [8] и негрупповой (13,6) код Грина [6]. Код (15,8) обладает интересными свойствами, а именно он является систематическим и удовлетворяет верхней границе Джонсона [7] для числа кодовых слов в коде длины 15 и с расстоянием 5. Впоследствии (15,8) код был описан в терминах полиномиальных (т. е. линейных) кодов над полем GF (2) (Препарата [11]). Это описание оказалось полезным, так как оно привело к формальному доказательству (Препарата [12]) некоторых метрических свойств этого кода, определенных ранее эвристически.

---

[1]) Preparata F. P., A class of optimum nonlinear double-error-correcting codes, *Information and Control*, 13, № 4 (1968), 378 – 400.

Оставался открытым вопрос, впервые поставленный Нордстромом и Робинсоном [9], является ли (15,8) код элементом некоторого класса кодов. В настоящей статье на этот вопрос дается положительный ответ. Для каждого четного $n \geqslant 4$ существует негрупповой $(2^n - 1,\ 2^n - 2n)$ код с исправлением двойных ошибок, поэтому (15,8) код является частным случаем. Здесь снова полиномиальное описание оказалось существенным при построении этих кодов.

Интересные свойства этих кодов можно суммировать следующим образом: 1) они содержат удвоенное число кодовых слов по сравнению с БЧХ кодами той же длины с исправлением двойных ошибок, что является максимально возможным числом кодовых слов для данной длины и расстояния, т. е. они оптимальны; 2) их декодирование основано на вычислении величин, сходных с синдромом, и сложность его сравнима с соответствующими БЧХ кодами; 3) эти коды являются систематическими, и кодирование может быть выполнено очень просто с помощью регистров сдвига за то же число тактов, которое требуется для последовательной передачи информационных разрядов.

Следующие параграфы посвящены описанию этих кодов и доказательству сформулированных выше свойств.

2. ОПИСАНИЕ КОДОВ [1])

В дальнейшем будем считать, что все рассматриваемые полиномы принадлежат алгебре A_{n-1} полиномов над полем GF(2) по $\mathrm{mod}\left(x^{2^{n-1}-1}+1\right)$ $(n \geqslant 4)$. Для данных $a(x) \in A_{n-1}$ и $b(x) \in A_{n-1}$ пусть $W[a(x)]$ обозначает число ненулевых коэффициентов $a(x)$, и пусть $d[a(x),\ b(x)] = W[a(x) + b(x)]$ есть расстояние Хемминга между $a(x)$ и $b(x)$. Символом $a(x)$ мы будем также обозначать вектор-строку $[a_{2^{n-1}-2},\ a_{2^{n-1}-3},\ \ldots,\ a_0]$, где $a(x) = \sum a_j x^j$.

Пусть $\{m(x)\}$ есть БЧХ код длины $2^{n-1} - 1$ с исправлением одиночных ошибок, порожденный примитивным полиномом $g_1(x)$ степени $(n-1)$, т. е. если α — примитивный элемент поля $\mathrm{GF}\left(2^{n-1}\right)$, то $g_1(\alpha) = 0$. Рассмотрим далее код $\{s(x)\}$, порождающий полином которого имеет корни α, α^3 и 1. Ясно, что $\{s(x)\}$ является БЧХ кодом с минимальным весом 6 [10] и $\{s(x)\} \subset \subset \{m(x)\}$. Ясно также, что $\{s(x)\}$ существует только тогда,

[1]) Материал этого параграфа отчасти пересекается со статьей [12], поскольку настоящая работа является естественным ее обобщением.

когда $2^{n-1}-1 \geqslant 2(n-1)+1$, т. е. при $n \geqslant 4$; при $n=4$ полином $s(x)$ тождественно равен 0. Наконец, обозначим через $u(x)$ полином $\left(x^{2^{n-1}-1}+1\right)/(x+1)$.

При заданных полиномах $a(x)$ и $b(x)$ ($a(x) \in A_{n-1}$, $b(x) \in A_{n-1}$) и двоичном параметре i построим (2^n-1)-компонентные векторы над полем GF(2) вида

$$[a(x),\ i,\ b(x)].$$

При заданных $m(x) \in \{m(x)\}$ и $s(x) \in \{s(x)\}$ и произвольном i положим теперь $a(x)=m(x)$ и $b(x)=m(x)+(m(1)+i)u(x)+s(x)$. Получаем

$$\mathbf{v}=[m(x),\ i,\ m(x)+(m(1)+i)u(x)+s(x)]. \tag{1}$$

Лемма 1. *Векторы* $\mathbf{v}$, *задаваемые* (1), *образуют некоторый линейный код* $\mathscr{C}_n$.

Доказательство. Утверждение немедленно следует из проверки того, что $\mathscr{C}_n$ является группой относительно сложения над GF(2). Действительно: 1) $\mathscr{C}_n$ содержит аддитивную единицу [0, 0, 0], получаемую из (1) при $m(x)=0$, $s(x)=0$, $i=0$; 2) $\mathscr{C}_n$ замкнут относительно сложения, так как $\{m(x)\}$ и $\{s(x)\}$ являются групповыми кодами.

Лемма 2. *Минимальное расстояние между любыми двумя кодовыми словами из* $\mathscr{C}_n$ *равно по крайней мере* 6[1]).

Доказательство. Так как $\mathscr{C}_n$ является линейным кодом, его минимальное расстояние совпадает с минимальным весом W его ненулевых кодовых слов, к определению которого мы переходим. Предположим сначала, что $m(x)=0$. Если также $i=0$, то $W=W[s(x)] \geqslant 6$. Если $i=1$, то $W=1+W[u(x)+s(x)] \geqslant 1+W[u(x)]-\max W[s(x)]$. Мы знаем, что $W[u(x)]=2^{n-1}-1$ и что $\max W[s(x)]$ равен $2^{n-1}-6$ при $n>4$ и равен 0 при $n=4$ (так как $\max W[s(x)]$ равен максимальному *четному* весу кодовых слов БЧХ кода с исправлением двойных ошибок). Следовательно, имеем

$$W \geqslant 1+2^{n-1}-1-2^{n-1}+6=6.$$

Предположим теперь, что $m(x) \neq 0$. Если $m(x) \notin \{s(x)\}$, то $m^*(x)=m(x)+(m(1)+i)u(x)+s(x) \neq 0$ и $m^*(x) \in \{m(x)\}$. От-

[1]) В действительности минимальное расстояние между словами кода $\mathscr{C}_n$ равно 5. В частности, вектор (1) имеет вес 5, если $m(x)$ есть элемент БЧХ кода с исправлением двойных ошибок, имеющий вес 5, $i=0$ и $s(x)=m(x)+u(x)$. Однако это обстоятельство не влияет на справедливость основного результата статьи (теорема 1). — *Прим. перев.*

сюда следует, что $W \geqslant W[m(x)] + W[m^*(x)] \geqslant 3+3=6$, так как и $m(x)$, и $m^*(x)$ являются ненулевыми, а $\{m(x)\}$ имеет минимальный вес 3. Если же наоборот $m(x) \in \{s(x)\}$, то $W[m(x)] \geqslant 6$ и $W \geqslant W[m(x)] \geqslant 6$. Доказательство закончено.

Число информационных разрядов кода $\mathscr{C}_n$ легко получить, если учесть, что независимо выбираемые $m(x)$, $s(x)$ и i дают соответственно $(2^{n-1}-n)$, $(2^{n-1}-2n)$ и 1 информационных разрядов. Следовательно, $\mathscr{C}_n$ есть $(2^n-1,\ 2^n-3n+1)$ линейный код с минимальным расстоянием 6.

Рассмотрим далее полином $\varphi(x)=(x^{2^{n-1}-1}+1)/g_1(x)$, т. е. последовательность длины $2^{n-1}-1$, являющуюся последовательностью максимальной длины. Покажем сначала, что справедлива следующая лемма.

Лемма 3. *Существует число* s $(0 \leqslant s \leqslant 2^{n-1}-2)$, *такое, что* $(x^s\varphi(x))^2 = x^s\varphi(x)$.

Доказательство. Вычислим произведение $\varphi(x)\varphi(x)$. Так как $\varphi(x)$ не делится на $g_1(x)$, полином $\varphi^2(x)$ не равен нулю; кроме того, $\varphi^2(x)$ принадлежит коду, порожденному $\varphi(x)$, т. е.

$$\varphi^2(x) = x^r\varphi(x) \tag{2}$$

при некотором r, $0 \leqslant r \leqslant 2^{n-1}-2$. Если умножить (2) на x^{2s}, то получится $x^{2s}\varphi^2(x) = x^{r+2s}\varphi(x)$, т. е. $(x^s\varphi(x))^2 = x^s\varphi(x)\cdot x^{r+s}$. Отсюда вытекает утверждение леммы, если $x^{r+s}=1$, т. е. если $r+s=0 \pmod{2^{n-1}-1}$, или, что эквивалентно этому, $s=2^{n-1}-1-r \pmod{2^{n-1}-1}$. Доказательство закончено.

Положим [1]) $f(x) \triangleq x^s\varphi(x)$.

Ясно, что полином $q(x)=ax^j$ $(a=0,\ 1;\ j=0,\ 1,\ \ldots,\ 2^{n-1}-2)$ при $a=1$ является главным элементом смежного класса кода $\{m(x)\}$, имеющим наименьший вес. Построим теперь вектор $\mathbf{u}$ вида

$$\mathbf{u} = [q(x),\ 0,\ q(x)f(x)]. \tag{3}$$

Справедливы следующие леммы.

Лемма 4. *Полином* $q(x)+q(x)f(x)$ *принадлежит* $\{m(x)\}$.

Доказательство. Утверждение немедленно следует из леммы 3, так как

$$f(x)\{q(x)+q(x)f(x)\} = f(x)q(x)+f^2(x)q(x) = 0,$$

---

[1]) Знаком $\triangleq$ автор обозначает равенство по определению. — *Прим. перев.*

т. е. полином $q(x)+q(x)f(x)$, будучи ортогональным к $f(x)$, делится на $g_1(x)$.

Лемма 5. *Сумма двух векторов* $\mathbf{u}_1$ *и* $\mathbf{u}_2$ *вида* (3) *допускает при* $n\geqslant 4$ *представление*

$$\mathbf{u}_1+\mathbf{u}_2=\mathbf{v}+\mathbf{q}+\mathbf{p}, \tag{4}$$

где

$$\mathbf{v}=[m'(x),\ 0,\ m'(x)+m'(1)u(x)],\ m'(x)\in\{m(x)\},\ \text{т. е. } \mathbf{v}\in\mathscr{C}_n,$$

$$\mathbf{q}=[q(x),\ 0,\ q(x)], \tag{5}$$

$$\mathbf{p}=[0,\ 0,\ m''(x)],\ m''(x)\in\{m(x)\}. \tag{6}$$

Если $q(x)=0$, *то* $m'(x)=0$; *если* $q(x)\neq 0$, *то либо* $m'(x)=0$, *либо* $m'(x)$ *является трехчленом.*

Доказательство. Пусть $\mathbf{u}_1=[q_1(x),\ 0,\ q_1(x)f(x)]$ и $\mathbf{u}_2=[q_2(x),\ 0,\ q_2(x)f(x)]$. Имеем

$$\mathbf{u}_1+\mathbf{u}_2=[q_1(x)+q_2(x),\ 0,\ (q_1(x)+q_2(x))f(x)]. \tag{7}$$

Пусть $q(x)f(x)=(q_1(x)+q_2(x))f(x)$ и $m'(x)\triangleq q_1(x)+q_2(x)+q(x)$. Так как $(q(x)+q_1(x)+q_2(x))f(x)=0$, ясно, что $m'(x)\in\{m(x)\}$. Если $q_1(x)=q_2(x)$, то отсюда следует, что $q(x)=0$ и $m'(x)=0$. Если $q_1(x)\neq q_2(x)$, то либо $q(x)=q_i(x)$ $(i=1,\ 2)$, либо полиномы $q_1(x)$, $q_2(x)$ и $q(x)$ различны и не равны нулю; в первом случае $m'(x)=0$, в последнем случае $m'(x)$ является трехчленом.

Это дает возможность написать

$$q_1(x)+q_2(x)=m'(x)+q(x)$$

и переписать (7) следующим образом:

$$\mathbf{u}_1+\mathbf{u}_2=[m'(x),\ 0,\ m'(x)+m'(1)u(x)]+[q(x),\ 0,\ q(x)]+$$
$$+[0,\ 0,\ q(x)+q(x)f(x)+m'(x)+m'(1)u(x).$$

Теперь очевидно, что полином $m''(x)\triangleq q(x)+q(x)f(x)+m'(x)+m'(1)u(x)$ принадлежит $\{m(x)\}$, так как он является суммой полиномов, принадлежащих $\{m(x)\}$.

Лемма 6. *Для любого трехчлена* $m(x)=x^s+x^l+x^j\in\{m(x)\}$ *справедливо равенство*

$$m(\alpha^3)=\alpha^{3s}(\alpha^h+\alpha^{2h}),$$

где h *— некоторое число* $(0<h\leqslant 2^{n-1}-1)$.

Доказательство. Пусть $m(x) = x^s + x^i + x^j$, где s, i и j различны. Тогда $m(\alpha^3) = \alpha^{3s} + \alpha^{3i} + \alpha^{3j}$. Вспоминая, что $m(\alpha) = \alpha^s + \alpha^i + \alpha^j = 0$, имеем

$$\alpha^{3s} = (\alpha^i + \alpha^j)^3 = \alpha^{3i} + \alpha^{3j} + \alpha^i\alpha^j(\alpha^i + \alpha^j),$$

что эквивалентно $m(\alpha^3) = \alpha^i\alpha^j\alpha^s = \alpha^{3s}\alpha^{i-s}\alpha^{j-s}$. Но $\alpha^{j-s} = 1 + \alpha^{i-s}$, следовательно, полагая $i - s = h \neq 0$, получаем утверждение леммы.

Рассмотрим теперь матрицы

$$H_1 = [\alpha^{2^{n-1}-2}, \ldots, \alpha, 1],$$
$$H_3 = [(\alpha^3)^{2^{n-1}-2}, \ldots, \alpha^3, 1],$$
$$U = [1, \ldots, 1, 1].$$

Матрица $H = [H_1^T, H_3^T, U^T]^T$ является проверочной матрицей кода $\{s(x)\}$ (индекс T обозначает операцию транспонирования). Для заданного полинома $h(x)$ определим $h(x)H^T \triangleq [\beta_1, \beta_3, c]$ как *характеристики* $h(x)$. Для произвольного $s(x) \in \{s(x)\}$ имеем

$$W[h(x) + s(x)] \geqslant W[k(x)],$$

где $k(x)$ — элемент с минимальным весом из смежного класса по $\{s(x)\}$, которому принадлежит $h(x)$. Вычислим теперь характеристики некоторых полиномов, которые часто будут использоваться в дальнейшем:

$$\begin{gathered} q(x) = x^s, \qquad q(x)H^T = [\alpha^s, \alpha^{3s}, 1], \\ m(x) \in \{m(x)\},\ m(x)H^T = [0,\ m(\alpha^3),\ m(1)], \\ m''(x) \text{ (см. (6))},\ m''(x)H^T = [0,\ m'(\alpha^3) + \alpha^{3s},\ 1]. \end{gathered} \tag{8}$$

Первое соотношение очевидно. Второе следует из того, что $m(\alpha) = 0$, так как $m(x) \in \{m(x)\}$. Чтобы доказать третье из соотношений (8), вспомним, что $m''(x) = q(x) + q(x)f(x) + m'(1)u(x) + m'(x)$ и что $q(x)f(x)H^T = [\alpha^s, 0, 0]$, так как $f(x)$ делится на минимальную функцию $g_3(x)$ от α^3 и на $(x + 1)$; а также примем во внимание, что $m'(1)u(x)H^T = [0, 0, m'(1)]$, так как $u(x)$ делится на $g_3(x)$ и минимальную функцию $g_1(x)$ от α.

Лемма 7. *Для* $m''(x) = q(x) + q(x)f(x) + m'(1)u(x) + m'(x)$ *и произвольного* $s(x) \in \{s(x)\}$ *справедливо*

$$W[m''(x) + q(x) + s(x)] \geqslant \begin{cases} 4 \text{ при четном } n, \\ 2 \text{ при нечетном } n. \end{cases}$$

Доказательство. Характеристики $(m''(x)+q(x))$ суть $[\alpha^s, m'(\alpha^3), 0]$ (см. (8)). Далее, $W[m''(x)+q(x)+s(x)]$ есть минимальное число столбцов H, которые прибавляются к $[\alpha^s, m'(\alpha^3), 0]^T$. Так как $c=0$, это число четно. Предположим, что имеются два элемента x_1 и x_2 из $\mathrm{GF}(2^{n-1})$, которые удовлетворяют уравнениям

$$x_1+x_2=\alpha^s,$$
$$x_1^3+x_2^3=m'(\alpha^3).$$

Учитывая, что $\alpha^s\neq 0$, сделаем подстановку $y_1=(x_1/\alpha^s)$, $y_2=(x_2/\alpha^s)$. После простых преобразований убеждаемся, что y_1 и y_2 являются решениями единственного уравнения

$$y^2+y+1+m'(\alpha^3)/\alpha^{3s}=0.$$

Так как либо $m'(x)=0$, либо $m'(x)$ является трехчленом (лемма 5), то лемма 6 дает $m'(\alpha^3)/\alpha^{3s}=\alpha^h+\alpha^{2h}$ $(0\leqslant h\leqslant 2^{n-1}-2)$; предыдущее уравнение преобразуется в уравнение

$$(y+\alpha^h)^2+(y+\alpha^h)+1=0,$$

которое при $z=y+\alpha^h$ эквивалентно уравнению

$$z^2+z+1=0. \tag{9}$$

Но решениями уравнения (9) являются первообразные корни третьей степени из единицы, в силу чего уравнение (9) имеет решения в $\mathrm{GF}(2^{n-1})$ лишь при нечетном n. Доказательство закончено.

Построим теперь (2^n-1)-компонентные векторы следующего вида:

$$\mathbf{w}=[m(x)+q(x),\ i,\ m(x)+q(x)f(x)+(m(1)+i)u(x)+s(x)], \tag{10}$$

где $m(x)$, $q(x)$, i, $s(x)$ выбираются независимо и дают соответственно $(2^{n-1}-n)$, $(n-1)$, 1, $(2^{n-1}-2n)$ информационных разрядов, так что общее число информациоиных разрядов равно (2^n-2n). Векторы $\mathbf{w}$ образуют некоторый $(2^n-1,\ 2^n-2n)$ код $\mathscr{K}_n$. Общий вектор $\mathbf{w}$ можно разложить следующим образом:

$$\mathbf{w}=\mathbf{v}+\mathbf{u}, \tag{11}$$

где $\mathbf{v}$ и $\mathbf{u}$ определены соотношениями (1) и (3) соответственно. Пусть $\mathbf{w}_1=\mathbf{v}_1+\mathbf{u}_1$ и $\mathbf{w}_2=\mathbf{v}_2+\mathbf{u}_2$ — два различных кодовых слова из $\mathscr{K}_n$. Используя соотношение (4) (лемма 5), имеем

$$\mathbf{w}_1+\mathbf{w}_2=(\mathbf{v}_1+\mathbf{v}_2)+(\mathbf{u}_1+\mathbf{u}_2)=\mathbf{v}_1+\mathbf{v}_2+\mathbf{v}+\mathbf{q}+\mathbf{p}$$

или

$$\mathbf{w}_1+\mathbf{w}_2=\mathbf{v}'+\mathbf{q}+\mathbf{p}, \tag{12}$$

где $\mathbf{v}' \triangleq \mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}$. Ясно, что $\mathbf{v}'$ является некоторым произвольным элементом кода $\mathscr{C}_n$, а $\mathbf{q}+\mathbf{p}$ можно представить следующим образом:

$$\begin{aligned}\mathbf{q}+\mathbf{p} &= [q(x),\ 0,\ q(x)f(x)+m'(x)+m'(1)u(x)] = \\ &= [q(x),\ 0,\ q(x)f(x)] + [0,\ 0,\ m'(x)+m'(1)u(x)] = \\ &= \mathbf{u}' + [0,\ 0,\ m'(x)+m'(1)u(x)].\end{aligned}$$

В случае, когда $m'(x) \neq 0$, вспомним, что $m'(x) \notin \{s(x)\}$ (лемма 5), т. е. $[0,\ 0,\ m'(x)+m'(1)u(x)] \notin \mathscr{C}_n$. Следовательно, код $\mathscr{K}_n$ является нелинейным кодом. Кроме того, каждый ненулевой вектор $\mathbf{u}$ в (11) определяет некоторый смежный класс кода $\mathscr{C}_n$, так как отличный от нуля полином $q(x)$ определяет некоторый смежный класс кода $\{m(x)\}$. Следовательно, код $\mathscr{K}_n$ можно рассматривать как теоретико-множественное объединение кода $\mathscr{C}_n$ и некоторого подмножества его смежных классов, мощность которого равна $2^{n-1}-1$.

Пусть W обозначает вес суммы $(\mathbf{w}_1+\mathbf{w}_2)$. Мы можем теперь доказать основной результат этой статьи.

Теорема 1. *При четном $n \geqslant 4$ код $\mathscr{K}_n$ является нелинейным $(2^n-1,\ 2^n-2n)$ кодом с минимальным расстоянием* 5.

Доказательство. Если $q(x)=0$, то $\mathbf{w}_1+\mathbf{w}_2 \in \mathscr{C}_n$ и $W \geqslant 6$ по лемме 2[1]). Предположим теперь, что $q(x)=x^s$ $(0 \leqslant s \leqslant 2^{n-1}-2)$. В общем случае имеем

$$W = i + W[m(x)+q(x)] + W[m(x)+(m(1)+i)u(x)+ \\ + s(x)+m''(x)+q(x)]. \quad (13)$$

В зависимости от значений $m(1)$ и i мы различаем три случая:

1) $m(1)=0$, $i=1$. Соотношение (13) преобразуется следующим образом:

$$W \geqslant 1 + W[m(x)+q(x)] + W[u(x)+s(x)+m''(x)] - \\ - W[m(x)+q(x)] = 1 + W[u(x)+s(x)+m''(x)].$$

Из соотношения (8) и $u(x)H^T = [0,\ 0,\ 1]$ получаем

$$(u(x)+m''(x))H^T = [\beta_1,\ \beta_2,\ c] = [0,\ m'(\alpha^3)+\alpha^{3s},\ 0].$$

Но, согласно леммам 5 и 6, имеем $m'(\alpha^3) = \alpha^{3s}(\alpha^k+\alpha^{2k})$, $0 \leqslant k \leqslant 2^{n-1}-2$. Следовательно, $m'(\alpha^3)+\alpha^{3s} = \alpha^{3s}(1+\alpha^k+\alpha^{2k}) \neq 0$ в $GF(2^{n-1})$ при четном n. Отсюда следует, что $W[u(x)+s(x)+m''(x)] \neq 0$. Кроме того, $W[u(x)+s(x)+m''(x)]$ есть четное число, превышающее 3, так как $c=0$ и $\beta_1=0$ (H_1 является

[1]) См. примечание на стр 20. — *Прим. перев.*

проверочной матрицей кода с исправлением одиночных ошибок). Мы заключаем, что $W \geqslant 1+4=5$.

2) $m(1)=0$, $i=0$. Если $m(x) \neq 0$, то соотношение (13) дает

$$W \geqslant W[m(x)] + W[m(x)+m''(x)+s(x)] - 2W[q(x)].$$

Соотношения (8) дают

$$(m(x)+m''(x))\,H^T = [\beta_1,\ \beta_3,\ c] = [0,\ m(\alpha^3)+m'(\alpha^3)+\alpha^{3s},\ 1],$$

т. е. вес $W[m(x)+m''(x)+s(x)]$ — нечетное число ($c=1$), не меньшее 3 ($\beta_1=0$). Кроме того, соотношения $m(x)\neq 0$ и $m(1)=0$ означают, что $W[m(x)] \geqslant 4$, откуда $W \geqslant 4+3-2=5$. Если $m(x)=0$, то соотношение (13) преобразуется к виду

$$W \geqslant W[q(x)] + W[m''(x)+q(x)+s(x)].$$

Из леммы 7 имеем, что $W[m''(x)+q(x)+s(x)] \geqslant 4$ при четном n, откуда $W \geqslant 1+4=5$.

3) $m(1)=1$. В этом случае число $W[m(x)+q(x)]$ четно и не менее 2. Предположим сначала, что $W[m(x)+q(x)]=2$. Это означает, что $m(x)=x^s+x^i+x^j$ (s, i, j различны); отсюда по лемме 6 имеем $m(\alpha^3)=\alpha^{3s}(\alpha^h+\alpha^{2h})$ при некотором h, $1 \leqslant h \leqslant 2^{n-1}-2$. Соотношение (13) дает

$$W = i+2+W[m(x)+(1+i)\,u(x)+m''(x)+q(x)+s(x)].$$

Для простоты положим $k(x) \triangleq m(x)+(1+i)\,u(x)+m''(x)+q(x)+s(x)$. С помощью соотношений (8) легко получаются следующие характеристики $k(x)$:

$$k(x)\,H^T = [\beta_1,\ \beta_3,\ c] = [\alpha^s,\ m(\alpha^3)+m'(\alpha^3),\ i].$$

Так как $m'(\alpha^3)=\alpha^{3s}(\alpha^k+\alpha^{2k})$ ($0 \leqslant k \leqslant 2^{n-1}-2$), отсюда следует, что $m(\alpha^3)+m'(\alpha^3)=\alpha^{3s}(\alpha^r+\alpha^{2r})$ при $r=h+k$. Следовательно, по лемме 7 имеем $W[k(x)]>2$ при четном n, т. е. $W[k(x)] \geqslant 4-i$ (так как i есть четность $W[k(x)]$). Мы заключаем, что

$$W \geqslant i+2+4-i=6.$$

Наконец, предположим, что $W[m(x)+q(x)] \geqslant 4$. Так как $\beta_1=\alpha^s$, находим $W[k(x)] \geqslant 1$, откуда следует $W \geqslant i+4+1 = i+5$. Доказательство окончено.

Замечание. Интересно рассмотреть задачу расширения примененного метода с целью построения кодов $\mathcal{K}_n$ для других значений числа исправляемых ошибок, а именно для $t=1$ и $t>2$.

Имеются две различные схемы, которые кажутся перспективными для обобщений. Рассмотрим сначала соотношение (10),

которое описывает код $\mathscr{K}_n$ с исправлением двойных ошибок, т. е.

$$\mathbf{w} = [m(x) + q(x),\ i,\ m(x) + (m(1) + i)\,u(x) + s(x) + q(x)f(x)].$$

Здесь $\mathscr{K}_n$ строится с помощью двух кодов $\{m(x)\}$ и $\{s(x)\}$, где $\{s(x)\} \subset \{m(x)\}$. В частности, если α — первообразный корень в $\mathrm{GF}(2^{n-1})$, то $\{m(x)\}$ характеризуется корнем α, а $\{s(x)\}$ — корнями 1, α, α^3. Поэтому имеются два потенциальных обобщения на случай t-кратных ошибок.

А. $\{m(x)\}$ имеет корень α, а $\{s(x)\}$ имеет корни 1, α, α^3, ..., ..., α^{2t-1}.

В. $\{m(x)\}$ имеет корни α, α^3, ..., α^{2t-3}, а $\{s(x)\}$ имеет корни 1, α, α^3, ..., α^{2t-1}.

При $t = 1$ обе схемы приемлемы и порождают, как легко показать, одни и те же коды. В частности, $m(x)$ в схеме В является общим элементом A_{n-1}, а $\{s(x)\}$ имеет $(x+1)g_1(x)$ в качестве своего порождающего полинома. Получающийся код $\mathscr{K}_n^{(1)}$ состоит из векторов

$$\mathbf{w}_B = [m(x),\ i,\ m(x) + (m(1) + i)\,u(x) + s(x)]. \qquad (14)$$

Удивительно, что код $\mathscr{K}_n^{(1)}$ является линейным кодом, как это легко видеть из (14). Кроме того, можно показать, что он совпадает с кодом Васильева [2]. В действительности (14) можно выразить следующим образом:

$$\mathbf{w}_B = [m(x),\ i,\ m(x) + p(x)],$$

где

$$p(x) = (m(1) + i)\,u(x) + s(x).$$

Если теперь наложить условие, что $p(x)$ принадлежит коду, порожденному полиномом $g_1(x)$, то это соотношение становится уравнением относительно неизвестных $s(x)$ и i, которое всегда можно разрешить, если

$$i = \text{четность } W[m(x)] + \text{четность } W[p(x)].$$

Тем самым получается линейный код Васильева (эквивалентный коду Хемминга).

При $t > 2$ вопрос о том, дает ли какая-либо из двух описанных схем приемлемое обобщение, остается полностью открытым.

3. ВИД ИЗБЫТОЧНЫХ ФУНКЦИЙ

Рассмотрим выражение (10) общего вектора кода $\mathscr{K}_n$, т. е.

$$\mathbf{w} = [m(x) + q(x),\ i,\ m(x) + (m(1) + i)\,u(x) + s(x) + q(x)f(x)].$$

Легко видеть, что $\mathcal{H}_n$ можно кодировать как систематический код, т. е. (2^n-2n) двоичных информационных разрядов можно произвольно приписать в фиксированных позициях, а остальные $(2n-1)$ избыточных разрядов можно подсчитать как функции информационных разрядов. В этом параграфе мы исследуем природу этих функций. Для удобства представим теперь $\mathbf{w}$ в виде

$$\mathbf{w}=\left[i^{(0)}_{2^n-1-2},\ \ldots,\ i^{(0)}_1,\ i^{(0)}_0,\ i,\ i^{(1)}_{2^n-1-2},\ \ldots,\ i^{(1)}_{2n-1},\ p_{2n-2},\ \ldots,\ p_1,\ p_0\right],$$

где буквы i и p обозначают информационные и избыточные разряды соответственно.

Предположим сначала, что $s(x)=0$. Тогда левые 2^{n-1} разрядов $\left[i^{(0)}_{2^n-1-2},\ \ldots,\ i\right]$ полностью определяют $2^{n-1}-1$ правых разрядов. Обозначим последние через $\left[\varphi_{2^n-1-2},\ \ldots,\ \varphi_0\right]$ и проанализируем их зависимость от предыдущего множества разрядов. Пусть

$$i(x)\triangleq\sum i^{(0)}_j x^j,\quad q(x)\,f(x)\triangleq c(x)=\sum c_j x^j,\quad f(x)=\sum f_j x^j,$$
$$m(x)=\sum m_j x^j,$$

где все суммы берутся по $j=0,\ 1,\ \ldots,\ 2^{n-1}-2$. Если $q(x)=0$, то $c_j=0$ при каждом j. Если $q(x)=x^s$, то, согласно свойству однозначности последовательности максимальной длины [10], имеем $c_{s+b}c_{s+b-1}\cdots c_{s+b-n+2}=1$ и $c_{j+b}\cdots c_{j+b-n+2}=0$ при $j\neq s$, где число b таково, что $f_b f_{b-1}\cdots f_{b-n+2}=1$. Легко видеть, что

$$q(x)=\sum c_{j+b}\cdots c_{j+b-n+2}x^j,\quad m_j=i^{(0)}_j+c_{j+b}\cdots c_{j+b-n+2}$$

и

$$\varphi_j=i^{(0)}_j+c_{j+b}\cdots c_{j+b-n+2}+i+\sum_k\left(i^{(0)}_k+c_{k+b}\cdots c_{k+b-n+2}\right)+c_j$$

или после перегруппировки членов

$$\varphi_j=\left\{\sum_{k\neq j}i^{(0)}_k+i+c_j\right\}+\left\{\sum_{h\neq j+b}c_h\cdots c_{h-n+2}\right\}. \tag{15}$$

Вспомним теперь, что поскольку $c(x)=q(x)f(x)=i(x)f(x)$, то $c_j=\sum f_{j-k}i^{(0)}_k$ есть линейная функция от переменных $i^{(0)}_0,\ i^{(0)}_1,\ \ldots$ $\ldots,\ i^{(0)}_{2^n-1-2}$. В частности, так как $f(x)$ есть последовательность максимальной длины, то для различных r и s существует число t, такое, что $c_r+c_s=c_t$. Если $q_1(x)$ есть трехчлен, то при $s=(r+n-1)$ число t удовлетворяет неравенствам $r<t<r+n-1$. Следовательно, имеем

$$c_h c_{h-1}\cdots c_{h-n+2}+c_{h-1}c_h\cdots c_{h-n+1}=$$
$$=c_{h-1}\cdots c_{h-n+2}(c_h+c_{h-n+1})=c_{h-1}\cdots c_{h-n+2}$$

Отсюда следует, что в последнем члене выражения (15), который является суммой $(2^{n-1}-2)$ произведений, каждая пара последовательных произведений из $(n-1)$ множителей преобразуется в одно произведение из $(n-2)$ множителей, так что получается $(2^{n-2}-1)$ произведений. В заключение мы получаем выражение

$$\varphi_j=\left\{i+i_j^{(0)}+\sum_{k=0}^{2^{n-1}-2}(1+f_{j-k})\,i_k^{(0)}\right\}+\sum_{h=0}^{2^{n-2}-2}\prod_{s=0}^{n-3}c_{j+b+1+2h-s}, \quad (16)$$

которое показывает, что $\varphi_j\left(i_{2^{n-1}-2}^{(0)}, \ldots, i_0^{(0)}, i\right)$ является суммой некоторой строго линейной функции и некоторой нелинейной функции степени не выше $(n-2)$. В качестве проверки заметим, что для (15,8) кода $(n=4)$ последняя функция является квадратичной.

Пусть h_{ij} — общий вход проверочной матрицы H^* кода $\{s(x)\}$, записанной в систематическом виде, т. е. $(2n-1)$ правых столбцов H^* образуют единичную матрицу и индекс j пробегает значения справа налево. Тогда соотношения

$$p_i=\varphi_i+\sum_{j=2n-1}^{2^{n-1}-2}h_{ij}\left(i_j^{(1)}+\varphi_j\right) \quad (i=0, 1, \ldots, 2n-2) \quad (17)$$

дают искомые избыточные функции.

Выражения (15) и (17) дают возможность предложить очень простую схему кодирования. Действительно, φ_j является циклической функцией от своих аргументов. Следовательно, ее можно реализовать с помощью нелинейного сверточного кодера, состоящего из циклического регистра сдвига и комбинационной схемы, реализующей $\varphi=\varphi_{2^{n-1}-2}$ (см. рис. 1). Полный кодер состоит из трех регистров сдвига PC1, PC2, PC3, имеющих 1, $(2^{n-1}-1)$ и $(2n-1)$ разрядов соответственно. Операция разбивается на четыре фазы G_1, G_2, G_3, G_4, продолжительность которых равна 1, $(2^{n-1}-1)$, $(2^{n-1}-2n)$ и $(2n-1)$ тактам соответственно. Указанные вентили проводят, когда приложенные сигналы равны 1. Все регистры первоначально заполнены нулями. Информационные разряды посылаются в виде последовательности $i, i_{2^{n-1}-2}^{(0)}, \ldots, i_0^{(0)}, i_{2^{n-1}-2}^{(1)}, \ldots, i_{2n-1}^{(1)}$ по одному разряду на каждый такт. Тогда в течение фазы G_1 разряд i посылается на PC1 и в течение фазы G_2 разряды $i_{2^{n-1}-2}^{(0)}, \ldots, i_0^{(0)}$ посылаются на PC2 (при этом они одновременно посылаются

на выход). Как показано, РС1 и РС2 являются регистрами с циркуляцией. В течение фазы G_3 функции $\varphi_j + i_j^{(1)}$ появляются в точке A для посылки в регистр РС3, который является регистром сдвига с обратной связью, осуществляющий деление полинома на $(x+1)g_1(x)g_3(x)$ (см. [10]). Тогда в конце фазы G_3 регистр РС3 содержит проверки на четность $\sum_i h_{ij}(i_i^{(1)} + \varphi_j)$.

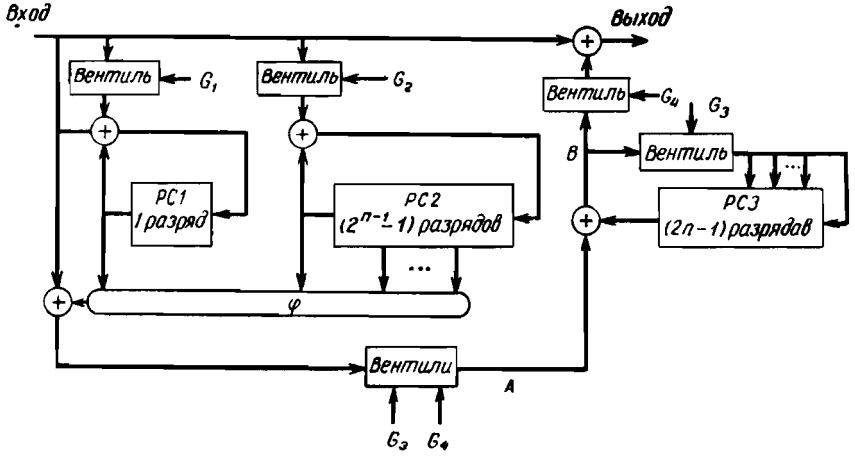


Рис. 1. Кодер для кода $\mathcal{K}_n$.

В течение фазы G_4 вход равен 0 и в точке B образуются функции p_i и посылаются на выход. Следовательно, вычисление избыточных разрядов происходит не дольше, чем последовательная передача информационных разрядов.

4. ОПТИМАЛЬНОСТЬ КОДОВ $\mathcal{K}_n$

Код $\mathcal{K}_n$ является $(2^n-1,\ 2^n-2n)$ кодом с исправлением двойных ошибок. Он содержит на один информационный разряд больше, чем соответствующий линейный код, т. е. БЧХ код той же самой длины с исправлением двойных ошибок (который является $(2^n-1,\ 2^n-2n-1)$ кодом).

В этом параграфе мы докажем более сильное утверждение, а именно, что код $\mathcal{K}_n$ имеет наибольшее число кодовых слов для своей длины и своего минимального расстояния, так как достигается граница Джонсона [7] для $A(N, d)$ при $N = 2^n - 1$

(n четно) и $d=5$ [1]). Граница Джонсона при $d=2t+1$ задается следующим образом:

$$A(N,d)\leqslant\frac{2^N}{\sum_{l=0}^{t}C_N^l+\dfrac{C_N^{t+1}-C_d^tR(N,d,t)}{[N/(t+1)]}}, \tag{18}$$

где $[a]$ есть целая часть числа a, а $R(N, d, t)$ удовлетворяет верхней оценке

$$R(N,d,t)\leqslant\left[\frac{N}{d}\left[\frac{N-1}{d-1}\left[\ldots\left[\frac{N-t}{d-1}\right]\ldots\right]\right]\right]. \tag{19}$$

При $N=2^n-1$ и $t=2$ соотношения (18) и (19) приобретают вид

$$A(2^n-1,5)\leqslant\frac{2^{2^n-1}}{1+C_{2^n-1}^1+C_{2^n-1}^2+\dfrac{C_{2^n-1}^3-C_5^2R(2^n-1,\,5,\,2)}{[(2^n-1)/3]}}, \tag{20}$$

$$R(2^n-1,\,5,\,2)\leqslant\left[\frac{2^n-1}{5}\left[\frac{2^n-2}{4}\left[\frac{2^n-3}{3}\right]\right]\right]. \tag{21}$$

Рассмотрим соотношение (21). При четном n число (2^n-4) делится на 3, так что $[(2^n-3)/3]=(2^n-4)/3$. Кроме того, (2^n-4) делится на 4. Мы должны теперь показать, что $(2^n-1)(2^n-2)\times$ $\times(2^{n-2}-1)$ делится на 5. Это немедленно следует из того замечания, что вычеты числа 2^n (n четно) по mod 5, чередуясь, равны 1 и 4, т. е. вычеты чисел (2^n-1), чередуясь, равны 0 и 3. Поскольку число $(2^n-1)(2^n-2)(2^{n-2}-1)$ содержит две последовательные четные степени 2, имеет место неравенство

$$R(2^n-1,\,5,\,2)\leqslant\frac{(2^n-1)(2^n-2)(2^n-4)}{60},$$

из которого при четном n легко получить соотношение

$$\frac{C_{2^n-1}^3-C_5^2R(2^n-1,5,2)}{[(2^n-1)/3]}\geqslant 3\,\frac{((2^n-1)(2^n-2))/6}{(2^n-1)}=2^{n-1}-1. \tag{22}$$

На основании этого мы заключаем, что при n четном имеем

$$A(2^n-1,5)\leqslant\frac{2^{2^n-1}}{2^{2n-1}-2^{n-1}+1+(2^{n-1}-1)}=2^{2^n-2n}.$$

---

[1]) Дж. П. Робинсон первый заметил, что $A(2^n-1,5)$ (n четно) есть степень 2. Еще ранее автор сформулировал предложение, основанное на довольно туманных геометрических аргументах, что нелинейные коды длины (2^n-1) и с расстоянием 5, аналогичные (15,8) коду, существуют лишь при четных n (частные сообщения, январь и март 1968 г.).

Последнее число в точности совпадает с числом кодовых слов кода $\mathcal{K}_n$. Ясно, что при нечетном n отношение в (22) строго больше, чем $(2^{n-1}-1)$, так как $[(2^n-3)/3]=(2^n-5)/3$. Это обстоятельство также показывает, что при нечетных n коды $\mathcal{K}_n$ не существуют.

5. ДЕКОДИРОВАНИЕ КОДА $\mathcal{K}_n$

В этом параграфе мы покажем, что декодирование кода $\mathcal{K}_n$ можно легко выполнить с помощью вычисления и исследования величин, подобных синдрому.

Вектором $\mathbf{e}=[e_0(x),\ e,\ e_1(x)]$ мы представим общий вид ошибки, где $e_j(x)\in A_{n-1}$ и e есть двоичный параметр. Используя свойства расстояния кода $\mathcal{K}_n$, заключаем, что ошибка $\mathbf{e}$ может быть исправлена при выполнении следующего условия:

$$W[e_0(x)]+W[e_1(x)]+e\leqslant 2. \tag{23}$$

В общем случае полученный вектор есть $\mathbf{r}=[r_0(x),\ r,\ r_1(x)]=\mathbf{w}+\mathbf{e}$, где $\mathbf{w}\in\mathcal{K}_n$. Вычислим теперь следующие функции:

$$\begin{aligned}
\sigma_0&\triangleq r_0(x)\,H_1^T,\\
\sigma_1&\triangleq r_1(x)\,H_1^T,\\
\sigma&\triangleq (r_0(x)+r_1(x))\,H_3^T,\\
d&\triangleq r+r_1(x)\,U^T.
\end{aligned}$$

Так как $r_0(x)=m(x)+q(x)+e_0(x)$, а $m(x)\,H_1^T=0$, то при $q(x)=bx^s$ имеем $\sigma_0=b\alpha^s+e_0(\alpha)$. Аналогично из $r_1(x)=m(x)+(m(1)+i)\,u(x)+q(x)f(x)+s(x)+e_1(x)$ и $u(x)\,H_1^T=0$, $s(x)\,H_1^T=0$, $q(x)f(x)\,H_1^T=b\alpha^s$ получаем, что $\sigma_1=b\alpha^s+e_1(\alpha)$. Из $r_0(x)+r_1(x)=q(x)+q(x)f(x)+s(x)+e_0(x)+e_1(x)+(m(1)+i)\,u(x)$, вспоминая, что $s(x)\,H_3^T=0$, $f(x)\,H_3^T=0$, $u(x)\,H_3^T=0$, находим, что $\sigma=b\alpha^{3s}+e_1(\alpha^3)+e_0(\alpha^3)$. Наконец, так как числа $W[q(x)\times f(x)]$, $W[s(x)]$, $W[m(x)+m(1)\,u(x)]$ четны, то $d=r+i+e_1(1)=e+e_1(1)$. Все это можно выразить следующим образом:

$$\begin{aligned}
\sigma_0&=b\alpha^s+e_0(\alpha),\\
\sigma_1&=b\alpha^s+e_1(\alpha),\\
\sigma&=b\alpha^{3s}+e_0(\alpha^3)+e_1(\alpha^3),\\
d&=e+e_1(1).
\end{aligned}\tag{24}$$

Четверку $\Sigma\equiv(\sigma_0,\ \sigma_1,\ \sigma,\ d)$ удобно называть *синдромом* вектора r.

Теперь мы приведем лемму, которая основана на достаточно известных результатах из теории конечных полей [1]).

Лемма 8. *Множество* Θ *всех* $\theta \in \mathrm{GF}(2^{n-1})$, *для которых уравнение* $y^2 + y + \theta = 0$ *имеет решения в* $\mathrm{GF}(2^{n-1})$, *образует векторное пространство размерности* $(n-2)$, *определяемое четными линейными комбинациями элементов нормального базиса* $\beta, \beta^2, \beta^4, \ldots, \beta^{2^{n-2}}$ *поля* $\mathrm{GF}(2^{n-1})$.

Доказательство. Известно (см., например, [1]), что существуют базисы поля $\mathrm{GF}(2^{n-1})$, состоящие из полных множеств сопряженных элементов (нормальные базисы). Пусть $\beta, \beta^2, \ldots, \beta^{2^{n-2}}$ — одно из таких множеств линейно независимых сопряженных элементов. Тогда каждый элемент $\gamma \in \mathrm{GF}(2^{n-1})$ однозначно выражается следующим образом:

$$\gamma = c_0\beta + c_1\beta^2 + \ldots + c_{n-2}\beta^{2^{n-2}} \quad (c_j \in \mathrm{GF}(2)).$$

Так как

$$\beta^{2^{n-1}} = \beta,$$

имеем

$$\gamma^2 = c_{n-2}\beta + c_0\beta^2 + \ldots + c_{n-3}\beta^{2^{n-2}}$$

и

$$\gamma^2 + \gamma = d_0\beta + d_1\beta^2 + \ldots + d_{n-2}\beta^{2^{n-2}} \quad (d_j \in \mathrm{GF}(2)), \qquad (25)$$

где $d_j = c_j + c_{j-1}$ (индексы вычисляются по $\bmod n-1$). Но правая часть (25) есть общий элемент Θ. Действительно, предположим, что $d_0, d_1, \ldots, d_{n-2}$ заданы. Тогда мы имеем $c_{n-2} = d_0 + c_0 = d_0 + d_1 + c_1 = \ldots = d_0 + \ldots + d_{n-2} + c_{n-2}$, следовательно,

$$d_0 + d_1 + \ldots + d_{n-2} = 0,$$

т. е. число ненулевых d_j четно. Доказательство закончено.

Эта лемма дает правило для проверки того, является ли $\gamma \in \mathrm{GF}(2^{n-1})$ элементом Θ. Фактически мы должны сначала найти нормальный базис $\beta, \beta^2, \ldots, \beta^{2^{n-2}}$ поля $\mathrm{GF}(2^{n-1})$ (см., например, [4]). Пусть через $\boldsymbol{\gamma}$ обозначен вектор-столбец представления над полем $\mathrm{GF}(2)$ элемента $\gamma \in \mathrm{GF}(2^{n-1})$ относительно базиса $1, \alpha, \ldots, \alpha^{n-2}$, и пусть $M \triangleq [\beta, \ldots, \beta^{2^{n-2}}]$ —

---

[1]) Рассуждения, приведенные ниже, по существу заимствованы из работы [1]. Весьма сходная теорема была доказана Берлекампом и другими ([3], теорема 1). Довольно подробное изложение рассматриваемого вопроса имеется в статье [4], в которой содержится также обобщение этой леммы. Так как приведенное здесь утверждение тесно связано с дальнейшими результатами, лемма и ее доказательство приводятся полностью.

невырожденная $(n-1)\times(n-1)$ матрица. Тогда $\boldsymbol{\gamma}$ связан с представлением $[d_0, \ldots, d_{n-2}]$ элемента γ относительно базиса $\beta, \beta^2, \ldots, \beta^{2^{n-2}}$ следующим образом:

$$\boldsymbol{\gamma} = M\,[d_0, \ldots, d_{n-2}]^T,$$

т. е. $M^{-1}\boldsymbol{\gamma} = [d_0, \ldots, d_{n-2}]^T$. Умножая слева обе стороны равенства на вектор-строку $\mathbf{u} = [1, 1, \ldots, 1]$, получаем условие

$$\mathbf{u}\cdot[d_0, \ldots, d_{n-2}]^T = \begin{cases} 0, & \text{если } \gamma \in \Theta, \\ 1, & \text{если } \gamma \notin \Theta. \end{cases}$$

которое, если обозначить через $\boldsymbol{\lambda}^T$ сумму строк M^{-1}, переходит в условие

$$\boldsymbol{\lambda}^T\cdot\boldsymbol{\gamma} = \begin{cases} 0, & \text{если } \gamma \in \Theta, \\ 1, & \text{если } \gamma \notin \Theta. \end{cases} \tag{26}$$

Следующая лемма проливает некоторый свет на взаимосвязь расстояний между общим вектором $\mathbf{r}$ и элементами $\mathbf{w}$ кода $\mathcal{K}_n$.

Лемма 9. *Для любого заданного* $\mathbf{r} = [r_0(x), r, r_1(x)]$ *существует* $\mathbf{w} \in \mathcal{K}_n$, *такое, что* $\mathbf{r} + \mathbf{w} = [0, e, e(x)]$, *причем* $W[e(x)] \leqslant 3$.

Доказательство. Пусть $\{t(x)\}$ есть БЧХ код с исправлением двойных ошибок, порожденный полиномом $g_1(x)\,g_3(x)$. Разложим $r_0(x)$ следующим образом: $r_0(x) = m_0(x) + q_0(x)$ и образуем $r_1^*(x) = r_1(x) + m_0(x) + (m_0(1) + r)\,u(x) + q_0(x)\,f(x)$. Далее $r_1^*(x)$ представим как $r_1^*(x) = t(x) + e(x)$, где $t(x) \in \{t(x)\}$ и $e(x)$ есть элемент с минимальным весом смежного класса по $\{t(x)\}$. Известно [5], что $W[e(x)] \leqslant 3$. Легко также проверить, что $(t(x) + t(1)\,u(x)) \in \{s(x)\}$. Образуем далее кодовое слово

$$\mathbf{w} = [m_0(x) + q_0(x),\ r + t(1),\ m_0(x) + (m_0(1) + r + t(1))\,u(x) + q_0(x)\,f(x) + t(x) + t(1)\,u(x)] = [r_0(x),\ r + t(1),\ r_1(x) + e(x)].$$

Полагая $t(1) = e$, получаем, что $\mathbf{r} + \mathbf{w} = [0, e, e(x)]$. Доказательство закончено.

В дальнейшем индекс j у σ_j или $e_j(x)$ следует рассматривать по mod 2. Положим $\rho \triangleq \sigma + (\sigma_0 + \sigma_1)^3$ и докажем следующую основную лемму.

Лемма 10. *Условия* $\rho + \sigma_j^3 = 0$ $(j = 0$ *и* $1)$, $d = 0$ *выполняются тогда и только тогда, когда* $\mathbf{r} \in \mathcal{K}_n$, *т. е. они характеризуют код* $\mathcal{K}_n$.

Доказательство. В силу леммы 9 без ограничения общности можно предположить, что разность между $\mathbf{r}$ и некоторым $\mathbf{w} \in \mathcal{K}_n$ имеет вид $[0, e, e(x)]$, $W[e(x)] \leqslant 3$. Тогда соотношения (24) приобретают вид

$$\begin{aligned} \sigma_0 &= b\alpha^s, \\ \sigma_1 &= b\alpha^s + e(\alpha), \\ \sigma &= b\alpha^{3s} + e(\alpha^3), \\ d &= e + e(1). \end{aligned} \tag{27}$$

Прямое утверждение немедленно следует, если в (27) положить $e(x) = 0$, $e = 0$. Чтобы доказать обратное утверждение, предположим, что $\rho + \sigma_j^3 = 0$ $(j = 0, 1)$. Это влечет $\sigma_0^3 = \sigma_1^3$, и в силу единственности корня третьей степени в поле $\mathrm{GF}(2^{n-1})$ (число n четно) имеет место $\sigma_0 = \sigma_1$. Из (27) следует, что $e(\alpha) = 0$. Теперь имеем $\rho + \sigma_j^3 = \sigma + \sigma_j^3 = e(\alpha^3) = 0$. Так как $[H_1^T, H_3^T]$ является проверочной матрицей кода с исправлением двойных ошибок, а $W[e(x)] \leqslant 3$ (лемма 9), из равенства $e(\alpha) = e(\alpha^3) = 0$ заключаем, что $e(x) = 0$. Наконец, условие $d = 0$ дает, что $e = e(1) = 0$. Доказательство окончено.

Легко проверить, что условия $\rho + \sigma_j^3 = 0\,(j = 0, 1)$, $d = 0$ эквивалентны условиям

$$\sigma_0 = \sigma_1, \quad \sigma = \sigma_0^3 = \sigma_1^3, \quad d = 0, \tag{28}$$

которые также характеризуют код.

Перейдем к доказательству трех теорем (2.1, 2.2 и 2.3), которые устанавливают соответствие между множествами синдромов и множествами конфигураций исправимых ошибок. Утверждения и соответствующие доказательства почти идентичны. Необходимое условие („тогда“) доказывается путем проверки с помощью соотношений (24) того, что конфигурация ошибки определенного типа производит синдром определенного типа. Обращение („только тогда“) доказывается следующим образом: образуется вектор „коррекции“ $\mathbf{c} = [c_0(x), c, c_1(x)]$, который является функцией только синдрома Σ и такой, что $c + W[c_0(x)] + W[c_1(x)] \leqslant 2$; затем показывается, что $\mathbf{r} + \mathbf{c} \in \mathcal{K}_n$, поскольку синдром $\Sigma^* \equiv (\sigma_0^*, \sigma_1^*, \sigma^*, d^*)$, вычисленный для $(\mathbf{r} + \mathbf{c})$, т. е.

$$\begin{aligned} \sigma_j^* &= \sigma_j + c_j(\alpha) \quad (j = 0, 1), \\ \sigma^* &= \sigma + c_0(\alpha^3) + c_1(\alpha^3), \\ d^* &= d + c + c_1(1), \end{aligned} \tag{29}$$

удовлетворяет условиям леммы 10 (или эквивалентным условиям (28)); наконец, благодаря свойствам расстояния кода $\mathcal{K}_n$

устанавливается, что $\mathbf{e}=\mathbf{c}$ есть единственная конфигурация исправимой ошибки, которая может произвести $\mathbf{r}$. Ясно, что каждая из этих теорем дает также декодирующее правило, заключающееся в вычислении вектора $\mathbf{c}$ на основании Σ. После этих пояснений мы только наметим доказательство каждой из следующих теорем.

Теорема 2.1. *В случае исправимой ошибки* $\mathbf{e}$ *условие* $\rho+\sigma_j^3=0$ *выполняется лишь для одного из значений* $j=0,\ 1$ *тогда и только тогда, когда* $W[e_0(x)]+W[e_1(x)]=1$.

Доказательство. *„Тогда“*. Соотношения $e_j(x)=x^k$, $e_{j+1}(x)=0$ дают $\sigma_j=b\alpha^s+\alpha^k$, $\sigma_{j+1}=b\alpha^s$, $\sigma=b\alpha^{3s}+\alpha^{3k}$, следовательно, $\rho=b\alpha^{3s}$. Отсюда имеем $\rho+\sigma_{j+1}^3=0$ и

$$\rho+\sigma_j^3=\alpha^{3k}[(b\alpha^{s-k})^2+b\alpha^{s-k}+1]\neq 0,$$

так как $z^2+z+1\neq 0$ при любом значении $z\in \mathrm{GF}(2^{n-1})$ (n четно).

„Только тогда“. Если $\rho+\sigma_j^3\neq 0$, $\rho+\sigma_{j+1}^3=0$, то подсчитаем $\sigma_0+\sigma_1=\alpha^h$, а затем положим $c_j(x)=x^h$, $c_{j+1}(x)=0$, $c=d+c_1(1)$ и вычислим Σ^*, т. е.

$$\sigma_j^*=\sigma_j+\alpha^h=\sigma_{j+1}=\sigma_{j+1}^*,\quad d^*=0,$$
$$\sigma^*=\sigma+\alpha^{3h}=\sigma+(\sigma_0+\sigma_1)^3=\rho=\sigma_{j+1}^3=\sigma_{j+1}^{*3}.$$

Итак, условия (28) выполняются, причем $c+W[c_0(x)]+W[c_1(x)]\leqslant 2$.

Теорема 2.1 дает следующее декодирующее правило.

Правило 1. Если $\sqrt[3]{\rho}=\sigma_j$, $\sqrt[3]{\rho}\neq\sigma_{j+1}$, то $c_{j+1}(x)=x^h$ и $c=d+c_1(1)$, где $\alpha^h=\sigma_0+\sigma_1$.

Теорема 2.2. *В случае исправимой ошибки* $\mathbf{e}$ *условия* $\rho+\sigma_j^3\neq 0$ $(j=0,\ 1)$, $d=1$ *выполняются тогда и только тогда, когда* $e=0$ *и* $W[e_j(x)]=1$ $(j=0,\ 1)$.

Доказательство. *„Тогда“*. Соотношения $e_j(x)=x^{k_j}$, $e=0$ дают $\sigma_j=b\alpha^s+\alpha^{k_j}$, $\sigma=b\alpha^{3s}+\alpha^{3k_j}+\alpha^{3k_{j+1}}$, $d=1$, следовательно, $\rho=b\alpha^{3s}+\alpha^{k_j}\alpha^{k_{j+1}}(\alpha^{k_j}+\alpha^{k_{j+1}})$. Отсюда имеем

$$\rho+\sigma_j^3=\alpha^{3k_j}\left[\left(\frac{\sigma_{j+1}}{\alpha^{k_j}}\right)^2+\frac{\sigma_{j+1}}{\alpha^{k_j}}+1\right]\neq 0$$

в $\mathrm{GF}(2^{n-1})$ (n четно).

„*Только тогда*". Если $d=1$, $\rho+\sigma_j^3\neq 0$ $(j=0, 1)$, то подсчитываем $\rho' \triangleq \sigma+\sigma_0\sigma_1(\sigma_0+\sigma_1)$ и получаем $\sigma_{j+1}+\sqrt[3]{\rho'}=\alpha^{k_j}$. Затем полагаем $c=[x^{k_0}, 0, x^{k_1}]$ и вычисляем Σ^*, т. е.

$$\sigma_j^*=\sigma_j+\alpha^{k_j}=\sigma_j+\sigma_{j+1}+\sqrt[3]{\rho'}=\sigma_{j+1}^*,$$

$$\sigma^*=\sigma+\alpha^{3k_j}+\alpha^{3k_{j+1}}=\sigma+(\sigma_{j+1}+\sqrt[3]{\rho'})^3+(\sigma_j+\sqrt[3]{\rho'})^3=$$

$$=(\sigma+\sigma_j\sigma_{j+1}(\sigma_{j+1}+\sigma_j)+\rho')+(\sigma_j+\sigma_{j+1}+\sqrt[3]{\rho'})^3=\sigma_j^{*3},$$

так как $\sigma+\sigma_j\sigma_{j+1}(\sigma_{j+1}+\sigma_j)=\rho'$. Наконец, $d^*=d+c_1(1)=0$. Соотношения (28) удовлетворяются, причем $c=W[c_0(x)]+W[c_1(x)]=2$.

Отсюда мы имеем следующее декодирующее правило.

Правило 2. Если $\sqrt[3]{\rho}\neq\sigma_0$, $\sqrt[3]{\rho}\neq\sigma_1$, $d=1$, то $c=0$ и $c_j(x)=x^{k_j}$, где $\alpha^{k_j}=\sigma_{j+1}+\sqrt[3]{\sigma+\sigma_0\sigma_1(\sigma_0+\sigma_1)}$.

Перед формулировкой теоремы 2.3 заметим, что при условии $(\sigma_0+\sigma_1)\neq 0$ функции $\tau_j \triangleq (\rho+\sigma_j^3)/(\sigma_0+\sigma_1)^3$ $(j=0, 1)$ связаны равенством

$$\tau_j+\tau_{j+1}+\frac{\sigma_0\sigma_1}{(\sigma_0+\sigma_1)^2}+1=0.$$

Выражая эти элементы поля $\mathrm{GF}(2^{n-1})$ как вектор-столбцы относительно базиса $1, \alpha, \ldots, \alpha^{n-2}$ и умножая слева на $\boldsymbol{\lambda}^T$ (см. (26)), имеем

$$\boldsymbol{\lambda}^T\boldsymbol{\tau}_j+\boldsymbol{\lambda}^T\boldsymbol{\tau}_{j+1}=1,$$

так как $\boldsymbol{\lambda}^T\cdot\mathbf{1}=1$ и $(\sigma_0\sigma_1)/(\sigma_0+\sigma_1)^2\in\Theta$ [фактически, $\sigma_0/(\sigma_0+\sigma_1)$ решает уравнение $y^2+y+(\sigma_0\sigma_1/(\sigma_0+\sigma_1)^2)=0$]. Этим доказана следующая лемма.

Лемма 11. *Если* $(\sigma_0+\sigma_1)\neq 0$, *то точно одна из двух функций* τ_j, τ_{j+1} *принадлежит* Θ.

Теорема 2.3. *В случае исправимой ошибки* e *условия* $\rho+\sigma_j^3\neq 0$ $(j=0, 1)$, $d=0$, $(\sigma_0+\sigma_1)\neq 0$ *выполняются тогда и только тогда, когда* $e_j(x)=0$, $W[e_{j+1}(x)]=2$, $e=0$.

Доказательство. „*Тогда*". Соотношение $e=0$, $e_j(x)=0$, $e_{j+1}(x)=x^{k_1}+x^{k_2}$ дают $\sigma_j=b\alpha^s$, $\sigma_{j+1}=b\alpha^s+e_{j+1}(\alpha)$, $\sigma=b\alpha^{3s}+e_{j+1}(\alpha^3)$. Отсюда (так как $e_{j+1}(\alpha)\neq 0$) имеем

$$\rho+\sigma_j^3=e_{j+1}^3(\alpha)+e_{j+1}(\alpha^3),$$

$$\rho+\sigma_{j+1}^3=e_{j+1}^3(\alpha)+e_{j+1}(\alpha^3)+e_{j+1}^3(\alpha)\left(1+\frac{\sigma_j}{e_{j+1}(\alpha)}+\left(\frac{\sigma_j}{e_{j+1}(\alpha)}\right)^2\right).$$

Вспоминая, что $e^3_{j+1}(\alpha)+e_{j+1}(\alpha^3)=\alpha^{k_1}\alpha^{k_2}(\alpha^{k_1}+\alpha^{k_2})$, и полагая $\gamma \triangleq \alpha^{k_1}/\alpha^{k_2} \neq 0$, $y \triangleq \sigma_j/e_{j+1}(\alpha)$, имеем $\rho+\sigma_j^3=\alpha^{3k_2}\gamma(1+\gamma)\neq 0$, поскольку $\gamma \neq 1$ $(k_1 \neq k_2)$, а также имеем

$$\rho+\sigma^3_{j+1}=\alpha^{3k_2}(1+\gamma)^3\left(y^2+y+1+\frac{\gamma}{1+\gamma^2}\right)\neq 0,$$

поскольку $1 \notin \Theta$ и $\gamma/(1+\gamma)^2 \in \Theta$ означают, что $1+\gamma(1+\gamma^2)\notin\Theta$. Кроме того, $d=0$ и $\sigma_0+\sigma_1=e_{j+1}(\alpha)\neq 0$.

„Только тогда". Если $d=0$, $\rho+\sigma_j^3\neq 0$ $(j=0, 1)$, $(\sigma_0+\sigma_1)\neq 0$, то получаем $\alpha^{k_1}/(\sigma_0+\sigma_1)$ и $\alpha^{k_2}/(\sigma_0+\sigma_1)$ в качестве решений уравнения

$$y^2+y+\frac{\rho+\sigma_j^3}{(\sigma_0+\sigma_1)^3}=0. \tag{30}$$

(Тот факт, что уравнение (30) имеет решения в $\mathrm{GF}(2^{n-1})$ в точности для одного из значений j, гарантируется леммой 11.) Положим $c_{j+1}(k)=x^{k_1}+x^{k_2}$, $c_j(x)=0$, $c=0$ и вычислим Σ^*, т. е.

$$\begin{aligned}
d^* &= d+e_1(1)=0,\\
\sigma^*_{j+1} &= \sigma_{j+1}+\frac{\alpha^{k_1}+\alpha^{k_2}}{\sigma_j+\sigma_{j+1}}(\sigma_j+\sigma_{j+1})=\sigma_j=\sigma_j^*,\\
\sigma^* &= \sigma+\frac{\alpha^{3k_1}+\alpha^{3k_2}}{(\sigma_0+\sigma_1)^3}(\sigma_0+\sigma_1)^3=\\
&= \sigma+(\sigma_0+\sigma_1)^3\left(1+\frac{\alpha^{k_1}\alpha^{k_2}}{(\sigma_0+\sigma_1)^2}\,\frac{\alpha^{k_1}+\alpha^{k_2}}{\sigma_0+\sigma_1}\right)=\\
&= \sigma+(\sigma_0+\sigma_1)^3\left(1+\frac{\rho+\sigma_j^3}{(\sigma_0+\sigma_1)^3}\right)=\sigma_j^3=\sigma_j^{*3},
\end{aligned}$$

так как $(\alpha^{k_1}+\alpha^{k_2})/(\sigma_0+\sigma_1)=1$ и $\alpha^{k_1}\alpha^{k_2}/(\sigma_0+\sigma_1)^2=(\rho+\sigma_j^3)/(\sigma_0+\sigma_1)^3$ являются суммой и произведением решений уравнения (30) соответственно. Соотношения (28) выполняются, причем $c+W[c_0(x)]+W[c_1(x)]=2$.

Это дает следующее декодирующее правило.

Правило 3. Если $\sqrt[3]{\rho}\neq\sigma_0$, $\sqrt[3]{\rho}\neq\sigma_1$, $d=0$, $\sigma_0+\sigma_1\neq 0$, то положим $c=0$, $c_j(x)=0$ и $c_{j+1}(x)=x^{k_1}+x^{k_2}$, где α^{k_1} и α^{k_2} являются решениями уравнения $z^2+(\sigma_0+\sigma_1)z+(\rho+\sigma_j^3)/(\sigma_0+\sigma_1)=0$.

Правила 1 — 3 составляют алгоритм, который производит коррекцию всех образцов исправимых ошибок. Выясним теперь, каково поведение этого алгоритма, когда полученный вектор $\mathbf{r}$ находится на расстоянии $\geqslant 3$ от любого $\mathbf{w}\in\mathscr{K}_n$. Ответ на этот вопрос неявно следует из предыдущих теорем, которые дают необходимые и достаточные условия для существования кодо-

вого слова, находящегося на расстоянии 2 или менее от полученного вектора $\mathbf{r}$. В силу этих теорем вектор $\mathbf{r}$ лежит на расстоянии $\geqslant 3$ от любого кодового слова тогда и только тогда, когда $\rho + \sigma_j^3 \neq 0$ $(j = 0, 1)$ (теорема 2.1), $d = 0$ (теорема 2.2) и $\sigma_0 + \sigma_1 = 0$ (теорема 2.3).

Ясно, что когда Σ удовлетворяет этим условиям, произвести коррекцию нельзя. В самом деле, хотя свойства расстояния кода $\mathcal{K}_n$ гарантируют, что существующий вектор коррекции с с весом $\leqslant 2$ является также единственным, тем не менее в случае, когда правила 1—3 неприменимы, можно построить более чем один вектор $\mathbf{c}$ с весом 3. Это показывается с помощью следующего рассуждения. Предположим, что условия $\sqrt[3]{\rho} \neq \sigma_j$ $(j = 0, 1)$, $d = 0$, $\sigma_0 = \sigma_1$ выполняются для $\mathbf{r}$. Определим α^h так, что $\left(1 + (\sigma + \sigma_0^3)/\alpha^{3h}\right) \in \Theta$. (Имеется 2^{n-2} значений h, которые удовлетворяют этому требованию, поскольку α^3 порождает мультипликативную группу поля $\mathrm{GF}(2^{n-1})$, и при фиксированном $(\sigma + \sigma_0^3)$ элемент $\left(1 + (\sigma + \sigma_0^3)/\alpha^{3h}\right)$ пробегает множество $\{0, \alpha, \alpha^2, \ldots, \alpha^{2^n-2}\}$, которое содержит Θ при четном n (лемма 8).) Образуем затем вектор коррекции с следующим образом: $c = 0$, $c_0(x) = x^h$, $c_1(x) = x^{k_1} + x^{k_2}$, где $\alpha^{k_i} = \sigma_0 + z_i\alpha^h$ $(i = 1, 2)$ и z_1, z_2 являются решениями уравнения

$$z^2 + z + 1 + \frac{\sigma + \sigma_0^3}{\alpha^{3h}} = 0.$$

Заметим, что $\alpha^{k_1} + \alpha^{k_2} = (z_1 + z_2)\alpha^h = \alpha^h$, так как $z_1 + z_2 = 1$. С учетом того, что $\sigma_0 = \sigma_1$, синдром Σ^* дает (см. (29))

$$\sigma_0^* = \sigma_0 + \alpha^h = \sigma_1 + \alpha^{k_1} + \alpha^{k_2} = \sigma_1^*,$$

$$\sigma^* = \sigma + \alpha^{3h} + \alpha^{3k_1} + \alpha^{3k_2} = \sigma_1 + \alpha^{k_1}\alpha^{k_2}(\alpha^{k_1} + \alpha^{k_2}) =$$

$$= \sigma + \alpha^h(\sigma_0 + z_1\alpha^h)(\sigma_0 + z_2\alpha^h) = (\sigma_0 + \alpha^h)^3 = \sigma_0^{*3},$$

$$d^* = c + c_1(1) = 0,$$

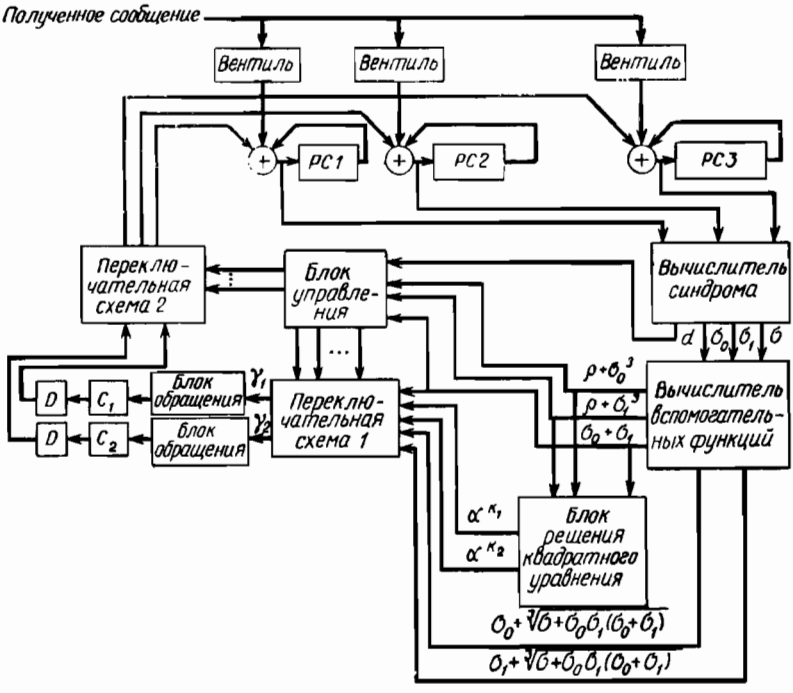
т. е. $\mathbf{r} + \mathbf{c} \in \mathcal{K}_n$. Это рассмотрение доказывает, что имеется несколько[1]) кодовых слов на расстоянии 3 от $\mathbf{r}$ (но ни одного слова на расстоянии < 3), и дает следующее правило обнаружения ошибки.

Правило 4. Если $\sqrt[3]{\rho} \neq \sigma_0$, $\sqrt[3]{\rho} \neq \sigma_1$, $d = 0$, $\sigma_0 + \sigma_1 = 0$, то полученный вектор $\mathbf{r}$ находится на расстоянии $\geqslant 3$ от любого кодового слова.

---

[1]) Другой вектор коррекции с весом 3 получается по правилу 2, т. е. $\mathbf{c} = [\alpha^k, 1, \alpha^k]$, где $\alpha^k = \sigma_0 + \sqrt[3]{\sigma} = \sigma_1 + \sqrt[3]{\sigma}$.

Дополнительный результат проведенного обсуждения состоит в том, что для любого заданного $\mathbf{r}$ имеются кодовые слова на расстоянии $\leqslant 3$ от $\mathbf{r}$. Это свойство аналогично свойству, обнаруженному в работе [5] для БЧХ кодов с исправлением двойных ошибок.



Р и с. 2. Декодер для кода $\mathcal{K}_n$.

На рис. 2 мы сделали набросок возможной блок-схемы декодера для кода $\mathcal{K}_n$. Последовательно получаемое сообщение запоминается в трех регистрах с циркуляцией РС1, РС2, РС3, соответствующих аналогичным регистрам на рис. 1. Получаемое сообщение посылается также на *вычислитель синдрома*, который после того, как прием завершен, хранит функции d, σ, σ_0, σ_1, т. е. синдром Σ. Эти функции являются входными сигналами комбинационных схем, которые мы теперь опишем (на рисунке жирными линиями обозначены пучки из $(n-1)$ двоичных линий).

Вычислитель вспомогательных функций (рис. 2) находит $\rho+\sigma_0^3$, $\rho+\sigma_1^3$, $\sigma_0+\sigma_1$, $\sigma_0+\sqrt[3]{\sigma+\sigma_0\sigma_1(\sigma_0+\sigma_1)}$ и $\sigma_1+\sqrt[3]{\sigma+\sigma_0\sigma_1(\sigma_0+\sigma_1)}$.

Первые три из них $\rho+\sigma_0^3$, $\rho+\sigma_1^3$, $\sigma_0+\sigma_1$ вместе с d посылаются в *блок управления*, который определяет, какое декодирующее правило должно быть применено. Параллельно с этим величины $\sigma+\rho_0^3$, $\rho+\sigma_1^3$, $\sigma_0+\sigma_1$ посылаются в *блок решения квадратного уравнения*, где вычисляются $\tau_0=(\rho+\sigma_0^3)/(\sigma_0+\sigma_1)^3$ и $\tau_1=(\rho+\sigma_1^3)/(\sigma_0+\sigma_1)^3$ и проверяется принадлежность множеству Θ. Если $\tau_j\in\Theta$, то $T\tau_j$ и $(1+T\tau_j)$ являются решениями уравнения $y^2+y+\tau_j=0$, где T — соответствующая квадратная матрица (см., например, [3]); затем $T\tau_j$ и $(1+T\tau_j)$ умножаются в $\mathrm{GF}(2^{n-1})$ на $(\sigma_0+\sigma_1)$ для получения решений α^{k_1} и α^{k_2} уравнения $z^2+(\sigma_0+\sigma_1)z+(\rho+\sigma_j^3)/(\sigma_0+\sigma_1)=0$ (см. правило 3). Так как $W[c_0(x)]+W[c_1(x)]\leqslant 2$, должно быть произведено исправление самое большее двух разрядов. Это выполняется следующим образом:

1) величины $\sigma_0+\sigma_1$, $\sigma_0+\sqrt[3]{\sigma+\sigma_0\sigma_1(\sigma_0+\sigma_1)}$, $\sigma_1+\sqrt[3]{\sigma+\sigma_0\sigma_1(\sigma_0+\sigma_1)}$, α^{k_1}, α^{k_2} посылаются в *переключательную схему* 1; здесь сигналы из блока управления определяют выбор двух корректирующих функций γ_1, γ_2 в виде векторных представлений двух элементов поля $\mathrm{GF}(2^{n-1})$;

2) комбинационная схема *блок обращения* (см. рис. 2) вычисляет обращение элемента $\gamma_j=\alpha^{h_j}$ (если $\gamma_j=0$, то выход блока обращения условно считается равным 0), и α^{-h_j} вводится в счетчик C_j, работающий в поле Галуа. Следует заметить, что если предположить отсутствие задержек в комбинационных элементах, то ввод α^{-h_j} $(j=1, 2)$ в C_j происходит одновременно с выработкой d, σ, σ_0, σ_1. На этом этапе содержимое регистров РС2 и РС3 циркулирует синхронно с работой счетчиков C_1 и C_2. Когда в C_j регистрируется последовательность $10\ldots0$, в блоке D образуется сигнал протяженностью в один такт и направляется через *переключательную схему* 2 для выполнения требуемой коррекции содержимого регистров. Поэтому операция декодирования заканчивается через $(2^{n-1}-1)$ тактов после завершения последовательного приема сообщения.

Этим заканчивается описание метода декодирования.

ЛИТЕРАТУРА

1. **Albert A. A.**, Fundamental concepts of higher algebra, University of Chicago Press, Chicago.
2. **Васильев Ю. Л.**, О негрупповых плотно упакованных кодах, Проблемы кибернетики, вып. 8, Физматгиз, 1962, стр. 337—339.

3. Berlekamp E. R., Rumsey H., Solomon G., On the solution of algebraic equations over finite fields, *Inform. and Control*, **10**, № 6 (1967), 553—564.
4. Berlekamp E. R., Algebraic coding theory, McGraw-Hill, New York, 1968. (В издательстве «Мир» готовится русский перевод этой книги.)
5. Gorenstein D., Peterson W. W., Zierler N., Two-error correcting Bose-Chaudhuri codes are quasi-perfect, *Inform. and Control*, **3**, № 3 (1960), 291—294. (Русский перевод: Кибериетический сборник, вып. 6, ИЛ, М., 1963.)
6. Green M. V., Two heuristic techniques for block-code construction, *IEEE Trans. on Inform. Theory*, **IT-12**, № 2 (1966), 273.
7. Johnson S. M., A new upper bound for error-correcting codes, *IRE Trans. on Inform. Theory*, **IT-8**, № 3 (1962), 203—207. (Русский перевод: сб. Теория кодирования, «Мир», М., 1964.)
8. Nadler M., A 32-point $n=12$, $d=5$ code, *IRE Trans. on Inform. Theory*, **IT-8**, № 1 (1962), 58.
9. Nordstrom A. W., Robinson J. P., An optimum nonlinear code, *Inform. and Control*, **11**, № 5/6 (1967), 613—616.
10. Петерсон У., Коды, исправляющие ошибки, «Мир», М., 1964.
11. Preparata F. P., An alternate description and a new decoding procedure of Nordstrom — Robinson optimum code, Proc. 2nd Princeton Conf. on Information Sciences and Systems, 1968, pp. 131—134.
12. Preparata F. P., Weight and distance structure of Nordstrom — Robinson quadratic code, *Inform. and Control*, **12**, № 5/6 (1968), 466—473 (см. также: Исправление, там же, **13**, № 7 (1968).)

О проблеме Дедекинда: число монотонных булевых функций[1])

Д. Клейтмен

Проблема определения числа $\psi(n)$ элементов свободной дистрибутивной структуры с n образующими была поставлена Дедекиндом [1] в 1897 г. Она была им решена для $n=4$. Р. Чёрч [2] в 1940 г. и М. Уорд [3] в 1946 г. получили решения соответственно для $n=5$ и $n=6$.

В 1954 г. Э. Н. Гильберт [4] показал, что $\psi(n)$ удовлетворяет неравенствам

$$2^{C_n^{[n/2]}} \leqslant \psi(n) \leqslant n^{C_n^{[n/2]}} + 2,$$

где $C_n^{[n/2]}$ – биномиальный коэффициент.

В. К. Коробков в нескольких статьях [5, 8*, 9*], опубликованных в 1962–1965 гг., сумел улучшить верхнюю границу для $\psi(n)$ до

$$2^{4{,}23 \cdot C_n^{[n/2]}}.$$

В 1966 г. Ж. Ансель [6] понизил верхнюю границу еще дальше, до

$$3^{C_n^{[n/2]}}.$$

В данной статье мы покажем, что $\log_2 \psi(n)$ асимптотически равен $C_n^{[n/2]}$, а именно мы покажем, что

$$2^{(1+\alpha_n) C_n^{[n/2]}} \leqslant \psi(n) \leqslant 2^{(1+\beta_n) C_n^{[n/2]}},$$

где $\alpha_n = ce^{-n/4}$, $\beta_n = \dfrac{c' \log n}{\sqrt{n}}$.

Число $\psi(n)$ равно числу идеалов, или антицепей, или монотонно возрастающих функций со значениями 0 и 1, определенных

---

[1]) Kleitman D., On Dedekind's problem: the number of monotone Boolean functions, *Proceedings of the American Mathematical Society*, **21**, № 3 (1969), 677–682.

на наборах длины n с элементами из $E^2=\{0, 1\}$[1]). Здесь идеал — это совокупность I наборов, такая, что если $\beta\in I$ и $\alpha<\beta$, то $\alpha\in I$; антицепь — это совокупность наборов, никакие два из которых несравнимы. Идеалу может быть однозначно поставлена в соответствие антицепь (его максимальные элементы) и монотонная функция (которая принимает значение 0 внутри и 1 вне его).

Ж. Ансель разбивал некоторым образом наборы длины n на $C_n^{[n/2]}$ цепей (вполне упорядоченных совокупностей наборов). Затем он определял всевозможные монотонные функции на каждой цепи по очереди. Так как оказалось, что можно не более чем тремя способами определить функции на каждой цепи, он получил следующий результат:

$$\psi(n)\leqslant 3^{C_n^{[n/2]}}.$$

Мы также разобьем наборы на цепи и определим монотонные функции на них по очереди. Однако мы будем допускать только два возможных определения на большинстве цепей (асимптотически на всех цепях). Поскольку такая процедура не дает всех возможных функций, повторим ее, используя $n!$ различных разбиений на цепи и $\alpha(n)$ различных упорядочений цепей. Докажем затем, что каждая функция получается при этом по крайней мере один раз. Таким образом, мы получим оценку для $\psi(n)$ в виде

$$\psi(n)\leqslant n!\,\alpha(n)\,2^{C_n^{[n/2]}(1+o(1))}.$$

Ниже дается процедура построения монотонных (со значениями 0 и 1) функций на наборах длины n. Затем доказывается, что каждая функция получается при этом хотя бы один раз.

Для удобства сначала построим монотонные функции на наборах, имеющих не менее $n/2-\tau$ и не более $n/2+\tau$ единиц. Такое же построение может быть применено для других интервалов числа единиц; так как каждая монотонная функция монотонна внутри каждого интервала, общее число таких функций не может превысить произведения числа таких функций, полученных для каждого интервала.

Рассмотрим некоторое определенное разбиение P наборов (с числом единиц в интервале, описанном выше) на $C_n^{[n/2]}$ цепей.

---

[1]) В оригинале „на структуре подмножеств n-элементного множества S_n". Здесь и в дальнейшем терминология автора заменена терминологией, употребляемой в отечественной литературе. — *Прим. перев.*

Мы считаем, что каждый член цепи, кроме наименьшего, покрывает другой член [1]). Согласно теоремам Дилворта и Шпернера, такое разбиение существует, а его пример дан Анселем [2]). Пусть π — любая подстановка компонент наборов. Она индуцирует подстановку среди наборов и, следовательно, среди цепей и среди разбиений на цепи. Процедура построения, описанная ниже, будет применена ко всем $n!$ разбиениям, полученным применением всех подстановок π к данному разбиению P.

Эта процедура включает упорядоченные разбиения на цепи, а именно такие разбиения, как в предыдущем абзаце, в которых группы цепей расположены в определенном порядке. Для каждого разбиения $\pi(P)$ мы будем рассматривать некоторое количество различных упорядочений цепей. Зададим сначала особое упорядочение следующим образом. Каждая цепь содержит один набор с $[n/2]$ единицами; мы зададим упорядочение для них и тем самым порядок цепей. Упорядочим наборы $\alpha_1, \alpha_2, \ldots, \alpha_{C_n^{[n/2]}}$ с $[n/2]$ единицами таким образом, чтобы как можно дальше наборы α_j удовлетворяли условию $|\alpha_j \cap \alpha_k| < \frac{n}{2} - \tau - 1$ для всех k из интервала $j - l \leqslant k < j$ [3]). Так как α_j при таком ограничении не может совпадать самое большее с

$$l \sum_{r=0}^{\tau+1} C_{[n/2]}^r C_{n-[n/2]}^r$$

наборами [см. п. 1 дополнений переводчика в конце статьи], можно считать, что α_j удовлетворяют данному условию для

$$j \leqslant j_0 \equiv C_n^{[n/2]} - l \sum_{r=0}^{\tau+1} C_{[n/2]}^r C_{n-[n/2]}^r.$$

Разобьем целые числа $j \leqslant j_0$ на блоки длины l; целые числа между j_0 и $C_n^{[n/2]}$ будем рассматривать как отдельные блоки. Возьмем в качестве упорядочений цепей (в упорядоченных разбиениях на цепи, используемых в построении ниже) все

$$\left([j_0/l] + 1 + l \sum_{r=0}^{\tau+1} C_{[n/2]}^r C_{n-[n/2]}^r\right)!$$

---

[1]) Выражение „набор x покрывает набор y“ в данной работе имеет следующий смысл: y получается из x заменой одной 1 на 0. Таким образом, рассматриваются цепи без пропусков и, кроме того, они предполагаются непересекающимися. — *Прим. перев.*

[2]) См. работу [6]. Из свойств цепей Анселя в данной статье используются только следующие: 1) цепи не имеют пропусков, 2) цепи не пересекаются, 3) их число равно $C_n^{[n/2]}$. — *Прим. перев.*

[3]) Здесь $\alpha_j \cap \alpha_k$ означает покомпонентную конъюнкцию, а $|\alpha|$ — число единиц в наборе α. — *Прим. перев.*

упорядочений блоков цепей, связанных с только что определенными блоками наборов α_j. Для каждых n и τ выберем l так, чтобы минимизировать выписанный выше факториал, минимальное значение факториала будем обозначать через $\alpha(n, \tau)$.

К каждому из $n!\alpha(n, \tau)$ упорядоченных разбиений на цепи применим следующую процедуру. На каждой цепи по порядку зададим значения 0 или 1 на каждом наборе любым способом в соответствии с уже выбранными заданиями на предшествующих цепях так, чтобы построенная функция была монотонно возрастающей. Однако рассматривать будем только те функции, в процессе задания которых не более чем на t цепях встречается следующая ситуация: в множестве членов цепи, на которых значения функции не предопределены (то есть еще не определены по монотонности заданием функции на предшествующих цепях), на предмаксимальном члене значение функции положено равным 1. Число функций, построенных согласно этому предписанию, не может превысить величины

$$C^t_{C_n^{[n/2]}}(2\tau+2)^t 2^{C_n^{[n/2]}-t}, \qquad (1)$$

так как на всех цепях (кроме t цепей, на которых функция на этом предмаксимальном непредопределенном наборе может принимать значение 1) можно задать только две функции: на наибольшем непредопределенном наборе может быть положен 0 или 1.

Желаемый результат (на интервале $n/2-\tau$, $n/2+\tau$) будет получен, если мы покажем, что для некоторого t, такого, что

$$n!\,\alpha(n, \tau) C^t_{C_n^{[n/2]}}(2\tau+2)^t 2^{C_n^{[n/2]}-t} = 2^{C_n^{[n/2]}(1+o(1))},$$

будут построены все монотонные функции при помощи процедуры, описанной выше и примененной ко всем вышеописанным упорядоченным разбиениям.

Рассмотрим монотонную функцию f на наборах, число единиц в которых заключено между $[n/2]-\tau$ и $[n/2]+\tau$[1]). Разобьем наборы x, удовлетворяющие условию $f(x)=1$, на два класса: $x \in A_k$, если среди наборов y, покрываемых набором x, не более чем k удовлетворяют условию $f(y)=1$, и $x \in B_k$ в противном случае. Если $x \in A_k$, то не более чем в $k/|x|$ всех разбиений, описанных выше, набор y, покрываемый набором x и лежащий в цепи, содержащей x, удовлетворяет условию $f(y)=1$ [см. п. 2 дополнений], тогда как по крайней мере в $1-k/|x|$

[1]) Считается, что вне этого интервала при меньшем числе единиц функция равна 0, а при большем — равна 1. — *Прим. перев.*

всех разбиений для набора y, входящего в цепь с x и покрываемого набором x, будет $f(y)=0$ [1]). Последнее будет иметь место, следовательно, по крайней мере для

$$n!\left(1-\frac{k}{[n/2]-\tau}\right)|A_k|$$

цепей в неупорядоченных разбиениях. Отсюда следует, что [см. п. 3 дополнений]

$$\left(1-\frac{k}{[n/2]-\tau}\right)|A_k|\leqslant C_n^{[n/2]}$$

и поэтому условие

$$x\in A_k,\quad x \text{ покрывает } y,\quad f(y)=1$$

выполняется в среднем по всем разбиениям не более **чем на**

$$C_n^{[n/2]}\frac{k}{[n/2]-\tau}\left(1-\frac{k}{[n/2]-\tau}\right)^{-1}$$

цепях.

Следовательно, должно существовать разбиение **на цепи**, в котором это условие выполняется не более чем на

$$C_n^{[n/2]}\frac{k}{[n/2]-\tau}\left(1-\frac{k}{[n/2]-\tau}\right)^{-1}$$

цепях. Выберем одно из таких разбиений P_f. Пусть в дальнейшем x есть наименьший член цепи из P_f, удовлетворяющий условию $x\in B_k$. Значение $f(x)$ будет предопределено, если некоторая из цепей, содержащих наборы y, покрываемые набором x и удовлетворяющие условию $f(y)=1$ (таких цепей не менее k), появится в упорядоченном разбиении раньше, чем цепь, содержащая x. Конструкция блоков, данная выше, обеспечивает то, что никакие два набора, покрываемые набором x, не могут лежать в одном и том же блоке цепей, образованном при формировании упорядоченных разбиений, рассмотренном выше [см. п. 4 дополнений]. Таким образом, набор x будет впереди k цепей, содержащих такие y, не более чем для $1/k$ всех упорядочений. Следовательно, $f(x)$ будет предопределено по крайней мере для $(k-1)/k$ всех упорядоченнй разбиения P_f.

Так как имеется самое большее $C_n^{[n/2]}$ членов B_k, которые являются наименьшими членами цепи с таким свойством, то в среднем по всем упорядочениям число цепей, содержащих наборы x из B_k, такие, что $f(x)$ не предопределено, не более

---

[1]) Либо x есть минимальный элемент цепи. — *Прим. перев.*

чем $\frac{1}{k}C_n^{[n/2]}$. Следовательно, должно существовать упорядочение O_f, в котором $f(x)$ предопределено в B_k на всех, кроме, быть может, $\frac{1}{k}C_n^{[n/2]}$, цепях.

Если при построении f на данной цепи предмаксимальному непредопределенному члену приписано значение 1, то наибольший непредопределенный член должен:

а) либо быть в A_k и покрывать набор, удовлетворяющий условию $f(y)=1$,

б) либо быть в B_k и быть непредопределенным.

По доказанному выше, если выбрано упорядочение O_f разбиения P_f, то эта альтернатива может выполняться не более чем на

$$C_n^{[n/2]}\left(\frac{1}{k}+\frac{k}{[n/2]-\tau}\left(1-\frac{k}{[n/2]-\tau}\right)^{-1}\right)$$

цепях. Таким образом, для

$$t=C_n^{[n/2]}\left(\frac{1}{k}+\frac{k}{[n/2]-\tau}\left(1-\frac{k}{[n/2]-\tau}\right)^{-1}\right)$$

процедура, описанная выше, должна дать все монотонные функции на данном интервале. Оптимальное значение для k здесь приблизительно равно $([n/2]-\tau)^{1/2}$, так что мы можем выбрать

$$t=C_n^{[n/2]}\frac{2}{([n/2]-\tau)^{1/2}}\left(1+O\left(\frac{1}{\sqrt{n}}\right)\right).$$

Легко убедиться в том, что $n!$ и $\alpha(n,\tau)$ крайне незначительны по сравнению с (1) при $\tau\sim n^{\beta}$, $\beta<1$, и что число монотонных функций в рассмотренном интервале не может превосходить число

$$2^{C_n^{[n/2]}\left(1+Cn^{-1/2}\log n\right)}.$$

Второй член в показателе появляется из выражения (1) при значении t, данном выше [см. п. 5 дополнений].

Подобные рассуждения могут быть применены к вычислению числа монотонных функций на других интервалах числа единиц. Выбирая подходящим образом эти интервалы, можно показать, что рассмотрение остальных наборов дает много меньший вклад в общее число монотонных функций, чем определяемый выше вклад за счет второго члена в показателе [см. п. 6 дополнений]. Таким образом, имеем

$$\psi(n)\leqslant 2^{C_n^{[n/2]}\left(1+Cn^{-1/2}\log n\right)}.$$

Можно легко составить $2^{C_n^{[n/2]}}$ антицепей, рассматривая все совокупности наборов с $[n/2]$ единицами. Большинство таких антицепей будет содержать приблизительно половину всех наборов с $[n/2]$ единицами и будет содержать наборы, которые покрываются всеми наборами, кроме приблизительно $C_n^{[n/2]+1}e^{-n/4}$ наборов с $[n/2]+1$ единицами. Любые из оставшихся наборов с $[n/2]+1$ единицами могут быть добавлены к антицепи. Следовательно,

$$\psi(n) \geqslant 2^{C_n^{[n/2]}\left(1+e^{-n/4}\right)}.$$

Эта нижняя оценка может быть улучшена [7].

ДОПОЛНЕНИЯ ПЕРЕВОДЧИКА

1. Число рассматриваемых наборов, имеющих $[n/2]-r$ общих единиц с набором α_k, равно $C^r_{[n/2]}C^r_{n-[n/2]}$ (r единиц заменяется 0 и r нулей заменяется 1). Поэтому α_k запрещает ставить на место α_l не более чем $\sum\limits_{r=0}^{\tau+1} C^r_{[n/2]}C^r_{n-[n/2]}$ наборов. А всего на место α_l нельзя ставить не более чем $l\sum\limits_{r=0}^{\tau+1} C^r_{[n/2]}C^r_{n-[n/2]}$ наборов.

2. Пусть в x' единицы соответствуют компонентам с номерами $i_1, \ldots, i_m$, а в x — компонентам с номерами $j_1, \ldots, j_m$. Если в первоначальном разбиении на цепи x' покрывает в своей цепи y', а x' и y' различаются на i_l-й компоненте, то среди всех подстановок, переводящих x' в x, i_l-я компонента будет с одинаковой частотой (в силу симметрии) переходить во все компоненты с номерами j_p. Поэтому звено (x', y') будет с одинаковой частотой переходить во все ребра (x, y), выходящие из x. Кроме того, x' может быть минимальным в своей цепи; тогда при подстановках, переводящих x' в x, из x не будет исходить звено цепи. Отсюда следует нужное утверждение.

3. При любом разбиении на цепи на каждой цепи не более чем один набор x, такой, что $f(x)=1$, является минимальным элементом цепи или покрывает в своей цепи набор y, на котором $f(y)=0$. Поэтому при каждом разбиении не более чем $C_n^{[n/2]}$ элементов из A_k покрывают в своей цепи набор y, такой, что $f(y)=0$, или ничего не покрывают. Следовательно,

$$n!\left(1-\frac{k}{[n/2]-\tau}\right)|A_k| \leqslant C_n^{[n/2]}n!,$$

откуда по сокращении на $n!$ и получается требуемое неравенство.

4. Пусть x покрывает y_1 и y_2, а цепи, проходящие через y_1 и y_2, содержат наборы с $[n/2]$ единицами соответственно α_{i_1} и α_{i_2}. Если $|x|>[n/2]$, то $\alpha_{i_1}<x$ и $\alpha_{i_2}<x$. Так как $|\alpha_{i_1}\cup\alpha_{i_2}|=|\alpha_{i_1}|+|\alpha_{i_2}|-|\alpha_{i_1}\cap\alpha_{i_2}|$ и $\alpha_{i_1}\cup\alpha_{i_2}\leqslant x$, то $|\alpha_{i_1}\cap\alpha_{i_2}|=|\alpha_{i_1}|+|\alpha_{i_2}|-|\alpha_{i_1}\cup\alpha_{i_2}|\geqslant|\alpha_{i_1}|+|\alpha_{i_2}|-|x|\geqslant 2\cdot[n/2]-n/2-\tau\geqslant n/2-\tau-1$. Если $|x|\leqslant[n/2]$, то y_1 и y_2 покрывают один и тот же набор z. Так как $\alpha_{i_1}\cap\alpha_{i_2}\geqslant z$, то $|\alpha_{i_1}\cap\alpha_{i_2}|\geqslant|z|\geqslant n/2-\tau-1$. Таким образом, в любом случае цепи, содержащие y_1 и y_2, лежат в разных блоках.

5. Произведем указанные вычисления:

$$\log n!\leqslant n\cdot\log n=o\left(C_n^{[n/2]}\cdot\frac{\log n}{\sqrt{n}}\right).$$

Оценим $\log\alpha(n,\tau)$. Легко получить, раскрыв факториалы в формуле $C_n^{\alpha n}=\frac{n!}{(\alpha n)!\,(n-\alpha n)!}$ по формуле Стирлиига, что $C_n^{\alpha n}=(1+\gamma)^n$, где $\gamma\to 0$ при $\alpha\to 0$. Так как $\tau/n\to 0$, то $R=\sum\limits_{r=0}^{\tau+1}C_{[n/2]}^r C_{n-[n/2]}^r\leqslant$ $\leqslant(\tau+2)\left(C_{n-[n/2]}^{\tau+1}\right)^2\sim n^\beta(1+\gamma)^n=(1+\delta)^n$, где $\gamma\to 0$ и $\delta\to 0$ при $n\to\infty$. Пусть $l\sim\sqrt{\frac{C_n^{[n/2]}}{R}}$ $\left(\text{например, } l=\left[\sqrt{\frac{C_n^{[n/2]}}{R}}\right]\right)$, тогда

$$L=\left[\frac{j_0}{l}\right]+1+lR\leqslant\frac{C_n^{[n/2]}}{l}+lR\sim 2\sqrt{C_n^{[n/2]}R}.$$

Отсюда

$$\log\alpha(n,\tau)\leqslant L\log L\lesssim 2\sqrt{C_n^{[n/2]}R}\,\frac{1}{2}n=$$

$$=C_n^{[n/2]}\sqrt{\frac{R}{C_n^{[n/2]}}}\,n=o\left(C_n^{[n/2]}\frac{\log n}{\sqrt{n}}\right),$$

так как $R=(1+\delta)^n$.

Упростив (1), получим $C^t_{C_n^{[n/2]}}(\tau+1)^t 2^{C_n^{[n/2]}}$. Имеем

$$\log(\tau+1)^t=t\log(\tau+1)\sim C_n^{[n/2]}\frac{2}{\sqrt{\frac{n}{2}}}\beta\log n\sim C_n^{[n/2]}\frac{C_1\log n}{\sqrt{n}}.$$

При $a\to\infty$ и $b=o(a)$ находим $\log C_a^b\sim b\log\frac{a}{b}$, что также легко получить, используя формулу Стирлинга. Отсюда

$$\log C^t_{C_n^{[n/2]}}\sim t\log\frac{C_n^{[n/2]}}{t}\sim C_n^{[n/2]}\sqrt{\frac{8}{n}}\,\frac{1}{2}\log n=C_n^{[n/2]}\frac{C_2\log n}{\sqrt{n}}.$$

6. Рассмотрим все наборы с числом единиц вне интервала $\left(\frac{n}{2}-\tau,\ \frac{n}{2}+\tau\right)$. Их число не более чем $2\sum_{r=0}^{[n/2]-\tau} C_n^r \leqslant$ $\leqslant 2\cdot([n/2]-\tau+1)\,C_n^{[n/2]-\tau}$. Имеем

$$\frac{C_n^{[n/2]-\tau}}{C_n^{[n/2]}}=\frac{([n/2]-\tau+1)\cdot\ \ldots\ \cdot\left[\frac{n}{2}\right]}{\left(n-\left[\frac{n}{2}\right]+\tau\right)\cdot\ \ldots\ \cdot\left(n-\left[\frac{n}{2}\right]+1\right)}=$$

$$=\frac{\left[\frac{n}{2}\right]}{n-\left[\frac{n}{2}\right]+\tau}\cdot\ \ldots\ \cdot\frac{\left[\frac{n}{2}\right]-\tau+1}{n-\left[\frac{n}{2}\right]+1}.$$

Так как при $a<b$ имеем $\frac{a}{b}<\frac{a+1}{b+1}$, то

$$\frac{C_n^{[n/2]-\tau}}{C_n^{[n/2]}}\leqslant\left(\frac{n/2}{\frac{n}{2}+\tau}\right)^{\tau}=\left(1-\frac{\tau}{\frac{n}{2}+\tau}\right)^{\tau}=$$

$$=\left(1-\frac{\tau}{\frac{n}{2}+\tau}\right)^{\frac{(n/2)+\tau}{\tau}\cdot\frac{\tau^2}{(n/2)+\tau}}\leqslant\left(\frac{1}{e}+\varepsilon\right)^{\frac{\tau^2}{(n/2)+\tau}}.$$

При $\tau\sim n^{\beta}$ **и** $\beta>\frac{1}{2}$ находим

$$2\sum_{r=0}^{[n/2]-\tau} C_n^r\leqslant 2\,([n/2]-n^{\beta}+1)\cdot\left(\frac{1}{e}+\varepsilon\right)^{\frac{n^{2\beta}}{(n/2)-n^{\beta}+1}}C_n^{[n/2]}=o\left(C_n^{[n/2]}\frac{\log n}{\sqrt{n}}\right).$$

ЛИТЕРАТУРА [1])

1. Dedekind R., Über Zerlegungen von Zahlen durch ihre grössten gemeinsamen Teilor, Festschrift Hoch. Braunschweig u. ges. Werke, II, 1897, S. 103—148.
2. Church R., Numerical analysis of certain free distributive structures, *Duke Math. J.*, **6** (1940), 732—734.
3. Ward M., Note on the order of free distributive lattices, Abstract 135, *Bull. Amer. Math. Soc.*, **52** (1946), 423.
4. Gilbert E. N., Lattice theoretic properties of frontal switching functions, *J. Math. Phys.*, **33** (1954), 57—67. (Русский перевод: Гильберт Э. Н., Теоретико-структурные свойства замыкающих переключательных функций, Кибернетический сборник, вып. 1, ИЛ, М., 1960.)

[1]) Работы, отмеченные звездочкой, добавлены при подготовке русского перевода. — *Прим. ред.*

5. Коробков В. К., О монотонных функциях алгебры логики, сб. Проблемы кибернетики, вып. 13, изд-во «Наука», М., 1965, стр. 5—28.
6. Hansel G., Sur le nombre des fonctions booléennes monotones de *n* variables, *C. R. Acad. Sci. Paris*, **262** (1966), 1088—1090. (Русский перевод: Анисель Ж., О числе монотонных булевых функций *n* переменных, Кибернетический сборник, новая серия, вып. 5, «Мир», М., 1968.)
7. Yamamoto K., Logarithmic order of free distributive lattices, *J. Math. Soc. Japan*, **6** (1954), 343—353.
8*. Коробков В. К., К вопросу о числе монотонных функций алгебры логики, Дискретный анализ, вып. 1, 1963, сборник трудов Института математики СО АН СССР.
9*. Коробков В. К., Оценка числа монотонных функций алгебры логики и сложности алгоритма отыскания разрешающего множества для произвольной монотонной функции алгебры логики, *ДАН СССР*, **150**, № 4 (1963), 744—747.

Время, требующееся для умножения в группе[1])

Ф. М. Спира

В работах Винограда [1, 2] было исследовано время, требующееся для выполнения сложения и умножения натуральных чисел и для выполнения умножения в группе логической схемой из элементов с ограниченным числом входов и задержкой, равной 1. В данной статье принята та же самая модель. Получена новая нижняя оценка для умножения в группе, совпадающая для абелевых групп с оценкой Винограда, но, вообще говоря, более сильная. Приводится также схема для выполнения умножения, которая в отличие от схемы Винограда может использоваться и для неабелевых групп. В случае абелевой группы эта схема оказалась по крайней мере столь же быстрой, как и схема Винограда. Следуя его методу применения схемы, построенной для абелевой группы, оказывается возможным также понизить верхнюю оценку времени, требующегося для выполнения сложения и умножения натуральных чисел.

1. МОДЕЛЬ

Принятая нами модель — по существу модель Винограда [1, 2]. Мы рассматриваем логические схемы, построенные из элементов, каждый из которых имеет не более r входных линий, одну расщепляемую выходную линию и задержку при вычислении выхода, равную 1. По каждой линии передаются значения из множества $Z_d = \{0, 1, \ldots, d-1\}$. Входные линии схемы разбиваются на n множеств; через $I_{C,j}$ обозначается множество всевозможных наборов значений на входах j-го множества ($j = 1, 2, \ldots, n$). Через O_C обозначается множество всевозможных наборов значений на выходных линиях схемы. Такая схема называется (d, r)-схемой.

Определение 1.1. Пусть φ: $X_1 \times X_2 \times \ldots \times X_n \to Y$ — функция, определенная на конечных множествах. Говорят, что схема C вычисляет φ в момент времени τ, если существуют отображения g_j: $X_j \to I_{C,j}$ ($j = 1, 2, \ldots, n$) и взаимно однозначная функция h: $Y \to O_C$, такие, что если C с момента времени 0 и до момента времени $\tau - 1$ получает постоянный вход $[g_1(x_1), \ldots, g_n(x_n)]$, то ее выход в момент времени τ есть $h(\varphi(x_1, \ldots, x_n))$.

---

[1]) Spira Ph. M., The time required for group multiplication, *Journal of the Association for Computing Machinery*, **16**, № 2 (1969), 235–243.

2. ОСНОВНАЯ ЛЕММА

Здесь мы получим общую нижнюю оценку времени работы (d, r)-схемы, вычисляющей заданную конечную функцию φ. Эта оценка проясняет метод, лежащий в основе результатов Винограда. Она зависит от свойств выходного кода h, введенного в предыдущем разделе, и использует новое понятие, которое мы здесь вводим, — понятие разделимых множеств. Сначала необходимо дать некоторые предварительные определения.

Определение 2.1. Пусть $]x[$ обозначает наименьшее целое число, большее или равное x; пусть $[x]$ обозначает наибольшее целое число, меньшее или равное x; пусть $|S|$ обозначает мощность множества S.

Определение 2.2. В любой (d, r)-схеме всегда, когда выходной набор значений есть $h(y)$, через $h_j(y)$ будем обозначать значение на j-й выходной линии.

Определение 2.3. Пусть φ: $X_1 \times \ldots \times X_n \to Y$, и пусть C вычисляет φ. Тогда $S \subseteq X_m$ называется h_j-разделимым множеством для C в m-м аргументе φ, если для любых двух различных элементов s_1 и s_2 из S мы можем найти $x_1, x_2, \ldots, x_{m-1}, x_{m+1}, \ldots, x_n$, где $x_i \in X_i$, такие, что

$$h_j(\varphi(x_1, \ldots, x_{m-1}, s_1, x_{m+1}, \ldots, x_n)) \neq$$
$$\neq h_j(\varphi(x_1, \ldots, x_{m-1}, s_2, x_{m+1}, \ldots, x_n)).$$

Лемма 2.1. *В (d, r)-схеме выход любого элемента в момент времени τ может зависеть не более чем от r^τ входных линий.*

Доказательство. Достаточно рассмотреть древообразную схему глубины τ из элементов с r входными линиями[1]). ▶

Это наблюдение, сделанное сначала Виноградом, в совокупности с понятием разделимых множеств позволяет доказать следующую лемму.

Лемма 2.2 (основная лемма). *Пусть C есть (d, r)-схема, вычисляющая φ в момент времени τ. Тогда*

$$\tau \geqslant \max_j \{\,]\log_r(\,]\log_d |S_1(j)|\, [+ \ldots +]\log_d |S_n(j)|\, [\,)[\,\},$$

где $S_i(j)$ есть h_j-разделимое множество для C в i-м аргументе φ.

Доказательство. В момент времени τ j-я выходная линия должна зависеть не менее, чем от $]\log_d |S_i(j)|[$ входных

[1]) Знак ▶ означает конец доказательства. — *Прим. перев.*

линий, определяющих $I_{C,i}$, так как в противном случае в $S_i(j)$ нашлись бы два элемента, которые не были бы h_j-разделимыми. Таким образом, j-й выход зависит не менее чем от $]\log_d|S_1(j)|[+\ \ldots\ +]\log_d|S_n(j)|[$ входных линий, а это число не превосходит r^τ. ▶

С помощью леммы 2.2 мы раскроем ход рассуждений, который в изложении Винограда, посвященном времени сложения и умножения, выражен не в столь явной форме. За счет этого в оставшейся части раздела мы не только быстро получим некоторые результаты Винограда, но и дадим более глубокий анализ других понятий и исследуем значительно более широкий класс функций впоследствии.

Следствие 2.1. *Пусть* φ: $Z_N \times Z_N \to \{0, 1\}$ *есть*

$$\varphi(x, y) = \begin{cases} 1, & \text{если } x \leqslant y, \\ 0, & \text{если } x > y. \end{cases}$$

Тогда если C *есть* (d, r)*-схема, вычисляющая* φ *в момент времени* τ, *то* $\tau \geqslant]\log_r 2]\log_d N[[$.

Доказательство. Возьмем j, такое, что $h_j(0) \neq h_j(1)$. Тогда Z_N есть h_j-разделимое множество для C в обоих — первом и втором — аргументах φ, так как если $x > y$, то $\varphi(x, y) \neq \varphi(y, y)$ и $\varphi(x, y) \neq \varphi(x, x)$. ▶

Следствие 2.2. *Пусть* φ: $Z_N \times Z_N \to Z_N$ *есть*

$$\varphi(x_1, x_2) = [x_1 x_2 / N].$$

Тогда если C *вычисляет* φ *в момент времени* τ, *то*

$$\tau \geqslant]\log_r 2]\log_d [N^{1/2}][[.$$

Доказательство. Возьмем j, такое, что $h_j(0) \neq h_j(1)$. Пусть $m = [N^{1/2}]$. Тогда $\{1, 2, \ldots, m\}$ есть h_j-разделимое множество для C в обоих аргументах φ, так как для каждой пары x, y, в которой $x < y$ и $x, y \in \{1, 2, \ldots, m\}$, мы можем выбрать $w \in Z_N$, такой, что $xw < N \leqslant yw < 2N$, с тем, чтобы получить $\varphi(x, w) = 0$, $\varphi(y, w) = 1$. В силу симметрии это утверждение верно также и для второго аргумента. Применение леммы 2.2 завершает доказательство. ▶

Данный раздел мы закончим примером, который показывает, что размеры разделимых множеств могут сильно зависеть от выходного кода схемы, вычисляющей заданную φ.

Пример 2.1. Пусть φ: $Z_N \times Z_N \to Z_{N^2}$ есть умножение натуральных чисел, причем $N = 2^8$. Рассмотрим выходной код,

при котором всегда, когда выход равен M, i-я выходная линия передает i-ю цифру двоичного представления M. При этом имеется шестнадцать выходных линий. Возьмем произвольную пару x, y, в которой $x \neq y$ и $x, y \in Z_N$. Двоичные представления этих чисел отличаются хотя бы в одном разряде, скажем в k-м. Выберем $z = 2^{8-k}$. Тогда $h_8(\varphi(y, z)) \neq h_8(\varphi(x, z))$ и $h_8(\varphi(z, y)) \neq h_8(\varphi(z, x))$. Таким образом, существует h_8-разделимое множество в обоих аргументах φ мощности 2^8.

Рассмотрим теперь ту же самую φ, но в качестве выходного кода для z возьмем набор двоичных представлений показателей степени в разложении z на простые сомножители. Пусть первые шесть выходных линий [1]) используются для кодирования показателя степени числа два в результате. Возьмем $x, y \in Z_N$, такие, что x и y в разложении на простые сомножители содержат различные степени числа два; пусть, например, показатели этих степеней при двоичном представлении отличаются в k-м разряде. Тогда, полагая $z = 2^{3-k}$, получаем $h_3(\varphi(x, z)) \neq h_3(\varphi(y, z))$ и $h_3(\varphi(z, x)) \neq h_3(\varphi(z, y))$. Таким образом, поскольку элементы из Z_N могут иметь в разложении на простые сомножители восемь различных степеней числа два, существует h_3-разделимое множество в обоих аргументах φ мощности 8. Поскольку число два — наименьшее простое число, то, как нетрудно заметить, это максимальный размер разделимого множества. Отметим, однако, что для этого выходного кода требуется тридцать девять выходных линий.

3. ОБЗОР РЕЗУЛЬТАТОВ

Рядом авторов исследовалось время, необходимое для выполнения (d, r)-схемой сложения по модулю N. Офман [3] построил схему для частного случая $N = 2^n$. Важные результаты получил Виноград [1, 2]. Он установил нижнюю оценку, которую мы здесь приведем, и построил (d, r)-схему с временем работы, близким к нижней оценке. Поскольку любая конечная абелева группа является прямым произведением циклических групп [4, стр. 50], результаты Винограда применимы также и к умножению в абелевой группе.

Определение 3.1. Пусть H — группа. Говорят, что H обладает свойством P, и пишут $P(H) = 1$, если существует элемент $a \in H$, причем $a \neq e$, такой, что каждая нетривиальная подгруппа группы H содержит a. Последнее обозначается в виде

---

[1]) По-видимому, многие числа в этом абзаце указаны автором неверно. — *Прим. перев.*

$P(a, H)=1$. Через $\alpha(G)$ обозначим максимальный порядок $H \subseteq G$, такой, что $P(H)=1$.

Лемма 3.1 (Виноград [1]). *Если G — абелева группа, то $\alpha(G)$ равно максимальному, являющемуся степенью простого числа порядку циклической подгруппы, содержащейся в G.*

Доказательство см. в работе [1]. ▶

Теперь мы полностью охарактеризуем $\alpha(G)$.

Определение 3.2. Обобщенной группой кватернионов Q_n называется группа порядка 2^n с двумя порождающими элементами a и b, удовлетворяющими соотношениям

$$a^{2^{n-1}}=e; \quad b^2=a^{2^{n-2}}; \quad ba=a^{-1}b.$$

Теорема 3.1. *p-группа содержит единственную подгруппу порядка p тогда и только тогда, когда она является циклической или обобщенной группой кватернионов. (Если p нечетно, то она должна быть циклической.)*

Доказательство см. в работе [4, стр. 211]. ▶

Следствие 3.1. *Пусть G — произвольная конечная группа. Тогда $\alpha(G)$ равно либо порядку наибольшей циклической p-подгруппы группы G, либо порядку наибольшей обобщенной группы кватернионов, содержащейся в G, а именно тому из них, который больше.*

Доказательство. Пусть H — произвольная подгруппа группы G. Если $P(H)=1$, то $|H|=p^n$ для некоторого простого числа p, так как в противном случае нашлось бы другое простое число q, делящее $|H|$, и, следовательно, нашлись бы элементы u и v из H с порядками $o(u)=p$ и $o(v)=q$ [1]). Но тогда подгруппа $\langle u\rangle \cap \langle v\rangle$ содержала бы только единицу. Итак, принимаем $|H|=p^n$. В таком случае каждая нетривиальная подгруппа группы H содержит подгруппу порядка p. Таким образом, $P(H)=1$ тогда и только тогда, когда H содержит единственную подгруппу порядка p, т. е. H является циклической группой или обобщенной группой кватернионов. ▶

Величина $\alpha(G)$ играет решающую роль в нижней оценке Винограда для времени умножения в группе. Эту оценку мы сейчас приведем. В разд. 4 мы дадим новую нижнюю оценку, которая, вообще говоря, выше оценки Винограда, но в случае абелевой группы совпадает с ней.

---

[1]) См., например, [4, стр. 54]. — *Прим. перев.*

Теорема 3.2 (Виноград [1]). *Пусть G — произвольная конечная группа. Пусть C есть (d, r)-схема, вычисляющая φ: $G \times G \to G$, где $\varphi(a, b) = ab$. Тогда для выполнения вычислений схемой C требуется время τ, где*

$$\tau \geqslant \rceil \log_r 2 \rceil \log_d \alpha(G) \lceil \lceil .$$

Доказательство см. в работе [1]. ▶

Виноград указал также способ построения схемы для выполнения умножения в абелевой группе G с временем вычисления

$$\tau = 2 + \left\rceil \log_{[(r+1)/2]} \right\rceil \frac{1}{[r/2]} \rceil \log_d \alpha(G) \lceil \left[\left[,$$

причем этот способ применим при $r \geqslant 3$ и $d \geqslant 2$. Мы дадим совершенно другой метод построения схем, который применим при $r \geqslant 2$ и $d \geqslant 2$ и действует независимо от того, является ли группа абелевой или нет. Более того, для любой заданной абелевой группы и для любых заданных d и r время вычисления у нас окажется меньше, чем у Винограда.

4. НИЖНЯЯ ОЦЕНКА

В этом разделе мы дадим новую нижнюю оценку времени, требующегося для выполнения умножения в группе (d, r)-схемой, и сравним ее с оценкой Винограда. Пусть G — пронзвольная конечная группа, φ: $G \times G \to G$ — умноженне в группе и C есть (d, r)-схема, вычисляющая φ. Обозначим через $h_j(g)$ значение на j-й выходной линии схемы C при выходе, равном $h(g)$.

Определение 4.1. Пусть x, $y \in G$. Будем говорить, что x и y являются R_j-эквнвалентными, если при всех $g \in G$ выполняется равенство $h_j(xg) = h_j(yg)$, и будем говорить, что они являются L_j-эквивалентными, если при всех $g \in G$ выполняется равенство $h_j(gx) = h_j(gy)$. Ясно, что R_j и L_j являются отношениями эквнвалентности. Класс, R_j-эквивалентный элементу g, будем обозначать через $R_j(g)$, а класс, L_j-эквивалентный элементу g, — через $L_j(g)$.

Лемма 4.1. *Классы $R_j = R_j(e)$ и $L_j = L_j(e)$ являются группами для всех выходных линий схемы C. Более того, для любого $g \in G$ справедливы соотношения $R_j(g) = R_j g$ и $L_j(g) = gL_j$.*

Доказательство. Допустим, что $a, b \in R_j$. Пусть c — произвольный элемент из G. Тогда $h_j(ab^{-1}c) = h_j(bb^{-1}c) = h_j(c)$. Следовательно, $ab^{-1} \in R_j$ и R_j является группой. Возьмем теперь произвольный $g \in G$. Тогда $d \in R_j(g)$ в том и только том случае, если $h_j(dc) = h_j(gc)$ при всех $c \in G$. Но последнее имеет место тогда и только тогда, когда $h_j(dg^{-1}c) = h_j(c)$, т. е. $dg^{-1} \in R_j$. Вторая половина леммы доказывается двойственным образом. ▶

Максимальные разделнмые множества выявляются с помощью следующей леммы.

Лемма 4.2. *h_j-разделимое множество максимальной мощности в первом аргументе φ содержит в точности по одному представителю из каждого правого смежного класса по подгруппе R_j в G. Таким образом, его мощность равна $|G|/|R_j|$. Двойственный результат имеет место для разделимых множеств во втором аргументе.*

Доказательство следует непосредственно из леммы 4.1 и определения разделимых множеств. ▶

Теперь мы имеем все необходимое для получення нижней оценки, зависящей от выходного кода.

Лемма 4.3. *Пусть C есть (d, r)-схема, предназначенная для умножения в группе G за время τ. Тогда*

$$\tau \geqslant \max_j \left\lceil \log_r \left(\left\lceil \log_d \frac{|G|}{|R_j|} \right\rceil + \left\lceil \log_d \frac{|G|}{|L_j|} \right\rceil \right) \right\rceil .$$

Доказательство непосредственно следует из лемм 2.2 и 4.2. ▶

Оценка по всем кодам получается путем минимизации максимальных размеров R_j и L_j для заданной группы G.

Определение 4.2. Если $G = \{e\}$, то положим $\delta(G) = 1$. В противном случае пусть $\delta(c)$ обозначает максимальный порядок подгруппы G, не содержащей c. В этом случае положим $\delta(G) = \min_{c \in G \setminus \{e\}} \delta(c)$.

Поскольку мы имеем дело только с конечными группами, величина $\delta(G)$ всегда определена и конечна. Заметнм, что если $P(a, G) = 1$, то $\delta(a) = 1$ и, такнм образом, $\delta(G) = 1$. Отметим также, что если G — нетривиальная группа и $P(G) \neq 1$, то всегда $\delta(G) > 1$. **Впоследствии нам понадобится следующая простая лемма.**

Лемма 4.4. *Пусть H и K — подгруппы конечной группы G, такие, что $H \cap K = \{e\}$. Тогда $|H||K| \leqslant |G|$.*

Доказательство. Пусть h_1, $h_2 \in H$ и k_1, $k_2 \in K$, такие, что $h_1 k_1 = h_2 k_2$. Тогда $h_2^{-1} h_1 = k_2 k_1^{-1} \in H \cap K$. Следовательно, $h_1 = h_2$ и $k_1 = k_2$. Таким образом, $|\{hk\colon h \in H,\ k \in K\}| \geqslant |H||K|$. Но ведь $\{hk\colon h \in H,\ k \in K\}$ также является подмножеством группы G. ▶

Важнейшим для нас свойством величины $\delta(G)$ является следующее.

Лемма 4.5. *Для любой конечной группы G справедливо соотношение $\alpha(G)\delta(G) \leqslant |G|$.*

Доказательство. Если $\delta(G) = 1$, то лемма, очевидно, верна. Поэтому предположим, что $\delta(G) \neq 1$. Возьмем $H \subset G$ и $a \in H$, такие, что $a \neq e$, $P(a, H) = 1$ и $|H| = \alpha(G)$. Выберем подгруппу $K \subset G$, такую, что $a \notin K$ и $|K| = \delta(a)$. Тогда, поскольку $H \cap K$ является подгруппой группы G, не содержащей a, имеем $H \cap K = \{e\}$. Следовательно, по лемме 4.4 и с учетом того, что $\delta(G) \leqslant \delta(a)$, имеем $\alpha(G)\delta(G) \leqslant \alpha(G)\delta(a) = |H||K| \leqslant |G|$. ▶

Теперь может быть установлена универсальная нижняя оценка для произвольной (d, r)-схемы, выполняющей умножение в конечной группе G.

Теорема 4.1. *Пусть G — конечная группа, $\varphi\colon G \times G \to G$ — умножение в группе и C есть (d, r)-схема, вычисляющая φ, причем $d \geqslant 2$ и $r \geqslant 2$. Тогда, если время вычисления схемы C есть τ, то*

$$\tau \geqslant \left\lceil \log_r 2 \right\rceil \left\lceil \log_d \frac{|G|}{\delta(G)} \right\rceil \left\lceil \right\rceil .$$

Доказательство. Допустим, что $\delta(G) > 1$, и выберем $a \in G$, такой, что $\delta(a) = \delta(G)$. Найдется выходная линия схемы C, скажем j-я, такая, что $h_j(e) \neq h_j(a)$. Но тогда оба класса R_j и L_j являются подгруппами группы G, не содержащими a. Следовательно, их порядки не превосходят $\delta(G)$. Таким образом, в этом случае утверждение следует из леммы 4.3. Если $\delta(G) = 1$, то либо $G = \{e\}$, либо $|G| = \alpha(G)$. В первом из этих случаев теорема устанавливается тривиально. Во втором случае выберем $g \in G$, такой, что $P(g, G) = 1$, и возьмем выходную линию, скажем i-ю, такую, что $h_i(e) \neq h_i(g)$. Тогда $R_j = L_j = \{e\}$ и утверждение снова следует из леммы 4.3. ▶

Лемма 4.5 показывает, что эта нижняя оценка не слабее результата Винограда, сформулированного в теореме 3.2;

в действительности же, как показывает следующий пример, эта оценка сильнее.

Пример 4.1. Пусть p — нечетное простое число. Тогда существует группа с тремя образующими a, b и c и определяющими соотношениями [4, стр. 63]

$$a^p = b^p = c^p = e, \quad ab = bac, \quad ca = ac, \quad cb = bc,$$

не содержащая элементов порядка p^2. Нетрудно показать, что любая ее подгруппа порядка p^2 обязана содержать c. Таким образом, $\delta(G) = \delta(c) = p$. Но ясно, что $\alpha(G) = p$. Поэтому $\alpha(G)\,\delta(G) < |G|$. Тем не менее в одном важном случае обе оценки совпадают.

Лемма 4.6. *Пусть G — конечная абелева группа. Тогда* $\alpha(G)\,\delta(G) = |G|$.

Доказательство. Согласно теореме о разложении абелевой группы [4, стр. 50], имеем $G = Z_1 \times \ldots \times Z_n$, где каждое Z_i является циклической p-группой; пусть, например, $|Z_i| = p_i^{r_i}$. Без ограничения общности будем считать, что

$$i < j \Rightarrow p_i^{r_i} \geqslant p_j^{r_j}. \qquad (*)$$

Если $n = 1$, то теорема верна, так как $P(G) = 1$ и $\delta(G) = 1$. Допустим, что $n > 1$, и пусть a_i порождает Z_i $(i = 1, 2, \ldots, n)$. Теперь если мы выберем произвольный $g \neq e$,

$$g = \left(a_1^{k_1}, \ \ldots, \ a_n^{k_n}\right),$$

у которого **хотя бы** один из показателей **степени, скажем** k_i, отличен от **0, то оказ**ается, что

$$g \notin \left(\prod_{\substack{j \neq i \\ j=1}}^{n} Z_j\right) \times \{e_i\},$$

где e_i **— единица гру**ппы Z_i. Отсюда в силу **(*) следует, что**

$$\delta(g) \geqslant \prod_{\substack{j \neq i \\ j=1}}^{n} p_j^{r_j} \geqslant \prod_{j=2}^{n} p_j^{r_j}.$$

Таким образом,

$$\delta(G) \geqslant \prod_{j=2}^{n} p_j^{r_j}.$$

Однако любая подгруппа, порядок которой больше $\prod_{j=2}^{n} p_j^{r_j}$, обязана пересекаться с Z_1 нетривиальным образом и поэтому

должна содержать

$$\left(a_1^{p_1^{r_1-1}}, e_2, \ldots, e_n\right) \notin \{e_1\} \times Z_2 \times \ldots \times Z_n.$$

Таким образом,

$$\delta(G) \leqslant \delta\left(\left(a_1^{p_1^{r_1-1}}, e_2, \ldots, e_n\right)\right) = \prod_{j=2}^{n} p_j^{r_j}. \blacktriangleright$$

В целях полноты изложения мы приведем здесь несколько примеров неабелевых групп G_i, для каждой из которых $\alpha(G_i)\delta(G_i) = |G_i|$.

Пример 4.2. Пусть p — нечетное простое число. Пусть G_1 — группа с образующими a и b и определяющими соотношениями [4, стр. 63] $a^{p^2} = b^p = e$; $b^{-1}ab = a^{1+p}$. Тогда $\alpha(G_1) = p^2$ и любая подгруппа порядка p^2 обязана содержать a^p.

Пример 4.3. Пусть G_2 — прямое произведение двух групп A и B, таких, что $\alpha(A)\delta(A) = |A|$; $\alpha(B)\delta(B) = |B|$. Тогда нетрудно заметить, что

$$\alpha(G_2) = \max\{\alpha(A), \alpha(B)\}; \quad \delta(G_2) = \min\{|B|\delta(A), |A|\delta(B)\}$$

и, таким образом, $\alpha(G_2)\delta(G_2) = |G_2|$. В частности, эти свойства имеют место, если G_2 неабелева, но все ее подгруппы инвариантны [4, стр. 213].

5. СХЕМА ДЛЯ УМНОЖЕНИЯ В ГРУППЕ

В этом разделе мы дадим метод построения (d, r)-схемы для умножения в произвольной конечной группе G, применимый при $d \geqslant 2$ и $r \geqslant 2$. Время вычисления такой схемы самое большее на единицу превосходит только что полученную нижнюю оценку. Если G — абелева группа и $r \geqslant 3$, то нашу схему можно сравнить со схемой Винограда. Можно заметить, что время вычисления у нашей схемы меньше, чем у схемы Винограда, и что на самом деле мы можем указать группу, для которой разность между этими значениями времени произвольно велика.

Лемма 5.1. *Пусть K — произвольная подгруппа группы G. Тогда существует (d, r)-схема, вычисляющая* φ: $G \times G \to \{0, 1\}$, *где*

$$\varphi(a, b) = \begin{cases} 0, & \text{если } ab \in K, \\ 1, & \text{если } ab \notin K, \end{cases}$$

за время [1])

$$\tau = 1 + \left\rceil \log_r \left\rceil \frac{1}{[r/2]} \right\rceil \log_d \frac{|G|}{|K|} \right\rceil \left[\left[\left[\cdot \right. \right. \right.$$

Доказательство. Пусть $M = |G|/|K|$. Для каждого правого смежного класса по подгруппе K в группе G возьмем по представителю $v_i \in Kv_i$. Тогда $\{v_i^{-1}\}$ будет множеством представителей левых смежных классов, так как $v_i^{-1}K = v_j^{-1}K$ тогда и только тогда, когда $v_i v_j^{-1} \in K$. Возьмем отображение z_1 группы G в множество $\left] \log_d M \right[$-ных векторов над Z_d, такое, что $z_1(g_1) = z_1(g_2)$ тогда и только тогда, когда $Kg_1 = Kg_2$. Затем определим другое отображение z_2 посредством равенства $z_1(g) \oplus z_2(g^{-1}) = 0$, где 0 — вектор, все координаты которого равны 0, а $\oplus$ — операция покомпонентного сложения по модулю d. Заметим, что z_2 отображает любые два элемента из одного левого смежного класса в один и тот же вектор. В первом ярусе схемы содержится $\left] \frac{1}{[r/2]} \right] \log_d M \left[\right.$ однотипных элементов. При вычислении ab каждый из этих элементов складывает $[r/2]$ компонент векторов $z_1(a)$ и $z_2(b)$ по модулю d (последний сумматор складывает менее $[r/2]$ компонент, если $[r/2]$ не делит M). Выход элемента равен 0, если для всех пар входных компонент суммы сравнимы с 0 по модулю d. В противном случае выход равен 1. Таким образом, выходы всех элементов равны 0 тогда и только тогда, когда существует j, такое, что $a \in Kv_j$ и $b \in v_j^{-1}K$. Оставшаяся часть схемы представляет собой „дерево“ из элементов с r входами. Ее выход равен 0, если все входы равны 0, и равен 1, если хотя бы один вход отличен от 0. Глубина этого дерева равна $\left\rceil \log_r \left\rceil \frac{1}{[r/2]} \right] \log_d M \left[\right[\right[$. Таким образом, схема вычисляет φ за время

$$\tau = 1 + \left\rceil \log_r \left\rceil \frac{1}{[r/2]} \right] \log_d M \left[\right[\right[\cdot$$

Следствие 5.1. *Для любого $v \in G$ существует (d, r)-схема с тем же временем вычисления, которая определяет, выполняется ли соотношение $ab \in Kv$.*

Определение 5.1. Полным множеством подгрупп группы G называется такое множество $\{K_i\}$ подгрупп, для которого $\bigcap_i K_i = \{e\}$.

---

[1]) В первоначальной формулировке леммы было $\tau = 1 + \left\rceil \log_r \right\rceil \log_d \frac{|G|}{|K|} \left[\right[\cdot$ Улучшение было указано автору Виноградом.

Лемма 5.2. *Если $\{K_i\}$ — полное множество подгрупп группы G, то для любого $a \in G$ знания правых смежных классов, содержащих a, достаточно для определения a.*

Доказательство. $\bigcap_i (K_i a) = \left(\bigcap_i K_i\right) a = a$. ▶

Отметим, что полное множество подгрупп существует для любой группы G; таким, например, является множество, состоящее из единственной подгруппы $\{e\}$. Если $P(G) \neq 1$, то существуют также и другие полные множества.

Лемма 5.3. *Пусть $\{K_i\}$ — полное множество подгрупп группы G. Тогда существует (d, r)-схема, выполняющая умножение в группе G за время*

$$\tau = 1 + \max_i \left\rceil \log_r \left\rceil \frac{1}{\lceil r/2 \rceil} \right\rceil \log_d \frac{|G|}{|K_i|} \right\rceil \left[\left[\left[\cdot \right. \right. \right.$$

Доказательство следует из леммы 5.1, следствия 5.1 и леммы 5.2. ▶

Теперь мы в состоянии доказать следующую теорему.

Теорема 5.1. *Пусть G — произвольная конечная группа. Тогда для любых $d \geqslant 2$ и $r \geqslant 2$ существует (d, r)-схема, выполняющая умножение в конечной группе G за время*

$$\tau = 1 + \left\rceil \log_r \left\rceil \frac{1}{\lceil r/2 \rceil} \right\rceil \log_d \frac{|G|}{\delta(G)} \right\rceil \left[\left[\left[\cdot \right. \right. \right.$$

Более того, время вычисления схемы превосходит нижнюю оценку не более, чем на одну единицу времени.

Доказательство. Допустим, что $\delta(G) > 1$. Для любого $g \in G$, $g \neq e$, существует подгруппа K_g порядка $\delta(g)$, не содержащая g. Таким образом, $\{K_g : g \in G \setminus \{e\}\}$ есть полное множество подгрупп, для которого $\min \{|K_g| : g \in G \setminus \{e\}\} = \delta(G)$. Если же $\delta(G) = 1$, то мы пользуемся полным множеством, состоящим из $\{e\}$. Второе утверждение теоремы следует из того, что при $r \geqslant 2$ имеем

$$\left\rceil \log_r \left\rceil \frac{1}{\lceil r/2 \rceil} \right\rceil \log_d x \left[\left[\left[\leqslant \right\rceil \log_r 2 \right] \log_d x \right[\left[\cdot \right. \right. \right. \right. \text{▶}$$

Следствие 5.2. *Если G — абелева группа или если $\delta(G) = 1$, то существует (d, r)-схема, выполняющая умножение в группе G за время*

$$\tau = 1 + \left\rceil \log_r \left\rceil \frac{1}{\lceil r/2 \rceil} \right\rceil \log_d a(G) \left[\left[\left[\cdot \right. \right. \right.$$

Как уже отмечалось, схема Винограда для абелевой группы G требовала времени

$$\tau = 2 + \left\lceil \log_{[(r+1)/2]} \right\rceil \frac{1}{[r/2]} \left\rceil \log_d \alpha(G) \right[\left[\left[.$$

Поскольку при $r \geqslant 3$ имеем

$$[(r+1)/2] < r,$$

отсюда следует, что время вычисления у нас меньше, чем у Винограда.

Пример 5.1. Возьмем, скажем, $r = 4$ и $\left\rceil \log_d \alpha(G) \right[= 2^{2^k}$ для некоторого $k \geqslant 1$. Тогда время вычисления у Винограда будет равно $2 + 2k$, а у нас равно $1 + k$, т. е. его схема будет требовать вдвое большего времени. Читатель легко сможет построить самостоятельно подобные примеры.

Виноград [2] распространил свои результаты, полученные для группы, на сложение и умножение натуральных чисел, заметив, что схема, выполняющая умножение в циклической группе порядка $2N - 1$, может также складывать пару чисел, заключенных между 0 и N, и что умножение натуральных чисел может выполняться путем сложения показателей степени в разложениях обоих сомножителей на простые множители. Поскольку мы сумели понизить время, необходимое для умножения в циклических группах, мы можем также соответственно уменьшить время для сложения и умножения натуральных чисел. Мы представим этот результат на основе определений Винограда. Читателя, который заинтересуется подробностями взаимосвязи между умножением в группе и двумя другими операциями, мы отсылаем к исходной статье Винограда.

Определение 5.2. При любом целом m положим $Q_m =$ $=$ н. о. к. $\{1, 2, \ldots, m\}$ и $\gamma(N) = \min\{m: Q_m \geqslant N\}$.

Теперь, следуя методу применения верхней оценки для времени умножения в группе, предложенному Виноградом, мы воспользуемся следствием 5.2 и получим теоремы 5.2 и 5.3.

Теорема 5.2. *Пусть* φ: $Z_N \times Z_N \to Z_{2N-1}$ *есть* $\varphi(a, b) = a + b$. *Тогда существует* (d, r)*-схема, вычисляющая* φ *за время*

$$\tau_\varphi = 1 + \left\rceil \log_r \right\rceil \frac{1}{[r/2]} \left\rceil \log_d \gamma(2N-1) \right[\left[\left[;\ r \geqslant 2,\ d \geqslant 2.$$

Теорема 5.3. *Пусть* ψ: $\{1, 2, \ldots, N\} \times \{1, 2, \ldots, N\} \to$ $\to \{1, 2, \ldots, N^2\}$ *есть* $\psi(a, b) = ab$. *Тогда для любых* $r \geqslant 2$ *и* $d \geqslant 2$ *существует* (d, r)*-схема, вычисляющая* ψ *за время*

$$\tau_\psi = 1 + \left\rceil \log_r \right\rceil \frac{1}{[r/2]} \left\rceil \log_d \gamma(2\,[\log_2 N] - 1) \right[\left[\left[.$$

В заключение для справки отметим, что Виноград оценил снизу τ_φ и τ_ψ следующим образом.

Теорема 5.4 (Виноград [2]). *При любых* $d \geqslant 2$ *и* $r \geqslant 2$ *для любой* (d, r)*-схемы, вычисляющей* φ, *требуется время* τ_φ, *где*

$$\tau_\varphi \geqslant \left\rceil \log_r 2 \right\rceil \log_d \gamma\left(\left\rceil \frac{N}{2} \right\lfloor \right) \left[\left[\right.,$$

а для любой (d, r)*-схемы, вычисляющей* ψ, *требуется время*

$$\tau_\psi \geqslant \left\rceil \log_r 2 \right\rceil \log_d \gamma\left(\left\rceil \frac{[\log_2 2N]}{2} \right\lfloor \right) \left[\left[\right..$$

О близости результатов, сформулированных в теоремах 5.2 и 5.3, к этим нижним оценкам говорит тот факт, что $\gamma(4x) \leqslant \leqslant 2 + \gamma(x)$.

Автор выражает глубокую признательность за помощь своему научному консультанту проф. М. А. Арбибу, а также благодарит д-ра С. Винограда за весьма полезные обсуждения.

ЛИТЕРАТУРА

1. Winograd S., On the time required to perform addition, *J. ACM*, 12, № 2 (1965), 277—285. (Русский перевод: Виноград С., О времени, требующемся для выполнения сложения, Кибернетический сборник, новая серия, вып. 6, «Мир», М., 1969, стр. 41—54.)
2. Winograd S., On the time required to perform multiplication, *J. ACM*, 14, № 4 (1967), 793—802. (Русский перевод: Виноград С., О времени, требующемся для выполнения умножения, Кибернетический сборник, новая серия, вып. 6, «Мир», М., 1969, стр. 55—71.)
3. Офман Ю. П., Об алгоритмической сложности дискретных функций, *ДАН СССР*, 145, № 1 (1962), 48—51.
4. Холл М., Теория групп, ИЛ, М., 1962.

Алгоритм Гаусса не оптимален[1])

В. Штрассен

1. Ниже мы приведем алгоритм для вычисления элементов произведения двух квадратных матриц A и B порядка n по заданным элементам матриц A и B с числом арифметических операций, меньшим чем $4{,}7 \cdot n^{\log 7}$ (все логарифмы в этой статье берутся по основанию 2, так что $\log 7 \approx 2{,}8$; обычный метод требует приблизительно $2n^3$ арифметических операций). Этот алгоритм индуцирует алгоритмы для обращения матрицы, для решения системы n линейных уравнений с n неизвестными, для вычисления детерминанта порядка n и т. д. Все они требуют меньше чем $\text{const} \cdot n^{\log 7}$ арифметических операций[2]).

Этот результат следует сопоставить с результатом В. В. Клюева и Н. И. Коковкина-Щербака [1] о том, что алгоритм Гаусса для решения систем линейных уравнений является оптимальным, если допускаются лишь операции над целыми строками и столбцами. Заметим также, что Виноград [2] модифицирует обычные алгоритмы для умножения матриц, обращения их и решения систем линейных уравнений, используя около половины умножений по сравнению со сложениями и вычитаниями [3]).

2. Определим алгоритмы $\alpha_{m,k}$, которые перемножают матрицы порядка $m2^k$, индукцией по k: $\alpha_{m,0}$ — это обычный алгоритм для умножения матриц, требующий m^3 умножений и $m^2(m-1)$ сложений. Если $\alpha_{m,k}$ уже определен, то $\alpha_{m,k+1}$ определим следующим образом.

---

[1]) Strassen V., Gaussian elimination is not optimal, *Numerische Mathematik*, **13**, H. 4 (1969), 354—356.

[2]) Все результаты работы сохраняются для операций с матрицами над произвольным (в частности, конечным) полем. — *Прим. ред.*

[3]) См. также работы: Клосс Б. М., Оценки сложности решения систем линейных уравнений, *ДАН СССР*, **171**, № 4 (1966), 781—783; Коновальцев И. В., Об одном алгоритме решения систем линейных уравнений в конечных полях, сб. „Проблемы кибернетики", вып. 19, 1967, стр. 269—274. — *Прим. ред.*

Для умножения матриц A и B порядка $m2^{k+1}$ запишем их в виде

$$A=\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \quad B=\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}, \quad AB=\begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix},$$

где A_{ij}, B_{ij}, C_{ij} – матрицы порядка $m2^k$. Используя алгоритм $\alpha_{m,k}$ для умножения и обычные алгоритмы для сложения и вычитания матриц порядка $m2^k$, вычислим:

$$\begin{aligned} \mathrm{I} &= (A_{11}+A_{22})(B_{11}+B_{22}), \\ \mathrm{II} &= (A_{21}+A_{22})B_{11}, \\ \mathrm{III} &= A_{11}(B_{12}-B_{22}), \\ \mathrm{IV} &= A_{22}(-B_{11}+B_{21}), \\ \mathrm{V} &= (A_{11}+A_{12})B_{22}, \\ \mathrm{VI} &= (-A_{11}+A_{21})(B_{11}+B_{12}), \\ \mathrm{VII} &= (A_{12}-A_{22})(B_{21}+B_{22}), \\ C_{11} &= \mathrm{I}+\mathrm{IV}-\mathrm{V}+\mathrm{VII}, \\ C_{21} &= \mathrm{II}+\mathrm{IV}, \\ C_{12} &= \mathrm{III}+\mathrm{V}, \\ C_{22} &= \mathrm{I}+\mathrm{III}-\mathrm{II}+\mathrm{VI}. \end{aligned}$$

Индукцией по k легко установить следующие утверждения.

Утверждение 1. *Алгоритм $\alpha_{m,k}$ вычисляет произведение двух матриц порядка $m2^k$ с $m^3 7^k$ умножениями и $(5+m)m^2 7^k - 6(m2^k)^2$ сложениями и вычитаниями чисел.*

Таким образом, умножение двух матриц порядка 2^k возможно с 7^k умножениями и менее чем $6\cdot 7^k$ сложениями и вычитаниями.

Утверждение 2. *Произведение двух матриц порядка n можно вычислить менее чем с $4{,}7\cdot n^{\log 7}$ арифметическими операциями.*

Доказательство. Положим

$$k=[\log n-4], \quad m=[n2^{-k}]+1.$$

Тогда

$$n\leqslant m2^k.$$

Погружая матрицы порядка n в матрицы порядка $m2^k$, сводим нашу задачу к оценке числа операций в алгоритме $\alpha_{m,k}$.

Из утверждения 1 следует, что это число равно

$$(5+2m)\,m^2 7^k - 6\,(m2^k)^2 <$$
$$< (5+2(n2^{-k}+1))(n2^{-k}+1)^2 7^k <$$
$$< 2n^3(7/8)^k + 12{,}03n^3(7/4)^k \leqslant$$

(здесь мы пользуемся тем, что $16 \cdot 2^k \leqslant n$)

$$\leqslant (2(8/7)^{\log n-k} + 12{,}03(4/7)^{\log n-k})\,n^{\log 7} \leqslant$$
$$\leqslant \max_{4 \leqslant t \leqslant 5} (2(8/7)^t + 12{,}03(4/7)^t)\,n^{\log 7} \leqslant$$
$$\leqslant 4{,}7 \cdot n^{\log 7}$$

из соображений выпуклости.

Перейдем теперь к обращению матриц. Для приложимости наших алгоритмов требуется предположить не только обратимость матрицы, но и то, что появляющиеся деления имеют смысл (аналогичное предположение необходимо, конечно, и для алгоритма Гаусса).

Мы определим алгоритмы $\beta_{m,k}$, которые обращают матрицы порядка $m2^k$, индукцией по k: $\beta_{m,0}$ — это обычный алгоритм Гаусса; алгоритм $\beta_{m,k}$ уже определен; теперь определим $\beta_{m,k+1}$ следующим образом.

Если A — матрица порядка $m2^{k+1}$, для которой необходимо найти обратную, то запишем

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \quad A^{-1} = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix},$$

где A_{ij}, C_{ij} — матрицы порядка $m2^k$. Используя $\alpha_{m,k}$ для умножения, $\beta_{m,k}$ для обращения и обычные алгоритмы для сложения и вычитания матриц порядка $m2^k$, вычисляем:

$$\begin{aligned} \mathrm{I} &= A_{11}^{-1}, & C_{12} &= \mathrm{III} \cdot \mathrm{VI}, \\ \mathrm{II} &= A_{21}\,\mathrm{I}, & C_{21} &= \mathrm{VI} \cdot \mathrm{II}, \\ \mathrm{III} &= \mathrm{I}\,A_{12}, & \mathrm{VII} &= \mathrm{III} \cdot C_{21}, \\ \mathrm{IV} &= A_{21}\,\mathrm{III}, & C_{11} &= \mathrm{I} - \mathrm{VII}, \\ \mathrm{V} &= \mathrm{IV} - A_{22}, & C_{22} &= -\,\mathrm{VI}. \\ \mathrm{VI} &= \mathrm{V}^{-1}, \end{aligned}$$

Индукцией по k легко **установить следующее утверждение.**

Утверждение 3. *Алгоритм* $\beta_{m,k}$ *вычисляет обратную матрицу к матрице порядка* $m2^k$ *с* $m2^k$ *делениями,* $\leqslant \left(\frac{6}{5}m^3 7^k - m2^k\right)$ *умножениями и* $\leqslant \left[\frac{6}{5}(5+m)\,m^2 7^k - 7(m2^k)^2\right]$ *сложениями и вычитаниями чисел.*

Следующее утверждение доказывается почти так же, как и утверждение 2.

Утверждение 4. *Обращение матрицы порядка* n *возможно с не более чем* $5{,}64n^{\log 7}$ *арифметическими операциями.*

Аналогичные результаты справедливы для решения систем линейных уравнений и вычисления детерминанта, если использовать равенство

$$\operatorname{Det} A = (\operatorname{Det} A_{11}) \operatorname{Det}\left(A_{22} - A_{21}A_{11}^{-1}A_{12}\right).$$

ЛИТЕРАТУРА

1. Клюев В. В., Коковкин-Щербак Н. И., О минимизации числа арифметических операций при решении линейных алгебраических систем уравнений, *Журнал вычисл. матем. и матем. физики*, 5, № 1 (1965), 21—23.
2. Winograd S., A new algorithm for inner product, IBM Research Report RC-1943, Nov. 21, 1967.

Некоторые аспекты последовательностного построения экспериментов [1])

Г. Роббинс

1. ВВЕДЕНИЕ

До недавнего времени в статистической теории ограничивались построением и анализом экспериментов, в которых количество и способы выбора испытаний полностью определялись до начала эксперимента. Причины этого частично имеют историческое происхождение и относятся ко времени, когда статистика проводилась только после окончания эксперимента, если это вообще было возможно, и частично связаны с математическими трудностями, которые возникают при работе с произвольным, но фиксированным числом независимых случайных величин. Основной сдвиг в этом направлении, по-видимому, состоит в создании теории *последовательностного конструирования* экспериментов, в которых количество и способы выбора испытаний не фиксированы заранее, а зависят от предшествующих испытаний.

Первое важное отступление от фиксирования количества испытаний появилось при контроле качества в промышленности (метод контроля Доджа и Ромига, связанный с двойной выборкой [1]). В данном случае выбирается только одна совокупность изделий и вопрос заключается в том, превышает ли количество брака в этой совокупности данный уровень. Сначала выбирается произвольно n_1 объектов из совокупности и через x обозначается число бракованных объектов. Если x меньше некоторого фиксированного значения a, то вся совокупность изделий принимается без дальнейшей проверки; если x больше некоторого фиксированного значения b $(a < b)$, то вся совокупность бракуется; однако если $a \leqslant x \leqslant b$, то второй раз выбирается n_2 изделий из совокупности, и заключение принимать или браковать всю совокупность делается на основании числа бракованных изделий в общем количестве выбранных $n_1 + n_2$ объектов. Таким образом, число n выбранных объектов из совокупности есть случайная величина, принимающая два значения n_1 и $n_1 + n_2$ и зависящая от испытаний.

---

[1]) Robbins H., Some aspects of the sequential design of experiments, *Bull. Amer. Math. Soc.*, 58. № 5 (1952), 527–535.

Во время второй мировой войны было развито направление, главным образом Вальдом, получившее в дальнейшем название *последовательностного анализа* [2], в основе которого лежит идея, обобщающая идею двойной выборки. Грубо говоря, она состоит в том, что испытания проводятся друг за другом, и заключение о том, принять или забраковать данную совокупность (или, в более общем виде, сомнительную статистическую гипотезу), может быть сделано на любом шаге. Количество испытаний n становится теперь случайной величиной, связанной с данной совокупностью и принимающей в принципе бесконечно много значений, хотя в большинстве практических задач оно может быть ограничено. Преимущество этого метода состоит в том, что при некоторых обстоятельствах разумный выбор последовательности испытаний может привести к значительному уменьшению среднего числа испытаний, необходимого для того, чтобы вероятность ошибки была минимальной. Теория последовательностного анализа еще очень несовершенна [1]), и нужно преодолеть значительные трудности, чтобы использовать ее для решения проблем статистики.

Использование последовательностных методов при выборе способа испытаний освобождает статистическую теорию от ограничений, связанных с фиксацией количества испытаний. Однако не только *количество* испытаний характеризует эффективную конструкцию эксперимента. Большинство статистических проблем, встречающихся на практике, связано с рассмотрением нескольких совокупностей, и нужно определить, *из какой совокупности* производить выборку на каждом шаге.

Следующий пример иллюстрирует сказанное. Пусть имеются две совокупности, соответствующие случайные величины которых нормально распределены. Пусть неизвестны математические ожидания μ_1 и μ_2 и дисперсии σ_1^2 и σ_2^2 этих случайных величин. Требуется оценить значение разности $\mu_1 - \mu_2$. Чтобы подчеркнуть обсуждаемую точку зрения, предполагается, что n фиксировано. Возникает вопрос, как распределить n испытаний между двумя совокупностями. Пусть $\bar{x}_1$ и $\bar{x}_2$ суть средние арифметические значения случайных величин, получающихся при n_1 и n_2 испытаниях с двумя соответствующими совокупностями, тогда естественно считать, что $\bar{x}_1 - \bar{x}_2$ является оценкой для $\mu_1 - \mu_2$ с дисперсией $\sigma^2 = (\sigma_1^2/n_1) + (\sigma_2^2/n_2)$. Для фиксированного $n = n_1 + n_2$ дисперсия σ^2 достигает минимума при $n_1/n_2 = \sigma_1/\sigma_2$. Если σ_1/σ_2 известно, то сразу строится соответствующий эксперимент;

---

[1]) Нужно **иметь в виду, что настоящая статья была опубликована** в 1952 г. – *Прим. ред.*

если же это отношение неизвестно, то процесс испытаний может быть разбит на два этапа. Производится некоторое число m испытаний с каждой совокупностью, и таким способом полученные значения случайных величин используются в качестве оценки для σ_1/σ_2; оставшиеся $n-2m$ испытаний распределяются между двумя совокупностями в соответствии с оценкой для σ_1/σ_2. Возникает вопрос, как лучше выбрать m. Если m мало, то получается грубая оценка для σ_1/σ_2. При достаточно больших m малость разности $n-2m$ не позволяет полностью использовать полученную оценку σ_1/σ_2. (Подобная дилемма весьма характерна для всех задач последовательностного конструирования.) Вообще говоря, возможно рассмотрение схем, в которых испытания проводятся одно за другим, и заключение, с какой совокупностью производить каждое испытание, делается на основании предыдущих испытаний; общее количество испытаний n может быть фиксированным или может быть случайной величиной, зависящей от проведенных испытаний.

Несмотря на полное отсутствие теории, Махалонобис [3] в 1938 г. первым попытался в духе последовательностных построений определить в Бенгалии площади, пригодные для выращивания джута. Первоначальное исследование было проведено в малом масштабе, чтобы оценить значения определенных параметров, знание которых было весьма существенно для перехода к эффективным исследованиям в большом масштабе.

В последующих публикациях Махаланобис [4] обратил внимание на то, что было бы желательно иметь способ изменения эксперимента по мере накопления новых данных.

Мы обязаны Вальду за большой вклад в теорию последовательностного анализа. В его книге [5] ставится проблема в самом общем виде и очерчиваются методы ее решения. Затронутые в этой книге вероятностные проблемы обширны, поскольку вероятностные зависимости встречаются в ней во всей их сложности, в то время как еще не найдены удовлетворительные средства для исследования практически интересных задач. Тем не менее достаточно явно подтверждается предположение, что дальнейшие результаты в теории последовательностного конструирования будут весьма важны в математической статистике и в науке в целом.

Ниже будет рассмотрено несколько простых задач последовательностного конструирования, которые теперь привлекают внимание и которые отличаются от тех задач, которые обычно встречаются в статистической литературе. Оптимальное решение этих задач неизвестно. Однако часто лучше иметь достаточно хорошее решение нужной задачи, чем оптимальное ре-

шение ненужной. В настоящее время этот принцип с особенной пользой применяется в проблемах последовательностного экспериментирования.

2. ПРОБЛЕМА ВЫБОРА ИЗ ДВУХ МНОЖЕСТВ

Пусть A и B — два множества (монеты, урны, промышленные изделия и т. д.) с заданными на них функциями распределения $F(x)$ и $G(x)$. Об этих функциях известно только, что они принадлежат некоторому классу D. Мы будем предполагать, что интегралы

$$\alpha = \int_{-\infty}^{\infty} x\,dF(x), \quad \beta = \int_{-\infty}^{\infty} x\,dG(x) \tag{1}$$

существуют. *Возникает вопрос, как сделать выборку $x_1, x_2, \ldots, x_n$ из двух указанных множеств, чтобы сумма $S_n = x_1 + \ldots + x_n$ достигала наибольшей возможной величины.*

Например, пусть A и B — две произвольные монеты, и предположим, что мы совершаем n бросаний, причем при выпадении орла мы получаем выигрыш в 1 единицу, а при выпадении решетки не получаем ничего. Если $x_i = 1$ или $x_i = 0$ соответствуют выпадению орла или решетки при i-м бросании, то S_n — общая сумма, которую мы можем получить, а α и β ($0 \leqslant \alpha, \beta \leqslant 1$) суть вероятности выпадения орла при единичном бросании монет A и B соответственно.

Интуитивно ясно, что всякий раз, когда на основании предыдущих испытаний возникает уверенность, что одно из двух чисел α или β больше другого, в будущем следует предпочесть соответствующее множество. Заметим, что мы при этом не интересуемся оценкой разности $\alpha - \beta$. Вся задача состоит в том, *какое множество A или B* выбрать для i-го испытания. Безусловно, существуют практические ситуации, в которых задача ставится скорее в процессе эксперимента, чем до него. Фактически проблема представляет собой в упрощенном виде общую задачу о том, как мы учимся или как бы мы могли учиться на основании прошлого опыта. По-видимому, решение этой проблемы сводится к построению средствами математической статистики конструкции, которая была названа фон Нейманом [6] *целесообразным поведением.*

Начнем с рассмотрения уже упомянутого специального случая, в котором множества A и B суть монеты, а соответствующие вероятности α и β выпадения орла ($x_i = 1$) при единичном бросании неизвестны. Возьмем, например, следующее правило построения последовательности испытаний.

Правило R_1. Для первого бросания выбираем A или B произвольно. Затем, если при i-м бросании выпал орел, то при $(i+1)$-м бросании используем ту же монету, что и при i-м бросании; если же при i-м бросании выпала решетка, то для $(i+1)$-го бросания берем другую монету; здесь $i = 1, 2, \ldots$.

Чем можно охарактеризовать правило R_1? Последовательность бросаний можно описать простой цепью Маркова с четырьмя состояниями (A, H), (A, T), (B, H), (B, T)[1]) и с вероятностными переходами, которые легко записываются. Например, вероятность перехода из (A, H) в (A, T) равна $(1-\alpha)$ при $(i+1)$-м бросании. Пусть p_i обозначает вероятность выпадения орла при i-м бросании. Исключая тривиальный случай, будем предполагать, что α и β обе не равны 0 и обе не равны 1; тогда $|\alpha+\beta-1|<1$. Легко показать, что

$$p_{i+1} = (\alpha+\beta-1)\,p_i + (\alpha+\beta-2\alpha\beta), \tag{2}$$

отсюда следует, что

$$p_i = (\alpha+\beta-1)^{i-1}\left[p_1 - \frac{\alpha+\beta-2\alpha\beta}{2-(\alpha+\beta)}\right] + \frac{\alpha+\beta-2\alpha\beta}{2-(\alpha+\beta)}, \tag{3}$$

и, следовательно,

$$\lim_{i\to\infty} p_i = \frac{\alpha+\beta-2\alpha\beta}{2-(\alpha+\beta)} = \gamma + \frac{\delta^2}{1-\gamma}, \tag{4}$$

где

$$\gamma = \frac{\alpha+\beta}{2}, \quad \delta = \frac{|\alpha-\beta|}{2}. \tag{5}$$

Используя правило R_1, получаем

$$\lim_{n\to\infty} E\left(\frac{S_n}{n}\right) = \lim_{n\to\infty} \frac{p_1 + \ldots + p_n}{n} = \gamma + \frac{\delta^2}{1-\gamma}. \tag{6}$$

Теперь, если *известно*, какое из двух чисел α или β больше, то, используя только ту монету, у которой вероятность выпадения орла больше, получаем

$$E\left(\frac{S_n}{n}\right) = \max(\alpha, \beta) = \gamma + \delta. \tag{7}$$

Таким образом, разность

$$L(A, B, R_1) = (\gamma+\delta) - \left(\gamma + \frac{\delta^2}{1-\gamma}\right) = \delta\left(1 - \frac{\delta}{1-\gamma}\right) \geqslant 0 \tag{8}$$

может быть принята за меру асимптотической *потери* при использовании правила R_1, *которая обусловлена незнанием истинных свойств монет*. Легко видеть, что $L(A, B, R_1)$ принимает

[1]) Здесь H (от англ. head) — орел, T (от англ. tail) — решетка. — *Прим. ред.*

свое максимальное значение $M_1 = 3 - (\sqrt{2})^3 \cong 0{,}172$, когда $\alpha = 0$ и $\beta = 2 - \sqrt{2} \cong 0{,}586$, или наоборот. Таким образом, при использовании правила R_1 для больших n потеря при одном бросании не превышает 0,172 единицы. Возможно также использование правила R_0, которое состоит в том, что на каждом шаге монета выбирается произвольно (или монеты A и B бросаются поочередно). Соответствующее значение $L(A, B, R_0)$ как легко видеть, равно величине $(\gamma + \delta) - \gamma = \delta$, которая принимает свое максимальное значение $M_0 = 1/2$ при $\alpha = 0$ и $\beta = 1$ или наоборот. Ясно, что правило R_1 значительно „лучше", чем R_0.

Правило R_0 таково, что выбор монеты при i-м бросании не зависит от результатов предыдущих бросаний, в то время как при использовании правила R_1 этот выбор зависит только от $(i-1)$-го бросания. В более общем случае можно считать, что правило R таково, что выбор монеты при i-м бросании зависит от значений $x_1, \ldots, x_{i-1}$ для всех предыдущих бросаний. Для любого такого правила R пусть

$$L_n(A, B, R) = \max(\alpha, \beta) - E\left(\frac{S_n}{n}\right), \tag{9}$$

где E обозначает математическое ожидание, вычисленное на основе α, β и R, и пусть

$$M_n(R) = \max_{(\alpha, \beta)} [L_n(A, B, R)], \tag{10}$$

$$\varphi(n) = \min_{(R)} [M_n(R)]. \tag{11}$$

Интересно изучить поведение $\varphi(n)$ и описать „минимаксное" правило R, для которого значение $\varphi(n)$ достигается.

Более простая задача: существует ли правило R, такое, что

$$\lim_{n \to \infty} L_n(A, B, R) = 0 \tag{12}$$

для всяких A и B? Мы увидим в следующем разделе, что ответ на этот вопрос положителен не только в случае монет, но и в случае двух произвольных множеств.

Возвращаясь к общему случаю, в котором A и B — произвольные множества, для которых значения (1) существуют, рассмотрим правило $\bar{R}$ проведения испытаний, определенное следующим образом. Пусть

$$\begin{aligned} 1 = a_1 < a_2 < \ldots < a_n < \ldots, \\ 2 = b_1 < b_2 < \ldots < b_n < \ldots \end{aligned} \tag{13}$$

— две фиксированные несовпадающие возрастающие последовательности натуральных чисел плотности 0, т. е. такие, доля которых в натуральном ряду стремится к 0. Правило $\bar{R}$ определим индуктивно: если i — одно из чисел a_n, то соответствующее значение x_i получается как результат испытания с A; если i — одно из чисел b_n, то x_i получается как результат испытания с B; если i не совпадает ни с одним из чисел a_n или b_n, то x_i получается как результат испытания с A или B в соответствии с тем, превышает или нет среднее арифметическое всех предыдущих испытаний с A среднее арифметическое всех предыдущих испытаний с B. Можно показать, что из усиленного закона больших чисел следует, что с вероятностью 1 имеет место равенство

$$\lim_{n\to\infty} \frac{S_n}{n} = \max(\alpha, \beta). \tag{14}$$

Это в свою очередь означает, что

$$\lim_{n\to\infty} E\left(\frac{S_n}{n}\right) = \max(\alpha, \beta), \tag{15}$$

так что

$$\lim_{n\to\infty} L_n(A, B, \bar{R}) = \max(\alpha, \beta) - \lim_{n\to\infty} E\left(\frac{S_n}{n}\right) = 0 \tag{16}$$

для всех A, B, для которых существуют α, β.

3. НЕКОТОРЫЕ ДРУГИЕ ПРОБЛЕМЫ ПОСЛЕДОВАТЕЛЬНОСТНОГО КОНСТРУИРОВАНИЯ

Задачи, рассмотренные в разд. 2, могут быть обобщены различными способами. В некоторых случаях можно считать, что общее число испытаний n является случайной величиной, как зависящей, так и не зависящей от предыдущих испытаний. Например, в задаче о двух монетах предположим, что следует платить некоторую сумму c за возможность произвести каждое бросание. В этом случае можно прекращать бросание всякий раз, когда кажется достаточно достоверным, что $\max(\alpha, \beta) < c$; это относится к специальному случаю задачи для трех множеств. Возможно рассмотрение случая континуума множеств. Предположим, что мы осуществляем уход за растениями или животными с интенсивностью θ из некоторого интервала, и пусть $F(x, \theta)$ — функция распределения реакций x на уход с интенсивностью θ. Математическое ожидание

$$\alpha(\theta) = \int_{-\infty}^{\infty} x\, dF(x, \theta) \tag{17}$$

„регрессии“ x от θ предполагается неизвестным. Пусть $\{\theta_i\}$ обозначает произвольную последовательность значений θ, выбранных последовательно экспериментатором, и пусть $\{x_i\}$ обозначает соответствующую последовательность реакций, причем каждое x_i имеет распределение

$$\Pr[x_i \leqslant x] = F(x, \theta_i).$$

(I) Пусть $\alpha(\theta)$ имеет единственный максимум в некоторой неизвестной точке θ_0. Спрашивается, каким образом экспериментатору выбирать последовательность $\{\theta_i\}$, чтобы максимизировать математическое ожидание суммы $S_n = x_1 + \ldots + x_n$ и тем самым оценить значение θ_0.

(II) Пусть $\alpha(\theta)$ — возрастающая функция от θ, которая принимает заданное значение α_0 в некоторой неизвестной точке θ_0. Спрашивается, каким образом экспериментатору выбрать последовательность $\{\theta_i\}$, для того чтобы оценить значение θ_0. Проблема (I) по существу состоит в экспериментальном определении максимума функции, когда испытания подвержены случайным ошибкам. Проблема (II) является основной при изучении чувствительности и в биологии.

Ясно, что в каждой из этих проблем выбор θ_i может зависеть от реакций $x_1, \ldots, x_{i-1}$ на предыдущие уровни $\theta_1, \ldots, \theta_{i-1}$ ухода. Таким образом, возникает задача последовательностного построения эксперимента. Изучение проблемы (I), но не в терминах последовательностного построения экспериментов, было предпринято Хотлингом [7, 8]. В то время по теории последовательностного анализа, однако, еще не было публикаций. Проблема (II) была рассмотрена Роббинсом и Монро [9]. Они предложили следующий метод. Пусть $\{a_n\}$ последовательность положительных чисел, такая, что

$$\sum_1^\infty a_n^2 < \infty, \quad \sum_1^\infty a_n = \infty. \tag{18}$$

Выберем произвольно θ_1 и положим

$$\theta_{n+1} = \theta_n + a_n(\alpha_0 - x_n) \quad (n = 1, 2, \ldots). \tag{19}$$

Тогда при некоторых слабых ограничениях на $F(x, \theta)$ может быть показано, что в вероятностном смысле имеет место соотношение

$$\lim_{n\to\infty} \theta_n = \theta_0. \tag{20}$$

В этой и других задачах может быть предложена последовательностная конструкция с достаточно хорошими свойствами, которая, возможно, встретит положительный отклик среди читателей. Это будет способствовать применению вероятностных

методов для нахождения эмпирических приближений путем построения последовательности экспериментов, когда полное математическое решение задачи затруднено. Эмпирическое исследование скорости сходимости в (20) было выполнено Техроевым [10].

4. ПРОБЛЕМА НЕПРЕДВИДЕННОЙ ОСТАНОВКИ

Пусть случайная величина x имеет нормальное распределение с неизвестным математическим ожиданием θ и единичной дисперсией. Пусть испытываются две гипотезы: $\theta = 0$ (гипотеза H_0) и $\theta > 0$ (гипотеза H_1). Стандартный статистический прием, основанный на фиксированном количестве испытаний n, состоит в следующем. Пусть $S_n = x_1 + \ldots + x_n$. Гипотеза H_0 отвергается в пользу H_1 тогда и только тогда, когда

$$S_n > \alpha n^{1/2}, \tag{21}$$

где α — некоторая константа. Вероятность отвергания H_0, если эта гипотеза верна, равна

$$\varepsilon(\alpha) = 1 - \Phi(\alpha), \tag{22}$$

где

$$\Phi(x) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{+\infty} e^{-t^2/2} dt, \tag{23}$$

причем выбирая величину α достаточно большой, можно величину $\varepsilon(\alpha)$ сделать как угодно малой. Например, если $\alpha = 3{,}09$, то $\varepsilon(\alpha) \cong 0{,}001$.

Предположим теперь, что гипотеза H_0 верна, но „экспериментатор" желает получить некоторую статистику для того, чтобы на ее основе отвергнуть H_0. Если количество испытаний не было оговорено заранее, то при выполнении неравенства (21) „экспериментатор" мог бы закончить эксперимент. Закон повторного логарифма в теории вероятностей означает, что с вероятностью 1 неравенство (21) будет справедливо для бесконечно большого числа значений n, если испытания продолжаются неограниченно, вне зависимости от величины α. Следовательно, „экспериментатор" должен стремиться достичь в конечном счете такого n, когда (21) имеет место и, останавливая эксперимент в этом месте, он получит нужный ему результат. Это обстоятельство показывает, что выполнение условия (21) не гарантирует проверки на истинность гипотезы H_0, так как существует некоторая вероятность того, что может иметь место непредвиденная остановка.

Простейший способ избежать эффекта непредвиденной остановки состоит в том, чтобы количество испытаний было

зафиксировано до начала эксперимента. Это ограничение было бы слишком жестким для практического использования. Статистики могли бы поэтому удовлетвориться установлением границ $n_1 \leqslant n \leqslant n_2$ для числа испытаний, которые будут достаточно гибкими, чтобы предотвратить случайности при экспериментировании, и довольно строгими, чтобы устранить нежелательный эффект непредвиденной остановки.

С этой целью статистикам было бы полезно знать значение функции

$$g(n_1, n_2, \alpha) = \Pr[S_n > \alpha n^{1/2}] \tag{24}$$

для некоторых $n_1 \leqslant n \leqslant n_2$, где x_i есть независимая и нормально распределенная на (0, 1) случайная величина. Совсем легко установить справедливость неравенства

$$g(n_1, n_2, \alpha) < \frac{1-\Phi(\alpha)}{1-\Phi\left(\alpha\dfrac{\lambda^{1/2}-1}{(\lambda-1)^{1/2}}\right)}, \quad \text{где} \quad \lambda = \frac{n_2}{n_1}, \tag{25}$$

которое полезно в случае, когда λ не очень велико; более точные оценки также могут быть получены.

Проблема непредвиденной остановки мало освещена в статистической теории (см. тем не менее [11], особенно стр. 286—292). Нет нужды, однако, предполагать, что экспериментатор сознательно выбирает неверную статистику, и, по мнению автора, было бы весьма желательно создание методов статистического анализа, нечувствительных к эффекту непредвиденной остановки.

ЛИТЕРАТУРА

1. Dodge H. F., Romig H. G., A method of sampling inspection, *BSTJ*, 8 (1929), 613—631; Single sampling and double sampling inspection tables, *ibid.*, 20 (1941), 1—61.
2. Wald A., Sequential analysis, New York, Wiley, 1947.
3. Mahalanobis P. C., A sample survey of the acreage under jute in Bengal with discussion of planning of experiments, *Snakhya*, 4 (1940), 511—531.
4. Mahalanobis P. C., On large-scale sample surveys, *Philos. Trans. Roy. Soc. London*, Ser. B, 231 (1944), 329—451.
5. Wald A., Statistical decision functions, New York, Wiley, 1950.
6. Neyman J., First course in probability and statistics, New York, Holt, 1950.
7. Hotelling H., Experimental determination of the maximum of a function, *Ann. of Math. Statist.*, 12 (1941), 20—45.
8. Friedman M., Savage L. J., Planning experiments seeking maxima, Chap. 3 of Selected techniques of statistical analysis, New York, McGraw-Hill, 1947.
9. Robbins H., Monro S., A stochastic approximation method, *Ann. of Math. Statist.*, 22 (1951), 400—407.
10. Teichroew D., не опубликовано.
11. Feller W., Statistical aspects of ESP, *Journal of Parapsychology*, 4 (1940), 271—298.

Построение последовательностных экспериментов с конечной памятью[1])

Г. Роббинс

Рассматривается задача последовательностного выбора одного из двух возможных способов действий, каждое из которых может привести к выигрышу или проигрышу. Спрашивается, как при этом максимизировать количество полученного выигрыша; выбор действий каждый раз основывается на результатах конечного числа предыдущих испытаний.

ВВЕДЕНИЕ

Экспериментатор имеет две монеты, монету 1 и монету 2, с соответствующими вероятностями выпадения орлов, равными $p_1 = 1 - q_1$ и $p_2 = 1 - q_2$, значения которых ему неизвестны. Он хочет произвести бесконечное число бросаний, используя в каждом бросании или монету 1, или монету 2 таким образом, чтобы получить максимально возможное число выпадения орлов. Задача экспериментатора заключается в том, чтобы найти правило выбора монеты для бросания на каждом шаге, основанное на результатах предыдущих бросаний.

Если бы он знал вначале, какое из чисел p_1 или p_2 больше, то он мог бы использовать только соответствующую монету, независимо от результатов предыдущих бросаний. В этом случае с вероятностью 1 имело бы место соотношение

$$\lim_{n\to\infty} \frac{\text{число орлов в первых } n \text{ испытаниях}}{n} = \max(p_1, p_2). \qquad (1)$$

В предыдущей статье[2]) [1] было показано, что даже для неизвестных p_1 и p_2 существует правило проведения бросаний, такое, что равенство (1) справедливо для любых p_1 и p_2. Однако любое правило, обладающее этим свойством, очевидно, должно быть таким, чтобы решение вопроса, какую монету использовать при n-м бросании, зависело от результатов всех предыдущих $n-1$ бросаний; другими словами, правило требует

[1]) Robbins H., A sequential decision problem with a finite memory, *Proc. Nat. Acad. Sci. USA*, 12 (1956), 920–923.

[2]) См. другую статью автора в этом же сборнике на стр. 71 — *Прим. ред.*

бесконечной „памяти“. В этой статье мы будем интересоваться тем, какие правила можно получить, если использовать „память“ *конечной* длины; будем говорить, что правило имеет тип r, если заключение, какую монету использовать при n-м бросании зависит от результатов только $n-r$, $n-(r+1)$, ..., $n-1$ предыдущих бросаний. (Под „результатом“ бросания мы подразумеваем и выбор монеты, и сообщение о том, какая сторона монеты выпала.)

Мы рассмотрим правило R_r (типа r), для которого с вероятностью 1 имеет место соотношение

$$\lim_{n\to\infty} \frac{\text{число орлов в первых } n \text{ испытаниях}}{n} = \frac{p_1 q_2^r + p_2 q_1^r}{q_1^r + q_2^r}. \tag{2}$$

Заметим, что при r, стремящемся к бесконечности, правая часть равенства (2), возрастая, стремится к правой части равенства (1).

Интересно было бы знать, насколько наше правило R_r „лучше“ любого другого правила типа r в том смысле, что левая часть соответствующего равенства (2) для этого правила существует для всех p_1 и p_2, является симметрической функцией от p_1 и p_2 и по крайней мере не больше $\frac{p_1 q_2^r + p_2 q_1^r}{q_1^r + q_2^r}$ для всех p_1 и p_2. Мы не знаем, так ли это.

ОПРЕДЕЛЕНИЕ ПРАВИЛА R_r И ДОКАЗАТЕЛЬСТВО ТЕОРЕМЫ

Мы напомним [2], что если для единственной монеты с вероятностью $q = 1 - p$ выпадения решетки производятся бросания до того момента, когда впервые появится последовательность выпавших подряд r решеток, то с вероятностью, близкой к единице, число бросаний должно быть не меньше

$$\frac{1-q^r}{pq^r}.$$

Теперь предположим, что повторные бросания производятся по следующему правилу: если при первом бросании выпала решетка, то заканчиваем бросания, в противном случае продолжаем бросания до первой встречи последовательности из r следующих друг за другом решеток. Тогда предполагаемое число бросаний с вероятностью, близкой к единице, будет не меньше

$$q \cdot 1 + p \cdot \left(1 + \frac{1-q^r}{pq^r}\right) = \frac{1}{q^r}. \tag{3}$$

Докажем следующую теорему.

Теорема. *Определим правило* R_r *следующим образом. Начинаем бросание с монеты* 1. *Если первое же выпадение — решетка, то останавливаемся; в противном случае продолжаем бросание монеты* 1 *до первого появления последовательности* r *следующих друг за другом решеток и только тогда останавливаемся. Таким образом, мы получим первый блок бросаний монеты* 1. *После этого начинаем бросать монету* 2, *используя тот же прием. Получим первый блок бросаний монеты* 2. *Потом опять начинаем бросать монету* 1. *Получим второй блок бросаний монеты* 1 *и т. д. В конце концов получим бесконечную последовательность блоков бросаний либо монеты* 1, *либо монеты* 2.

Для такого определения правила R_r *равенство* (2) *справедливо с вероятностью* 1.

Доказательство. Пусть $x_i(y_i)$ — длина i-го блока бросаний монеты 1 (2), $i = 1,2, \ldots$. В процессе бросания с вероятностью 1 генерируется бесконечная последовательность независимых случайных величин

$$x_1,\ y_1,\ x_2,\ y_2,\ \ldots,\ x_n,\ y_n,\ \ldots. \tag{4}$$

Тогда отношение числа бросаний монеты 1 к общему числу бросаний в первых $2n$ блоках есть

$$\frac{x_1 + \ldots + x_n}{x_1 + \ldots + x_n + y_1 + \ldots + y_n} = \frac{(x_1 + \ldots + x_n)/n}{(x_1 + \ldots + x_n)/n + (y_1 + \ldots + y_n)/n}. \tag{5}$$

Согласно усиленному закону больших чисел и равенству (3), это отношение с вероятностью 1 стремится к пределу

$$\frac{1/q_1^r}{1/q_1^r + 1/q_2^r} = \frac{q_2^r}{q_1^r + q_2^r}. \tag{6}$$

Через u_n обозначим отношение числа выпавших орлов к общему числу бросаний в первых $2n$ блоках. Тогда

u_n равно отношению числа бросаний монеты 1 к общему числу бросаний в первых $2n$ блоках, умноженному на отношение числа орлов, выпавших при использовании монеты 1, к общему числу бросаний монеты 1, плюс подобное же произведение для монеты 2. (7)

Заметим, что при этом с вероятностью 1 имеем

$$\lim_{n\to\infty} u_n = p_1 \frac{q_2^r}{q_1^r + q_2^r} + p_2 \frac{q_1^r}{q_1^r + q_2^r} = \frac{p_1 q_2^r + p_2 q_1^r}{q_1^r + q_2^r}. \tag{8}$$

Для всякого n определим целое положительное число $N = N(n)$ так, что

$$x_1 + y_1 + \ldots + x_N + y_N \leqslant n < x_1 + y_1 + \ldots + x_{N+1} + y_{N+1}. \tag{9}$$

Пусть w_n — число орлов, выпавших в первых n бросаниях. Тогда

число орлов, выпавших в первых $2N$ блоках бросаний, не больше w_n и w_n не больше числа орлов, выпавших в первых $2N+2$ блоках. (10)

Из (9) и (10) следует, что

$$\frac{u_N(x_1+y_1+\ldots+x_N+y_N)}{x_1+y_1+\ldots+x_{N+1}+y_{N+1}} \leqslant \frac{w_n}{n} \leqslant$$

$$\leqslant \frac{u_{N+1}(x_1+y_1+\ldots+x_{N+1}+y_{N+1})}{x_1+y_1+\ldots+x_N+y_N}. \quad (11)$$

Так как $(x_{N+1}+y_{N+1})/N$ стремится к 0 с вероятностью 1 при возрастании n (и, следовательно, N), то с вероятностью 1 справедливо равенство

$$\lim_{n\to\infty} \frac{x_1+y_1+\ldots+x_{N+1}+y_{N+1}}{x_1+y_1+\ldots+x_N+y_N} = 1.$$

Таким образом, с вероятностью 1 из соотношений (8) и (11) находим

$$\lim_{n\to\infty} \frac{w_n}{n} = \lim_{N\to\infty} u_N = \frac{p_1q_2^r+p_2q_1^r}{q_1^r+q_2^r}, \quad (12)$$

что и требовалось доказать.

ЛИТЕРАТУРА

1. Robbins H., Some aspects of the sequential design of experiments, *Bull. Am. Math. Soc.*, 58 (1952), 529—532. (Русский перевод см. на стр. 71 настоящего сборника.)
2. Феллер В., Введение в теорию вероятностей и ее приложения, т. 1, «Мир», М., 1967.

Правильность программ [1])

З. Манна

В статье рассматривается соотношение между правильностью программ и выполнимостью (или невыполнимостью) формул исчисления предикатов первой ступени. Кроме того, в статью включены некоторые результаты об эквивалентности программ.

ВВЕДЕНИЕ

В последнее время значительные усилия направлены на отыскание методов установления правильности (вычислительных) программ. Программа называется правильной, если ее выполнение заканчивается и дает желаемый окончательный результат.

В настоящей работе мы пытаемся формализовать эту проблему путем сведения ее к выполнимости (или невыполнимости) определенных формул первой ступени. Точнее говоря, мы описываем алгоритм, преобразующий любую программу P из данного класса $\{P\}$ программ в формулы первой ступени W_P и $\tilde{W}_P$, такие, что:

(i) W_P выполнима *тогда и только тогда*, когда P правильна или выполнение P не заканчивается, и

(ii) $\tilde{W}_P$ невыполнима *тогда и только тогда*, когда P правильна.

Подобный результат получен и для эквивалентности программ. Две программы называются эквивалентными, если для одинаковых входных величин выполнение обеих заканчивается и обе дают одинаковый окончательный результат [2]). Ясно, что проблемы правильности и эквивалентности программ связаны между собой, ибо вместо того, чтобы непосредственно доказывать правильность программы, мы можем доказать эквива-

---

[1]) Manna Z., The correctness of programms, *Journal of Computer and System Sciences*, **3**, № 2 (1969), 119—127.

[2]) Это определение отличается от обычного определения эквивалентности: для одинаковых входных величин *либо* выполнение обеих программ не заканчивается, *либо* обе программы дают одинаковый окончательный результат.

лентность этой программы некоторой другой программе, правильность которой заранее известна.

Чтобы сократить предварительные определения, мы будем предполагать, что читатель знаком с обычными понятиями, рассматриваемыми в исчислении предикатов первой ступени (см., например, Мендельсон [6]). Представленных здесь результатов касаются также работы Флойда [2, 3], Манны [4, 5] и Купера [1].

ОПРЕДЕЛЕНИЕ ПРОГРАММ

Программа P состоит из:

1) (a) *входного вектора* $\mathbf{x}$,
   (b) *программного вектора* $\mathbf{y}$,
   (c) *выходного вектора* $\mathbf{z}$,
2) (a) непустой *входной области* $D_{\mathbf{x}}$,
   (b) непустой *программной области* $D_{\mathbf{y}}$,
   (c) непустой *выходной области* $D_{\mathbf{z}}$,
3) *начальной функции* (initial assignment function) $g(\mathbf{x})$ — всюду определенной функции, отображающей $D_{\mathbf{x}}$ в $D_{\mathbf{y}}$, и
4) последовательности из N ($N \geqslant 1$) *предложений*, где i-е предложение ($1 \leqslant i \leqslant N$) имеет вид

$$i\text{: } \textit{если } p_i(\mathbf{x}, \mathbf{y}), \textit{ то } [\mathbf{y} \leftarrow f_i^1(\mathbf{x}, \mathbf{y}); \textit{ переходи к } i_1],$$
$$\textit{иначе } [\mathbf{y} \leftarrow f_i^2(\mathbf{x}, \mathbf{y}); \textit{ переходи к } i_2],$$

где:

(a) $p_i(\mathbf{x}, \mathbf{y})$ — всюду определенный предикат на $D_{\mathbf{x}} \times D_{\mathbf{y}}$ [1]),
(b) $f_i^1(\mathbf{x}, \mathbf{y})$ и $f_i^2(\mathbf{x}, \mathbf{y})$ — всюду определенные функции, отображающие $D_{\mathbf{x}} \times D_{\mathbf{y}}$ в $D_{\mathbf{y}}$, и
(c) $1 \leqslant i_1, i_2 \leqslant N$.

Каждая команда типа *переходи к* i_k: $[\mathbf{y} \leftarrow f_i^k(\mathbf{x}, \mathbf{y})$; *переходи к* $i_k]$, $k = 1$ или 2, может быть заменена *стоп-командой* $[\mathbf{z} \leftarrow h_i^k(\mathbf{x}, \mathbf{y}), \textit{стоп}]$, где $h_i^k(\mathbf{x}, \mathbf{y})$ — всюду определенная функция, отображающая $D_{\mathbf{x}} \times D_{\mathbf{y}}$ в $D_{\mathbf{z}}$.

ВЫПОЛНЕНИЕ ПРОГРАММ

Если дана программа P и входное значение $\xi \in D_{\mathbf{x}}$ для входного вектора $\mathbf{x}$, то программа может быть выполнена. Выполнение всегда начинается с первого предложения (помеченного цифрой 1) при начальном значении $\mathbf{y}$, равном $g(\xi)$.

Вообще если достигнуто i-е предложение ($1 \leqslant i \leqslant N$) при $\mathbf{y} = \eta$, то выполнение происходит следующим образом: *если*

[1]) То есть для каждой пары $(\xi, \eta) \in D_{\mathbf{x}} \times D_{\mathbf{y}}$ значением $p_i(\xi, \eta)$ является либо И (истина), либо Л (ложь).

$p_i(\xi, \eta) = И$, *то* [заменяем значение **y** на $f_i^1(\xi, \eta)$ и переходим к i_1-му предложению], *иначе* [заменяем значение **y** на $f_i^2(\xi, \eta)$ и переходим к i_2-му предложению]. Если команда *переходи к* i_k заменена *стоп-командой*, то выполнение осуществляется так: приписываем **z** значение $h_i^k(\xi, \eta)$ и останавливаемся.

Если выполнение заканчивается с $\mathbf{z} = \zeta$, то мы говорим, что $P(\xi)$ *определено и* $P(\xi) = \zeta$. В противном случае, т. е. если выполнение никогда не оканчивается, мы говорим, что $P(\xi)$ *не определено*. Иначе говоря, программу P можно рассматривать как представление частичной функции $\mathbf{z} = P(\mathbf{x})$, отображающей $D_\mathbf{x}$ в $D_\mathbf{z}$.

Пусть P — программа, $\varphi(\mathbf{x})$ — всюду определенный предикат на $D_\mathbf{x}$ (называемый *входным предикатом*) и $\psi(\mathbf{x}, \mathbf{z})$ — всюду определенный предикат на $D_\mathbf{x} \times D_\mathbf{z}$ (называемый *выходным предикатом*). Мы будем говорить, что:

1) P является *правильной относительно* φ *и* ψ, если для каждого ξ, такого, что $\varphi(\xi) = И$, программа $P(\xi)$ определена и $\psi(\xi, P(\xi)) = И$;

2) P является *частично правильной относительно* φ *и* ψ, если для каждого ξ, такого, что $\varphi(\xi) = И$ и $P(\xi)$ определена, имеет место $\psi(\xi, P(\xi)) = И$.

Пример. Рассмотрим программу P^* (для **умножения** целого числа x_1 на неотрицательное целое число x_2 **путем повторных** сложений):

$\mathbf{x} = (x_1, x_2)$ — входной вектор,
$\mathbf{y} = (y_1, y_2)$ — программный вектор,
$\mathbf{z} = z$ — выходной вектор,
$D_\mathbf{x} = I \times I$ — входная область (где I — **множество целых** чисел),
$D_\mathbf{y} = I \times I$ — программная область,
$D_\mathbf{z} = I$ — выходная область,
$g(\mathbf{x}) = (0, x_2)$ — начальная функция.

Единственным предложением является

1: *если* $y_2 = 0$, *то* $[z \leftarrow y_1$, *стоп*],
иначе $[(y_1, y_2) \leftarrow (y_1 + x_1, y_2 - 1)$, *переходи к* 1].

В дальнейшем мы рассмотрим правильность программы P^* относительно входного предиката $x_2 \geqslant 0$ и выходного предиката $z = x_1 x_2$.

АЛГОРИТМ

Если дана программа P, входной предикат $\varphi(\mathbf{x})$ и выходной предикат $\psi(\mathbf{x}, \mathbf{z})$, то можно построить **формулы** $W_P[\varphi, \psi]$ и $\tilde{W}_P[\varphi, \psi]$ следующим образом:

1. Для каждого предложения вида

i: *если* $p_i(\mathbf{x}, \mathbf{y})$, *то* $[\mathbf{y} \leftarrow f_i^1(\mathbf{x}, \mathbf{y})$, *переходи к* $i_1]$,
иначе $[\mathbf{y} \leftarrow f_i^2(\mathbf{x}, \mathbf{y})$, *переходи к* $i_2]$

определим W_i как

$$\forall \mathbf{y}\,(q_i(\mathbf{x}, \mathbf{y}) \supset \textit{если } p_i(\mathbf{x}, \mathbf{y}), \textit{ то } q_{i_1}(\mathbf{x}, f_i^1(\mathbf{x}, \mathbf{y})), \textit{ иначе } q_{i_2}(\mathbf{x}, f_i^2(\mathbf{x}, \mathbf{y}))\,)^{1}),$$

где:

а) q_i — различные предикатные символы, т. е. символы, представляющие неопределенные предикаты;

b) если команда *переходи к* i_k заменена в P *стоп-командой*, то $q_{i_k}(\mathbf{x}, f_i^k(\mathbf{x}, \mathbf{y}))$ заменяется в W_i на $\psi(\mathbf{x}, h_i^k(\mathbf{x}, \mathbf{y}))$.

2. Определим $[P, \psi](\mathbf{x})$ как

$$q_1(\mathbf{x}, g(\mathbf{x})) \wedge W_1 \wedge W_2 \wedge \ \ldots \ \wedge W_N.$$

3. Наконец, определим:

$$W_P[\varphi, \psi] \quad \text{как} \quad \forall \mathbf{x}\,(\varphi(\mathbf{x}) \supset [P, \psi](\mathbf{x})),$$

$$\tilde{W}_P[\varphi, \psi] \quad \text{как} \quad \exists \mathbf{x}\,(\varphi(\mathbf{x}) \wedge [P, \sim\psi](\mathbf{x})).$$

Формула W вида W_P или $\tilde{W}_P$ называется *выполнимой*, если каждому предикатному символу q_i, встречающемуся в ней, мы можем приписать всюду определенный предикат на $D_x \times D_y$, так, что W станет истинной. Формула W называется *невыполнимой*, если она не является выполнимой.

Пример. Рассмотрим программу P' с входным предикатом $x_2 \geqslant 0$ и выходным предикатом $z = x_1 x_2$. Следуя вышеописанному алгоритму [2]), мы получим, что $W_{P'}[x_2 \geqslant 0, z = x_1 x_2]$ есть

$$\forall x_1 \forall x_2 (x_2 \geqslant 0 \supset (q(0, x_2) \wedge \forall y_1 \forall y_2 (q(y_1, y_2) \supset \textit{если } y_2 = 0, \textit{ то } y_1 = x_1 x_2, \textit{ иначе } q(y_1 + x_1, y_2 - 1)))),$$

а $\tilde{W}_{P'}[x_2 \geqslant 0, z = x_1 x_2]$ есть

$$\exists x_1 \exists x_2 (x_2 \geqslant 0 \wedge q(0, x_2) \wedge \forall y_1 \forall y_2 (q(y_1, y_2) \supset \textit{если } y_2 = 0, \textit{ то } y_1 \neq x_1 x_2, \textit{ иначе } q(y_1 + x_1, y_2 - 1))).$$

---

[1]) То есть $\forall \mathbf{y}\,((q_i(\mathbf{x}, \mathbf{y}) \wedge p_i(\mathbf{x}, \mathbf{y})) \supset q_{i_1}(\mathbf{x}, f_i^1(\mathbf{x}, \mathbf{y}))) \wedge ((q_i(\mathbf{x}, \mathbf{y}) \wedge \sim p_i(\mathbf{x}, \mathbf{y})) \supset q_{i_2}(\mathbf{x}, f_i^2(\mathbf{x}, \mathbf{y})))$.

[2]) Для краткости мы вместо $q_1(x_1, x_2, y_1, y_2)$ пишем $q(y_1, y_2)$.

Можно легко проверить, что приписыванием предиката $y_1 = x_1(x_2 - y_2)$ в качестве значения символа $q(y_1, y_2)$ в W_{P^*} получаем истинное выражение. Таким образом, W_{P^*} выполнимо. С другой стороны, мы покажем позднее, что можно вывести противоречие из $\tilde{W}_{P^*}$, что означает невыполнимость $\tilde{W}_{P^*}$.

ЛЕММА

Результаты настоящей статьи доказываются с помощью следующих определений и леммы.

Пусть $\xi \in D_{\mathbf{x}}$. Всюду определенный предикат $\delta(\mathbf{y})$ на $D_{\mathbf{y}}$ называется:

1) *характеристическим* (valid) *предикатом i-го предложения для* $P(\xi)$, если для каждого η из $D_{\mathbf{y}}$, такого, что в процессе выполнения $P(\xi)$ мы достигаем i-го предложения при $\mathbf{y} = \eta$, имеем $\delta(\eta) = \text{И}$;

2) *минимальным характеристическим предикатом i-го предложения для* $P(\xi)$, если для каждого η из $D_{\mathbf{y}}$, такого, что в процессе выполнения $P(\xi)$ мы достигаем i-го предложения при $\mathbf{y} = \eta$ и ни для каких других η, имеем $\delta(\eta) = \text{И}$.

Пример. Рассмотрим программу P^* при значении входа $\xi = (3, 4)$. Предикат $(0 \leqslant y_1 \leqslant 12) \wedge (0 \leqslant y_2 \leqslant 4)$ является характеристическим предикатом (в то время как предикат $\delta(y_1, y_2)$, истинный только для (0, 4), (3, 3), (6, 2), (9, 1) и (12, 0), является минимальным характеристическим предикатом) первого предложения для P^* (3, 4).

Лемма 1.

(i) $P(\xi)$ *не определено или*

(ii) $P(\xi)$ *определено и* $\psi(\xi, P(\xi)) = \text{И}$ *тогда и только тогда, когда* $[P, \psi](\xi)$ *выполнимо.*

Доказательство.→: Для каждого предикатного символа q_i, $1 \leqslant i \leqslant N$, в $[P, \psi](\xi)$ придадим $q_i(\xi, \mathbf{y})$ значение минимального характеристического предиката i-го предложения для $P(\xi)$. Так как $P(\xi)$ либо не определено, либо определено и $\psi(\xi, P(\xi)) = \text{И}$ (другими словами, если $P(\xi)$ определено, то $\psi(\xi, P(\xi) = \text{И})$, то, согласно построению $[P, \psi]$, значением $[P, \psi](\xi)$ при вышеуказанном значении q_i является истина, т. е. $[P, \psi](\xi)$ выполнимо.

←: Если $[P, \psi](\xi)$ выполнимо, то это означает, что предикатным символам $q_i(\xi, \mathbf{y})$, $1 \leqslant i \leqslant N$, можно приписать в качестве значений такие всюду определенные на $D_{\mathbf{y}}$ предикаты $\delta_i(\mathbf{y})$, что значением $[P, \psi](\xi)$ будет истина. Согласно определению $[P, \psi]$, это означает, что каждое δ_i, $1 \leqslant i \leqslant N$, есть характеристический

предикат i-го предложения для $P(\xi)$, и, следовательно, если $P(\xi)$ определено, то $\psi(\xi, P(\xi)) = И$.

Из леммы 1 следует, что:

1) $[P, И](\xi)$ всегда выполнима и

2) $[P, Л](\xi)$ выполнима тогда и только тогда, когда $P(\xi)$ не определено.

ПРАВИЛЬНОСТЬ. ПРОГРАММ

Следующие результаты формализуют понятие правильности программ.

Теорема 1. *Программа P является частично правильной относительно φ и ψ тогда и только тогда, когда $\tilde{W}_P[\varphi, \psi]$ выполнима.*

Теорема 2. *Программа P является правильной относительно φ и ψ тогда и только тогда, когда $\tilde{W}_P[\varphi, \psi]$ невыполнима.*

Теорема 1 представляет собой результат Флойда [2] и доказывается непосредственно с помощью леммы 1. Теорема 2 также может быть доказана с помощью леммы 1, если показать, что программа P *не является* правильной относительно φ и ψ [т. е. $\exists\xi$ такое, что $\psi(\xi) = И$ и

(i) $P(\xi)$ не определено, либо

(ii) $P(\xi)$ определено и $\psi(\xi, P(\xi)) = Л$] тогда и только тогда, когда $\tilde{W}_P[\varphi, \psi]$ выполнима.

Из теоремы 1 (при $\psi \equiv Л$) и теоремы 2 (при $\psi \equiv И$) соответственно получаем следствия.

Следствие 1. *Для каждого ξ, такого, что $\varphi(\xi) = И$, значение $P(\xi)$ не определено тогда и только тогда, когда $W_P[\varphi, Л]$ выполнима.*

Следствие 2. *Для каждого ξ, такого, что $\varphi(\xi) = И$, значение $P(\xi)$ определено тогда и только тогда, когда $\tilde{W}_P[\varphi, И]$ невыполнима.*

Пример. Из теоремы 2 следует, что мы можем доказать правильность программы P^* относительно входного предиката $x_2 \geqslant 0$ и выходного предиката $z = x_1 x_2$, показав, что $\tilde{W}_{P^*}[x_2 \geqslant 0, z = x_1 x_2]$ невыполнима. Как уже было найдено, $\tilde{W}_{P^*}[x_2 \geqslant 0, z = x_1 x_2]$ есть

$$\exists x_1 \exists x_2 (x_2 \geqslant 0 \wedge q(0, x_2) \wedge \forall y_1 \forall y_2 (q(y_1, y_2) \supset \text{если } y_2 = 0, \text{ то } y_1 \neq x_1 x_2, \text{ иначе } q(y_1 + x_1, y_2 - 1))).$$

Мы покажем, что $\tilde{W}_{P^*}[x_2 \geqslant 0,\ z = x_1 x_2]$ невыполнима, путем вывода противоречия, используя следующие четыре предложения:

(1) $x_2 \geqslant 0$,

(2) $q(0, x_2)$,

(3) $\forall y_1 \forall y_2 ((q(y_1 y_2) \wedge y_2 > 0) \supset q(y_1 + x_1,\ y_2 - 1))$,

(4) $\forall y_1 \forall y_2 ((q(y_1,\ y_2) \wedge y_2 = 0) \supset y_1 \neq x_1 x_2)$.

Путем подстановки $x_1(x_2 - y_2)$ вместо y_1 в (3) получаем:

(3′) $\forall y_2 ((q(x_1(x_2 - y_2),\ y_2) \wedge y_2 > 0) \supset q(x_1(x_2 - y_2) + x_1,\ y_2 - 1))$.

Используя принцип индукции

$$\exists x (x \geqslant 0 \wedge Q(x)) \wedge \forall y ((Q(y) \wedge y > 0) \supset Q(y - 1)) \supset Q(0)$$

при $Q(t) \equiv q(x_1(x_2 - t),\ t)$, мы получим из (1), (2) и (3′)

$Q(0)$, т. е. $q(x_1 x_2,\ 0)$, что противоречит (4).

ЭКВИВАЛЕНТНОСТЬ ПРОГРАММ

Две программы P_1 и P_2 назовем *сравнимыми*, если они имеют одинаковые входную переменную $\mathbf{x}$, выходную переменную $\mathbf{z}$, входную область D_x и выходную область D_z.

Пусть P_1 и P_2 — две сравнимые программы, и пусть $\varphi(\mathbf{x})$ — всюду определенный предикат на D_x (называемый *входным предикатом*). Мы скажем, что P_1 *и* P_2 *эквивалентны относительно* φ, если для всякого ξ, такого, что $\varphi(\xi) = \text{И}$, как $P_1(\xi)$, так и $P_2(\xi)$ определены и $P_1(\xi) = P_2(\xi)$.

Эквивалентность программ можно формализовать с помощью формулы $W_{P_1, P_2}[\varphi]$, которая определяется так:

$$\exists \mathbf{x} (\varphi(\mathbf{x}) \wedge [P_1, r](\mathbf{x}) \wedge [P_2, \sim r](\mathbf{x})),$$

где:

(a) $r(\mathbf{x}, \mathbf{z})$ — произвольный предикатный символ,

(b) символы r, q_i' (использованные при построении формулы $[P, r]$) и q_i'' (использованные при построении формулы $[P_2, \sim r]$) должны быть различными.

Теорема 3. *Программы* P_1 *и* P_2 *эквивалентны относительно* φ *тогда и только тогда, когда* $W_{P_1, P_2}[\varphi]$ *невыполнима.*

Доказательство. Мы покажем, что P_1 и P_2 *не* эквивалентны относительно φ, т. е. существует ξ, такое, что $\varphi(\xi) = \text{И}$, и

(i) $P_1(\xi)$ не определено, либо

(ii) $P_2(\xi)$ не определено, либо

(iii) $P_1(\xi)$ и $P_2(\xi)$ определены, но $P_1(\xi) \neq P_2(\xi)$, тогда и только тогда, когда $W_{P_1, P_2}[\varphi]$ выполнима.

(i) $\rightarrow$: Придаем r значение Л и используем лемму 1.

(ii) $\rightarrow$: Придаем r значение И и используем лемму 1.

(iii) $\rightarrow$: Берем в качестве r предикат δ (где $\delta(\mathbf{x}, \mathbf{z}) = $ И тогда и только тогда, когда $(\mathbf{x}, \mathbf{z}) = (\xi, P_1(\xi))$). Согласно лемме 1 $[P_1, \delta](\xi)$ выполнима. Кроме того, так как $P_1(\xi) \neq P_2(\xi)$, имеем $\delta(\xi, P_2(\xi)) = $ Л, т. е. $\sim\delta(\xi, P_2(\xi)) = $ И. Следовательно, по лемме 1 $[P_2, \sim\delta](\xi)$ также выполнима. Это означает, что и $W_{P_1, P_2}[\varphi]$ выполнима.

$\leftarrow$: Предположим, что $W_{P_1, P_2}[\varphi]$ выполнима при ξ, подставленном вместо $\mathbf{x}$, и δ — вместо r. Так как $[P_1, \delta](\xi)$ выполнима, то из леммы 1 следует, что если $P_1(\xi)$ определено, то $\delta(\xi, P_1(\xi)) = $ И. Так как $[P_2, \sim\delta](\xi)$ выполнима, из леммы 1 также следует, что если $P_2(\xi)$ определена, то $\sim\delta(\xi, P_2(\xi)) = $ И. Это означает, что если определены $P_1(\xi)$ и $P_2(\xi)$, то $\delta(\xi, P_1(\xi)) = $ И, а $\delta(\xi, P_2(\xi)) = $ Л, т. е. $P_1(\xi) \neq P_2(\xi)$.

Пример. Рассмотрим две программы P_1^* и P_2^* (вычисляющие $x!$ для неотрицательных целых x), где:

1) в обеих программах

$$\mathbf{x} = x, \quad \mathbf{y} = (y_1, y_2), \quad \mathbf{z} = z, \quad D_{\mathbf{x}} = I, \quad D_{\mathbf{y}} = I \times I, \quad D_{\mathbf{z}} = I;$$

(2) кроме того,

(a) P_1^* состоит из начальной функции $g_1(x) = (1, x)$ и предложения

1: *если* $y_2 = 0$, *то* $[z \leftarrow y_1;$ *стоп*],
иначе $[(y_1, y_2) \leftarrow (y_1 y_2, y_2 - 1);$ *переходи к* 1];

(b) P_2^* состоит из начальной функции $g_2(x) = (1, 0)$ и предложения

1: *если* $y_2 = x$, *то* $[z \leftarrow y_1,$ *стоп*],
иначе $[(y_1, y_1) \leftarrow (y_1(y_2 + 1), y_2 + 1);$ *переходи к* 1].

Ясно, что P_1^* и P_2^* сравнимы.

Из теоремы 3 следует, что мы можем доказать эквивалентность программ P_1^* и P_2^* относительно входного предиката $x \geqslant 0$, показав, что $W_{P_1^*, P_2^*}[x \geqslant 0]$ невыполнима, где $W_{P_1^*, P_2^*}[x \geqslant 0]$ есть

$$\exists x \{x \geqslant 0 \wedge$$
$$\wedge\, q_1'(x, 1, x) \wedge \forall y_1 \forall y_2 [q_1'(x, y_1, y_2) \supset \textit{если } y_2 = 0, \textit{ то } r(x, y_1),$$
$$\textit{иначе } q_1'(x_1, y_1 y_2, y_2 - 1)]$$
$$\wedge\, q_1''(x, 1, 0) \wedge \forall y_1 \forall y_2 [q_1''(x, y_1, y_2) \supset \textit{если } y_2 = x, \textit{ то } \sim r(x, y_1),$$
$$\textit{иначе } q_1''(x, y_1(y_2 + 1), y_2 + 1)]\}.$$

ЛИТЕРАТУРА

1. Cooper D. C., Program scheme equivalences and second order logic, Presented at Fourth Annual Machine Intelligence Workshop, University of Edinburg, August 1968.

2. Floyd R. W., Assigning meaning to programms, Proceedings of Simposia in Applied Mathematics, American Mathematical Society, v. 19, 1967, p. 19—32.
3. Floyd R. W., The verifying compiler, Computer Science Research Review Carnegie-Mellon University, December 1967.
4. Manna Z., Termination of algorithms, Ph. D. Thesis, Computer Science Department, Carnegie-Mellon University, April 1968.
5. Manna Z., Properties of programs and the ferst-order predicate calculus, *JACM* (April 1969).
6. Mendelson F., Introduction to mathematical logic, Van Nostrand Company, Princeton, 1964.

О проблеме нахождения минимальных программ для таблиц [1])

Д. Пейджер

В этой статье мы накладываем очень слабое условие на „длину программы“ $Q(z)$. Не предполагая, что $Q(z)$ частично рекурсивна, мы доказываем несколько неожиданный результат: для некоторого целого числа m существует конечный набор программ, содержащий кратчайшие программы для множества характеристических фуикций, определенных на множествах яз m элементов, и такой, что проблэма отыскания по таблнце функции ее кратчайшей программы в этом наборе неразрешима.

Вопрос о сложности программ привлек внимание многих авторов. Блюм [1] рассматривал соотношение между сложностью программы и временем вычисления. Чейтин [2] получил оценку сложности кратчайших программ, порождающих последовательности данной длины [2]). Радо [4] показал, что невозможно эффективно определить для произвольного n наибольшее число, которое может вычислить программа с n командами. Мы интересовались обратной проблемой: найти кратчайшую программу для таблицы. Под таблицей мы понимаем конечное множество пар: $\{(x_i, y_i)\colon 1 \leqslant i \leqslant n;\ x_i = x_j \to i = j\}$.

Результат, который мы получили, обладает двумя несколько неожиданными свойствами:

(1) выбор кратчайшей программы для рассматриваемого множества таблиц оказывается невозможным, несмотря на то, что он производится каждый раз из одного и того же конечного множества программ;

(2) теорема верна для любого определения длины программы, такого, что данную длину имеет только конечное множество программ. В частности, *не* предполагается, что длина эффективно вычислима [3]).

---

[1]) P a g e r D., On the problem of finding minimal programs for tables, *Information and Control*, 14, № 6 (1969), 550–554.

[2]) Ранее аналогичный результат в терминах реализации функций алгебры логики машинами Тьюринга и нормальными алгоритмами (а также конечными автоматами) получил В. А. Кузьмин. (См. К у з ь м и н В. А., Реализация функций алгебры логики автоматами, нормальными алгорифмами и машинами Тьюринга, сб. „Проблемы кибернетики“, вып. 13, 1965, стр. 75–96.) — *Прим. ред.*

[3]) Однако предполагается, что множество всех программ данной длины известно. — *Прим. перев.*

Примерами невычислимой функции длины (применимой к данной программе Z) являются:

(a) номер первой программы (в некотором упорядочении программ), которая эквивалентна Z;

(b) номер первой таблицы вида

$$\{(x_1, Z(x_1)), (x_2, Z(x_2)), \ldots, (x_z, Z(x_z))\},$$

которая встречается в некотором упорядочении множества всех таблиц. Здесь $Z(x)$ есть выход программы Z для входа x.

В этом и других примерах нерекурсивной функции длины возможно, что существуют некоторые непрямые способы нахождения кратчайшей программы для таблицы, даже если мы не всегда можем вычислить длину произвольной программы.

Мы примем некоторую синтаксическую систему для представления алгоритмов, причем такую, что к ней применима гёделевская нумерация. Мы будем называть эти синтаксические представления *программами*, обозначая их прописными буквами, а их гёделевские номера будем обозначать соответствующими строчными буквами.

Под *длиной* программы мы понимаем некоторую числовую характеристику программ $Q(z)$, обладающую следующим свойством: для любого k множество программ X, таких, что $Q(x) < k$, конечно.

Длина $Q(z)$ программы может быть некоторой мерой „объема памяти", занимаемого программой, например числом символов, содержащихся в Z [1]). Нетрудно видеть, как меры такого рода могут быть применены к таким алгоритмическим системам, как машины Тьюринга, вычислительные машины, системы уравнений Клини [3], U.R.M. Шефердсона и Стургиса [6] и т. д.

Под *таблицей* мы понимаем конечное множество записей:

$$\{(x_i, y_i)\colon\ 1 \leqslant i \leqslant n,\ x_i = x_j \rightarrow i = j\},$$

а *программой* для этой таблицы является программа Z, такая, что

$$Z(x_i) = y_i \quad \text{для} \quad 1 \leqslant i \leqslant n.$$

В частности, мы рассмотрим таблицы для характеристических функций конечной мощности, называемые *разрешающими таблицами*. Такие таблицы состоят из записей вида

$$(x_i, t_i), \quad \text{где} \quad t_i \quad \text{есть} \quad 0 \quad \text{или} \quad 1.$$

Будем называть такие записи *характеристиками*.

---

[1]) Под символом мы понимаем элементарный символ. Таким образом, составной „символ" q_{12} машины Тьюринга должен рассматриваться как состоящий из трех символов.

Теорема. *Существует конечный набор программ, содержащий кратчайшие программы для некоторого множества разрешающих таблиц одинаковой (конечной) длины и такой, что проблема получения по таблице ее кратчайшей программы из этого набора неразрешима.*

Мы докажем эту теорему, показав, что *существует конечная разрешающая таблица* λ_0, *такая, что проблема нахождения кратчайшей программы для таблицы, полученной добавлением к* λ_0 *одной характеристики, не является равномерно разрешимой* [1]).

Заметим, что в любом случае программа может выбираться из одного и того же *конечного множества* U программ длины

$$\leqslant \max(Q(Z_0),\ Q(Z_1)),$$

где Z_0, Z_1 — произвольные программы для λ_0, такие, что $Z_0(x)=0$ и $Z_1(x)=1$ для всякого x, не принадлежащего области определения λ_0.

Доказательство

Определение (Роджерс). Два непересекающихся рекурсивно перечислимых (р. п.) множества G_1 и G_2 называются *сильно неотделимыми* тогда и только тогда, когда:

(i) $\overline{G_1 \cup G_2}$ бесконечно и

(ii) если A есть р. п. множество и $A \cap \overline{G_1 \cup G_2}$ бесконечно, то $(A \cap G_1) \neq \varnothing$ и $(A \cap G_2) \neq \varnothing$.

Пусть G_1 и G_2 — сильно неотделимые множества [2]), и пусть $f(x)$ — частично рекурсивная функция, такая, что

$$f(x) = \begin{cases} 0, & \text{если } x \in G_1, \\ 1, & \text{если } x \in G_2. \end{cases}$$

Любое частично рекурсивное расширение $F(x)$ функции $f(x)$ определено только на конечном множестве дополнительных элементов. Действительно, область определения функции $F(x)$ можно разбить на два непересекающихся рекурсивно перечислимых множества K_1, K_2, а именно:

$$K_1 = \{x:\ F(x) = 0\},\ K_2 = \{x:\ F(x) \neq 0\},$$

---

[1]) То есть не существует алгоритма, который для любого такого пополнения таблицы λ_0 давал бы кратчайшую программу. — *Прим. перев.*

[2]) Пример пары сильно неотделимых множеств содержится в работе: Мучник А. А., Об отделимости рекурсивно перечислимых множеств, *ДАН СССР*, **109**, № 1 (1956), 29—32. Множества в построенном А. А. Мучником примере неотделимы даже в более сильном смысле. — *Прим. ред.*

так, что $G_1 \subseteq K_1$, $G_2 \subseteq K_2$. Но тогда $K_1 - G_1$ и $K_2 - G_2$ конечны и, следовательно, конечно множество

$$(K_1 - G_1) \cup (K_2 - G_2),$$

т. е.

$$(K_1 \cup K_2) - (G_1 \cup G_2),$$

которое является разностью области определения $F(x)$ и области определения $f(x)$.

Пусть H — конечное множество программ, вычисляющих функцию $f(x)$ на ее области определения с минимальной Q-мерой, и пусть k есть значение Q на элементах H. Если $Z \in H$, то $Z(x)$ является расширением $f(x)$. Пусть V является (конечным) объединением всех конечных множеств $D_z - D_f$, где $Z \in H$, D_z — область определения $Z(x)$, D_f — область определения $f(x)$.

Если Z есть произвольная программа, которая не вычисляет $f(x)$ на ее области определения, то существует x из D_f, такое, что $Z(x)$ либо не определено, либо $Z(x) \neq f(x)$. Пусть x_z — наименьшее такое значение для программы Z, и пусть S обозначает конечное множество $\{x_z\colon Q(z) \leqslant k\}$.

Множество характеристик $\{(x, f(x))\colon x \in S\}$ обозначим через λ_0. Если $t \notin V$, то $t \in D_f$ *тогда и только тогда, когда самая короткая программа для таблицы* λ_0, *пополненной характеристикой* $(t, 0)$ *либо характеристикой* $(t, 1)$, *является элементом* H.

Если бы существовал единый метод нахождения кратчайшей программы для λ_0, пополненной одной дополнительной характеристикой, то, так как V конечно, из последнего утверждения следовало бы, что множество $\{t\colon t \notin V \text{ и } t \notin D_f\}$, т. е. $\{t\colon t \in \overline{G_1 \cup G_2} - V\}$, рекурсивно перечислимо. А так как множество V конечно, это противоречит сильной неотделимости G_1 и G_2. Теорема доказана.

Фактически мы получили более сильный результат, чем утверждается в теореме: *пусть* U — *конечное множество программ, такое, что* H — *его (собственное) подмножество. Пусть* ζ — *рекурсивное множество разрешающих таблиц, состоящее из таблиц, являющихся пополнением* λ_0 *одной характеристикой. Тогда кратчайшая программа для каждого члена* ζ *принадлежит* U, *но мы не можем эффективно решить, принадлежит ли она* H *или* $U - H$.

Предположим теперь, что существует эффективное многооднозначное отображение, которое соотносит программам Z эквивалентные программы Z^* с тем же числом команд, так, что

множество программ Z^* с данным числом команд конечно [1]). Тогда можно получить вариант теоремы, в котором слова „кратчайшая программа" заменены словами „программа с наименьшим числом команд".

Подобный результат имеет место и по отношению к шенноновской мере сложности машин Тьюринга, т. е. произведению числа состояний на число ленточных символов [5].

С другой стороны, если рассмотреть более слабый случай, когда 1) множество кратчайших программ для рассматриваемых таблиц бесконечно и 2) длина эффективно вычислима, то можно, используя теорему о рекурсии, доказать: *если справедливы* 1) *и* 2) *и если множество рассматриваемых таблиц рекурсивно перечислимо, то не существует эффективного способа находить кратчайшие программы.*

ЛИТЕРАТУРА

1. Blum M., On the size of machines, *Inform. and Control*, **11** (1967), 257—265. (Русский перевод: Блюм М., Об объеме машин, сб. «Проблемы математической логики», стр. 423—431, «Мир», М., 1970.)
2. Chaintin G. T., On the length of programs for computing finite binary sequences, *J. ACM*, **13** (1966), 547—569.
3. Клини С. К., Введение в метаматематику, ИЛ, М., 1957.
4. Rado T., On noncomputable functions, *Bell Sistem Techn. J.* (1966), 877—884.
5. Shannon C. E., A universal Turing machine with two internal states, in «Automata Studies», Princeton University Press, Princeton, New Jersey, 1956. (Русский перевод: см. в сб. «Автоматы», ИЛ, М., 1956, а также в сб.: Шеннон К., Работы по теории информации и кибернетике, ИЛ, М., 1963.)
6. Shepherdson J. C., Sturgis H. E., Computability of recursive functions, *J. ACM*, **10** (1963), 217—225.
7. Yong P. R., Towards a theory of enumerations, IEEE Conference Record, Symposium on Switching and Automata Theory, 1968, pp. 334—350.

[1]) Например, оно получается с помощью замены отсылок к бо́льшим адресам памяти в реальных вычислительных машинах или в U. R. M. отсылками к неиспользованиым меньшим адресам либо заменой составных символов q_j и S_i в командах машин Тьюринга символами с неиспользованными меньшими индексами.

Об арифметических выражениях и деревьях[1])

Р. Р. Реджеёвски

В работе описывается, каким образом дерево, представляющее вычисление арифметического выражения, можно изобразить так, что количество требуемых для счета ячеек памяти легко определяется. Это представление сводит проблему выбора наилучшего порядка вычисления к одной из задач теории графов Предлагается алгоритм решения этой задачи.

1. ВВЕДЕНИЕ

В работе [1] обсуждается алгоритм вычисления арифметического выражения, минимизирующий требуемое для вычисления количество ячеек памяти. Как показано ниже, дерево, используемое в [1], может быть модифицировано таким образом, что требуемое количество ячеек памяти получает явное представление. Это представление приводит к специфической проблеме „минимальной ширины" упорядочения дерева, которая и является темой настоящей статьи.

Чтобы вычислить выражение, например

$$((a+b)-c\cdot d)/(e\,(f-g)),$$

потребуется выполнить ряд арифметических операций: $A=a+b$, $B=c\cdot d$, $C=f-g$, $D=A-B$, $E=e\cdot C$, $F=D/E$, вычисляя таким образом промежуточные результаты A, B, ..., E и конечный результат F. Этот процесс может быть представлен в виде дерева, как на рис. 1. Каждая вершина этого дерева представляет одну операцию; дуга из вершины x в вершину y представляет промежуточный результат, который используется в операции x, а вычисляется операцией y.

Выбрать возможный порядок операций — значит топологически упорядочить дерево, т. е. расположить его вершины в такой последовательности, что вершина x предшествует вершине y, если существует дуга из y в x. (Это условие эквивалентно требованию, чтобы каждый промежуточный результат был вычислен раньше, чем использован.)

---

[1]) Redziejowski R. R., On arithmetic expressions and trees, *Communications of the ACM*, 12, № 2 (1969), 81—84.

Результат такого упорядочения удобно представлять в виде диаграммы, как это сделано на рис. 2. В дальнейшем мы будем называть такую диаграмму „укладкой“ дерева, изображенного на рис. 1. Она может рассматриваться как представление вычисления на шкале времени.

Предположим, что мы имеем „моментальный снимок“ вычисления в момент выполнения операции x. Такой моментальный снимок покажет некоторые результаты предыдущих операций, хранящиеся для дальнейшего использования и не

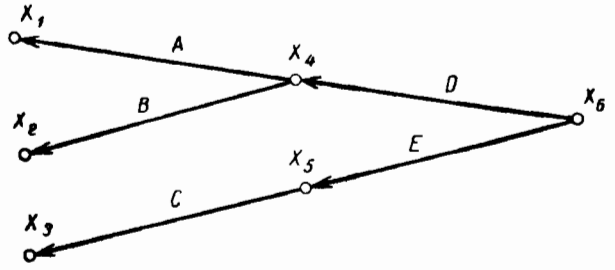


Рис. 1.

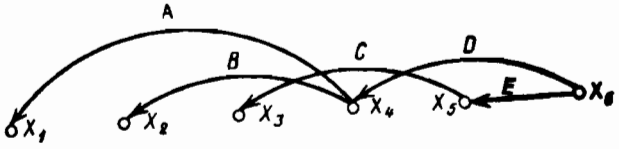


Рис. 2.

участвующие в операции x. Легко видеть, что на рис. 2 эти хранящиеся результаты представляются дугами, проходящими над вершиной x. Если число этих дуг есть w, то мы должны иметь не менее $w+1$ запоминающих элементов (ячеек в оперативной памяти или на магнитной ленте) в нашем распоряжении в данный момент: w для результатов, уже хранящихся, и один для результата x. Число запоминающих элементов, которые мы должны иметь для всего вычисления, определяется, таким образом, как максимальное число дуг, проходящих над одной вершиной на рис. 2; мы будем называть это число „шириной“ нашей укладки. Выбор наилучшего порядка операций сводится, таким образом, к построению укладки минимальной ширины для дерева на рис. 1.

В третьем разделе предлагается алгоритм построения такой укладки; во втором разделе даны определения.

Хотя деревья, соответствующие арифметическим выражениям, всегда бинарны, решение задачи для общего случая не является более сложным. Поэтому в дальнейшем рассматривается общий случай.

2. ОПРЕДЕЛЕНИЯ

В дальнейшем используется понятие (направленного) *графа*, определенного как упорядоченная пара (X, U), состоящая из непустого множества X *вершин* и множества U *дуг*. Дуга из вершины $x \in X$ в вершину $y \in X$ обозначается здесь $(x, y) \in U$. Для $x, y \in X$, таких, что существует дуга $(x, y) \in U$, вершину y назовем *последователем* x. Предполагается, что читатель знаком с элементарными понятиями пути в графе, контура и т. д.

Граф (X, U) называется *деревом*, если он конечен, не содержит контуров и содержит вершину r, такую, что:

a) r не является концом никакой дуги и

b) каждая вершина $x \in X$, $x \neq r$ является концом в точности одной дуги.

Вершина r, обладающая этими свойствами, называется *корнем* дерева. Вершина $x \in X$, не являющаяся началом никакой дуги, называется *листом* дерева [1]); любая вершина, не являющаяся листом, называется *узлом*. Для дерева $G = (X, U)$ и его вершины $x \in X$ мы обозначаем через $G(x)$ поддерево, состоящее из x и всех вершин, достижимых из x путями в G. *Укладка* дерева $G = (X, U)$ формально определяется как последовательность $\varphi = (x_1, x_2, \ldots, x_n)$ всех вершин дерева G, такая, что для каждой дуги $(x_i, x_j) \in U$ имеет место $j < i$. В укладке $\varphi = (x_1, x_2, \ldots, x_n)$ дуга $(x_i, x_j) \in U$ называется *проходящей* над вершиной $x_k \in X$, если $j < k < i$.

Для этой укладки пусть w_i обозначает количество дуг, проходящих над вершиной $x_i \in X$. *Ширина* укладки φ определяется тогда как $\max w_i (i = 1, 2, \ldots, n)$ и обозначается $W(\varphi)$.

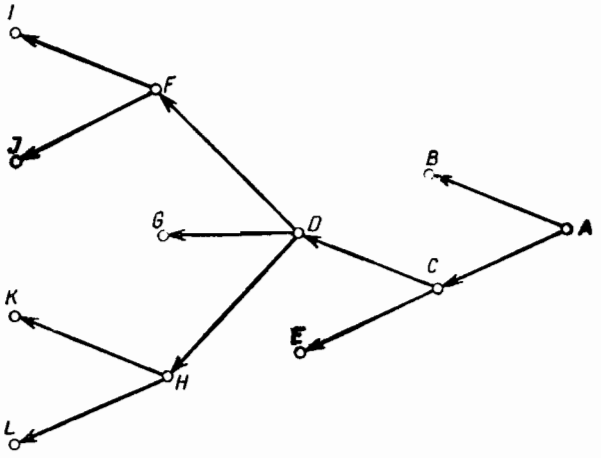
3. АЛГОРИТМ

Укладка минимальной ширины данного дерева G может быть получена заполнением таблицы вида, указанного на рис. 3. Таблица имеет 3 колонки и n строк. Заполнение начинается

---

[1]) В советской литературе более распространен термин „висячая вершина". — *Прим. перев.*

с перечисления в колонке 1 всех вершин дерева G. Мы записываем здесь снизу вверх сначала корень дерева G, затем записываем всех его последователей, затем — последователей этих последователей и т. д. Каждой вершине, таким образом, сопоставляется строка таблицы, причем если вершина y есть



Колонка № 1, x	Колонка № 2, $v(x)$	Колонка № 3, $\varphi(x)$		
L	0	*L*		
K	0	*K*		
J	0	*J*		
I	0	*I*		
H	1	*KLH*	*K, L;*	0, 1;
G	0	*G*		
F	1	*IJF*	*I, J;*	0, 1;
E	0	*E*		
D	2	*IJFKLHGD*	*F, H, G;*	1, 2, 2;
C	2	*IJFKLHGDEC*	*D, E;*	2, 1;
B	0	*B*		
A	2	*IJFKLHGDECBA*	*C, B;*	2, 1;

Рис. 3.

последователь вершины x, то строка, сопоставленная x, лежит ниже строки, сопоставленной y. Заполнив колоику 1, заполняем таблицу по строкам.

Предположим, что мы уже заполнили $m \geqslant 0$ строк и собираемся заполнять следующую $(m+1)$-ю строку. Пусть вершина, соответствующая этой строке, есть x. Мы заполняем $(m+1)$-ю

строку по одному из следующих правил, в зависимости от того, является x листом или узлом дерева G.

Правило 1. Если x является листом дерева G, то пишем 0 в колонке 2 и пишем x в колонке 3.

Правило 2. Если x есть узел дерева G, то находим в G всех последователей x; пусть это $y_0, y_1, \ldots, y_k$, где $k \geqslant 0$. Находим в таблице строки, соответствующие $y_0, y_1, \ldots, y_k$; благодаря способу заполнения колонки 1, все они должны быть среди m уже заполненных строк. Пусть $v(y_i)$ для $i = 0, 1, \ldots, k$ обозначает содержимое колонки 2 в строке, соответствующей y_i, и $\varphi(y_i)$ — содержимое колонки 3 в той же строке. Расположим $y_0, y_1, \ldots, y_k$ в последовательность по убыванию величины $v(y_i)$ для $i = 0, 1, \ldots, k$. Пусть эта последовательность есть $(y_{i_0}, y_{i_1}, \ldots, y_{i_k})$; тогда $v(y_{i_0}) \geqslant v(y_{i_1}) \geqslant \ldots \geqslant v(y_{i_k})$. Подсчитаем число $p = \max(p_0, p_1, \ldots, p_k)$, где $p_j = v(y_{i_j}) + j$ для $j = 0, 1, \ldots, k$. Запишем p в $(m+1)$-й строке второй колонки; в колонке 3 запишем друг за другом $\varphi(y_{i_0})\varphi(y_{i_1}) \ldots \varphi(y_{i_k})$ в таком порядке и сзади припишем x.

Для иллюстрации применения правила 2 на рис. 3 показаны последовательности $(y_{i_0}, y_{i_1}, \ldots, y_{i_k})$ и числа $p_0, p_1, \ldots, p_k$; они расположены справа от каждой строки, заполнявшейся по этому правилу. Например, для $x = D$ мы имеем $y_{i_0} = F$, $y_{i_1} = H$, $y_{i_2} = G$ и $p_0 = v(y_{i_0}) + 0 = 1$, $p_1 = v(y_{i_1}) + 1 = 2$, $p_2 = v(y_{i_2}) + 2 = 2$.

Покажем теперь, что после заполнения таблицы последовательность, появившаяся в последней строке колонки 3, есть укладка минимальной ширины дерева G.

4. ДОКАЗАТЕЛЬСТВО АЛГОРИТМА

Предположим, что таблица заполнена по вышеуказанным правилам для дерева $G = (X, U)$ с корнем r. Для $x \in X$ пусть $v(x)$ есть число, стоящее в строке таблицы, соответствующей x, в колонке 2; пусть $\varphi(x)$ — последовательность, стоящая в той же строке в колонке 3. Последняя строка соответствует корню r; для этой строки мы имеем $v(r)$ и $\varphi(r)$ соответственно. Могут быть доказаны следующие два утверждения относительно содержимого таблицы.

Теорема 1. *Для каждой вершины $x \in X$ последовательность $\varphi(x)$ есть укладка поддерева $G(x)$, а ее ширина равна $v(x)$.*

Теорема 2. *Для каждой укладки дерева G ее ширина не меньше, чем $v(r)$.*

Из теоремы 1 и того факта, что $G(r) = G$, немедленно следует, что последовательность $\varphi(r)$, стоящая в последней строке третьей колонки, есть укладка дерева G, имеющая ширину $v(r)$. Теорема 2 утверждает, что эта ширина минимальна.

Доказательство теоремы 1. Чтобы доказать теорему, достаточно показать, что (1) теорема имеет место для каждого листа дерева G, и (2) если теорема верна для всех последователей узла x, то она верна и для x.

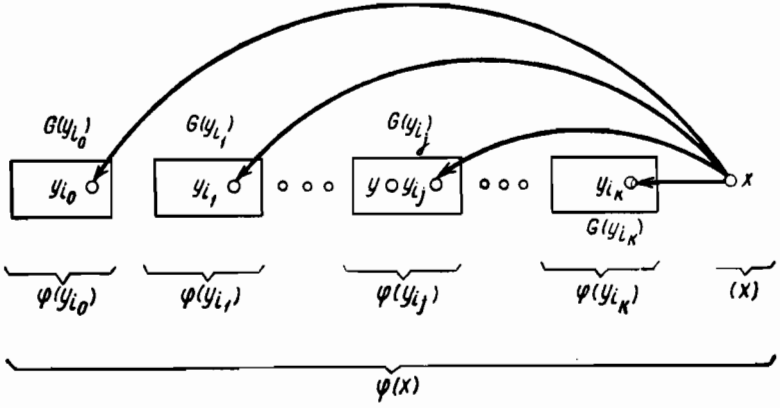


Рис. 4.

То, что утверждение (1) верно, следует из правила 1. Если x есть лист, то поддерево $G(x)$ состоит из единственной вершины x. Последовательность (x) есть укладка такого дерева, и ее ширина равна 0.

Справедливость утверждения (2) может быть показана следующим образом. Пусть x есть узел дерева G и $y_{i_0}, y_{i_1}, \ldots, y_{i_k}$ — его последователи, как в правиле 2. Предположим, что теорема верна для всех этих последователей, т. е. $\varphi(y_{i_j})$ для $j = 0, 1, \ldots, k$ есть укладка поддерева $G(y_{i_j})$, и ее ширина равна $v(y_{i_j})$. Сочленение $\varphi(y_{i_0}), \varphi(y_{i_1}), \ldots, \varphi(y_{i_k})$ и (x) в правиле 2 соответствует конструкции, показанной на рис. 4.

Имея $(k+1)$ дерево $G(y_{i_0}), G(y_{i_1}), \ldots, G(y_{i_k})$, мы строим дерево $G(x)$, добавляя вершину x и $(k+1)$ дуг $(x, y_{i_0}), (x, y_{i_1}), \ldots, (x, y_{i_k})$. Легко видеть (из рис. 4), что полученная таким образом последовательность $\varphi(x)$ должна быть укладкой $G(x)$. Чтобы показать, что ширина этой укладки есть $v(x)$,

рассмотрим вершину y, принадлежащую дереву $G(y_{i_j})$, где $0 \leqslant j \leqslant k$. Легко видеть, что число дуг, проходящих над y, не превышает числа p_j, определенного по правилу 2: имеется j дуг (x, y_{i_0}), (x, y_{i_1}), ..., $(x, y_{i_{j-1}})$ вне дерева $G(y_{i_j})$ и ровно $v(y_{i_j})$ дуг в этом дереве [1]). Отсюда ширина $\varphi(x)$ должна быть равной $\max(p_0, p_1, \ldots, p_k) = v(x)$. Теорема, таким образом, верна также и для x.

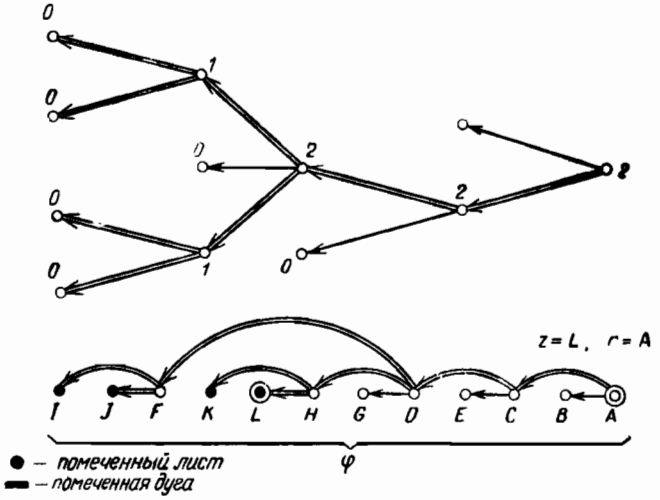


Рис. 5.
Числа в вершинах – значении $v(x)$.

Доказательство теоремы 2. Теорема, очевидно, верна в случае $v(r) = 0$. Пусть $v(r) > 0$, и, предположим, существует укладка φ дерева G, имеющая ширину $W(\varphi) < v(r)$. Для каждого узла $x \in G$ сделаем следующее: найдем последовательность $(y_{i_0}, y_{i_1}, \ldots, y_{i_k})$ и числа $p_0, p_1, \ldots, p_k$, вычисленные во время заполнения по правилу 2 строки, соответствующей x; затем найдем наименьшее j, $0 \leqslant j \leqslant k$, для которого $p_j = v(x)$, и пометим дуги (x, y_{i_0}), (x, y_{i_1}), ..., (x, y_{i_j}). Например, если на рис. 3 $x = D$, то $j = 1$, и мы пометим дуги (D, F) и (D, H); результат таких пометок дерева, изображенного на рис. 3, показан на рис. 5. Легко видеть [2]), что (1) каждый узел x есть начало по меньшей мере одной помеченной дуги и (2) для

[1]) Точнее, не более чем $v(y_{i_j})$ дуг. – *Прим. перев.*

[2]) Из определения $v(x)$. – *Прим. перев.*

каждой помеченной дуги $(x, y) \in U$ узел x есть начало по меньшей мере $(q+1)$ помеченных дуг, где $q = v(x) - v(y)$ [1]).

Пометим теперь все листы дерева G, достижимые из корня r путями, состоящими только из помеченных дуг; из утверждения (1) следует, что существует по крайней мере один такой лист. Пусть $z \in X$ — помеченный лист, самый правый в последовательности φ; пусть μ — путь из r в z (μ состоит только из помеченных дуг). Возьмем дугу $u = (x, y) \in U$, принадлежащую μ. Пусть $v(x) - v(y) = q$; согласно утверждению (2), узел x в этом случае является началом не менее чем $(q+1)$ помеченных дуг. Пусть $q > 0$, и пусть тогда $u_1, u_2, \ldots, u_q$ — помеченные дуги, начинающиеся в узле x и отличные от u. Для

Рис. 6.

$i = 1, 2, \ldots, q$ пойдем из x по пути, состоящему целиком из помеченных дуг и начинающемуся дугой u_i сколь возможно далеко (для иллюстрации см. рис. 6). Этот путь μ_i должен кончаться в помеченном листе, т. е. слева от z в последовательности φ (по выбору z). Так как φ есть укладка дерева G, то узел x должен быть справа от z в φ, т. е. путь μ_i должен содержать дугу u'_i, проходящую над z. Это дает q дуг, проходящих над z: $u'_1, u'_2, \ldots, u'_q$. Беря в качестве u по очереди все дуги μ, получаем, таким образом, $v(r) - v(z) = v(r)$ дуг, проходящих над z, что противоречит предположению $W(\varphi) < v(r)$. (Аналогично можно доказать, что $v(x)$ для каждого $x \in X$ есть минимальная ширина всевозможных укладок поддерева $G(x)$.)

5. ОБСУЖДЕНИЕ ПОЛУЧЕННЫХ РЕЗУЛЬТАТОВ

Если дерево G представляет процесс вычисления, как это описано во введении, то объекты, обсуждавшиеся в разд. 2—4, имеют следующую интерпретацию. Лист дерева G представляет операцию, выполняемую исходя только из начальных данных, а узел дерева G — операцию, использующую по крайней

[1]) Действительно, пусть $y = y_{i_s}$; поскольку дуга (x, y) помечена, то $s \leqslant j$, т. е. $v(y_{i_f}) \leqslant v(y_{i_s})$; $q = v(x) - v(y) = v(y_{i_f}) + j - v(y_{i_s}) \leqslant j$; но в узле x начинается $(j+1)$ помеченных дуг, и из $q \leqslant j$ следует (2). — *Прим. перев.*

мере один промежуточный результат в качестве своего аргумента. Корень дерева G представляет операцию подсчета конечного результата. Укладка φ дерева G изображает допустимый порядок выполнения операций; она может рассматриваться как некоторая программа для вычисления, изображаемого деревом G. Ширина $W(\varphi)$ представляет собой уменьшенное на 1 число клеток памяти, требуемых для выполнения этой программы. Для операции x поддерево $G(x)$ изображает часть вычисления, состоящую из x и всех операций, требуемых для установления значений аргументов x. Теорема 1 утверждает, что последовательность $\varphi(x)$, появляющаяся в строке таблицы, которая соответствует вершине x, есть программа для „частичного" вычисления $G(x)$ и что число $v(x)$, появляющееся в той же строке, есть уменьшенное на 1 количество клеток памяти, требуемое для выполнения этой программы. Предложенный здесь алгоритм состоит в построении таких „частичных" программ. Сначала мы строим наиболее элементарные „частичные" программы, используя правило 1, а затем сочетаем их по правилу 2, получая все более и более сложные программы.

Последовательности $\varphi(y_{i_j})$, имеющиеся в правиле 2, представляют собой программы нахождения значений аргументов операции x. Объединяя их в одну программу $\varphi(x)$, мы упорядочиваем их по убыванию требования количества памяти. Этот принцип совпадает с принципом, изложенным в работе [1]; наш алгоритм в основе такой же, как и предложенный там. Доказательство, проведенное в разд. 4, может рассматриваться как формальное доказательство алгоритма, приведенного в работе [1].

ЛИТЕРАТУРА

1. **Nakata I.**, On compiling algorithms for arithmetic expressions, *Comm. ACM*, **10**, № 8 (1967), 492—494.

Плоские карты с заданными степенями вершин и граней[1])

Б. Грюнбаум

1. Введение. В 3-связном графе G, уложенном на плоскости, обозначим через $v_k(G)$ число вершин степени k, а через $p_k(G)$ — число граней степени k (число k-угольников карты, которая определяется графом G). Рассмотрим здесь только графы без петель и параллельных ребер, поэтому $v_k(G)$ и $p_k(G)$ определены лишь для $k \geqslant 3$. Следующее выражение получаем из формулы Эйлера:

$$\sum_k (4-k)\, p_k(G) + \sum_k (4-k)\, v_k(G) = 8.$$

Продолжая и обобщая различные исследования, касающиеся теоремы Эберхарда (см. работы [1—5, 7], а также работу [6], в которой даны исторические замечания об этой теореме и ее вариантах), рассмотрим следующую проблему: пусть даны последовательности $p = (p_3, \ldots, p_n)$ и $v = (v_3, \ldots, v_m)$ неотрицательных целых чисел, удовлетворяющих уравнению

$$\sum_{k=3}^{n} (4-k)\, p_k + \sum_{k=3}^{m} (4-k)\, v_k = 8. \tag{1}$$

Выясним, можно ли пару p, v *реализовать*, т. е. существует ли 3-связный плоский граф G, такой, что

$$p_k(G) = p_k \quad \text{и} \quad v_k(G) = v_k$$

для всех $k \neq 4$.

По теореме Штейница [4, 8, 9] граф G изоморфен графу, образованному ребрами и вершинами 3-мерного выпуклого многогранника тогда и только тогда, когда G плоский и 3-связный; поэтому легко понять значение этой проблемы в комбинаторной теории выпуклых многогранников.

Так как для плоского графа G число ребер равно половине суммы

$$\sum_{k \geqslant 3} k p_k(G) = \sum_{k \geqslant 3} k v_k(G),$$

---

[1]) Grünbaum B., Planar maps with prescribed types of vertices and faces, *Mathematika*, **16**, № 1 (1969), 28—36.

то еще одним необходимым условием реализуемости пары p, v является четность сумм

$$\sum_{k=3}^{n} kp_k \quad \text{и} \quad \sum_{k=3}^{m} kv_k. \tag{2}$$

Докажем следующую теорему.

Теорема. *Если последовательности* $p=(p_3, \ldots, p_n)$ *и* $v=(v_3, \ldots, v_m)$ *неотрицательных целых чисел удовлетворяют условиям* (1) *и* (2), *то пара* p, v *реализуема.*

2. Доказательство теоремы. Для доказательства теоремы построим 3-связный плоский граф для каждой пары p, v, удовлетворяющей (1) и (2). Из построения будет видно, что все получаемые графы плоские и 3-связные. Поэтому эти свойства особо рассматриваться не будут, а все внимание мы сосредоточим на получении заданных степеней вершин и граней.

Удобно разбить доказательство теоремы на ряд случаев.

Случай 1. $v_k=0$ для $k\neq 4$. Уравнение (1) примет вид

$$\sum_{k\geqslant 3} (4-k)\, p_k = 8. \tag{3}$$

Мы должны построить однородный граф G степени 4, такой, что $p_k(G)=p_k$ для всех $k\neq 4$.

Начнем с построения специальных графов, которые будем называть *блоками*. Для построения *r-блока* ($r\geqslant 4$) возьмем квадрат, на каждой из его двух смежных сторон отметим $r-4$ точек, которые соединим, как указано на рис. 1 (для $r=9$). Если точки выбраны соответствующим образом [1]), то каждый r-блок будет содержать один r-угольник, $(r-4)$ треугольников и C_{r-4}^2 четырехугольников. При этом все внутренние вершины имеют степень 4, а на двух сторонах квадрата будет 4 вершины степени 2 и $2(r-4)$ вершины степени 3.

Пусть дана последовательность p, удовлетворяющая уравнению (3). Для каждого $k\geqslant 5$ образуем p_k k-блоков. Из этих блоков и дополнительных четырехугольников построим *большой блок* (superblock), так, как указано на рис. 2. (Общий случай мы не будем описывать ввиду очевидности построения.)

„Внешний" вид большого блока такой же, как и самого блока. Большой блок включает необходимое число k-угольников для всех $k\geqslant 5$, но для выполнения уравнения (1) еще требуется 8 треугольников.

---

[1]) Линии можно проводить **как угодно, лишь бы выполнялось** условие: любые две линии пересекаются **в одной точке.** — ***Прим. перев.***

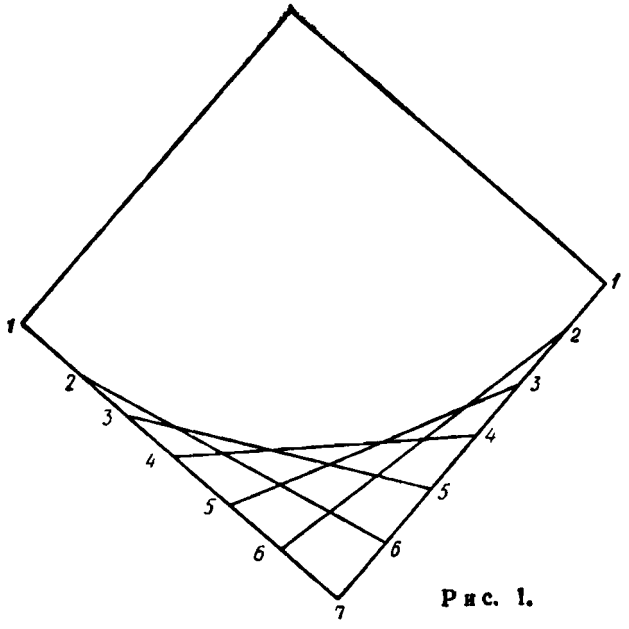


Рис. 1.

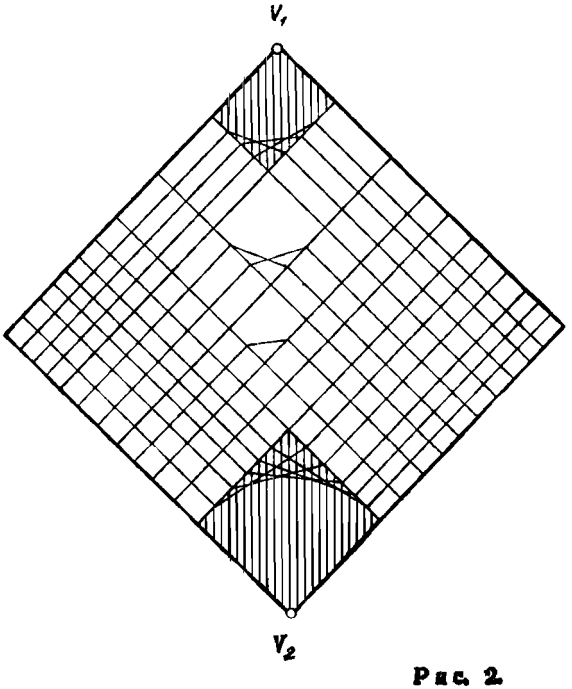


Рис. 2.

Построение теперь легко завершить способом, указанным на рис. 3 для большого блока, представленного на рис. 2. Так как на этом этапе добавляются только четырехугольники и 8 треугольников, обеспечивающие всем вершинам степень 4, мы получаем требуемый граф. Доказательство теоремы в случае 1 завершено.

Замечание. Рассмотренную выше конструкцию можно найти в работе [4; § 13.3]. Она была приведена здесь для полноты

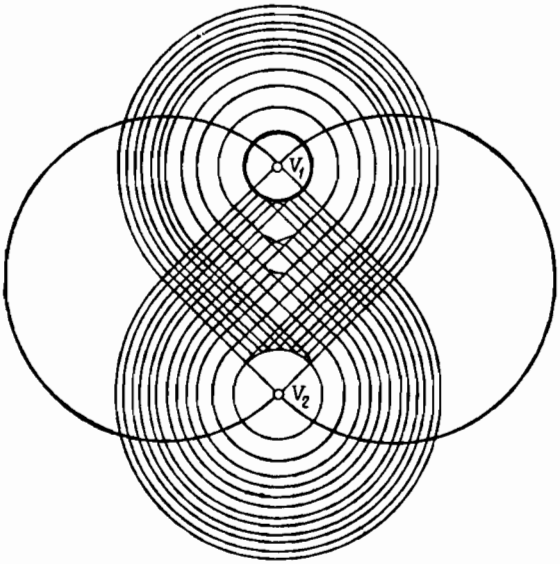


Рис. 3.

изложения, а также для выделения следующих этапов решения, необходимых при изучении оставшихся случаев:

(i) Два блока, заштрихованные на рис. 2, назовем *верхним* и *нижним*. Подчеркнем здесь важность того, какие именно k-блоки нужно разместить в эти выделенные области. (Отметим, что ориентация нижнего блока отлична от ориентации других блоков.)

(ii) Произвольно большой набор четырехугольников (далее используется термин *четырехугольная решетка*) можно организовать в большой блок, реализующий любую последовательность при добавлении необходимого числа 4-блоков в конструкцию больших блоков. Чтобы в некоторых особых случаях (таких, как $p_k = 0$ для всех $k \geqslant 5$) не менять эту конструкцию,

будем всегда в большой блок включать по крайней мере два 4-блока.

Случай 2.

$$\sum_{k=3}^{n}(4-k)\,p_k=\sum_{k=3}^{m}(4-k)\,v_k=4, \quad \text{где} \quad v_3\geqslant 5.$$

Обозначим через v' последовательность: $v'_3=v_3+3$, $v'_{m-1}=v_{m-1}+1$, $v'_m=v_m-1$, $v'_k=v_k$ для $k\neq 3$, $m-1$, m. Последовательность v' удовлетворяет уравнению (3) (последователь-

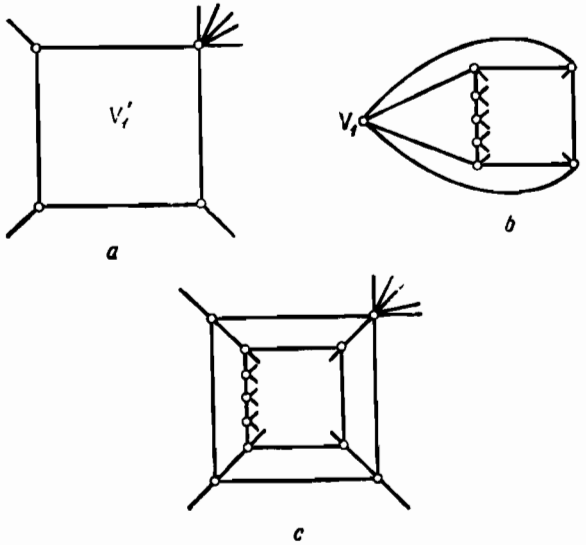


Рис. 4.

ность v' соответствует p в случае 1). Построим, как и в случае 1, однородный граф G степени 4, имеющий v_k граней степени k. Возьмем в качестве верхнего блока $(m-1)$-блок (далее будем использовать термин „вершинная компонента" конструкции). Теперь рассмотрим граф G', двойственный к G. Все грани графа G' суть четырехугольники, и v'_k вершии имеют степень k $(k\neq 4)$. Верхняя вершина V_1 графа G соответствует в графе G' четырехугольнику V'_1, который имеет 3 вершины степени 3 и одну вершиу степени $(m-1)$ (рис. 4, *a*, где $m=8$).

Пусть p' — последовательность, определяемая соотношениями: $p'_3=p_3+3$, $p'_{n-1}=p_{n-1}+1$, $p'_n=p_n-1$ и $p'_k=p_k$ для $k\neq 3$, $n-1$, n. Эта последовательность также удовлетворяет уравне-

нию (3), и, как в случае 1, мы можем построить однородный граф G'' степени 4 с p'_k гранями степени k ($k \neq 4$), имеющий в качестве верхнего блока $(n-1)$-блок. Представим G'' на плоскости так, как это показано на рис. 4, *b* (на котором квадрат соответствует жирной линии, окружающей вершину V_1 на рис. 3). Наконец, объединим графы G' и G'', выбросив V_1 из G'', помещая оставшуюся часть G'' в четырехугольник V'_1 в G' и соединяя ребрами четыре пары вершин, расположенные друг против друга (рис. 4, *c*). Как легко проверить, полученный граф имеет p_k граней степени k и v_k вершин степени k для всех $k \neq 4$. Этим завершается построение в случае 2.

Случай 3.

$$\sum_{k=3}^{n} (4-k)\, p_k > 4 > \sum_{k=3}^{m} (4-k)\, v_k \quad \text{для} \quad v_3 \geqslant 5.$$

Теперь начнем с выбора последовательности w целых чисел, в которой $w_3 = 0$, $w_5 \geqslant 0$, ..., $w_m \geqslant 0$ и такой, что будем иметь

$$\sum_{k \geqslant 3} (4-k)\, v'_k = 4,$$

положив $v'_k = v_k - w_k$ [1]). Дальнейшее построение зависит от четности суммы $\sum_{k \geqslant 5} k w_k$.

Подслучай 3.1. Сумма $\sum_{k \geqslant 5} k w_k$ четна.

Определим последовательность p' с помощью

$$p'_3 = p_3 - \sum_{k \geqslant 5} (4-k)\, w_k \quad \text{и} \quad p'_k = p_k \quad \text{для} \quad k \geqslant 5.$$

Легко проверить, что последовательности p', v' удовлетворяют условиям, наложенным на p и v в случае 2. К графу, построенному так, как это делается в случае 2, реализующему p', v' и содержащему достаточно большую четырехугольную решетку, применим следующие преобразования, которые, однако, еще не обеспечивают конструкцию треугольниками и вершинами с большими степенями.

---

[1]) Существование такой последовательности можно доказать следующим образом. Обозначим $\sum_{k=5}^{n} (4-k)\, p_k = t$, тогда $p'_3 = p_3 - (t-4)$ и $t - 4 = -\sum_{k=5}^{n} (4-k)\, p_k$. Например, можно взять $w_5 = t-4$, $w_k = 0$ для $k \geqslant 6$. Тогда $p'_3 > 0$, так как иначе получаем $t = p_3 + \sum_{k \geqslant 5}^{n} (4-k)\, p_k < p_3$ и $p_3 < t-4$, т. е. $t < p_3 < t-4$. — *Прим. перев.*

Первое преобразование можно описать следующим образом. Для любого целого числа $r \geqslant 3$ в решетке стянем в точку простую цепь, содержащую $r-2$ ребра; на рис. 5 приведен случай $r=5$. Эта операция дает одну вершину степени $2r$ и

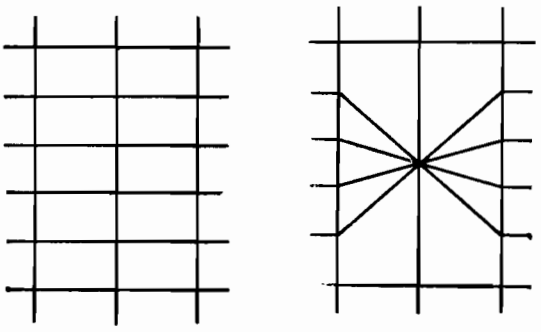


Рис. 5.

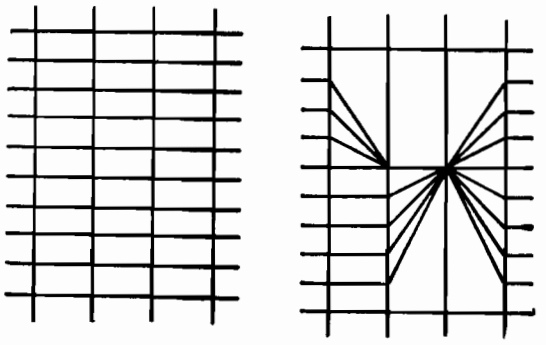


Рис. 6.

$2r-4$ треугольников, а остальная часть графа остается без изменений.

Другим преобразованием для нечетных чисел $t \geqslant s \geqslant 5$ стягиваются в точки $s-4$ ребер и $(t+s-8)/2$ ребер данной решетки способом, представленным на рис. 6 для $s=7$, $t=15$. В результате получаем одну вершину степени s, одну вершину степени t и $s+t-8=(s-4)+(t-4)$ треугольников; остальная часть графа опять остается без изменений.

Повторяя преобразования указанных двух типов необходимое число раз, завершаем доказательство.

Подслучай 3.2. Сумма $\sum_{k \geqslant 5} k w_k$ нечетна.

Пусть $r \geqslant 5$ — нечетное число, такое, что $w_r \geqslant 1$. Определим последовательности v'', p'' с помощью соотношений $v''_3 = v_3 + 1$, $v''_{r-1} = v_{r-1} + 1$, $v''_r = v_r - 1$, $v''_k = v_k$ для $k \neq 3$, $r-1$, r и $p''_3 = p_3 - 2$, $p''_k = p_k$ для $k \neq 3$. Эти последовательности удовлетворяют предположениям подслучая 3.1. Пусть H — граф, по-

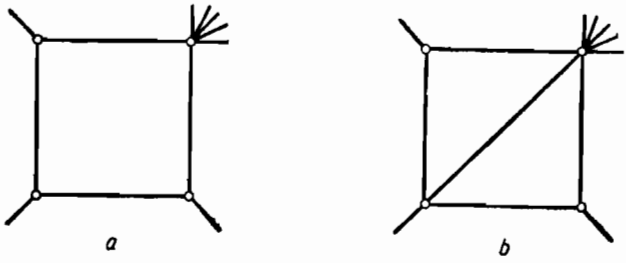


Рис. 7.

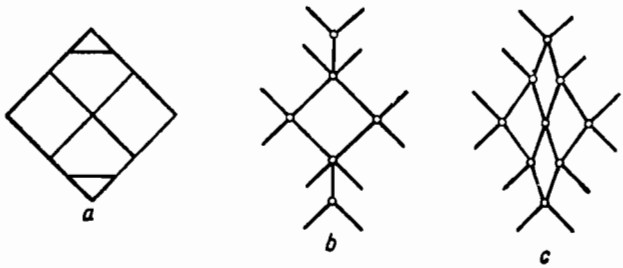


Рис. 8.

строенный, как в подслучае 3.1, и реализующий p'', v'' таким образом, что нижним блоком, который используется в построении „вершинной компоненты", является $(r-1)$-блок. Из задания графа H следует, что он включает подграф (изображен на рис. 7, *a*), содержащий четырехугольник V'_2 с тремя вершинами степени 3 и одной — степени $(r-1)$. Проведя в V'_2 диагональ (рис. 7, *b*), получаем из H граф, реализующий данные последовательности p, v. Этим завершается доказательство этого подслучая.

Случай 4. $\sum\limits_{k \geqslant 3} (4-k)\, p_k \geqslant 4 \geqslant \sum\limits_{k \geqslant 3} (4-k)\, v_k$ и $0 \leqslant v_3 \leqslant 4$.

Заменим последовательность v новой последовательностью $v' = v'_3 = v_3 + 6$, $v'_5 = v_5 + 6$ и $v'_k = v_k$ для $k \neq 3, 5$. Последовательности p и v' удовлетворяют предположениям случаев 2 или 3, поэтому, как и выше, можно построить реализующий их граф G. Единственным дополнительным шагом в этом по-

строении является размещение 6 „новых“ 5-блоков „вершинной компоненты“ в трех противоположных парах, вид которых показан на рис. 8, *a*. Каждой такой паре в графе G соответствует подграф, показанный на рис. 8, *b*. Каждый из трех подграфов затем преобразуется так, как указано при переходе от рис. 8, *b* к рис. 8, *c*. Полученный граф реализует данные последовательности p, v.

Случай 5. $\sum_{k\geqslant 3}(4-k)p_k<4<\sum_{k\geqslant 3}(4-k)v_k$.

Доказательство в этом случае можно получить из уже рассмотренных случаев, используя переход к двойственным графам.

Таким образом, все случаи разобраны и доказательство теоремы закончено.

Замечания. Построение, используемое при доказательстве теоремы, приводит, вообще говоря, к очень большому числу четырехугольников и вершин степени 4. Было бы интересно найти нижнюю границу числа $p_4(G)+v_4(G)$ в зависимости от p и v.

Эберхард [3] получил следующий результат: пусть $p_3, \ldots, p_n$, r — целые неотрицательные числа, удовлетворяющие уравнению $\sum_{k\geqslant 3}(6-k)p_k=12+2r$, тогда существует 3-связный плоский граф G, такой, что $p_k(G)=p_k$ для всех $k\neq 6$ и $\sum_{k\geqslant 4}(3-k)v_k(G)=-r$.

Как обобщение этого результата и по аналогии с доказанной теоремой выдвигается следующее предположение.

Предположение 1. *Если* $p_3, \ldots, p_n$; $v_3, \ldots, v_m$ — *неотрицательные целые числа, такие, что сумма* $\sum_{k\geqslant 3}kv_k$ *четна и*

$$\sum_{k\geqslant 3}(6-k)p_k+2\sum_{k\geqslant 3}(3-k)v_k=12, \tag{4}$$

то существует 3-*связный плоский граф* G, *у которого* $p_k(G)=p_k$ *для всех* $k\neq 6$ *и* $v_k(G)=v_k$ *для всех* $k\neq 3$.

Если допустить кратные ребра и петли и отказаться от 3-связности, то можно выдвинуть второе предположение.

Предположение 2. *Пусть* $p_1, \ldots, p_n$; $v_1, \ldots, v_m$ — *неотрицательные целые числа, такие, что сумма* $\sum_{k\geqslant 1}kv_k$ *четна и*

$$\sum_{k\geqslant 1}(4-k)p_k+\sum_{k\geqslant 1}(4-k)v_k=8$$

$\left(\text{*соответственно* } \sum_{k\geqslant 1}(6-k)p_k+2\sum_{k\geqslant 1}(3-k)v_k=12\right)$. *Тогда существует связный плоский граф* G, *такой, что* $p_k(G)=p_k$ *и*

$v_j(G) = v_j$ *для всех* $k \neq 4$ *и* $j \neq 4$ (*соответственно для всех* $k \neq 6$ *и* $j \neq 3$). *Однако если* $p_1 = v_1 = 0$, *то можно предположить, что будет существовать 2-связный плоский граф* G.

В частном случае, когда $v_k = 0$ для $k \neq 4$ (соответственно для $k \neq 3$), предположение 2 было доказано Рауландом [7] [1]).

По-видимому, доказанная теорема и приведенные выше предположения можно обобщить на графы, которые укладываются на 2-мерном многообразии рода g, поставив в правой части условий (1), (4) и др. множитель $(1-g)$.

ЛИТЕРАТУРА

1. Barnette D., On p-vectors of simple 3-polytopes, *J. Combinat. Theory* (в печати).
2. Böttger G., Harders H., Note on a problem by S. L. Hakimi concerning planar graphs without parallel elements, *J. SIAM*, 12 (1964), 838—839.
3. Eberhard V., Zur Morphologie der Polyeder, Teubner, Leipzig, 1891.
4. Grünbaum B., Convex polytopes, New York, 1967.
5. Grünbaum B., A companion to Eberhard's theorem, *Israel J. Math.* (в печати).
6. Grünbaum B., Polytopes, graphes, and complexes, *Bull. Amer. Math. Soc.* (в печати).
7. Rowland D., An extension of Eberhard's theorem, M. Sc. Thesis, University of Washington, 1968.
8. Steinitz E., Polyeder und Raumeinteilung, Enzykl. math. Wiss., v. 3 (Geometrie), Part 3AB 12, 1922, S. 1—139.
9. Steinitz E., Rademacher H., Vorlesungen über die Theorie der Polyeder, Berlin, 1934.
10*. Chvátal V., Planarity of graphs with given degrees of vertices, *Nieuw arch. wiskunde*, 17, № 1 (1969), 47—60.

---

[1]) В работе [10*] было получено необходимое условие реализуемости. Невозрастающую последовательность положительных целых чисел $d_1, d_2, \ldots, d_n$, $n \geqslant 3$, можно реализовать степенями n вершин плоского графа, если выполняется неравенство

$$\sum_{i=1}^{k} d_i \leqslant s(k, n),$$

где

$$s(k, n) = \begin{cases} k(n-1), & \text{если } 1 \leqslant k \leqslant 2, \\ 2n + 6k - 16, & \text{если } 3 \leqslant k < \dfrac{n+4}{3}, \\ 3n + 3k - 12, & \text{если } \dfrac{n+4}{3} \leqslant k \leqslant n. \end{cases}$$

Для заданных k, r существует такая последовательность $d_1, d_2, \ldots, d_n$, допускающая плоскую реализацию, что имеет место $\sum_{i=1}^{k} d_i = s(k, n)$ для $1 \leqslant k \leqslant n$. — *Прим. перев.*

Вложение графов в ориентируемые поверхности[1])[2])

Дж. Ч. Боланд

§ 1. В работе [1] мы дали необходимые и достаточные условия вложимости графа в проективную плоскость. Возникает вопрос, применимы ли методы, использованные в [1], для получения необходимых и достаточных условий вложимости графа в другие поверхности. В данной статье мы будем называть замкнутое ориентируемое 2-мерное многообразие рода g ориентируемой поверхностью и обозначать ее T_g. Можно было бы ожидать, что, например, в случае тора T_1 справедливо следующее утверждение, аналогичное теореме Маклейна для T_0 [2] и теореме в работе [1] для проективной плоскости.

Утверждение. *Граф X вложим в T_1 тогда и только тогда, когда существует множество циклов $C_i (i = 1, \ldots, n)$ в X, такое, что:*

1) *каждое ребро графа X принадлежит не более чем двум циклам из этого множества;*

2) *циклы $C_i (i = 1, \ldots, n)$ порождают подгруппу $N(C_1, \ldots, C_n)$ группы гомологий $H_1(X, Z)$, такую, что $H_1(X, Z)/N(C_1, \ldots, C_n)$ — свободная группа ранга 2.*

Однако это утверждение не имеет места, как показывает следующий контрпример. Пусть C — меридиан тора T_1, а p — некоторая точка T_1, $p \notin C$. Факторпространство $T_1' = T_1/(C \cup \{p\})$ есть ориентируемое псевдомногообразие, у которого первая группа гомологий с целыми коэффициентами $H_1(T_1')$ есть свободная абелева группа ранга 2.

---

[1]) Boland J. Ch., Embedding of graphs into orientable surfaces, *Indagationes Mathematicae*, **XXIX**, Fasc 1 (1967), 33—44 (*Proceedings Koninklijke Nederlandse Akademie van Wetenschappen*, **LXX**, ser. A, № 1).

[2]) В статье предполагаются известными некоторые понятия и теоремы алгебраической топологии. Для справок можно рекомендовать, например, книги [5*, 6*]. При доказательстве основной теоремы существенно используются результаты статей [2, 4], переводы которых помещены в настоящем сборнике (стр. 133, 145). — *Прим. ред.*

Как показывает рис. 1, факторпространство T_1' содержит два графа Куратовского K_1 и K_2, которые имеют не более одной общей точки. Рассмотрим триангуляцию факторпространства T_1', такую, что ее 1-остов X содержит $K_1 \cup K_2$. Возьмем границы 2-симплексов в качестве циклов $C_i \subset X\,(i = 1, \ldots, n)$; ясно, что граф X удовлетворяет условиям высказанного утверждения. Однако X нельзя вложить в T_1, поскольку этого нельзя сделать уже с графом $K_1 \cup K_2$. Для доказательства предположим,

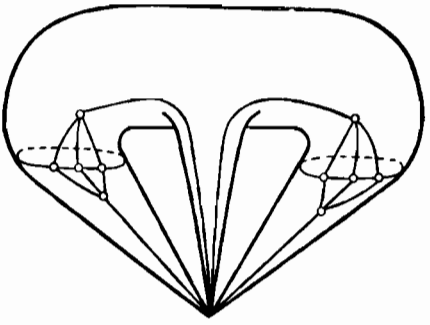


Рис. 1.

что существует вложение f: $K_1 \cup K_2 \to T_1$. Легко убедиться, что компоненты из $T_1 \setminus f(K_1) = \{x \mid x \in T_1,\ x \notin f(K_1)\}$ суть 2-клетки.

Следовательно, $f(K_2)$ содержится в замыкании некоторой 2-клетки E, с границей ∂E которой $f(K_2)$ имеет ровно одну общую точку. Рассмотрим последовательность

$$K_2 \xrightarrow{f} \overline{E} \xrightarrow{\varphi} \overline{E}/\partial E,$$

в которой $\overline{E}$ — замыкание E в топологии T_1, а φ — естественное отображение $\overline{E}$ на факторпространство $\overline{E}/\partial E$. Мы получили гомеоморфизм φf графа K_2 в 2-сферу $\overline{E}/\partial E$, что невозможно. Отсюда следует, что высказанное нами утверждение не имеет места, и мы не сможем получить обобщение теоремы Маклейна, работая с $H_1(X, Z)$. Однако мы добьемся успеха, если заменим группу $H_1(X, Z)$ фундаментальной группой $\pi_1(X)$. Прежде чем формулировать основную теорему, введем следующие *обозначения*:

1) F_r — свободная группа ранга r,

2) $A * B$ — свободное произведение групп A и B,

3) $\prod_{i=1}^{n}{}_{*} A_i$ — свободное произведение групп $A_1, \ldots, A_n$.

4) Если $C_1, \ldots, C_n$ — циклы графа X, то $N(C_1, \ldots, C_n)$ — наименьшая нормальная подгруппа в $\pi_1(X)$, определенная циклами $C_1, \ldots, C_n$.

Используя эти обозначения, сформулируем следующую теорему.

Теорема. *Связный граф X вложим в поверхность T_g тогда и только тогда, когда X содержит множество циклов $C_1, \ldots, C_n$, такое, что:*

1) *каждое ребро в X принадлежит не более чем двум циклам из этого множества,*

2) $\pi_1(X)/N(C_1, \ldots, C_n) = F_{g_0} * \prod_{i=1}^{r}{}_* \pi_1(T_{g_i})$, *причем* $\sum_{i=0}^{r} g_i \leqslant g$.

В § 3 мы докажем эту теорему, а в § 2 установим леммы, необходимые для доказательства.

§ 2. Графом мы называем конечный 1-мерный комплекс. Связность понимается как обычно. В данной статье мы рассматриваем только связные графы.

Циклом будем называть связный граф, каждая вершина которого принадлежит ровно двум ребрам. Если цикл C с базисной точкой c_0 содержится в графе X с базисной точкой x_0, то отображение вложения $i\colon C \to X$ не обязательно индуцирует гомоморфизм $\pi_1(C, c_0)$ в $\pi_1(X, x_0)$, поскольку C не обязательно содержит точку x_0. Мы выбираем путь η из x_0 в c_0 в графе X и определяем $N(C)$ как наименьшую нормальную подгруппу в $\pi_1(X, x_0)$, содержащую все пути $\eta \cdot \xi \cdot \eta^{-1}$, где $\xi \in \pi_1(C, c_0)$. Ясно, что подгруппа $N(C)$ не зависит от выбора η. Говорят, что элемент α в $\pi_1(X, x_0)$ соответствует циклу C, если α сопряжен с элементом $\eta\xi\eta^{-1}$, где ξ есть образующий в $\pi_1(C, c_0)$. Пусть $\{C_1, \ldots, C_n\}$ — множество циклов в X. Определим $N(C_1, \ldots, C_n)$ как наименьшую нормальную подгруппу в $\pi_1(X, x_0)$, содержащую все подгруппы $N(C_i)$ $(i = 1, \ldots, n)$. Если не возникает неясности, мы будем писать $\pi_1(X)$ вместо $\pi_1(X, x_0)$.

Вложение $f\colon X \to T_g$ называется *2-клеточным вложением* тогда и только тогда, когда компоненты $T_g \setminus f(X)$ суть 2-клетки.

Лемма 2.1. *Пусть X есть связный граф, $x_0 \in X$, $z \in T_g$, и пусть*

$$f\colon (X, x_0) \to (T_g, z)$$

есть 2-клеточное вложение. Тогда индуцированный гомоморфизм

$$f_*\colon \pi_1(X, x_0) \to \pi_1(T_g, z)$$

есть эпиморфизм.

Доказательство следует из рассмотрения точной последовательности

$$\pi_1(X,\ x_0) \to \pi_1(T_g,\ z) \to \pi_1(T_g,\ f(X)) \to 0,$$

поскольку $\pi_1(T_g,\ f(X)) = \pi_1(T_g/f(X)) = \pi_1$ (букет 2-сфер) $= 0$.

Определение. Пусть X — граф, и пусть $S = \{C_1, \ldots, C_n\}$ — множество циклов в X. Множество S называется τ-системой в X тогда и только тогда, когда каждое ребро в X принадлежит не более чем двум элементам из S.

Разделяющая точка связного пространства Y — это такая точка $y \in Y$, для которой $Y \setminus \{y\}$ несвязно. Локально разделяющая точка пространства Y — это точка $y \in Y$, которая является разделяющей для всех достаточно малых окрестностей y.

Далее, 2-мерное псевдомногообразие есть 2-мерный комплекс со следующими свойствами:

a) каждый 1-симплекс лежит на границе по крайней мере одного, но не более двух 2-симплексов;

b) для любых двух 2-симплексов ξ и η существует последовательность $\xi_1, \ldots, \xi_r$ 2-симплексов, такая, что $\xi_1 = \xi$, $\xi_r = \eta$, и два соседних элемента этой последовательности имеют общий 1-симплекс.

Лемма 2.2. *Пусть X — связный граф, и пусть $S = \{C_1, \ldots, C_n\}$ есть τ-система циклов в X. Существует свободная группа F и множество псевдомногообразий $Y_1, \ldots, Y_r$, такие, что*

$$\pi_1(X)/N(C_1, \ldots, C_n) = F * \prod_{i=1}^{r}{}_{*}\,(\pi_1(Y_i)).$$

Доказательство. Приклеим 2-клетку E_i к каждому из циклов C_i $(i = 1, \ldots, n)$. Пусть $Y = X \cup \bigcup_{i=1}^{n} E_i$, и пусть $\bar{E}_i$ — замыкание E_i в Y $(i = 1, \ldots, n)$. В множестве 2-клеток E_i $(i = 1, \ldots, n)$ введем отношение эквивалентности $\Re$ следующим образом: $E_i \Re E_j$ $(i, j = 1, \ldots, n)$ выполнено тогда и только тогда, когда существует последовательность $F_k = \bar{E}_{ik}$ $(k = 1, \ldots, s)$, такая, что $F_1 = \bar{E}_i$, $F_s = \bar{E}_j$ и $F_k \cap F_{k+1}$ содержит не меньше одного ребра из X $(k = 1, \ldots, s-1)$. Пусть $H_1, \ldots, H_r$ — классы эквивалентности по отношению $\Re$, и пусть $Y_i = \bigcup_{E_k \in H_i} \bar{E}_k$ $(i = 1, \ldots, r)$. Поскольку S есть τ-система циклов в X, ясно, что каждое Y_i

($i=1, \ldots, r$) является псевдомногообразием [1]). Пусть X_0 — подграф в X, определенный теми ребрами, которые не вошли ни в один цикл из S, и пусть X_i — максимальное дерево, содержащееся в $X \cap Y_i$ ($i=1, \ldots, r$).

Пусть теперь $Y_0 = \bigcup_{i=0}^{r} X_i$. Ясно, что $Y = \bigcup_{i=0}^{r} Y_i$. Каждое псевдомногообразие Y_i ($i=1, \ldots, r$) имеет в пересечении с $Y_0 \cup \bigcup_{i \neq j} Y_j$ ацикличное множество X_i.

Согласно теореме Ван Кампена, отсюда следует, что

$$\pi_1(Y) = \prod_{i=0}^{r}{}_{*}(\pi_1(Y_i)).$$

Поскольку Y_0 — граф, $\pi_1(Y_0)$ — свободная группа. Обозначив $\pi_1(Y_0) = F$, получим

$$\pi_1(Y) = F * \prod_{i=1}^{r}{}_{*}(\pi_1(Y_i)).$$

С другой стороны, пространство Y было построено путем приклеивания 2-клетки к каждому циклу C_i ($i=1, \ldots, n$). Поэтому

$$\pi_1(Y) = \pi_1(X)/N(C_1, \ldots, C_n) = F * \prod_{i=1}^{r}{}_{*}(\pi_1(Y_i)).$$

Лемма 2.3. *Для всякого 2-мерного псевдомногообразия Y существуют свободная группа F и замкнутое 2-мерное многообразие $\tilde{Y}$, такие, что $\pi_1(Y) = F * \pi_1(\tilde{Y})$.*

Доказательство. Если Y не имеет локально разделяющих точек, то ясно, что Y есть компактное 2-мерное многообразие, быть может, с границей. Если граница Y не пуста, то она состоит из некоторого числа циклов $C_1, \ldots, C_r$. Пусть циклу C_i ($i=1, \ldots, r$) соответствует элемент $\xi_i \in \pi_1(Y)$. Известно, что в этом случае всегда найдутся такие элементы $x_1, \ldots, x_k$, что $x_1, \ldots, x_k, \xi_1, \ldots, \xi_r$ будут системой образующих для $\pi_1(Y)$, которые удовлетворяют только одному соотношению вида

$$\alpha\xi_1 \cdot \ldots \cdot \xi_r = e,$$

[1]) Рассмотрим Y. Пользуясь гомотопической эквивалентностью, можно заменить каждую (изолированную) разделяющую точку отрезком и получить гомотопически эквивалентное пространству Y пространство, состоящее из букета замкнутых 2-многообразий, окружностей и 2-многообразий с границей. Заметим, что 2-многообразие с границей гомотопически эквивалентно букету окружностей. Окончательно получаем: пространство Y гомотопически эквивалентно букету g_0 окружностей и замкнутых многообразий.

где $\alpha = [x_1, x_2] \cdot \ldots \cdot [x_{k-1}, x_k]$, если Y ориентируемо, и $\alpha = x_1^2 \ldots x_k^2$, если Y не ориентируемо; здесь $[x_i, x_j] = x_i x_j x_i^{-1} x_j^{-1}$.

Итак, в этом случае $\pi_1(Y)$ есть свободная группа ранга $k+r-1$, и достаточно положить $F = \pi_1(Y)$, $\tilde{Y} = T_0$, чтобы убедиться в справедливости леммы 2.3.

В случае, когда граница Y пуста, лемма выполняется, если взять $\tilde{Y} = Y$ и $F = \{e\}$, где $\{e\}$ — группа, содержащая только один элемент. Таким образом, лемма доказана для псевдомногообразий без локально разделяющих точек. Возьмем теперь целое n и псевдомногообразие Y с n локально разделяющими точками, и пусть лемма выполняется для всех псевдомногообразий с не более чем $(n-1)$ такими точками. Ясно, что для локально разделяющей точки $p \in Y$ существует достаточно малая окрестность V этой точки, такая, что:

a) $V = \bigcup_{i=1}^{s} E_i$, каждое E_i есть 2-клетка,

b) $E_i \cap E_j = \{p\}$ $(i, j = 1, \ldots, n)$,

c) каждая компонента из $\bar{V} \setminus V$ есть цикл или дуга ($\bar{V}$ означает замыкание V).

Пусть $C_1, \ldots, C_r$ — те компоненты из $\bar{V} \setminus V$, которые являются циклами, а $L_{r+1}, \ldots, L_s$ — остальные компоненты. Приклеим 2-клетку E_i' к каждому из циклов C_i $(i = 1, \ldots, r)$ и рассмотрим пространство $Y' = (Y \setminus V) \cup \bigcup_{i=1}^{r} E_i'$, которое, очевидно, есть псевдомногообразие с не более чем $(n-1)$ локально разделяющими точками.

Согласно предположению индукции, имеем

$$\pi_1(Y') = F' * \pi_1(\tilde{Y}').$$

Выберем теперь точки $p_i \in E_i$ $(i = 1, \ldots, r)$, $p_j \in L_j$ $(j = r+1, \ldots, s)$ и обозначим $P = \{p_1\} \cup \ldots \cup \{p_s\}$. Рассмотрим факторпространство $Y'' = Y'/P$. Известно, что $\pi_1(Y'') = F_{s-1} * \pi_1(Y')$. Легко доказать, что Y'' гомотопически эквивалентно первоначально взятому псевдомногообразию Y, откуда получаем

$$\pi_1(Y'') = \pi_1(Y),$$

а, значит, $\pi_1(Y) = F_{s-1} * F' * \pi_1(\tilde{Y}') = F * \pi_1(\tilde{Y})$, где $F = F_{s-1} * F'$ и $\tilde{Y} = \tilde{Y}'$.

Лемма 2.4. *Пусть X есть связный граф, и пусть $S = \{C_1, \ldots, C_n\}$ есть τ-система циклов в X. Тогда:*

a) *существуют свободная группа* F *и замкнутые* 2*-мерные многообразия* $Y_1, \ldots, Y_r$, *такие, что*

$$\pi_1(X)/N(C_1, \ldots, C_n) = F * \pi_1(Y_1) * \ldots * \pi_1(Y_r); \qquad (1)$$

b) *более того, множество* S *определяет группу* F *с точностью до изоморфизма, а многообразия* $Y_1, \ldots, Y_r$ *— с точностью до гомеоморфизма.*

Доказательство. Первое утверждение следует из лемм 2.2 и 2.3. Ден [3] показал, что в разложении (1) сомножители определены с точностью до изоморфизма и, значит, соответствующие многообразия — с точностью до гомеоморфизма, что и доказывает второе утверждение.

Определение. Пусть X есть связный граф; $S = \{C_1, \ldots, C_n\}$ есть τ-система циклов в нем; F — свободная группа ранга g_0; Y_i $(i = 1, \ldots, r)$ есть 2-мерное многообразие рода g_i $(i = 1, \ldots, r)$, и пусть

$$\pi_1(X)/N(C_1, \ldots, C_n) = F * \pi_1(Y_1) * \ldots * \pi_1(Y_r).$$

Определим род $g(S)$ для τ-системы S как

$$g(S) = g_0 + g_1 + \ldots + g_r.$$

Лемма 2.5. *Пусть* S_1 *и* S_2 *есть* τ*-системы циклов связного графа* X. *Если* $S_1 \subset S_2$, *то* $g(S_2) \leqslant g(S_1)$.

Доказательство. Пусть $S_1 = \{C_1, \ldots, C_n\}$, и пусть $C \in S_2 \setminus S_1$. Рассмотрим множество $S_1' = \{C_1, \ldots, C_n, C\}$-циклов в X. Поскольку $S_1' \subset S_2$, ясно, что S_1' есть τ-система. Достаточно доказать, что $g(S_1') \leqslant g(S_1)$, так как отсюда по индукции следует, что $g(S_2) \leqslant g(S_1)$.

Приклеим 2-клетку E_i к циклу C_i $(i = 1, \ldots, n)$ и рассмотрим пространство $Y = X \cup \bigcup_{i=1}^{r} E_i$.

Как и при доказательстве леммы 2.2, пишем

$$Y = Y_0 \cup Y_1 \cup \ldots \cup Y_r,$$

где $Y_1, \ldots, Y_r$ — псевдомногообразия (быть может, с границей), а Y_0 — граф, который пересекается с каждым Y_i $(i = 1, \ldots, r)$ по ацикличному множеству. Разумеется, $\pi_1(Y) = \prod_{i=0}^{r} * (\pi_1(Y_i))$.

Предположим сначала, что C содержится в одном из пространств Y_i $(i = 1, \ldots, r)$, например $C \subset Y_1$. Согласно лемме 2.3, имеем

$$\pi_1(Y_1) = F * \pi_1(\tilde{Y}_1),$$

где $\tilde{Y}_1$ есть 2-мерное многообразие. Рассмотрим отображение вложения $i\colon C \to Y_1$ и индуцированный им гомеоморфизм $i_*\colon \pi_1(C) \to \pi_1(Y_1)$ (базисная точка в Y выбрана лежащей на C). Ясно, что C — граничный цикл псевдомногообразия Y_1. Поэтому, если ξ есть образующий группы $\pi_1(C)$, то $i_*(\xi) \in \pi_1(\tilde{Y}_1)$ [1]).

Приклеим 2-клетку E к циклу C. Получим пространство $Y_1' = Y_1 \cup E$. Снова пишем $\pi_1(Y_1') = F * \pi_1(\tilde{Y}_1')$. Ясно, что $\pi_1(\tilde{Y}_1')$ получается из $\pi_1(\tilde{Y}_1)$ факторизацией по наименьшей нормальной подгруппе, содержащей $i_*(\xi)$ [2]), а $F' = F$. Если $\tilde{Y}_1'$ имеет непустую границу, то непустую же границу имеет и $\tilde{Y}_1$. Как $\pi_1(Y_1')$, так и $\pi_1(\tilde{Y}_1)$ в этом случае суть свободные группы, причем ранг $\pi_1(\tilde{Y}_1')$ равен рангу $\pi_1(\tilde{Y}_1)$ без единицы.

Если же $\tilde{Y}_1'$ — замкнутое многообразие рода g, то $\pi_1(\tilde{Y}_1)$ — свободная группа ранга $2g$ [3]). В обоих случаях имеем по определению рода τ-системы $g(S_1') \leqslant g(S_1)$.

Пусть теперь C не содержится полностью ни в одном из псевдомногообразий $Y_1, \ldots, Y_r$. Тогда при $i = 1, \ldots, r$ $C \cap Y_i$ есть ацикличное множество. Легко показать, что в этом случае в $X \cap Y_i$ существует максимальное дерево L_i, содержащее $C \cap Y_i$ $(i = 1, \ldots, r)$. Если обозначить X_0 подграф в X, определенный теми ребрами X, которые не входят ни в один цикл τ-системы S_1, и обозначить $Y_0' = X_0 \cup \bigcup_{i=1}^{r} L_i$, то можно записать, что $Y = Y_0' \cup \bigcup_{i=1}^{r} Y_i$ и что

$$\pi_1(Y) = \pi_1(Y_0') * \prod_{i=1}^{r}{}_* (\pi_1(Y_i)).$$

Ясно, что теперь $C \subset Y_0'$. Снова рассмотрим $i\colon C \to Y_0'$. Выберем базисные точки в Y, Y_0', C так, чтобы они совпадали, тогда $i_*(\xi) \in \pi_1(Y_0')$ (здесь ξ есть образующий в $\pi_1(C)$). Оче-

---

[1]) Автор пользуется здесь леммой 2.3 не в том виде, в котором она сформулирована, а следующим утверждением, легко вытекающим из доказательства леммы 2.3: если псевдомногообразие Y имеет p локально разделяющих точек, то существует многообразие $\tilde{Y}$ (возможно, с границей), такое, что $\pi_1(Y) = \pi_1(\tilde{Y}) * F_p$. — *Прим. перев.*

[2]) Эта подгруппа может состоять лишь из единицы. — *Прим. перев.*

[3]) В этом случае $\tilde{Y}_1$ получается из многообразия рода g вырезанием дыры. — *Прим. перев.*

видно, что $\pi_1(Y_0)$ изоморфно $\pi_1(Y_0')$[1]. Приклеим 2-клетку E к циклу C, тогда $\pi_1(Y \cup E) = A * \prod_{i=1}^{r} * (\pi_1(Y_i))$, где A получено из $\pi_1(Y_0')$ факторизацией по наименьшей нормальной подгруппе, содержащей $i_*(\xi)$. Но $\operatorname{rank} A \leqslant \operatorname{rank} \pi_1(Y_0') = \operatorname{rank} \pi_1(Y_0)$, так что и в этом случае $g(S_1') \leqslant g(S_1)$. Лемма 2.5 доказана.

По определению τ-система S в графе X называется *максимальной*, если S не содержится ни в какой τ-системе, отличной от S.

§ 3. В этом параграфе мы докажем следующую теорему.

Теорема 3.1. *Связный граф X может быть вложен в ориентируемую поверхность T_g рода g тогда и только тогда, когда в X существует τ-система циклов $S = \{C_1, \ldots, C_n\}$, для которой выполнены условия*

1) $g(S) \leqslant g$,

2) $\pi_1(X)/N(C_1, \ldots, C_n) = F * \prod_{i=1}^{r} * (\pi_1(T_{g_i}))$[2].

Доказательство. Сначала докажем необходимость этих условий индукцией по g. При $g = 0$ поверхность T_g является 2-сферой. Пусть связный граф X вкладывается в T_0. В этом случае, как утверждает теорема Маклейна, существует τ-система $S = \{C_1, \ldots, C_n\}$ циклов в X, для которой

$$\pi_1(X)/N(C_1, \ldots, C_n) = 0.$$

Очевидно, что условия 1) и 2) при этом выполнены.

Возьмем теперь натуральное число g. Предположим, что в каждом связном графе, который вкладывается в поверхность $T_{g'}$, $0 \leqslant g' < g$, существует τ-система циклов, удовлетворяющая условиям теоремы. Рассмотрим связный граф X, вкладывающийся в поверхность T_g. Согласно Янгсу [4], в этом случае существует 2-клеточное вложение X в поверхность $T_{g'}$, где $0 \leqslant g' \leqslant g$. Если $g' < g$, то в X по предположению индукции существует такая τ-система, род которой не превосходит g', а следовательно, не превосходит и g, причем выполнено также условие 2). Пусть теперь $g' = g$, т. е. существует 2-клеточное вложение f графа X в T_g. По лемме 2.1 индуцированное отображение $f_*\colon \pi_1(X) \to \pi_1(T_g)$ является эпиморфизмом. Поэтому

[1]) Это следует, например, из однозначности разложения в лемме 2.3. — *Прим. перев.*

[2]) Здесь требуется больше, чем утверждает лемма 2.4, а именно, чтобы все поверхности, участвующие в разложении, были бы ориентируемыми. — *Прим. перев.*

в X существует $2g$ циклов $D_1, \ldots, D_{2g}$, таких, что для соответствующих элементов $x_1, \ldots, x_{2g}$ в $\pi_1(X)$ их образы $f_*(x_1), \ldots, f_*(x_{2g})$ можно взять в качестве образующих группы $\pi_1(T_g)$, связанных единственным соотношением

$$[y_1, y_2] \cdot \ldots \cdot [y_{2g-1}, y_{2g}] = e \qquad (\text{здесь } y_i = f_*(x_i)).$$

Разрезав T_g по каждому из циклов $f(D_i)$ $(i = 1, \ldots, 2g)$, мы получим замкнутую 2-клетку P и непрерывное отображение $\alpha\colon P \to T_g$, сужение которого на внутренность P есть гомеоморфизм.

Более того, $\alpha(\partial P) \subset f(X)$.

Рассмотрим множество $X' = \alpha^{-1}(f(X))$. Очевидно, что X' есть плоский граф, поэтому в X' существует τ-система циклов $S' = \{C_1', \ldots, C_n'\}$, для которой

$$\pi_1(X')/N(C_1', \ldots, C_n') = 0.$$

Согласно [4], каждый цикл C_i' $(i = 1, \ldots, n)$ лежит на границе некоторой компоненты из $P \setminus X'$ [1]). Каждое ребро границы ∂P принадлежит ровно одному циклу C_i' $(i = 1, \ldots, n)$. Предположим сначала, что $\alpha|_{C_i'}$ есть гомеоморфизм при всех $i = 1, \ldots, n$. Положим $C_i = f^{-1}\alpha(C_i')$ $(i = 1, \ldots, n)$. Ясно, что множество $S = \{C_1, \ldots, C_n\}$ является τ-системой в X, причем $N(C_1, \ldots, C_n)$ есть ядро отображения $f_*\colon \pi_1(X) \to \pi_1(T_g)$. Поскольку f_* эпиморфизм, имеем $\pi_1(X)/N(C_1, \ldots, C_n) = \pi_1(T_g)$, а значит, S удовлетворяет условиям 1) и 2) нашей теоремы.

Пусть теперь $\alpha|_{C_i'}$ не есть гомеоморфизм при некотором i, например, $\alpha|_{C_1'}$ не есть гомеоморфизм. Тогда существуют точки a и b на C_1', $a \neq b$, такие, что $\alpha(a) = \alpha(b)$. Поскольку C_1' лежит на границе некоторой компоненты из $P \setminus X'$, существует дуга $L \subset P$, для которой

$$L \cap X' = \{a\} \cup \{b\}.$$

Очевидно, $D = \alpha(L)$ есть цикл на T_g, причем $D \cap X = \{p\}$, где $p = \alpha(a) = \alpha(b)$. Разрежем поверхность T_g по циклу D. Пусть D' и D'' — циклы, полученные из D при разрезании. В зависимости от того, разделяет или нет цикл D поверхность T_g, рассмотрим два случая.

---

[1]) То есть на границе **некоторой 2-клетки.** — ***Прим. перев.***

Пусть D разделяет поверхность T_g [1]). При этом мы получаем две поверхности $T_{g'}$ и $T_{g''}$, причем $g'+g''=g$. Поскольку цикл D не ограничивает на поверхности T_g никакой 2-клетки [2]), то $g'>0$ и $g''>0$, и, следовательно, $g'<g$, $g''<g$. Пусть X_1 и X_2 означают подграфы графа X, попавшие соответственно на $T_{g'}$ и $T_{g''}$. По предположению индукции существуют τ-система $S_1=\{C_1, \ldots, C_r\}$ циклов из X_1 и τ-система $S_2=\{C_{r+1}, \ldots, C_m\}$ циклов из X_2, для которых

1) $g(S_1)\leqslant g'$,
$g(S_2)\leqslant g''$,

2) $$\pi_1(X_1)/N(C_1, \ldots, C_r)=F_1*\prod_{i=1}^{r_1}{}_*(\pi_1(T_{g_i})),$$

$$\pi_1(X_2)/N(C_{r+1}, \ldots, C_m)=F_2*\prod_{i=1}^{r_2}{}_*(\pi_1(T_{g_i})).$$

Приклеим к каждому циклу C_i $(i=1, \ldots, m)$ 2-клетку E_i и рассмотрим $Y_1=X_1\cup\bigcup_{i=1}^{r}E_i$, $Y_2=X_2\cup\bigcup_{i=r+1}^{m}E_i$. Имеем

$$\pi_1(Y_1)=\pi_1(X_1)/N(C_1, \ldots, C_r) \text{ и } \pi_1(Y_2)=\pi_1(X_2)/N(C_{r+1}, \ldots, C_m).$$

Если отождествить теперь точку, взятую на $X_1\cap D'$, с точкой, взятой на $X_2\cap D''$, то получится пространство Y, для которого $\pi_1(Y)=\pi_1(Y_1)*\pi_1(Y_2)$.

Рассмотрим множество $S=\{C_1, \ldots, C_m\}$ как τ-систему в X. Ясно, что $\pi_1(X)/N(C_1, \ldots, C_m)=\pi_1(Y_1)*\pi_1(Y_2)$, т. е. $g(S)\leqslant g'+g''=g$. Условия нашей теоремы выполнены.

Остается случай, когда D не разделяет поверхность T_g. Разрезание T_g по D дает ориентируемую незамкнутую поверхность T'_{g-1} рода $g-1$, для которой D' и D'' являются граничными циклами. Разрезанию соответствует отображение η: $T'_{g-1}\to T_g$, каждой точке $x\in T'_{g-1}$ оно сопоставляет точку $\eta(x)\in T_g$, из которой x получено при разрезании. Приклеив к D' и D'' соответственно 2-клетки E' и E'', получим замкнутую поверхность $T_{g-1}=T'_{g-1}\cup E'\cup E''$. Рассмотрим $X'=\eta^{-1}(f(X))$. По предположению индукции в X' существует τ-система циклов $S''=\{C''_1, \ldots, C''_m\}$, для которой

$$\pi_1(X')/N(C''_1, \ldots, C''_m)=F*\prod_{i=1}^{r}{}_*(\pi_1(T_{g_i})),$$

[1]) То есть при разрезании по D поверхность T_g распадается на две компоненты. — *Прим. перев.*

[2]) Это легко показать, вспомнив, как выбирались циклы $D_1, \ldots, D_{2g}$. — *Прим. перев.*

причем $g(S'') \leqslant g-1$. Приклеим к каждому циклу C_i 2-клетку E_i $(i=1, \ldots, m)$ и рассмотрим $Y = X \cup \bigcup_{i=1}^{m} E_i$. Ясно, что $\pi_1(Y) = \pi_1(X')/N(C_1'', \ldots, C_m'')$. Если отождествить теперь две точки из $\eta^{-1}(p)$, получится пространство Y', для которого $\pi_1(Y') = F_1 * \pi_1(Y)$ (F_1 есть свободная циклическая группа). Если взять теперь в качестве τ-системы S множество циклов $C_i = f^{-1}\eta(C_i'')$, $i=1, \ldots, m$, то получим

$$\pi_1(X)/N(C_1, \ldots, C_m) = F * \pi_1(Y'),$$

т. е. $g(S) \leqslant 1+(g-1) = g$. Таким образом, условия 1) и 2) оказываются выполненными и здесь. Это завершает доказательство необходимости условий нашей теоремы.

Докажем теперь достаточность. Пусть X — связный граф, и пусть $S = \{C_1, \ldots, C_n\}$ есть τ-система циклов из X, удовлетворяющая условиям 1) и 2) нашей теоремы. Мы предполагаем, что среди всех таких τ-систем S обладает минимальным родом $g(S)$. Приклеим к каждому циклу C_i 2-клетку E_i $(i=1, \ldots, n)$. Рассмотрим $Y' = X \cup \bigcup_{i=1}^{n} E_i$.

В обозначениях леммы 2.2 имеем $Y' = \bigcup_{i=0}^{r} Y_i$, где Y_0 содержит, в частности, все ребра X, не вошедшие ни в один цикл из S. Поскольку $g(S)$ минимально, каждое псевдомногообразие Y_i $(i=1, \ldots, r)$ замкнуто[1]) и каждая компонента связности M_j графа $Y' \setminus \bigcup_{i=1}^{r} Y_i$ есть дерево $(j=1, \ldots, s)$.

Можно считать поэтому, что каждое M_j вложено в 2-сферу Y_{r+j} $(j=1, \ldots, s)$ так, что $Y_{r+j} \cap Y' = M_j$. Рассмотрим $Y = \bigcup_{i=1}^{r+s} Y_i$. По теореме Ван Кампена, имеем

$$\pi_1(Y) = \pi_1(Y') = \pi_1(X)/N(C_1, \ldots, C_n) = F_{g_0} * \prod_{i+1}^{r}{}_{*}(\pi_1(T_{g_i})) = \prod_{i=0}^{r+s}{}_{*}(\pi_1(Y_i)).$$

[1]) Это неточно. Лемма 2.5 утверждает лишь, что при расширении τ-системы ее род *не убывает*, а не обязательно возрастает. Легко построить контрпример, взяв сферу с дырой и парой тождественных точек. Заклейка дыры в этом случае не изменяет фундаментальной группы, равной Z. Мы можем, однако, считать, что выбранная нами τ-система соответствующим образом расширена, так что все псевдомногообразия действительно замкнуты. — *Прим. перев.*

Если вспомнить, что $\pi_1(Y_i) = F_{h_i} * \pi_1(T_{g_i})$ $(i = 1, \ldots, r+s)$, то мы получим $\sum_{i=1}^{r+s} h_i = g_0$ [1]), $g_{r+1} = \ldots = g_{r+s} = 0$.

Для некоторых целых u, v, $1 \leqslant u, v \leqslant r+s$, пусть $Y_u \cap Y_v \neq \varnothing$. Пусть $p \in Y_u \cap Y_v$. Рассмотрим замкнутые 2-клетки E и E', содержащиеся соответственно в Y_u и Y_v и такие, что $E \cap X = E' \cap X = \{p\}$. Пусть $C = \partial E$ и $C' = \partial E'$ — граничные циклы этих клеток, $\tilde{E}$ и $\tilde{E}'$ — их внутренности, и пусть $Z' = Y \setminus (\tilde{E} \cup \tilde{E}')$, $Y'_u = Y_u \setminus \tilde{E}$, $Y'_v = Y_v \setminus \tilde{E}'$. Ясно, что

$$\pi_1(Y'_u) = F_{h_u} * A,$$

где A — группа с образующими $x_1, \ldots, x_{2g_u}$, l и **единственным** соотношением

$$[x_1, x_2] \cdot \ldots \cdot [x_{2g_u-1}, x_{2g_u}] \cdot l = e,$$

а l — элемент из $\pi_1(Y'_u)$, соответствующий циклу C. **Точно так же** получаем

$$\pi_1(Y'_v) = F_{h_v} * B,$$

где B — группа с образующими $y_1, \ldots, y_{2g_v}$, m и **единственным** соотношением

$$[y_1, y_2] \cdot \ldots \cdot [y_{2g_v-1}, y_{2g_v}] \cdot m = e,$$

а m — элемент из $\pi_1(Y'_v)$, соответствующий циклу C'. **Для Z' имеем**

$$\pi_1(Z') = F_{g_0} * \coprod_{i=1}^{r+s} G_i,$$

где $G_i = \pi_1(T_{g_i})$, если $i \neq u, v$ и $G_u = A$, $G_v = B$.

Приклеив 2-клетку к Z' вдоль пути $l \cdot m$, мы получим пространство Z. Его фундаментальная группа имеет те же образующие и соотношения, что и группа $\pi_1(Z')$, и еще одно соотношение $l \cdot m = e$. Добавление этого соотношения не изменяет групп G_i при $i \neq u, v$, но приводит к замене группы $G_u * G_v$ группой $\pi_1(T_{g_u+g_v})$. Отсюда получаем

$$\pi_1(Z) = F_{g_0} * \coprod_{\substack{i=1 \\ i \neq u, v}}^{r+s} (\pi_1(T_{g_i})) * \pi(T_{g_u+g_v}).$$

---

[1]) Здесь допущена неточность. На самом деле $g_0 = h_0 + \sum_{i=1}^{r+s} h_i$, где h_0 равно рангу $\pi_1(Y_0)$. Именно так и следует понимать это в дальнейшем. — *Прим. перев.*

Очевидно, что $\sum_{i=0}^{r+s} g_i$ при этой операции не изменяется, поэтому, повторяя ее, мы в конце концов получим вложение X в псевдомногообразие Z_1, для которого

$$\pi_1(Z_1) = F_{g_0} * \pi_1(T_h), \quad g_0 + h \leqslant g.$$

Пусть q — локально разделяющая точка Z_1; U — окрестность q; U_1 и U_2 — две различные компоненты $U \setminus \{q\}$; D — цикл, содержащий точку q и такой, что $D \cap U_1 \neq \varnothing$ и $D \cap U_2 \neq \varnothing$. В группе F_{g_0} этому циклу соответствует элемент ξ. Рассмотрим 2-клетки E и E', для которых $E \cap X = E' \cap X = \{q\}$, $E \subset U_1 \cup \{q\}$, $E' \subset U_2 \cup \{q\}$. Удалим из Z_1 внутренности клеток E и E' и обозначим полученное пространство Z_1'. Имеем $\pi_1(Z_1') = F_{g_0-1} * A$, где A — группа с $2h+3$ образующими $x_1, \ldots, x_{2h}, l, m, \xi$, связанными единственным соотношением

$$[x_1, x_2] \cdot \ldots \cdot [x_{2h-1}, x_{2h}] \cdot l \cdot \xi \cdot m \cdot \xi^{-1} = e.$$

Здесь l и m — элементы группы $\pi_1(Z_1')$, соответствующие граничным циклам ∂E и $\partial E'$ клеток E и E'. Приклеив 2-клетку к Z_1' вдоль пути $l \cdot m$, мы получим пространство Z_1''. Его фундаментальная группа $\pi_1(Z_1'')$ отличается от группы $\pi_1(Z_1')$ лишь наличием одного дополнительного соотношения $l \cdot m = l$, т. е. имеем

$$\pi_1(Z_1'') = F_{g_0-1} * A',$$

где A' задается образующими $x_1, \ldots, x_{2h}, \xi, l$ и единственным соотношением

$$[x_1, x_2] \cdot \ldots \cdot [x_{2h-1}, x_{2h}] \cdot [l, \xi] = e.$$

Следовательно, $A' = \pi_1(T_{h+1})$. Таким образом, мы получили вложение X в Z_1'', где

$$\pi_1(Z_1'') = F_{g_0-1} * \pi_1(T_{h+1}).$$

Повторяя описанную операцию, можно уничтожить все локально разделяющие точки и получить вложение X в T_{g_0+h}. Поскольку $g_0 + h \leqslant g$, достаточность условий 1) и 2) доказана, а тем самым полностью закончено доказательство теоремы 3.1.

Замечание 1. В лемме 2.4 мы доказали, что для любой τ-системы $S = \{C_1, \ldots, C_n\}$ связного графа X выполнено следующее равенство:

$$\pi_1(X)/N(C_1, \ldots, C_n) = F * \pi_1(Y_1) * \ldots * \pi_1(Y_r),$$

где F — свободная группа, $Y_1, \ldots, Y_r$ — замкнутые многообразия. Отсюда следует, что группа $\pi_1(X)/N(C_1, \ldots, C_n)$ определена при помощи конечного числа образующих и конечного числа соотношений, причем каждый из образующих появляется в соотношениях этой группы не более двух раз. Ден [3] показал, что в таких группах проблема тождества слов алгоритмически разрешима, а следовательно, и наше описание графов, вкладывающихся в ориентируемую поверхность рода g, эффективно.

Замечание 2. Пусть X_1 и X_2 — два графа, p_1 — вершина в X_1, p_2 — вершина в X_2. Обозначим $X_1 \vee X_2$ граф, полученный отождествлением точек p_1 и p_2. Известно, что род графа $X_1 \vee X_2$ равен сумме родов графов X_1 и X_2. Это можно теперь легко доказать с помощью теоремы 3.1.

ЛИТЕРАТУРА

1. Boland J. Ch., Embeddings of graphs in the projective plane, *Fund. Math.*, 57 (1965), 195—203.
2. MacLane S., A combinatorial condition for planar graphs, *Fund. Math.*, 28 (1937), 271—283. (Русский перевод см. на стр. 133 настоящего сборника.)
3. Dehn M., Ueber unendliche diskontinuierliche Gruppen, *Math. Ann.*, 71 (1911), 116—144.
4. Youngs J. W. T., Minimal embeddings and the genus of a graph, *J. of Math. and Mech.*, 12, № 2 (1963). (Русский перевод см. на стр. 145 настоящего сборника.)
5*. Хилтон П., Уайли С., Теория гомологий, «Мир», М., 1966.
6*. Кроуэл Р., Фокс Р., Введение в теорию узлов, «Мир», М., 1964.

Комбинаторное условие для плоских графов[1])

С. Маклейн

1. ВВЕДЕНИЕ

Куратовский [1][2]) доказал, что топологический[3]) граф является плоским (т. е. может быть взаимно однозначно отображен в плоскость) тогда и только тогда, когда он не содержит подграфа одного из двух особенных видов. Уитни [2] показал, что граф плоский тогда и только тогда, когда у него есть комбинаторный „двойник". В данной статье устанавливается еще один критерий того, что граф плоский. Он формулируется при помощи обычных комбинаторных понятий.

Теорема I. *Комбинаторный граф является плоским тогда и только тогда, когда он содержит такую полную систему циклов, для которой никакое ребро не принадлежит более чем двум циклам.*

Граф, расположенный на плоскости, делит ее на некоторое число областей, каждая из которых ограничена одним или более циклами. Легко показать, что множество таких циклов содержит полную систему, а каждое ребро графа принадлежит не более чем двум циклам этой системы. Отсюда следует необходимость высказанных условий (разд. 5). Доказательство достаточности носит более комбинаторный характер. Для удобства мы сначала сводим задачу к случаю несепарабельных графов (разд. 3), ранее рассмотренных Уитни. Удалив из несепарабельного графа подходящее ребро, мы получаем более простой (и опять несепарабельный) граф. Его мы укладываем на плоскость, а затем показываем, что условия теоремы позволяют добавить удаленное ребро уже на плоскости (разд. 4, 5).

Чисто комбинаторными рассуждениями мы показываем затем эквивалентность критерия теоремы I критерию в терминах существования двойника (разд. 6). В свою очередь последний критерий, как известно [3], эквивалентен (в комбинаторном смысле) условию Куратовского.

---

[1]) MacLane S., A combinatorial condition for planar graphs, *Fundamenta Mathematika*, 28 (1937), 22–32.

[2]) В [1] рассматриваются не графы, а множества с более общими свойствами.

[3]) Терминология автора часто отличается от современной, однако изменять ее было бы нецелесообразно. Исключение сделано для термина „2-fold set", который переводится как „τ-система" (см. стр. 136 и стр. 139). — *Прим. ред.*

2. ОПРЕДЕЛЕНИЯ

Комбинаторный граф G состоит из конечного множества элементов $a, b, c, \ldots$, называемых *ребрами*, и конечного множества *вершин* $p, q, r, \ldots$, такого, что каждое ребро b соединяет в точности две вершины p и q (в этом случае говорят, что p и q — „концы“ ребра b, или p и q лежат на b). Мы предполагаем, что каждая вершина лежит хотя бы на одном ребре [1]). Любое множество ребер вместе со всеми лежащими на них вершинами образует *подграф* графа G. Каждый подграф полностью определяется своими ребрами.

Если $m > 1$ и если $p_1, p_2, \ldots, p_m$ — различные вершины, то подграф C, определяемый ребрами $p_1p_2, p_2p_3, \ldots, p_{m-1}p_m, p_mp_1$, называется *циклом*, а подграф D с ребрами $p_1p_2, \ldots, p_{m-1}p_m$ *цепью с концами* p_1 и p_m. Цепь называется *подвешенной*, если из всех ее вершин только p_1 и p_m лежат не менее чем на трех ребрах из G. Если A и B — подграфы, то $A \cap B$ — подграф, определенный пересечением, $A + B$ — объединением, $G - A$ — разностью соответствующих множеств ребер.

Если граф G имеет $E(G)$ ребер, $V(G)$ вершин и $P(G)$ компонент связности, то

$$R(G) = V(G) - P(G), \quad N(G) = E(G) - V(G) + P(G) \qquad (1)$$

называются соответственно *рангом* и *цикломатическим числом* графа G. *Сумма по* mod 2 циклов $C_1 + C_2 + \ldots + C_m$ — это подграф, содержащий все ребра, присутствующие в нечетном числе циклов написанной суммы. Циклы $C_1, \ldots, C_n$ образуют *полную систему* в G, если каждый цикл из G единственным образом представляется в виде суммы mod 2 некоторых из циклов $C_1, \ldots, C_n$. Каждый граф G содержит по крайней мере одну полную систему из $n = N(G)$ циклов.

Плоский *топологический граф* H состоит из конечного числа дуг (гомеоморфных образов отрезка) на плоскости. Две таких дуги либо не пересекаются, либо точка их пересечения является концом каждой из двух дуг. Дуги графа H и их концы, очевидно, образуют комбинаторный граф H', который тоже называется *плоским*.

3. НЕСЕПАРАБЕЛЬНЫЕ ГРАФЫ

Граф G называется *сепарабельным*, если он обладает двумя подграфами F_1 и F_2, такими, что $F_1 + F_2 = G$, а F_1 и F_2 не имеют общих ребер и могут иметь не более одной общей вершины.

[1]) Исключение „изолированных“ вершин не влияет, очевидно, на выводы теоремы I.

Если F_1 и F_2 не имеют общей вершины, то они не связаны [1]), а если такая вершина p есть, ее называют *разделяющей* вершиной графа G. Про сепарабельный граф G говорят, что его можно *разделить* на графы F_1 и F_2. Граф F_1 либо несепарабелен, либо его можно разделить на графы F_3 и F_4. То же верно и для F_2. Повторяя этот процесс разделения графа, мы в конце концов получим подграфы $G_1, \ldots, G_m$, которые будут несепарабельными, причем, проводя разделения другим способом, мы все равно получим те же самые графы ([2], теорема 12). С другой стороны, два ребра a и b графа G назовем *циклически связанными*, если либо $a=b$, либо в G существует цикл, содержащий как a, так и b. Можно доказать, что отношение „a и b циклически связаны" симметрично, рефлексивно и транзитивно и что его выполнение равносильно условию „принадлежит той же несепарабельной компоненте, что и b" (см. [2], теорема 7). Тем самым несепарабельность оказывается определенной инвариантно относительно процесса разделения, что доказывает единственность несепарабельных компонент графа.

Теорема 3.1. *Если G — несепарабельный граф, $N(G)>1$, а R есть цикл в G, то в R существует* [2]) *цепь A, подвешенная в G, для которой $G-A$ есть несепарабельный граф и $N(G-A)=N(G)-1$.*

Доказательство заключается в построении графа G из последовательности несепарабельных подграфов $H_1 \subset H_2 \subset \ldots \subset G$. Поскольку $N(G)>1$, в G существует ребро a_1, не лежащее на R. Возьмем в качестве H_1 цикл, содержащий a_1 и некоторое ребро из R. Если $H_{m-1} \neq G$ уже выбрано, то H_m строим так: возьмем ребро a_m в $G-H_{m-1}$, не содержащееся в R (если такое имеется). Поскольку граф G несепарабелен, существует цикл D, содержащий a_m и некоторое ребро из H_{m-1}. Пусть теперь A_m — максимальная из цепей, лежащих на цикле D, содержащих a_m и не содержащих ребер из H_{m-1}. Остается положить $H_m = H_{m-1} + A_m$.

Каждый подграф H_m несепарабельный; мы докажем это по индукции. Цикл H_1, очевидно, является несепарабельным графом. Если H_{m-1} несепарабельный, то ребра из A_m циклически связаны с остальными ребрами H_m, поэтому и H_m — несепарабельный граф.

Покажем теперь, что из $R \subset H_m$ следует $H_m = G$. По построению H_1 имеем $R \not\subset H_1$. Возьмем наименьшее $m>1$, для ко-

[1]) В этом случае граф G называется несвязным. — *Прим. перев.*

[2]) Это утверждение включает частный случай, доказанный в работе [2], теорема 18, в котором не фигурировало условие $A \subset R$. Теорему 3.1 можно также вывести по индукции из теоремы Уитни.

торого $R \subset H_m$. В этом случае $R \not\subset H_{m-1}$ и в R существует ребро b, не принадлежащее H_{m-1}. Обозначим через E максимальную из цепей, лежащих на R, содержащих b и не содержащих ребер из H_{m-1}. Поскольку $R \subset H_m$, то и $E \subset H_m = H_{m-1} + A_m$. По построению в E нет ребер из H_{m-1}, поэтому $E \subset A_m$. Но A_m выбиралось так, что единственными вершинами A_m, общими с H_{m-1}, были концы этой цепи. Тем же свойством обладает и E, следовательно, E — подцепь цепи A_m с теми же концами, что и у A_m. Отсюда $E = A_m$. Если бы $G - H_{m-1}$ не содержалось в R, то A_m имело бы ребро a_m, не лежащее в R, откуда $A_m \neq E \subset R$. Полученное противоречие доказывает, что $G - H_{m-1} \subset R$. Но H_m содержит по построению H_{m-1} и по предположению содержит R, т. е. $H_m = G$.

Каждое A_m представляет собой цепь, подвешенную в H_m, поэтому $N(H_m) = m$. Процесс построения заканчивается при $H_n = G$, так что $n = N(G)$. Последняя из добавленных цепей $A = A_n$ подвешена в $H_n = G$ и содержится в R (как показано выше), следовательно, $N(G - A) = n - 1$, что и завершает доказательство теоремы.

Сепарабельные графы можно также строить стандартным способом.

Теорема 3.2. *Если граф G является сепарабельным, то в G существует такая несепарабельная компонента H, что H и $G - H$ имеют не более одной общей вершины.*

Доказательство. Возьмем любую несепарабельную компоненту H_1 в G. Либо H_1 обладает нужными свойствами, либо в $G - H_1$ существует несепарабельная компонента H_2, которая имеет с H_1 общую вершину p_1. Если H_2 также не обладает требуемыми качествами, существует компонента $H_3 \neq H_2$, у которой с H_2 имеется общая вершина $p_2 \neq p_1$. Если бы $H_3 = H_1$, то цепи в H_2 и H_1, соединяющие точки p_1 и p_2, образовывали бы цикл, не содержащийся целиком ни в одной из этих компонент, что невозможно. Следовательно, H_1, H_2 и H_3 различны, а поэтому, если H_3 нас не удовлетворит, можно взять новую компоненту H_4 и т. д., пока (в силу конечности графа) не получим компоненту H_m с ровно одной вершиной, общей с $G - H_m$.

4. ПРОЦЕСС ИНДУКЦИИ

Рассмотрим комбинаторный граф G, который удовлетворяет условиям теоремы I, т. е. предположим, что G содержит полную систему циклов

$$C_1, C_2, \ldots, C_n, \quad n = N(G), \tag{2}$$

в которую любое ребро входит не более двух раз. Такую систему назовем полной *τ-системой*. Циклы C_i независимы, а поэтому сумма

$$R = C_1 + C_2 + \ldots + C_n \pmod 2 \qquad (3)$$

не равна нулю. Будем называть R *ободом* графа G. Обод обладает следующим свойством.

Лемма 4.1. *Если G — несепарабельный граф, то его обод является циклом.*

Доказательство. Пусть, напротив, R не есть цикл. Тогда во всяком случае каждая вершина из R имеет четную степень, а следовательно, R в собственном смысле содержит некоторый цикл D. Можно написать разложение D по полной системе циклов (2), возможно изменив нумерацию,

$$D = C_1 + \ldots + C_m \pmod 2.$$

Поскольку $D \neq R$, имеем $n > m$. Поэтому ни один из графов

$$F_1 = C_1 + \ldots + C_m, \qquad F_2 = C_{m+1}, + \ldots + C_n$$

не пуст. Мы покажем, что G можно разделить на F_1 и F_2. Ребро b из $F_1 \cap F_2$ должно содержаться в одном из первых m циклов $C_1, \ldots, C_m$ и в одном из последних $n-m$ циклов $C_{m+1}, \ldots, C_n$. Поскольку (2) есть полная τ-система, это единственные ее циклы, в которые входит b. Таким образом, b входит в одно слагаемое в разложении D и в два слагаемых в разложении R. Мы получили, что b входит в D, но не в R, хотя, с другой стороны, $D \subset R$. Это противоречие доказывает, что $F_1 \cap F_2$ не содержит ни одного ребра.

Так как граф G несепарабелен, существует цикл E, содержащий как ребро из F_1, так и ребро из F_2. Разложим этот цикл по полной системе (2): $E = \sum' C_i + \sum'' C_j \pmod 2$ (суммирование в $\sum'$ выполняется по некоторым индексам от 1 до m, а в $\sum''$ — по некоторым из оставшихся индексов). Каждая из сумм $\sum'$, $\sum''$ содержит какие-то слагаемые, поскольку в E есть ребра как из F_1, так и из F_2. Таким образом, первая сумма $E' = \sum' C_i$ не равна E. Но E' содержится в F, так что ни одно ребро из E' не входит в циклы $C_{m+1}, \ldots, C_n$. Поэтому $E' \subset E$. Мы получили цикл E, в котором есть не совпадающий с ним подцикл, что невозможно. Следовательно, R есть цикл.

В несепарабельном плоском (топологическом) графе всегда есть цикл, который ограничивает конечную область и примыкает к границе бесконечной области по некоторой цепи.

Лемма 4.2. *Если G — несепарабельный граф, $N(G)>1$, то существует подвешенная цепь A, для которой*

(i) *$G-A$ несепарабален, $N(G-A)=N(G)-1$;*

(ii) *A содержится в R и ровно в одном из G_i, например, в C_n;*

(iii) *$C_1, \ldots, C_{n-1}$ образуют полную τ-систему для $G-A$;*

(iv) *концы p и q цепи A принадлежат R', где*

$$R' = C_1 + C_2 + \ldots + C_{n-1} \pmod 2; \tag{4}$$

(v) *R' состоит из двух цепей $R-A$ и C_n-A, соединяющих p и q.*

Доказательство. Возьмем, как и в теореме 3.1, подвешенную цепь $A \subset R$. Тогда $G-A$ несепарабелен, как и утверждает (i). Поскольку граф G несепарабелен, каждое ребро из A, а поэтому и вся цепь A, содержится в некотором цикле. Любой цикл представляется суммой по mod 2 циклов C_i, так что A содержится в некотором C_i. Изменим нумерацию, чтобы получить $A \subset C_n$. Так как A содержится в R и не более чем в двух из циклов $C_1, \ldots, C_n$, ясно, что A не содержится ни в каком $C_i \neq C_n$, что и утверждается в (ii). Циклы $C_1, \ldots, C_{n-1}$, таким образом, входят в $G-A$, причем они остаются там независимыми, как и в G. В $G-A$ существует не более $n-1=N(G-A)$ независимых циклов, поэтому $C_1, \ldots, C_{n-1}$ образуют полную τ-систему, как и отмечалось в (iii).

Покажем теперь, что $A = C_n \cap R$. Предположим, что существует ребро b в $C_n \cap R$, не входящее в A. Тогда b содержится в $G-A$, а следовательно, и в некотором цикле D из $G-A$. Согласно доказанному утверждению (iii), цикл D можно представить в виде суммы (mod 2) некоторых из циклов $C_1, \ldots, C_{n-1}$. Но b содержится в C_n и в R и не содержится ни в каких других C_i, поэтому представление цикла D в виде суммы по mod 2 циклов, не содержащих C_n и R, невозможно. Полученное противоречие показывает, что всякое ребро из $C_n \cap R$ входит и в A, а поэтому $C_n \cap R = A$.

Из (4) имеем $R' = R + C_n \pmod 2$, так что R' состоит из ребер, принадлежащих либо R, либо C_n, но не принадлежащих $R \cap C_n$. Поскольку $R \cap C_n = A$, получаем, что R' состоит из ребер $R-A$ и C_n-A. Так как R есть цикл, $R-A$ (а также C_n-A) есть цепь, соединяющая p и q, т. е. соединяющая концы цепи A. Для доказательства (iv) и (v) остается заметить, что $R-A$ лежит в R', следовательно, концы этой цепи также лежат в R'.

5. ДОКАЗАТЕЛЬСТВО ДОСТАТОЧНОСТИ

Теорема 5.1. *Несепарабельный граф G, в котором имеется полная τ-система циклов* (2), *можно вложить на плоскость так, что каждый из циклов $C_1, \ldots, C_n$ будет ограничивать одну из конечных областей, на которые G делит плоскость, а R будет ограничивать внешнюю (бесконечную) область.*

Доказательство будем вести индукцией по $N(G)$. Чтобы не привлекать не относящиеся к делу сведения из топологии, мы покажем, что G можно вложить на плоскость так, чтобы *каждому ребру из G на плоскости соответствовала бы ломаная линия, а циклу R — равносторонний треугольник.* Сначала рассмотрим случай $N(G)=1$. В этом случае G представляет собой цикл, так как G — несепарабельный граф (см. [2], теорема 10); следовательно, G можно вложить на плоскость описанным образом. Пусть теперь $N(G)>1$. Тогда, как утверждает лемма 4.2, в графе G имеется такая подвешенная цепь A, что $G-A$ несепарабелен, содержит полную τ-систему циклов и $N(G-A)<N(G)$.

По предположению индукции $G-A$ можно вложить на замкнутую плоскую треугольную область, границей которой является R'. Согласно лемме 4.2 (iv), концы p и q цепи A лежат на R', а, согласно (v), цепи C_n-A на плоскости соответствует дуга, соединяющая точки p и q. Поэтому вне треугольной области можно так провести ломаную A^*, соединяющую точки p и q, что A^* и C_n-A вместе будут образовывать границу новой конечной области. Тогда из (v) следует, что граница новой внешней (бесконечной) области есть $R=A^*+(R-A)$. Доказательство будет завершено, если мы покажем, что R можно сделать равносторонним треугольником.

Для этого предположим, что $G-A$ вложен на замкнутую треугольную область с вершинами r, s, t. Сначала рассмотрим случай, когда C_n-A содержит сторону треугольника с вершинами r и t. На плоскости цепь A будет представляться ломаной $puvq$ (рис. 1). Нам нужно сделать обод $spuvqs$ треугольником. Рассмотрим полуплоскость, содержащую s, край которой проходит через точки p и q, и подвергнем ее преобразованию сдвига. *Сдвигом* полуплоскости мы называем такое преобразование, при котором точка P, лежащая на расстоянии d от края полуплоскости, сдвигается параллельно этому краю на расстояние kd (k не зависит от P). В данном случае мы применим такой сдвиг, при котором точка s перейдет в s' так, что up и ps' будут лежать на одной прямой. Теперь применим к полуплоскости, содержащей точку v, край которой проходит через uq, такой сдвиг, чтобы $s'q$ и qv оказались бы на одной прямой

$s'qv'$. Новым ободом графа будет теперь треугольник $us'v'$. Его можно сделать равнобедренным при помощи подходящего сдвига полуплоскости с краем $v's'$, а затем и равносторонним, сжимая его по высоте. Преобразование сдвига переводит прямые линии в ломаные, так что мы получили вложение графа G с требуемыми свойствами. Другие случаи, когда $C_n - A$ содержит остальные вершины r, s или t, можно рассмотреть аналогично.

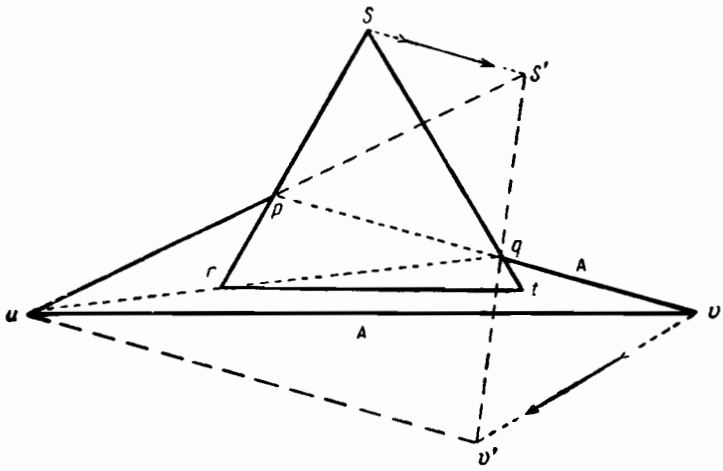


Рис. 1.

Теорема 5.2. *Граф G с полной τ-системой циклов можно вложить на плоскость так, чтобы каждому ребру соответствовала ломаная линия.*

Доказательство. Если G — сепарабельный граф, его можно разложить на несепарабельные компоненты $H_1, H_2, \ldots, H_m$. Каждая из этих компонент, которая не является ребром, содержит некоторые циклы из системы (2), например, $C_1, \ldots, C_k$ содержатся в H_i и образуют там полную систему, так как, во-первых, они независимы, во-вторых, любой цикл D в H_i представляется в виде $D = \sum' C_j + \sum'' C_j \pmod 2$, где первое суммирование производится по некоторым из индексов $j \leqslant k$, а второе — по некоторым $j > k$. Тогда $D - \sum' C_j$ состоит из ребер, содержащихся в H_i, и равно $\sum'' C_j$, которое не содержит ребра, входящие в H_i, и поэтому равно нулю.

Мы получили, что $D = \sum' C_j \pmod 2$, а следовательно, $C_1, \ldots, C_k$ образуют в H_i полную τ-систему. По теореме 5.1 заключаем, что H_i есть плоский граф. При доказательстве нашей теоремы достаточно теперь рассматривать связные графы.

Доказательство будем проводить индукцией по числу компонент. Выберем любую компоненту H_m. Согласно теореме 3.2, H_m и $G-H_m=F$ имеют одну общую вершину p. Предположение индукции будет заключаться в том, что граф F уже вложен на плоскость. Если $N(H_m)=0$, то H_m состоит из единственного ребра, которое легко добавляется к F на плоскости. Если же $N(H_m)>0$, то p принадлежит по крайней мере одному циклу C_l в компоненте H_m. Если p при этом не попадает на обод графа, мы заменяем в выбранной полной системе цикл C_l ободом графа H_m. Очевидно, что получится снова полная τ-система, но p уже будет лежать на новом ободе. Используя теорему 5.1, вложим H_m на плоскость так, чтобы получился топологический граф H_m^*, а вершина p перешла в точку p_1^* на внешней границе. Но вершине p соответствует также точка p_2^* при вложении на плоскость графа $G-H_m$. Осталось устроить такие вложения, чтобы точки p_1^* и p_2^* совпали.

Для этого разделим плоскость прямой, проходящей через p_1^*, и применим к каждой полуплоскости такие сдвиги, чтобы в точке p_1^* угол α между отрезками границы H_m^* стал меньше угла β между отрезками в точке p_2^*. Теперь сожмем H_m^* до таких размеров, чтобы его можно было поместить внутри угла β; поместим его туда и объединим точки p_1^* и p_2^*, получив, таким образом, требуемую разделяющую вершину p^* для $G=(G-H_m)+H_m$. Все использованные отображения переводят ломаные линии в ломаные [1]).

Мы установили достаточность критерия теоремы I. Остается проверить необходимость. Если каждая несепарабельная компонента графа G имеет полную τ-систему, то объединение всех таких систем дает полную τ-систему для G, поэтому достаточно рассмотреть несепарабельный случай.

Теорема 5.3. *Если G — несепарабельный плоский граф, $N(G)>0$, то каждая из конечных областей, на которые G делит плоскость, имеет своей границей цикл из G; множество всех таких циклов образует в G полную τ-систему, а их сумма по* mod 2 *является границей внешней (бесконечной) области.*

Доказательство. При $N(G)=1$ это есть теорема Жордана о плоских кривых. В общем случае применяем индукцию по $N(G)$ и теорему 3.1, а также тот факт, что разрез внутренней (внешней) области жордановой кривой делит эту

[1]) Другое, более „топологическое“ доказательство см. в работе [2], теорема 27.

область на две области с соответствующими границами. Последнее нужно для доказательства того, что каждая добавляемая дуга лежит на границе не более чем двух компонент.

6. КОМБИНАТОРНОЕ ПОСТРОЕНИЕ ДВОЙСТВЕННОГО ГРАФА

Установленный нами критерий, конечно, должен быть эквивалентен критерию Уитни, заключающемуся в требовании существования комбинаторного двойника. Мы дадим вывод этой эквивалентности при помощи комбинаторных рассуждений, получив, таким образом, новое доказательство критериев Уитни и Куратовского для плоских графов.

Граф G' называется *двойником* графа G, если существует взаимно однозначное соответствие между ребрами этих графов, при котором для любого подграфа H в G и его образа H' в G' будет выполняться равенство $R(G'-H')=R(G')-N(H)$ (разд. 1). Множество S ребер из G называется *сечением*, если у $G-S$ имеется либо большее, чем у G, число компонент связности, либо меньшее, чем у G, число вершин[1]) и если это не выполняется ни для какого собственного подмножества в S (см. [3], стр. 76). Граф обладает двойником тогда и только тогда, когда двойником обладает каждая его несепарабельная компонента ([2], теорема 23 и 25), с другой стороны, мы показали, что аналогичная ситуация возникает и при рассмотрении полных т-систем циклов. Поэтому мы ограничимся рассмотрением несепарабельных графов.

Теорема II. *Несепарабельный граф G обладает двойником тогда и только тогда, когда в G имеется полная т-система циклов.*

Тривиальный случай, когда G состоит из одного ребра, опустим. Пусть теперь G имеет полную т-систему (2), $n>0$. Обозначим обод R в (3) через C_{n+1}. Поскольку каждое ребро принадлежит не более двум циклам полной системы, ясно, что оно принадлежит в точности двум циклам системы $C_1, C_2, \ldots, C_n, C_{n+1}$. Построим новый граф G' с вершинами $p_1, \ldots, p_{n+1}$, соответствующими циклам $C_1, \ldots, C_{n+1}$, и с ребрами $b'_1, \ldots, b'_l$, которые взаимно однозначно сопоставлены ребрам G так, что ребро b'_l имеет концы p_i и p_k, если соответствующее ребро b_l принадлежит циклам C_i и C_k. Докажем, что этот „граф циклов“ G' является двойником графа G.

---

[1]) Практически второй случай означает, что граф $G-S$ несвязен и в качестве одной из компонент имеет изолированную вершину. Мы исключили из рассмотрения такую ситуацию.

Основную роль будет играть комбинаторный аналог теоремы Жордана о плоских кривых.

Лемма 6.1. *Если D есть цикл в G, то соответствующий подграф D' в G' является сечением.*

Доказательство. Цикл D можно разложить по полной системе циклов C_i. Перенумеруем эти циклы так, чтобы в разложение для D входили только первые k циклов. Используя определение (3), получаем

$$D = C_1 + C_2 + \ldots + C_k = C_{k+1} + \ldots + C_{n+1} \pmod 2. \qquad (5)$$

Этим разложениям соответствует разбнение множества вершин графа G' на два множества $p_1, \ldots, p_k$ и $p_{k+1}, \ldots, p_{n+1}$. Согласно (5), D' состоит из тех ребер, один конец которых находится в первом, а другой — во втором из упомянутых множеств. Таким образом, в $G' - D'$ два этих множества не связаны между собой. Однако в каждом из множеств вершины связаны друг с другом. Для доказательства этого предположим противное. Пусть, например, $p_1, \ldots, p_j$ не связаны в $G' - D'$ никакими ребрами с вершинами $p_{j+1}, \ldots, p_k$. Тогда $E = C_1 + \ldots + C_j \pmod 2$ не содержит ребер, входящих в $C_{j+1}, \ldots$ $\ldots, C_k$, но в то же время содержит только ребра из D. Получилось, что в цикле D есть собственный подцикл E, что невозможно. Совершенно так же рассматривается множество $p_{k+1}, \ldots, p_{n+1}$. Добавление любого ребра из D' восстанавливает связность. Следовательно, D' является сечением, что и утверждалось в лемме.

Теперь, чтобы доказать[1]), что G' есть двойник G, достаточно убедиться ([3], теорема 2), что циклу в G соответствует сечение в G' и обратно. В одну сторону мы доказали это в предыдущей лемме. Покажем обратное. Пусть S' — сечение в G'. При доказательстве леммы мы убедились, что сечение D' делит связный граф G' на две компоненты связности. Ясно, что и S' поэтому разделяет G' на две компоненты связности. Пусть первая компонента содержит (перенумерованные) вершины $p_1, \ldots, p_k$, вторая — вершины $p_{k+1}, \ldots, p_{n+1}$. Соответствующие циклы в G образуют подграф

$$D = C_1 + \ldots + C_k = C_{k+1} + \ldots + C_{n+1} \pmod 2.$$

Каждое ребро D принадлежит одному из циклов $C_1, \ldots, C_n$ и одному из остальных циклов, и оно соответствует, таким образом, ребру, соединяющему в графе G' первое множество

[1]) Мы будем проводить такие же рассуждения, как и в [2], теорема 29. Там доказывается, что плоский граф имеет двойник. Мы заменяем только теорему Жордана ее комбинаторным аналогом.

вершин со вторым. Такое ребро должно принадлежать сечению в графе G', поэтому $D \subset S$. Но подграф D непременно содержит цикл D_1 (см. лемму 4.1). Согласно лемме 6.1, граф D_1' есть сечение в G', а по определению сечения имеем $D_1' = S'$, откуда получаем $S = D_1$. Мы доказали, что каждому сечению в графе G' соответствует цикл в графе G. Отсюда следует, что G' — двойник G.

Пусть теперь G имеет двойник G'. Мы будем искать в G полную τ-систему. Заметим сначала, что если граф G несепарабельный, то и G' несепарабельный [1]).

Согласно определению двойника, $n = N(G) = V(G) - 1$ (см. (1)), поэтому G' имеет $(n+1)$ вершину $p_1, \ldots, p_{n+1}$. Множество D_i' всех ребер, инцидентных данной вершине p_i, является сечением, так как, удалив его, мы получим изолированную вершину p_i. Соответствующий подграф D_i является циклом в G по цитированной выше теореме Уитни. Каждое ребро циклов $D_1, \ldots, D_{n+1}$ содержится ровно в двух циклах, так что их сумма (mod 2) равна нулю. Никакими другими соотношениями по mod 2 эти циклы не связаны. Действительно, если бы

$$D_1 + D_2 + \ldots + D_m = 0 \pmod 2, \quad m < n+1,$$

то любое ребро одного из этих циклов принадлежало бы еще ровно одному из них. Это означает, что любое ребро в G', у которого один конец находится среди вершин $p_1, \ldots, p_m$, имеет среди этих вершин и второй конец; следовательно, вершины $p_1, \ldots, p_m$ не связаны с остальными вершинами. Это противоречит несепарабельности G'. Поэтому $D_1, \ldots, D_n$ независимы (mod 2) и их ровно $n = N(G)$. Мы получили полную τ-систему циклов. Теорема II доказана.

ЛИТЕРАТУРА

1. Kuratowski C., *Fund. Math.*, **15** (1930), 271—283.
2. Whitney H., Non separable and planar graphs, *Trans. Amer. Math. Soc.*, **34** (1932), 339—362.
3. Whitney H., Planar graphs, *Fund. Math.*, **21** (1933), 74—83.
4. Kuratowski C., Whyburn, Sur les éléments cycliques et leurs applications. *Fund. Math.*, **16** (1930), 305—331.

[1]) См. [2], теорема 26; в данном случае существенно, что графы не имеют изолированных вершин.

Минимальные вложения и род графа[1])

Дж. Янгс

1. ВВЕДЕНИЕ

1.1. Достаточно очевидно, что связный граф можно вложить в ориентируемое 2-многообразие. Однако вовсе не очевидно, существует ли для данного вложения вложение в ориентируемое 2-многообразие меньшего рода. Если такого вложения не существует, данное вложение называется *минимальным* вложением; род поверхности при таком вложении называется *родом графа*.

1.2. Эта статья посвящена проблеме описания минимальных вложений и вычислению рода графа.

1.3. Важную роль будет играть вспомогательное понятие 2-*клеточного вложения*. Это такое вложение, при котором каждая компонента связности дополнения образа графа на поверхности представляет собой открытую 2-клетку. Если для данного графа число компонент в дополнении графа на поверхности при некотором вложении не меньше, чем при любом другом, такое вложение называется *максимальным*.

1.4. Теорема характеризации утверждает, что вложение минимально тогда и только тогда, когда оно максимальное 2-клеточное.

1.5. При этом значительную роль играет операция перекройки[2]) (имеющая, по-видимому, и самостоятельную ценность). Эта операция позволяет при любом вложении графа так „перекроить" компоненты его дополнения, что в результате получается 2-клеточное вложение (см. п. 3.5).

1.6. Теорема характеризации имеет интересное ответвление. Можно ли считать, что 1-мерный остов триангуляции ориентируемого 2-многообразия вложен в это многообразие при помощи минимального вложения? Для рассматриваемой тематики это традиционный вопрос. Теорема характеризации и „вычислительный процесс" решают его утвердительно, как и ожидалось (см. п. 5.2—5.5).

---

[1]) Youngs J. W. T., Minimal imbeddings and the genus of a graph, *Journal of Mathematics and Mechanics*, **12**, № 2 (1963), 303—315.

[2]) В подлиннике «capping operation». — *Прим. перев.*

1.7. Приведенные результаты выделены потому, что они касаются классических вопросов. В данной статье аналогичные утверждения доказаны в более общем виде, когда 2-многообразие не обязательно ориентируемо.

1.8. Статью завершает алгоритм вычисления рода графа, использующий теорему характеризации и изящный результат Дж. Эдмондса [1, 2]. К сожалению, этот алгоритм очень трудоемкий, а методы сокращения перебора неизвестны. Поэтому он мало пригоден для практических вычислений.

2. ГРАФЫ

2.1. *Граф* — это упорядоченная пара (G^0, G^1), в которой G^0 — непустое конечное множество объектов, а G^1 — множество неупорядоченных пар различных элементов из G^0. Элементы множества G^0 называются *вершинами*, а элементы множества G^1 — *ребрами* графа G.

Число элементов в G^i обозначается $\| G^i \|$, $i = 0, 1$.

2.2. *Дугой* в произвольном топологическом пространстве X называют гомеоморф замкнутого единичного отрезка. После удаления из дуги образов концов этого отрезка получается *открытая дуга*.

Пусть G — граф, у которого

$$G^0\colon\ v_1, \ldots, v_n,$$
$$G^1\colon\ a_1, \ldots, a_m.$$

Вложением G в X называется подпространство $G(X)$ пространства X, такое, что

$$G(X) = \bigcup v_i(X) \cup \bigcup a_k(X),$$

где:

1) $v_1(X), \ldots, v_n(X)$ есть n различных точек в X;

2) $a_1(X), \ldots, a_m(X)$ есть m попарно непересекающихся открытых дуг в X;

3) $a_i(X) \cap v_j(X) = \varnothing$, $i = 1, \ldots, n$, $j = 1, \ldots, m$;

4) если $a_k = (v_{k_1}, v_{k_2})$, то открытая дуга $a_k(X)$ имеет $v_{k_1}(X)$ и $v_{k_2}(X)$ своими концами; $k = 1, \ldots, m$.

Мы нередко пользуемся сокращениями: пишем v_i вместо $v_i(X)$ и a_k вместо $a_k(X)$ ($i = 1, \ldots, n$, $k = 1, \ldots, m$) и говорим „вложение $G(X)$“ вместо „вложение $G(X)$ графа G в X“.

2.3. Очевидно, что два вложения $G(X)$ и $G(Y)$ одного и того же графа G гомеоморфны.

Граф G *связен*, если для некоторого вложения подпространство $G(X)$ связно. Предыдущее замечание показывает, что равносильным будет определение, в котором связность подпро-

странства $G(X)$ требуется при любом вложении. *В дальнейшем рассматриваются только связные графы.*

2.4. Очевидно, что для любого графа G существует вложение G в 3-мерное евклидово пространство E_3. На самом деле можно добиться даже того, чтобы любая открытая дуга была бы открытым интервалом на некоторой прямой. Очевидно также, что для данного графа G можно найти такое (зависящее от G) ориентируемое 2-многообразие M, что существует вложение $G(M)$. Например, можно взять $\|G^0\|$ различных точек 2-сферы и поместить каждую дугу a_k на отдельную ручку, подходящим образом прикрепленную к 2-сфере. Род такого многообразия M будет слишком велик, и задача нахождения многообразия M с минимальным числом ручек весьма естественна.

2.5. Уточним сделанные замечания. Обозначим $\chi(M)$ характеристику Эйлера 2-многообразия M. Напоминаем, что $\chi(M)$ является топологическим инвариантом M; для данного клеточного разбиения M, у которого число i-клеток равно α^i $(i = 0, 1, 2)$, имеем

$$\chi(M) = \alpha^0 - \alpha^1 + \alpha^2.$$

Кроме того, $\chi(M) \leqslant 2$. Если M ориентируемо, то $\chi(M)$ четно. *Род* $\gamma(M)$ 2-многообразия M определяется формулой

$$\gamma(M) = [2 - \chi(M)]/2.$$

Для каждого четного значения характеристики Эйлера, меньшего двух, имеется два 2-многообразия с этой характеристикой: одно ориентируемое, а другое нет. Любое ориентируемое 2-многообразие M можно представлять себе как 2-сферу с некоторым числом ручек, тогда род $\gamma(M)$ попросту равен числу этих ручек.

2.6. Буквами M и N мы будем обозначать 2-многообразия; $M_\#$ и $N_\#$ означают *ориентируемые* 2-многообразия; $G(M)$ и $G(M_\#)$ означают вложение графа G в не обязательно ориентируемое и ориентируемое 2-многообразие соответственно. Используя эти обозначения, мы назовем вложение $G(M)$ *простейшим*, если $\chi(M) \geqslant \chi(N)$ для любого вложения $G(N)$. Если $G(M)$ — простейшее вложение, то *показатель вложения* $\mu(G)$ графа G по определению равен $\chi(M)$. Употребление термина „простейшее" оправдывается тем, что с возрастанием характеристики Эйлера 2-многообразие усложняется. Если рассматриваются только ориентируемые 2-многообразия, то вместо *простейшее вложение* говорят *минимальное вложение*. В силу 2.5 вложение $G(M_\#)$ *минимально*, если для любого вложения $G(N_\#)$ имеем $\gamma(M_\#) \leqslant \gamma(N_\#)$. *Род графа* $\gamma(G)$ в этом случае полагают по определению равным $\gamma(M_\#)$.

Если $G(M)$ — простейшее вложение, а $G(M_\#)$ — минимальное, то $\chi(M) \geqslant \chi(M_\#)$. Следовательно,

$$\gamma(G) \geqslant [2 - \mu(G)]/2.$$

Для примера возьмем в качестве графа G любой из двух знаменитых графов Куратовского. Его простейшим вложением будет вложение в проективную плоскость, а минимальным — в тор. Соответствующие характеристики Эйлера суть 1 и 0, так что приведенное неравенство выполняется.

2.7. Оказывается важным рассматривать для данного вложения компоненты связности открытого множества $[M - G(M)]$. Каждая такая компонента является открытым множеством, поскольку открытым множеством является M, и ясно, что этих компонент конечное число.

Обозначим $\| G(M) \|$ *число* компонент в $[M - G(M)]$. Если каждая компонента из $[M - G(M)]$ является открытой 2-клеткой, то $G(M)$ называется 2-*клеточным вложением*.

2.8. Вложение $G(M)$ *максимально*, если $\| G(M) \| \geqslant \| G(N) \|$ для любого вложения $G(N)$; оно — *максимальное ориентируемое вложение*, если $M_\#$ ориентируемо и $\| G(M_\#) \| \geqslant \| G(N_\#) \|$ для любого $N_\#$ (вспомните 2.6).

3. ОПЕРАЦИЯ ПЕРЕКРОЙКИ

3.1. Результаты этого раздела не имеют такой простой и естественной формы, как результаты остальных разделов. Вызвано это тем, что мы вынуждены принять меры предосторожности, необходимость которых некоторые авторы проглядели. Грубо говоря, мы показываем, как из произвольного вложения получнть 2-клеточное. Метод, который применяется при этом, имеет большое значение и будет применяться в других работах. Добиваясь предельной краткости, мы собрали все существенные результаты в одном месте, даже если некоторые из них не находят здесь применения. Те, которые понадобятся в дальнейшем (в других работах), заключены в скобки.

Чтобы ввести основные идеи, рассмотрим некоторое вложение $G(M)$ и некоторую компоненту S из $[M - G(M)]$. Если S не является 2-клеткой, можно пытаться заменить S некоторым набором 2-клеток, изменив тем самым 2-многообразие M. Именно здесь и необходима осторожность. Рингель [3, стр. 61], например, рассматривает $(M - S)$ и утверждает, что это 2-многообразие с краем, граничные циклы которого $J_1, \ldots, J_s$ он заклеивает соответствующими 2-клетками $C_1, \ldots, C_s$. Множество $G(M)$ является вложением G в 2-многообразие $N \equiv$

$\equiv (M - S) \cup \bigcup C_i$, при этом мешавшая ранее компонента S из $[M - G(M)]$, так сказать, заменяется 2-клетками $C_1, \ldots, C_s$ из $[N - G(M)]$.

К сожалению, $(M - S)$ далеко не всегда есть 2-многообразие с краем. Оно может даже в точности совпадать с множеством $G(M)$. Возьмем, например, в качестве графа G треугольник и уложим его на тор M так, чтобы $G(M)$ нельзя было стянуть в точку; $[M - G(M)]$ здесь имеет ровно одну компоненту S, т. е. само $[M - G(M)]$, так что $(M - S) = G(M)$, а $G(M)$, конечно, не является 2-многообразием с краем.

Исправление ошибки требует тонких дополнительных рассмотрений, в результате которых появляется лемма Робертса и Стинрода [4]. Они показывают следующее.

Лемма (Робертса и Стинрода). *Если M есть (не обязательно ориентируемое) 2-многообразие, K есть непустой континуум, являющийся собственным подмножеством M, а S — компонента из $(M - K)$, то S содержит комплекс T, который является 2-многообразием с краем. Более того, если $J_1, \ldots, J_2$ есть жордановы кривые, образующие границу T, то:*

1) *компоненты $(S - T)$ суть открытые цилиндры $L_1, \ldots, L_s$;*

2) *$\mathrm{Fr}(L_i)$, т. е. граница L_i, имеет две компоненты, J_i и подмножество из K, $i = 1, \ldots, s$.*

Пусть теперь $G(M)$ — вложение графа G в 2-многообразие M, а S — компонента из $[M - G(M)]$. Множество $G(M)$ можно рассматривать как непустой континуум в M, причем с очень хорошими дополнительными свойствами. Поэтому можно не только применять лемму Робертса и Стинрода, но даже получить в этом частном случае ее доказательство, основанное на простых комбинаторных рассмотрениях. Для доказательства возьмем такую триангуляцию τ многообразия M, чтобы $G(M)$ было подкомплексом ее 1-мерного остова. Возьмем также τ_2 — барицентрическое подразделение τ. Рассмотрим теперь R — открытую звезду границы $\mathrm{Fr}(S)$ относительно τ_2, и пусть $J_1, \ldots, J_s$ — компоненты $\mathrm{Fr}(R) \cap S$.

Легко видеть, что у графа J_i каждая вершина имеет степень 2, а значит J_i есть элементарный цикл $(i = 1, \ldots, s)$. Если Q — открытая звезда подкомплекса $\bigcup J_i$ относительно τ_2, то компонентами множества $Q \cap R$ являются открытые цилиндры $L_1, \ldots, L_s$. Граница $\mathrm{Fr}(L_i)$ имеет две компоненты: J_i и цикл из $G(M)$. В качестве 2-многообразия T, фигурировавшего в лемме, достаточно взять $\left(S - \bigcup L_i\right)$.

Совершенно очевидно, что число s в этой лемме не меньше числа компонент границы $\mathrm{Fr}(S)$.

3.2. Пусть T^0 означает внутренность T. Ясно, что в отличие от $(M-S)$ пространство $(M-T^0)$ является 2-многообразием с краем (вспомните замечания в 3.1). Рассмотрим 2-многообразие M_S, полученное из $(M-T^0)$ заклейкой каждого граничного цикла J_i 2-клеткой C_i, $i=1, \ldots, s$, другими словами, $M_S = (M-T^0) \cup \bigcup C_i$. Поскольку $G(M) \subset (M-T^0)$, получаем вложение $G(M_S)$ графа G в M_S. Процесс получения M_S из M обозначим

$$M \to M_S$$

и назовем *операцией перекройки* компоненты S. Заметим, что $\| G(M_S) \| \geqslant \| G(M) \|$, ибо $s \geqslant 1$ (см. 2.7).

3.3. Если T_S есть 2-многообразие $T \cup \bigcup C_i$, то элементарные вычисления показывают, что

$$\chi(M) = \chi(M_S) + \chi(T_S) - 2s.$$

Напомним, что для любого 2-многообразия N имеем

$$\chi(N) \leqslant 2,$$

причем, равенство имеет место в том и только в том случае, когда N есть 2-сфера.

Поскольку $s \geqslant 1$ и

$$\chi(M_S) = \chi(M) + [2s - \chi(T_S)],$$

получаем

$$\chi(M_S) \geqslant \chi(M).$$

Равенство здесь выполняется тогда и только тогда, когда $s=1$, а T_S есть 2-сфера, другими словами, когда T^0 (а следовательно, и S) есть *открытая 2-клетка*. В этом случае *в качестве* T *можно взять 2-клетку, которой заклеен единственный граничный цикл в* $(M-T^0)$, причем только в этом случае $M_s = M$.

(Соответствующие замечания можио сделать и относительно T_S; действительно,

$$\chi(T_S) \geqslant \chi(M).$$

Равенство выполнено тогда и только тогда, когда $(M-T)$ есть открытая 2-клетка.)

(Докажем для использования в дальнейшем такой факт. Если $G(M)$ есть жорданова кривая, а среди компонент $[M-G(M)]$ нет открытых 2-клеток, то $\chi(T_S) > \chi(M)$. Предположим, что это не так, тогда $\chi(T_S) = \chi(M)$, а, следовательно, $(M-T)$ есть открытая 2-клетка. Имеем $G(M) \subset (M-T)$, следовательно, по теореме Жордана о плоских кривых в $[(M-T) - G(M)]$ существует такая компонента R, которая является

открытой 2-клеткой. Но R также есть компонента из $[M-G(M)]$, где по предположению открытых 2-клеток нет. Мы пришли к противоречию.)

3.4. В случае, когда M ориентируемо, ориентируемы также M_S и T_S. Далее,

$$\gamma(M)=\gamma(M_S)+\gamma(T_S)+(s-1).$$

Это равенство получено из 3.3 с помощью соотношения $2\gamma = 2-\chi$. Аналогично

$$\gamma(M_S)\leqslant\gamma(M),$$

причем равенство выполняется в том и только в том случае, когда S — открытая 2-клетка.

(Если $G(M)$ — жорданова кривая, а среди компонент $[M-G(M)]$ нет открытых 2-клеток, то $\gamma(T_S)<\gamma(M)$.)

(В заключение этого пункта заметим, что из леммы п. 3.1 следует существование последовательности $\{T_n\}$ 2-многообразий T_n, каждое из которых обладает теми же свойствами, что и T, причем $T_n\subset T^0_{n+1}$ и $\bigcup T_n = S$. Поэтому для любого замкнутого подмножества F в S можно выбрать такое 2-многообразие T, что $F\subset T$, а, следовательно, $F\subset T_S$.)

3.5. Пусть $S_1, \ldots, S_p$ — все компоненты $[M-G(M)]$. Рассмотрим операцию перекройки $M\to M_{S_1}\equiv M^1$. Ясно, что $G(M)\subset M^1$, а при перекройке изменениям подвергалось только S_1. Поэтому компонентами $[M^1-G(M)]$ будут $S_2, \ldots, S_p$ и несколько открытых 2-клеток, причем $\|G(M^1)\|\geqslant\|G(M)\|$ (см. 3.2).

Теперь рассмотрим $M^1\to M^1_{S_2}\equiv M^2$. Здесь снова $G(M)\subset M^2$ и компонентами $[M^2-G(M)]$ являются $S_3, \ldots, S_p$ и некоторые открытые 2-клетки, причем опять $\|G(M^2)\|\geqslant\|G(M^1)\|$. Повторяя эту операцию p раз, получаем в результате 2-многообразие M^*, $M^{p-1}\to M^{p-1}_{S_p}\equiv M^*$, с такими свойствами:

1) $G(M)\subset M^*$. Следовательно, имеется вложение $G(M^*)$ графа G в M^*; подмножество $G(M)$ в M — это то же самое, что $G(M^*)$ в M^* (см. 2.2).

2) Компоненты $[M^*-G(M^*)]$ суть открытые 2-клетки.

3) $S_i\subset M^*$ тогда и только тогда, когда S_i есть открытая 2-клетка (см. 3.3).

4) $\|G(M^*)\|\geqslant\|G(M)\|$.

Композиция операций перекройки $M\to M^1\ldots M^{p-1}\to M^*$ называется *операцией перекройки* многообразия M; обозначим ее

$$M\to M^*.$$

Очевидно, что для любых двух нумераций компонент $[M-G(M)]$ в результате операции перекройки многообразия M

получаются гомеоморфные 2-многообразия, т. е. эта операция не зависит от порядка компонент.

3.6. Полезно записать некоторые свойства M^*.

1) Если M ориентируемо, то ориентируемо и M^* (см. 3.4).

2) $G(M^*)$ есть 2-клеточное вложение [см. 1) и 2) в 3.5].

3) $M^* = M$ тогда и только тогда, когда $G(M)$ есть 2-клеточное вложение [см. 3.3 и 3) в 3.5].

4) Если $G(M)$ не есть 2-клеточное вложение, то $\chi(M^*) > \chi(M)$, поэтому, если M ориентируемо, то $\gamma(M^*) < \gamma(M)$ (см. 3.3 и 3.4).

5) $\| G(M^*) \| \geqslant \| G(M) \|$ [см. 4) в 3.5].

[6) Если F — замкнутое подмножество $G(M)$ и некоторая компонента из $(M - F)$ есть открытая 2-клетка C, то C является компонентой из $(M^* - F)$.]

Для доказательства 6) заметим, что $M = G(M) \cup \bigcup S_i$, $i = 1, \ldots, p$. Следовательно, $C = C \cap M = [G(M) \cap C] \cup \bigcup_1^p (S_i \cap C)$. Однако, поскольку C есть компонента $(M - F)$, а S_i есть связное подмножество $(M - F)$, то утверждение $S_i \cap C \neq \varnothing$ эквивалентно утверждению $S_i \subset C$. Следовательно, $C = [G(M) \cap C] \cup \bigcup S_i$, $S_i \subset C$. Но C — открытая 2-клетка, поэтому если $S_i \subset C$, то S_i — открытая 2-клетка, поскольку G связно. Таким образом, согласно 3) из 3.5, имеем $M^* \supset \bigcup S_i$, $S_i \subset C$. С другой стороны, $[G(M) \cap C] \subset G(M) \subset M^*$, согласно 1) из 3.5. Следовательно, $C \subset M^*$. Ясно, что $C \subset M^i$, $i = 1, \ldots, p$, где $M^p = M^*$. Используя определение операции перекройки $M \to M^1$, легко убедиться, что C — компонента из $(M - F)$ есть также компонента из $(M^1 - F)$. Отсюда сразу же следует по индукции, что C есть компонента $(M^* - F)$, что и завершает доказательство.

4. ТЕОРЕМА ХАРАКТЕРИЗАЦИИ

4.1. Теперь легко доказать обещанную во введении теорему характеризации.

Сначала заметим, что если $G(M)$ есть 2-*клеточное вложение*, то оно определяет клеточное разбиение M, для которого $\alpha^0 = \| G^0 \|$, $\alpha^1 = \| G^1 \|$, $\alpha^2 = \| G(M) \|$ (см. 2.1, 2.5 и 2.7). Отсюда

$$\chi(M) = \| G^0 \| - \| G^1 \| + \| G(M) \|.$$

Если $G(N)$ есть другое 2-*клеточное вложение*, то

$$\chi(M) - \chi(N) = \| G(M) \| - \| G(N) \|.$$

Теорема 4.2. *Вложение графа является простейшим тогда и только тогда, когда оно максимальное* 2-*клеточное.*

Доказательство. Пусть $G(M)$ — простейшее вложение, а $G(N)$ — максимальное 2-клеточное. Тогда $G(M)$ есть 2-клеточное вложение, согласно 4) из 3.6. Используя 4.1 и условия теоремы, получаем

$$0 \leqslant \chi(M) - \chi(N) = \| G(M) \| - \| G(N) \| \leqslant 0.$$

Следовательно,

$$\chi(M) = \chi(N), \quad \| G(M) \| = \| G(N) \|.$$

Отсюда следует, что $G(M)$ есть максимальное 2-клеточное вложение, а $G(N)$ — простейшее.

Следствие. *Если $G(M)$ — максимальное вложение, то в результате операции перекройки $M \to M^*$ получается простейшее вложение $G(M^*)$.*

Доказательство непосредственно следует из 3.6, 2), 5) и теоремы 4.2.

Теорема 4.3. *Вложение графа минимально тогда и только тогда, когда оно есть максимальное ориентируемое 2-клеточное вложение.*

Доказательство очевидным образом получается из доказательства предыдущей теоремы, надо только заметить, что в этом случае минимальное вложение $G(M_\#)$ является 2-клеточным, согласно 1) и 4) из 3.6.

Следствие. *Если $G(M_\#)$ — максимальное вложение, то операция перекройки $M_\# \to M_\#^*$ дает минимальное вложение.*

Доказательство основано на использовании 1), 2), 5) из 3.6 и теоремы 4.3.

5. ПРОБЛЕМА ТРИАНГУЛЯЦИИ

5.1. Пусть T есть 1-мерный остов триангуляции 2-многообразия M, т. е. совокупность 0- и 1-мерных симплексов этой триангуляции. Очевидно, что существует граф G и вложение $G(M)$, такие, что $T = G(M)$. Возникает два вопроса:

1) Является ли $G(M)$ простейшим вложением?

2) Если M ориентируемо, является ли $G(M)$ минимальным вложением?

Второй вопрос стал широко известен благодаря Ауслендеру, но во всяком случае он издавна бытовал в математическом фольклоре, связанном с теорией графов. Ответы на оба вопроса единодушно считались утвердительными, оставалось только доказать эти естественные предположения. Результаты разд. 4

и приводимые ниже вычисления, основным достижением которых является формула 5.3, (2), позволяют провести весьма прозрачное доказательство.

5.2. Пусть $G(M)$ есть 2-*клеточное вложение* графа G в 2-многообразие M. Пусть $\mathfrak{S}$ — совокупность компонент $[M - G(M)]$.

Если $S \in \mathfrak{S}$, $a \in G^1$, то из $a(M) \cap \mathrm{Fr}(S) \neq \emptyset$ следует $a(M) \subset \mathrm{Fr}(S)$. Поскольку M есть 2-многообразие, существует *не более одного* $S_0 \in \mathfrak{S}$, $S_0 \neq S$, для которого $a(M) \subset \mathrm{Fr}(S_0)$. Следовательно, для $S \in \mathfrak{S}$ и $a \in G^1$ выполняется одно из трех соотношений:

Случай 0. $a(M) \cap \mathrm{Fr}(S) = \emptyset$.

Случай 1. $a(M) \subset \mathrm{Fr}(S)$ и $a(M) \subset \mathrm{Fr}(S_0)$ для некоторого единственного $S_0 \in (\mathfrak{S} - S)$.

Случай 2. $a(M) \subset \mathrm{Fr}(S)$, но $a(M) \subset \mathrm{Fr}(S_0)$ не выполняется ни при каком $S_0 \in (\mathfrak{S} - S)$.

Для $a \in G^1$ и $S \in \mathfrak{S}$ определим $f_S(a) = i$, если имеет место случай i. Заметим, что при *фиксированном* $a \in G^1$ имеем

$$\sum_{S \in \mathfrak{S}} f_S(a) = 2. \tag{1}$$

5.3. Компонента $S \in \mathfrak{S}$ называется *k-угольником*, если

$$\sum_{a \in G^1} f_S(a) = k.$$

Пусть $\mathfrak{S}_k = \{S \mid S \in \mathfrak{S}$ и S есть k-угольник$\}$. Введем определения

$$N_k = \text{число элементов в } \mathfrak{S}_k,$$
$$p = \min(k),\ \mathfrak{S}_k \neq \emptyset,$$
$$q = \max(k),\ \mathfrak{S}_k \neq \emptyset.$$

Используя (1), получаем прямым вычислением

$$2 \| G^1 \| = \sum_{k=p}^{q} k N_k. \tag{2}$$

Но

$$p \sum_{k=p}^{q} N_k \leqslant \sum_{k=p}^{q} k N_k \leqslant q \sum_{k=p}^{q} N_k$$

и

$$\| G(M) \| = \sum_{k=p}^{q} N_k.$$

Поэтому из (2) следует

$$p \| G(M) \| \leqslant 2 \| G^1 \| \leqslant q \| G(M) \|, \tag{3}$$

причем

$$p \| G(M) \| = 2 \| G^1 \| \tag{4}$$

тогда и только тогда, когда $p = q$, т. е. когда $\mathfrak{S}$ состоит исключительно из p-угольников.

5.4. Предположим теперь, что G есть такой граф, для которого некоторое вложение $G(M)$ *является триангуляцией* многообразия M, т. е. каждая компонента из $[M - G(M)]$ есть треугольник. В силу (4) имеем

$$3\,\|G(M)\| = 2\,\|G^1\|. \tag{5}$$

Если $G(N)$ есть 2-клеточное вложение, то в силу (3) находим

$$p\,\|G(N)\| \leqslant 2\,\|G^1\|, \tag{6}$$

причем равенство имеет место тогда и только тогда, когда $p = q$ при вложении $G(N)$.

Мы так определили граф, что 1-угольники невозможны: концы ребра не должны совпадать. Более того, никакие две вершины графа не связаны двумя или более ребрами, так что и 2-угольники невозможны, за исключением тривиального случая, когда граф состоит из единственного ребра и двух вершин.

Поскольку $G(M)$ определяет некоторую триангуляцию M, граф G должен иметь по крайней мере 3 ребра. Следовательно, ни 1-, ни 2-угольников нет, и для вложения $G(N)$ имеем

$$p \geqslant 3. \tag{7}$$

Используя (5) — (7), получаем

$$\|G(N)\| \leqslant \|G(M)\|, \tag{8}$$

причем равенство имеет место тогда и только тогда, когда $p = q = 3$ для вложения $G(N)$.

5.5. Если предполагать, что в 5.4 имеется в виду *ориентируемое* многообразие M, а $G(N)$ — минимальное вложение, то по теореме характеризации 4.3 имеем

$$\|G(N)\| \geqslant \|G(M)\|.$$

Отсюда, согласно (8), получаем

$$\|G(N)\| = \|G(M)\|$$

и $p = q = 3$ для вложения $G(N)$.

Следовательно, $G(M)$ является также максимальным ориентируемым 2-клеточным вложением, а поэтому в силу 4.3 — минимальным вложением, причем $G(N)$ определяет триангуляцию многообразия N. Тем самым доказана следующая теорема.

Теорема. *Если $G(M)$ есть 1-мерный остов триангуляции ориентируемого 2-многообразия M, то $G(M)$ — минимальное вложение, причем если $G(N)$ есть вложение в ориентируемое*

2-*многообразие* N *того же рода, что и* M, *то* $G(N)$ *является* 1-*мерным остовом триангуляции* N.

5.6. Теперь не будем предполагать, что M ориентируемо. Пусть $G(M)$ — вложение, определяющее некоторую триангуляцию M, а $G(N)$ — простейшее вложение. В силу 4.2 $G(N)$ есть максимальное 2-клеточное вложение и, согласно 5.4 и 5.5, имеем

$$\| G(N) \| = \| G(M) \|,$$

причем $p = q = 3$ для вложения $G(N)$.

Аналогично 5.5 получаем теорему.

Теорема. *Если* $G(M)$ *есть* 1-*мерный остов триангуляции* 2-*многообразия* M, *то* $G(M)$ *есть простейшее вложение, причем если* $G(N)$ *есть вложение в* 2-*многообразие* N *с той же характеристикой Эйлера, что и у* M, *то* $G(N)$ *является* 1-*мерным остовом триангуляции* N.

Замечание. В случае, когда $\chi(M)$ нечетно, а $G(N)$ — простейшее вложение, M и N есть гомеоморфные неориентируемые 2-многообразия. Если $\chi(M)$ четно, можно утверждать лишь, что $\chi(M) = \chi(N)$, так как логически имеется возможность, когда одно из многообразий ориентируемо, а другое нет, т. е. они не гомеоморфны. Однако вложение представляется настолько жестко определенным, что кажется, будто такая возможность исключена. Вначале так считал и автор этих строк. Но это предположение не оправдалось. Пусть K_n означает полный граф с n вершинами (т. е. K_n^0: $v_1, \ldots, v_n$; K_n^1: (v_i, v_j), $i \neq j$, $i, j = 1, \ldots, n$). Граф K_{12} вкладывается как в ориентируемое, так и в неориентируемое 2-многообразие характеристики -10, причем в обоих случаях он определяет некоторую триангуляцию (см. [3], стр. 74, 78).

6. ОПРЕДЕЛЕНИЕ РОДА ГРАФА

6.1. Теорема характеризации не дает способа нахождения минимального вложения, т. е. не решает классической задачи определения рода графа. Однако, если использовать эту теорему вместе с теоремой Эдмондса (см. [1, 2], а также [3], где на стр. 68 — 70 имеется подобный результат), можно построить соответствующий алгоритм.

6.2. Обозначим вершины данного графа $G = (G^0, G^1)$ не $v_1, \ldots, v_n$, а $1, \ldots, n$. *Во избежание тривиальных случаев предположим, что вершин степени* 1 *нет*. Определим $V(i)$ $(i = 1, \ldots, n)$ — множество таких k, для которых (i, k) есть ребро графа, т. е.

$$V(i) = \{k \mid (i, k) \in G^1\}.$$

Число элементов в $V(i)$ обозначим n_i. Пусть

$$p_i\colon V(i) \to V(i)$$

есть циклическая перестановка множества $V(i)$, $i = 1, \ldots, n$, т. е. взаимно однозначное преобразование с единственной орбитой длины n_i.

Теорема Эдмондса утверждает, что каждый набор $(p_1, \ldots, p_n)$ определяет 2-клеточное вложение $G(M)$ графа G в ориентируемое 2-многообразие M, для которого существует ориентацня M, индуцирующая циклическое упорядочение ребер (i, k), инцидентных i, причем за ребром (i, k) непосредственно следует ребро $(i, p_i(k))$, $i = 1, \ldots, n$. Более того, для любого 2-клеточиого вложения $G(M)$ графа G в ориентируемое 2-многообразие M с данной ориентацией существует набор $(p_1, \ldots, p_n)$, определяющий это вложение описанным образом.

Действительно, существует алгоритм, который по данному $(p_1, \ldots, p_n)$ строит соответствующее вложение.

6.3. Опишем этот алгоритм. Определим W как множество *упорядоченных* пар $[a, b]$, для которых соответствующая неупорядоченная пара $(a, b) \in G^1$. Определим

$$P\colon W \to W$$

формулой

$$P[a, b] = [b, p_b(a)].$$

Преобразование P определено корректно, поскольку из $(a, b) \in G^1$ следует $a \in V(b)$, а поэтому $p_b(a)$ имеет смысл. Легко видеть, что преобразованне P взаимно однозначио.

Рассмотрим любую орбиту R преобразования P и предположим, что ее длина есть k.

Возьмем ориентируемую 2-клетку S в виде многоугольника с k сторонами. Обозиачим какую-иибудь сторону $[a, b]$, следующую за ней в иаправлеиии ориентации, через $P[a, b]$ и т. д. Получим такие обозначеиия:

$$[a, b],\ P[a, b],\ \ldots,\ P^{k-1}[a, b].$$

Если $R_1, \ldots, R_N$ есть совокупность различных орбит P, то описаиный выше процесс определяет 2-клетки $S_1, \ldots, S_N$. Заметим, что миожество $R_1, \ldots, R_N$ обладает таким свойством: если $[a, b]$ попадает в одну из этих орбит, то и $[b, a]$ попадает, причем каждое по одиому разу. Более того, поскольку граф G связный, никакое собственное непустое подмножество множества $R_1, \ldots, R_N$ этим свойством не обладает.

Теперь остается получить ориентируемое 2-многообразие при помощи обычного склеиваиия, как в элементариой топологии.

Из определения P и из того, что $W = \bigcup R_i$, следует, что стороны 2-клеток $S_1, \ldots, S_N$ образуют 2-клеточное вложение $G(M)$ графа G в M, при котором компонентами $[M - G(M)]$ являются внутренности этих клеток. Ориентация клеток $S_1, \ldots, S_N$ определяет ориентацию многообразия M. Противоположная ориентация M повлечет за собой такое циклическое упорядочение ребер (i, k), инцидентных i, при котором (i, k) непосредственно предшествует $(i, p_i(k))$, $i = 1, \ldots, n$.

6.4. Получение орбит $R_1, \ldots, R_N$ можно систематизировать следующим образом.

Возьмем наименьшее k_1 в $V(1)$ и запишем R_1^1, орбиту, определенную $[1, k_1]$.

Найдем наименьшее $k_2 \in V(1)$, при котором $[1, k_2]$ не содержится в R_1, и запишем R_2^1, орбиту, определенную $[1, k_2]$.

Возьмем наименьшее $k_3 \in V(1)$, при котором $[1, k_3]$ не содержится ни в R_1, ни в R_2, и запишем R_3^1, орбиту, определенную $[1, k_3]$.

После конечного числа таких шагов мы запишем все орбиты, определенные парами $[1, k]$, $k \in V(1)$.

Теперь рассмотрим $V(2)$. Либо мы уже записали все орбиты, определяемые парами $[2, j]$, $j \in V(2)$, либо существует наименьшее $j_1 \in V(2)$, при котором $[2, j_1]$ в записанных орбитах отсутствует. В первом случае перейдем к $V(3)$, во втором — запишем орбиту R_1^2, определенную $[2, j_1]$. Продолжая этот процесс, мы получим все орбиты преобразования P.

6.5. Не вдаваясь в подробности, заметим, что число N орбит P есть функция $N = N(p_1, \ldots, p_n)$ выбранных заранее циклических перестановок p_i: $V(i) \to V(i)$, $i = 1, \ldots, n$, а 6.4 дает способ вычисления N.

Мы отмечали в 6.3, что 2-клеточное вложение $G(M)$, определенное набором $(p_1, \ldots, p_n)$, таково, что $[M - G(M)]$ имеет $N(p_1, \ldots, p_n)$ компонент.

Теорема Эдмондса утверждает, в частности, что для любого 2-клеточного вложения $G(M)$ в ориентируемое 2-многообразие M существует такой набор перестановок $(p_1, \ldots, p_n)$, который определяет это вложение описанным образом. С другой стороны, теорема характеризации 4.3 утверждает, что набор $(p_1, \ldots, p_n)$, на котором $N(p_1, \ldots, p_n)$ достигает максимума, определяет минимальное вложение. Обозначим $\|G\| = \max N(p_1, \ldots, p_n)$.

Существует множество одинаково утомительных эффективных способов выписывания всех циклических перестановок p_i: $V(i) \to V(i)$, $i = 1, \ldots, n$. Таким образом, мы получили ре-

шение классической проблемы о вычислении рода графа $\gamma(G)$ (см. 2.6). Действительно, согласно 2.5, имеем

$$\gamma(G) = [2 - \| G^0 \| + \| G^1 \| - \| G \|]/2.$$

Трудности применения этого алгоритма, конечно, возникают из-за огромного числа $[n_i - 1]!$ циклических перестановок p_i: $V(i) \to V(i)$, $i = 1, \ldots, n$, причем при вычислении $\| G \|$ (а, следовательно, и $\gamma(G)$) пока не удалось изобрести ничего существенно лучшего, чем полный перебор этих перестановок.

ЛИТЕРАТУРА

1. Edmonds, Jr., A combinational representation for polyhedral surfaces, *Amer. Math. Soc. Notices*, 7 (1960), 646.
2. Edmonds, Jr., Symmetric imbeddings of complete graphs, *Amer. Math. Soc. Notices*, 7 (1960), 948.
3. Ringel G., Färbungsprobleme auf Flächen und Graphen, Math. Monographien, v 2, VEB Deutscher Verlag der Wissenschaften, Berlin, 1959.
4. Roberts J. H., Steenrod N. E., Monotone transformations of two-dimensional manifolds, *Annals of Math.*, 39 (1938), 851—862.

Устранение лишнего из механических доказательств[1])

М. Девис

В математике исключение *опровергает* правило. Следовательно, общее утверждение арифметики (например, последнюю теорему Ферма) можно опровергнуть явным указанием единственного контрпримера. Когда каждый частный случай общего утверждения допускает проверку с помощью единообразного алгорифма (как это имеет место для последией теоремы Ферма), такой контрпример, если он существует, в конце концов может быть найден (в предположении, что нет никаких ограничений на время или память) с помощью вычислительной машины, запрограммированной для последовательного исследования всех возможностей. Возникает вопрос, имеется ли подобиая процедура поиска доказательства (а не опровержения) таких общих предложений. Если полностью описаны аксиомы, из которых должно исходить доказательство, то ответ на вопрос *положительный*. Единообразные процедуры поиска можно определить как процедуры, которые дают возможность машине (соответственно запрограммированной) обнаруживать доказательство данного предложения из данных аксиом, если такое доказательство существует. Такие процедуры поиска называются *логическими процедурами доказательства*. В последнее время проявлен значительный интерес к улучшению таких процедур и программированию их для вычислительных машин. Здесь мы дадим общую схему для сравнения различных существующих в настоящее время процедур доказательств, а также рассмотрим их преимущества и недостатки. В конце будет описана новая улучшенная процедура доказательства. Мы не будем обсуждать работу Ван Хао, потому что она не укладывается просто в нашу общую схему. Работа Ван Хао, а также и другая новейшая литература по рассматриваемому вопросу, указана в конце статьи.

---

[1]) Davis M., Eliminating the irrelevant from mechanical proofs, Proceedings of the Fifteenth Symposium in Applied Mathematics of the American Mathemat.cal Society, v. XV, Experimental arithmetic, high speed computing and mathematics, 1963, p. 15–30.

1. ЛОГИЧЕСКИЙ ЯЗЫК МАТЕМАТИКИ

В книгах или исследовательских статьях по математике утверждения формулируются на обычном русском[1]) (или немецком, французском, английском и пр.) языке с использованием математических символов. Развитие современной символической логики позволило формулировать математические утверждения, используя только технические символы и не используя слов. Удивительно, насколько экономичен этот „словарь" символической логики. Здесь мы дадим краткий логический словарь, который отвечает требованиям математики, но никоим образом не является максимально экономичным. Мы разбиваем логические символы на следующие группы: *знаки пунктуации, пропозициональные связки, переменные и кванторы.*

Знаки пунктуации: ( ) , . Это круглые скобки и запятая.

Пропозициональные связки: $\neg\ \vee\ \wedge\ \rightarrow\ \leftrightarrow$. Эти символы означают определенные операции над утверждениями; они приводят к новым утверждениям. Если p и q данные утверждения, то:

„$\neg p$" обозначает утверждение: „не p," т. е. „p не имеет места".

„$p \wedge q$" обозначает утверждение: „p и q".

„$p \vee q$" обозначает утверждение: „p или q".

„$p \rightarrow q$" обозначает утверждение: „Если p, то q".

„$p \leftrightarrow q$" обозначает утверждение: "p тогда и только тогда, когда имеет место q".

Двусмысленность и отсутствие точности в разговорном языке устраняется в случае пропозициональных связок следующими требованиями:

a) $\neg p$, $p \vee q$, $p \wedge q$, $p \rightarrow q$ и $p \leftrightarrow q$ должны рассматриваться, как вполне определенные утверждения, если таковы p и q, независимо от того, связано содержание p с содержанием q или нет;

b) истинность и ложность этих утверждений полностью определяется истинностью или ложностью p и q в соответствии с приводимой ниже „истинностной" таблицей (табл. I).

Таблица 1

p	q	$\neg p$	$p \vee q$	$p \wedge q$	$p \rightarrow q$	$p \leftrightarrow q$
t	t	f	t	t	t	t
f	t	t	t	f	t	f
t	f	f	t	f	f	f
f	f	t	f	f	t	t

[1]) В оригинале — „английском". Далее в тексте термин „английский" заменен на термин „русский". — *Прим. перев.*

Это требование имеет следствие, отчасти противоречащее интуиции: мы должны считать *истинным* такое неудобоваримое утверждение, как „Если $2+2=5$, то снег горячий". Но это лишь новый пример общематематической тенденции по возможности расширять область определения операций, даже если при этом придается смысл тому, что прежде считалось бессмыслицей (вспомним, например, $2-5$ или $\sqrt{-4}$). Дальнейшее обсуждение этого вопроса читатель найдет в любом учебнике логики, см., например [1, 2, 7, 10].

Переменные. Это символы, о которых предполагается, что они принимают произвольные значения из некоторого заранее фиксированного множества. Здесь мы будем использовать в качестве переменных буквы x, y, z, u, v и w с нижними индексами или без них.

Кванторы. Выражение, образованное заключением переменной в круглые скобки [например, (x) или (y)], называется *квантором общности.* Подразумевается, что выражение $(x)A$ утверждает истинность A для всех значений x. Аналогично выражение, образованное заключением в круглые скобки буквы E, за которой следует переменная [например, (Ex) или (Ey)], называется *квантором существования*; подразумевается, что выражение $(Ex)A$ утверждает истинность A по крайней мере для одного значения x.

В дополнение к только что перечисленным *логическим символам* мы должны ввести некоторые специальные *математические символы.* Примерами таких символов будут: 0, 1, $+$, $=$, $<$, $\in$. Какие из этих символов действительно встречаются, будет зависеть от рассматриваемой конкретной области математики. Однако все эти символы разделяются на три категории: *символы констант*, *символы функций* и *символы отношений.*

Предполагается, что каждый *символ константы* (обычные примеры 0 и 1) обозначает некоторую определенную константу. Предполагается, что каждый *символ функции* (например, $+$) обозначает функцию от одного или более аргументов (например, обычно предполагается, что $+$ обозначает функцию от двух аргументов). *Символ отношения* (например, $=$, $<$ или $\in$) обозначает отношение с одним или более аргументами.

В таблице 2 дается перевод различных утверждений из обычной математики в логический символизм.

Теперь мы можем систематически исследовать все утверждения, которые можно записать, используя этот символизм. Во-первых, исходя из переменных и символов констант, мы можем образовывать *термы*, т. е. выражения, построенные из этих символов с помощью различных символов функций. В частности, если g — символ функции от n аргументов и уже известно,

Таблица 2

Утверждение	Перевод в логический символизм	Подразумеваемая область значений переменных	Символы констант	Символы функций	Символы отношений
(1) Существует бесконечно много простых чисел	$(x)(Ey)[y>x\wedge P(y)]$	Положительные целые числа	Нет	Нет	P (... есть простое число) $>$ (... больше, чем ...)
(2) Сложение коммутативно	$(x)(y)(x+y=y+x)$	Элементы некоторой абелевой группы	Нет	$+$	$=$(... равно ...)
(3) Множество положительных целых чисел, содержащее 1 и вместе с каждым из своих элементов содержащее следующий за ним элемент, состоит из всех положительных целых чисел	$(v)\{[1\in v\wedge\wedge(x)(x\in v\rightarrow s(x)\in v)]\rightarrow(x)[I(x)\rightarrow x\in v]\}$	Множество всех положительных целых чисел и все множества положительных целых чисел	1	s (следующее за)	$\in$ (... принадлежит ...) I (... положительное целое число)
(4) Между n и $2n$ всегда имеется простое число	$(x)(Ey)[y>x\wedge x+x>y\wedge P(y)]$	См. (1)	Нет	$+$	$>$, P См. (1)

что $u_1, u_2, \ldots, u_n$ — термы, то $g(u_1, u_2, \ldots, u_n)$ — также *терм*. Обычно допускают небольшое изменение этой записи.

Для символа функции g от двух аргументов (вроде символа $+$) пишем (ugv) вместо $g(uv)$.

Примеры термов:

$$(x+y),\ (0+(x+y)),\ s(0+y).$$

Если R — символ отношения от n аргументов и $u_1, \ldots, u_n$ — термы, то выражение $R(u_1, \ldots, u_n)$ называется *атомарной формулой*. [Введенные выше особые соглашения о записи для символов функций от двух аргументов используются также и для символов отношений; например, пишут $(x=y)$ вместо $=(xy)$.] Итак, *утверждения, которые мы собираемся рассматривать — это утверждения, которые могут быть построены*

из атомарных формул с помощью пропозициональных связок и кванторов. Нетрудно увидеть, что все утверждения, приведенные в табл. 2, попадают в эту категорию. Кроме того, мы утверждаем, что *все математические суждения находятся среди утверждений, которые можно выразить таким способом.* Это утверждение иногда приводит к определенной путанице. Именно, виды логических выражений, которые мы допустили, часто называются выражениями „первого порядка", и утверждается, что для математики необходимы логики „более высокого порядка". Однако сами эти логики „более высокого порядка" могут быть выражены описанным способом (например, с помощью символа $\in$ из теории множеств). В частности, различные формулировки аксиоматической теории множеств могут быть выражены описанным способом (ср. [2]), и тогда, как известно, для развития классического анализа может быть использована программа Пеано — Дедекинда.

2. ЭРБРАНОВСКИЕ ДОКАЗАТЕЛЬСТВА

Пусть C — утверждение, выраженное в только что описанном символизме. В C входят логические символы (именно: знаки пунктуации, пропозициональные связки, переменные и кванторы) и математические символы (именно: символы констант, символы функций и символы отношений). *Интерпретация* утверждения C задается путем указания „смысла" математических символов. Точнее, интерпретация утверждения C задана, если указаны:

(1) непустое множество u, называемое *универсумом*;

(2) определенный элемент из u для каждого символа константы из C;

(3) функция от n переменных из u в u (т. е. отображение из u^n в u) для каждого символа функции от n аргументов;

(4) n-местное отношение на u для каждого символа отношения от n аргументов.

Когда интерпретация утверждения C указана, C становится определенным математическим утверждением — истинным или ложным. Например, пусть в качестве C выбрано (2) из табл. 2:

$$(x)(y)(x + y = y + x).$$

Ниже приводятся три интерпретации утверждения C.

(1) Универсум есть множество рациональных чисел; $+$ есть сложение; $=$ есть равенство.

(2) Универсум есть множество векторов в трехмерном евклидовом пространстве; $+$ есть векторное произведение; $=$ есть равенство.

(3) Универсум есть множество положительных целых чисел; $+$ есть умножение; $=$ есть „меньше чем“.

Здесь C истинно при интерпретации (1) и ложно при интерпретациях (2) и (3). Интерпретация, которая делает утверждение C истинным, называется *моделью* утверждения C. Под *моделью утверждений* $C_1, C_2, \ldots, C_n$ мы подразумеваем модель для одного утверждения $C_1 \wedge C_2 \wedge \ldots \wedge C_n$; иными словами, такая модель есть интерпретация математических символов из $C_1, \ldots, C_n$, которая делает все эти утверждения истинными. Например, рассмотрим следующий список утверждений, который мы обозначим через $\mathfrak{G}_1$:

$$(x)(e \circ x = x),$$
$$(x)(I(x) \circ x = e),$$
$$(x)(y)(z)[(x \circ (y \circ z)) = ((x \circ y) \circ z)],$$
$$(x)(x = x),$$
$$(x)(y)[x = y \rightarrow y = x],$$
$$(x)(y)(z)[((x = y) \wedge (y = z)) \rightarrow (x = z)],$$
$$(x)(y)[x = y \rightarrow I(x) = I(y)],$$
$$(x)(y)(u)(v)[((x = y) \wedge (u = v)) \rightarrow (x \circ u = y \circ v)].$$

То, что обычно называется *группой*, есть просто модель этих утверждений, в которой „$=$“ интерпретируется как равенство. Интерпретациями символов e, $\circ$ и I тогда будут соответственно тождественный элемент группы, умножение группы и функция, которая отображает каждый элемент группы в обратный ему. В действительности эти утверждения составляют формальную систему *аксиом теории групп*. В любой модели этих аксиом „$=$“ должно интерпретироваться как отношение эквивалентности, сохраняющее операции, которые являются интерпретациями символов I и $\circ$. Тогда соответствующие классы эквивалентности будут образовывать группу. Результат групповой операции (класс эквивалентности, содержащий $a \circ b$) будет при этом естественно получаться умножением класса эквивалентности, содержащего a, на класс, содержащий b.

Другой список утверждений, который также представляет интерес в связи с группами, это следующий список, который мы назовем $\mathfrak{G}_2$:

$$(x)P(e, x, x),$$
$$(x)(y)P(I(x), x, e),$$
$$(x)(y)(z)(u)(v)(w)[(P(x, y, u) \wedge P(u, z, w) \wedge P(y, z, v)) \rightarrow P(x, v, w)],$$
$$(x)(y)(z)(u)(v)(w)[(P(y, z, v) \wedge P(x, v, w) \wedge P(x, y, u)) \rightarrow P(u, z, w)].$$

Именно, если мы возьмем любую группу, то ее можно использовать как модель для $\mathfrak{G}_2$, интерпретируя e как тождественный элемент, I как функцию обращения и $P(x, y, z)$ как отношение $(x \circ y) = z$. Мы не будем обсуждать здесь обратный вопрос о сопоставлении некоторой группы каждой модели для $\mathfrak{G}_2$; заметим только, что любая модель для $\mathfrak{G}_2$, в которой интерпретация символа P такова, что для каждого x, y имеется *точно* одно z, для которого истинно $P(x, y, z)$, есть группа с операцией умножения:

$x \circ y$ есть то единственное z, для которого истинно $P(x, y, z)$.

Теперь предположим, что мы хотим доказать некоторое утверждение T. В интересующей нас ситуации мы имеем *конечный* список аксном $A_1, \ldots, A_m$, исходя из которых мы хотели бы доказать T. Это означает, что мы хотим показать, что

Каждая модель для $A_1, \ldots, A_m$ есть также модель для T.

Например, предположим, что мы хотим доказать элементарную теорему теории групп $(x)(x \circ I(x) = e)$, т. е. что постулированный левый обратный есть также и правый обратный. Мы хотим показать, что это предложение истинно во всех группах. Так как каждая группа есть модель для $\mathfrak{G}_1$, достаточно показать, что это предложение истинно в каждой модели для $\mathfrak{G}_1$. Или иначе, мы можем, используя обозначения системы $\mathfrak{G}_2$, записать это предложение в виде

$$(x)\, P(x,\ I(x),\ e)$$

и пытаться показать, что это последнее истинно в каждой модели для $\mathfrak{G}_2$. Такая точка зрения на доказуемость из аксиом обычно называется *семантической* или *теоретико-модельной*. В учебниках логики обычно подчеркивается другая, *синтаксическая* точка зрения, согласно которой выводимость T из $A_1, \ldots, A_m$ — это возможность использовать некоторый список формальных правил вывода для получения T из $A_1, \ldots, A_m$. Но тогда основным результатом оказывается теорема Гёделя о полноте, которая утверждает, что указанные правила вывода *полны* в следующем смысле:

T может быть получено из $A_1, \ldots, A_m$ с помощью указанных правил вывода тогда и только тогда, когда каждая модель для $A_1, \ldots, A_m$ есть также модель для T.

Итак, синтаксическая и семантическая точки зрения оказываются эквивалентными. Здесь мы будем пользоваться семантической точкой зрения. Более подробную информацию о теореме Гёделя можно получить в [1, 2, 7, 10].

При каждой интерпретации утверждение T либо истинно, либо ложно. Следовательно, каждая модель для $A_1, \ldots, A_m$ есть либо модель для T, либо модель для $\neg T$. Таким образом, утверждение, что каждая модель для $A_1, \ldots, A_m$ есть также модель для T, эквивалентно утверждению, что *никакая модель для* $A_1, \ldots, A_m$ *не есть также модель для* $\neg T$. Все процедуры доказательства, предложенные здесь, можно рассматривать с этой точки зрения: т. е. мы предлагаем *доказывать, что* T *есть следствие* $A_1, \ldots, A_m$, *показывая, что список предложений* $A_1, \ldots, A_m, \neg T$ *не имеет модели.*

Мы начинаем с предварительной обработки отдельных утверждений $A_1, \ldots, A_m, \neg T$, согласно шагам 1 — 8 приводимого ниже описания. Каждый шаг будет выполняться снова и снова до тех пор, пока это возможно; когда данный шаг нельзя больше применить, переходим к следующему шагу. Читатель должен заметить, что:

(1) Эти шаги могут быть легко сделаны чисто механическими и, следовательно, приспособлены для программирования на машине. (Практически эти предварительные шаги легко проделать вручную.)

(2) Наличие модели для некоторого списка утверждений эквивалентно наличию модели для списка, получающегося из рассматриваемого в результате одного из данных шагов.

Шаг 1. Переименование переменных. Если одна и та же переменная встречается более чем в одном кванторе в одном и том же утверждении, то используем новую переменную для одного из этих кванторов. Например, $(x)R(x) \vee (x)S(x)$ следует переписать в виде $(x)R(x) \vee (y)S(y)$.

Шаг 2. Исключение $\rightarrow$ *и* $\leftrightarrow$. Всякий раз, когда встречаются $\rightarrow$ и $\leftrightarrow$, делаем замену:

Заменяем $A \rightarrow B$ на $(\neg A) \vee B$.

Заменяем $A \leftrightarrow B$ на $(A \wedge B) \vee [(\neg A) \wedge (\neg B)]$.

Шаг 3. Продвижение символа $\neg$ *внутрь.* **Везде, где возможно, делаем замены:**

Заменяем $\neg(x)M$ на $(Ex)\neg M$.

Заменяем $\neg(Ex)M$ на $(x)\neg M$.

Заменяем $\neg(M \wedge N)$ на $(\neg M) \vee (\neg N)$.

Заменяем $\neg(M \vee N)$ на $(\neg M) \wedge (\neg N)$.

Заменяем $\neg\neg M$ на M.

В конце концов получаются утверждения, где каждое $\neg$ встречается непосредственно перед атомарной формулой.

Для атомарной формулы мы пишем $\bar{R}(U_1, \ldots, U_m)$ вместо $\neg R(U_1, \ldots, U_m)$; здесь $\bar{R}(U_1, \ldots, U_m)$ и $R(U_1, \ldots, U_m)$ называются литерами.

Таким образом, начав с

$$(x)\{R(x)\vee\neg(Ey)[S(x,\ y)\wedge(R(x)\vee U(x))]\},$$

мы получим

$$(x)\{R(x)\vee(y)[\overline{S}(x,\ y)\vee(\overline{R}(x)\wedge\overline{U}(x))]\}.$$

Шаг 4. Исключение квантора существования. Вычеркиваем поочередно кванторы существования. Соответствующая переменная (обозначим ее через y) заменяется на $g(x_1, \ldots, x_m)$, где g — новый символ функции, а $x_1, \ldots, x_m$ — все переменные, встречающиеся в кванторах общности, находящихся слева от вычеркиваемого квантора существования (у нас — (Ey)).

В случае, когда слева от рассматриваемого квантора существования нет кванторов общности, переменная заменяется на новый символ константы g; этот случай может быть включен в общий, если мы будем считать, что *символ константы — это просто символ функции от* 0 *аргументов.*

Чтобы доказать, что шаг 4 сохраняет свойство иметь или не иметь модель, мы заметим, что если до использования шага 4 рассматриваемые утверждения имели модель для любых $x_1, \ldots, x_m$ из рассматриваемого универсума, то можно найти хотя бы одно значение y, для которого обрабатываемое утверждение истинно. Следовательно, мы можем взять в качестве интерпретации для g такую функцию, что для любых $x_1, \ldots, x_m$ функция $g(x_1, \ldots, x_m)$ равна упомянутому значению y. (Можно избежать молчаливого использования аксиомы выбора, но мы не будем этим заниматься.) Обратно, если рассматриваемые утверждения имеют модель после обработки, то $y = g(x_1, \ldots, x_m)$ обращает исходное утверждение в истину для всех $x_1, \ldots, x_m$, так что первоначальное существовательное условие также имеет модель.

Шаг 5. Вынесение кванторов общности. Выносим все кванторы общности влево, так что рассматриваемое утверждение переходит в выражение стандартного вида (последовательность кванторов общности, за которой следует бескванторное выражение).

Шаг 6. Применение (*до тех пор, пока это возможно*) *дистрибутивности* $\vee$ *относительно* $\wedge$. Иными словами, заменяем $(A\wedge B)\vee C$ на $(A\vee C)\wedge(B\vee C)$.

После выполнения шагов 1 — 6 каждое утверждение принимает вид

$$(x_1)(x_2)\ldots(x_n)[B_1\wedge B_2\wedge\ldots\wedge B_k],$$

где для каждого B_i справедливо

$$B_i = l_1^{(i)} \vee l_2^{(i)} \vee \ldots \vee l_{r_i}^{(i)}$$

и где каждая $l_j^{(i)}$ есть литера. В этом выражении каждое B_i называется *простой дизъюнкцией* (*дизъюнктом*), $l_j^{(i)}$ называется *литерой* дизъюнкта B_i.

Шаг 7. Упрощение. Если дизъюнкт B_i содержит атомарную формулу вместе с ее отрицанием, то вычеркиваем весь дизъюнкт. Если литера встречается дважды в одном и том же дизъюнкте B_i, то вычеркиваем одно из вхождений (например, дизъюнкт $l_1 \vee l_2 \vee l_1$ следует заменить на $l_1 \vee l_2$).

Этим заканчивается предварительная обработка дизъюнктов. Нужно подчеркнуть, что здесь не утверждается, будто результат, получающийся после этой обработки, эквивалентен (по смыслу) $A_1, \ldots, A_m, \neg T$.

Утверждается лишь, что модель существует для одного из этих утверждений тогда и только тогда, когда она существует и для другого.

Мы воспользуемся преимуществами весьма простой формы, к которой приведены теперь рассматриваемые утверждения, чтобы ввести некоторые сокращения. Мы вычеркиваем кванторы общности. Мы также устраняем явное использование символов $\wedge$ и $\vee$. Таким образом, пишем $l_1 l_2 \ldots l_r$ вместо $l_1 \vee l_2 \vee \ldots \vee l_r$. Конъюнкция дизъюнктов представляется вертикальной колонкой, составленной из них. Дизъюнкты, происходящие от каждого из утверждений $A_1, \ldots, A_m, \neg T$, можно теперь собрать в единый список дизъюнктов. Этот единый список можно дальше упростить, используя следующий шаг.

Шаг 8. Исключение лишних дизъюнктов. Вычеркнем все дизъюнкты, частью которых являются другие (или которые совпадают с другими). (Действительно, конъюнкция всех дизъюнктов истинна тогда и только тогда, когда все дизъюнкты истинны; но если некоторый дизъюнкт истинен, то истинен также и любой дизъюнкт, частью которого является данный.)

Например, пусть $A_1, \ldots, A_4$ есть аксиомы теории групп $\mathfrak{G}_2$, и пусть T есть $(x)P(x, I(x), e)$. Применяя шаги 1—8 к $\mathfrak{G}_2$ вместе с $\neg T$, мы получим дизъюнкты:

$$P(e, x, x), \tag{1}$$

$$P(I(x), x, e), \tag{2}$$

$$\bar{P}(x, y, u)\,\bar{P}(u, z, w)\,\bar{P}(y, z, v)\,P(x, v, w), \tag{3}$$

$$\bar{P}(y, z, v)\,\bar{P}(x, v, w)\,\bar{P}(x, y, u)\,P(u, z, w), \tag{4}$$

$$\bar{P}(a, I(a), e), \tag{5}$$

где a — символ константы. Чтобы увидеть, как обрабатывается, например, $\neg T$, мы начинаем с $\neg(x)P(x, I(x), e)$. Шаги 1 и 2 не применимы. Шаг 3 дает

$$(Ex)\bar{P}(x, I(x), e).$$

Шаг 4 дает $\bar{P}(a, I(a), e)$, и оставшиеся шаги не применимы.

Объединенный список дизъюнктов $C_1, C_2, \ldots, C_k$ можно мыслить как единое бескванторное выражение

$$C_1 \wedge C_2 \wedge \ldots \wedge C_k.$$

Такое выражение находится в *конъюнктивной нормальной форме*; это означает, что оно есть конъюнкция дизъюнктов, к которым применено упрощение шага 7. Обычно выражение в конъюнктивной нормальной форме (или более общо, всякое бескванторное выражение) называется *тождественно ложным*, если это выражение ложно при любом сопоставлении значений „истина", „ложь" его атомарным формулам (это эквивалентно требованию, чтобы его отрицание было тавтологией).

С каждым полным списком дизъюнктов, полученных из $A_1, \ldots, A_m, \neg T$, мы связываем некоторый список H термов, который мы называем *эрбрановским универсумом* для этого списка. Сперва определим H_0 как множество символов констант и переменных, встречающихся в этом списке. Далее полагаем по определению, что H_{n+1} состоит из элементов H_n вместе со всеми термами, которые могут быть получены путем применения символов функций, встречающихся в рассматриваемом списке дизъюнктов, к элементам из H_n. Наконец, мы полагаем

$$H = \bigcup_n H_n.$$

Для только что рассмотренного примера (теории групп) имеем

$H_0 = \{e, a, x, y, z, u, v, w\}$,

$H_1 = \{e, a, x, y, z, u, v, w, I(e), I(a), I(x), I(y), I(z), I(u), I(v), I(w)\}$.

Все рассматриваемые нами процедуры поиска доказательств основываются на следующей форме теоремы Эрбрана.

Теорема Эрбрана. *Пусть $A_1, \ldots, A_m$ — данные аксиомы, и T — предполагаемая теорема. Пусть $C_1, C_2, \ldots, C_k$ — список дизъюнктов, полученных из $A_1, \ldots, A_m$ и $\neg T$ с помощью шагов 1—8, и пусть H — результирующий эрбрановский универсум. В этом случае T есть следствие $A_1, \ldots, A_m$ тогда и только тогда, когда существует такой список дизъюнктов $Q_1, \ldots, Q_r$,*

каждый из которых получен из некоторого C_i, $i = 1, 2, \ldots, k$, *путем замены его переменных членами* H, *что формула* $Q_1 \wedge Q_2 \wedge \ldots \wedge Q_r$ *— тождественно ложная.*

Эта теорема доказывается следующим образом:

1) Предположим, что $Q_1 \wedge Q_2 \wedge \ldots \wedge Q_r$ тождественно ложно, но тем не менее $A_1, \ldots, A_m, \neg T$ имеет модель. Следовательно, $C_1, \ldots, C_k$ имеют модель с универсумом U. Припишем каждой переменной и каждому символу константы в H некоторый фиксированный элемент из U. Тогда каждому элементу множества H будет сопоставлен единственный элемент множества U. При этой интерпретации каждое Q_i $(i = 1, 2, \ldots, r)$ истинно. Таким образом, $Q_1 \wedge Q_2 \wedge \ldots \wedge Q_r$ не может быть тождественно ложным.

2) Обратно, предположим, что $A_1, \ldots, A_m, \neg T$ не имеют модели. Тогда $C_1, \ldots, C_k$ не имеют модели. Пусть Q_1, Q_2, Q_3, ... — всевозможные дизъюнкты, полученные из $C_1, \ldots, C_k$ путем замены его переменных членами универсума H. Тогда мы утверждаем, что при любом сопоставлении истинностных значений атомарным формулам из Q_1, Q_2, Q_3, ..., по крайней мере одно из Q_i будет ложным. Допустив временно, что это утверждение верно, мы закончим доказательство, так как, согласно [10, стр. 254—256] („закон бесконечной конъюнкции"), мы тогда будем иметь, что $Q_{s_1} \wedge Q_{s_2} \wedge \ldots \wedge Q_{s_r}$ тождественно ложно для некоторых $s_1, s_2, \ldots, s_r$. Наконец, чтобы обосновать истинность нашего утверждения, достаточно заметить, что если бы оно было ложно, то мы могли бы построить модель для $C_1, \ldots, C_k$ с универсумом H. Интерпретация символов констант и функций в этой модели определяется автоматически. И если дано сопоставление истинностных значений атомарным формулам из Q_1, Q_2, Q_3, ..., которое делает все Q_i истинными, то мы можем интерпретировать каждый символ отношения R, считая $R(U_1, \ldots, U_m)$ истинным для $U_1, \ldots, U_m \in H$, тогда и только тогда, когда данное сопоставление истинностных значений делает атомарную формулу $R(U_1, \ldots, U_m)$ истинной. Таким образом, теорема доказана.

Используя теорему Эрбрана, мы можем построить любое число полных процедур поиска доказательств следующим способом.

Если даны аксиомы $A_1, \ldots, A_m$ и предполагаемая теорема T, то мы сперва применяем шаги 1—7 для получения списка дизъюнктов $C_1, \ldots, C_k$. Затем применяем процедуру *порождения*, которая порождает дизъюнкты $Q_1, Q_2, \ldots$, получаемые из $C_1, \ldots, C_k$ путем замены переменных на члены эрбрановского универсума. Время от времени мы прерываем процедуру

порождения и проверяем конъюнкцию $Q_1 \wedge Q_2 \wedge \ldots \wedge Q_N$ дизъюнктов, порожденных к этому моменту, на тождественную ложность. Если эта конъюнкция тождественно ложна, то мы

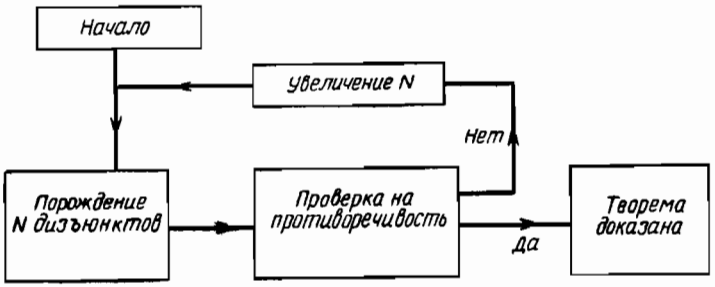


Рис. 1.

имеем доказательство. Если нет, порождаем **новые дизъюнкты.** Эта общая схема иллюстрируется на рис. 1.

3. ПРОЦЕДУРА ГИЛЬМОРА

Процедура Гильмора [6] может быть описана (с несущественными изменениями) следующим образом. Пусть переменные, присутствующие в $C_1, \ldots, C_k$, — это $x_1, \ldots, x_n$. Пусть $(t_1^{(j)}, \ldots, t_n^{(j)})$, $j = 1, 2, 3, \ldots$, — некоторое определенное перечисление всех элементов универсума H. Пусть $C_1^{(j)}, \ldots, C_k^{(j)}$ есть результат замены $x_1, \ldots, x_n$ на $t_1^{(j)}, \ldots, t_n^{(j)}$ в $C_1, \ldots, C_k$ соответственно. Формулу

$$\left.\begin{array}{c} C_1^{(1)} \wedge C_2^{(1)} \wedge \ldots \wedge C_k^{(1)} \wedge \\ \wedge C_1^{(2)} \wedge C_2^{(2)} \wedge \ldots \wedge C_k^{(2)} \wedge \\ \cdots\cdots\cdots\cdots \\ \wedge C_1^{(rN)} \wedge C_2^{(rN)} \wedge \ldots \wedge C_k^{(rN)} \end{array}\right\} \qquad (*)$$

приводим к *дизъюнктивной* нормальной форме для каждого из последовательных значений r; здесь N — некоторое данное целое число (допустим 10).

Приведение (*) к дизъюнктивной нормальной форме означает итерирование распределения $\vee$ относительно $\wedge$ (т. е. замену, где возможно, $(A \vee B) \wedge C$ на $(A \wedge C) \vee (B \wedge C)$) до тех пор, пока формула не примет вид

$$V_1 \vee V_2 \vee \ldots \vee V_k,$$

где для каждого V_i справедливо

$$V_i = l_1^{(i)} \wedge l_2^{(i)} \wedge \ldots \wedge l_{r_i}^{(i)}.$$

Такая формула тождественно ложна тогда и только тогда, когда в каждом V_i одна из встречающихся литер есть отрицание другой. (В противном случае можно так выбрать истин-

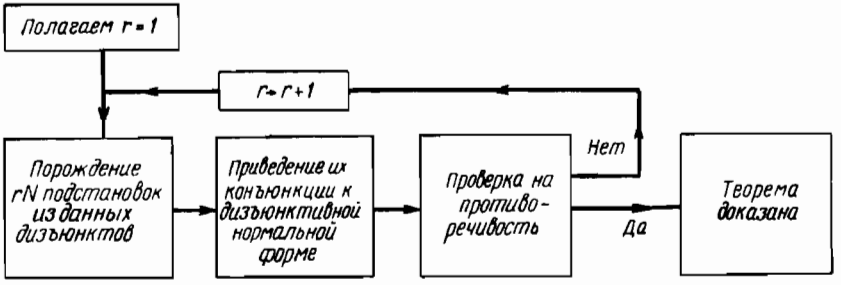


Рис. 2.

ностные значения атомарных формул, что взятое V_i, и следовательно вся формула, станет истинной.)

Вся процедура может быть представлена схемой, показанной на рис. 2.

4. ПРОЦЕДУРА ДЕВИСА — ПАТНЭМА

Процедура Гильмора, которая запрограммирована на IBM-704, не смогла дать доказательство следующего примера:

$$(Ex)(Ey)(z)\{[F(x, y) \rightarrow (F(y, z) \wedge F(z, z))] \wedge$$
$$\wedge [(F(x, y) \wedge G(x, y)) \rightarrow (G(x, z) \wedge G(z, z))]\}.$$

Эта теорема должна быть доказана без использования специфических аксиом, т. е. *это логическая теорема.* Применяя шаги 1—8 к отрицанию этой теоремы (заметим, что шаг 8 исключает три дизъюнкта), мы получаем список дизъюнктов:

$$F(x, y),$$
$$\bar{F}(y, f(x, y))\, \bar{F}(f(x, y), f(x, y))\, G(x, y),$$
$$\bar{F}(y, f(x, y))\, \bar{F}(f(x, y), f(x, y))\, \bar{G}(x, f(x, y))\, \bar{G}(f(x, y), f(x, y)).$$

Сообщалось, что программа Гильмора испытала ровно 7 подстановок, т. е. $rN = 7$. Легко видеть, что при этом получается дизъюнктивная нормальная форма из 12^7 конъюнктивных

дизъюнктов, каждый из которых содержит 21 литер. Процедура из работы [3] (так же, как и несколько улучшенная версия из работы [4]) совершенно аналогична процедуре Гильмора, но не использует сведения к дизъюнктивной нормальной форме. Процедура проверки на противоречивость формулы в конъюнктивной нормальной форме происходит по следующим правилам:

1. *Правило однолитерных дизъюнктов.*

(a) Если имеются однолитерные дизъюнкты p и $\bar{p}$, то формула противоречива.

(b) Если (a) не применимо и имеется однолитерный дизъюнкт l, то вычеркиваются все вхождения $\neg l$, а также все дизъюнкты, в которых встречается l. (Под $\neg l$ мы понимаем $\bar{p}$, если l — атомарная формула p, и понимаем p, если l есть атомарная формула $\bar{p}$.) Если после этого совсем не остается дизъюнктов, то формула непротиворечива.

2. *Положительно-отрицательное правило.* Если имеется литера l, но нет $\neg l$, то можно вычеркнуть все дизъюнкты, содержащие l. Если после этого совсем не остается дизъюнктов, то рассматриваемая формула непротиворечива.

3. *Правило расщепления.* Пусть p — атомарная формула, такая, что имеются как p, так и $\bar{p}$. Пусть A_1 и A_2 получены из исходной формулы путем вычеркивания всех дизъюнктов, в которых встречаются соответственно p или $\bar{p}$. Тогда исходная формула противоречива тогда и только тогда, когда A_1 и A_2 обе противоречивы.

Пример Гильмора легко выводится (даже вручную) с использованием этой процедуры (ср. [3]); но задача из теории групп, рассматриваемая выше, оказывается трудной (обсуждение трудностей см. в [4]).

5. ПРОЦЕДУРА ПРАВИЦА

Недостатком процедуры Девиса — Патнэма является порождение значительно большего числа дизъюнктов, чем нужно для завершения доказательства. Правиц в работе [9] предлагает алгорифм, свободный от этого дефекта.

Пусть дан список дизъюнктов $C_1, \ldots, C_k$, и пусть входящие в него перемеиные — это $x_1, \ldots, x_n$. Пусть теперь $C_1^{(j)}, \ldots, C_k^{(j)}$ суть соответственно результаты замены $x_1, \ldots, x_n$ новыми различными переменными $x_{1j}, \ldots, x_{nj}$. Сиова мы выбираем целое число N, и для последовательных значений r мы приводим формулу (*) из разд. 3 к *дизъюнктивной нормальной форме.* Затем ищем подстановки элементов эрбрановского универсума H вместо переменных, которые могут перевести две литеры в одном и том же (конъюнктивном) дизъюнкте в отрицание друг друга.

Если не существует таких подстановок для всех дизъюнктов, то увеличиваем r и начинаем снова.

В качестве примера [9] рассмотрим логическую теорему

$$(x)(y)[P(x) \vee Q(y)] \rightarrow [(x)P(x) \vee (y)Q(y)].$$

Шаги 1—8 дают список дизъюнктов:

$$P(x)Q(y),$$
$$\bar{P}(a),$$
$$\bar{Q}(b).$$

В дизъюнктивной нормальной форме это переходит в формулу

$$[P(x) \wedge \bar{P}(a) \wedge \bar{Q}(b)] \vee [Q(y) \wedge \bar{P}(a) \wedge \bar{Q}(b)],$$

которая будет противоречивой только в случае $x = a$ и $y = b$. Здесь для доказательства достаточно положить $rN = 1$.

Большим достоинством процедуры Правица является порождение только таких подстановок в $C_1 \wedge C_2 \wedge \ldots \wedge C_k$, которые действительно нужны для противоречия. Но использование дизъюнктивной нормальной формы неизбежно порождает те же самые трудности, что и в процедуре Гильмора.

6. НОВАЯ ПРОЦЕДУРА

Мы теперь опишем новый тип процедуры, в которой попытаемся объединить достоинства процедуры Правица и процедуры Девиса — Патнэма.

Определение. *Формула в конъюнктивной нормальной форме называется связной конъюнкцией, если для каждого вхождения литеры l в данный дизъюнкт существует по крайней мере одно вхождение $\neg l$ в какой-нибудь другой дизъюнкт. Назовем такое вхождение $\neg l$ партнером для данного вхождения l.*

Теорема. *Пусть $C_1, \ldots, C_k$ — список (дизъюнктивных) дизъюнктов, конъюнкция которых находится в конъюнктивной нормальной форме и является противоречием. Тогда существует подмножество $\{C_1, \ldots, C_k\}$, конъюнкция которого противоречива и является связной конъюнкцией.*

Предположим, что некоторый дизъюнкт содержит литеру l и что нет другого дизъюнкта, содержащего $\neg l$. Тогда все дизъюнкты, содержащие l, можно вычеркнуть. Действительно, полагая l истинным, мы сделаем истинными также все дизъюнкты, содержащие l. Следовательно, всякое сопоставление истинностных значений атомарным формулам должно делать

ложным некоторый дизъюнкт, не содержащий l. (Это замечание является просто обоснованием положительно-отрицательного правила.) Повторение этого процесса должно привести к связной конъюнкции, которая также противоречива.

Эта теорема о связной конъюнкции (или, скорее, двойственная ей) была независимо замечена Данхэмом и Нортом [5].

Следовательно, мы можем модифицировать процедуру Девиса – Патнэма так, чтобы искать именно связные конъюнкции; проверку на противоречивость надо делать только после получения связной конъюнкции. Найдем подходящую связную конъюнкцию в рассматриваемой задаче теории групп.

Искомая связная конъюнкция должна содержать дизъюнкт (5), так как он – единственный след доказываемой теоремы. Подстановка в (1) или (2) никогда не даст (5). Следовательно, в связной конъюнкции должен присутствовать дизъюнкт, который получается подстановкой из (3) или (4). Попытка использовать (3) таким способом приведет в тупик, как легко может проверить читатель. Используя (4), получим:

$$\bar{P}(a, I(a), e),$$
$$\bar{P}(y, I(a), v)\,\bar{P}(x, v, e)\,\bar{P}(x, y, a)\,P(a, I(a), e).$$

При подстановке в этот список мы должны заставить первые три литеры иметь ту же форму, что и какая-нибудь подстановка в (1) или (2) или в последнюю литеру из (3) (или, быть может, другая подстановка в (4)). Достичь этого можно различными путями; мы выбираем подстановку $y = e$, $v = I(a)$. Тогда первая литера будет иметь вид (1), и если мы далее положим $x = I(I(a))$, то вторая литера будет иметь вид (2). Это приводит к списку дизъюнктов:

$$\bar{P}(a, I(a), e),$$
$$\bar{P}(e, I(a), I(a))\,\bar{P}(I(I(a)), I(a), e)\,\bar{P}(I(I(a)), e, a)\,P(a, I(a), e),$$
$$P(e, I(a), I(a)),$$
$$P(I(I(a)), I(a), e).$$

Но так как третья литера второго дизъюнкта не имеет себе партнера, то это еще не есть связная конъюнкция. Дизъюнкты (1) и (2) не помогают, поэтому испытаем (3):

$$\bar{P}(I(I(a)), y, u)\,\bar{P}(u, z, a)\,\bar{P}(y, z, e)\,P(I(I(a)), e, a).$$

Здесь первая литера подсказывает подстановку $y = I(a)$, $u = e$, и тогда второй дизъюнкт подсказывает подстановку $z = a$.

Таким образом, получим связную конъюнкцию:

$\bar{P}(a, I(a), e)$,
$\bar{P}(e, I(a), I(a))\,\bar{P}(I(I(a)), I(a), e)\,\bar{P}(I(I(a)), e, a)\,P(a, I(a), e)$,
$P(e, I(a), I(a))$,
$P(I(I(a)), I(a), e)$,
$\bar{P}(I(I(a)), I(a), e)\,\bar{P}(e, a, a)\,\bar{P}(I(a), a, e)\,P(I(I(a)), e, a)$,
$P(e, a, a)$,
$P(I(a), a, e)$.

Нескольких применений правила однолитерного дизъюнкта достаточно, чтобы показать, что эта связная конъюнкция противоречива.

Остается предложить исчерпывающую процедуру поиска связных конъюнкций. Мы начнем с замечания, что в связной конъюнкции, полученной из (1) — (5), мы могли бы ожидать больше дизъюнктов, происходящих от (1) и (2), чем от (3) и (4). Дело в том, что каждое использование (3) или (4) требует от нас нахождения партнеров для трех литер вида $\bar{P}(U, V, W)$. Таким образом, мы можем предположить, что от (1) и (2) происходит примерно втрое больше дизъюнктов, чем от (3) или (4). Мы можем выразить это предположение в виде отношения:

$$3:3:1:1:1.$$

Вообще, пусть даны дизъюнкты $C_1, C_2, \ldots, C_k$; сопоставим им целые числа $n_1, n_2, \ldots, n_k$, где n_i/n_j мыслится как отношение числа ожидаемых дизъюнктов, происходящих от C_i, к числу дизъюнктов, происходящих от C_j. Для каждого фиксированного значения r, $r = 1, 2, 3, \ldots$, существует только *конечное число* существенно различных связных конъюнкций, содержащих не более чем rn_i дизъюнктов, происходящих от каждого C_i, $i = 1, 2, \ldots, k$. (Две связные конъюнкции *несущественно различны*, если одна может быть получена из другой подходящей заменой переменных.) Их можно систематически порождать в процессе поиска всевозможных путей нахождения партнера для различных литер.

Д. Мак-Илрой предложил интересную схему для выполнения поиска всех связных конъюнкций для каждого фиксированного значения r. Эта схема подобна некоторым игровым алгорифмам, так как она основана на *оценке* „позиций“ и *переходе* к „позиции“ с наилучшей оценкой.

Список дизъюнктов оценивается следующим образом. Каждое вхождение литеры, которое еще не имеет партнера, полу-

чает в качестве оценки суммарное число литер из других дизъюнктов, которые могут быть использованы в качестве партнеров для него (после подходящей подстановки). Оценка задачи теории групп начинается следующим образом:

$$P(e, \overset{6}{x}, x), \tag{1}$$

$$P(\overset{6}{I}(x), x, e), \tag{2}$$

$$\overline{P}(x, \overset{3}{y}, u)\,\overline{P}(u, \overset{3}{z}, w)\,\overline{P}(y, \overset{3}{z}, v)\,P(x, \overset{3}{v}, w), \tag{3}$$

$$\overline{P}(y, \overset{3}{z}, v)\,\overline{P}(x, \overset{3}{v}, w)\,\overline{P}(x, \overset{3}{y}, u)\,P(u, \overset{3}{z}, w), \tag{4}$$

$$\overline{P}(a, \overset{2}{I}(a), e). \tag{5}$$

Наша стратегия состоит в нахождении партнера сначала для литеры (или одной из литер) с наилучшей оценкой. Если, как в нашем случае, эта наилучшая оценка больше, чем 1, то имеется несколько возможностей. Мы испытаем их все одну за другой (просмотром вперед) и произведем переоценку. При этом мы сохраняем литеры с наименьшей минимальной оценкой. В нашем случае имеются две возможности:

$$P(e, \overset{5}{x}, x),$$

$$P(\overset{4}{I}(x), x, e),$$

$$\overline{P}(a, \overset{1}{y}, u)\,\overline{P}(u, \overset{3}{z}, e)\,\overline{P}(y, \overset{2}{z}, I(a))\,P(a, I(a), e),$$

$$\overline{P}(y, \overset{3}{z}, v)\,\overline{P}(x, \overset{3}{v}, w)\,\overline{P}(x, \overset{3}{y}, u)\,P(u, \overset{4}{z}, w),$$

$$\overline{P}(a, I(a), e)$$

с минимальной оценкой 1 и

$$P(e, \overset{5}{x}, x),$$

$$P(\overset{5}{I}(x), x, e),$$

$$\overline{P}(x, \overset{3}{y}, u)\,\overline{P}(u, \overset{3}{z}, w)\,\overline{P}(y, \overset{3}{z}, v)\,P(x, \overset{4}{v}, w),$$

$$\overline{P}(y, \overset{3}{I}(a), v)\,\overline{P}(x, \overset{3}{v}, e)\,\overline{P}(x, \overset{2}{y}, a)\,P(a, I(a), e),$$

$$\overline{P}(a, I(a), e)$$

с минимальной оценкой 2.

Следовательно, мы продолжим рассмотрение первого варианта. В конце концов нам, может быть, придется вернуться

ко второму либо потому, что не удастся получить никакой связной конъюнкции, либо потому, что полученная связная конъюнкция будет непротиворечива. В рассматриваемой задаче теории групп первой встречается связная конъюнкция, найденная выше.

Усовершенствование предложенной процедуры, которое, возможно, обладает некоторыми практическими преимуществами, состояло бы в попытке объединить связные конъюнкции с помощью новых подстановок таким образом, чтобы у них появились общие литеры. Это не является теоретически необходимым, но на практике могло бы привести к более быстрым доказательствам некоторых теорем.

Процедура доказательства, основанная на этих идеях, в настоящее время программируется на вычислительной машине IBM 7090 (Лаборатории компании Bell Telephone).

ЛИТЕРАТУРА

1. Черч А., Введение в математическую логику, т. I, ИЛ, М., 1960.
2. Davis M., Mathematical logic, Mimeographed lecture notes, New York Univ., Inst. Math. Sciences, New York, 1960.
3. Davis M., Putnam H., A computing procedure for quantification theory, *J. ACM*, 7 (1960), 201—215.
4. Davis M., Logemann G., Loveland D., A machine program for theorem-proving, *Comm. ACM*, 5 (1962), 394—397.
5. Dunham B., North J. H., Theorem testing by computer, Proc. Sympos. Math. Theory Automata, 1962, Polytechnic Press, Brooklyn, N. Y. 1963, pp. 173—177.
6. Gilmore P. C., A proof method for quantification theory, *IBM J. Res. Develop.*, 4 (1960), 28—35.
7. Гильберт Д., Аккерман В., Основы теоретической логики, ИЛ, М., 1947.
8. Prawitz D., An improved proof procedure, *Theoria*, 26 (1960), 102—139.
9. Prawitz D., Prawitz H., Voghera N., A mechanical proof procedure and its realization in an electronic computer, *J. ACM*, 7 (1960), 102—128.
10. Quine W. V. O., Methods of logic, Henry Holt, New York, 1959.
11. Robinson A., Proving a theorem (as done by man, logician or machine), Summaries of talks at Cornell Summer 1957, Institute for Symbolic Logic, IDA, Princeton, N. J., 1960, p. 350—352.
12. Wang Hao, Towards mechanical mathematics, *IBM J. Res. Develop*, 4 (1960), 2—22. (Русский перевод: Бан Хао, На пути к механической математике, Кибернетический сборник, вып. 5, ИЛ, М., 1962.)
13. Wang Hao, Proving theorems by pattern recognition, I, *Comm. ACM*, 3 (1960), 220—234.
14. Wang Hao, Proving theorems by pattern recognition, II, *BSTJ*, 40 (1961), 1—41.

Формализация и автоматическое доказательство теорем [1])

Ван Хао

ОБЩИЙ ОБЗОР

Главная цель этой статьи — привести ряд новых примеров механизации доказательств в арифметике и исчислении предикатов, которые, как мы полагаем, позволяют несколько продвинуться в развитии механической математики. Прежде чем приступить к детальному изложению, дадим в соответствии с пожеланиями организаторов конференции краткую сводку известных результатов.

Возможны различные пути использования вычислительных машин при доказательстве теорем. Чрезвычайно интересная работа Лемера [23] по доказательству теоретико-числовых теорем заключается по существу в сведении (человеком) общей теоремы к ряду сложных численных примеров, которые проверяются на машинах. Это несомненно разумный путь применения машин, при помощи которого уже доказаны новые теоремы. Но здесь мы будем рассматривать применения иного сорта, отличные от указанного пути, так как машина может выполнять только численные операции.

В качестве исходных данных не обязательно брать только аксиомы Пеано (с этим логики, по-видимому, согласятся). В связи с этим автор рассматривает вопрос о накоплении запаса стандартных математических результатов, а также допускает возможность частого вмешательства человека. Наш подход отличает желание применить хорошо известные способности машин к численным расчетам в области построения логических выводов.

Среди попыток осуществить такое применение различаются две основные тенденции, которые можно назвать психоцентрической и логоцентрической. В принципе различие между ними невелико. На практике психоцентрический подход обладает обманчивой привлекательностью, которая часто ведет к разочарованиям. Во всяком случае, лучше, по-видимому, оценивать

---

[1]) Wang Hao, Formalization and automatic theorem-proving, Proceedings of IFIP Congress 65, v. 1, 1965, p. 51—58.

каждую индивидуальную работу по ее конкретным результатам вместо того, чтобы вдаваться в общие рассуждения.

Известная работа Ньюэла, Шоу и Саймона [25, 26] имела важные побочные результаты в изучении языков программирования; что же касается доказательства теорем, она, по-видимому, мало пригодна для этой цели. Кроме того, хотя размышления по вопросам механического моделирования, вероятно, могут быть полезны для психолога, трудно предсказать, каким образом результаты осуществления этих идей на вычислительных машинах могли бы способствовать прогрессу в области психологии.

Схема доказательства геометрических теорем, разработанная Гелернтером и др. [17], дает результаты, которые сначала озадачивали. Но тщательный анализ Гильмора, по-видимому, показывает, что достигнуто было много меньше, чем первоначально казалось. „Авторы составили обычную машинную программу, в которой были реализованы совершенно определенные алгорифмы" (Гильмор [19], стр. 26).

Слегл [34] написал удачную программу для решения на машинах проблемы интегрирования в анализе. Было бы интересно изучить более подробно всю проблему интегрирования. Программа, которая составлена Бобровым [2] для решения на машине задач описательной алгебры, тщательно разработана и снабжена пояснениями. Интересно, что небольшой запас простых средств дает ему возможность манипулировать достаточно широким кругом английских предложений. Однако вопросы рассматриваемого нами типа являются побочными для его статьи.

В рамках логоцентрического подхода главное внимание пока что уделяется исчислению предикатов. Вне исчисления предикатов некоторые авторы, в частности Коллинз [5], рассматривали разрешающую процедуру Штурма — Тарского для элементарной алгебры. Робинсон предложил в несколько ином направлении систему для теории уравнений [29, 30]. Браун развил эффективную практическую систему для алгебраических преобразований [3].

Первая серьезная попытка поставить исчисление предикатов на машины, по-видимому, относится к лету 1958 г. [36]. Она дала довольно обнадеживающие результаты. С того времени было сделано много близких работ как по процедурам доказательства [4, 6—8, 18, 27, 28, 31, 32, 40], так и по разрешающим процедурам [14—16, 37] для интересных подклассов. Приятно наблюдать медленный, но неуклонный прогресс по направлению ко все более и более эффективным процедурам.

Основополагающий для обоих подходов теоретический результат восходит к творению Сколема и Эрбрана, известному под

названием теоремы Эрбрана. Однако для того, чтобы сделать теоретически возможные разложения [1]) практически пригодными, возникает новая серия вопросов, которая требует более сложных рассмотрений.

В области разрешимых подклассов исчисления предикатов Фридман [16] не только дает эффективную программу для решения вопроса о том, будет ли формула вида

$$(x_1) \ldots (x_m)(Ey_1)(Ey_2)(z_1) \ldots (z_n)\, M$$

теоремой, но также систематически рассматривает вопросы расширения (*дизъюнкты, которые сами о себе заботятся*). Непосредственные обобщения, которые желательно было бы получить, — это включение равенства и снятие ограничения на размерность предикатов.

Среди процедур доказательства, изученных на машинах, наиболее полезными методами, достаточно завершенными к настоящему моменту, являются, по-видимому, методы Чинлунда и др. [4] и Робинсона [31]. Более эффективная процедура разработана Робинсоном в работе [32], подобные идеи были независимо изложены Аандера [1]. Ни один из этих методов не был реализован на машине, хотя, по-видимому, работа Аандера содержит больше подробностей, нужных программисту. Для обоих методов желательно было бы включение равенства.

Ни одна из программ, законченных к настоящему времени, не смогла вывести примеры $ExQ1$, $ExQ2$, $ExQ3$ (в порядке возрастания трудности), рассматриваемые ниже. Надеемся, что эти примеры будут приняты в качестве тестов для новых программ.

Более общий вопрос состоит в том, является ли метод эрбрановских разверток достаточным в своей основе. Может быть необходимо ввести специальные стратегии, которые не имеют прямых связей с проверкой эрбрановских разверток. Например, ниже рассматривается стратегия устранения кванторов подстановкой Fu вместо $(Ex)(x = u \wedge Fx)$ или $(x)(x = u \supset Fx)$.

Для механизации рассуждений в случае арифметики сделано мало. Очень вероятно, что многие пытались сделать это и нашли задачу слишком трудной. Следующий раздел с примерами из арифметики посвящается предварительному исследованию некоторых возможностей. По-видимому, для недалекого будущего подходящей целью должна быть механизация доказательств из работы Сколема [33] [2]).

---

[1]) Имеются в виду эрбрановские развертки. — *Прим. ред.*

[2]) См. также: Гудстейн Р. Л., Рекурсивная теория чисел (в сб. «Рекурсивный математический анализ», „Наука", М., 1970). — *Прим. ред.*

Автор в других работах [35—39] уже много рассуждал о будущем автоматического доказательства теорем и не хочет здесь повторяться. По-видимому, необходимо подчеркнуть четыре коротких тезиса.

Хотя формализация есть только часть решения проблемы (так как для механизации нам нужен, кроме того, метод выбора следующего шага), весьма желательно выполнить больше работ по формализации, имея в виду расширение набора точных приемов. Привлекательная черта этого направления работ — возможность обойтись без машин. Следовательно, мы имеем область, *где не обязательно пачкать руки*. Это помогает разделению труда.

Нереалистично с самого начала ожидать практических методов, которые были бы полностью „автоматическими", так как цель состоит в том, чтобы расширить человеческие способности, поручив машинам наиболее утомительные шаги (см. [37], стр. 221, а также [31]).

Поучительно сопоставить возможность механизировать некоторый процесс с возможностью обучить его выполнению.

Монтаж различных методов в более широко применимую структуру представляет много серьезных трудностей. Например, булевы операции быстро выполнимы, но если мы введем такие операции на многих различных стадиях, то существенно, чтобы на каждой стадии требовалось не слишком много таких операций.

ПРИМЕРЫ ИЗ АРИФМЕТИКИ

Чтобы доказать некоторую формулу арифметики, мы допускаем, что она ложна и берем минимальный контрпример МК, который представляется *неопределенной константой*. В общем случае существовательное утверждение $(Ex)Fx$ дает неопределенную константу, которая произвольна, но обладает свойством F. Кванторы общности опускаются, и используются свободные переменные. Мы будем использовать a, b, c, x, y, z и т. д. в качестве переменных, m, n, k, x_m, y_m и т. д. — в качестве неопределенных констант. Основная стратегия состоит в том, чтобы выводить утверждения об этих неопределенных константах в надежде найти противоречивые. Переменные пробегают только положительные целые числа (исключая 0).

Чтобы сделать такой подход практичным, необходимо использовать некоторый запас известных математических результатов ЗР и список TF простых пропозициональных преобразований.

Следующие четыре примера предназначены просто для иллюстрации того, какого рода преобразования можно делать. Доказательства становятся все более и более эскизными и менее механическими. Можно обобщить методы, намеченные ниже, и получить *полную* систему арифметики в том смысле, что всякая теорема обычной аксиоматической арифметической системы теоретически может быть доказана; однако этот аспект здесь не будет рассматриваться. Например, чтобы выполнить повторные индукции, необходим, кроме МК, обычный выбор минимальной неопределенной константы.

Пропозициональные постулаты (включая постулаты для равенства) ограничены пока следующими:

TF1. Если A и $A \supset B$, то B.
TF2. Если A, C, $(A \wedge C) \supset B$, то B.
TF3. Если $A \supset B$ и $\neg B$, то $\neg A$.
TF4. Если $A \supset (B \vee C)$ и $\neg C$, то $A \supset B$.
TF5. Если $A \equiv B$ и A, то B (обычно для определений).
TF6. Замена равных на равные.
TF7. Если $\neg A$, вычеркнуть $A \supset B$.

Основные инструкции (для первого уровня сложности) таковы:

МК. Записать условия для минимальных контрпримеров.

НК. Сформировать неопределенные константы и подставить их вместо свободных переменных в предыдущие строчки доказательства.

TFL. Применять TF1 — TF7, пока это возможно, к полученным до сих пор строчкам доказательства, не содержащим свободных переменных.

TFG. Применять TF1 — TF7 как на основе строчек доказательства, так и на основе ЗР, но с ограничением, что хотя бы одна из посылок взята из доказательства и не содержит свободных переменных и что заключение не содержит функций и предикатов, не встретившихся в доказательстве.

ИП. Искать противоречия среди уже полученных строчек доказательства.

Вводится ряд механизируемых неявных соглашений. Например, сложение и умножение коммутативно и ассоциативно. Так, $x \nmid y$ и $x \nleq y$ для $\neg x \mid y$, $\neg x < y$; $x < y < z$ для $x < y \wedge y < z$; $a = b$ и $b = a$ — все считаются совпадающими.

За начальный ЗР мы взяли произвольно следующий список (с широкими возможностями усовершенствования):

B1. $a \nleq a$;
B2. $(a < b \wedge b < c) \supset a < c$;

B3. $a \leqq a$;
B4. $ab = ac \supset b = c$;
B5. $x \leqq y \equiv (x = y \vee x < y)$;
B6. $x = y \vee x < y \vee y < x$;
B7. $a \leqq b \supset b \not< a$.
D1. $x \mid x$;
D2. $(x \mid y \wedge y \mid z) \supset x \mid z$;
D3. $x \mid y \equiv y = xu_{xy}$;
D4. $xy = z \supset (x \mid z \wedge y \mid z)$;
D5. $x \mid y \supset x \leqq y$.
P1. $(Px \wedge x \mid yz) \supset (x \mid y \vee x \mid z)$;
P2. $y > 1 \supset [\neg Py \equiv (1 < x_y < y \wedge x_y \mid y)]$.

На первом этапе предпринимаемого доказательства мы поступаем следующим образом. Начало: (1) МК, (2) НК, (3) TFL, (4) ИП, (5) TFG, (6) ИП, (7) возвращаемся к (2) и повторяем. Выход: либо когда противоречие найдено на этапе ИП, либо когда НК не дает новых подстановок. В последнем случае требуется более сложное обращение к ЗР.

Пример 1. $ExN1.\ x > 1 \supset (Ey)(Py \wedge y \mid x)$.
Используя МК, получаем следующее:

$$m > 1, \tag{1}$$
$$Pb \supset b \not\mid m, \tag{2}$$
$$1 < a < m \supset (Py_a \wedge y_a \mid a). \tag{3}$$

Применяем НК:

$$Pm \supset m \not\mid m, \tag{4}$$
$$1 < m < m \supset (Py_m \wedge y_m \mid m). \tag{5}$$

Применение TFL и ИП ничего не дает. **Применяем** TFG. Используя TF7 и B1, можно вычеркнуть (5).
Используя (4), D1 и TF3, получаем

$$\neg Pm. \tag{6}$$

Далее,

$$\neg Pm \equiv (1 < x_m < m \wedge x_m \mid m), \text{ согласно (1), } \mathbf{P2}, \mathbf{TF1}. \tag{7}$$

В силу (6), (7), TF5 имеем

$$1 < x_m < m, \tag{8}$$
$$x_m \mid m. \tag{9}$$

Применение ИП не дает никакого результата. Возвращаемся к НК:

$$Px_m \supset x_m \nmid m \text{ после подстановки в (2) } x_m \text{ вместо } b, \quad (10)$$

$$1 < x_m < m \supset (Py_{x_m} \wedge y_{x_m} \mid x_m) \text{ после подстановки в (3)}$$
$$x_m \text{ вместо } a. \quad (11)$$

Применяем TFL:

$$\neg Px_m, \ [(9), 10), \text{TF3}], \quad (12)$$

$$Py_{x_m}, \ [(8), (11), \text{TF1}], \quad (13)$$

$$y_{x_m} \mid x_m, \ [(8), (11), \text{TF1}]. \quad (14)$$

Применение ИП не дает никакого результата. Применяем TFG:

$$\neg Px_m \equiv (1 < u_{x_m} < x_m \wedge u_{x_m} \mid x_m) \text{ в силу (8), P2, TF1}, \quad (15)$$

$$1 < u_{x_m} < x_m, \ [(12), (15), \text{TF5}], \quad (16)$$

$$u_{x_m} \mid x_m, \ [(12), (15), \text{TF5}], \quad (17)$$

$$y_{x_m} \mid m, \ [(9), (14), \text{D2, TF2}], \quad (18)$$

$$u_{x_m} \mid m, \ [(9), (17), \text{D2, TF2}]. \quad (19)$$

Применение ИП не дает никакого результата. Возвращаемся к НК:

$$Py_{x_m} \supset y_{x_m} \nmid m, \quad (20)$$

$$Pu_{x_m} \supset u_{x_m} \nmid m, \quad (21)$$

$$1 < y_{x_m} < m \supset (Py_{y_{x_m}} \wedge y_{y_{x_m}} \mid y_{x_m}), \quad (22)$$

$$1 < u_{x_m} < m \supset (Py_{u_{x_m}} \wedge y_{u_{x_m}} \mid u_{x_m}). \quad (23)$$

Применяем TFL:

$$y_{x_m} \nmid m, \ [(20), (13), \text{TF1}], \quad (24)$$

$$\neg Py_{x_m}, \ [(20), (18), \text{TF3}], \quad (25)$$

$$\neg Pu_{x_m}, \ [(21), (19), \text{TF3}]. \quad (26)$$

Применяем ИП и находим, что (24) противоречит (18), а (25) противоречит (13). Следовательно, $ExN1$ есть теорема арифметики.

Для того чтобы доказать следующую теорему, мы расширяем ЗР, присоединяя:

P3. $y > 1 \supset [\neg Py \equiv (1 < x_y < y \wedge Px_y \wedge x_y \mid y)]$,
B8. $(a \mid b \wedge a \mid (b + c)) \supset a \mid c$,
F1. $a \leqq b \supset a \mid b!$;
F2. $1 < a! + 1$;
F3. $a < a! + 1$;
F4. $\neg \mid a! + 1 \leqq a$.

Пример 2. $ExN2.\ Px \supset (Ey)(Py \wedge x < y \leqq x! + 1)$.

Условие о минимальности оказывается бесполезным, и мы для краткости его опускаем.

В силу МК:

$$Pm, \tag{1}$$

$$Pa \supset (a \leqq m \vee m! + 1 < a). \tag{2}$$

Применяем НК:

$$Pm \supset (m \leqq m \vee m! + 1 < m), \tag{3}$$

$$P(m! + 1) \supset (m! + 1 \leqq m \vee m! + 1 < m! + 1). \tag{4}$$

Применение TFL и ИП не дает ничего, кроме

$$m \leqq m \vee m! + 1 < m, \ [(1), (3), \text{TF1}]. \tag{5}$$

Применяем TFG:

$$P(m! + 1) \supset m! + 1 \leqq m, \ [(4), \text{B1}, \text{TF4}], \tag{6}$$

$$\neg P(m! + 1), \ [(6), \text{F4}, \text{TF3}], \tag{7}$$

$$1 < m! + 1, \ [\text{F2 (не обосновывается ограниченным TFG)}], \tag{8}$$

$$1 < x_{m!+1} < m! + 1, \ [(8), (7), \text{P3}, \text{TF1}, \text{TF5}], \tag{9}$$

$$Px_{m!+1}, \tag{10}$$

$$x_{m!+1} \mid m! + 1. \tag{11}$$

Применение ИП ничего не дает. Вернемся к НК:

$$Px_{m!+1} \supset (x_{m!+1} \leqq m \vee m! + 1 < x_{m!+1}). \tag{12}$$

Применяем TFL:

$$m! + 1 \not< x_{m!+1} \text{ в силу (9)}, \tag{13}$$

$$Px_{m!+1} \supset x_{m!+1} \leqq m, \ [(12), (13), \text{TF4}], \tag{14}$$

$$x_{m!+1} \leqq m, \ [(14), (10), \text{TF1}]. \tag{15}$$

Применение ИП не дает никакого результата. Применяем TFG:

$$x_{m!+1} \mid m!, \ [(15), \text{F1}, \text{TF1}], \tag{16}$$

$$x_{m!+1} \mid 1, \ [(16), (11), \text{B8}, \text{TF2}], \tag{17}$$

$$x_{m!+1} \leqq 1, \ [(17), \text{D5}, \text{TF1}], \tag{18}$$

$$1 \not< x_{m!+1}, \ [(18), \text{B7}, \text{TF1}]. \tag{19}$$

Но (19) противоречит (9).

Следующие два примера наводят на мысль, что необходимы более сложные стратегии, такие, как упорядочение всех встречающихся неопределенных констант, разбиение произведений, рассмотрение отдельных случаев $a = b$, $a < b$, $b < a$ и т. д.

Пример 3. *ExN*3. $Pw \supset wx^2 \neq y^2$.

Применение МК неоднозначно, так как здесь не одна свободная переменная. Выбираем следующее:

$$Pk, \tag{1}$$

$$km^2 = n^2, \tag{2}$$

$$(a < m \wedge b < n) \supset ka^2 \neq b^2. \tag{3}$$

Существуют различные пути применения НК, большей частью с тривиальными результатами. Применение TFL сравнительно бесполезно. Перейдем к либерализованному TFG, ведущий принцип которого — разбиение произведения. Имеем

$$k \mid n^2, \text{ [D4, (2), TF1]}, \tag{4}$$

$$m^2 \mid n^2, \tag{5}$$

$$k \mid n, \text{ [(1), (4), P1, TF2 (и соглашение о замене } A \vee A \text{ на } A)], \tag{6}$$

$$n = x_{kn}k, \text{ [(6), D3, TF5]}, \tag{7}$$

$$n^2 = k^2 (x_{kn})^2, \text{ используя } a = b \supset a^2 = b^2, \text{ так как } n^2 \text{ встречается выше}, \tag{8}$$

$$km^2 = k^2 (x_{kn})^2, \text{ [(8), (2), TF6]}, \tag{9}$$

$$m^2 = k (x_{kn})^2, \text{ [(9), B4, TF1]}, \tag{10}$$

$$m < n, \text{ согласно (2), (1) и } (Pa \wedge ab^2 = c^2) \supset b < c, \tag{11}$$

$$x_{kn} < n, \text{ согласно (7), (1) и } (Pa \wedge ab = c) \supset b < c. \tag{12}$$

Подставляем x_{kn} вместо a и m вместо b в (3):

$$(x_{kn} < m \wedge m < n) \supset k (x_{kn})^2 \neq m^2, \tag{13}$$

$$k (x_{kn})^2 \neq m^2, \text{ согласно (13), (11), (12), TF2}. \tag{14}$$

Но (14) противоречит (10).

Пример 4. *ExN*4.

$$\left[x = \prod_{i=1}^{u} y_i = \prod_{i=1}^{v} z_i \wedge (i)_1^{u-1} (y_i \leqq y_{i+1} \wedge Py_i \wedge Py_{i+1}) \wedge \right.$$

$$\left. \wedge (i)_1^{v-1} (z_i \leqq z_{i+1} \wedge Pz_i \wedge Pz_{i+1}) \right] \supset \left(u = v \wedge (i)_1^{u} y_i = z_i\right).$$

Возьмем минимальный контрпример $x = k$. Если $y_1 = z_1$, то $\prod_{i=2}^{u} y_i = \prod_{i=2}^{v} z_i < k$. Если $y_1 < z_1$, то $y_1 \prod_{i=2}^{v} z_i < \prod_{i=1}^{v} z_i$ и y_1 делит обе части неравенства. Отсюда $y_1 \mid (z_1 - y_1) \prod_{i=2}^{v} z_i$. Так как Pz_1, то $y_1 \nmid (z_1 - y_1)$ и $y_1 \mid z_2, \ldots z_n$. Отсюда $y_1 \prod_{i=1}^{w} t_i = (z_1 - y_1) \prod_{i=2}^{u} z_i < k$.

ПРИМЕРЫ НА ИСЧИСЛЕНИЕ ПРЕДИКАТОВ

Вместо того чтобы доказывать, что данная формула является теоремой, мы берем ее отрицание и выводим противоречие. Будем использовать для неопределенных констант запись, подобную записи их в разделе с примерами из арифметики. Основной пример настоящего раздела следующий:

Пример 1. *ExQ*1. Вывести противоречие из конъюнкции:

$$m \neq n, \tag{1}$$

$$n \neq k, \tag{2}$$

$$k \neq m, \tag{3}$$

$$y = m \vee [Fym \equiv (Ez)(z \neq m \wedge z \neq y \wedge Fyz \wedge Fzy)], \tag{4}$$

$$y = n \vee [Fyn \equiv \neg(Ez)(z \neq n \wedge z \neq y \wedge Fyz \wedge Fzy)], \tag{5}$$

$$y = k \vee [Fyk \equiv (y = m \vee y = n)]. \tag{6}$$

Имеются еще два родственных, но более трудных примера.

Пример 2. *ExQ*2. Заменить (2) и (3) в *ExQ*1 на

$$(n = k \vee k = m) \text{ и} \tag{2'}$$

$$y = j \vee (Fyj \equiv y = k). \tag{3'}$$

Пример 3. *ExQ*3. Вычеркнуть (2′) из *ExQ*2.

Не будем обсуждать здесь эти довольно трудные примеры. Для более интуитивного понимания указанных примеров следует мыслить F как отношение принадлежности $\in$.

Существуют три до некоторой степени различных метода решения *ExQ*1.

Первый метод — использовать разрешающую процедуру для специального префикса. Таким способом мы приводим конъюнкцию к предваренной форме с префиксом

$$(Em)(En)(Ek)(y)(z)(Eu)(Ev).$$

Если в программе, которую дает Фридман [16], удалить искусственные ограничения и присоединить методы обращения с равенством, то с примером *ExQ*1, по-видимому, можно справиться.

Второй метод. Подставим m, n, k вместо y в (4), (5), (6), т. е. действуем почти так же, как прежде с НК. Из девяти возможных подстановок три дают тривиальные результаты (m вместо y в (4) и т. д.). Мы рассмотрим ровно четыре из

остальных шести и упростим их с помощью (1), (2), (3), $k=k$, $n=n$:

$$Fkm \equiv (Ez)(z \neq m \wedge z \neq k \wedge Fkz \wedge Fzk),\ k \text{ вместо } y \text{ в (4)}, \quad (7)$$

$$Fkn \equiv \neg(Ez)(z \neq n \wedge z \neq k \wedge Fkz \wedge Fzk),\ k \text{ вместо } y \text{ в (5)}, \quad (8)$$

$$Fmk,\ m \text{ вместо } y \text{ в (6)}, \quad (9)$$

$$Fnk,\ n \text{ вместо } y \text{ в (6)}. \quad (10)$$

Попытаемся использовать (Ez), применяя эквивалентность $Gw \equiv (Ez)(z = w \wedge Gz)$. Из (6) получаем:

$$(Fyk \wedge y \neq k \wedge y \neq m) \supset y = n;$$

$$(Fyk \wedge y \neq k \wedge y \neq n) \supset y = m.$$

Поэтому мы можем добавить $y=n$ к правой части (7) и $y=m$ к правой части (8). Таким образом, исключая Ez в обоих случаях, получаем:

$$Fkm \equiv (n \neq m \wedge n \neq k \wedge Fkn \wedge Fnk),$$

$$Fkn \equiv \neg(m \neq n \wedge m \neq k \wedge Fkm \wedge Fmk).$$

Используя (1) и (2), сделаем упрощения:

$$Fkm \equiv (Fkn \wedge Fnk), \quad (11)$$

$$Fkn \equiv (\neg Fkm \vee \neg Fmk), \quad (12)$$

$$Fkn \equiv \neg Fkm, \text{ согласно (9) и (12)}, \quad (13)$$

$$Fkm \equiv Fkn, \quad \text{согласно (10) и (11)}. \quad (14)$$

Согласно (13) и (14), $Fkm \equiv \neg Fkm$, имеем противоречие.

Третий метод. Кванторы и $\equiv$ в (4) и (5) исключаем более механическим способом, что дает:

$$[y = m \vee Fym \vee (u = m \vee u = y \vee \neg Fyu \vee \neg Fuy)] \wedge$$
$$\wedge [y = m \vee \neg Fym \vee (x_y \neq m \wedge x_y \neq y \wedge Fyx_y \wedge Fx_y y)], \quad (4')$$

$$[y = n \vee \neg Fyn \vee (v = n \vee v = y \vee \neg Fyv \vee \neg Fvy)] \wedge$$
$$\wedge [y = n \vee Fyn \vee (w_y \neq n \wedge w_y \neq y \wedge Fyw_y \wedge Fw_y y)]. \quad (5')$$

Задача заключается в выводе противоречия из (1), (2), (3) (4'), (5') и (6). Мы записываем каждый дизъюнкт в конъюнктивной нормальной форме, так что результат является конъюнкцией многих дизъюнктов:

(1);
(2);
(3);

(4.1) $y = m \vee \neg Fym \vee x_y \neq m$;
(4.2) $y = m \vee \neg Fym \vee x_y \neq y$;
(4.3) $y = m \vee \neg Fym \vee Fyx_y$;
(4.4) $y = m \vee \neg Fym \vee Fx_y y$;
(4.5) $y = m \vee Fym \vee u = m \vee u = y \vee \neg Fyu \vee \neg Fuy$;
(5.1) $y = n \vee Fyn \vee w_y \neq n$;
(5.2) $y = n \vee Fyn \vee w_y \neq y$;
(5.3) $y = n \vee Fyn \vee Fyw_y$;
(5.4) $y = n \vee Fyn \vee Fw_y y$;
(5.5) $y = n \vee \neg Fyn \vee v = n \vee v = y \vee \neg Fyv \vee \neg Fvy$;
(6.1) $y = k \vee y \neq m \vee Fyk$;
(6.2) $y = k \vee y \neq n \vee Fyk$;
(6.3) $y = k \vee y = m \vee y = n \vee \neg Fyk$.

Подставляем m вместо y в (6.1), используя $m \neq k$ и $m = m$:

$$Fmk, \tag{15}$$

Fnk, [n вместо y в (6.2), $n \neq k$, $n = n$], (16)

$k = n \vee \neg Fkn \vee m = n \vee m = k \vee \neg Fkm \vee \neg Fmk$,

k вместо y и m вместо v в (5.5), (17)

$\neg Fkn \vee \neg Fkm$, используя (1) – (3), (15), (17), (18)

$Fkm \vee \neg Fkn$, k вместо y и n вместо u в (4.5),

используя (1), (2), (3), (16), (19)

$\neg Fkn$, используя (18), (19), (20)

$w_k \neq n$, $w_k \neq k$, Fkw_k, $Fw_k k$, k вместо y в (5.1) – (5.4),

используя (2) и (20), (21)

$w_k = m$, w_k вместо y в (6.3), используя (21), (22)

Fkm, используя (21) и (22), (23)

$x_k \neq m$, $x_k \neq k$, Fkx_k, $Fx_k k$, k вместо y в (4.1) – (5.4),

используя (3) и (23) (24)

$x_k = n$, x_k вместо y в (6.3), используя (21), (25)

Fkn, используя (24), (25). (26)

Отсюда видно, что (20) и (26) противоречат друг другу.

Это доказательство довольно близко по характеру к доказательству, полученному Чинлундом и др. [4]. Действительно, если мы обобщим метод Чинлунда и др. [4], присоединив равенство и исключив лишние шаги, машина, вероятно, сможет выдать в разумное время доказательство, более или менее похожее на данное. Если оставить в стороне вопрос о равенстве, имеется существенное теоретическое различие между методом

Чинлунда и др. [4] и приведенным доказательством, а именно: приведенное доказательство не использует предваренной нормальной формы. Такое усовершенствование предложено в диссертации Эрбрана (см. также [9, 20, 38]) и требует несколько более тщательного обоснования [10], чем метод, использующий предваренную форму.

ЛИТЕРАТУРА

1. Aanderaa S., A deterministic proof procedure (manuscript), Harvard, May 1964.
2. Bobrow D. G., Natural language input for a computer problem-solving system, Ph. D. thesis, MIT, September 1964.
3. Brown W. S., The ALPAK system, MM-63-1214-3, Bell Laboratories, April 1963
4. Chinlund T., Davis M., Hinman P. G., McIlroy D., Theorem-proving by machine, Bell Laboratories, Spring 1964.
5. Collins G. E., Computational reductions in Tarski's decision method for elementary algebra, IBM Corporation, Yorktown Heights, July 1962.
6. Davis M., Putnam H., A computing procedure for quantification theory, *J. ACM*, 7 (1960), 201—215.
7. Davis M., Logemann G., Loveland D., A machine program for theorem-proving, *Comm. ACM*, 5 (1962), 394—397.
8. Davis M., Eliminating the irrelevant from mechanical proofs, Proceedings of Symposium in Applied Mathematics. American Mathematical Society, v. 15, 1963. (Русский перевод: Девис М., Устранение лишнего из механических доказательств, см. настоящий сборник, стр. 160.)
9. Dreben B., Andrews P., Aanderaa S., False lemmas in Herbrand, *Bull. Am. Math Soc.*, **69** (1963), 699—706.
10. Dreben B., Wang H., A refutation procedure and its model-theoretic justification (manuscript), Harvard University, November 1964.
11. Dunham B., Fridshal R., Sward G. L., A nonheuristic program for proving elementary logical theorems, Proc. of the First International Conference on Information Processing, Paris, 1959; pub. Unesco, 1960, pp. 282—285.
12. Dunham B., Fridshal R., North J., Exploratory mathematics by machine, Recent Development in Information and Decision Processes, Robert E. Machol and Paul Grey (eds.), MacMillan, N. Y., 1962. (Proceedings of a Symposium at Purdue University, April 1961).
13. Dunham B., North J. H., Theorem testing by computer, paper presented April 24—26, 1962, Mathematical Theory of Automata, Polytechnic Press, Brooklyn, 1963, p. 173—177.
14. Friedman J., A semidecision procedure for the functional calculus, *J. ACM*, **10** (1963), 1—24.
15. Friedman J., A computer program for a solvable case of the decision problem, *J. ACM*, **10** (1963), 348—356.
16. Friedman J., A new decision procedure in logic and its computer realization, Ph. D. thesis, Harvard University, September 1964.
17. Gelernter H., Hanson J. R., Loveland D. W., Empirical investigations of the geometry theorem machine, Proceedings of the Western Joint Computer Conference, San Francisco, 1960, pp. 143—149.
18. Gilmore P. C., A proof method for quantification theory — its justification and realization, *IBM Journal of Research and Development*, **4** (1960), 28—35.

19. Gilmore P. C., An examination of the geometry theorem machine, IBM Corporation, Yorktown Heights, April 1962.
20. Herbrand J., Recherches sur la theorie de la demonstration, *Travaux de la Sociétě des Sciences et des Lettres de Varsovie, Cl. III, Math. Phys.*, **33** (1930).
21. Kanger S., A simplified proof method for elementary logic, Computer Programming and Formal Systems, P. Braffort and D. Hirschberg (eds.), North-Holland Publishing Co., Amsterdam, 1963, pp. 87—94. (Proceedings of Seminars at Blaricum, Holland in 1961.)
22. Kuroda S., An investigation of the logical structure of mathematics, XIII, A method of programming proofs in mathematics for electronic computers, *Nagoya Math. J.*, **16** (1960), 195—203.
23. Lehmer D. H., Some high speed logic, Proceedings symposium in Applied Mathematics, American Mathematical Society, v. 15, 1963.
24. McCarthy J., Computer programs for checking mathematical proofs, AMS Symposium on Recursive Function Theory, New York, April 1961.
25. Newell A., Shaw J. C., Simon H. A., Empirical explorations of the logic theory machine, Proceedings of the Western Joint Computer Conference, 1957.
26. Newell A., Shaw J. C., A variety of intelligent learning in a general problem solver, Self-Organizing Systems, M. C. Yovits and S. Cameron (eds.), New York, Pergamon Press, 1960, pp. 153—189. См. русский перевод в сб. «Самоорганизующиеся системы», стр. 211—261, «Мир», М., 1964.
27. Prawitz D., Prawitz H., Voghera N., A mechanical proof procedure and its realization in an electronic computer, *J. ACM*, 7 (1960), 102—128.
28. Prawitz D., An improved proof procedure, *Theoria (a Swedish Journal of Philosophy and Psychology)*, **26** (1960), 102—139.
29. Robinson A., On the mechanization of the theory of equations, *Bulletin of the Research council of Israel*, **9F** (1960), 47—70.
30. Robinson A., A basis for the mechanization of the theory of equations, Computer Programming and Formal Systems, pp. 95—99.
31. Robinson J. A., Theorem-proving on the computer, *J. ACM*, **10** (1963), 163—174.
32. Robinson J. A., A machine-oriented logic based on the resolution principle, *J. ACM*, **12** (1965), 23—41. (Русский перевод: Робинсон Дж. А., Машинно-ориентированная логика, основанная на принципе резолюции, см. настоящий сборник, стр. 194.)
33. Skolem Th., Begrűndung der Elementaren Arithmetik, 38 pp., Oslo, 1923.
34. Slagle J. R., A heuristic program that solves symbolic integration problems, Ph. D. thesis, MIT, May 1961.
35. Wang H., A variant to Turings theory, *J. ACM*, **4** (1957), 63—92.
36. Wang H., Toward mechanical mathematics, *IBM Journal of Research and Development*, 4 (1960), 2—22. (Русский перевод: см. [12] на стр. 179.)
37. Wang H., Proving theorems by pattern recognition, «I», *Comm. ACM*, **3** (1960), 220—234; «II», *Bell System Tech. J.*, **40** (1960), 1—41.
38. Wang H., Mechanical mathematics and inferential analysis, Computer programming and Formal Systems, P. Braffort and D. Hirschberg (eds.), North-Holland Publishing Co., Amsterdam, 1963, pp. 1—20 (Proceedings of the seminars at Blaricum, Holland 1961.)
39. Wang H., The mechanization of mathematical arguments, Proceedings of the Symposium in Applied Mathematics, v. 15, 1963, pp. 31—40.
40. Wos L., Carson D., Robinson G., The unit preference strategy in theorem proving, Proceedings Fall Joint Computer Conference, 1964, pp. 615—621; IBM 704, Argonne National Laboratory Report ANL-6447, 1961, 46 pp.

Машинно-ориентированная логика, основанная на принципе резолюции[1])

Дж. А. Робинсон

Исследуется возможность увеличения эффективности и расширения области практической применимости машинного доказательства теорем, использующего процедуры, базирующиеся на основополагающей теореме Эрбрана для исчисления предикатов первого порядка. Тщательное исследование процесса подстановки (термов вместо переменных) и процесса проверки результатов таких подстановок на тавтологичность показывает, что оба процесса могут быть соединены в один новый процесс (называемый *резолюцией*), итерирование которого гораздо более эффективно, чем прежняя циклическая процедура, состоящая в чередовании шагов подстановки и шагов проверки на тавтологичность.

Теория резолюционного процесса представлена в виде системы логики первого порядка, состоящей из единственного правила вывода (принципа резолюции). Доказана полнота этой системы; простейшей процедурой поиска доказательства, основанной на этой системе, является прямое использование доказательства полноты. Однако эта процедура совершенно неэффективна, и работа завершается обсуждением нескольких принципов (называемых принципами поиска), которые применимы для построения эффективных процедур поиска доказательства, использующих резолюцию как основной логический процесс.

1. ВВЕДЕНИЕ

В настоящей работе приведена формулировка логики первого порядка, специально предназначенная для использования в качестве теоретического аппарата машинных программ доказательства теорем. Предшествующие программы доказательства теорем были основаны на системах логики первого порядка, которые первоначально предназначались для других целей. Характерной чертой этих логических систем, которая отсутствует у описанной в этой работе системы, является относительная *простота* их правил вывода.

По традиции в силу прагматических и психологических при-

---

[1]) Robinson J. A., A machine-oriented logic based on the resolution principle, *Journal of the Association for Computing Machinery*, 12, № 1 (1965), 23–41.

чин обычно требуют, чтобы каждый отдельный шаг в дедукции был, вообще говоря, прост настолько, чтобы человек понял его правильность за один интеллектуальный акт. Эта традиция, вне сомнения, связана с желанием, чтобы каждый отдельный шаг был неоспоримым, даже если вся дедукция состоит из длинной цепи таких шагов. Окончательное заключение дедукции, если она правильна, логически следует из посылок, использованных в дедукции, но человеческий ум вполне может найти этот скачок от посылок к заключению удивительным, а потому (психологически) сомнительным. Отчасти поэтому логический анализ дедуктивных рассуждений состоит в сведении сложных выводов, которые не могут быть охвачены человеческим умом за один шаг, к цепочкам более простых выводов, каждый из которых может быть постигнут человеческим умом за один шаг.

Однако с теоретической точки зрения от правила вывода требуется только, чтобы оно было *корректным* (т. е. позволяло выводить лишь логические следствия посылок) и *эффективным* (т. е. должно быть алгорифмически проверяемо, является ли предполагаемое применение этого правила вывода в действительности его применением). Когда правило вывода применяет современная вычислительная машина, традиционное ограничение на сложность правил вывода более не является разумным. Становятся доступными более мощные правила, требующие, быть может, много больше комбинаторной обработки информации для одного применения.

В системе, описанной в этой работе, используется одно такое правило. Оно называется *принципом резолюции*, и в свете предыдущего замечания оно скорее ориентировано на машину, чем на человека. Принцип резолюции является очень мощным как в психологическом смысле, ибо он допускает отдельные шаги, которые человеческий ум часто может постичь лишь путем рассуждения, так и в теоретическом смысле, ибо он один в качестве единственного правила вывода образует полную систему логики первого порядка. Хотя второе свойство не очень важно, интересно отметить, что (насколько известно автору) никакая другая полная система логики первого порядка не состоит из единственного правила вывода, если считать правилами вывода (беспосылочными) логические аксиомы, заданные явно или в виде схемы.

Основное преимущество принципа резолюции состоит в том, что он позволяет избежать одного из основных комбинаторных препятствий к достижению эффективности, которое погубило предыдущие процедуры доказательства теорем.

В следующем разделе объясняются синтаксис и семантика того формализма, который используется в настоящей работе.

2. ПРЕДВАРИТЕЛЬНЫЕ ФОРМАЛЬНОСТИ

Формализм, используемый в настоящей работе, основан скорее на понятиях невыполнимости и опровержения, чем на понятиях общезначимости и доказательства. Известно (ср. [2] и [5]), что задачу о проверке произвольного множества высказываний на выполнимость можно свести к проверке на выполнимость конечного множества S, состоящего из высказываний в предваренной форме без кванторов существования в префиксе [1]); кроме того, можно предполагать, что матрица каждого высказывания из S есть дизъюнкция формул, каждая из которых или является атомарной формулой, или отрицанием атомарной формулы. Поэтому наш синтаксис устроен таким образом, что естественной синтаксической единицей является конечное множество S высказываний, имеющих эту специальную форму. Кванторный префикс в каждом высказывании опускается, поскольку он состоит только из кванторов общности, связывающих все переменные этого высказывания; далее, матрица каждого высказывания рассматривается просто как множество ее дизъюнктивных членов, поскольку порядок и кратность членов дизъюнкции несущественны.

В соответствии с этим мы даем следующие определения (отчасти следуя терминологии работ [2] и [5]).

2.1. *Переменные.* Следующие символы являются переменными:

$$u,\ v,\ w,\ x,\ y,\ z,\ u_1,\ v_1,\ w_1,\ x_1,\ y_1,\ z_1,\ u_2,\ \ldots \text{ и т. д.};$$

порядок, в котором они перечислены, называется алфавитным.

2.2. *Символы функций.* Следующие символы являются символами n-местных функций для каждого $n \geqslant 0$:

$$a^n,\ b^n,\ c^n,\ d^n,\ e^n,\ f^n,\ g^n,\ h^n,\ k^n,\ a_1^n,\ b_1^n,\ \ldots \text{ и т. д.};$$

порядок, в котором они перечислены, называется алфавитным. В случае $n = 0$ верхний индекс может быть опущен. Символы 0-местных функций являются *индивидными константами.*

2.3. *Символы предикатов.* Следующие символы являются символами n-местных предикатов для каждого $n \geqslant 0$:

$$P^n,\ Q^n,\ R^n,\ P_1^n,\ Q_1^n,\ R_1^n,\ P_2^n,\ \ldots \text{ и т. д.};$$

порядок, в котором они перечислены, называется алфавитным. Верхний индекс может быть опущен при $n = 0$.

2.4. *Символ отрицания.* Следующий символ является символом отрицания: $\neg$.

---

[1]) См. статью М. Девиса в настоящем сборнике, стр. 160. — *Прим. ред.*

2.5. *Алфавитный порядок символов.* Множество всех символов вполне упорядочено в алфавитном порядке посредством добавления к предыдущим соглашениям о порядке следующего правила: переменные предшествуют символам функций, символы функций предшествуют символам предикатов, символы предикатов предшествуют символу отрицания, символы функций меньшей размерности предшествуют символам функций большей размерности, а символы предикатов меньшей размерности предшествуют символам предикатов большей размерности.

2.6. *Термы.* Переменная является термом; последовательность символов, состоящая из символа n-местной функции ($n \geqslant 0$), за которым следует n термов, является термом.

2.7. *Атомарные формулы.* Последовательность символов, состоящая из символа n-местного предиката ($n \geqslant 0$), за которым следует n термов, является атомарной формулой.

2.8. *Литеры.* Атомарная формула является литерой; если A — атомарная формула, то $\neg A$ является литерой.

2.9. *Дополнения.* Если A — атомарная формула, то две литеры A и $\neg A$ называются дополнениями друг друга и образуют в любом порядке контрарную (complementary) пару.

2.10. *Дизъюнкты.* Конечное (быть может, пустое) множество литер называется дизъюнктом (clause). Пустой дизъюнкт обозначается через $\square$.

2.11. *Фундаментальные литеры.* Литера, не содержащая переменных, называется фундаментальной литерой.

2.12. *Фундаментальные дизъюнкты.* Дизъюнкт, все члены которого являются фундаментальными литерами, называется фундаментальным дизъюнктом. В частности, $\square$ является фундаментальным дизъюнктом.

2.13. *Правильно построенные выражения.* Термы и литеры являются (единственными) правильно построенными выражениями.

2.14. *Лексикографический порядок правильно построенных выражений.* Множество всех правильно построенных выражений вполне упорядочено в лексикографическом порядке посредством следующего правила: A предшествует B только в том случае, когда A короче B, или A и B имеют равную длину, но на первом месте, где A и B имеют разные символы, в A стоит более ранний, согласно алфавитному порядку.

При написании правильно построенных выражений с целью иллюстрации мы будем придерживаться более удобного для чтения способа, заключая n термов, следующих за символом n-местной функции или предиката, в круглые скобки и разделяя термы запятыми, если их более двух. В этом случае

мы можем опустить все верхние индексы, не вызывая недоразумений. При записи конечных множеств мы придерживаемся обычного соглашения о заключении членов в фигурные скобки и разделении членов запятыми, помня, что порядок записи членов безразличен.

2.15. *Эрбрановский универсум.* Любому множеству S дизъюнктов следующим образом сопоставляется множество фундаментальных термов, называемое эрбрановским универсумом для S: пусть F — множество всех символов функций, входящих в S. Если F содержит какой-либо символ 0-местной функции, то функциональный словарь для S совпадает с F; в противном случае он есть множество $\{a\} \cup F$. Эрбрановский универсум для S состоит из всех фундаментальных термов, в которых встречаются лишь символы из функционального словаря для S.

2.16. *Насыщение.* Если S — множество дизъюнктов, а P — множество термов, то через $P(S)$ мы обозначаем насыщение S над P, которое является множеством всех фундаментальных дизъюнктов, получаемых из членов S такими заменами переменных членами P, при которых все вхождения одной и той же переменной в какой-либо дизъюнкт заменяются на вхождение одного и того же терма.

2.17. *Модели.* Множество фундаментальных литер, не содержащее контрарных пар, называется моделью. Пусть M — модель, а S — множество фундаментальных дизъюнктов; тогда M называется моделью для S, если каждый дизъюнкт из S содержит некоторый член M. В общем случае, если S — произвольное множество дизъюнктов, а H — эрбрановский универсум для S, мы говорим, что M есть модель для S тогда и только тогда, когда M есть модель для $H(S)$.

2.18. *Выполнимость.* Множество S дизъюнктов является выполнимым, если существует модель для S; в противном случае S невыполнимо.

Из определения выполнимости ясно, что любое множество дизъюнктов, содержащее $\square$, невыполнимо и что пустое множество дизъюнктов выполнимо. Эти два обстоятельства покажутся вполне естественными в ходе развития нашей системы. Ясно также, что в соответствии с нашими семантическими определениями каждый непустой дизъюнкт интерпретируется (согласно объяснению в неформальных замечаниях в начале этого раздела) как замыкание всеобщности дизъюнкции литер, из которых этот дизъюнкт состоит.

2.19. *Фундаментальные резольвенты.* Если C и D — два фундаментальных дизъюнкта и $L \subseteq C$, $M \subseteq D$ — два одночленных множества, элементы которых образуют контрарную пару,

то фундаментальный дизъюнкт $(C-L)\cup(D-M)$ называется фундаментальной резольвентой C и D.

Ясно, что любая модель для $\{C, D\}$ является также моделью для $\{C, D, R\}$, где R — фундаментальная резольвента C и D. Не все пары фундаментальных дизъюнктов имеют фундаментальные резольвенты, а некоторые имеют более одной; однако в любом случае, как явствует из определения, два фундаментальных дизъюнкта имеют конечное число фундаментальных резольвент.

2.20. *Фундаментальная резолюция.* Если S — множество фундаментальных дизъюнктов, то фундаментальной резолюцией S, обозначаемой через $\mathscr{R}(S)$, является множество фундаментальных дизъюнктов, состоящее из всех элементов S и всех фундаментальных резольвент всех пар элементов S.

2.21. *N-я фундаментальная резолюция.* Если S — множество фундаментальных дизъюнктов, то n-я фундаментальная резолюция S, обозначаемая через $\mathscr{R}^n(S)$, определяется для каждого $n\geqslant 0$ следующим образом: $\mathscr{R}^0(S)=S$; для каждого $n\geqslant 0$ $\mathscr{R}^{n+1}(S)=\mathscr{R}(\mathscr{R}^n(S))$.

Этим завершается первая часть определений. Следующий раздел посвящен различным формам, которые принимает теорема Эрбрана в нашей системе. Для каждой такой формы имеется некоторый тип процедуры опровержения, которую эта форма подсказывает и оправдывает. Основная формулировка такова (ср. [2, 4]):

Теорема Эрбрана. *Если S — произвольное конечное множество дизъюнктов, H — его эрбрановский универсум, то S невыполнимо тогда и только тогда, когда некоторое конечное подмножество эрбрановского универсума $H(S)$ является невыполнимым.*

3. ПРОЦЕДУРА НАСЫЩЕНИЯ

Как отмечено в работе [5], теорему Эрбрана можно выразить в следующей форме.

Теорема 1. *Если S — произвольное конечное множество дизъюнктов, то S невыполнимо тогда и только тогда, когда $P(S)$ невыполнимо для некоторого конечного подмножества P эрбрановского универсума для S.*

Эта формулировка теоремы Эрбрана подсказывает следующий тип процедуры опровержения, которую мы назовем *процедурой насыщения*: по данному конечному множеству S дизъюнктов выбирается последовательность P_0, P_1, P_2, ... конечных подмножеств эрбрановского универсума H для S, та-

кая, что $P_j \subseteq P_{j+1}$ для каждого $j \geqslant 0$ и $\bigcup_{j=0}^{\infty} P_j = H$. Затем поочередно проверяются на выполнимость множества $P_0(S)$, $P_1(S)$, $P_2(S)$, ... Ясно, что каково бы ни было конечное подмножество P множества H, имеем $P \subseteq P_j$ для некоторого j и, следовательно, $P(S) \subseteq P_j(S)$. Следовательно, по теореме 1, если S невыполнимо, то для некоторого j подмножество $P_j(S)$ невыполнимо.

Конечно, любая конкретная процедура такого типа должна осуществлять выбор P_0, P_1, P_2, ... одинаково для всех конечных множеств дизъюнктов. Самым естественным методом такого выбора является использование так называемых уровней H_0, H_1, H_2, ... эрбрановского универсума H, где H_0 состоит из всех индивидных констант эрбрановского универсума H, а H_{n+1} для $n \geqslant 0$ состоит из всех термов эрбрановского универсума H, которые входят в H_n или чьи аргументы входят в H_n. В работе [5] мы назвали процедуру, использующую этот метод, *процедурой насыщения уровня*. Там же было отмечено, что процедуры Гильмора [4] и Девиса — Патнэма [2] по существу являются процедурами насыщения уровня.

Основным комбинаторным препятствием для достижения эффективности процедур насыщения уровня является огромная скорость роста конечных множеств H_j и $H_j(S)$ с ростом j по крайней мере для большинства интересных множеств S. Эта скорость роста до некоторой степени проанализирована в [5], там же приведены примеры совсем простых невыполнимых S, для которых первое невыполнимое $H_j(S)$ столь велико, что лежит далеко за пределами достижимости.

Интересное эвристическое замечание состоит в том, что для каждого конечного множества S дизъюнктов, которое невыполнимо и для которого фактически можно построить опровержение, существует по крайней мере одно конечное подмножество эрбрановского универсума для S, имеющее приемлемые размеры, такое, что $P(S)$ невыполнимо, причем P *минимально* (в том смысле, что $Q(S)$ выполнимо для любого собственного подмножества Q множества P). Такое P было в [5] названо *доказательным множеством* для S. Если бы в этих условиях нашелся всезнающий и благосклонный демон, который мог бы снабжать нас в приемлемое время доказательным множеством P для каждого невыполнимого конечного множества S дизъюнктов, которое мы рассматриваем, то мы могли бы просто начать насыщать S над P и потом извлечь подходящее опровержение множества S из окончательного конечного невыполнимого множества $P(S)$ фундаментальных дизъюнктов.

Такова была в сущности основная схема машинной программы, описанной в [5], где роль такого демона играет в меру своей изобретательности математик, использующий эту программу. Однако в действительности хотелось бы моделировать на машине самого „демона доказательства"; однако об этом, казалось интуитивно, не может идти и речи.

Тем не менее ситуация здесь не такая уж безнадежная. В частности, метод, развитый в остальных разделах этой статьи, по-видимому, достаточно хорошо моделирует требуемого демона посредством вычислительного процесса. В четвертом разделе мы сделаем первый основной шаг к развитию этого метода, доказав новые формы теоремы Эрбрана. Мы также дадим предварительный неформальный обзор остальных идей, откладывая строгое рассмотрение до следующих разделов.

4. ТЕОРЕМЫ О РЕЗОЛЮЦИИ И ОСНОВНАЯ ЛЕММА

Как специальный метод проверки конечных множеств фундаментальных дизъюнктов на выполнимость, метод Девиса — Патнэма [4] вряд ли можно улучшить с точки зрения эффективности. Тем не менее здесь мы приведем другой метод, гораздо менее эффективный, который играет только теоретическую роль в нашем построении и который формулируется много проще: по данному конечному множеству S фундаментальных дизъюнктов строятся последовательно множества S, $\mathscr{R}(S)$, $\mathscr{R}^2(S)$, ..., пока какое-нибудь $\mathscr{R}^n(S)$ будет содержать $\square$ или не будет содержать $\square$, но будет совпадать с $\mathscr{R}^{n+1}(S)$. В первом случае S невыполнимо, во втором случае S выполнимо. Рано или поздно одно из этих условий конца должно выполниться, поскольку число различных дизъюнктов, построенных из конечного множества тех литер, которые входят в S, конечно и, следовательно, в бесконечной последовательности вложенных множеств

4.1. $S \subseteq \mathscr{R}(S) \subseteq \mathscr{R}^2(S) \subseteq \ldots \subseteq \mathscr{R}^n(S) \subseteq \ldots$ не все включения являются собственными, так как резолюция не вводит новых литер.

Ввиду того что описанный процесс рано или поздно оборвется, мы можем доказать его корректность в смысле, сформулированном выше, в виде теоремы о фундаментальной резолюции.

Теорема о фундаментальной резолюции. *Если S есть произвольное конечное множество фундаментальных дизъюнктов, то S невыполнимо тогда и только тогда, когда $\mathscr{R}^n(S)$ для некоторого $n \geqslant 0$ содержит* $\square$.

Доказательство. Часть „тогда“ очевидна. Чтобы доказать часть „только тогда“, предположим, что T — заключительное множество в последовательности (4.1), так что T замкнуто относительно фундаментальной резолюции. Нам надо лишь показать, что если T не содержит $\square$, то T выполнимо и, следовательно, S выполнимо, ибо $S \subseteq T$. Пусть $L_1, \ldots, L_k$ — все различные атомарные формулы, которые входят в T (быть может, со знаком $\neg$). Пусть M — модель, определяемая следующим образом: M_0 — пустое множество, и для каждого j, $0 < j \leqslant k$, множество M_j есть множество $M_{j-1} \cup \{L_j\}$, если никакой дизъюнкт из T не состоит лишь из дополнений литер из множества $M_{j-1} \cup \{L_j\}$; в противном случае M_j есть множество $M_{j-1} \cup \{\neg L_j\}$. Наконец, M есть M_k. Если T не содержит $\square$, то M есть модель для T. Действительно, в противном случае существует наименьшее j, $0 < j \leqslant k$, такое, что некоторый дизъюнкт (скажем, C) в T состоит целиком из дополнений литер множества M_j. Следовательно, по определению M_j есть $M_{j-1} \cup \{\neg L_j\}$. Поэтому ввиду минимальности j дизъюнкт C содержит L_j. Но поскольку M_j есть $M_{j-1} \cup \{\neg L_j\}$, существует некоторый дизъюнкт (скажем, D) в T, который состоит целиком из дополнений литер множества $M_{j-1} \cup \{L_j\}$. Поэтому ввиду минимальности j дизъюнкт D содержит $\neg L_j$. В таком случае дизъюнкт $B = (C - \{L_j\}) \cup (D - \{\neg L_j\})$ состоит целиком из дополнений литер множества M_{j-1}, если B не есть $\square$. Но B — это фундаментальная резольвента C и D, поэтому B входит в T и, следовательно, B отлично от $\square$. Но это противоречит минимальности j, и теорема доказана.

Теорема о фундаментальной резолюции позволяет сформулировать более специальную форму теоремы 1.

Теорема 2. *Если S — произвольное конечное множество дизъюнктов, то S невыполнимо тогда и только тогда, когда для некоторого конечного подмножества P эрбрановского универсума для S и некоторого $n \geqslant 0$ резолюция $\mathscr{R}^n(P(S))$ содержит* $\square$.

Теперь стало возможным изложить неформально основные шаги оставшейся части построения. Мы обобщим понятия фундаментальной резольвенты и фундаментальной резолюции до понятий резольвенты и резолюции соответственно, устранив условие фундаментальности дизъюнктов. Любые два дизъюнкта будут тогда иметь нуль, один или более дизъюнктов в качестве своих резольвент, но в любом случае лишь конечное множество. В специальном случае, когда C и D *являются* фундаментальными дизъюнктами, их резольвенты, если они существуют,

являются в точности их фундаментальными резольвентами, согласно вышеприведенному определению. Аналогично обозначения $\mathscr{R}(S)$, $\mathscr{R}^n(S)$ будут сохранены, но S будет произвольным множеством дизъюнктов. Тогда $\mathscr{R}(S)$ будет обозначать резолюцию S, которая будет множеством, состоящим из всех элементов S и всех резольвент всех пар элементов S. Если S окажется множеством фундаментальных дизъюнктов, то $\mathscr{R}(S)$ будет фундаментальной резолюцией S в смысле предыдущего определения.

Детали обобщения и формальные определения приведены в разд. 5. Тем не менее неформальное представление об общем понятии резолюции можно получить уже сейчас, до его точного определения, если просто допустить, что оно обладает следующим важным свойством (это будет показано ниже): *резолюция полукоммутативна с насыщением*. Более точно это свойство сформулировано в следующей основной лемме, которая будет доказана в разд. 5.

Лемма. *Если S — произвольное множество дизъюнктов, а P — произвольное подмножество эрбрановского универсума для S, то* $\mathscr{R}(P(S)) \subseteq P(\mathscr{R}(S))$.

В действительности, как будет показано ниже, любой фундаментальный дизъюнкт, который может быть получен в результате подстановки термов из P в пару дизъюнктов C и D из S и *последующего* образования фундаментальной резольвенты получившихся дизъюнктов, может также быть получен в результате подстановки термов из P в одну из конечного числа резольвент C и D.

Простым следствием основной леммы является полукоммутативность n-й резолюции и насыщения.

Следствие. *Если S — произвольное конечное множество дизъюнктов, а P — произвольное подмножество эрбрановского универсума для S, то* $\mathscr{R}^n(P(S)) \subseteq P(\mathscr{R}^n(S))$ *для всех* $n \geqslant 0$.

Доказательство проводится индукцией по n. Имеем $\mathscr{R}^0(P(S)) = P(S) = P(\mathscr{R}^0(S))$, так что случай $n = 0$ тривиален. А если $\mathscr{R}^n(P(S)) \subseteq P(\mathscr{R}^n(S))$ для $n \geqslant 0$, то имеем

$$\begin{aligned}
\mathscr{R}^{n+1}(P(S)) &= \mathscr{R}(\mathscr{R}^n(P(S))) \subseteq && \text{по определению } \mathscr{R}^{n+1},\\
&\subseteq \mathscr{R}(P(\mathscr{R}^n(S))) \subseteq && \text{по предположению индукции, ибо } \mathscr{R} \text{ сохраняет включение,}\\
&\subseteq P(\mathscr{R}(\mathscr{R}^n(S))) = && \text{по лемме,}\\
&= P(\mathscr{R}^{n+1}(S)) && \text{по определению } \mathscr{R}^{n+1}
\end{aligned}$$

и следствие доказано.

Теперь с помощью вышеприведенного следствия мы можем непосредственно получить из теоремы 2 третью форму теоремы Эрбрана.

Теорема 3. *Если S — произвольное конечное множество дизъюнктов, то S невыполнимо тогда и только тогда, когда для некоторого конечного подмножества P эрбрановского универсума для S и некоторого $n \geqslant 0$ подмножество $P(\mathscr{R}^n(S))$ содержит* $\square$.

Здесь изменен порядок операций насыщения и n-й резолюции. Теперь стало возможным неожиданное упрощение теоремы 3 ввиду замечания, что подстановка термов вместо переменных не может преобразовать непустой дизъюнкт в $\square$. Следовательно, $P(\mathscr{R}^n(S))$ будет содержать $\square$ тогда и только тогда, когда $\mathscr{R}^n(S)$ содержит $\square$. Из теоремы 3, следовательно, мы немедленно получаем нашу окончательную форму теоремы Эрбрана, которая является основным результатом этой работы.

Теорема о резолюции. *Если S — произвольное конечное множество дизъюнктов, то S невыполнимо тогда и только тогда, когда $\mathscr{R}^n(S)$ содержит $\square$ для некоторого $n \geqslant 0$.*

Утверждение теоремы о резолюции есть в точности утверждение теоремы о фундаментальной резолюции без какого-либо упоминания о фундаментальности. Следовательно, за исключением несколько более сложного способа вычисления резольвент двух дизъюнктов (описанного в разд. 5), подсказываемый теоремой о резолюции метод проверки конечного множества S дизъюнктов на выполнимость в точности таков, как приведенный ранее метод для случая, когда S есть множество фундаментальных дизъюнктов; в действительности он автоматически сводится к этому методу, будучи применен к конечному множеству фундаментальных дизъюнктов. Однако теперь, вообще говоря, неверно, что последовательность вложенных множеств

$$S \subseteq \mathscr{R}(S) \subseteq \mathscr{R}^2(S) \subseteq \ldots \subseteq \mathscr{R}^n(S) \subseteq \ldots$$

должна оборваться для каждого конечного S. По теореме Чёрча это не может быть так, ибо иначе мы имели бы разрешающую процедуру для нашей формулировки логики первого порядка.

Перейдем теперь к рассмотрению „демона доказательства“, о котором шла речь в разд. 3. Мы предположили там, что если бы нам было дано доказательное множество P для невыполнимого множества S дизъюнктов, то все, что нам надо было бы делать, — это вычислять резолюции до тех пор, пока мы не встретим первую резолюцию $\mathscr{R}^n(P(S))$, которая содержит $\square$, и извлечь из нее формальное опровержение S. Но теорема о резолюции гарантирует нам, что к тому времени, когда мы вычислили бы $\mathscr{R}^n(S)$, если не раньше, мы встретили бы $\square$,

несмотря на наше незнание P. В этом смысле теорема о резолюции делает лишней роль демона доказательства.

В пятом разделе мы введем немного более формальный аппарат посредством второй группы определений и выполним наш долг по определению общего понятия резолюции и доказательству основной леммы.

5. ПОДСТАНОВКА, УНИФИКАЦИЯ И РЕЗОЛЮЦИЯ

Следующие определения касаются операции конкретизации (т. е. подстановки термов вместо переменных в правильно построенных выражениях и множествах правильно построенных выражений) и различных вспомогательных понятий, необходимых для определения общего понятия резолюции.

5.1. *Подстановочные компоненты.* Подстановочная компонента — это выражение вида T/V, где V — переменная, а T — терм, отличный от V, причем V называется *переменной компоненты* T/V, а T — *термом компоненты* T/V.

5.2. *Подстановки.* Подстановка — это конечное (быть может, пустое) множество подстановочных компонент с попарно различными переменными. Если P — произвольное множество термов, а термы всех компонент подстановки θ входят в P, то говорят, что θ есть подстановка над P. Мы записываем подстановку с компонентами $T_1/V_1, \ldots, T_k/V_k$ в виде $\{T_1/V_1, \ldots, T_k/V_k\}$, помня, что порядок компонент безразличен. Строчные греческие буквы используются для обозначения подстановок. В частности, ε есть *пустая подстановка*.

5.3. *Конкретизация.* Если E — произвольная конечная цепочка символов, а

$$\theta = \{T_1/V_1, \ldots, T_k/V_k\}$$

— подстановка, то конкретизацией E посредством θ называется операция, состоящая в замене всех вхождений в E переменной V_i, $1 \leqslant i \leqslant k$, на вхождение терма T_i. Получающаяся цепочка символов, обозначаемая через $E\theta$, называется θ-примером E. Иными словами, если E есть цепочка $E_0V_{i_1}E_1 \ldots V_{i_n}E_n$, то $E\theta$ есть цепочка $E_0T_{i_1}E_1 \ldots T_{i_n}E_n$. Здесь ни одна из подцепочек E_j цепочки E не содержит вхождений переменных $V_1, \ldots, V_k$, некоторые из E_j могут быть пустыми; n, возможно, равно 0, а каждое V_{i_j} является вхождением одной из переменных $V_1, \ldots, V_k$. Всякая цепочка вида $E\theta$ называется примером цепочки E. Если C — произвольное множество цепочек, а θ — подстановка, то θ-примером C является множество всех цепочек вида $E\theta$, где E входит в C. Мы будем обозначать это множество через $C\theta$ и говорить, что это есть пример C.

5.4. *Стандартизация.* Если C — конечное множество цепочек, а $V_1, \ldots, V_k$ — все попарно различные переменные, входящие в цепочки из C, расположенные в алфавитном порядке, то x-стандартизацией C, обозначаемой через ξ_C, называется подстановка, состоящая из всех подстановочных компонент вида x_j/V_j, где $1 \leqslant j \leqslant k$, $V_j \neq x_j$, а y-стаидартизацией C, обозначаемой через η_C, называется подстановка, состоящая из всех подстановочных компонент вида y_j/V_j, где $1 \leqslant j \leqslant k$, а $V_j \neq y_j$.

5.5. *Композиции подстановок.* Если $\theta = \{T_1/V_1, \ldots, T_k/V_k\}$ и λ — две произвольные подстановки, то множество $\theta' \cup \lambda'$ (где λ' — множество всех компонент λ, переменные которых не встречаются среди $V_1, \ldots, V_k$, а θ' — множество всех компонент $T_i\lambda/V_i$, $1 \leqslant i \leqslant k$, таких, что $T_i\lambda$ отлично от V_i) называется композицией θ и λ и обозначается через $\theta\lambda$.

Непосредственно проверяется, что $\varepsilon\theta = \theta\varepsilon = \theta$ для любой подстановки θ. Кроме того, композиция подстановок ассоциативна, $(\theta\lambda)\mu = \theta(\lambda\mu)$, так что мы можем опускать скобки при записи кратных композиций подстановок.

Сущность композиции операций подстановки состоит в том, что если E — произвольная цепочка, а $\sigma = \theta\lambda$, то цепочка $E\sigma$ есть в точности цепочка $E\theta\lambda$, т. е. λ-пример $E\theta$.

Эти свойства композиции подстановок устанавливаются с помощью следующих предложений.

5.5.1. $(E\sigma)\lambda = E(\sigma\lambda)$ *для всех цепочек* E *и всех подстановок* σ, λ.

Доказательство. Пусть $\sigma = \{T_1/V_1, \ldots, T_k/V_k\}$, $\lambda = \{U_1/W_1, \ldots, U_m/W_m\}$, а $E = E_0V_{i_1}E_1 \ldots V_{i_n}E_n$ в обозначениях пункта (5.3). Тогда по определению $E\sigma = E_0T_{i_1}E_1 \ldots T_{i_n}E_n$, а $(E\sigma)\lambda = \bar{E}_0\bar{T}_{i_1}\bar{E}_1 \ldots \bar{T}_{i_n}\bar{E}_n$, где каждое $\bar{T}_{i_j}$ есть $T_{i_j}\lambda$, а каждое $\bar{E}_j$ есть $E_j\lambda'$, где λ' есть множество всех тех компонент λ, переменные которых не встречаются среди $V_1, \ldots, V_k$ (поскольку ни одна из этих переменных не входит ни в одну из этих E_j). Но $\sigma\lambda = \sigma' \cup \lambda'$, где каждая компонента σ' есть в точности $\bar{T}_i/V_i$, если только $\bar{T}_i$ отлично от V_i. Следовательно, $E(\sigma\lambda) = \bar{E}_0\bar{T}_{i_1}\bar{E}_1 \ldots \bar{T}_{i_n}\bar{E}_n$.

5.5.2. *Для любых подстановок* σ, λ *справедливо следующее: если для каждой цепочки* E *имеет место равенство* $E\sigma = E\lambda$, *то* $\sigma = \lambda$.

Доказательство. Пусть $V_1, \ldots, V_k$ включает переменные всех компонент σ и λ; тогда $V_j\sigma = V_j\lambda$ для $1 \leqslant j \leqslant k$. В таком случае σ и λ состоят из одних и тех же компонент.

5.5.3. *Для любых подстановок* σ, λ, μ *справедливо* $(\sigma\lambda)\mu = \sigma(\lambda\mu)$.

Доказательство. Пусть E — произвольная цепочка. Тогда по 5.5.1 имеем

$$E((\sigma\lambda)\mu) = (E(\sigma\lambda))\mu = ((E\sigma)\lambda)\mu = (E\sigma)(\lambda\mu) = E(\sigma(\lambda\mu)).$$

Отсюда по 5.5.2 $(\sigma\lambda)\mu = \sigma(\lambda\mu)$.

У нас будет в дальнейшем случай использовать следующую дистрибутивность.

5.5.4. *Для любых множеств* A, B *цепочек и произвольной подстановки* λ *справедливо* $(A \cup B)\lambda = A\lambda \cup B\lambda$.

5.6. *Множества несогласованности.* Если A — некоторое множество правильно построенных выражений, то множеством несогласованности множества A называется множество всех тех правильно построенных подвыражений элементов множества A, которые начинаются с того места, где не все правильно построенные выражения из A имеют один и тот же символ.

Пример:

$$A = \{P(x, h(x, y), y),\ P(x, k(y), y),\ P(x, a, b)\}.$$

Множество несогласованности для A есть $\{h(x, y), k(y), a\}$.

Ясно, что если A содержит более одного элемента, то множество несогласованности множества A также содержит более одного элемента.

5.7. *Унификация.* Если A — множество правильно построенных выражений, а θ — подстановка, то говорят, что θ унифицирует A или что θ является унификатором A, если $A\theta$ есть одночленное множество. Множество правильно построенных выражений, имеющее унификатор, называется унифицируемым.

Ясно, что если θ унифицирует A, а A содержит более одного элемента, то θ унифицирует множество несогласованности множества A.

5.8. *Алгорифм унификации.* Следующий процесс, применимый к любому конечному непустому множеству A правильно построенных выражений, называется алгорифмом унификации.

Шаг 1. Положить $\sigma_0 = \varepsilon$ и $k = 0$ и перейти к шагу 2.

Шаг 2. Если $A\sigma_k$ не есть одноэлементное множество, то перейти к шагу 3. В противном случае положить $\sigma_A = \sigma_k$ и окончить работу.

Шаг 3. Пусть V_k — самый первый, а U_k — следующий за ним в лексикографическом порядке элементы множества несогласованности B_k для множества $A\sigma_k$. Если V_k — переменная, не входящая в U_k, то положить $\sigma_{k+1} = \sigma_k\{U_k/V_k\}$, увеличить k на 1 и возвратиться к шагу 2; в противном случае окончить работу.

Это определение нуждается в доказательстве того, что описанный процесс всегда завершается. Действительно, этот про-

цесс всегда заканчивается для любого конечного непустого множества правильно построенных выражений, в противном случае порождалась бы бесконечная последовательность A, $A\sigma_1$, $A\sigma_2$, ... конечных непустых множеств правильно построенных выражений, в которой каждый последующий член содержит на одну переменную меньше, чем предшествующий (а именно $A\sigma_k$ содержит V_k, а $A\sigma_{k+1}$ не содержит V_k). Но это невозможно, поскольку A содержит лишь конечное множество переменных.

5.9. *Наиболее общий унификатор.* Если A — конечное непустое множество правильно построенных выражений, для которого алгорифм унификации оканчивает работу на шаге 2, то построенная подстановка σ_A называется наиболее общим унификатором A и говорят, что A является общеунифицируемым множеством.

5.10. *Ключевые тройки.* Упорядоченная тройка $\langle L, M, N\rangle$ конечных множеств литер называется ключевой тройкой упорядоченной пары $\langle C, D\rangle$ дизъюнктов, если выполнены следующие условия:

5.10.1. Множества L и M не пусты, и $L \subseteq C$, $M \subseteq D$.

5.10.2. Множество N есть множество всех атомарных формул, которые входят (быть может, со знаком $\neg$) в множество $L\xi_C \cup M\eta_D$ (ср. определение (5.4)).

5.10.3. Множество N является общеунифицируемым множеством, σ_N обозначает его наиболее общий унификатор.

5.10.4. Множества $L\xi_C\sigma_N$ и $M\eta_D\sigma_N$ одночленны, и их элементы образуют контрарную пару.

Ясно, что пара $\langle C, D\rangle$ дизъюнктов имеет лишь конечное (быть может, пустое) множество ключевых троек.

5.11. *Резольвенты.* Резольвентой двух дизъюнктов C и D называется любой дизъюнкт вида $(C - L)\xi_C\sigma_N \cup (D - M)\eta_D\sigma_N$, где $\langle L, M, N\rangle$ есть ключевая тройка пары $\langle C, D\rangle$.

Ввиду замечания к определению 5.10 ясно, что два дизъюнкта C и D имеют конечное число резольвент или не имеют их вообще.

5.12. *Резолюции.* Если S — множество дизъюнктов, то резолюцией S, обозначаемой через $\mathcal{R}(S)$, называется множество дизъюнктов, которые являются или членами S, или резольвентами членов S.

5.13. *N-я резолюция.* Если S — множество дизъюнктов, то n-я резолюция S обозначается через $\mathcal{R}^n(S)$ и определяется для всех $n \geqslant 0$ аналогично (2.21).

На этом завершается наша вторая группа определений. Определение $\mathcal{R}(S)$ вполне подходит для наших теоретических целей, но при фактическом вычислении не следует включать в $\mathcal{R}(S)$ одновременно резольвенты пары $\langle C, D\rangle$ и пары $\langle D, C\rangle$,

поскольку они совпадают друг с другом с точностью до переименования переменных. Если C и D — фундаментальные дизъюнкты, то, как легко проверить, резольвенты пар $\langle C, D\rangle$ и $\langle D, C\rangle$ в точности совпадают друг с другом и с фундаментальными резольвентами C и D.

Теперь осталось лишь доказать основную лемму, что и будет сделано после того, как мы докажем следующую теорему, устанавливающую основное свойство унификации, которое потребуется нам для доказательства леммы и в дальнейшем в нашей теории.

Теорема об унификации. *Пусть A — произвольное конечное непустое множество правильно построенных выражений. Если A унифицируемо, то A общеунифицируемо; кроме того, если θ — произвольный, а σ_A — наиболее общий унификаторы A, то существует такая подстановка λ, что $\theta = \sigma_A\lambda$.*

Доказательство. Достаточно доказать, что в условиях теоремы алгорифм унификации заканчивает работу над A на шаге 2 и если алгорифм унификации еще не окончил работу и построена σ_k, $k \geqslant 0$, то равенство

5.14. $\theta = \sigma_k\lambda_k$

справедливо на шаге 2 для некоторой подстановки λ_k. Для $k = 0$ это верно при $\lambda_0 = \theta$, поскольку $\sigma_0 = \varepsilon$. Допустим, что при некотором $k \geqslant 0$ справедливо (5.14) для некоторой подстановки λ_k. Тогда если $A\sigma_k$ есть одночленное множество, то алгорифм унификации кончает работу на шаге 2, при этом $\sigma_A = \sigma_k$ — наиболее общий унификатор и $\lambda = \lambda_k$ есть требуемая подстановка; в противном случае алгорифм переходит к шагу 3. В последнем случае, поскольку λ_k унифицирует $A\sigma_k$ (в силу (5.14), ибо θ унифицирует A), то λ_k должно унифицировать множество несогласованности B_k множества $A\sigma_k$. Поэтому V_k и U_k, определенные на шаге 3 алгорифма унификации, удовлетворяют равенству

5.15. $V_k\lambda_k = U_k\lambda_k$.

Поскольку B_k есть множество несогласованности, то правильно построенные выражения из B_k не могут все начинаться с одного и того же символа; следовательно, не все они начинаются с символов, отличных от переменных, ибо B_k унифицируемо. Поэтому по крайней мере одно правильно построенное выражение в B_k начинается с переменной и потому является переменной, ибо оно правильно построено. Так как в лексикографическом порядке переменные предшествуют всем другим правильно построенным выражениям, а V_k является самым первым правильно построенным выражением в B_k, то V_k есть переменная. Если V_k входит в U_k, то $V_k\lambda_k$ входит в $U_k\lambda_k$, что невозможно, ибо V_k и U_k не тождественны друг другу и имеет

место равенство (5.15). Следовательно, V_k не входит в U_k. Поэтому алгорифм унификации не кончает работу на шаге 3, а возвращается к шагу 2, и при этом $\sigma_{k+1} = \sigma_k\{U_k/V_k\}$. Положим $\lambda_{k+1} = \lambda_k - \{V_k\lambda_k/V_k\}$. Тогда

$$\begin{aligned}
\lambda_k &= \{V_k\lambda_k/V_k\} \cup \lambda_{k+1} && \text{по определению } \lambda_{k+1},\\
&= \{U_k\lambda_k/V_k\} \cup \lambda_{k+1} && \text{по (5.15)},\\
&= \{U_k\lambda_{k+1}/V_k\} \cup \lambda_{k+1} && \text{ибо } V_k \text{ не входит в } U_k,\\
&= \{U_k/V_k\}\lambda_{k+1} && \text{по определению (5.5)}.
\end{aligned}$$

Отсюда по (5.14) $\theta = \sigma_{k+1}\lambda_{k+1}$. Таким образом, равенство (5.14) выполнено для всех σ_k, которые строит алгорифм унификации, и теорема доказана.

Теперь мы можем доказать основную лемму, которую для удобства сформулируем еще раз.

Лемма. *Если S — произвольное множество дизъюнктов, а P — произвольное подмножество эрбрановского универсума для S, то $\mathscr{R}(P(S)) \subseteq P(\mathscr{R}(S))$.*

Доказательство. Пусть $A \in \mathscr{R}(P(S))$. Тогда либо $A \in P(S)$ [в этом случае $A \in P(\mathscr{R}(S))$, ибо $S \subseteq \mathscr{R}(S)$], либо A есть фундаментальная резольвента двух фундаментальных дизъюнктов $C\alpha$, $D\beta$, где $C \in S$, $D \in S$, $\alpha = \{T_1/V_1, \ldots, T_k/V_k\}$, $\beta = \{U_1/W_1, \ldots, U_m/W_m\}$. Здесь $V_1, \ldots, V_k$ — полный список всех переменных, входящих в C и расположенных в алфавитном порядке, $W_1, \ldots, W_m$ — полный список всех переменных, входящих в D и расположенных в алфавитном порядке, а $T_1, \ldots, T_k$, $U_1, \ldots, U_m$ — элементы P. В этом случае $A = (C - L)\alpha \cup (D - M)\beta$, где $L \subseteq C$, $M \subseteq D$, L и M не пусты, а $L\alpha$ и $M\beta$ — одночленные множества, элементы которых образуют контрарную пару. Пусть

$$\theta = \{T_1/x_1, \ldots, T_k/x_k, U_1/y_1, \ldots, U_m/y_m\}.$$

Тогда $A = (C - L)\xi_C\theta \cup (D - M)\eta_D\theta$ и $L\xi_C\theta = L\alpha$, $M\eta_D\theta = M\beta$. Следовательно, θ унифицирует множество N тех атомарных формул, которые либо входят в множество $L\xi_C \cup M\eta_D$, либо являются дополнениями формул из этого множества. По теореме об унификации N имеет наиболее общий унификатор σ_N, и существует такая подстановка λ над P, что $\theta = \sigma_N\lambda$. Отсюда $L\xi_C\sigma_N\lambda = L\alpha$, $M\eta_D\sigma_N\lambda = M\beta$ и, следовательно, $L\xi_C\sigma_N$ и $M\eta_D\sigma_N$ суть одночленные множества, элементы которых являются дополнениями друг друга. Поэтому $\langle L, M, N\rangle$ является ключевой тройкой для $\langle C, D\rangle$, а дизъюнкт

$$B = \{C - L\}\xi_C\sigma_N \cup \{D - M\}\eta_D\sigma_N$$

является резольвентой C и D; следовательно, $B \in \mathscr{R}(S)$. Поскольку $\theta = \sigma_N\lambda$, то, согласно (5.5.4), $A = B\lambda$ и, следовательно, $A \in P(\mathscr{R}(S))$. Доказательство завершено.

Условия леммы не влекут обратного включения $P(\mathscr{R}(S)) \subseteq \mathscr{R}(P(S))$. Рассмотрим простой пример:

$$S = \{\{Q(x, f(y))\}, \{\neg Q(g(y), x)\}\}, \quad P = \{a\}.$$

Короткие выкладки показывают, что $P(\mathscr{R}(S))$ содержит $\square$ (ибо $\mathscr{R}(S)$ содержит $\square$), в то время как $\mathscr{R}(P(S))$ не содержит $\square$. Таким образом, S невыполнимо, но P не является доказательным множеством для S.

6. ПРИНЦИП РЕЗОЛЮЦИИ. ОПРОВЕРЖЕНИЯ

Единственным правилом вывода в нашей логической системе, упомянутой в разд. 1, является *принцип резолюции: из любых двух дизъюнктов C и D можно вывести резольвенту C и D.*

Под *опровержением* множества S дизъюнктов мы понимаем конечную последовательность дизъюнктов $B_1, \ldots, B_n$, такую, что: a) каждое B_i, $1 \leqslant i \leqslant n$, либо входит в S, либо является резольвентой двух предшествующих дизъюнктов и b) B_n есть $\square$.

Из теоремы о резолюции немедленно следует, что конечное множество S дизъюнктов невыполнимо тогда и только тогда, когда существует опровержение S. Таким образом, теорема о резолюции является теоремой о полноте нашей логической системы.

Два примера опровержения демонстрируют работу нашей системы.

Пример 1. Множество, состоящее из двух дизъюнктов C_1 и C_2, где

$$C_1 = \{Q(x, g(x), y, h(x, y), z, k(x, y, z))\},$$
$$C_2 = \{\neg Q(u, v, e(v), w, f(v, w), x)\},$$

имеет опровержение C_1, C_2, $\square$. Отметим, что $\langle C_1, C_2\rangle$ имеет ключевую тройку $\langle C_1, C_2, N\rangle$, где N — множество

$$\{Q(x_1, g(x_1), x_2, h(x_1, x_2), x_3, k(x_1, x_2, x_3)), Q(y_1, y_2, e(y_2), y_3, f(y_2, y_3), y_4)\}.$$

Читатель легко может проверить за несколько минут вычислений, согласно алгорифму унификации, что σ_N — подстановка со следующими компонентами:

$$y_1/x_1, \quad h(y_1, e(g(y_1)))/y_3,$$
$$g(y_1)/y_2, \quad f(g(y_1), h(y_1, e(g(y_1))))/x_3,$$
$$e(g(y_1))/x_2, \quad k(y_1, e(g(y_1)), f(g(y_1), h(y_1, e(g(y_1))))))/y_4,$$

а $C_1\xi_{C_1}\sigma_N$ и $C_2\eta_{C_2}\sigma_N$ являются одночленными множествами, элементы которых являются дополнениями друг друга.

Этот пример показывает, как доказательное множество автоматически стронтся в качестве побочного результата операции резолюции. Термы вышеприведенных подстановочных компонент становятся термами доказательного множества для $\{C_1, C_2\}$, если переменную y_1 всюду заменить на какой-либо терм эрбрановского универсума для $\{C_1, C_2\}$, например, на индивидную константу „a“. Интересно отметить, что наименьшим уровнем эрбрановского универсума, содержащим это доказательное миожество, является H_5, который имеет порядка 10^{64} членов. Следовательно, $H_5(\{C_1, C_2\})$ имеет порядка 10^{256} членов. Этот пример недоступен процедуре насыщения уровней.

Пример 2. Более интересный пример был рассмотрен в работе [5]. Он возник нз следующей алгебраической проблемы.

Доказать, что в любой ассоциативной системе, в которой имеются решения всех уравнений $x \cdot a = b$ *и* $a \cdot y = b$, *существует правый единичный элемент.*

Чтобы формализовать эту проблему в нашей логике, мы отрицаем предполагаемое заключение и пытаемся опровергнуть множество, состоящее из следующих дизъюнктов (здесь $Q(x, y, z)$ понимается как $x \cdot y = z$):

C_1: $\{\neg Q(x, y, u), \neg Q(y, z, v), \neg Q(x, v, w), Q(u, z, w)\}$,
C_2: $\{\neg Q(x, y, u), \neg Q(y, z, v), \neg Q(u, z, w), Q(x, v, w)\}$,
} ассоциативность

C_3: $\{Q(g(x, y), x, y)\}$,
C_4: $\{Q(x, h(x, y), y)\}$,
} существование левого и правого решений,

C_5: $\{Q(x, y, f(x, y))\}$, замкнутость относительно умножения,

C_6: $\{\neg Q(k(x), x, k(x))\}$ отсутствие правой единицы.

Добавляя следующие резольвенты, мы получаем опровержение:

C_7: $\{\neg Q(y_1, x_6, y_1), Q(y_2, x_6, y_2)\}$,
C_8: $\{\neg Q(y_1, y_2, y_1)\}$,
C_9: $\square$.

Комментарий. Резольвента C_7 — резольвента пары $\langle C_1, C_3\rangle$ с ключевой тройкой

$$\langle\{\neg Q(x, y, u), \neg Q(x, v, w)\}, \{Q(g(x, y), x, y)\}, N\rangle,$$

где N есть множество $\{Q(x_4, x_5, x_1), Q(x_4, x_2, x_3), Q(g(y_1, y_2), y_1, y_2)\}$.

Как легко проверить, σ_N, вычисляемая с помощью алгорифма унификации, такова:

$$\{y_2/x_1,\ y_1/x_2,\ y_2/x_3,\ g(y_1,\ y_2)/x_4,\ y_1/x_5\}.$$

Резольвента C_8 — единственная резольвента пары $\langle C_6, C_7\rangle$, а $\square$ — единственная резольвента пары $\langle C_4, C_8\rangle$.

Этот пример показывает, что отдельный шаг опровержения, выполняемый по правилу резолюции, может являться столь сложным, что человеческий ум не может убедиться в его правильности за один интеллектуальный акт. Работая крупноблочно, правило резолюции является очень компактным, если не сказать элегантным, способом рассуждений. Резольвенты C_2 и C_5 не были использованы в опровержении в качестве посылок, хотя это не имеет ничего общего с правилом резолюции. Поэтому экономным опровержением для этого примера является последовательность C_1, C_3, C_4, C_6, C_7, C_8, $\square$.

7. ПРОЦЕДУРЫ ОПРОВЕРЖЕНИЯ; ТАКТИКИ ПОИСКА

Целью вышеизложенного было лишь установление теоретической базы в виде специальной логической системы для построения различных программ доказательства теорем, т. е. в данном случае процедур опровержения. До сих пор не рассматривался вопрос о нахождении эффективной процедуры опровержения; в данном разделе мы кратко обсудим вопрос.

Если строить процедуру опровержения непосредственно по теореме о резолюции, то получится совсем неэффективная процедура. Эта процедура состоит в вычислении для данного конечного множества S дизъюнктов последовательности множеств S, $\mathscr{R}(S)$, $\mathscr{R}^2(S)$, ... до тех пор, пока не встретится $\mathscr{R}^n(S)$, такое, что оно или содержит $\square$, или не содержит $\square$, но совпадает с $\mathscr{R}^{n+1}(S)$. В первом случае опровержение S можно получить, проследив порождение $\square$; во втором случае S выполнимо. В силу теоремы Чёрча [1] мы знаем, что для некоторых множеств S эта процедура, как и вообще все корректные процедуры опровержения, не окончит работу ни тем, ни другим способом, а будет продолжать вычисления бесконечно.

В некоторых случаях мы можем предвидеть бесконечность процесса. Рассмотрим пример множества, состоящего из двух дизъюнктов

$$C_1\colon \{Q(a)\},\ C_2\colon \{\neg Q(x),\ Q(f(x))\}.$$

(Это формулировка в нашей логике фрагмента аксиом Пеано для натуральных чисел, если интерпретировать „$Q(x)$“ как „x есть натуральное число“, „a“ — как „0“, а „$f(x)$“ — как „число,

следующее за x".) Легко видеть, что описанная выше процедура будет порождать последовательно резольвенты

$$\{Q(f(a))\},\ \{Q(f(f(a)))\},\ \{Q(f(f(f(a))))\},\ \ldots$$

и т. д. *до бесконечности.*

Данный пример наводит на мысль о правиле, которое позволяло бы нам эффективно распознать этот специальный вид бесконечного процесса в случае его возникновения, так, чтобы мы могли включить это правило в процедуру опровержения, в результате чего она окончила бы работу над S и другими аналогичнымн примерами. Такое правило возможно, оно основано на понятии литеры, чистой в данном множестве S дизъюнктов, и мы назовем его *правилом чистоты.*

7.1. *Чистые литеры.* Если S — произвольное конечное множество дизъюнктов, C — дизъюнкт из S, а L — литера из C, причем не существует ключевой тройки $\langle\{L\}, M, N\rangle$ ни для какой пары $\langle C, D\rangle$ дизъюнктов, где D — дизъюнкт из $S-\{C\}$, то говорят, что литера L чиста в S.

Правило чистоты основано на следующей теореме.

Теорема о чистоте. *Если S — произвольное множество дизъюнктов, $L\in C\in S$, и литера L чиста в S, то S выполнимо тогда и только тогда, когда $S-\{C\}$ выполнимо.*

Доказательство. Если S выполнимо, то и $S-\{C\}$ выполнимо, ибо $S-\{C\}$ есть подмножество S. Если $S-\{C\}$ выполнимо, то существует модель A для $S-\{C\}$, каждая литера которой входит в некоторый дизъюнкт из $H(S)$, где H — эрбрановский универсум для S. Пусть N — множество всех фундаментальных литер вида $L\theta$, где θ — подстановка над H, и пусть K состоит из дополнений всех членов N. Тогда множество $P = N\cup(A-K)$ является моделью, более того, оно есть модель для S, поскольку каждый дизъюнкт в $H(\{C\})$ содержит член P (а именно какой-либо член N) и каждый дизъюнкт из $H(S-\{C\})$ содержит член P, а именно некоторый член $A-K$. Действительно, никакой дизъюнкт из $H(S-\{C\})$ не содержит членов K, ибо в противном случае если $D\beta$ был бы таким дизъюнктом (где $D\in(S-\{C\})$), то нашлось бы такое M, что $M\subseteq D$, а $M\beta$ — одноэлементное множество, содержащее член K. Тогда имелась бы некоторая подстановка α над H, такая, что $\{L\}\alpha$ и $M\beta$ — это одночленные множества, элементы которых являются дополнениями друг друга. Рассуждая так же, как в доказательстве основной леммы, мы нашли бы ключевую тройку $\langle\{L\}, M, N\rangle$ для пары $\langle C, D\rangle$, что противоречит чистоте литеры L в S. Теорема доказана.

Правило чистоты формулируется так: *можно вычеркивать из конечного множества S дизъюнктов любой дизъюнкт, содержащий литеру, чистую в S.*

Если S — рассмотренный выше небольшой фрагмент аксиоматики Пеано, т. е. множество, состоящее из двух дизъюнктов

$$C_1\colon \{Q(a)\},\ C_2\colon \{\neg Q(x),\ \underline{Q(f(x))}\},$$

то мы видим, что подчеркнутая литера чиста в S. Поэтому мы можем вычеркнуть C_2, получив множество $S-\{C_2\}$, единственным дизъюнктом которого является

$$C_1\colon \{\underline{Q(a)}\}.$$

Очевидно, что подчеркнутая литера чиста в $S-\{C_2\}$. Поэтому мы можем вычеркнуть C_1, получив множество $S-\{C_1\}-\{C_2\}$, которое пусто и поэтому выполнимо. Отсюда по теореме о чистоте S выполнимо.

Таким образом, процедура опровержения, объединяющая правило резолюции и правило чистоты, „сходится" для большего числа конечных множеств дизъюнктов, чем процедура, основанная лишь на одном правиле резолюции. Такие правила, как правило чистоты, мы называем *тактиками поиска*, чтобы отличать их от правил вывода.

Существует другая тактика поиска, которая хотя и не увеличивает область сходимости, но полезна для увеличения скорости сходимости процедуры опровержения. Мы называем эту тактику *тактикой поглощения.*

7.2. *Поглощение.* Если C и D — два различных непустых дизъюнкта, то мы говорим, что C поглощает D, если существует такая подстановка σ, что $C\sigma \subseteq D$.

Следующая теорема устанавливает основное свойство поглощения.

Теорема о поглощении. *Если S — произвольное конечное множество дизъюнктов, а D — дизъюнкт из S, который поглощается некоторым дизъюнктом из $S-\{D\}$, то S выполнимо тогда и только тогда, когда выполнимо $S-\{D\}$.*

Доказательство. Достаточно показать, что если M — модель для $S-\{D\}$, то M является также моделью для S. Пусть M — модель для $S-\{D\}$, и предположим, что $C\in(S-\{D\})$ поглощает D. Тогда существует такая подстановка σ, что $C\sigma\subseteq D$. Так как $D\in S$, то термы компонент σ построены из символов функций, входящих в функциональный словарь для S и, быть может, переменных. Следовательно, каждый фундаментальный пример $C\sigma$ над H является фундаментальным примером C над H

и потому содержит член M. Но каждый фундаментальный пример $D\lambda$ дизъюнкта D содержит фундаментальный пример $C\sigma\lambda$ дизъюнкта C и потому содержит член M. Таким образом, M является моделью для S и теорема доказана.

Тактика поглощения формулируется следующим образом: *можно вычеркивать из конечного множества S дизъюнктов любой дизъюнкт D, который поглощается некоторым дизъюнктом из $S-\{D\}$.*

Для того чтобы тактику поглощения можно было включить в процедуру опровержения, мы должны указать алгорифм распознавания того, что один дизъюнкт C поглощает другой дизъюнкт D. Таким алгорифмом является следующий алгорифм поглощения:

Шаг 1. Пусть $V_1, \ldots, V_m$ — полный список всех переменных, входящих в D и расположенных в алфавитном порядке. Пусть $I_1, \ldots, I_m$ — различные индивидные константы, ни одна из которых не входит ни в C, ни в D. Пусть $\theta = \{I_1/V_1, \ldots, I_m/V_m\}$. Вычислить $D\theta$ и перейти к шагу 2.

Шаг 2. Положить $A_0 = \{C\}$, $k = 0$ и перейти к шагу 3.

Шаг 3. Если A_k не содержит $\square$, то A_{k+1} будет множеством, состоящим из всех дизъюнктов вида $(K\sigma_N - M\sigma_N)$, где $K \in A_k$, $M \subseteq K$, $N = M \cup \{P\}$, $P \in D\theta$, σ_N — наиболее общий унификатор N; в противном случае прекратить работу.

Шаг 4. Если A_{k+1} непусто, то увеличить k на 1 и вернуться к шагу 2. В противном случае прекратить работу.

Ясно, что этот алгорифм всегда кончает работу, ибо каждый дизъюнкт в A_{k+1} меньше по крайней мере на одну литеру того дизъюнкта в A_k, из которого он был получен. Поскольку единственный дизъюнкт в A_0 содержит лишь конечное число литер, то в последовательности $A_0, A_1, \ldots$ рано или поздно встретится множество, которое либо содержит $\square$, либо пусто.

Следующее рассуждение показывает, что алгорифм поглощения заканчивает работу на шаге 3 тогда и только тогда, когда C поглощает D.

Если C поглощает D, то $C\sigma \subseteq D$ для некоторой σ. Следовательно, $C\sigma\theta \subseteq D\theta$. Таким образом, $C\mu \subseteq D\theta$ для некоторых μ. Допустим, что для $k \geqslant 0$ существует K, такое, что $K \in A_k$ и $K\mu \subseteq D\theta$ для некоторой μ. Если K непусто, то пусть P будет литерой из $K\mu$ и, следовательно, из $D\theta$. Тогда существует $M \subseteq K$, такое, что $N = M \cup \{P\}$ унифицируется μ. Построим с помощью алгорифма унификации наиболее общий унификатор σ_N множества N. Дизъюнкт $G = K\sigma_N - M\sigma_N$ входит в A_{k+1}. По теореме об унификации $\mu = \sigma_N\lambda$ для некоторой λ, следовательно, $K\sigma_N\lambda \subseteq D\theta$. Поэтому $G\lambda \subseteq D\theta$. Так как $C \in A_0$ непусто, то и каждое A_k, $k \geqslant 0$, либо непусто, либо содержит $\square$. Поэтому алгорифм поглощения оканчивает работу не на шаге 4, а на шаге 3.

Если алгорифм поглощения оканчивает работу над C и D на шаге 3, то существует конечная последовательность дизъюнктов $C_0, C_1, \ldots, C_{n+1}$, такая, что $C_0 = C$, $C_{n+1} = \square$ и для каждого j, $0 \leqslant j \leqslant n$, имеем $C_{j+1} = C_j\sigma_j - M_j\sigma_j$, где $M_j \subseteq C_j$, а σ_j — наиболее общий унификатор $M_j \cup \{P\}$, где $P \in D\theta$. Отсюда следует (поскольку $M_j\sigma_j$ не содержит переменных, $0 \leqslant j \leqslant n$), что

$$C_{n+1} = \square = C\sigma_0\sigma_1 \ldots \sigma_n - M_0\sigma_0 - M_1\sigma_1 - \ldots - M_n\sigma_n,$$

т. е. $C\sigma_0\sigma_1 \ldots \sigma_n \subseteq (M_0\sigma_0 \cup M_1\sigma_1 \cup \ldots \cup M_n\sigma_n) \subseteq D\theta$. Положим $\lambda = \sigma_0\sigma_1 \ldots \sigma_n$. Тогда $C\lambda \subseteq D\theta$. Пусть σ — подстановка, получающаяся из λ заменой в каждой компоненте J_i на V_i для $1 \leqslant i \leqslant m$. Тогда $C\sigma \subseteq D$.

Особенно полезным применением тактики поглощения является следующее. Предположим, что резольвента R дизъюнктов C и D поглощает один из них; тогда, добавив R по правилу резолюции, мы можем одновременно вычеркнуть, согласно тактике поглощения, тот из дизъюнктов C или D, который поглощается R. Эта объединенная операция сводится к замене C или D на R; мы назовем это правило вывода *правилом замещения*, а соответствующую тактику — *тактикой замещения*.

Следующий пример, который рассмотрели Гильмор [4], Девис и Патнэм [2] и Фридман [3], демонстрирует пользу этих тактик для увеличения скорости сходимости. Пусть множество S состоит из следующих дизъюнктов:

C_1: $\{\underset{(6)}{P(x_1, x_2)}\}$,

C_2: $\{\underset{(1)}{\neg P(y_2, f(y_1, y_2))}, \underset{(2)}{P(f(y_1, y_2), f(y_1, y_2))}, Q(y_1, y_2)\}$,

C_3: $\{\underset{(3)}{\neg P(y_2, f(y_1, y_2))}, \underset{(4)}{\neg P(f(y_1, y_2), f(y_1, y_2))}, \underset{(5)}{\neg Q(y_1, f(y_1, y_2))}, \neg Q(f(y_1, y_2), f(y_1, y_2))\}$.

Мы получим множество S', имеющее лишь два члена

$$C_4: \{Q(y_1, y_2)\},$$
$$C_5: \{\neg Q(f(y_1, y_2), f(y_1, y_2))\},$$

всего за шесть шагов, которые состоят в вычеркивании подчеркнутых литер и дизъюнктов в указанном порядке. Вычеркивания с (1) по (5) проводятся согласно правилу замещения; вычеркивание (6) целого дизъюнкта проводится согласно правилу чистоты. Мы немедленно получаем, что S' невыполнимо, ибо $\square$ — единственная резольвента C_4 и C_5.

Программа Гильмора на IBM 704 не окончила работу над этим примером за 21 *мин* счета. Более эффективная процедура Девиса и Патнэма закончила работу над этим примером за 30 *мин* ручного счета.

Применение одной из трех описанных тактик к конечному множеству S дизъюнктов дает множество S', которое либо содержит меньше дизъюнктов, либо содержит столько же дизъюнктов, но один или несколько из них являются более короткими. Поэтому обычный способ использовать эти тактики в процедуре опровержения состоит в том, чтобы добавлять новый дизъюнкт по правилу резолюции лишь тогда, когда ни одна из этих трех тактик неприменима. Мы можем назвать такие множества *несводимыми*. Если множество S выполнимо, то такая процедура может кончить работу либо на пустом множестве (как в случае аксиом Пеано), либо на непустом несводимом множестве, обладающем следующим свойством: любая резольвента любой пары дизъюнктов из этого множества поглощается одним из дизъюнктов этого множества.

Имеются и другие тактики поиска того же типа, но менее простые, чем рассмотренные. Предполагается, что эта работа послужит началом серии работ, в которых развитая здесь теоретическая база будет использована для дальнейшего изучения тактик поиска и построения процедур опровержения. Этот раздел был всего лишь наброском общей картины проблемы и кратким взглядом на возможные подходы к ней.

В заключение я хотел бы выразить свою благодарность моим коллегам д-ру Дж. А. Робинсону и д-ру Л. Т. Восу из Аргонской национальной лаборатории и проф. В. Давидону из колледжа Хаверфорда за обсуждение статьи и критические замечания по основным понятиям этой работы. Я благодарю также рецензентов из *ACM* и д-ра Т. Гальперина, чьи замечания по рукописи статьи во многом способствовали написанию окончательного варианта.

ЛИТЕРАТУРА

1. Church A., A note on the Entscheidungsproblem, *J. Symb. Logic*, 1 (1936), 40—41.
2. Davis M., Putnam H., A computing procedure for quantification theory, *J. ACM*, 7 (1960), 201—215.
3. Friedman J., A semi-decision procedure for the functional calculus, *J. ACM*, 10 (1963), 1—24.
4. Gilmore P. C., A proof method for quantification theory, *IBM J. Res. Develop.*, 4 (1960), 28—35.
5. Robinson J. A., Theorem proving on the computer, *J. ACM*, 10 (1963), 163—174.

Алгоритм моделирования динамических систем с помощью вычислительных машин, основанный на теории потоковых графов[1])

Р. Якубовски

1. ВВЕДЕНИЕ

Сравнительно недавно в литературе появилось много статей о применении аналоговой техники для моделирования динамических систем с помощью вычислительных машин [1—4]. Аналоговая техника воспроизводит ясно очерченную картину для исследуемой задачи. В частности, это имеет место при моделировании нелинейных динамических систем. В последнее время наблюдается также прогресс в применении потоковых графов в программировании для аналоговых вычислительных машин [5—7]. Учитывая обе эти тенденции, в настоящей статье мы хотим изложить метод, основанный на графах с многоаргументными функциональными ребрами, который позволял бы нам подготавливать данные для вычислительной машины; причем последняя в свою очередь могла бы использовать их в самоорганизующейся программе для решения системы дифференциальных уравнений, задаваемой графом. Из-за простоты получающегося алгоритма он может найти широкое применение в технических задачах и задачах обучения, связанных с моделированием динамических систем. Так как упомянутый алгоритм основан на теории потоковых графов, мы вначале определим элементы графа и сформулируем зависимости между этими элементами.

---

[1]) Jakubowski R., An algorithm for the simulation of dynamical systems by means of digital computers, based on signal-flow graph, *Journal of Mathematical Analysis and Applications*, **22** (1968), 172—187.

2. ГРАФЫ И ИХ ПРИМЕНЕНИЕ В ПРОГРАММИРОВАНИИ ДЛЯ ВЫЧИСЛИТЕЛЬНЫХ МАШИН

В гра́фе, который служит базой для реализации программы работы вычислительной машины, зависимые и независимые переменные изображаются вершинами. Умножение на постоянный коэффициент и другие математические операции изображаются ребрами (математические операции суммирования ассоциируются со стоками — выходами). Мы будем различать два типа ребер и вершин. Ребра разделяются на нефункциональные и функциональные, а вершины — на источники и стоки. Вершины, которые не являются ни источниками, ни стоками (эти вершины имеют как входящие, так и выходящие ребра), здесь не существуют. Операция, которая сопоставляется этим вершинам, изображается специальным функциональным ребром, описанным в разд. 2.3.

2.1. НЕФУНКЦИОНАЛЬНЫЕ РЕБРА

Нефункциональное ребро есть ориентированная линия с направлением (указанным стрелкой), связывающая две вершины. Это ребро характеризуется коэффициентом усиления, причем

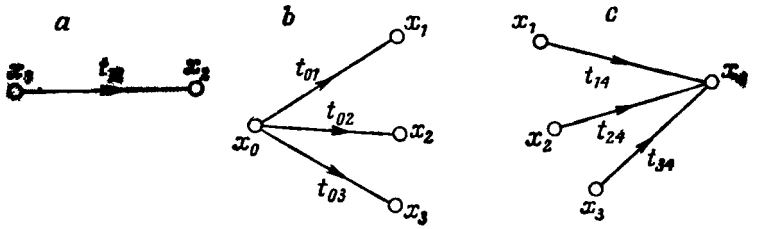


Рис. 1. Элементы нефункциональных ребер.

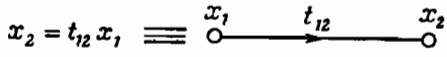
коэффициент является константой и обозначается символом, приписываемым ребру. На рис. 1, *a* показано нефункциональное ребро, связывающее вершины x_1 и x_2; коэффициент усиления этого ребра обозначается через t_{12}.

Если стрелка данного ребра направлена к вершине, то мы будем называть это ребро заходящим ребром по отношению к данной вершине и исходящим ребром в противном случае. Вершина, имеющая только исходящие ребра, называется источником (рис. 1, *b*), а вершина, имеющая только заходящие ребра, называется стоком (рис. 1, *c*). Список правил для нефункциональных ребер [6] приведен в табл. 1.

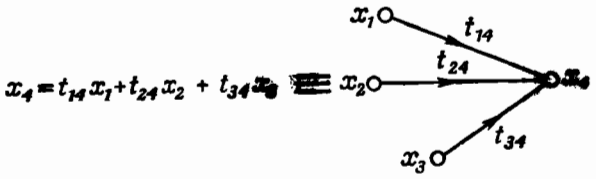
Таблица 1

Список правил для нефункциональных ребер

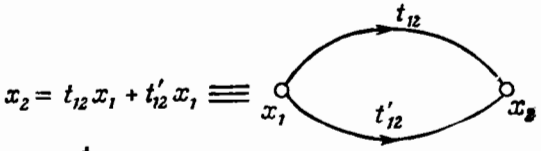
А. Правило соответствия для единственного нефункционального ребра:

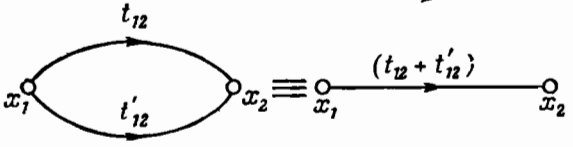


В. Правило соответствия для нескольких нефункциональных **ребер:**



С. Дополнительное правило:





2.2. ОДНОАРГУМЕНТНЫЕ ФУНКЦИОНАЛЬНЫЕ РЕБРА

В рассматриваемом здесь гра́фе можно различить два типа функциональных ребер: ребра с одним аргументом и ребра со многими аргумеитами. Рассмотрим первый тип. Одноаргументное фуикциональное ребро есть ориентированная линия (нарисованная толще, чтобы отличить от нефункциональных ребер), связывающая две вершины. Этому ребру всегда приписана определеииая операция. Пара вершин x_1 и x_2 и связывающее их ребро, которому приписана операция P, изображает следующее соотношение (рис. 2, *а*):

$$x_2 = P(x_1). \qquad (1)$$

Вершины, связанные функциональным ребром, тем самым определяются единственным образом; вершина x_1, изображаю-

щая независимую переменную, становится источником, а вершина x_2, изображающая зависимую переменную, — стоком.

Рассмотрим граф, имеющий два функциональных ребра (рис. 2, *b*). Согласно правилу (А) из табл. 1, для этого графа имеем

$$x_3 = x_2.$$

Из определения функционального ребра следует:

$$x_2 = P(x_1),\ x_4 = N(x_3).$$

Таким образом, в результате получаем:

$$x_4 = N(P(x_1)).$$

Так как $x_2 = x_3$, мы можем изобразить граф на рис. 2, *b* в упрощенной форме, показанной на рис. 2, *c*. В соответствии

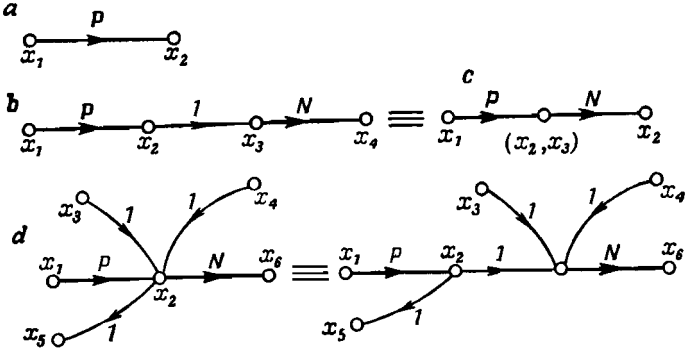


Рис. 2. Связи между однозначными функциональными ребрами.

с принятым правилом вершина (x_2, x_3) в графе на рис. 2, *c* есть сток для ребра N и источник для ребра P. Таким образом, для графа на рис. 2, *d* мы имеем:

$$x_5 = P(x_1),\ x_6 = N(P(x_1) + x_3 + x_4).$$

2.3. МНОГОАРГУМЕНТНЫЕ ФУНКЦИОНАЛЬНЫЕ РЕБРА

Многоаргументное функциональное ребро есть ориентированная геометрическая структура — „куст" (рис. 3, *a*), связывающая n вершин $x_1, x_2, \ldots, x_{n-1}, x_n$. Одна из них, x_n, изображает зависимую переменную. Этому ребру всегда приписана определенная операция P. Множество вершин $x_1, x_2, \ldots, x_n$ и ребро, связывающее эти вершины, изображает следующее отношение:

$$x_n = P(x_1, x_2, x_3, \ldots, x_{n-1}). \qquad (2)$$

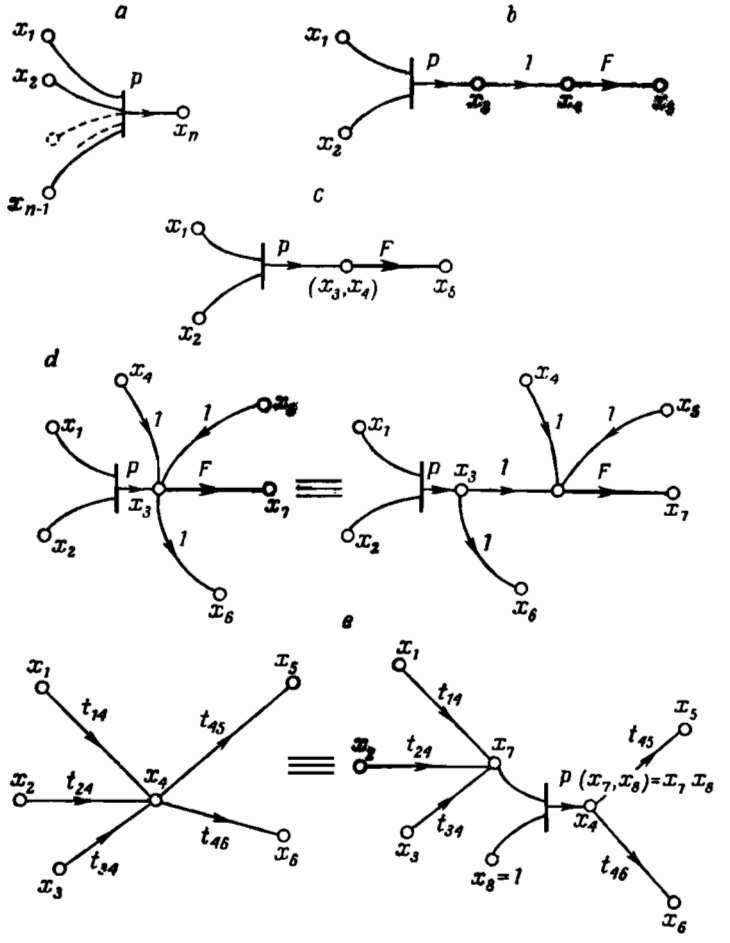


Рис. 3. Связи между функциональными ребрами.

Подобно предыдущему, вершины, связанные с многоаргументным ребром, определены единственным образом. Вершины x_1, x_2, ..., x_{n-1}, изображающие независимые переменные, становятся источникамн, а вершина x_n, изображающая зависимую переменную, становится стоком.

Рассмотрим теперь граф, содержащий два функциональных ребра (рис. 3, *b*). Согласно правилу А, из табл. 1 для этого графа мы имеем

$$x_4 = x_3,$$

а из определения функционального ребра получаем

$$x_3 = P(x_1, x_2), \quad x_5 = F(x_4).$$

Таким образом, в результате находим

$$x_5 = F(P(x_1, x_2)).$$

Этот последний граф можно изобразить также в другой форме (рис. 3, *c*). При этом следует помнить, однако, что вершина (x_3, x_4) является стоком для ребра P и источником для ребра F. Учитывая это, мы получаем для графа, показанного на рис. 3, *d*, следующие связи:

$$x_6 = P(x_1, x_2), \quad x_7 = F(P(x_1, x_2) + x_4 + x_5).$$

В дополиение к рассмотренному выше можно отметить, что функциональное ребро на рис. 3, *e* описывает подобную операцию, так как она поставлена в соотвстствие (в классическом потоковом графе) вершине, которая имеет как входящие, так и выходящие ребра.

2.4. МАТЕМАТИЧЕСКОЕ ОПИСАНИЕ ФУНКЦИОНАЛЬНЫХ РЕБЕР

Как уже упоминалось, на базе графов (описывающих обыкновенные линейные и нелинейные функциональные уравнения) будет разработан метод подготовки данных для ЭВМ, так что, базируясь на этих данных и на заранее заданной программе, с помощью ЭВМ можно решать такие уравнения.

Имея в виду указанные применения, мы ограничимся графами, характеризуемыми следующими элементами:

1. Единичный источник. Величина, изображаемая этой вершиной, принимается постоянной и равной 1. Как и для аналоговых ЭВМ, эта вершина будет играть роль ввода константы.

2. Нефункциональные ребра вместе с соответствующими коэффициентами усиления. Эти ребра будут связывать вершины, смежные с функциональными ребрами, или же связывать вершины с единичным источником.

3. Одноаргументные и многоаргументные функциональные ребра. Во множестве функциональных ребер, имеющих заданное число аргументов, существует несколько типов ребер. Каждому из этих ребер приписаны различные операции.

Среди функциональных ребер одного и того же типа существует несколько видов ребер; каждому из них приписана одна и та же операция, но описанная другими параметрами (на-

пример, если некоторое ребро соответствует интегрированию, то параметрами являются константы интегрирования и начальные условия).

В общем случае операция, приписанная данному функциональному ребру, определяется символом $F_{[\alpha, \beta, \gamma]}$, где α — число аргументов для данного ребра; β — тип функционального ребра с α аргументами; γ — вид функционального ребра, встречающегося в графе и принадлежащего к β-типу с данным α.

Пусть для данного функционального ребра, обозначенного символом $F_{[\alpha, \beta, \gamma]}$ (рис. 4, *а*), вершины, изображающие независи-

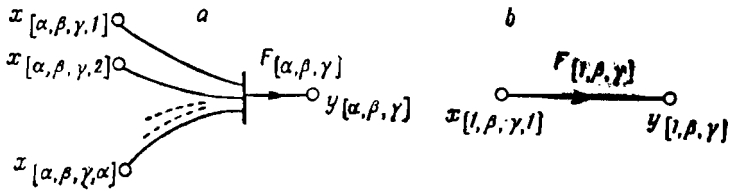


Рис. 4. Описание функциональных ребер.

мые и зависимые переменные, обозначаются следующими символами: $x_{[\alpha, \beta, \gamma, \delta]}$ — вершины, изображающие независимые переменные; δ — индекс вершины, он обозначается операторным символом; $y_{[\alpha, \beta, \gamma]}$ — вершина, изображающая зависимую переменную. Тогда, согласно (2), имеем

$$y_{[\alpha, \beta, \gamma]} = F_{[\alpha, \beta, \gamma]}\left(x_{[\alpha, \beta, \gamma, 1]}; x_{[\alpha, \beta, \gamma, 2]}; \ldots; x_{[\alpha, \beta, \gamma, \alpha]}\right). \quad (3)$$

При этом в частном случае, когда $\alpha = 1$ (рис. 4, *b*), имеем

$$y_{[1, \beta, \gamma]} = F_{[1, \beta, \gamma]}\left(x_{[1, \beta, \gamma, 1]}\right). \quad (4)$$

В дальнейшем мы введем некоторые другие обозначения:

α_m — максимальное значение для α, встречающееся в программе;

$\beta_{m\,[\alpha]}$ — максимальное число ребер типа β, встречающихся в программе (для данного α);

$\gamma_{m\,[\alpha, \beta]}$ — максимальное число α-аргументных ребер вида γ среди ребер типа β, встречающихся в графе;

$m_{[\alpha, \beta]}$ — число параметров, необходимых для описания операции $F_{[\alpha, \beta, \gamma]}$;

$P_{[\alpha, \beta, \gamma, \lambda]}$ — параметры, описывающие операцию $F_{[\alpha, \beta, \gamma]}$, где

$$\lambda = 1, 2, 3, \ldots, m_{[\alpha, \beta]}.$$

Предположим, что первые α параметров

$$P_{[\alpha,\beta,\gamma,1]};\ P_{[\alpha,\beta,\gamma,2]};\ \ldots;\ P_{[\alpha,\beta,\gamma,\alpha]}$$

определяют число функциональных и нефункциональных ребер, входящих в вершины

$$x_{[\alpha,\beta,\gamma,1]};\ x_{[\alpha,\beta,\gamma,2]};\ \ldots;\ x_{[\alpha,\beta,\gamma,\alpha]}$$

соответственно.

В программе, описанной ниже, числа α_m, $\beta_{m[\alpha]}$ и $m_{[\alpha,\beta]}$ задаются заранее, так же как и все операции. При рассмотрении частного графа возможно определить числа $\gamma_{m[\alpha,\beta]}$ и параметры $P_{[\alpha,\beta,\gamma,\lambda]}$.

Договоримся, что частные α-аргументные ребра типа β и их частные виды γ рассматриваются в следующей последовательности:

$$F_{[1,1,1]};\ F_{[1,1,2]};\ \ldots;\ F_{[1,1,\gamma_{m[1,1]}]};\ F_{[1,2,1]};\ F_{[1,2,2]};\ \ldots;\ F_{[1,2,\gamma_{m[1,2]}]};\ \ldots;$$
$$F_{[1,\beta_{m[1]},1]};\ F_{[1,\beta_{m[1]},2]};\ \ldots;\ F_{[1,\beta_{m[1]},\gamma_{m[1,\beta_{m[1]}]}]};\ F_{[2,1,1]};\ F_{[2,1,2]};\ \ldots;$$
$$F_{[2,1,\gamma_{m[2,1]}]};\ \ldots;\ F_{[\alpha_m,\beta_{m[\alpha_m]},\gamma_{m[\alpha_m,\beta_{m[\alpha_m]}]}]}. \qquad (5)$$

Используя эту последовательность, можно единственным образом приписать каждой операции $F_{[\alpha,\beta,\gamma]}$ номер, под которым эта операция встречается в последовательности. Это соответствует преобразованию индексов α, β, γ в новый индекс i, так что

$$F_{[\alpha,\beta,\gamma]}(\ ) \equiv F_{[i]}(\ ). \qquad (6)$$

Эти операции, очевидно, действуют на одни и те же аргументы. Зависимость $i = f(\alpha, \beta, \gamma)$ задается формулой

$$i \triangleq \gamma + \sum_{v=0}^{\beta-1} \gamma_m[\alpha, v] + \sum_{u=0}^{\alpha-1} \sum_{v=0}^{\beta_m[u]} \gamma_m[u, v];\ \gamma_m[\alpha, 0] = \beta_m[0] = 0. \qquad (7)$$

Предположим, что рассматриваемые графы описываются согласно вышеприведенной последовательности. Поскольку существуют и другие системы обозначения, связанные с символом операции $F_{[\alpha,\beta,\gamma]}$, мы будем производить преобразование индексов также и в этих случаях; таким образом,

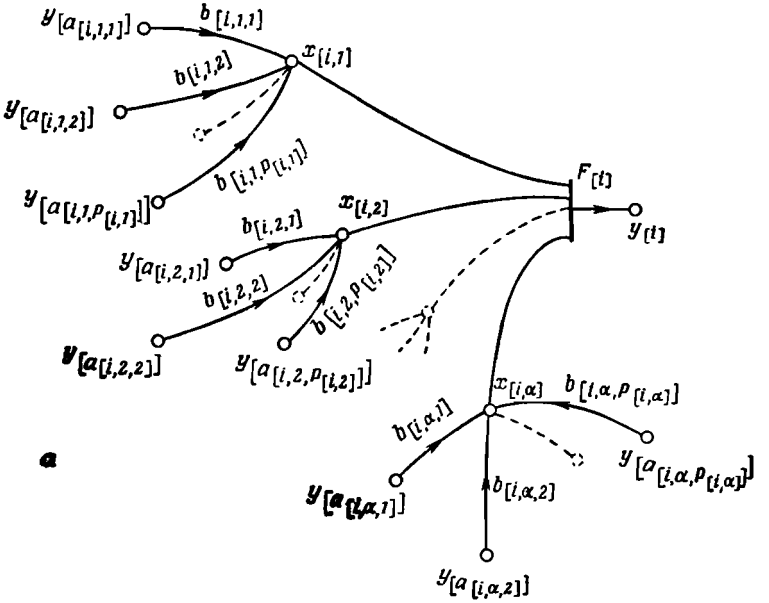
$$x_{[\alpha,\beta,\gamma,\delta]} \equiv x_{[i,\delta]}, \qquad (8)$$

$$y_{[\alpha,\beta,\gamma]} \equiv y_{[i]}, \qquad (9)$$

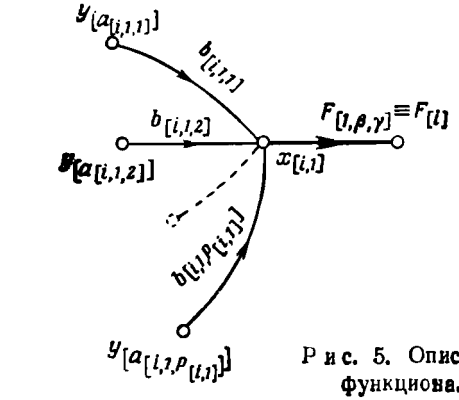
$$P_{[\alpha,\beta,\gamma,\lambda]} \equiv P_{[i,\lambda]}. \qquad (10)$$

2.5. МАТЕМАТИЧЕСКОЕ ОПИСАНИЕ СВЯЗЕЙ ФУНКЦИОНАЛЬНЫХ РЕБЕР

Предварительно было установлено, что в рассматриваемых графах вершины, изображающие независимые переменные, связаны с вершинами, изображающими зависимые переменные, с помощью нефункциональных ребер.



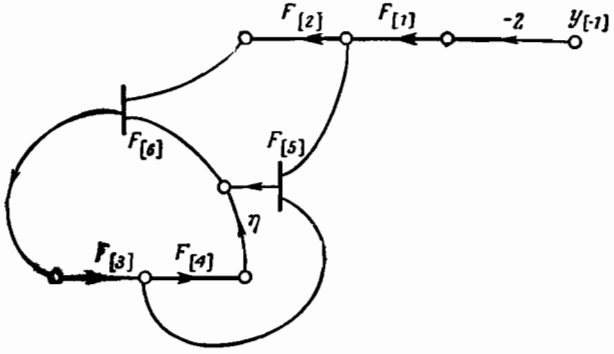
a



b

Рис. 5. Описания связей между функциональными ребрами.

Для того чтобы дать ясное описание этих связей, введем два члена $a_{[i,\,\delta,\,\tau]}$ и $b_{[i,\,\delta,\,\tau]}$. Первый из них дает индекс вершины, изображающей зависимую переменную, которая связана с вершиной $x_{[i,\,\delta]}$ посредством нефункционального ребра. Второй дает



Р и с. 6. Пример графа.

Таблица 2

Члены $a_{[i,\,\delta,\,\tau]}$ и $b_{[i,\,\delta,\,\tau]}$ для графа на рис. 6

i, δ, τ	$a_{[i,\,\delta,\,\tau]}$	$b_{[i,\,\delta,\,\tau]}$
1, 1, 1	−1	−2
2, 1, 1	1	1
3, 1, 1	6	1
4, 1, 1	3	1
5, 1, 1	1	1
5, 2, 1	3	1
6, 1, 1	4	η
6, 1, 2	5	1
6, 2, 1	2	1

коэффициент усиления этой связи. Индекс τ определяет переменный номер ребра, входящего в вершину $x_{[i,\,\delta]}$ (рис. 5, *а*). Полное число таких ребер определяется, согласно предварительному соглашению, параметром $P_{[i,\,\delta]}$.

Рассмотрим ребро с α аргументами (рис. 5, *а*), в котором каждая из вершин, изображающая независимые переменные, связана посредством нефункциональных ребер с определенным числом фуикциоиальных ребер. Согласно введенным ранее

соотношениям, для такого ребра мы имеем

$$y_{[i]} = F_{[i]}\left(\sum_{\tau=1}^{P_{[i,\,1]}} y_{[a_{[i,\,1,\,\tau]}]} b_{[i,\,1,\,\tau]};\ \sum_{\tau=1}^{P_{[i,\,2]}} y_{[a_{[i,\,2,\,\tau]}]} b_{[i,\,2,\,\tau]};\ \ldots\right.$$
$$\left.\ldots;\ \sum_{\tau=1}^{P_{[i,\,\alpha]}} y_{[a_{[i,\,\alpha,\,\tau]}]} b_{[i,\,\alpha,\,\tau]}\right). \quad (11)$$

В частности, для одноаргументного ребра (рис. 5, *б*) имеем

$$y_{[i]} = F_{[i]}\left(\sum_{\tau=1}^{P_{[i,\,1]}} y_{[a_{[i,\,1,\,\tau]}]} b_{[i,\,1,\,\tau]}\right). \quad (12)$$

В добавление к приведенным ранее рассуждениям укажем, что вершину (рассматриваемый единичный источник) будем обозначать символом $y_{[-1]}$. Таким образом, когда описание графа составлено из нескольких функциональных ребер различных типов (одноаргументных и многоаргументных) и нефункциональных ребер в форме подготовленных данных для вычислительной машины, оно должно дать независимо от таких данных, как

$$\gamma_{m\,[\alpha,\,\beta]},\quad P_{[\alpha,\,\beta,\,\gamma,\,\lambda]} \equiv P_{[i,\,\lambda]},$$

также члены $a_{[i,\,\delta,\,\tau]}$ и $b_{[i,\,\delta,\,\tau]}$. Чтобы проиллюстрировать описание графа, выполненное заданием членов $a_{[i,\,\delta,\,\tau]}$ и $b_{[i,\,\delta,\,\tau]}$, рассмотрим граф, показанный на рис. 6. Члены этого графа приведены в табл. 2.

2.6. ГРАФЫ С ИНТЕГРИРУЮЩИМИ РЕБРАМИ

Рассмотрим граф, состоящий из нескольких функциональных ребер, связанных посредством нефункциональных ребер, в котором независимые и зависимые переменные, изображаемые вершинами, суть функции переменной t, т. е.

$$x_{[\alpha,\,\beta,\,\gamma,\,\delta]} = x_{[\alpha,\,\beta,\,\gamma,\,\delta]}(t),\quad y_{[\alpha,\,\beta,\,\gamma]} = y_{[\alpha,\,\beta,\,\gamma]}(t).$$

Пусть, кроме того, ребра, для которых $\alpha = 1$ и $\beta = 1$, реализуют операцию интегрирования. В связи с этим можно записать

$$y_{[1,\,1,\,\gamma]}(t) = \frac{1}{T_{[\gamma]}} \int_0^t x_{[1,\,1,\,\gamma,\,1]}(t)\,dt + y_{[1,\,1,\,\gamma]}(0), \quad (13)$$

где $T_{[\gamma]}$ — постоянная интегрирования, связанная с операцией $F_{[1,\,1,\,\gamma]}$ и определенная соотношением

$$T_{[\gamma]} = P_{[1,\,1,\,\gamma,\,3]}, \quad (14)$$

а $y_{[1,\,1,\,\gamma]}(0)$ — начальное условие, определенное равенством

$$y_{[1,\,1,\,\gamma]}(0) = P_{[1,\,1,\,\gamma,\,2]}. \tag{15}$$

После дифференцирования обеих частей (13) мы получим

$$\frac{dy_{[1,\,1,\,\gamma]}(t)}{dt} = \frac{1}{T_{[\gamma]}}\, x_{[1,\,1,\,\gamma,\,1]}(t); \tag{16}$$

отсюда после перехода к индексу „i“ находим

$$\frac{dy_{[i]}(t)}{dt} = \frac{1}{T_{[i]}}\, x_{[i,\,1]}(t), \tag{17}$$

где $i = 1,\ 2,\ 3,\ \ldots,\ \gamma_{m\,[1,\,1]}$.

Вершина $x_{[i,\,1]}$ связана посредством нефункциональных ребер с вершинами, изображающими зависимые переменные, определенные единственным образом структурой графа. Эти связи определяются, как всегда устанавливалось выше, членами $a_{[i,\,\delta,\,\tau]}$ и $b_{[i,\,\delta,\,\tau]}$, так что, согласно (12) и (17), можно написать, что

$$\frac{dy_{[i]}(t)}{dt} = \frac{1}{T_{[i]}}\, x_{[i,\,1]}(t) = \frac{1}{T_{[i]}} \sum_{\tau=1}^{P_{[i,\,1]}} y_{[a_{[i,\,1,\,\tau]}]}\, b_{[i,\,1,\,\tau]}, \tag{18}$$

где

$$i = 1,\ 2,\ \ldots,\ \gamma_{m\,[1,\,1]}.$$

Эта формула описывает систему дифференциальных уравнений типа

$$\frac{dy_{[i]}(t)}{dt} = f_{[i]}\left(y_{[0]}(t);\ y_{[1]}(t);\ \ldots;\ y_{[\gamma_{m\,[1,\,1]}]}(t)\right), \tag{19}$$

где

$$y_{[0]}(t) = t; \qquad \left(\frac{dy_{[0]}}{dt}\right) = 1. \tag{20}$$

(Величину $y_{[0]}$ можно трактовать как второй единичный источник.) Численное интегрирование системы уравнений (18) описывается в следующем разделе.

3. АЛГОРИТМ ДЛЯ РЕШЕНИЯ СИСТЕМЫ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ, ЗАДАВАЕМЫХ ГРАФОМ

Программа численного решения системы дифференциальных уравнений, определяемой структурой графа, будет дана в виде блок-схемы. Применяемая при описании этой схемы терминология аналогична использованной в работе [8].

Рассматриваемая программа требует подготовки данных в следующей последовательности:

1. $\gamma_{m\,[\alpha,\,\beta]}$ для всех α и β;

2. $\mathscr{H}_m$ — число вершин, изображающих зависимые переменные интегрирующих ребер: значения $\mathscr{H}$ должны выдаваться на печать;

3. $W_{[\mathscr{H}]}$ — индексы вершин, изображающих зависимые переменные интегрирующих ребер, которые должны выдаваться на печать; $\mathscr{H} = 1, 2, \ldots, \mathscr{H}_m$;

4. h — шаг интегрирования;

5. B — значение переменной t, для которой может быть выполнено интегрирование;

6. $P_{[i,\,\lambda]}$ — для всех λ;

7. $a_{[i,\,\delta,\,\tau]}$ — для всех δ, τ;

8. $b_{[i,\,\delta,\,\tau]}$ — для всех δ, τ;

— для всех i.

Мы предполагаем, что в программе значения α_m, $\beta_{m\,[\alpha]}$, $m_{[\alpha,\,\beta]}$ задаются заранее, так же как и процедуры, соответствующие операциям $F_{[\alpha,\,\beta,\,\gamma]}$. Эти процедуры описываются параметрами

$$P_{[i,\,\lambda]} \equiv P_{[\alpha,\beta,\gamma,\lambda]},$$

где

$$\lambda = \alpha + 1,\ \alpha + 2,\ \ldots,\ m_{[\alpha,\,\beta]},$$

так как мы знаем, что параметры с индексами от $\lambda = 1$ до $\lambda = \alpha$ резервируются для описания количества ребер, входящих в соответствующую вершину, изображающую независимую переменную данной операции. Блок-схема программы состоит из следующих трех основных частей.

Часть I (рис. 7). В этой части программы выполняется ввод данных, а также преобразование индексов α, β, γ в единственный индекс „i“, согласно уравнению (7). Тройка индексов α, β, γ, соответствующая данному „i“, хранится в ячейках $\varepsilon_{[i]}$, $\sigma_{[i]}$, $\omega_{[i]}$.

Часть II (рис. 8). Эта часть программы вырабатывает последовательность, в которой вычисляются выражения $y_{[i]}$ для $i > \gamma_{m\,[1,\,1]}$. Последующие индексы членов $y_{[i]}$ упорядочены для того, чтобы требуемые члены были записаны в ячейках $d_{[j]}$, где $j = 1, 2, 3, \ldots, (\nu - \gamma_{m\,[1,\,1]})$. Предполагается также, что вершина

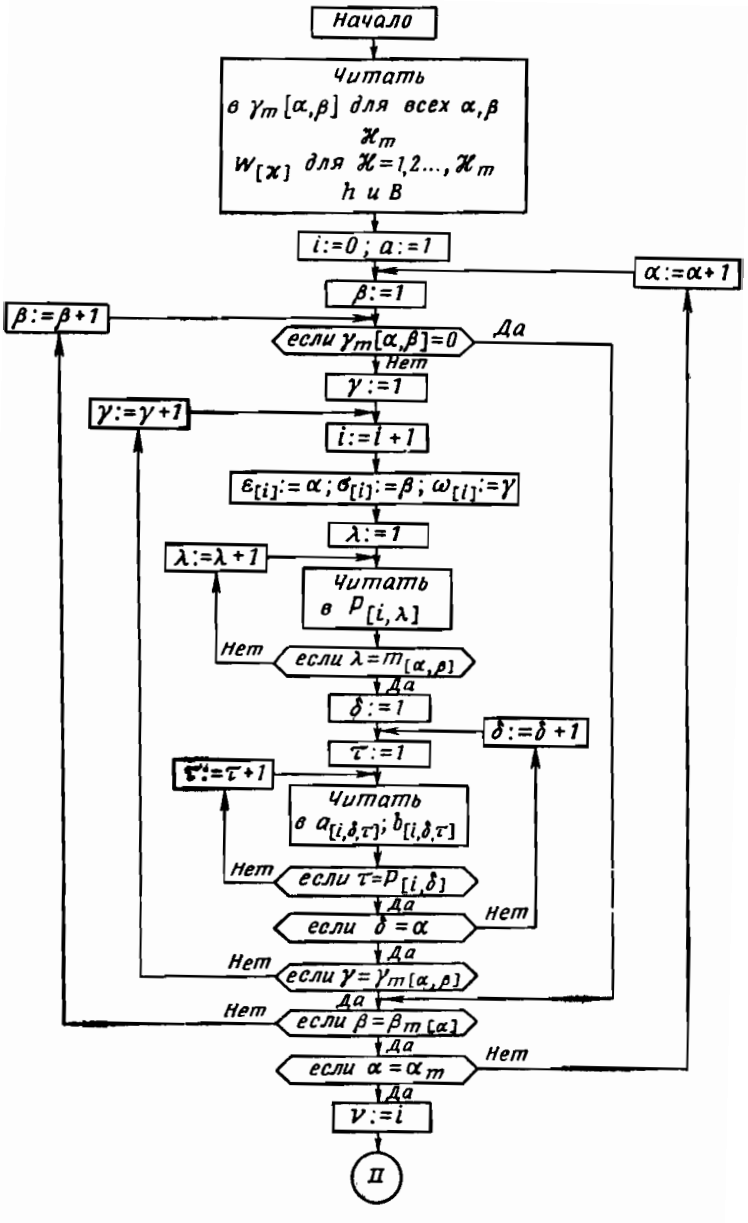
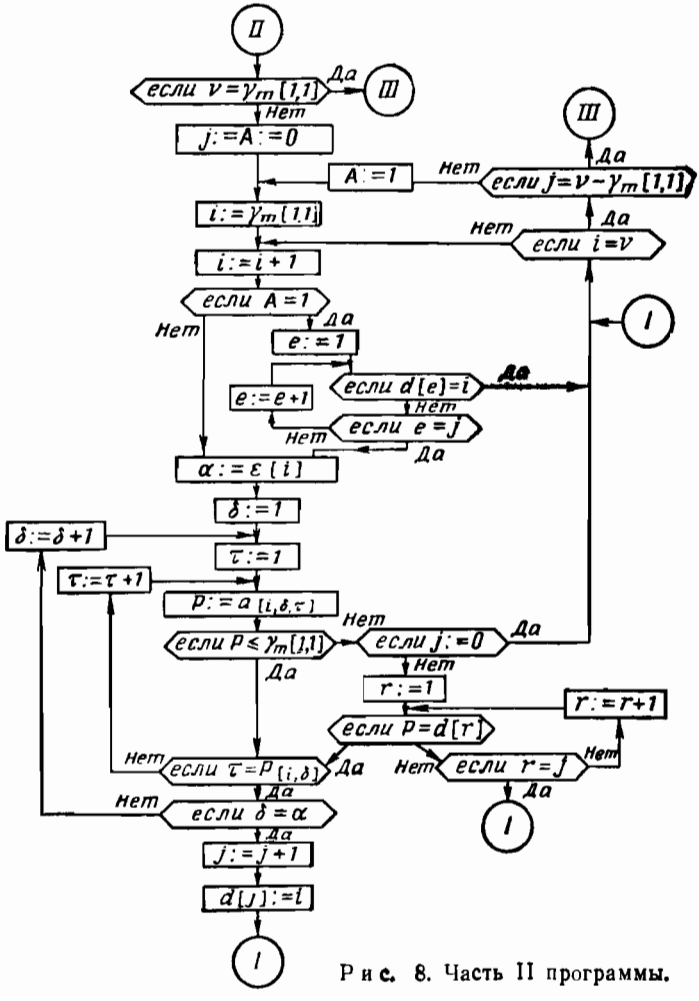


Рис. 7. Часть I программы.

$y_{[i]}$, соответствующая зависимой переменной, не связана ни с какой вершиной $x_{[i,\delta]}$ одиого и того же ребра.

Эту трудность можно обойти путем добавления функционального ребра (реализующего тождественную операцию) в виде изолирующего блока, описанного в [1].

Часть III (рис. 9). В этой части программы совокупность блоков, обозначенная символом A, производит вычисление $y_{[i]}$ для $i > \gamma_{m\,[1,\,1]}$ в последовательности, установленной выше.



Р и с. 8. Часть II программы.

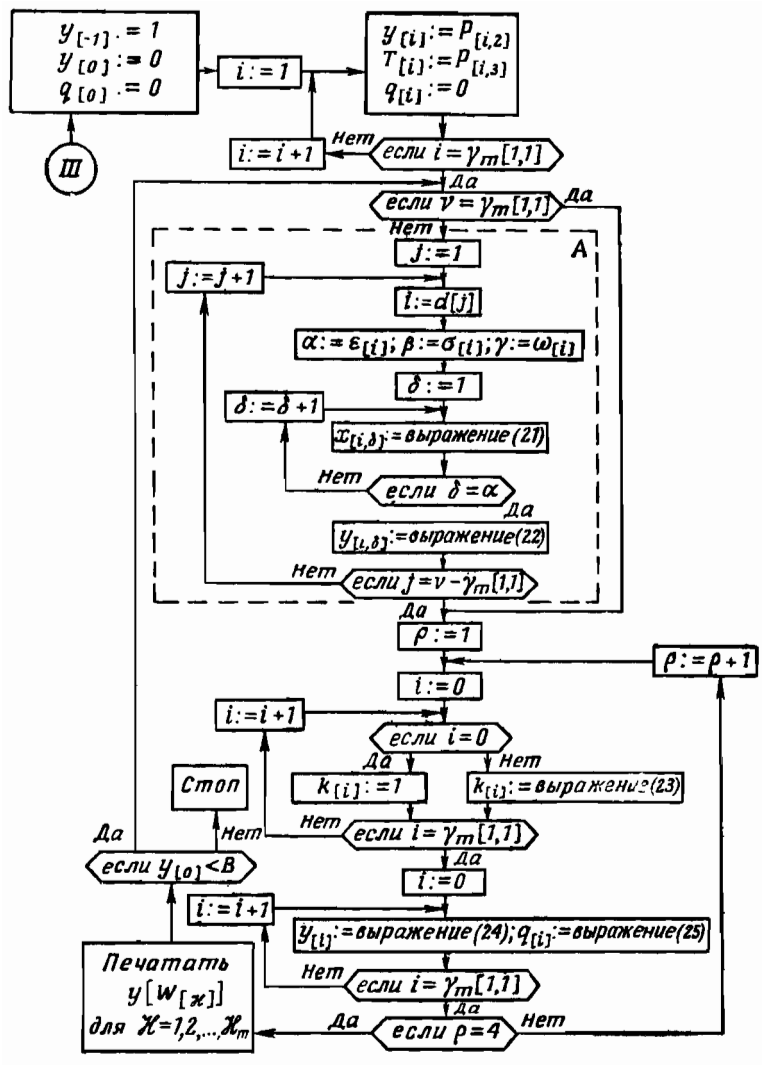


Рис. 9. Часть III программы.

Отношения, которые не включены в блоки этой совокупности, имеют форму

$$x_{[i,\,\delta]}: = \sum_{\tau=1}^{P_{[i,\,\delta]}} y_{[a_{[i,\,\delta,\,\tau]}]} b_{[i,\,\delta,\,\tau]}, \tag{21}$$

$$y_{[i]}: = F_{[i]}\left(x_{[i,\,1]};\ x_{[i,\,2]};\ \ldots;\ x_{[i,\,\alpha]}\right). \tag{22}$$

Нужно указать, что в вычислении последнего уравнения мы можем использовать соотношения (6) и (10).

Последующие вычисления выполняются путем решения системы дифференциальных уравнений, задаваемых графом, по методу Рунге – Кутта [9], модифицированному Джиллом. Этот метод выбран из-за его простоты; вполне возможно применение других методов.

Основной упор в этой статье был сделан не на выбор оптимального численного решения и не на оценку точности, а на подготовку данных и на построение программы. Соотношения, не задаваемые явно в блоках этой программы, перечислены ниже:

$$k_{[i]}: = \frac{1}{T_{[i]}} \sum_{\tau=1}^{P_{[i,\,1]}} y_{[a_{[i,\,1,\,\tau]}]} b_{[i,\,1,\,\tau]}, \qquad i > 0, \tag{23}$$

$$y_{[i]}: = y_{[i]} + h\left[\Psi_{[\rho]}\left(k_{[i]} - \mu_{[\rho]} q_{[i]}\right)\right], \tag{24}$$

$$q_{[i]}: = q_{[i]} + 3\left[\Psi_{[\rho]}\left(k_{[i]} - \mu_{[\rho]} q_{[i]}\right)\right] - \varphi_{[\rho]} k_{[i]}, \tag{25}$$

где

$$\Psi_{[1]} = \frac{1}{2}, \qquad \mu_{[1]} = 2, \qquad \varphi_{[1]} = \frac{1}{2},$$

$$\Psi_{[2]} = 1 - \sqrt{\frac{1}{2}}, \qquad \mu_{[2]} = 1, \qquad \varphi_{[2]} = 1 - \sqrt{\frac{1}{2}},$$

$$\Psi_{[3]} = 1 + \sqrt{\frac{1}{2}}, \qquad \mu_{[3]} = 1, \qquad \varphi_{[3]} = 1 + \sqrt{\frac{1}{2}},$$

$$\Psi_{[4]} = \frac{1}{6}, \qquad \mu_{[4]} = 2, \qquad \varphi_{[4]} = \frac{1}{8}.$$

ЛИТЕРАТУРА

1. Steel G. H., Programming of digital computers for transient studies in control systems, *Int. J. Electr. Eng. Educ.*, 3 (1965), 261—278.
2. Brannan R. D., Fanidy T. D., Digital simulation, *Instruments and Control Systems*, № 3 (1966), 133—137.
3. Janoski R. M., Schaeffer R. L., COBLOC — a program for all-digital simulation of a hybrid computer, *IEEE Trans. Electronic Computers*, EC-15, № 1 (1966), 74—82.
4. Berle F. J., Haberstock F., Blockorientierte Programmierung bei der Untersuchung dynamischer Systeme, *Elektronische Rechenanlagen*, Heft 3 (1966), 135—140.

5. Becker S. M., Try signal flow graphs to simplify analog programming, *Contr. Eng.*, 11, № 9 (1964), 109—111.
6. Robichaud L. P. A., Boisvert M., Robert J., Graphs de fluence, Applications à l'électrotechnique et à l'électronique. Calculateurs analogiques et digitaux, Presses de l'Université Laval, 1961.
7. Chow Y., Cassignol E., Linear signal-flow graphs and applications, Wiley, New York, 1962.
8. Burnett-Hall D. G., Dressel L. A. G., Samet P. A., Computer programming and autocodes, The English Universities Press L. T. D., 1961.
9. Ralston A., Wilf H. S., Mathematical methods for digital computers, Wiley, New York, 1962.

СОДЕРЖАНИЕ

Математические вопросы

Прикладные вопросы

КИБЕРНЕТИЧЕСКИЙ СБОРНИК

Новая серия

Выпуск 7

Редактор *А. С. Попов*
Художник *Н. К. Сапожников*
Художественный редактор *В. И. Шаповалов*
Технический редактор *И. К. Дерва*
Корректор *Н. И. Баранова*

Сдано в производство 19/III 1970 г. Подписано к печати 21/VIII 1970 г. Бумага тип. № 1, 60 × 90¹/₁₆=7,5 бум. л. 15 усл. печ. л. Уч.-изд. л. 12,82. Изд. № 1/5747. Цена 1 р. 24 к. Зак. 560.

ИЗДАТЕЛЬСТВО «МИР»
Москва, 1-й Рижский пер., 2

Ордена Трудового Красного Знамени Ленинградская типография № 2 имени Евгении Соколовой Главполиграфпрома Комитета по печати при Совете Министров СССР. Измайловский проспект, 29

В ИЗДАТЕЛЬСТВЕ «МИР»

готовится к печати:

Берлекэмп Э. **Алгебраическая теория кодирования,** Нью-Йорк, 1968, перев. с англ., 25 л.

Книга американского ученого, освещающая основные вопросы общей теории линейных кодов, исследования циклических (двоичных и недвоичных) кодов для метрик Хэмминга и Ли, вычисление параметров оптимальных кодов, а также вопросы построения кодирующих и декодирующих устройств. Теоретические исследования сопровождаются большим количеством примеров и задач, что делает книгу интересной и доступной не только для математиков, но и для широкого круга специалистов, связанных с разработкой систем передачи цифровой информации, а также для аспирантов и студентов соответствующих специальностей.

Минский М., Пейперт С. **Персептроны,** Нью-Йорк, 1969, перев. с англ., 15 л.

Книга видных американских ученых посвящена параллельным вычислительным устройствам, известным под названием персептронов. В ней подробно проанализированы общие алгебраические и геометрические свойства подобных схем, рассмотрены вопросы, связанные с обучением персептронов, в частности длительность процесса обучения, эффективность схемы как адаптивного запоминающего устройства и т. п., а также исследованы потенциальные возможности персептронов как обучающихся распознающих устройств.

Книга представляет несомненный интерес для специалистов по современной кибернетике, в частности зарождающейся общей теории вычислений и вычислительных схем.