РОССИЙСКАЯ АКАДЕМИЯ НАУК

ПРОБЛЕМЫ ПЕРЕДАЧИ ИНФОРМАЦИИ

Журнал основан в январе 1965 г. **ISSN: 0555-2923** **Выходит 4 раза в год**

Том 57, 2021 **Вып. 3**

Июль–Август–Сентябрь

М о с к в а

С О Д Е Р Ж А Н И Е

CONTENTS

ПРОБЛЕМЫ ПЕРЕДАЧИ ИНФОРМАЦИИ

Том 57 **2021** **Вып. 3**

УДК 621.391 : 519.725

© 2021 г. **А.М. Романов**

О СОВЕРШЕННЫХ КОДАХ И КОДАХ РИДА – МАЛЛЕРА НАД КОНЕЧНЫМИ ПОЛЯМИ[1]

Рассматриваются коды с исправлением ошибок над конечным полем с q элементами (q-ичные коды). Изучается связь q-ичных совершенных кодов с исправлением одной ошибки и q-ичных кодов Рида – Маллера. При $q \geqslant 3$ найдены параметры аффинных кодов Рида – Маллера порядка $(q-1)m-2$. Показано, что аффинные коды Рида – Маллера порядка $(q-1)m-2$ являются квазисовершенными кодами. Предложена конструкция, которая позволяет строить q-ичные совершенные коды, исправляющие одну ошибку, из кодов с параметрами аффинных кодов Рида – Маллера. Модификация этой конструкции позволяет строить q-ичные квазисовершенные коды с параметрами аффинных кодов Рида – Маллера.

Ключевые слова: код Рида – Маллера, аффинный код Рида – Маллера, проективный код Рида – Маллера, код Хэмминга, совершенный код, квазисовершенный код, МДР-код, конечное поле.

DOI: 10.31857/S0555292321030013

§ 1. Введение

В данной статье рассматриваются аффинные и проективные коды Рида – Маллера и совершенные коды с исправлением одной ошибки (1-совершенные коды). Все эти коды определяются над конечным полем $\mathbb{F}_q$ и являются q-ичными кодами. В статье изучается связь между q-ичными кодами Рида – Маллера и q-ичными 1-совершенными кодами.

Классические коды Рида – Маллера являются двоичными кодами и определяются над конечным полем $\mathbb{F}_2$ (см. [1, гл. 13]). Без преувеличения можно сказать, что коды Рида – Маллера являются наиболее известным и изученным классом двоичных кодов с исправлением ошибок. Аффинные коды Рида – Маллера были открыты в работе [2] и были названы авторами этой работы обобщенными кодами Рида – Маллера. Аффинные коды Рида – Маллера являются прямым обобщением двоичных кодов Рида – Маллера и определяются над любым конечным полем $\mathbb{F}_q$. В данной статье при $q \geqslant 3$ найдены параметры аффинных кодов Рида – Маллера порядка $(q-1)m-2$.

Значительно позднее были открыты так называемые проективные коды Рида – Маллера [3]. Из работ [3, 4] следует, что при $q \geqslant 3$ проективный код Рида – Маллера порядка $(q-1)m-1$ является q-ичным кодом Хэмминга длины $n = (q^{m+1}-1)/(q-1)$.

Группы автоморфизмов аффинных кодов Рида – Маллера известны [5]. Исследование групп автоморфизмов проективных кодов Рида – Маллера предложено в [6].

[1] Работа выполнена при поддержке программы фундаментальных научных исследований СО РАН № I.5.1., проект № 0314-2019-0017.

Хорошо известна связь между расширенными двоичными кодами Хэмминга и кодами Рида – Маллера. Известно, что расширенный двоичный код Хэмминга длины 2^m является кодом Рида – Маллера длины 2^m и порядка $m-2$. Также известно, что код Рида – Маллера порядка $m-2$ (двоичный расширенный код Хэмминга) можно построить с помощью $(\boldsymbol{u}\,|\,\boldsymbol{u}+\boldsymbol{v})$-конструкции [1, с. 76]. Вертикальная черта $(\cdot\,|\,\cdot)$ обозначает конкатенацию. Следует заметить, что с помощью $(\boldsymbol{u}\,|\,\boldsymbol{u}+\boldsymbol{v})$-конструкции можно построить код Рида – Маллера любого порядка [7, с. 34].

Одним из наиболее простых и распространенных методов построения нелинейных двоичных 1-совершенных кодов является комбинаторное обобщение $(\boldsymbol{u}\,|\,\boldsymbol{u}+\boldsymbol{v})$-конструкции (см., например, [8]). При этом этот метод позволяет строить как нелинейные двоичные 1-совершенные коды, так и расширенные нелинейные двоичные 1-совершенные коды. В [9] Фелпс представил комбинаторную конструкцию расширенных двоичных 1-совершенных кодов. В [10] он предложил многомерное обобщение двоичной комбинаторной конструкции из [9] также на двоичный случай. Работы Фелпса [9, 10] являются классическими работами по теории нелинейных совершенных кодов. В [8] дано обобщение двоичной комбинаторной $(\boldsymbol{u}\,|\,\boldsymbol{u}+\boldsymbol{v})$-конструкции на q-ичный случай. В настоящей статье предложена конструкция, обобщающая конструкции из [8, 10] и позволяющая строить q-ичные 1-совершенные коды из q-ичных кодов с параметрами аффинных кодов Рида – Маллера порядка $(q-1)m-2$. Модификация этой конструкции позволяет строить q-ичные квазисовершенные коды с параметрами аффинных кодов Рида – Маллера порядка $(q-1)m-2$. Приведены также нижние оценки для числа неэквивалентных q-ичных 1-совершенных кодов и для числа неэквивалентных q-ичных квазисовершенных кодов с параметрами аффинных кодов Рида – Маллера порядка $(q-1)m-2$.

Методы построения кодов из [8–10] и метод, предлагаемый в настоящей статье, являются примерами конструкции обобщенных каскадных кодов [11].

Из результатов настоящей статьи следует, что аффинные коды Рида – Маллера порядка $(q-1)m-2$ играют такую же роль в теории q-ичных 1-совершенных кодов, что и расширенные двоичные коды Хэмминга в теории двоичных 1-совершенных кодов.

Аффинные коды Рида – Маллера порядка $(q-1)m-2$, так же как и расширенные двоичные коды Хэмминга, являются линейными квазисовершенными кодами, и следовательно, равномерно упакованными и полностью регулярными кодами.

Известно, что коды, имеющие параметры расширенных двоичных кодов Хэмминга, являются квазисовершенными, равномерно упакованными и полностью регулярными [12].

В § 2 приведены необходимые определения и обозначения. В § 3 дано определение аффинных кодов Рида – Маллера и найдены параметры аффинных кодов Рида – Маллера длины q^m и порядка $(q-1)m-2$. В § 4 дано определение проективных кодов Рида – Маллера и показано, что при $q \geqslant 3$ код Хэмминга длины $n=(q^{m+1}-1)/(q-1)$ является проективным кодом Рида – Маллера порядка $(q-1)m-1$. В § 5 приведены конструкции 1-совершенных кодов и кодов с параметрами аффинных кодов Рида – Маллера.

§ 2. Определения и обозначения

Пусть $\mathbb{F}_q^n$ – векторное пространство размерности n над конечным полем $\mathbb{F}_q$, где q – степень простого числа. Произвольное подмножество $\mathcal{C} \subseteq \mathbb{F}_q^n$ называется *q-ичным кодом с исправлением ошибок* (кратко – q-ичным кодом). *Длина* кода $\mathcal{C} \subseteq \mathbb{F}_q^n$ равна размерности пространства $\mathbb{F}_q^n$. Мы предполагаем, что нулевой вектор $\mathbf{0}$ всегда принадлежит коду, если не указано иное. Код называется *линейным*, если он образует линейное подпространство над $\mathbb{F}_q$. В противном случае код называется *нелинейным*. Векторы, принадлежащие пространству $\mathbb{F}_q^n$, будем рассматривать как слова длины n

над алфавитом $\mathbb{F}_q$. Слова, принадлежащие коду $\mathcal{C} \subseteq \mathbb{F}_q^n$, будем называть *кодовыми словами*.

Расстояние Хэмминга $d(\boldsymbol{x}, \boldsymbol{y})$ между двумя векторами $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{F}_q^n$ равно числу ненулевых координатных позиций в векторе $\boldsymbol{x} - \boldsymbol{y}$. *Вес* вектора $\boldsymbol{x} \in \mathbb{F}_q^n$ равен $d(\boldsymbol{x}, \boldsymbol{0})$. *Минимальное расстояние* кода $\mathcal{C}$ равно наименьшему расстоянию Хэмминга между двумя различными кодовыми словами из $\mathcal{C}$ и обозначается через $d(\mathcal{C})$. Линейный q-ичный код длины n и размерности $\dim(\mathcal{C}) = k$ с минимальным расстоянием d называется $[n, k, d]_q$-кодом. Нелинейный q-ичный код длины n с M кодовыми словами и минимальным расстоянием d называется $(n, M, d)_q$-кодом.

Для кода с параметрами $(n, M, d)_q$ справедливо неравенство

$$M \leqslant q^{n-d+1}. \tag{1}$$

Неравенство (1) называется *границей Синглтона*. Коды, достигающие границы Синглтона, называются *разделимыми кодами с максимальным расстоянием*, или кратко МДР-кодами. В данной статье будут рассматриваться МДР-коды только с минимальным расстоянием 2.

Для кода с параметрами $(n, M, d)_q$ справедливо также неравенство

$$q^n \geqslant M \sum_{i=0}^{\lfloor \frac{d-1}{2} \rfloor} (q-1)^i \binom{n}{i}, \tag{2}$$

где $\lfloor \cdot \rfloor$ обозначает целую часть числа. Неравенство (2) называется *границей сферической упаковки*. Коды, достигающие границы сферической упаковки, называются *совершенными кодами*. Совершенные q-ичные коды с минимальным расстоянием 3 (т.е. коды, исправляющие одну ошибку) называются q-ичными 1-*совершенными* кодами. Линейные q-ичные 1-совершенные коды называются q-ичными *кодами Хэмминга* и обозначаются через $\mathcal{H}_q(m)$. Коды Хэмминга имеют следующие параметры:

1. Длина q-ичного кода Хэмминга $\mathcal{H}_q(m)$ равна $n = \dfrac{q^m - 1}{q - 1}$, где $m \geqslant 2$;
2. Размерность q-ичного кода Хэмминга $\mathcal{H}_q(m)$ равна $n - m$;
3. Минимальное расстояние q-ичного кода Хэмминга $\mathcal{H}_q(m)$ равно 3.

Параметры нелинейных q-ичных 1-совершенных кодов длины n совпадают с параметрами q-ичного кода Хэмминга длины n.

Коды $\mathcal{C}_1, \mathcal{C}_2 \subseteq \mathbb{F}_q^n$ называются *различными*, если $\mathcal{C}_1 \neq \mathcal{C}_2$. Два кода $\mathcal{C}_1, \mathcal{C}_2 \subseteq \mathbb{F}_q^n$ называются *эквивалентными*, если существует вектор $\boldsymbol{v} \in \mathbb{F}_q^n$ и мономиальная матрица M размера $n \times n$ над полем $\mathbb{F}_q$, такие что

$$\mathcal{C}_2 = \{(\boldsymbol{v} + \boldsymbol{c}M) \mid \boldsymbol{c} \in \mathcal{C}_1\}.$$

Известно, что существует единственный с точностью до эквивалентности q-ичный код Хэмминга длины n (см. [7, теорема 1.8.2]).

Определим *радиус упаковки* $e(\mathcal{C})$ кода $\mathcal{C} \subseteq \mathbb{F}_q^n$. Положим $e(\mathcal{C}) = \lfloor (d-1)/2 \rfloor$, где d – минимальное расстояние кода $\mathcal{C}$. Определим также *радиус покрытия* $\rho(\mathcal{C})$ кода $\mathcal{C}$. Положим

$$\rho(\mathcal{C}) = \max_{\boldsymbol{x} \in \mathbb{F}_q^n} \min_{\boldsymbol{c} \in \mathcal{C}} d(\boldsymbol{x}, \boldsymbol{c}).$$

Код $\mathcal{C}$ является совершенным тогда и только тогда, когда $\rho(\mathcal{C}) = e(\mathcal{C})$. Код $\mathcal{C}$ называется *квазисовершенным*, если $\rho(\mathcal{C}) = e(\mathcal{C}) + 1$ (см. [1, с. 19]).

Далее определим функцию $p\colon \mathbb{F}_q^n \to \mathbb{F}_q$. Если $\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{F}_q^n$, то

$$p(\boldsymbol{x}) = \sum_{i=1}^{n} x_i.$$

Функция $p(\boldsymbol{x})$ называется *функцией четности*. Пусть q-ичный код $\mathcal{C}$ имеет параметры $(n, M, d)_q$. Тогда определим *расширенный* код $\widehat{\mathcal{C}}$. Положим

$$\widehat{\mathcal{C}} = \big\{\boldsymbol{x} = (x_1, x_2, \ldots, x_n, x_{n+1}) \in \mathbb{F}_q^{n+1} \mid (x_1, x_2, \ldots, x_n) \in \mathcal{C} \text{ и } p(\boldsymbol{x}) = 0\big\}.$$

Расширенный код $\widehat{\mathcal{C}}$ имеет параметры $(n+1, M, \widehat{d})_q$, где $\widehat{d} = d$ или $d+1$ [7, с. 14]. Говорят, что код $\widehat{\mathcal{C}}$ получен из кода $\mathcal{C}$ добавлением общей проверки на четность.

Вектор $\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{F}_q^n$ называется *четным*, если $p(\boldsymbol{x}) = \sum\limits_{i=1}^{n} x_i = 0$. Код называется *четным*, если он имеет только четные кодовые слова. Четный код длины n является подкодом линейного МДР-кода с параметрами $[n, n-1, 2]_q$.

Все не определенные в данной статье понятия можно найти в [1, 7].

§ 3. Аффинные коды Рида – Маллера

Рассмотрим алгебру $\mathbb{F}_q[X_1, X_2, \ldots, X_m]$ многочленов от m переменных над полем $\mathbb{F}_q$. Через $\deg(f)$ обозначим полную степень многочлена f из $\mathbb{F}_q[X_1, X_2, \ldots, X_m]$. Пусть $n = q^m$, и пусть точки $P_1, P_2, \ldots, P_n$ аффинного пространства $AG(m, q)$ размерности m над полем $\mathbb{F}_q$ некоторым образом упорядочены. Пусть r – целое число, такое что $0 \leqslant r \leqslant (q-1)m$. Тогда аффинным (или обобщенным [2]) кодом Рида – Маллера порядка r над полем $\mathbb{F}_q$ называется подпространство

$$RM_q(r, m) = \big\{(f(P_1), f(P_2), \ldots, f(P_n)) \mid f \in \mathbb{F}_q[X_1, X_2, \ldots, X_m],\ \deg(f) \leqslant r\big\}.$$

Аффинные коды Рида – Маллера $RM_q(r, m)$ порядка r над полем $\mathbb{F}_q$ имеют следующие параметры (см. [2; 13, теорема 5.5]):

1. Длина кода $RM_q(r, m)$ равна q^m;
2. Размерность кода $RM_q(r, m)$ равна

$$\sum_{k=0}^{m} (-1)^k \binom{m}{k} \binom{m + r - kq}{r - kq}; \tag{3}$$

3. Минимальное расстояние кода $RM_q(r, m)$ равно

$$(q-b)q^{m-a-1}, \tag{4}$$

где $r = (q-1)a + b$ и $0 \leqslant b < q-1$.

При $q = 2$ и $0 \leqslant r \leqslant m$ аффинный код Рида – Маллера порядка r является классическим кодом Рида – Маллера порядка r (см. [13, пример 5.4]).

Утверждение 1. *При $q \geqslant 3$ аффинный код Рида – Маллера над полем $\mathbb{F}_q$ длины $n = q^m$ и порядка $(q-1)m - 2$ имеет размерность $n - m - 1$ и минимальное расстояние* 3.

Доказательство. Если код $\mathcal{C}$ принадлежит классу аффинных кодов Рида – Маллера, то и дуальный код $\mathcal{C}^{\perp}$ также принадлежит этому классу [14]. Допустим, что длина кода $\mathcal{C}$ равна n. Тогда

$$\dim(\mathcal{C}) + \dim(\mathcal{C}^{\perp}) = n. \tag{5}$$

Из [13, теорема 5.8] следует, что при $r < (q-1)m$

$$RM_q(r,m)^{\perp} = RM_q((q-1)m-1-r,m).$$

Следовательно, порядок кода, дуального к коду $RM_q((q-1)m-2,m)$, равен 1. В силу формул (3) и (5) размерность аффинного кода Рида – Маллера $RM_q((q-1)m-2,m)$ равна $n-m-1$.

Теперь покажем, что минимальное расстояние кода $RM_q((q-1)m-2,m)$ равно 3. Поскольку

$$r = (q-1)m-2 = (q-1)a+b \quad \text{и} \quad 0 \leqslant b < q-1,$$

то при $q \geqslant 3$ имеем $a = m-1$ и $b = q-3$. Следовательно, в силу формулы (4) код $RM_q((q-1)m-2,m)$ имеет минимальное расстояние 3. ▲

Аффинный код Рида – Маллера порядка $(q-1)m-1$ является линейным q-ичным МДР-кодом длины q^m с минимальным расстоянием 2. Очевидно, что если $r < v$, то $RM_q(r,m) \subset RM_q(v,m)$. Следовательно, код $RM_q((q-1)m-2,m)$ является подкодом линейного МДР-кода длины q^m с минимальным расстоянием 2, т.е. четным кодом.

Утверждение 2. *Аффинный код Рида – Маллера $RM_q((q-1)m-2,m)$ является квазисовершенным кодом.*

Доказательство. Аффинный код Рида – Маллера 1-го порядка является линейным двухвесовым кодом [14]. Согласно [15, теорема 5.10] код, дуальный к линейному двухвесовому коду, имеет радиус покрытия 2. Следовательно, аффинные коды Рида – Маллера порядка $(q-1)m-2$ являются квазисовершенными. ▲

В силу [16, теорема 3.11] линейные квазисовершенные коды являются равномерно упакованными. Равномерно упакованные коды достигают границы Джонсона и являются оптимальными кодами [16, теорема 1.3]. Известно также, что равномерно упакованный код полностью регулярен [17].

§ 4. Проективные коды Рида – Маллера

Пусть r – целое число, такое что $1 \leqslant r \leqslant (q-1)m$. Рассмотрим множество $\mathbb{S}_r$ однородных многочленов полной степени r над полем $\mathbb{F}_q$ от переменных $X_0, X_1, \ldots, X_m$. Пусть $n = \dfrac{q^{m+1}-1}{q-1}$, и пусть точки $P_0, P_1, \ldots, P_{n-1}$ проективного пространства $PG(m,q)$ размерности m над полем $\mathbb{F}_q$ некоторым образом упорядочены. Точки проективного пространства $PG(m,q)$ задаются так, чтобы крайняя левая ненулевая координата была равна единице. Тогда проективный код Рида – Маллера порядка r над полем $\mathbb{F}_q$ состоит из слов

$$PRM_q(r,m) = \big\{(f(P_0), f(P_1), \ldots, f(P_{n-1})) \mid f \in \mathbb{S}_r \cup \{\mathbf{0}\}\big\}.$$

Длина проективного кода Рида – Маллера $PRM_q(r,m)$ равна $n = \dfrac{q^{m+1}-1}{q-1}$. Известны и другие параметры проективных кодов Рида – Маллера [3, 4].

Утверждение 3. *При $q \geqslant 3$ проективный код Рида – Маллера над полем $\mathbb{F}_q$ длины $n = \dfrac{q^{m+1}-1}{q-1}$ и порядка $(q-1)m-1$ является кодом Хэмминга $\mathcal{H}_q(m+1)$.*

Доказательство. Проективный q-ичный код Рида – Маллера первого порядка является q-ичным симплексным кодом [3]. Известно, что q-ичный симплексный код дуален q-ичному коду Хэмминга [7, с. 30].

Пусть $\mu = (q-1)m - \nu$. Тогда в [4] доказано, что

$$PRM_q(\nu, m)^{\perp} = PRM_q(\mu, m) \quad \text{при} \quad \nu \neq 0 \mod (q-1),$$
$$PRM_q(\nu, m)^{\perp} = \overline{PRM_q(\mu, m)} \quad \text{при} \quad \nu = 0 \mod (q-1),$$

где код $\overline{PRM_q(\mu, m)}$ порождается проективным кодом Рида – Маллера $PRM_q(\mu, m)$ и вектором, состоящим из единиц. Таким образом, мы получаем, что при $q \geqslant 3$ код Хэмминга длины $n = \dfrac{q^{m+1}-1}{q-1}$ является проективным кодом Рида – Маллера порядка $(q-1)m - 1$. ▲

При $q = 2$, $m \geqslant 2$ проективный код Рида – Маллера $PRM_2(m-1, m)$ является четной половиной двоичного кода Хэмминга $\mathcal{H}_2(m+1)$ и является полностью регулярным кодом [12].

§ 5. Конструкции кодов

В этом параграфе представлена конструкция q-ичных 1-совершенных кодов. Эта конструкция основана на разбиениях и является обобщением конструкций из [8,10]. Модификация этой конструкции позволяет строить квазисовершенные коды с параметрами аффинных кодов Рида – Маллера. Приведены также нижние оценки для числа неэквивалентных q-ичных 1-совершенных кодов и для числа неэквивалентных q-ичных квазисовершенных кодов с параметрами аффинных кодов Рида – Маллера.

1. Пусть $\mathcal{B}_0, \mathcal{B}_1, \ldots, \mathcal{B}_{(q-1)n}$ – разбиение пространства $\mathbb{F}_q^n$ на q-ичные 1-совершенные коды длины $n = \dfrac{q^{s_1}-1}{q-1}$, где $s_1 \geqslant 2$.
2. Пусть пространство $\mathbb{F}_q^{(q-1)n+1}$ разбито на смежные классы, образованные аффинным кодом Рида – Маллера $RM_q((q-1)s_1 - 1, s_1)$ порядка $(q-1)s_1 - 1$. Поскольку $n = \dfrac{q^{s_1}-1}{q-1}$, то $(q-1)n + 1 = q^{s_1}$.
3. Пусть $\mathcal{C}_0^k, \mathcal{C}_1^k, \ldots, \mathcal{C}_{(q-1)n}^k$ – разбиение k-го смежного класса на коды с параметрами аффинных кодов Рида – Маллера $RM_q((q-1)s_1 - 2, s_1)$ порядка $(q-1)s_1 - 2$, коды могут быть как линейные, так и нелинейные, $0 \leqslant k \leqslant q-1$.
4. Пусть $\mathcal{R}$ является q-ичным 1-совершенным кодом длины $m = \dfrac{q^{s_2}-1}{q-1}$, где $s_2 \geqslant 2$. Для каждого кодового слова $\boldsymbol{r} \in \mathcal{R}$ определим m-арную квазигруппу $q_{\boldsymbol{r}}$ порядка $(q-1)n + 1$. Квазигруппа определяется на множестве $\{0, 1, \ldots, (q-1)n\}$. Квазигруппу $q_{\boldsymbol{r}}$ можно рассматривать как код длины $m+1$ с минимальным расстоянием 2 над алфавитом из $(q-1)n + 1$ элементов.

Тогда образуем q-ичный код $\mathcal{H}$. Положим

$$\mathcal{H} = \Big\{ (\boldsymbol{u} \,|\, \boldsymbol{v}_1 \,|\, \boldsymbol{v}_2 \,|\, \ldots \,|\, \boldsymbol{v}_m) \mid \boldsymbol{u} \in \mathcal{B}_{j_0},\ \boldsymbol{v}_i \in \mathcal{C}_{j_i}^{r_i} \text{ при } 1 \leqslant i \leqslant m,$$
$$\boldsymbol{r} = (r_1, r_2, \ldots, r_m) \in \mathcal{R},\ j_0 = q_{\boldsymbol{r}}(j_1, j_2, \ldots, j_m),$$
$$j_i \in \{0, 1, \ldots, (q-1)n\} \text{ при } i = 0, 1, \ldots, m \Big\}. \tag{6}$$

Поясним обозначение $\mathcal{C}_j^{r_i}$. Для этого рассмотрим i-ю компоненту r_i кодового слова $\boldsymbol{r} = (r_1, r_2, \ldots, r_m) \in \mathcal{R}$. По определению $r_i \in \mathbb{F}_q$. Пусть α – примитивный элемент конечного поля $\mathbb{F}_q$. Если $r_i \neq 0$, то $r_i = \alpha^k$, где $k \in \{1, \ldots, q-1\}$. Тогда обозначение $\mathcal{C}_j^{r_i}$ следует понимать как $\mathcal{C}_j^k$, при этом $0 \leqslant j \leqslant (q-1)n$. Если $r_i = 0$, то $\mathcal{C}_j^{r_i} = \mathcal{C}_j^0$ и $\mathcal{C}_j^{r_i}$ является элементом разбиения 0-го смежного класса.

Теорема 1. *Построенный выше код $\mathcal{H}$ является q-ичным 1-совершенным кодом длины $(q-1)nm+n+m$.*

Доказательство. Поскольку $\boldsymbol{u} \in \mathcal{B}_{j_0} \subseteq \mathbb{F}_q^n$ и $\boldsymbol{v}_i \in \mathcal{C}_{j_i}^{r_i} \subseteq \mathbb{F}_q^{(q+1)n+1}$, то

$$n + \big((q+1)n+1\big)m = (q+1)nm + n + m,$$

и поскольку $n = \dfrac{q^{s_1}-1}{q-1}$ и $m = \dfrac{q^{s_2}-1}{q-1}$, то

$$(q-1)nm + n + m = \frac{q^{s_1+s_2}-1}{q-1}.$$

Следовательно, код $\mathcal{H}$ имеет корректную длину.

Далее покажем, что код $\mathcal{H}$ содержит корректное число кодовых слов и его минимальное расстояние равно 3.

Поскольку разбиение $\mathcal{C}_0^k, \mathcal{C}_1^k, \ldots, \mathcal{C}_{(q-1)n}^k$ k-го смежного класса, образованного кодом $RM_q((q-1)s_1-1, s_1)$, содержит $(q-1)n+1$ элементов и размерность кода $RM_q((q-1)s_1-1, s_1)$ равна

$$q^{s_1-1} = (q-1)\frac{q^{s_1}-1}{q-1} = (q-1)n,$$

то

$$|\mathcal{C}_{j_i}^{r_i}|((q-1)n+1) = q^{(q-1)n}.$$

По определению $|\mathcal{B}_{j_0}| = q^{n-s_1}$ для каждого $j_0 \in \{0, 1, \ldots, (q-1)n\}$. Следовательно, для каждого $\boldsymbol{r} \in \mathcal{R}$ мы можем построить

$$|\mathcal{B}_{j_0}||\mathcal{C}_{j_i}^{r_i}|^m((q-1)n+1)^m = q^{n-s_1}q^{(q-1)nm} = q^{(q-1)nm+n-s_1}$$

кодовых слов. Поскольку мощность q-ичного 1-совершенного кода $|\mathcal{R}| = q^{m-s_2}$, то

$$|\mathcal{H}| = q^{(q-1)nm+n+m-(s_1+s_2)}.$$

Следовательно, код $\mathcal{H}$ содержит корректное число кодовых слов.

Теперь покажем, что минимальное расстояние кода $\mathcal{H}$ равно 3. Предположим, что векторы $\boldsymbol{x} = (\boldsymbol{x}_0 \,|\, \boldsymbol{x}_1 \,|\, \ldots \,|\, \boldsymbol{x}_m)$ и $\boldsymbol{y} = (\boldsymbol{y}_0 \,|\, \boldsymbol{y}_1 \,|\, \ldots \,|\, \boldsymbol{y}_m)$ принадлежат коду $\mathcal{H}$. Тогда

$$d(\boldsymbol{x}, \boldsymbol{y}) \geqslant \sum_{i=0}^{m} d(\boldsymbol{x}_i, \boldsymbol{y}_i),$$

где векторы $\boldsymbol{x}_0, \boldsymbol{y}_0$ имеют длину n, а при $1 \leqslant i \leqslant m$ векторы $\boldsymbol{x}_i, \boldsymbol{y}_i$ имеют длину $(q-1)n+1$. Пусть $r_i = p(\boldsymbol{x}_i)$ и $r_i' = p(\boldsymbol{y}_i)$, $i = 1, 2, \ldots, m$, где $p(\boldsymbol{x})$ – функция четности. Тогда векторы $\boldsymbol{r} = (r_1, r_2, \ldots, r_m)$ и $\boldsymbol{r}' = (r_1', r_2', \ldots, r_m')$ принадлежат коду $\mathcal{R}$.

Если $d(\boldsymbol{x}_i, \boldsymbol{y}_i) = 0$ для всех значений i, то $\boldsymbol{r} = \boldsymbol{r}'$. Если $r_i \neq r_i'$, то $d(\boldsymbol{x}_i, \boldsymbol{y}_i) \geqslant 1$. Следовательно, если $d(\boldsymbol{r}, \boldsymbol{r}') \geqslant 3$, то $d(\boldsymbol{x}_i, \boldsymbol{y}_i) \geqslant 1$ как минимум для трех значений i. Таким образом,

$$\sum_{i=0}^{m} d(\boldsymbol{x}_i, \boldsymbol{y}_i) \geqslant 3 \quad \text{при } \boldsymbol{r} \neq \boldsymbol{r}'.$$

Если $\boldsymbol{r} = \boldsymbol{r}'$, то $p(\boldsymbol{x}_i) = p(\boldsymbol{y}_i)$ и $d(\boldsymbol{x}_i, \boldsymbol{y}_i) \geqslant 2$ при $\boldsymbol{x}_i \neq \boldsymbol{y}_i$, $i = 1, 2, \ldots, m$, и $d(\boldsymbol{x}_0, \boldsymbol{y}_0) \geqslant 1$ при $\boldsymbol{x}_0 \neq \boldsymbol{y}_0$. Допустим, что $\boldsymbol{x}_0 \in \mathcal{B}_{j_0}$, $\boldsymbol{y}_0 \in \mathcal{B}_{k_0}$, $\boldsymbol{x}_i \in \mathcal{C}_{j_i}^{r_i}$ и $\boldsymbol{y}_i \in \mathcal{C}_{k_i}^{r_i}$,

где $i = 1, 2, \ldots, m$. Тогда равенство $d(\boldsymbol{x}_i, \boldsymbol{y}_i) = 0$ означает, что $j_i = k_i$, $i = 0, 1, \ldots, m$. Поскольку $\boldsymbol{j} = (j_0, j_1, \ldots, j_m)$ и $\boldsymbol{k} = (k_0, k_1, \ldots, k_m)$ могут совпадать только в $m-1$ позициях, то мы получаем, что хотя бы для одного значения $i \in \{0, 1, \ldots, m\}$ справедливо неравенство $d(\boldsymbol{x}_i, \boldsymbol{y}_i) \geqslant 2$, а для некоторого другого значения i справедливо неравенство $d(\boldsymbol{x}_i, \boldsymbol{y}_i) \geqslant 1$. Таким образом, $d(\boldsymbol{x}, \boldsymbol{y}) \geqslant 3$ за исключением случая, когда $\boldsymbol{j} = \boldsymbol{k}$. Однако в этом случае, если $\boldsymbol{x}_i \neq \boldsymbol{y}_i$, то в силу утверждения 1 справедливо неравенство $d(\boldsymbol{x}_i, \boldsymbol{y}_i) \geqslant 3$, и следовательно, $d(\boldsymbol{x}, \boldsymbol{y}) \geqslant 3$. ▲

При $m = 1$ конструкция (6) сводится к комбинаторной конструкции q-ичных 1-совершенных кодов [8].

Теорема 2. *Число неэквивалентных q-ичных 1-совершенных кодов длины $(q^{s_1+s_2}-1)/(q-1)$ превосходит*

$$q^{s_1 q^{s_1}(q^{s_1}-1)(q^{s_2-1}-2)q^{q^{s_2-1}-s_2-1}-(s_1+s_2+2)q^{s_1+s_2}+(s_1+s_2+1)},$$

где величина $(q^{s_1+s_2}-1)/(q-1)$ является достаточно большой.

Доказательство. Обозначим через $Q(m, (q-1)n+1)$ множество всех m-арных квазигрупп порядка $(q-1)n+1$. Множество $Q(m, (q-1)n+1)$ можно также рассматривать как множество $((q-1)n+1)$-ичных кодов с параметрами

$$\big(m+1, ((q-1)n+1)^m, 2\big)_{(q-1)n+1}.$$

В [10] показано, что при достаточно больших n справедливо неравенство

$$|Q(m, (q-1)n+1)| \geqslant ((q-1)n+1)^{\left[((q-1)n+1)^2-((q-1)n+1)\right](m-1)}. \tag{7}$$

Далее рассмотрим вышеописанную конструкцию 1-совершенных кодов. Поскольку для каждого кодового слова $\boldsymbol{r} \in \mathcal{R}$ мы можем выбрать любую m-арную квазигруппу $q_{\boldsymbol{r}} \in Q(m, (q-1)n+1)$, то исходя из одного фиксированного разбиения пространства $\mathbb{F}_q^n$ на коды $\mathcal{B}_0, \mathcal{B}_1, \ldots, \mathcal{B}_{(q-1)n}$ и одного фиксированного разбиения пространства $\mathbb{F}_q^{(q-1)n+1}$ на коды $\mathcal{C}_0^k, \mathcal{C}_1^k, \ldots, \mathcal{C}_{(q-1)n}^k$ мы можем построить

$$|Q(m, (q-1)n+1)|^{|\mathcal{R}|} \tag{8}$$

различных q-ичных 1-совершенных кодов длины $(q^{s_1+s_2}-1)/(q-1)$.

Так как $(q-1)n+1 = q^{s_1}$ и $(q-1)m+1 = q^{s_2}$, то в силу формулы (7) получаем, что

$$|Q(m, (q-1)n+1)|^{|\mathcal{R}|} \geqslant \left(\left(q^{s_1}\right)^{\left[(q^{s_1})^2-q^{s_1}\right]\left(q^{s_2-1}-2\right)}\right)^{|\mathcal{R}|}.$$

Поскольку $\mathcal{R}$ является q-ичным 1-совершенным кодом длины $(q^{s_2}-1)/(q-1)$, то

$$|\mathcal{R}| = q^{\frac{q^{s_2}-1}{q-1}-s_2},$$

и так как

$$q^{q^{s_2-1}-s_2-1} < q^{\frac{q^{s_2}-1}{q-1}-s_2},$$

то

$$|Q(m, (q-1)n+1)|^{|\mathcal{R}|} \geqslant \left(q^{s_1}\right)^{\left[(q^{s_1})^2-q^{s_1}\right]\left(q^{s_2-1}-2\right)q^{q^{s_2-1}-s_2-1}}. \tag{9}$$

Множество различных q-ичных 1-совершенных кодов длины $(q^{s_1+s_2}-1)/(q-1)$ разбивается на классы эквивалентности. Произвольный класс эквивалентности в

этом разбиении содержит не более чем

$$(q-1)^{\frac{q^{s_1+s_2-1}}{q-1}}\left(\frac{q^{s_1+s_2}-1}{q-1}\right)!\,q^{\frac{q^{s_1+s_2-1}}{q-1}}$$

различных q-ичных 1-совершенных кодов длины $(q^{s_1+s_2}-1)/(q-1)$, где

$$(q-1)^{\frac{q^{s_1+s_2-1}}{q-1}}\left(\frac{q^{s_1+s_2}-1}{q-1}\right)!$$

– число мономиальных матриц размера $\dfrac{q^{s_1+s_2}-1}{q-1}\times\dfrac{q^{s_1+s_2}-1}{q-1}$ над полем $\mathbb{F}_q$, а

$$q^{\frac{q^{s_1+s_2-1}}{q-1}}$$

– число векторов в пространстве $\mathbb{F}_q^{\frac{q^{s_1+s_2-1}}{q-1}}$. Если коды из множества различных q-ичных 1-совершенных кодов длины $(q^{s_1+s_2}-1)/(q-1)$ содержат нулевой вектор $\mathbf{0}$, то произвольный класс эквивалентности содержит не более чем

$$(q-1)^{\frac{q^{s_1+s_2-1}}{q-1}}\left(\frac{q^{s_1+s_2}-1}{q-1}\right)!\,q^{\frac{q^{s_1+s_2-1}}{q-1}-(s_1+s_2)}$$

различных q-ичных 1-совершенных кодов длины $(q^{s_1+s_2}-1)/(q-1)$, где

$$q^{\frac{q^{s_1+s_2-1}}{q-1}-(s_1+s_2)}$$

– число слов в q-ичном 1-совершенном коде длины $(q^{s_1+s_2}-1)/(q-1)$.

В [10] показано, что неравенство (7) справедливо и для квазигрупп q, таких что $q(0,0,\ldots,0)=0$. Поскольку мы предполагаем, что нулевой вектор $\mathbf{0}$ всегда принадлежит коду, то коды $\mathcal{B}_0$, $\mathcal{C}_0^0$ и $\mathcal{R}$ содержат вектор $\mathbf{0}$. Следовательно, мы можем предполагать, что q-ичные 1-совершенные коды длины $(q^{s_1+s_2}-1)/(q-1)$, построенные вышеописанным методом, также содержат вектор $\mathbf{0}$.

Справедливость теоремы 2 вытекает из неравенства (9) и следующих неравенств:

$$q^{\frac{q^{s_1+s_2-1}}{q-1}-(s_1+s_2)}\leqslant q^{q^{s_1+s_2}-(s_1+s_2+1)},$$

$$(q-1)^{\frac{q^{s_1+s_2-1}}{q-1}}\left(\frac{q^{s_1+s_2}-1}{q-1}\right)!\,q^{q^{s_1+s_2}-(s_1+s_2+1)}\leqslant$$

$$\leqslant(q-1)^{q^{s_1+s_2}}(q^{s_1+s_2})!\cdot q^{q^{s_1+s_2}-(s_1+s_2+1)},$$

$$(q-1)^{q^{s_1+s_2}}(q^{s_1+s_2})!\,q^{q^{s_1+s_2}-(s_1+s_2+1)}\leqslant q^{q^{s_1+s_2}+(s_1+s_2)q^{s_1+s_2}+q^{s_1+s_2}-(s_1+s_2+1)}.\quad\blacktriangle$$

Следствие 1. *Число неэквивалентных q-ичных 1-совершенных кодов длины $n=(q^s-1)/(q-1)$ превосходит $q^{q^{cn}}$, где константа $c<1$, а s – достаточно большое натуральное число.*

В [10] Фелпс предложил конструкцию расширенных двоичных 1-совершенных кодов, т.е. кодов с параметрами расширенных двоичных кодов Хэмминга (кодов Рида – Маллера). Далее мы представим модификацию конструкции (6), которая позволяет строить коды с параметрами аффинных кодов Рида – Маллера и является прямым обобщением конструкции Фелпса [10] на q-ичный случай.

1. Пусть пространство $\mathbb{F}_q^{(q-1)n+1}$ разбито на смежные классы, образованные аффинным кодом Рида – Маллера $RM_q((q-1)s_1-1,s_1)$ порядка $(q-1)s_1-1$. При этом $n=\dfrac{q^{s_1}-1}{q-1}$, $(q-1)n+1=q^{s_1}$ и $s_1\geqslant 2$.

2. Пусть $\mathcal{C}_0^k, \mathcal{C}_1^k, \ldots, \mathcal{C}_{(q-1)n}^k$ – разбиение k-го смежного класса на квазисовершенные коды с параметрами аффинных кодов Рида – Маллера порядка $(q-1)s_1-2$, коды могут быть как линейные, так и нелинейные, $0 \leqslant k \leqslant q-1$.
3. Пусть $\mathcal{R}$ – четный квазисовершенный код с параметрами аффинного кода Рида – Маллера $RM_q((q-1)s_2-2, s_2)$ порядка $(q-1)s_2-2$, $s_2 \geqslant 2$.
4. Пусть $m = \dfrac{q^{s_2}-1}{q-1}$. Для каждого кодового слова $\boldsymbol{r} \in \mathcal{R}$ определим $(q-1)m$-арную квазигруппу $q_{\boldsymbol{r}}$ порядка $(q-1)n+1$. Квазигруппа определяется на множестве индексов $\{0,1,\ldots,(q-1)n\}$. В этом случае квазигруппу $q_{\boldsymbol{r}}$ можно рассматривать как код длины $(q-1)m+1$ с минимальным расстоянием 2 над алфавитом, состоящим из $(q-1)n+1$ элементов.

Тогда образуем q-ичный код $\mathcal{P}$. Положим

$$\mathcal{P} = \Big\{(\boldsymbol{c}_0 \,|\, \boldsymbol{c}_1 \,|\, \boldsymbol{c}_2 \,|\, \ldots \,|\, \boldsymbol{c}_{(q-1)m}) \mid \boldsymbol{c}_i \in \mathcal{C}_{j_i}^{r_i},$$
$$\boldsymbol{r} = (r_0, r_1, \ldots, r_{(q-1)m}) \in \mathcal{R},\ j_0 = q_{\boldsymbol{r}}(j_1, j_2, \ldots, j_{(q-1)m}),$$
$$j_i \in \{0,1,\ldots,(q-1)n\},\ i = 0,1,\ldots,(q-1)m\Big\}. \tag{10}$$

Теорема 3. *Построенный выше код $\mathcal{P}$ является q-ичным квазисовершенным кодом с параметрами аффинного кода Рида – Маллера порядка $(q-1)(s_1+s_2)-2$.*

Доказательство. При $q=2$ теорема 3 доказана в [10]. Докажем теорему 3 при $q \geqslant 3$.

В силу утверждений 1, 2 аффинный код Рида – Маллера $RM_q((q-1)(s_1+s_2)-2, s_1+s_2)$ порядка $(q-1)(s_1+s_2)-2$ имеет следующие параметры:

1. Длина кода Рида – Маллера $RM_q((q-1)(s_1+s_2)-2, s_1+s_2)$ равна $q^{s_1+s_2}$;
2. Число кодовых слов в коде $RM_q((q-1)(s_1+s_2)-2, s_1+s_2)$ равно $q^{q^{s_1+s_2}-(s_1+s_2)-1}$;
3. Минимальное расстояние $d(RM_q((q-1)(s_1+s_2)-2, s_1+s_2)) = 3$;
4. Радиус покрытия $\rho(RM_q((q-1)(s_1+s_2)-2, s_1+s_2)) = 2$.

Покажем, что код $\mathcal{P}$ имеет такие же параметры.

Поскольку $(q-1)n+1 = q^{s_1}$ и $(q-1)m+1 = q^{s_2}$, то

$$((q-1)n+1)((q-1)m+1) = q^{s_1+s_2}.$$

Следовательно, длина кода $\mathcal{P}$ равна $q^{s_1+s_2}$.

Поскольку размерность аффинного кода Рида – Маллера $RM_q((q-1)s_1-1, s_1)$ порядка $(q-1)s_1-1$ равна $q^{s_1-1} = (q-1)n$ и каждый смежный класс, образованный этим кодом, разбивается на $(q-1)n+1$ подкодов $\mathcal{C}_{j_i}^{r_i}$, то получаем, что

$$|\mathcal{C}_{j_i}^{r_i}|((q-1)n+1) = q^{(q-1)n}.$$

Так как $|\mathcal{C}_{j_i}^{r_i}| = q^{(q-1)n-s_1}$ при $j_i \in \{0,1,\ldots,(q-1)n\}$, $i = 0,1,\ldots,(q-1)m$, то для каждого $\boldsymbol{r} \in \mathcal{R}$ мы можем построить

$$|\mathcal{C}_{j_0}^{r_0}||\mathcal{C}_{j_i}^{r_i}|^{(q-1)m}((q-1)n+1)^{(q-1)m} = q^{(q-1)n(q-1)m+(q-1)n-s_1}$$

кодовых слов.

Поскольку $|\mathcal{R}| = q^{(q-1)m-s_2}$, то

$$|\mathcal{P}| = q^{(q-1)n(q-1)m+(q-1)n+(q-1)m+1-(s_1+s_2)-1} = q^{((q-1)n+1)((q-1)m+1)-(s_1+s_2)-1}.$$

Следовательно,

$$|\mathcal{P}| = q^{q^{s_1+s_2}-(s_1+s_2)-1}.$$

Далее мы покажем, что минимальное расстояние кода $\mathcal{P}$ равно 3. Пусть векторы $\boldsymbol{x} = (\boldsymbol{x}_0 \,|\, \boldsymbol{x}_1 \,|\, \ldots \,|\, \boldsymbol{x}_{(q-1)m})$ и $\boldsymbol{y} = (\boldsymbol{y}_0 \,|\, \boldsymbol{y}_1 \,|\, \ldots \,|\, \boldsymbol{y}_{(q-1)m})$ принадлежат $\mathcal{P}$. Тогда

$$d(\boldsymbol{x}, \boldsymbol{y}) \geqslant \sum_{i=0}^{(q-1)m} d(\boldsymbol{x}_i, \boldsymbol{y}_i),$$

где векторы $\boldsymbol{x}_i, \boldsymbol{y}_i$ имеют длину $(q-1)n+1$ при $0 \leqslant i \leqslant (q-1)m$. Пусть $r_i = p(\boldsymbol{x}_i)$ и $r'_i = p(\boldsymbol{y}_i)$, где $p(\boldsymbol{x})$ – функция четности и $i = 1, 2, \ldots, (q-1)m$. Тогда векторы $\boldsymbol{r} = (r_1, r_2, \ldots, r_{(q-1)m})$ и $\boldsymbol{r}' = (r'_1, r'_2, \ldots, r'_{(q-1)m})$ принадлежат коду $\mathcal{R}$. Если $d(\boldsymbol{x}_i, \boldsymbol{y}_i) = 0$, то $\boldsymbol{r} = \boldsymbol{r}'$. Если $r_i \neq r'_i$, то $d(\boldsymbol{x}_i, \boldsymbol{y}_i) \geqslant 1$. Поэтому если $d(\boldsymbol{r}, \boldsymbol{r}') \geqslant 3$, то $d(\boldsymbol{x}_i, \boldsymbol{y}_i) \geqslant 1$ для трех значений i.

Таким образом,

$$\sum_{i=0}^{(q-1)m} d(\boldsymbol{x}_i, \boldsymbol{y}_i) \geqslant 3 \quad \text{при } \boldsymbol{r} \neq \boldsymbol{r}'.$$

Если $\boldsymbol{r} = \boldsymbol{r}'$, то $p(\boldsymbol{x}_i) = p(\boldsymbol{y}_i)$ и $d(\boldsymbol{x}_i, \boldsymbol{y}_i) \geqslant 2$ при $\boldsymbol{x}_i \neq \boldsymbol{y}_i$. Допустим,что $\boldsymbol{x}_0 \in \mathcal{C}_{j_0}$, $\boldsymbol{y}_0 \in \mathcal{C}_{k_0}$, $\boldsymbol{x}_i \in \mathcal{C}^{r_i}_{j_i}$ и $\boldsymbol{x}_i \in \mathcal{C}^{r_i}_{k_i}$, где $i = 1, 2, \ldots, (q-1)m$. Тогда равенство $d(\boldsymbol{x}_i, \boldsymbol{y}_i) = 0$ означает, что $j_i = k_i$ для всех $i = 0, 1, \ldots, (q-1)m$. Поскольку $\boldsymbol{j} = (j_0, j_1, \ldots, j_{(q-1)m})$ и $\boldsymbol{k} = (k_0, k_1, \ldots, k_{(q-1)m})$ могут совпадать только в $(q-1)m-1$ позициях, то $d(\boldsymbol{x}_i, \boldsymbol{y}_i) \geqslant 2$ по крайней мере для двух значений i. Следовательно, $d(\boldsymbol{x}, \boldsymbol{y}) \geqslant 3$ за исключением случая, когда $\boldsymbol{j} = \boldsymbol{k}$. Однако в этом случае, если $\boldsymbol{x}_i \neq \boldsymbol{y}_i$, то в силу утверждения 1 имеем $d(\boldsymbol{x}_i, \boldsymbol{y}_i) \geqslant 3$, и следовательно, $d(\boldsymbol{x}, \boldsymbol{y}) \geqslant 3$.

Остается показать, что радиус покрытия кода $\mathcal{P}$ равен 2. В векторном пространстве размерности $q^{s_1+s_2} = ((q-1)n+1)((q-1)m+1)$ над полем $\mathbb{F}_q$ рассмотрим произвольный вектор $\boldsymbol{x} = (\boldsymbol{x}_0 \,|\, \boldsymbol{x}_1 \,|\, \ldots \,|\, \boldsymbol{x}_{(q-1)m})$. Код $\mathcal{R}$ – квазисовершенный код с параметрами аффинного кода Рида – Маллера порядка $(q-1)s_2-2$. Следовательно, радиус покрытия кода $\mathcal{R}$ равен 2 и существует кодовое слово $(\boldsymbol{c}_0 \,|\, \boldsymbol{c}_1 \,|\, \boldsymbol{c}_2 \,|\, \ldots \,|\, \boldsymbol{c}_{(q-1)m}) \in \mathcal{P}$, такое что

$$d\Big((p(\boldsymbol{x}_0) \,|\, p(\boldsymbol{x}_1) \,|\, \ldots \,|\, p(\boldsymbol{x}_{(q-1)m})), \big(p(\boldsymbol{c}_0) \,|\, p(\boldsymbol{c}_1) \,|\, p(\boldsymbol{c}_2) \,|\, \ldots \,|\, p(\boldsymbol{c}_{(q-1)m})\big)\Big) \leqslant 2.$$

Без ограничения общности мы можем полагать, что кодовое слово $(\boldsymbol{c}_0 \,|\, \boldsymbol{c}_1 \,|\, \boldsymbol{c}_2 \,|\, \ldots \,|\, \boldsymbol{c}_{(q-1)m})$ является нулевым, а вектор $\boldsymbol{x}$ таков, что $\boldsymbol{x}_2 = \boldsymbol{0}, \ldots, \boldsymbol{x}_{(q-1)m} = \boldsymbol{0}$. Тогда вектор $\boldsymbol{x} = \big(\boldsymbol{x}_0 \,|\, \boldsymbol{x}_1 \,|\, \ldots \,|\, \boldsymbol{x}_{(q-1)m}\big)$ принадлежит линейному подпространству

$$\Big\{\big(\boldsymbol{y}_0 \,|\, \boldsymbol{y}_1 \,|\, \boldsymbol{c}_2 \,|\, \ldots \,|\, \boldsymbol{c}_{(q-1)m}\big) \mid \boldsymbol{y}_0, \boldsymbol{y}_1 \in \mathbb{F}_q^{(q-1)n+1},\ \boldsymbol{c}_i = \boldsymbol{0},\ i = 2, 3, \ldots, (q-1)m\Big\}. \quad (11)$$

Если $\boldsymbol{y}_1 = \boldsymbol{0}$, то по построению кода $\mathcal{P}$ подпространство

$$\Big\{\big(\boldsymbol{y}_0 \,|\, \boldsymbol{0} \,|\, \boldsymbol{c}_2 \,|\, \ldots \,|\, \boldsymbol{c}_{(q-1)m}\big) \mid \boldsymbol{y}_0 \in \mathbb{F}_q^{(q-1)n+1},\ \boldsymbol{c}_i = \boldsymbol{0},\ i = 2, 3, \ldots, (q-1)m\Big\}$$

содержит код с параметрами кода $RM_q((q-1)s_1-2, s_1)$, радиус покрытия которого равен 2 (утверждение 2). Если $p(\boldsymbol{y}_1) = \boldsymbol{0}$ и вектор $\boldsymbol{y}_1$ является фиксированным,

тогда аналогичное утверждение справедливо и для всех линейных многообразий

$$\Big\{\big(\boldsymbol{y}_0 \,|\, \boldsymbol{y}_1 \,|\, \boldsymbol{c}_2 \,|\, \ldots \,|\, \boldsymbol{c}_{(q-1)m}\big) \mid \boldsymbol{y}_0, \boldsymbol{y}_1 \in \mathbb{F}_q^{(q-1)n+1},\ p(\boldsymbol{y}_1) = \boldsymbol{0},\ \boldsymbol{c}_i = \boldsymbol{0},\ i = 2, 3, \ldots, (q-1)m\Big\}.$$

Поскольку радиус покрытия аффинного кода Рида – Маллера $RM_q((q-1)s_1 - 1, s_1)$ порядка $(q-1)s_1 - 1$ равен 1, то по построению кода $\mathcal{P}$ для любого вектора из множества

$$\Big\{\big(\boldsymbol{y}_0 \,|\, \boldsymbol{y}_1 \,|\, \boldsymbol{c}_2 \,|\, \ldots \,|\, \boldsymbol{c}_{(q-1)m}\big) \mid \boldsymbol{y}_0, \boldsymbol{y}_1 \in \mathbb{F}_q^{(q-1)n+1},\ p(\boldsymbol{y}_1) \neq \boldsymbol{0},\ \boldsymbol{c}_i = \boldsymbol{0},\ i = 2, 3, \ldots, (q-1)m\Big\}$$

в подпространстве (11) существует кодовое слово из $\mathcal{P}$ на расстоянии не более 2. ▲

Утверждение 4. *Построенный выше код $\mathcal{P}$ является четным.*

Доказательство. Поскольку

$$p(\boldsymbol{x}) = p(p(\boldsymbol{x}_0), p(\boldsymbol{x}_1), \ldots, p(\boldsymbol{x}_{(q-1)m})) = p(r_0, r_1, \ldots, r_{(q-1)m}) = p(\boldsymbol{r})$$

и $p(\boldsymbol{r}) = 0$ для всех $\boldsymbol{r} \in \mathcal{R}$, то $p(\boldsymbol{x}) = 0$ для всех $\boldsymbol{x} \in \mathcal{P}$. ▲

Теорема 4. *Число неэквивалентных q-ичных четных квазисовершенных кодов с параметрами аффинного кода Рида – Маллера порядка $(q-1)(s_1+s_2)-2$ и длины $q^{s_1+s_2}$ превосходит*

$$q^{s_1 q^{s_1}(q^{s_1}-1)(q^{s_2}-2)q^{q^{s_2}-s_2-1}-(s_1+s_2+2)q^{s_1+s_2}+(s_1+s_2+1)},$$

где число $q^{s_1+s_2}$ является достаточно большим.

Доказательство. Рассмотрим множество $Q((q-1)m, (q-1)n+1)$ всех $(q-1)m$-арных квазигрупп порядка $(q-1)n+1$. Множество $Q((q-1)m, (q-1)n+1)$ можно также рассматривать как множество $((q-1)n+1)$-ичных кодов с параметрами

$$\Big((q-1)m+1, ((q-1)n+1)^{(q-1)m}, 2\Big)_{(q-1)n+1}.$$

В [10] показано, что при достаточно больших n справедливо неравенство

$$|Q((q-1)m, (q-1)n+1)| \geqslant ((q-1)n+1)^{\left[((q-1)n+1)^2-((q-1)n+1)\right]((q-1)m-1)}. \qquad (12)$$

Далее рассмотрим вышеописанную конструкцию квазисовершенных кодов. Поскольку для каждого кодового слова $\boldsymbol{r} \in \mathcal{R}$ мы можем выбрать любую квазигруппу $q_{\boldsymbol{r}} \in Q((q-1)m, (q-1)n+1)$, то исходя из одного фиксированного разбиения пространства $\mathbb{F}_q^{(q-1)n+1}$ на коды $\mathcal{C}_0^k, \mathcal{C}_1^k, \ldots, \mathcal{C}_{(q-1)n}^k$ мы можем построить более чем

$$|Q((q-1)m, (q-1)n+1)|^{|\mathcal{R}|}$$

различных q-ичных четных квазисовершенных кодов с параметрами аффинного кода Рида – Маллера порядка $(q-1)(s_1+s_2)-2$ и длины $q^{s_1+s_2}$.

Так как $(q-1)n+1 = q^{s_1}$ и $(q-1)m+1 = q^{s_2}$, то в силу формулы (12) получаем, что

$$|Q((q-1)m, (q-1)n+1)|^{|\mathcal{R}|} \geqslant \left(\left(q^{s_1}\right)^{\left[(q^{s_1})^2-q^{s_1}\right](q^{s_2}-2)}\right)^{|\mathcal{R}|}.$$

Из утверждения (1) следует, что $|\mathcal{R}| = q^{q^{s_2}-s_2-1}$. Следовательно,

$$|Q((q-1)m,(q-1)n+1)|^{|\mathcal{R}|} \geqslant \left(q^{s_1}\right)^{\left[(q^{s_1})^2-q^{s_1}\right](q^{s_2}-2)q^{q^{s_2}-s_2-1}}. \tag{13}$$

Множество различных q-ичных четных квазисовершенных кодов с параметрами аффинного кода Рида – Маллера порядка $(q-1)(s_1+s_2)-2$ и длины $q^{s_1+s_2}$ разбивается на классы эквивалентности. Справедливость теоремы 4 следует из формулы (13) и того факта, что произвольный класс эквивалентности в этом разбиении содержит не более чем

$$\left((q-1)^{q^{s_1+s_2}}(q^{s_1+s_2})!\right)\left(q^{q^{s_1+s_2}-(s_1+s_2+1)}\right) \leqslant q^{q^{s_1+s_2}+(s_1+s_2)q^{s_1+s_2}+q^{s_1+s_2}-(s_1+s_2+1)}$$

различных q-ичных четных квазисовершенных кодов с параметрами аффинного кода Рида – Маллера порядка $(q-1)(s_1+s_2)-2$ и длины $q^{s_1+s_2}$. ▲

Следствие 2. *Число неэквивалентных q-ичных четных квазисовершенных кодов с параметрами аффинного кода Рида – Маллера порядка $(q-1)s-2$ и длины $n=q^s$ превосходит $q^{q^{cn}}$, где константа $c<1$, а s – достаточно большое натуральное число.*

При $q=2,3$ существует единственный q-ичный код с параметрами аффинного кода Рида – Маллера $RM_q((q-1)s_1-1,s_1)$ [18, с. 224]. (Аффинный код Рида – Маллера $RM_q((q-1)s_1-1,s_1)$ является линейным q-ичным МДР-кодом с параметрами $[q^{s_1},q^{s_1}-1,2]_q$.)

Некоторая модификация конструкции (10) позволяет строить коды с параметрами аффинных кодов Рида – Маллера $RM_q((q-1)s_1-1,s_1)$.

Допустим, что код $\mathcal{R}$ является элементом разбиения кода $RM_q((q-1)s_2-1,s_2)$ на подкоды, имеющие параметры кода $RM_q((q-1)s_2-2,s_2)$. Такое разбиение содержит q^{s_2} элементов, поскольку код $RM_q((q-1)s_2-1,s_2)$ имеет параметры $[q^{s_2},q^{s_2}-1,2]_q$, а код $RM_q((q-1)s_2-1,s_2)$ в силу утверждения 1 имеет параметры $[q^{s_2},q^{s_2}-s_2-1,3]_q$. Разбиение кода $RM_q((q-1)s_1-1,s_1)$ на подкоды с параметрами кода $RM_q((q-1)s_1-2,s_1)$ содержит q^{s_1} элементов. Следовательно, каждый элемент разбиения кода $RM_q((q-1)s_2-1,s_2)$ на подкоды, имеющие параметры кода $RM_q((q-1)s_2-2,s_2)$, позволяет построить q^{s_1} попарно непересекающихся кодов с параметрами аффинного кода Рида – Маллера порядка $(q-1)(s_1+s_2)-2$. Следовательно, мы можем построить $q^{s_1+s_2}$ попарно непересекающихся кодов с параметрами аффинного кода Рида – Маллера порядка $(q-1)(s_1+s_2)-2$.

Таким образом, если код $\mathcal{R}$ является элементом разбиения аффинного кода Рида – Маллера $RM_q((q-1)s_2-1,s_2)$ на подкоды, имеющие параметры аффинного кода Рида – Маллера $RM_q((q-1)s_2-2,s_2)$, то и код $\mathcal{P}$ является элементом разбиения кода с параметрами аффинного кода Рида – Маллера $RM_q((q-1)(s_1+s_2)-1,s_1+s_2)$ на подкоды с параметрами аффинного кода Рида – Маллера порядка $(q-1)(s_1+s_2)-2$.

Автор искренне признателен рецензенту за полезные замечания и предложения, которые позволили существенно улучшить первоначальный вариант статьи.

СПИСОК ЛИТЕРАТУРЫ

1. *Мак-Вильямс Ф.Дж., Слоэн Н.Дж.А.* Теория кодов, исправляющих ошибки. М.: Связь, 1979.
2. *Kasami T., Lin S., Peterson W.* New Generalizations of the Reed–Muller Codes. I: Primitive Codes // IEEE Trans. Inform. Theory. 1968. V. 14. № 2. P. 189–199. https://doi.org/10.1109/TIT.1968.1054127
3. *Lachaud G.* The Parameters of Projective Reed–Müller Codes // Discrete Math. 1990. V. 81. № 2. P. 217–221. https://doi.org/10.1016/0012-365X(90)90155-B

4. *Sørensen A.B.* Projective Reed–Muller Codes // IEEE Trans. Inform. Theory. 1991. V. 37. № 6. P. 1567–1576. https://doi.org/10.1109/18.104317
5. *Berger T., Charpin P.* The Automorphism Group of Generalized Reed–Muller Codes // Discrete Math. 1993. V. 117. № 1–3. P. 1–17. https://doi.org/10.1016/0012-365X(93)90321-J
6. *Berger T.P.* Automorphism Groups of Homogeneous and Projective Reed–Muller Codes // IEEE Trans. Inform. Theory. 2002. V. 48. № 5. P. 1035–1045. https://doi.org/10.1109/18.995540
7. *Huffman W.C., Pless V.* Fundamentals of Error-Correcting Codes. Cambridge: Cambridge Univ. Press, 2003.
8. *Romanov A.M.* On Non-Full-Rank Perfect Codes over Finite Fields // Des. Codes Cryptogr. 2019. V. 87. № 5. P. 995–1003. https://doi.org/10.1007/s10623-018-0506-1
9. *Phelps K.T.* A Combinatorial Construction of Perfect Codes // SIAM J. Algebr. Discrete Methods. 1983. V. 4. № 3. P. 398–403. https://doi.org/10.1137/0604040
10. *Phelps K.T.* A General Product Construction for Error Correcting Codes // SIAM J. Algebr. Discrete Methods. 1984. V. 5. № 2. P. 224–228. https://doi.org/10.1137/0605023
11. *Зиновьев В.А.* Обобщенные каскадные коды // Пробл. передачи информ. 1976. Т. 12. № 1. С. 5–15. http://mi.mathnet.ru/ppi1670
12. *Боржес Ж., Рифа Ж., Зиновьев В.А.* О полностью регулярных кодах // Пробл. передачи информ. 2019. Т. 55. № 1. С. 3–50. https://doi.org/10.1134/S0134347519010017
13. *Assmus E.F., Key J.D.* Polynomial Codes and Finite Geometries // Handbook of Coding Theory. V. II. Amsterdam: Elsevier, 1998. P. 1269–1344.
14. *Delsarte P., Goethals J.-M., MacWilliams F.J.* On Generalized Reed–Muller Codes and Their Relatives // Inform. Control. 1970. V. 16. № 5. P. 403–442. https://doi.org/10.1016/S0019-9958(70)90214-7
15. *Delsarte P.* An Algebraic Approach to the Association Schemes of Coding Theory // Philips Res. Rep. Suppl. 1973. № 10.
16. *van Tilborg H.C.A.* Uniformly Packed Codes. PhD Thesis. Technische Hogeschool Eindhoven, Nederland, 1976.
17. *Goethals J.-M., van Tilborg H.C.A.* Uniformly Packed Codes // Philips Res. Rep. 1975. V. 30. P. 9–36.
18. *Laywine C.F., Mullen G.L.* Discrete Mathematics Using Latin Squares. New York: Wiley, 1998.

Романов Александр Михайлович
Институт математики им. С.Л. Соболева СО РАН,
Новосибирск
rom@math.nsc.ru

Поступила в редакцию
30.06.2020
После доработки
12.04.2021
Принята к публикации
04.06.2021

ПРОБЛЕМЫ ПЕРЕДАЧИ ИНФОРМАЦИИ

Том 57 **2021** **Вып. 3**

УДК 621.391 : 519.725

 Г. Маринге́р[1], Н.А. Полянский[2], И.В. Воробьев[3], Л. Вельтер[4]

КОДЫ С ОБРАТНОЙ СВЯЗЬЮ, ИСПРАВЛЯЮЩИЕ ВСТАВКИ И ВЫПАДЕНИЯ

Рассматривается задача передачи информации по комбинаторному каналу со вставками и выпадениями и обратной связью. Предположим, что отправитель передает n двоичных символов один за другим по каналу, в котором могут происходить вставки и выпадения. После передачи каждого символа отправитель узнает, какие вставки и выпадения произошли в канале, и адаптирует алгоритм кодирования. Целью статьи является разработка стратегии кодирования, позволяющей безошибочно передавать максимальное количество информации в предположении, что общее количество вставок и выпадений не превосходит τn для некоторой константы τ, $0 < \tau < 1$. Мы покажем, как эта задача может быть сведена к задаче передачи информации по каналу с замещениями. Таким образом, максимальная асимптотическая скорость кодов для комбинаторного канала со вставками и выпадениями с обратной связью полностью установлена. Вычисление максимальной асимптотической скорости кодов для комбинаторного канала с замещениями и обратной связью было частично выполнено Берлекэмпом и окончено позже Зигангировым. Однако доказательство Зигангирова нижней оценки скорости достаточно сложное. Мы возвращаемся к результату Зигангирова и представляем более подробное доказательство нижней границы.

Ключевые слова: кодирование с обратной связью, вставки и выпадения, асимптотическая скорость.

DOI: 10.31857/S0555292321030025

§ 1. Введение

Задача построения кодов, исправляющих вставки и выпадения, стала рассматриваться начиная с 1965 г., когда Левенштейн опубликовал свою работу [1], в которой показал, что всякий код исправляет t выпадений тогда и только тогда, когда он исправляет любую комбинацию из t выпадений и вставок. Более того, он показал, что коды Варшамова – Тененгольца [2] могут быть использованы для исправления одной вставки или выпадения. Примечательно, что большинство классических методов построения кодов, исправляющих замещения, не могут быть использованы для построения кодов для вставок и выпадений, поскольку данные виды ошибок приводят

[1] Работа выполнена при поддержке гранта немецкого научно-исследовательского сообщества (номер проекта WA3907/4-1).

[2] Работа выполнена при частичной поддержке гранта немецкого научно-исследовательского сообщества (номер проекта WA3907/1-1).

[3] Работа выполнена при частичной поддержке совместного гранта Российского фонда фундаментальных исследований (РФФИ) и Японского общества содействия науке (номер проекта 20-51-50007), а также гранта РФФИ (номер проекта 20-01-00559).

[4] Работа выполнена при поддержке гранта Европейского исследовательского совета из программы Horizon 2020 (номер проекта 801434).

к потери синхронизации между отправителем и получателем. Именно поэтому в последние годы были предложены новые подходы и развиты многочисленные методы построения кодов для комбинаторных и вероятностных каналов с выпадениями [3].

Сначала напомним некоторые известные результаты для комбинаторного канала с выпадениями без обратной связи. Пусть $\mathcal{C} \subseteq \{0,1\}^n$ является кодом, состоящим из 2^{Rn} двоичных кодовых слов длины n. Основная задача – построить код с максимально достижимой *скоростью* R, который может исправить долю τ комбинаторных ошибок типа выпадений, т.е. всевозможные комбинации τn выпадений.

Напомним, что корректирующую способность кода $\mathcal{C}$ можно определить через понятие наидлиннейшей общей подпоследовательности двух кодовых слов. Для того чтобы код исправлял долю τ ошибок, наидлиннейшая общая подпоследовательность между двумя различными кодовыми словами из $\mathcal{C}$ должна иметь длину менее $(1-\tau)n$. Очевидно, что при передаче произвольного слова через канал с долей $\tau \geqslant 1/2$ выпадений на выходе может получиться либо слово из всех единиц, либо слово из всех нулей. Таким образом, произвольный код $\mathcal{C}$ может содержать не более двух слов, и асимптотическая скорость для такого τ равна нулю. Первая конструкция кодов, исправляющих ненулевую долю ошибок типа выпадений и предполагающих эффективное кодирование и декодирование, была предложена в работе [4], авторы которой построили каскадный код, состоящий из недвоичного внешнего кода и хорошего двоичного кода, который может быть найден полным перебором. Последнее улучшение в данном направлении было получено в работах [5, 6], где было приведено семейство кодов с положительной скоростью для доли выпадений $\tau < \sqrt{2}-1 \approx 0{,}41$. Напомним, что для комбинаторного канала с ошибками типа замещений [7] нельзя построить коды с экспоненциально большим числом слов для доли ошибок $\tau \geqslant 0{,}25$. Насколько нам известно, не существует каких-либо других улучшений фундаментальных пределов для кодов, исправляющих выпадения. Таким образом, задача определения максимальной доли выпадений для кодов с положительной скоростью остается открытой в данной области.

Безошибочная обратная связь между отправителем и получателем может потенциально увеличить скорость кода, исправляющего долю выпадений τ. Модель с обратной связью для комбинаторных ошибок типа замещений была рассмотрена в работе Берлекэмпа [8]. В этой модели при передаче по каналу доля символов τ внутри передаваемого блока может поменяться на противоположные. Вдобавок после посылки очередного символа отправитель получает информацию о том, какой символ был получен на выходе канала с помощью безошибочной обратной связи. Таким образом, перед посылкой следующего символа отправитель может изменить свою стратегию на основании ранее полученных символов из канала. Максимальная асимптотическая скорость таких кодов была полностью охарактеризована Берлекэмпом [8] и Зигангировым [9]. Из их результатов следует, что скорость положительна при $\tau < 1/3$.

1.1. Обозначения. В этом пункте мы формально введем обозначения, которые будут использоваться на протяжении всей статьи. Множество целых чисел $\{1,\dots,n\}$ обозначим через $[n]$. Множество целых чисел $\{i+1, i+2, \dots, j\}$ будем кратко записывать через $(i,j]$. Под двоичным алфавитом будем подразумевать множество $\{0,1\}$. Двоичную строку длины n будем обозначать через $\boldsymbol{x} \in \{0,1\}^n$, где $\boldsymbol{x} = (x_1,\dots,x_n)$. Таким образом, символ $x_i \in \{0,1\}$ обозначает i-й элемент строки $\boldsymbol{x}$, где $i \in [n]$. Множество $\{0,1\}^*$ содержит все двоичные строки всевозможных длин и включает в себя пустую строку, которую мы обозначим через $(\ )$. Длина строки $\boldsymbol{x}$ обозначается символом $|\boldsymbol{x}|$. Для произвольных строк $\boldsymbol{x} \in \{0,1\}^{n_1}$ и $\boldsymbol{y} \in \{0,1\}^{n_2}$ обозначим конкатенацию строк $\boldsymbol{z} = (\boldsymbol{x} \,\|\, \boldsymbol{y})$, где $\boldsymbol{z} \in \{0,1\}^{n_1+n_2}$. Заметим, что двоичную строку $\boldsymbol{x} = (x_1,\dots,x_n) \in \{0,1\}^n$ можно также записать в виде $\boldsymbol{x} = (x_1 \,\|\, \dots \,\|\, x_n)$. Однако для n строк $\boldsymbol{y}_1,\dots,\boldsymbol{y}_n \in \{0,1\}^*$ мы разделяем конкатенацию строк, за-

писываемую в виде $\boldsymbol{y} = (\boldsymbol{y}_1 \,\|\, \ldots \,\|\, \boldsymbol{y}_n) \in \{0,1\}^*$, и упорядоченный набор строк $\widehat{\boldsymbol{y}} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n) \in (\{0,1\}^*)^n$. Набор строк $\widehat{\boldsymbol{y}}$ может быть единственным образом разложен на компоненты $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$, в то время как для строки $\boldsymbol{y}$, полученной в качестве конкатенации строк, это сделать невозможно.

Мы будем использовать две функции расстояния. Для двух произвольных строк $\boldsymbol{x} \in \{0,1\}^*$ и $\boldsymbol{y} \in \{0,1\}^*$ обозначим через $\Delta(\boldsymbol{x}, \boldsymbol{y})$ минимальное число выпадений и вставок, необходимых для получения $\boldsymbol{x}$ из $\boldsymbol{y}$. Мы также будем писать $d_H(\boldsymbol{x}, \boldsymbol{y})$ для обозначения расстояния Хэмминга между строками $\boldsymbol{x}$ и $\boldsymbol{y}$, имеющими одну длину. Все записываемые логарифмы берутся по основанию 2. Двоичная энтропия определяется как $h(x) := -x \log x - (1-x) \log(1-x)$.

1.2. Постановка задачи. В данной статье рассматривается задача передачи данных по комбинаторному каналу со вставками и выпадениями и обратной связью. Практический интерес к задаче обусловлен возможным применением к хранению данных с помощью ДНК последовательностей, где ошибки типа выпадений и вставок более распространены, чем замещения.

Для произвольного сообщения $m \in [M]$ отправитель кодирует m в двоичную строку $\boldsymbol{c} \in \{0,1\}^n$ таким образом, что после передачи строки по комбинаторному каналу со вставками и выпадениями получатель сможет декодировать исходное сообщение m по полученной строке $\boldsymbol{y}$. При передаче по каналу строки $\boldsymbol{c}$ могут происходить вставки и выпадения. Процесс кодирования и передачи сообщения состоит из n шагов, т.е. $\boldsymbol{c} = (c_1, \ldots, c_n)$. В момент времени i отправитель посылает новый двоичный символ $c_i \in \{0,1\}$ в канал. Внутри канала могут произойти вставки и выпадения. Полученную после внесения ошибок строку (возможно, пустую) обозначим через $\boldsymbol{y}_i \in \{0,1\}^*$. Отметим, что символы в процессе передачи по каналу могут быть вставлены в том числе и перед символом c_i. Получатель наблюдает только целую сконкатенированную строку $\boldsymbol{y} = (\boldsymbol{y}_1 \,\|\, \ldots \,\|\, \boldsymbol{y}_n)$, но не видит разбиения на строки $\boldsymbol{y}_i$. Иначе говоря, декодирующая функция $\mathrm{dec}(\boldsymbol{y})$ имеет следующий вид: $\mathrm{dec}\colon \{0,1\}^* \to [M]$. В момент времени $(i+1)$ отправитель корректирует свою стратегию отправки сообщения m, основываясь на упорядоченном наборе строк $\widehat{\boldsymbol{y}}^{(i)} := (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_i) \in (\{0,1\}^*)^i$. Таким образом, символ $c_{i+1}\colon [M] \times (\{0,1\}^*)^i \to \{0,1\}$ является функцией исходного сообщения m и набора строк $\widehat{\boldsymbol{y}}^{(i)}$, т.е. $c_{i+1} = c_{i+1}(m, \widehat{\boldsymbol{y}}^{(i)})$. Общее количество ошибок, которое может произойти в процессе передачи по каналу, ограничено величиной t, т.е.

$$\sum_{i=1}^{n} \Delta\big(c_i(m, \widehat{\boldsymbol{y}}^{(i-1)}), \boldsymbol{y}_i\big) \leqslant t. \tag{1}$$

Целью статьи является нахождение максимального количества сообщений, которое может быть передано отправителем таким образом, что получатель всегда может восстановить переданное сообщение на основе строки $\boldsymbol{y}$, которая получается из передаваемой строки с помощью не более чем t вставок и выпадений. Формально, пусть $M_{\mathrm{id}}(n,t)$ обозначает максимальное количество сообщений, которое отправитель может передать получателю в данной модели. Мы рассматриваем случай линейной зависимости числа ошибок t от длины передаваемого сообщения n, т.е. $t = \lfloor \tau n \rfloor$, $0 \leqslant \tau \leqslant 1$. Определим *максимальную асимптотическую скорость* кодов с обратной связью, исправляющих долю τ вставок и выпадений, как

$$R_{\mathrm{id}}(\tau) := \limsup_{n \to \infty} \frac{\log M_{\mathrm{id}}(n, \lfloor \tau n \rfloor)}{n}.$$

Отметим, что длина полученной строки может достигать длины вплоть до $n + \lfloor \tau n \rfloor$, в то время как число отправленных символов равно n. Однако асимптотическая

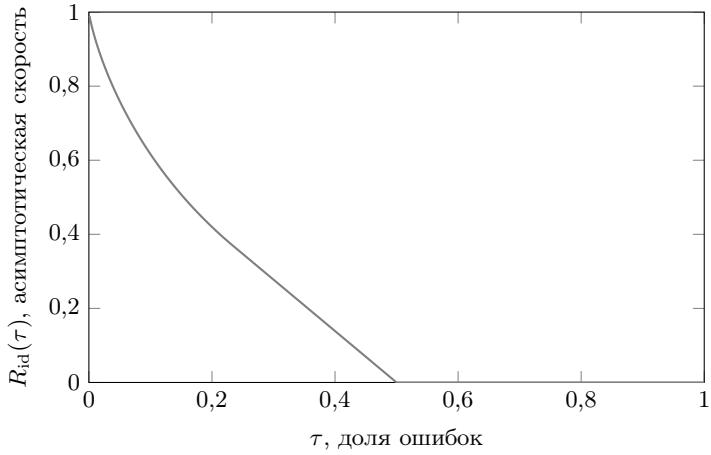

Рис. 1. Максимальная асимптотическая скорость двоичных кодов с обратной связью, исправляющих вставки и выпадения

скорость измеряется в количестве бит, нормированном на число переданных бит по каналу. В данной статье будет исследоваться величина $R_{\mathrm{id}}(\tau)$ для произвольного $\tau \in [0, 1]$.

1.3. Результаты статьи. В статье будет показано, что задача передачи информации по комбинаторному каналу со вставками и выпадениями может быть сведена к задаче передачи информации по комбинаторному каналу с замещениями. А именно, сначала мы доказываем, что любой алгоритм кодирования с n передачами по каналу с обратной связью, позволяющий исправить t вставок, может быть использован как алгоритм кодирования, исправляющий t замещений на длине $n + t$. Из этого результата будет выведена верхняя граница для $R_{\mathrm{id}}(\tau)$. Далее мы адаптируем алгоритм кодирования, изначально предложенный для канала с замещениями и обратной связью, для получения нижней границы на $R_{\mathrm{id}}(\tau)$. Эта нижняя граница совпадает с доказанной ранее верхней. Таким образом, будет получено точное значение $R_{\mathrm{id}}(\tau)$ для любой доли ошибок τ, $0 \leqslant \tau \leqslant 1$. Полученная асимптотическая скорость изображена на графике на рис. 1.

Теорема 1. *Максимальная асимптотическая скорость передачи информации по комбинаторному каналу со вставками и выпадениями и обратной связью описывается формулой*

$$R_{\mathrm{id}}(\tau) = \begin{cases} (1+\tau)\left(1 - h\left(\dfrac{\tau}{1+\tau}\right)\right) & \text{для } 0 \leqslant \tau \leqslant \sqrt{5} - 2, \\ (1-2\tau)\log\left(\dfrac{1+\sqrt{5}}{2}\right) & \text{для } \sqrt{5} - 2 < \tau \leqslant \dfrac{1}{2}, \\ 0 & \text{в противном случае.} \end{cases} \tag{2}$$

Вдобавок мы представим более подробную версию технического анализа Зигангирова [9] алгоритма Хорштейна [10] для комбинаторного канала с замещениями. Мы надеемся, что это сделает интуитивно понятный, но сложный для анализа алгоритм более доступным для широкой аудитории.

1.4. Структура статьи. Оставшаяся часть статьи организована следующим образом. В § 2 доказывается верхняя граница скорости $R_{\mathrm{id}}(\tau)$. В § 3 обсуждаются коды с обратной связью, исправляющие вставки и выпадения, и их сведение к кодам

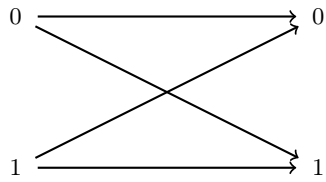

Рис. 2. Двоичный симметричный канал

с обратной связью, исправляющим замещения. Анализ кодов с обратной связью, исправляющих замещения, представлен в § 4. Статья завершается заключением в § 5.

§ 2. Верхняя граница на $R_{\mathrm{id}}(\tau)$

В этом параграфе мы вводим понятие комбинаторного канала с замещениями и доказываем, что коды с обратной связью, исправляющие долю τ вставок и выпадений, могут быть использованы как коды с обратной связью, исправляющие долю $\tau/(1+\tau)$ замещений. Это позволяет нам доказать верхнюю границу на максимальную достижимую скорость передачи по каналу со вставками и выпадениями, которая равна величине $R_{\mathrm{id}}(\tau)$, определенной в (2).

2.1. Комбинаторный канал с замещениями и обратной связью. В данной статье рассматривается только двоичный случай, поэтому и входной, и выходной алфавиты канала будут двоичными. Канал синхронизирован, и ошибкой мы называем событие, когда символ на выходе не равен соответствующему символу на входе (например, входной символ "0" заменяется на символ "1"). Передача одного бита изображена на рис. 2. Этот канал известен в теории кодирования как *двоичный симметричный канал.*

Пусть $[M] := \{1, \ldots, M\}$ обозначает множество сообщений. Задача отправителя состоит в безошибочной передаче получателю сообщения $m \in [M]$ вне зависимости от происходящих ошибок. Мы рассматриваем блочные коды с длиной блока n. Обозначим i-й символ, посылаемый отправителем в канал, через $c_i \in \{0,1\}$. Символ на выходе из канала обозначим через $y_i \in \{0,1\}$, причем $y_i \neq c_i$ в случае, если произошла ошибка. Перед посылкой i-го символа c_i в канал отправитель может скорректировать свою стратегию, основываясь на ранее принятых символах $\widehat{\boldsymbol{y}}^{(i-1)} := (y_1, \ldots, y_{i-1})$. Другими словами, $c_i \colon [M] \times \{0,1\}^{i-1} \to \{0,1\}$ является функцией от m и $\widehat{\boldsymbol{y}}^{(i-1)}$, т.е. $c_i = c_i(m, \widehat{\boldsymbol{y}}^{(i-1)})$. Отметим, что здесь $\widehat{\boldsymbol{y}}^{(i-1)}$ ничем не отличается от $\boldsymbol{y}^{(i-1)} := (y_1 \parallel \ldots \parallel y_{i-1}) \in \{0,1\}^{i-1}$, так как канал с замещениями синхронизирован. Основываясь на принятой из канала строке $\boldsymbol{y} := (y_1 \parallel \ldots \parallel y_n)$, получатель должен восстановить исходное передаваемое сообщение m. Другими словами, декодирующая функция $\mathrm{dec}(\boldsymbol{y})$ имеет вид $\mathrm{dec} \colon \{0,1\}^n \to [M]$. Мы ограничиваем величиной t общее количество ошибок, которое может произойти в канале, т.е. $d_H(\boldsymbol{c}, \boldsymbol{y}) \leqslant t$, где $\boldsymbol{c} := (c_1, \ldots, c_n)$. Пусть $M_s(n,t)$ равно максимальному количеству сообщений, которое отправитель может передать получателю безошибочно. Определим максимальную асимптотическую скорость кодов с обратной связью, исправляющих долю τ замещений, как

$$R_s(\tau) := \limsup_{n\to\infty} \frac{\log M_s(n, \lfloor \tau n \rfloor)}{n}.$$

2.2. Построение кодов, исправляющих замещения, из кодов, исправляющих вставки и выпадения. Для доказательства верхней оценки (2) максимальной скорости кодов с обратной связью, исправляющих вставки и выпадения, мы используем две

идеи. Сначала мы показываем, что код с обратной связью для вставок и выпадений может быть использован для исправления замещений.

Лемма 1. *Предположим, что нам дан код длины n и мощности M с обратной связью, способный исправлять t вставок и выпадений. Тогда существует код с обратной связью длины $n+t$ и мощности M, исправляющий t замещений.*

Далее воспользуемся верхней границей скорости $R_s(\tau)$, доказанной Берлекэмпом [8].

Теорема 2 (следствие из [8, гл. IV] и [9]). *Максимальная асимптотическая скорость для комбинаторного канала с замещениями удовлетворяет соотношению*

$$R_s(\tau) = \begin{cases} 1 - h(\tau) & \text{для } 0 \leqslant \tau \leqslant \dfrac{3-\sqrt{5}}{4}, \\ (1-3\tau)\log\left(\dfrac{1+\sqrt{5}}{2}\right) & \text{для } \dfrac{3-\sqrt{5}}{4} < \tau \leqslant \dfrac{1}{3}, \\ 0 & \text{в противном случае.} \end{cases}$$

Пользуясь леммой 1, мы получаем $M_s(n+t,t) \geqslant M_{\mathrm{id}}(n,t)$, и следовательно, $(1+\tau)R_s(\tau/(1+\tau)) \geqslant R_{\mathrm{id}}(\tau)$. Эта оценка и теорема 2 показывают, что асимптотическая скорость из теоремы 1 действительно является верхней границей на максимальную достижимую скорость комбинаторного канала со вставками и выпадениями. Таким образом, нам остается проверить справедливость леммы 1.

Доказательство леммы 1. Стратегии кодирования и декодирования данного кода с обратной связью, исправляющего вставки и выпадения, могут быть описаны функциями $c_i\colon [M] \times (\{0,1\}^*)^{i-1} \to \{0,1\}$ для $i \in [n]$ и $\mathrm{dec}\colon \{0,1\}^* \to [M]$. Мы определим кодирующую функцию $e_j\colon [M] \times \{0,1\}^{j-1} \to \{0,1\}$ для всех $j \in [n+t]$ для канала с замещениями, используя функции c_i в качестве вспомогательного инструмента. С помощью функций e_j отправитель сможет передать M сообщений по каналу с обратной связью и не более чем t замещениями на длине $n+t$.

Обозначим количество безошибочно полученных символов через i, количество переданных символов через j, индекс последнего правильно полученного символа через k, полученную на выходе канала строку через $\boldsymbol{z}$, упорядоченный набор полученных строк согласно выходу канала со вставками и выпадениями через $\widehat{\boldsymbol{y}}$.

Предлагаемая нами схема кодирования состоит из трех частей. Сначала инициализируются все счетчики и переменные, описывающие состояние схемы. Шаг 1 описывает процесс повторной передачи символа c_i, происходящей до тех пор, пока он не будет принят правильно. Шаг 2 используется лишь для того, чтобы передать нули в случае, если нами не была использована вся длина кода. Описывающие состояние алгоритма переменные i и k изменяются только на шаге 1.

Инициализация: Задаем $i \leftarrow 0$, $j \leftarrow 0$, $k \leftarrow 0$, $\boldsymbol{z} \leftarrow (\,)$ и $\widehat{\boldsymbol{y}} \leftarrow (\,)$.

Шаг 1: Определяем $e_{j+1}(m, \boldsymbol{z}) = c_{i+1}(m, \widehat{\boldsymbol{y}})$ и передаем этот символ. Обновляем $j \leftarrow j+1$. Пусть z_j – полученный символ. Обновляем $\boldsymbol{z} \leftarrow (\boldsymbol{z} \,\|\, z_j)$. Если $z_j \neq e_j(m, \boldsymbol{z})$, тогда повторяем шаг 1. Иначе, обновляем $i \leftarrow i+1$, определяем $\boldsymbol{y}_i \leftarrow \boldsymbol{z}_{[k+1,j]}$, обновляем $\widehat{\boldsymbol{y}} \leftarrow (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_i)$ и $k \leftarrow j$. Если $i = n$, то переходим к шагу 2. Иначе, повторяем шаг 1.

Шаг 2: Если $j = n+t$, обновляем $\boldsymbol{y}_n \leftarrow (\boldsymbol{y}_n \,\|\, \boldsymbol{z}_{[k+1,n+t]})$, $\widehat{\boldsymbol{y}} \leftarrow (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)$ и завершаем алгоритм. Иначе, определяем $e_{j+1}(m, \boldsymbol{z}) = 0$ и передаем его. Обновляем $j \leftarrow j+1$. Пусть полученный символ равен z_j. Обновляем $\boldsymbol{z} \leftarrow (\boldsymbol{z} \,\|\, z_j)$. Повторяем шаг 2.

Алгоритм кодирования успешно завершает свою работу после $n+t$ шагов, так как у нас будет не меньше n правильно принятых символов в блоке длины $n+t$.

В качестве алгоритма декодирования можно переиспользовать алгоритм для кода со вставками и выпадениями, так как полученная на выходе из канала строка $\boldsymbol{z}$ является возможным выходом и для канала со вставками и выпадениями с не более чем t ошибками при кодировании функциями c_i, $i \in [n]$. В самом деле, если ввести обозначение $\widehat{\boldsymbol{y}}^{(i)} := (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_i)$, то получится следующая цепочка равенств:

$$\sum_{i=1}^{n} \Delta\big(c_i(m, \widehat{\boldsymbol{y}}^{(i)}), \boldsymbol{y}_i\big) = \sum_{i=1}^{n} (|\boldsymbol{y}_i| - 1) = \sum_{i=1}^{n} |\boldsymbol{y}_i| - n = (n + t) - n = t.$$

Таким образом, декодирующая функция для кода со вставками и выпадениями выдаст правильное сообщение, т.е. $\operatorname{dec}(\boldsymbol{z}) = m$. ▲

§ 3. Нижняя оценка на $R_{\mathrm{id}}(\tau)$

В этом параграфе мы строим код с обратной связью, исправляющий вставки и выпадения и имеющий асимптотическую скорость, сколь угодно близкую к (2). Мы адаптируем алгоритм, изначально предложенный Хорштейном [10] и впоследствии доработанный Шалквейком [11] и Зигангировым [9]. Отметим, что этот же алгоритм уже использовался для комбинаторного канала с обратной связью и замещениями. Подчеркнем, что не любая кодирующая стратегия для комбинаторного канала с замещениями может быть так легко адаптирована для канала со вставками и выпадениями. Ключевым свойством алгоритма Хорштейна является то, что в случае ошибки отправитель перепосылает символ, посланный на прошлом шаге, до тех пор, пока он не будет принят безошибочно. Продемонстрируем важность этого свойства, рассмотрев общий вид кодирующего алгоритма для канала с обратной связью. Предположим, что в соответствии с ранее полученными символами, кодирующая стратегия предлагает послать по каналу символ $c_i = 0$, а в канале происходит вставка символа "1" перед c_i, т.е. $\boldsymbol{y}_i = 10$. Символ "1" в начале последовательности $\boldsymbol{y}_i$ можно интерпретировать как вставленный в конец строки $\boldsymbol{y}_{i-1}$. Для произвольного алгоритма кодирования возможна ситуация, при которой алгоритм кодирования предложит передавать символ $c_i = 1$ после наблюдения $\boldsymbol{y}_{i-1}$, что приводит к возможности создания двух ошибок ценой одной вставки. В алгоритме Хорштейна такое невозможно, что позволяет использовать этот алгоритм в модели со вставками и выпадениями.

3.1. Код с обратной связью, исправляющий вставки и выпадения. Предположим, что в канале может произойти не больше доли τ ошибок, и отправитель хочет передать $M = 2^{Rn}$ сообщений, где $R = R_{\mathrm{id}}(\tau) - \varepsilon$ и $\varepsilon > 0$. В зависимости от значений τ, ε и n отправитель и получатель выбирают два трансцендентных числа α и β, удовлетворяющих соотношению $\alpha + \beta = 2$. Для $\tau < \sqrt{5} - 2$ они выбирают α достаточно близким к $2\tau/(1 + \tau)$, а для $\sqrt{5} - 2 \leqslant \tau < 1/2$ параметр α выбирается в окрестности точки $(3 - \sqrt{5})/2$. Насколько близким должно быть α к указанным значениям, мы разъясняем в § 4. Далее отправитель делит отрезок $[0, 1]$ на M подотрезков длины $1/M$ и нумерует их элементами из $[M]$ слева направо. Длины этих M отрезков будут меняться в течение процесса передачи сообщения. Если отправитель планирует послать сообщение $m \in [M]$, тогда отрезок с номером m будем называть *истинным отрезком*. Пусть $T(i)$ обозначает истинный отрезок длины $t(i)$ после i-го шага. Обозначим объединение отрезков, находящихся слева и справа от отрезка $T(i)$ через $L(i)$ и $R(i)$. Теперь мы готовы к описанию процедур кодирования и декодирования предлагаемого алгоритма, который будем обозначать $\mathfrak{A}_{\mathrm{id}}(M, n, \alpha)$.

Замечание 1. Отметим, что так как трансцендентные числа всюду плотны в множестве действительных чисел, то мы можем приблизить ими любое действительное число с произвольной точностью.

Кодирование: В момент времени i отправитель проверяет, лежит ли центр истинного отрезка в полуинтервале $[0, 1/2)$. Если лежит, то по каналу передается символ $c_i = 0$. В противном случае посылается символ $c_i = 1$. Далее отправитель смотрит на принятую из канала строку $\boldsymbol{y}_i$ длины n_i и модифицирует длины всех M отрезков по следующему правилу. Отправитель перебирает все символы строки $\boldsymbol{y}_i = (y_{i,1}, \ldots, y_{i,n_i}) \in \{0,1\}^{n_i}$ слева направо. Если строка $\boldsymbol{y}_i$ пустая, то он переходит к передаче следующего символа. В противном случае, если $y_{i,j} = 0$, то длины всех отрезков, полностью лежащих внутри $[0, 1/2)$, умножаются на число β. Длины отрезков, полностью лежащих внутри $[1/2, 1]$, умножаются на α. Может существовать отрезок, содержащий точку $1/2$ как внутреннюю. Обозначим через x длину такого отрезка. Величина x может быть представлена в виде суммы $x = x_0 + x_1$, где x_0 – длина левой части отрезка, лежащей в $[0, 1/2)$. Тогда длина x этого отрезка заменяется длиной $\beta x_0 + \alpha x_1$. Если же полученный из канала символ $y_{i,j}$ равен 1, то отправитель умножает длины отрезков слева от точки $1/2$ на α, а отрезки с правой стороны увеличивает в β раз. По сравнению со случаем $y_{i,j} = 0$ роли чисел α и β меняются местами. В процессе передачи сообщения длины отрезков изменяются согласно алгоритму, однако порядок отрезков остается неизменным. В итоге отправитель делает n_i таких обновлений длин после передачи i-го символа. Легко заметить, что сумма длин отрезков всегда равна 1. Идея алгоритма состоит в том, что длина истинного отрезка постепенно увеличится, так что в конце процедуры он будет содержать точку $1/2$.

Сложность описанного алгоритма кодирования равна $\Theta(Mn)$. Однако из описания алгоритма ясно, что достаточно следить только за длинами отрезков $L(i)$, $T(i)$ и $R(i)$. Поэтому сложность может быть уменьшена до $\Theta(n)$.

Декодирование: После приема строки $\boldsymbol{y} = (y_1 \parallel \ldots \parallel y_{n'})$ длины $n' \in [n-t, n+t]$ получатель может восстановить точку зрения отправителя, обновляя длины отрезков тем же самым способом. После n' шагов отправитель находит сообщение, соответствующее отрезку, содержащему точку $1/2$. Сложность этого алгоритма декодирования равна $\Theta(Mn)$. Для уменьшения сложности получатель может обратить процедуру декодирования следующим образом. Перебирая символы строки $\boldsymbol{y}$ справа налево, можно отследить прообраз точки $1/2$. Более формально, пусть $P(n') := 1/2$. Для $j \in \{0, 1, \ldots, n'-1\}$ и $y_j = 0$ определим

$$P(j-1) := \begin{cases} \dfrac{P(j)}{\beta}, & \text{если } P(j) \leqslant \dfrac{\beta}{2}, \\ \dfrac{1}{2} + \dfrac{2P(j) - \beta}{2\alpha}, & \text{если } P(j) > \dfrac{\beta}{2}. \end{cases}$$

Для $y_j = 1$ пусть

$$P(j-1) := \begin{cases} \dfrac{P(j)}{\alpha}, & \text{если } P(j) \leqslant \dfrac{\alpha}{2}, \\ \dfrac{1}{2} + \dfrac{2P(j) - \alpha}{2\beta}, & \text{если } P(j) > \dfrac{\alpha}{2}. \end{cases}$$

В конце процесса декодирования получатель выберет сообщение $\widehat{m} := \lceil M \cdot P(0) \rceil$. Очевидно, что сложность такого упрощенного алгоритма декодирования также линейна, т.е. равна $\Theta(n)$.

3.2. Простое, но неправильное объяснение работы алгоритма. Приведем интуитивное объяснение, почему эта стратегия имеет смысл. В данном пункте мы рассматриваем только $\tau < \sqrt{5} - 2$, так как в этом случае

$$R_{\mathrm{id}}(\tau) = (1+\tau)\left(1 - h\left(\frac{\tau}{1+\tau}\right)\right).$$

Пусть $\alpha = 2\tau/(1+\tau)$. Предположим, что в процессе передачи количество раз, когда истинный отрезок содержит точку $1/2$, равно $o(n)$, а число вставок и выпадений равно $\tau_{\text{ins}}n + o(n)$ и $\tau_{\text{del}}n + o(n)$ соответственно, причем $\tau_{\text{ins}} + \tau_{\text{del}} \leqslant \tau$. Тогда после n шагов длина истинного отрезка $T(n)$ может быть ограничена снизу следующим образом:

$$
\begin{aligned}
&t(n) \geqslant M^{-1}\alpha^{\tau_{\text{ins}}n+o(n)}\beta^{(1-\tau_{\text{del}})n+o(n)} = \\
&= 2^{-(R_{\text{id}}(\tau)-\varepsilon)n+\tau_{\text{ins}}n\log(2\tau/(1+\tau))+(1-\tau_{\text{del}})n\log(2/(1+\tau))+o(n)} \geqslant \\
&\geqslant 2^{-(R_{\text{id}}(\tau)-\varepsilon)n+\tau n\log(2\tau/(1+\tau))+n\log(2/(1+\tau))+o(n)} = \\
&= 2^{\varepsilon n+o(n)}.
\end{aligned}
$$

В этой цепочке преобразований мы использовали тот факт, что $\tau_{\text{ins}}\log(2\tau/(1+\tau)) + (1-\tau_{\text{del}})\log(2/(1+\tau))$ при ограничении $\tau_{\text{ins}} + \tau_{\text{del}} \leqslant \tau$ достигает минимума при $\tau_{\text{ins}} = \tau$ и $\tau_{\text{del}} = 0$. Таким образом, в конце процедуры длина истинного отрезка будет очень велика, а сам отрезок будет содержать точку $1/2$. Однако на самом деле наше изначальное предположение неверно, и точка $1/2$ попадает в истинный отрезок гораздо чаще, чем $o(n)$ раз. Поэтому для доказательства корректности работы стратегии необходимы более сложные рассуждения.

3.3. Построение кодов, исправляющих вставки и выпадения, из кодов, исправляющих замещения. Алгоритм $\mathfrak{A}_{\text{id}}(M, n, \alpha)$, описанный в п. 3.1, был изначально предложен в очень похожем виде для канала с замещениями. Подчеркнем, что такой канал гарантирует синхронизацию между отправителем и получателем, т.е. после передачи символа c_i канал выдает ровно один символ y_i. Обозначим алгоритм кодирования для канала с замещениями, использующий коды длины n, как это описано в п. 3.1, через $\mathfrak{A}_s(M, n, \alpha)$. Формальное определение этого алгоритма для канала с замещениями будет дано в п. 4.1. В дальнейшем мы покажем, что используя алгоритм $\mathfrak{A}_{\text{id}}(M, n, \alpha)$, можно достичь скорости (2) при передаче данных по комбинаторному каналу со вставками и выпадениями. Для начала докажем следующую лемму.

Лемма 2. *Если для всех $n' \in [n-t, n+t]$ алгоритм $\mathfrak{A}_s(M, n', \alpha)$ может быть использован для безошибочной передачи произвольного сообщения $m \in [M]$ по комбинаторному каналу с обратной связью и не более чем $t' := \lfloor (n'-n+t)/2 \rfloor$ замещениями, то алгоритм $\mathfrak{A}_{\text{id}}(M, n, \alpha)$ может быть использован для корректной передачи любого сообщения $m \in [M]$ по каналу с обратной связью и не более чем t вставками и выпадениями.*

Доказательство. Пусть $\boldsymbol{y} = (\boldsymbol{y}_1 \,\|\, \ldots \,\|\, \boldsymbol{y}_n) = (y_1, \ldots, y_{n'})$ – возможная выходная строка длины $n' \in [n-t, n+t]$ при применении алгоритма $\mathfrak{A}_{\text{id}}(M, n, \alpha)$ для передачи сообщения $m \in [M]$ по комбинаторному каналу с не более чем t вставками и выпадениями. Мы покажем, что эта строка $\boldsymbol{y}$ также является возможным выходом канала при применении алгоритма $\mathfrak{A}_s(M, n', \alpha)$ для передачи того же сообщения $m \in [M]$ по комбинаторному каналу с не более чем $t' := \lfloor (n'-n+t)/2 \rfloor$ замещениями. Определим $\widehat{\boldsymbol{y}}^{(i-1)} := (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{i-1}) \in (\{0,1\}^*)^{i-1}$ и $\boldsymbol{y}^{(i-1)} := (y_1, \ldots, y_{i-1}) \in \{0,1\}^{i-1}$. Пусть $c_i = c_i(m, \widehat{\boldsymbol{y}}^{(i-1)})$ и $e_i = e_i(m, \boldsymbol{y}^{(i-1)})$ обозначают i-е биты, выданные алгоритмами кодирования кода для каналов с обратной связью и ошибками вида вставок и выпадений и ошибками-замещениями. Введем обозначение $\boldsymbol{e} := (e_1, \ldots, e_{n'})$. Теперь нам достаточно показать, что $d_H(\boldsymbol{e}, \boldsymbol{y}) \leqslant t'$. Пусть длина строки $\boldsymbol{y}_i$ равна n_i. Представим $\boldsymbol{e}$ в виде конкатенации $(\boldsymbol{e}_1 \,\|\, \ldots \,\|\, \boldsymbol{e}_n)$, так чтобы строка $\boldsymbol{e}_i$ имела длину n_i. Очевидно, что для "синхронизированных" моментов i и $N_i := n_1 + \ldots + n_{i-1} + 1$ оба алгоритма отправляют в канал одинаковые символы, т.е.

$$c_i(m, \widehat{\boldsymbol{y}}^{(i-1)}) = e_{N_i}(m, \boldsymbol{y}^{(N_i-1)}).$$

Из определения алгоритма следует, что если $y_j \neq e_j$, то $e_{j+1} = e_j$, т.е. в случае ошибки происходит переотправка символа. Таким образом,

$$d_H(\boldsymbol{e}_i, \boldsymbol{y}_i) \leqslant \begin{cases} n_i - 1, & \text{если } \Delta(c_i, \boldsymbol{y}_i) = n_i - 1, \\ n_i, & \text{если } \Delta(c_i, \boldsymbol{y}_i) = n_i + 1. \end{cases}$$

Отметим, что второй случай может произойти только в случае выпадения. Пусть общие количества вставок и выпадений равны t_{ins} и t_{del}. Напомним, что у нас имеются ограничения $t_{\text{ins}} + t_{\text{del}} \leqslant t$ и $n - t_{\text{del}} + t_{\text{ins}} = n'$. Отсюда следует, что $n' - n + 2t_{\text{del}} \leqslant t$ и $t_{\text{del}} \leqslant \lfloor (n - n' + t)/2 \rfloor$. В итоге получаем

$$d_H(\boldsymbol{e}, \boldsymbol{y}) = \sum_{i=1}^{n} d_H(\boldsymbol{e}_i, \boldsymbol{y}_i) \leqslant \sum_{i=1}^{n} n_i - \sum_{i=1}^{n} \mathbb{1}\{\Delta(c_i, \boldsymbol{y}_i) = n_i - 1\} \leqslant$$
$$\leqslant n' - (n - t_{\text{del}}) \leqslant t',$$

где $\mathbb{1}\{x = a\}$ обозначает индикатор события $x = a$. ▲

Теперь рассмотрим комбинаторный канал с замещениями. Для достижения максимальной асимптотической скорости комбинаторного канала с замещениями при доле ошибок $\tau \leqslant (3 - \sqrt{5})/4$ в качестве значения параметра α из алгоритма $\mathfrak{A}_s(M, n, \alpha)$ выбирается трансцендентное число, близкое к 2τ.

Лемма 3 (следует из [9]). *Пусть $\tau^* \leqslant (3 - \sqrt{5})/4$ является трансцендентным числом. Положим $\alpha^* = 2\tau^*$. Пусть $\mathfrak{A}_s(M, n, \alpha^*)$ используется для передачи данных по комбинаторному каналу с замещениями для различных τ. Тогда достижимая скорость как функция от различных τ выглядит как касательная к функции $R_s(\tau)$ в точке $(\tau^*, R_s(\tau^*))$.*

С помощью леммы 3 можно доказать следующее свойство монотонности.

Предложение 1. *Пусть $M_s(n, \tau, \alpha)$ обозначает максимальное количество сообщений M, которое может быть успешно передано алгоритмом $\mathfrak{A}_s(M, n, \alpha)$ по комбинаторному каналу с длиной блока n и не более чем τn замещениями. Пусть α и τ – трансцендентные числа, такие что $\alpha = 2\tau/(1+\tau)$ и $\alpha \leqslant (3 - \sqrt{5})/2$. Тогда выполняется следующее*:

$$\min_{k \in \{0,1,\ldots,\tau n\}} \log M_s\big(\lfloor n(1+\tau) \rfloor - 2k, \lfloor \tau n \rfloor - k, \alpha\big) =$$
$$= (1 + o(1)) \log M_s\big(\lfloor (1+\tau) n \rfloor, \lfloor \tau n \rfloor, \alpha\big) =$$
$$= n(1+\tau) R_s\Big(\frac{\tau}{1+\tau}\Big) + o(n).$$

Доказательство. Выбор $\alpha = 2\tau/(1+\tau)$ позволяет передавать с помощью алгоритма $\mathfrak{A}_s(M, \lfloor n(1+\tau) \rfloor, \alpha)$ количество сообщений, асимптотически эквивалентное максимальному возможному при условии, что число ошибок ограничено $\lfloor \tau n \rfloor$. Мы покажем, что несмотря на то что выбранное α, вероятно, не является оптимальным с точки зрения достижимой скорости для длины блока $\lfloor n(1+\tau) \rfloor - 2k$ и не более чем $\lfloor \tau n \rfloor - k$ ошибок, величина $\log M_s(\lfloor n(1+\tau) \rfloor - 2k, \lfloor \tau n \rfloor - k, \alpha)$ все равно асимптотически минимизируется при $k = 0$.

Согласно лемме 3, зафиксировав $\alpha = 2\tau/(1+\tau)$ и используя алгоритм $\mathfrak{A}_s(M, n', \alpha)$ для $n' = \lfloor n(1+\tau) \rfloor - 2k$ и не более чем $\lfloor \tau n \rfloor - k$ ошибок, возможно достичь скоростей, лежащих на касательной прямой к функции $R_s(x) = 1 - h(x)$ в точке $(\tau/(1+\tau), R_s(\tau/(1+\tau)))$.

Для доказательства утверждения о минимизации определим функцию

$$f(x) := \frac{\log M_s(n(1+\tau-2x),(\tau-x)n,\alpha)}{n} =$$
$$= (1+\tau-2x)\left(a\frac{\tau-x}{1+\tau-2x}+b\right)(1+o(1)),$$

где действительные числа a, b таковы, что достижимые скорости лежат на касательной в точке $(\tau/(1+\tau), R_s(\tau/(1+\tau)))$. Из этих соображений находим $a = \log\tau$ и $b = 1-\log(1+\tau)$.

Производная $f(x)$ при $n\to\infty$ равна

$$\frac{\partial f}{\partial x} = -a-2b+o(1) = \log\left(\frac{(1+\tau)^2}{\tau}\right)-2+o(1) > 0,$$

где неравенство следует из того, что выражение под логарифмом строго больше 4 при $\tau < 1/2$. Отсюда следует, что функция f асимптотически достигает минимума при $x = 0$. ▲

Мы уже показали, что выходная последовательность алгоритма $\mathfrak{A}_{\mathrm{id}}(M,n,\alpha)$ для комбинаторного канала со вставками и выпадениями также принадлежит множеству возможных выходных последовательностей для комбинаторного канала с замещениями при использовании алгоритма $\mathfrak{A}_s(M,n',\alpha)$ для длины $n'\in[n-t,n+t]$ и не более чем $t' = \lfloor(n'-n+t)/2\rfloor$ ошибок. Таким образом, из предложения 1 следует, что $R_{\mathrm{id}}(\tau)\geqslant(1+\tau)R_s(\tau/(1+\tau))$. Применяя теорему 2, получаем, что скорость $R_{\mathrm{id}}(\tau)$, определенная в (2), действительно достижима, и верхняя и нижняя границы скорости для комбинаторного канала со вставками и выпадениями совпадают.

§ 4. Анализ кода с обратной связью, исправляющего замещения

В этом параграфе мы напомним алгоритм передачи и его анализ, предложенный Зигангировым в [9]. Чтобы сделать этот параграф более самодостаточным, сначала будет описан процесс кодирования. Далее будут введены некоторые полезные понятия, и мы проанализируем алгоритм, не вдаваясь в технические подробности. Наконец, мы приведем доказательства вспомогательных технических утверждений, из которых будет следовать основной результат.

4.1. Код с обратной связью, исправляющий замещения. Предположим, что не более чем τn замещений может произойти при передаче n символов, и отправитель хочет передать одно из M сообщений. Если $0\leqslant\tau<(3-\sqrt{5})/4$, то сначала нужно определить действительные числа α и β, такие что $\alpha+\beta=2$, причем α достаточно близко к 2τ. Если $(3-\sqrt{5})/4\leqslant\tau\leqslant 1/3$, то α и β берутся таким образом, что $\alpha+\beta=2$, причем α достаточно близко к числу $(3-\sqrt{5})/2$, но при этом меньше этого числа. Таким образом, $\alpha<1$ и $\beta>1$. Также очевидно, что для такого выбора будет выполнено неравенство

$$\alpha\beta^2\leqslant 1. \tag{3}$$

Кодирование: Изначально отправитель берет отрезок $[0,1]$ и делит его на M подотрезков одной и той же длины. Занумеруем все эти отрезки слева направо. Таким образом, сообщение j, $j\in[M]$, связано естественным образом с j-м отрезком. Предположим, что отправитель хочет послать сообщение $m\in[M]$. Тогда отрезок с номером m будем называть *истинным отрезком*, и символом $T(i)$ мы будем обозначать истинный отрезок после отправки i символов. Определим объединение отрезков слева и справа от $T(i)$ через $L(i)$ и $R(i)$. Обозначим длины отрезков $L(i)$, $T(i)$ и $R(i)$

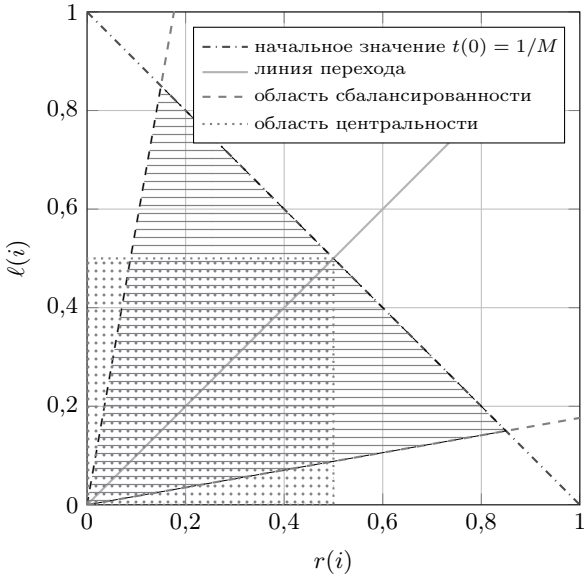

Рис. 3. Иллюстрация понятий перехода, сбалансированного и центрального истинного отрезка и соответствующих областей при $\tau = 0{,}15$, $\alpha = 2\tau$ и $\beta = 2 - \alpha$

через $\ell(i)$, $t(i)$ и $r(i)$. Очевидно, что $t(0) = 1/M$ и $\ell(i) + t(i) + r(i) = 1$. В i-й момент времени отправитель проверяет, находится ли центр отрезка $T(i-1)$ левее точки $1/2$, и отправляет "0" в канал при выполнении этого условия. В противном случае отправитель посылает "1" в канал. Если получен "0", то длины всех отрезков, находящихся левее точки $1/2$, умножаются на β, а длины отрезков правее точки $1/2$ умножаются на α. Более формально,

$$\ell(i) := \beta \min\Big(\frac{1}{2}, \ell(i-1)\Big) + \alpha \max\Big(\ell(i-1) - \frac{1}{2}, 0\Big),$$
$$r(i) := \alpha \min\Big(\frac{1}{2}, r(i-1)\Big) + \beta \max\Big(r(i-1) - \frac{1}{2}, 0\Big),$$
$$t(i) := 1 - \ell(i) - r(i).$$

Если получен символ "1", то длины всех отрезков левее точки $1/2$ умножаются на α, а длины всех отрезков правее точки $1/2$ умножаются на β. Таким образом,

$$\ell(i) := \beta \max\Big(\ell(i-1) - \frac{1}{2}, 0\Big) + \alpha \min\Big(\frac{1}{2}, \ell(i-1)\Big),$$
$$r(i) := \alpha \max\Big(r(i-1) - \frac{1}{2}, 0\Big) + \beta \min\Big(\frac{1}{2}, r(i-1)\Big),$$
$$t(i) := 1 - \ell(i) - r(i).$$

В последующих пунктах будет показано, что для произвольного $\varepsilon > 0$, правильно выбранного $\alpha \approx 2\min(\tau, (3-\sqrt{5})/4)$, достаточно большого n и $M = 2^{(R_s(\tau)-\varepsilon)n}$ истинный отрезок $T(n)$ содержит точку $1/2$.

4.2. Дополнительные обозначения. Определим $x(i) := \min(\ell(i), r(i))$ и $y(i) := \max(\ell(i), r(i))$. Будем говорить, что $T(i)$ является *центральным*, если $y(i) \leqslant 1/2$. Также будем говорить, что выполнен *переход* между моментом $i-1$ и моментом i, если либо (1) $\ell(i-1) \leqslant r(i-1)$ и $\ell(i) > r(i)$, либо (2) $\ell(i-1) > r(i-1)$ и $\ell(i) \leqslant r(i)$. Ис-

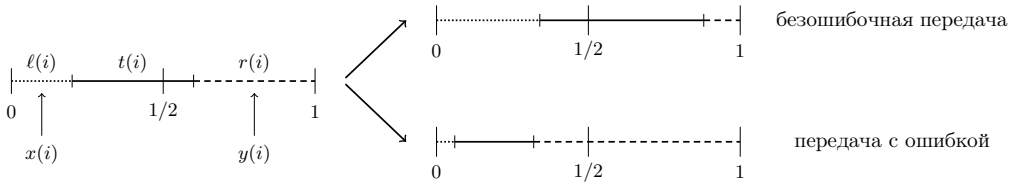

Рис. 4. Пример изменения отрезков $L(i), T(i)$ и $R(i)$ в случае безошибочной передачи и передачи с ошибкой

тинный отрезок $T(i)$ назовем *сбалансированным*, если $y(i)/x(i) \leqslant \beta/\alpha$. Иллюстрация данных понятий приведена на рис. 3.

Пример 1. На рис. 4 можно увидеть отрезки $L(i), R(i)$ и $T(i)$ с $\ell(i) = 0{,}2$, $r(i) = 0{,}4$ и $t(i) = 0{,}4$, а также локальные изменения, которые произойдут в случае ошибочной и безошибочной передачи согласно описанной выше процедуре кодирования при $\tau = 0{,}15$, $\alpha = 2\tau$ и $\beta = 2 - \alpha$. В частности, истинный отрезок $T(i)$ содержит точку $1/2$, и следовательно, является центральным. Более того, выполнено $x(i) = \ell(i)$ и $y(i) = r(i)$, а истинный отрезок $T(i)$ является сбалансированным, поскольку $2 = \dfrac{y(i)}{x(i)} \leqslant \dfrac{\beta}{\alpha} = \dfrac{1{,}7}{0{,}3}$. В этом примере безошибочная передача соответствует отправке и приему "0", что соответствует увеличению всего слева от точки $1/2$ в β раз и уменьшению всего справа в α^{-1} раз. В таком случае происходит переход, так как $\ell(i) \leqslant r(i)$ и $\ell(i+1) > r(i+1)$. Истинный отрезок в момент $i+1$ по-прежнему содержит точку $1/2$, а потому является центральным. Более того, $T(i+1)$ является сбалансированным. Если же при передаче происходит ошибка, то все, что находится слева от $1/2$, увеличивается в β раз, а другая часть уменьшается в α раз. В этом случае не происходит перехода, а истинный отрезок $T(i+1)$ не содержит точку $1/2$, а потому не является центральным. Более того, $T(i+1)$ не является сбалансированным, поскольку $\dfrac{34}{3} = \dfrac{y(i)}{x(i)} > \dfrac{\beta}{\alpha} = \dfrac{1{,}7}{0{,}3}$.

Определим функцию

$$g(i_0, i_1) := e(i_0, i_1) \log \alpha + f(i_0, i_1) \log \beta,$$

где $e(i_0, i_1)$ и $f(i_0, i_1)$ обозначают число ошибочных и безошибочных передач в промежутке времени $(i_0, i_1] = \{i_0 + 1, \ldots, i_1\}$. Отметим, что функция g является аддитивной в том смысле, что выполнено равенство $g(i_0, i_2) = g(i_0, i_1) + g(i_1, i_2)$. Для краткости будем писать $g(n)$, $e(n)$ и $f(n)$ для обозначения $g(0, n)$, $e(0, n)$ и $f(0, n)$. Можно думать о величине $g(i, n)$ как о количестве *энергии*, которая имеется в момент i и расходуется для изменения длины истинного отрезка. Для одной безошибочной передачи нужно потратить количество энергии, равное $\log \beta > 0$, в то время как при ошибочной передаче энергия увеличивается на $-\log \alpha > 0$.

4.3. План доказательства. В следующем утверждении показано, что если количество энергии достаточно велико, то в какой-то момент времени длина истинного отрезка будет отделена от нуля вне зависимости от длины кода n, а также останется некоторое положительное количество энергии, которая будет использоваться в дальнейшем для увеличения длины истинного отрезка.

Лемма 4. *Для любых действительных $\varepsilon_1 > \varepsilon_2 > 0$ существует $\delta > 0$, такое что для почти всех (за исключением конечного числа) $\alpha \in (0, 1)$ выполнено утверждение: для $n \in \mathbb{N}$ условие $g(0, n) > \varepsilon_1 n - \log t(0)$ гарантирует, что существует первый момент времени n_1, при котором $t(n_1) \geqslant \delta$; более того, $g(0, n_1) < \varepsilon_2 n - \log t(0)$ и $g(n_1, n) > (\varepsilon_1 - \varepsilon_2) n$.*

Следующая лемма показывает, что после того как было потрачено достаточное количество энергии, истинный отрезок становится центральным.

Лемма 5. *Пусть α и β удовлетворяют соотношению* (3). *Тогда существует константа $c > 0$, зависящая только от α, β и $t(n_1)$, такая что если $g(n_1, n) > c$, то истинный отрезок $T(n)$ является центральным. Более того, длина истинного отрезка в момент n равна $t(n) = 1 - o(1)$ при $c \to \infty$.*

Наконец, применяя указанные выше леммы, можно получить основной результат.

Теорема 3. *Пусть $\tau \in [0, 1/3)$. Для любого $\varepsilon > 0$ существует $\alpha \in [0, (3-\sqrt{5})/2]$, такое что для достаточно большого n с помощью описанной процедуры кодирования для комбинаторного канала с замещениями и обратной связью можно исправить τn ошибок. При этом скорость передачи при таком кодировании не меньше $R_s(\tau) - \varepsilon$, где*

$$R_s(\tau) = \begin{cases} 1 - h(\tau), & \text{если } 0 \leqslant \tau \leqslant \dfrac{3-\sqrt{5}}{4}, \\ (1-3\tau)\log\left(\dfrac{1+\sqrt{5}}{2}\right), & \text{если } \dfrac{3-\sqrt{5}}{4} < \tau \leqslant 1/3. \end{cases}$$

Доказательство. Выберем M и соответственно $t(0) = 1/M$ так, что выполнены неравенства $2^{(R_s(\tau)-\varepsilon)n} < M < 2^{(R_s(\tau)-\varepsilon/2)n}$ и $-\log t(0) < (R_s(\tau) - \varepsilon/2)n$. По определению имеем $e(0, n) \leqslant \tau n$ и $f(0, n) \geqslant (1 - \tau)n$. Следовательно,

$$g(0, n) = e(0, n)\log\alpha + f(0, n)\log\beta \geqslant (\tau\log\alpha + (1-\tau)\log\beta)n.$$

При условии $\alpha + \beta = 2$ и $\alpha\beta^2 \leqslant 1$ функция $\tau\log\alpha + (1-\tau)\log\beta$ достигает максимального значения $R_s(\tau)$ при $\alpha = 2\min(\tau, (3-\sqrt{5})/4)$ и $\beta = 2 - \alpha$. Для использования леммы 4 выберем α немного меньшим $2\min(\tau, (3-\sqrt{5})/4)$, так что

$$g(0, n) = e(0, n)\log\alpha + f(0, n)\log\beta \geqslant \left(R_s(\tau) - \frac{\varepsilon}{10}\right)n > \frac{2}{5}\varepsilon n - \log t(0).$$

Применяя лемму 4 с $\varepsilon_1 = 2\varepsilon/5$ и $\varepsilon_2 = \varepsilon/5$, получаем, что существует константа $\delta > 0$, для которой для почти всех (за исключением конечного числа) $\alpha \in (0, 1)$ существует первый момент времени n_1, такой что $t(n_1) \geqslant \delta$. Более того,

$$g(n_1, n) = g(0, n) - g(0, n_1) = e(n_1, n)\log\alpha + f(n_1, n)\log\beta > \frac{1}{5}\varepsilon n,$$

и следовательно, для достаточно большого n по лемме 5 можно заключить, что $T(n)$ является центральным. ▲

Наиболее технически сложный результат, используемый при доказательстве теоремы 1, заключен в лемме 5. Для ее доказательства определим несколько вспомогательных функций:

$$u_1(i) := \log(t(i)/x(i)),$$
$$v_1(i) := \log(2t(i)),$$
$$u_2(i) := \begin{cases} -\log(4x(i)y(i)), & \text{если } y(i) \leqslant 1/2, \\ \log((1-y(i))/x(i)), & \text{если } y(i) > 1/2, \end{cases}$$
$$v_2(i) := \begin{cases} -\log(2y(i)), & \text{если } y(i) \leqslant 1/2, \\ \log(2(1-y(i))), & \text{если } y(i) > 1/2. \end{cases}$$

Также определим константы $u_1' := \log(\beta - \alpha)$, $u_1'' := \log(\beta/\alpha - 1)$, $u_2' := \log(\beta/\alpha)$. Очевидно, что $u_1' < u_1''$. Наконец, определим

$$u(i) := \begin{cases} u_1(i), & \text{если } u_1(i) < u_1', \\ u_2(i), & \text{если } u_1(i) \geqslant u_1', \end{cases}$$

$$v(i) := \begin{cases} v_1(i), & \text{если } u_1(i) < u_1'', \\ v_2(i), & \text{если } u_1(i) \geqslant u_1''. \end{cases}$$

Для того чтобы показать корректность леммы 5, мы докажем, что значение $v(n)$ велико, что может быть выполнено лишь тогда, когда $\ell(n)$ и $r(n)$ достаточно малы (см. пример 2).

Предложение 2. *Если $T(i-1)$ является центральным и происходит переход, то $v(i) - v(i-1) \geqslant -\log\beta$. В противном случае выполнено*

$$v(i) - v(i-1) \geqslant \begin{cases} \log\beta, & \textit{если передача безошибочна}, \\ \log\alpha, & \textit{если произошла ошибка}. \end{cases}$$

Изменение $\Delta v(i) := v(i) - v(i-1)$ функции v больше или равно $g(i-1, i)$ за исключением случая, когда $T(i-1)$ является центральным и происходит переход. Если количество подобных центральных переходов достаточно мало, то значение $v(n) - v(n_1)$ достаточно близко к значению $g(n_1, n)$. Поскольку $v(n_1)$ ограничено константой, зависящей от $t(n_1)$, можно показать, что $v(n)$ достаточно велико, и все доказано. Чтобы разобраться со вторым случаем, при котором число центральных переходов велико, мы воспользуемся функцией u и следующим предложением.

Предложение 3. *Если $T(i-1)$ является центральным и происходит переход, то $u(i) - u(i-1) \geqslant \log\beta$. В противном случае $u(i) - u(i-1) \geqslant 0$.*

Таким образом, во втором случае функция u достаточно велика, откуда также следует, что функция u_2 велика. Однако этого недостаточно, чтобы заключить, что $T(n)$ является центральным (см. пример 2). Рассмотрим момент времени, когда произошел последний центральный переход. Можно показать, что в этот момент истинный отрезок является также сбалансированным. Отсюда будет следовать, что значение функции v_2 достаточно близко к значению функции u_2, т.е. тоже достаточно велико. Для того чтобы функция v_2 снова стала мала, должно произойти достаточно много ошибок в силу предложения 4.

Предложение 4. *Если произошел переход и $u_1(i-1) \geqslant u_1''$, то $v_2(i) - v_2(i-1) \geqslant -\log\beta$. Если не произошел переход, то*

$$v_2(i) - v_2(i-1) \geqslant$$

$$\geqslant \begin{cases} \log\beta, & \textit{если передача безошибочна}, \\ \log\alpha, & \textit{если произошла ошибка и } v_2(i-1) < \log\beta, \\ -\log\beta \ (\geqslant \log\alpha), & \textit{если произошла ошибка и } v_2(i-1) \geqslant \log\beta. \end{cases}$$

Большое количество ошибок означает, что осталось большое количество энергии. Можно снова воспользоваться предложением 2, поскольку больше не осталось центральных переходов, и следовательно, можно заключить, что функция v примет большое значение после последней передачи.

Пример 2. Рисунки 5 и 6 демонстрируют различные множества уровней функций $v(i)$ и $u(i)$ как функций, зависящих от $\ell(i)$ и $r(i)$. Множество уровня задается как множество пар $(\ell(i), r(i))$, для которых верно $v(i) = \text{const}$ ($u(i) = \text{const}$). При больших значениях функции $v(i)$ соответствующее множество уровня находится ближе к началу координат, т.е. значения $\ell(i)$ и $r(i)$ довольно малы, а значение $t(i)$ велико.

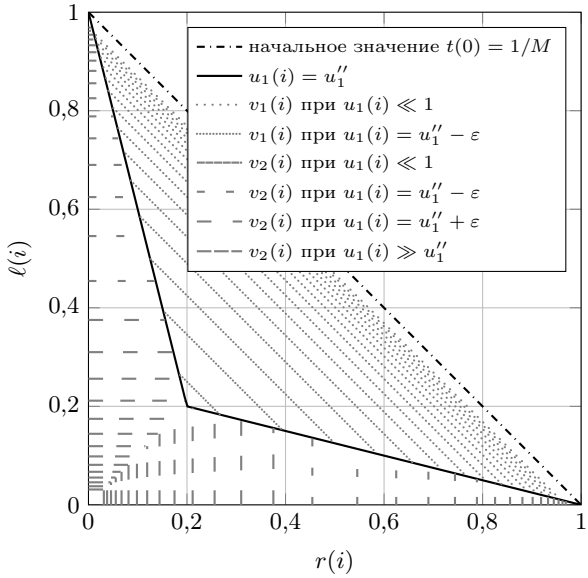

Рис. 5. Множество уровня функции $v(i)$ при $\tau = 0{,}15$, $\alpha = 2\tau$ и $\beta = 2 - \alpha$

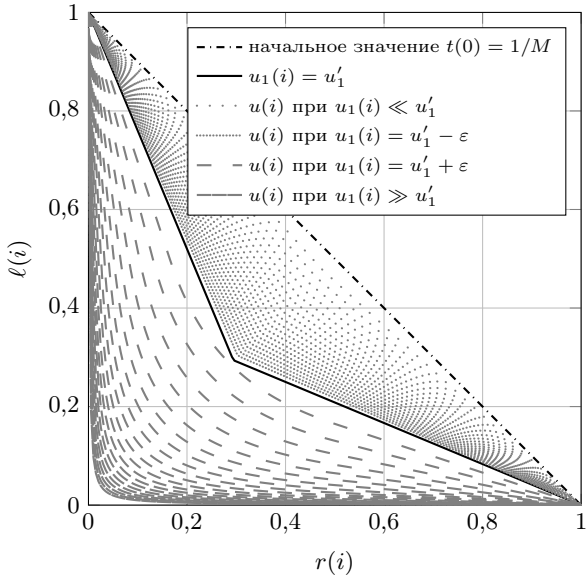

Рис. 6. Множество уровня функции $u(i)$ при $\tau = 0{,}15$, $\alpha = 2\tau$ и $\beta = 2 - \alpha$

В то же время при больших значениях функции $u(i)$ можно лишь утверждать, что множество уровня расположено близко к осям координат, т.е. значение $\ell(i)$ или $r(i)$ мало. Также отметим, что в случае, если $T(i)$ является центральным, то должно быть выполнено неравенство $\ell(i), r(i) < 1/2$. Исходя из приведенных рассуждений, достаточно показать, что значение $v(n)$ велико, чтобы заключить, что истинный отрезок $T(n)$ является центральным.

В вышеприведенных рассуждениях было опущено много технических деталей. Для того чтобы формально доказать лемму 5, нам понадобится несколько вспомогательных предложений.

Предложение 5. *Выполнены следующие (очевидные) свойства*:

1. *Если в какой-то момент произошел переход, то прием был безошибочным*;
2. *Если прием безошибочный, то длина истинного отрезка увеличилась, т.е.* $t(i) > t(i-1)$. *В противном случае длина уменьшилась*;
3. *Если* $T(i-1)$ *является центральным,* $T(i)$ *не является центральным и не произошел переход, то при передаче произошла ошибка*;
4. *Выполнены неравенства* $\alpha t(i-1) \leqslant t(i) \leqslant \beta t(i-1)$;
5. *Для всяких* $x(i)$, $y(i)$ *и* $t(i)$ *верны неравенства* $u_2(i) \geqslant u_1(i)$ *и* $u_2(i) \geqslant u(i)$;
6. *Для всяких* $x(i)$, $y(i)$ *и* $t(i)$ *верны неравенства* $v_2(i) \geqslant v_1(i)$ *и* $v_2(i) \geqslant v(i)$.

Предложение 6. *Выполнены следующие свойства*:

1. *Если* $T(i)$ *является центральным, то* $u_2(i) \geqslant 2v_2(i)$;
2. *Если* $T(i)$ *является центральным и сбалансированным, то* $u_2(i) < 2v_2(i) + \log(\beta/\alpha)$;
3. *Если происходит переход между моментами* $i-1$ *и* i, *то* $T(i-1)$ *и* $T(i)$ *являются сбалансированными*;
4. *Если* $u_1(i-1) < u_1''$ (*или* $u_1(i-1) < u_1'$), $T(i-1)$ *является центральным и передача безошибочна, то происходит переход между моментами* $i-1$ *и* i;
5. *Если* $u_1(i-1) \geqslant u_1'$ (*или* $u_1(i-1) \geqslant u_1''$) *и происходит переход между моментами* $i-1$ *и* i, *то* $T(i)$ *является центральным*;
6. *Если* $u_1(i-1) \geqslant u_1''$ *и происходит переходит между моментами* $i-1$ *и* i, *то* $T(i-1)$ *и* $T(i)$ *являются центральными*;
7. *Если* $u_2(i) \geqslant u_2'$ *и происходит переход между моментами* $i-1$ *и* i *или между моментами* i *и* $i+1$, *то* $T(i)$ *является центральным.*

Предложение 7. *Если* $u_1(i-1) < u_1'$, *то* $u_1(i) - u_1(i-1) \geqslant 0$. *Если вдобавок* $T(i-1)$ *является центральным и происходит переход между моментами* $i-1$ *и* i, *то* $u_1(i) - u_1(i-1) \geqslant \log\beta$. *С другой стороны, если* $u_1(i-1) \geqslant u_1'$, *то* $u_1(i) \geqslant u_1'$, *а если* $u_1(i-1) \geqslant u_1''$, *то* $u_1(i) \geqslant u_1''$.

Предложение 8. *Если* $u_1(i-1) \geqslant u_1'$, *то* $u_2(i) - u_2(i-1) \geqslant 0$. *Если оба отрезка* $T(i-1)$ *и* $T(i)$ *являются центральными, то* $u_2(i) - u_2(i-1) = -\log(\alpha\beta)$.

Предложение 9. *Предположим, что* $u_1(i-1) < u_1''$. *Если происходит переход, то* $v_1(i) - v_1(i-1) > 0$, *в противном случае*

$$v_1(i) - v_1(i-1) \geqslant \begin{cases} \log\beta, & \text{если передача безошибочна,} \\ \log\alpha, & \text{если произошла ошибка.} \end{cases}$$

Предложение 10. *Пусть* $c = -5\log(\alpha\beta)$. *Если оба отрезка* $T(i_0)$ *и* $T(i_1)$ *являются центральными и сбалансированными, а также верно* $u_2(i_0) \geqslant c$, *то*

$$g(i_0, i_1) \leqslant u_2(i_1) - u_2(i_0). \tag{4}$$

Доказательства лемм и предложений приведены в пп. 4.4, 4.5. Наконец, в блок-схеме, приведенной на рис. 7, показана схема доказательства основного результата.

Перед тем как будут даны доказательства, приведем пример 3, в котором рассмотрена передача сообщения с помощью описанной процедуры кодирования. В данном примере отслеживается положение истинного отрезка $T(i)$.

Пример 3. На рис. 8 продемонстрирован пример передачи случайного сообщения m из $M = 2^{14}$ возможных на длине $n' = 27$. Использованы параметры $\tau = 0{,}15$,

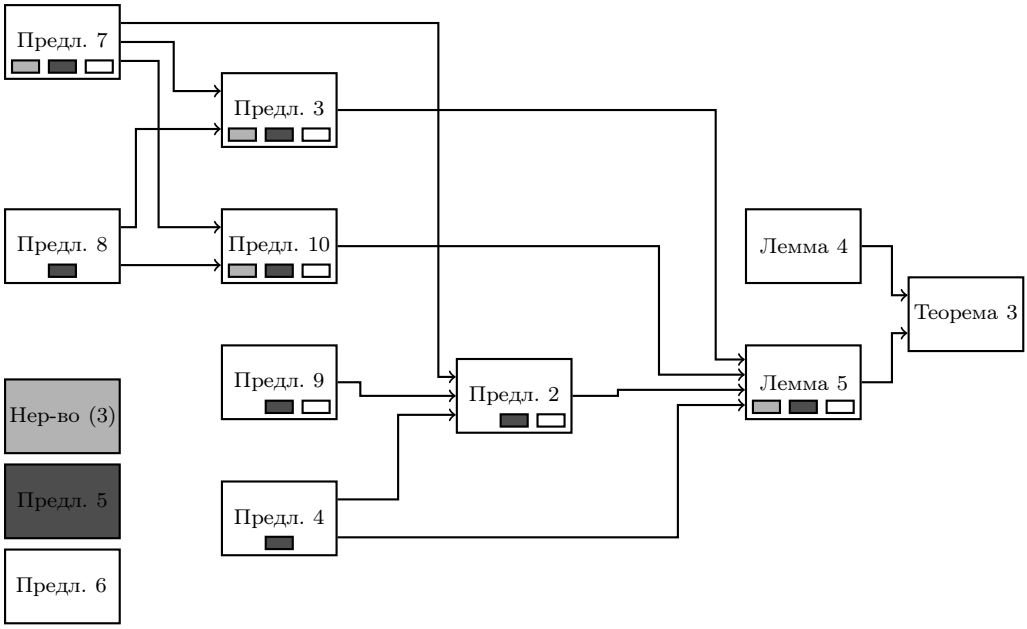

Рис. 7. Иллюстрация схемы доказательства теоремы 3. Для неравенства (3) и предложений 5, 6 не показаны стрелочки для простоты восприятия блок-схемы; маленькие прямоугольники указывают, как используются эти результаты в ходе доказательства

$\alpha = 2\tau$ и $\beta = 2 - \alpha$. Напомним, что $t(i) = 1 - \ell(i) - r(i)$. Три ошибки произошли в моменты $i \in \{16, 21, 22\}$. Поскольку перед началом передачи все отрезки одной длины, положение истинного отрезка $T(0)$, описываемое парой $(\ell(0), r(0))$, близко к прямой $\ell(i) + r(i) = 1$. При увеличении числа переданных символов видно, что положение истинного отрезка становится все ближе к началу координат. Другими словами, длина истинного отрезка увеличивается. Видно, что $T(i)$ становится центральным в момент времени $i = 15$. Поскольку в момент времени $i = 16$ произошла ошибка, истинный отрезок перестает быть центральным. Более того, в этот же момент траектория пересекает множество уровня u_1' для функции u, и в последующих шагах уже не пересекает этот уровень (см. предложение 7). Аналогичное поведение истинного отрезка можно видеть при пересечении множества уровня u_1'' несмотря на то, что ошибки произойдут в моменты времени $i = 21, 22$ (см. предложение 7). Оставшиеся передачи являются безошибочными, и потому можно видеть, что длина $t(i)$ увеличивается до самого последнего момента n'.

4.4. Доказательства лемм.

Доказательство леммы 4. Выберем $k \in \mathbb{N}$ таким, что $k > \dfrac{\log \beta}{\varepsilon_2}$, где $\varepsilon_1 > \varepsilon_2 > 0$. Зафиксируем некоторое значение δ, которое меньше $\dfrac{1}{\beta^2}$ и не превосходит наименьшей по модулю разницы между точкой $1/2$ и ее возможным образом после передачи $1, 2, \ldots, k-1$ символов. Например, после передачи одного символа точка $1/2$ может перейти в одну из точек $\left\{\dfrac{\alpha}{2}, \dfrac{\beta}{2}\right\}$, а после передачи двух – в одну из точек $\left\{\dfrac{\alpha^2}{2}, \dfrac{\alpha\beta}{2}, 1 - \alpha\Big(1 - \dfrac{\beta}{2}\Big), 1 - \beta\Big(1 - \dfrac{\beta}{2}\Big)\right\}$. Легко показать, что $\delta > 0$ для всех (за

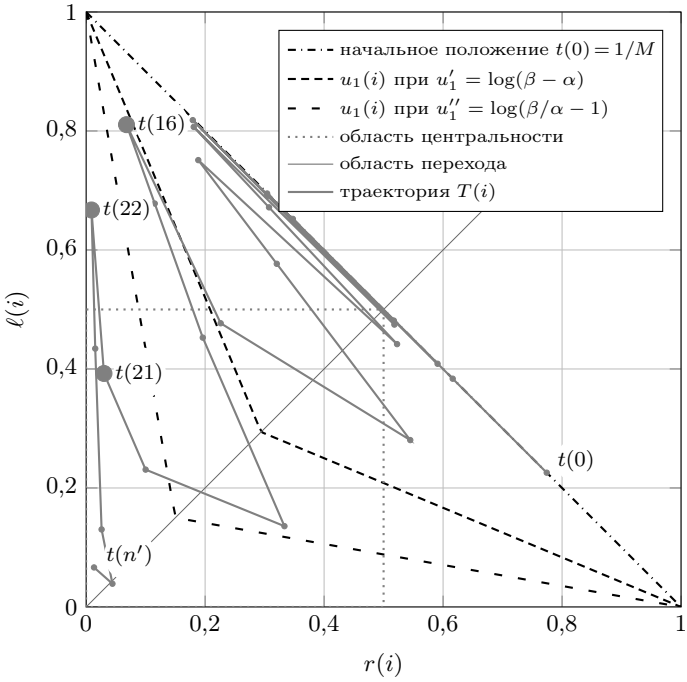

Рис. 8. Иллюстрация траектории значения функции $t(i) = 1 - \ell(i) - r(i)$. Положив $M = 2^{14}$, мы передали случайное сообщение, используя кодирование длины $n' = 27$. Ошибки произошли в моменты времени 16, 21 и 22. Используемые параметры: $\tau = 0{,}15$, $\alpha = 2\tau$ и $\beta = 2 - \alpha$

исключением конечного числа) значений $\alpha \in (0, 1)$. Отметим, что если α является трансцендентным числом, то $\delta > 0$ для любого k.

Предположим, что $t(i) < \delta$ для всех $i \in \{0, 1, \ldots, n\}$. По определению числа δ имеем, что если $T(i)$ является центральным (точка $1/2$ принадлежит отрезку $T(i)$), то $T(i+1), \ldots, T(i+k-1)$ не являются центральными. Обозначим через n' число моментов времени, когда $T(i)$ является центральным. Очевидно, что

$$n' \leqslant \frac{n}{k} + 1 < \frac{\varepsilon_1 n}{\log \beta} + 1.$$

Отсюда можно заключить, что

$$t(n) \geqslant t(0)\alpha^{e(0,n)}\beta^{f(0,n)-n'} \geqslant t(0)\alpha^{e(0,n)}\beta^{f(0,n)-\frac{\varepsilon_1 n}{\log \beta}-1} \geqslant \frac{1}{\beta} > \delta,$$

где мы воспользовались тем, что $g(0, n) = e(0, n)\log\alpha + f(0, n)\log\beta > \varepsilon_1 n - \log t(0)$. Таким образом, пришли к противоречию.

Пусть n_1 обозначает первый момент, при котором $t(n_1) \geqslant \delta$. Используя аналогичные рассуждения, получаем, что

$$\beta\delta > \beta t(n_1 - 1) \geqslant t(n_1) \geqslant t(0)\alpha^{e(0,n_1)}\beta^{f(0,n_1)-\frac{\varepsilon_2 n_1}{\log \beta}-1}.$$

Тогда из вышесказанного следует, что

$$g(0, n_1) = e(0, n_1)\log\alpha + f(0, n_1)\log\beta < \varepsilon_2 n_1 - \log t(0) + \log(\beta^2\delta) < \varepsilon_2 n - \log t(0).$$

Следовательно,

$$g(n_1, n) = g(0, n) - g(0, n_1) > (\varepsilon_1 - \varepsilon_2)n > 0. \quad \blacktriangle$$

Доказательство леммы 5. Определим множество $I := \{i \in (n_1, n] : T(i-1)$ является центральным и происходит переход между моментами $i-1$ и $i\}$. Пусть $m := |I|$, и пусть выбраны числа c, m', v_2', удовлетворяющие соотношению $c \gg m' \gg \gg v_2' \gg 1$. Например, можно выбрать эти числа в виде $c = x^3$, $m' = x^2$, $v_2' = x$, где x достаточно велико по сравнению с величинами α, β, $t(n_1)$. Итак, из условия леммы имеем, что $g(n_1, n) > c$.

Случай $m \leqslant m'$. Из п. 6 предложения 5 и определения функции v можно получить, что $v(n_1) \geqslant v_1(n_1) = \log(2t(n_1))$. Отсюда в силу предложения 2 имеем

$$v(n) \geqslant v(n_1) + e(n_1, n)\log\alpha + (f(n_1, n) - m)\log\beta - m\log\beta \geqslant$$
$$\geqslant \log(2t(n_1)) + g(n_1, n) - 2m'\log\beta > \log(2t(n_1)) + c - 2m'\log\beta \gg 1.$$

Здесь мы воспользовались выбором чисел c и m', а также тем фактом, что $g(n_1, n) = = e(n_1, n)\log\alpha + f(n_1, n)\log\beta$. Из определения функции v и из неравенства $v(n) \gg 1$ получаем, что $T(n)$ является центральным (см. рис. 5).

Случай $m > m'$. Пусть n_2 является m'-м наименьшим элементом в множестве I. Из п. 5 предложения 5 и определения функции u получаем $u_2(n_2) \geqslant u(n_2)$, а также $u(n_1) \geqslant u_1(n_1) = \log(t(n_1)/x(n_1))) \geqslant \log(2t(n_1))$, так как $x(n_1) \leqslant 1/2$. Отсюда в силу предложения 3 имеем

$$u_2(n_2) \geqslant u(n_2) \geqslant u(n_1) + m'\log\beta \geqslant \log(2t(n_1)) + m'\log\beta = \Omega(m').$$

Последнее равенство верно из-за выбора величины m'. Отсюда следует, что в силу п. 7 предложения 6 отрезок $T(n_2)$ является центральным, а в силу п. 3 предложения 6 отрезок $T(n_2)$ является сбалансированным. Из предложения 2, п. 6 предложения 5 и п. 1 предложения 6 имеем

$$g(n_1, n_2) \leqslant v(n_2) - v(n_1) + 2m'\log\beta \leqslant v_2(n_2) - v_1(n_1) + 2m'\log\beta \leqslant$$
$$\leqslant \frac{1}{2}u_2(n_2) - \log(2t(n_1)) + 2m'\log\beta \leqslant \frac{1}{2}u_2(n_2) + O(m'). \tag{5}$$

Пусть n_3 является самым последним элементом в множестве I. Мы знаем, что $T(n_2)$ является центральным и сбалансированным, а также $u_2(n_2) = \Omega(m')$. Аналогичный вывод верен и для отрезка $T(n_3)$, т.е. $T(n_3)$ является центральным и сбалансированным, а $u_2(n_3) = \Omega(m')$. По предложению 10 получаем

$$g(n_2, n_3) \leqslant u_2(n_3) - u_2(n_2). \tag{6}$$

В силу п. 2 предложения 6 имеем

$$v_2(n_3) \geqslant \frac{1}{2}u_2(n_3) - \frac{1}{2}\log(\beta/\alpha) = \Omega(m') \gg v_2'.$$

Если $v_2(n) \geqslant v_2'$, то все доказано. В противном случае пусть $n_4 > n_3$ обозначает первый момент, при котором $v_2(n_4) < v_2'$. Поскольку $v_2(i) > 0$ для всех $i \in [n_3, n_4)$, из определения функции v_2 следует, что отрезок $T(i)$ является центральным для всех $i \in [n_3, n_4)$. По определению множества I нет ни одного перехода между моментами n_3 и n_4. В силу предложения 4 и п. 2 предложения 6, а также неравенства $\alpha\beta^2 < 1$ легко получить, что

$$g(n_3, n_4) = e(n_3, n_4)\log\alpha + f(n_3, n_4)\log\beta \leqslant$$
$$\leqslant 2(-e(n_3, n_4) + f(n_3, n_4))\log\beta \leqslant 2v_2(n_4) - 2v_2(n_3) < -u_2(n_3) + O(v_2'). \tag{7}$$

По определению множества I нет ни одного момента $i \in (n_4, n]$, при котором $T(i-1)$ является центральным и происходит переход между моментами $i-1$ и i. Из предложения 2 можно вывести, что

$$g(n_4, n) \leqslant v(n) - v(n_4). \tag{8}$$

Поскольку не происходит переход между моментами $n_4 - 1$ и n_4, из предложения 4 следует, что $v_2(n_4) \geqslant v_2(n_4 - 1) - \log\beta \geqslant v_2' - \log\beta \gg 1$. По определению значения $v_2(n_4)$ имеем, что $y(n_4) \ll 1$, а также $v_1(n_4) > 0$. Отсюда получаем $v(n_4) \geqslant 0$. Суммируя (5)–(8) и используя тот факт, что $v(n_4) \geqslant 0$, получаем

$$v(n) \geqslant g(n_1, n) - O(m') - O(v_2') \gg 1.$$

Это означает, что отрезок $T(n)$ является центральным. ▲

4.5. Доказательство предложений.

Доказательство предложения 5. 1. Если произошел переход, то меньший отрезок из множества $\{L(i), R(i)\}$ увеличился. Такое может произойти только в случае безошибочной передачи символа.

2. Мы приводим доказательство только для случая безошибочного приема, поскольку случай ошибочного приема рассматривается аналогично.

В случае, если отрезок $T(i-1)$ не является центральным, утверждение выполнено, так как $t(i) = t(i-1)\beta$. В случае, если $T(i-1)$ является центральным отрезком, выполняется

$$t(i) = 1 - \beta x(i-1) - \alpha y(i-1),$$

откуда вытекает следующая цепочка равенств:

$$t(i) - t(i-1) = (1-\beta)x(i-1) + (1-\alpha)y(i-1) = \frac{\beta - \alpha}{2}(y(i-1) - x(i-1)) > 0,$$

где мы воспользовались тем фактом, что $\alpha + \beta = 2$.

3. Предположим, что передача произошла без ошибки и перехода не было. Тогда, так как $T(i-1)$ является центральным отрезком, то

$$\begin{aligned} &x(i-1)\beta = x(i), \\ &y(i-1)\alpha = y(i). \end{aligned} \tag{9}$$

Для того чтобы $T(i)$ не был центральным, необходимо выполнение неравенства $y(i) > 1/2$, но это приводит к противоречию с формулой (9), так как $\alpha < 1$ и $y(i-1) < 1/2$.

4. При безошибочном и ошибочном приеме символа происходит отображение отрезка $[0, 1]$ в отрезок $[0, 1]$, причем длина всякого подотрезка умножается на некоторое действительное число, находящееся в интервале $[\alpha, \beta]$.

5. Вспомним определение функции $u_2(i)$:

$$u_2(i) := \begin{cases} -\log(4x(i)y(i)), & \text{если } y(i) \leqslant 1/2, \\ \log((1-y(i))/x(i)), & \text{если } y(i) > 1/2. \end{cases}$$

Утверждается, что

$$-\log(4x(i)y(i)) \geqslant \log((1-y(i))/x(i)).$$

Действительно, это неравенство эквивалентно следующим двум:

$$\frac{1}{4x(i)y(i)} \geqslant \frac{1-y(i)}{x(i)},$$
$$1 \geqslant 4y(i)(1-y(i)).$$

Последнее неравенство верно, так как его правая часть максимизируется в точке $y(i) = 1/2$, где она принимает значение 1. Неравенство $u_2(i) \geqslant u_1(i) = \log(t(i)/x(i))$ выполняется, так как

$$\frac{1-y(i)}{x(i)} \geqslant \frac{t(i)}{x(i)} = 2^{u_1(i)}.$$

6. Напомним определение функции $v_2(i)$:

$$v_2(i) := \begin{cases} -\log(2y(i)), & \text{если } y(i) \leqslant 1/2, \\ \log(2(1-y(i))), & \text{если } y(i) > 1/2. \end{cases}$$

По аналогии с доказательством п. 5 предложения 5 мы покажем, что

$$-\log(2y(i)) \geqslant \log(2(1-y(i))). \tag{10}$$

Это неравенство эквивалентно следующим двум:

$$\frac{1}{2y(i)} \geqslant 2(1-y(i)),$$
$$1 \geqslant 4y(i)(1-y(i)).$$

Последнее неравенство верно, так как правая часть принимает максимальное значение 1 в точке $y(i) = 1/2$. Неравенство $v_2(i) \geqslant v_1(i) = \log(2t(i))$ выполняется, так как

$$\log(2(1-y(i))) \geqslant \log(2(1-x(i)-y(i))) = v_1(i). \quad \blacktriangle$$

Доказательство предложения 6. 1. Так как $T(i)$ – центральный, то выполняется $y(i) \leqslant 1/2$. Поэтому верно

$$u_2(i) = -\log(4x(i)y(i)) \geqslant -\log(4y^2(i)) = 2v_2(i).$$

2. Так как отрезок $T(i)$ – центральный и сбалансированный, то верны неравенства $y(i) \leqslant 1/2$ и $y(i)/x(i) \leqslant \beta/\alpha$. Следовательно,

$$u_2(i) = -\log(4x(i)y(i)) = -\log(4y^2(i)x(i)/y(i)) =$$
$$= -\log(4y^2(i)) + \log(y(i)/x(i)) \leqslant 2v_2(i) + \log(\beta/\alpha).$$

3. Без ограничения общности предположим, что $\ell(i-1) \leqslant 1/2 \leqslant r(i-1)$ и $\ell(i) > r(i)$. Из п. 1 предложения 5 следуют неравенства $\ell(i) = \beta\ell(i-1)$ и $r(i) \geqslant \geqslant \alpha r(i-1)$. Тогда выполнено

$$\frac{y(i)}{x(i)} = \frac{\ell(i)}{r(i)} \leqslant \frac{\beta\ell(i-1)}{\alpha r(i-1)} \leqslant \frac{\beta r(i-1)}{\alpha r(i-1)} = \frac{\beta}{\alpha}.$$

Полученное неравенство означает, что $T(i)$ сбалансирован. Аналогичным образом покажем, что отрезок $T(i-1)$ сбалансирован:

$$\frac{y(i-1)}{x(i-1)} = \frac{r(i-1)}{\ell(i-1)} \leqslant \frac{r(i)/\alpha}{\ell(i)/\beta} < \frac{\beta\ell(i)}{\alpha\ell(i)} = \frac{\beta}{\alpha}.$$

4. Для начала отметим, что $u_1' < u_1''$. Без ограничения общности можно считать, что $\ell(i-1) < r(i-1) \leqslant 1/2$. Условие $u_1(i-1) \leqslant u_1''$ эквивалентно следующему:

$$\frac{t(i-1)}{x(i-1)} = \frac{1-\ell(i-1)-r(i-1)}{\ell(i-1)} = \frac{1-r(i-1)}{\ell(i-1)} - 1 \leqslant \frac{\beta}{\alpha} - 1.$$

Полученное неравенство можно переписать в виде $\beta\ell(i-1)+\alpha r(i-1) \geqslant \alpha$ и $\beta\ell(i-1) \geqslant \alpha - \alpha r(i-1)$. Так как передача произошла без ошибки, то $\ell(i) = \beta\ell(i-1)$. Следовательно, $\ell(i) \geqslant \alpha - \alpha r(i-1)$, а так как $r(i-1) \leqslant 1/2$, то $\ell(i) > \alpha/2 > \alpha r(i-1) = r(i)$. Это доказывает, что между моментами $i-1$ и i произошел переход.

5. Без ограничения общности предположим, что $\ell(i-1) \leqslant r(i-1)$. Из условия $u_1(i-1) = \log(t(i-1)/x(i-1)) \geqslant u_1'$ выводим

$$\frac{1-2\ell(i-1)}{\ell(i-1)} \geqslant \frac{1-\ell(i-1)-r(i-1)}{\ell(i-1)} = \frac{t(i-1)}{x(i-1)} \geqslant 2^{u_1'} = \beta - \alpha = 2\beta - 2$$

и $\ell(i-1) \leqslant 1/(2\beta)$. Так как произошел переход, то $\beta\ell(i-1) = \ell(i) \geqslant r(i)$, и следовательно, $y(i) = \ell(i) \leqslant 1/2$.

6. Без ограничения общности предположим, что $\ell(i-1) < r(i-1)$ и $\ell(i) \geqslant r(i)$. Для начала отметим, что переход может произойти только в случае правильно принятого символа. Из условия $u_1(i-1) \geqslant u_1''$ получаем, что

$$\frac{t(i-1)}{x(i-1)} = \frac{1-\ell(i-1)-r(i-1)}{\ell(i-1)} = \frac{1-r(i-1)}{\ell(i-1)} - 1 \geqslant \frac{\beta}{\alpha} - 1.$$

Перепишем это неравенство в следующем виде: $\beta\ell(i-1) + \alpha r(i-1) \leqslant \alpha$. Так как передача произошла без ошибки, то $\ell(i) = \beta\ell(i-1)$, а значит, $\ell(i) + \alpha r(i-1) \leqslant \alpha$ или $r(i-1) \leqslant 1 - \ell(i)/\alpha$. Так как $\ell(i) \geqslant r(i)$, то $r(i-1) \leqslant 1 - r(i)/\alpha$. Из условия $r(i) \geqslant r(i-1)\alpha$ получаем неравенство $r(i-1) \leqslant 1 - r(i-1)$, означающее, что отрезок $T(i-1)$ является центральным. Для доказательства того, что отрезок $T(i)$ центральный, заметим, что $u_1(i-1) \geqslant u_1'' \geqslant u_1'$, и поэтому можно воспользоваться п. 5 предложения 6.

7. Предположим, что произошел переход между моментами $i-1$ и i или между моментами i и $i+1$. Также без ограничения общности будем считать, что $\ell(i) \geqslant r(i)$ и $\ell(i-1) < r(i-1)$ или $\ell(i+1) < r(i+1)$. Доказывая от противного, предположим, что отрезок $T(i)$ не централен, а значит, $y(i) = \ell(i) > 1/2$. Условие $u_2(i) \geqslant u_2'$ эквивалентно следующему:

$$\frac{1-y(i)}{x(i)} = \frac{1-\ell(i)}{r(i)} \geqslant \frac{\beta}{\alpha}.$$

Отсюда следует неравенство $\alpha\ell(i) + \beta r(i) \leqslant \alpha$. Так как $\ell(i) > 1/2$, то получаем, что выполняется неравенство $r(i) \geqslant \alpha r(i-1) > \alpha\ell(i-1) = \alpha\ell(i)/\beta \geqslant \alpha/(2\beta)$ или неравенство $\beta r(i) = r(i+1) > \ell(i+1) \geqslant \ell(i)\alpha \geqslant \alpha/2$. Таким образом, $\alpha\ell(i) + \beta r(i) > \alpha/2 + \alpha/2 = \alpha$. Полученное противоречие показывает, что наше предположение было неверно, а значит, отрезок $T(i)$ является центральным. ▲

Доказательство предложения 7. Сначала посмотрим на первую часть предложения, в которой выполняется условие $u_1(i-1) < u_1'$. Придется рассмотреть несколько разных случаев. Предположим, что при передаче произошла ошибка. Тогда $t(i) \geqslant \alpha t(i-1)$ (п. 4 предложения 5) и $x(i) = \alpha x(i-1)$, а значит, $u_1(i) = \log(t(i)/x(i)) \geqslant \log(t(i-1)/x(i-1)) = u_1(i-1)$. Теперь рассмотрим случай безошибочной передачи. Если перехода не было, то из п. 4 предложения 6 следует, что отрезок $T(i-1)$ должен быть не центральным, поэтому $t(i) = \beta t(i-1)$ и $x(i) = \beta x(i-1)$, что влечет $u_1(i) = u_1(i-1)$.

Теперь предположим, что ошибки не было и произошел переход. Сначала рассмотрим случай, в котором отрезок $T(i-1)$ не является центральным. Так как был переход, то $x(i) \leqslant \beta x(i-1)$ и

$$2^{u_1(i)} = \frac{t(i)}{x(i)} \geqslant \frac{\beta t(i-1)}{\beta x(i-1)} = \frac{t(i-1)}{x(i-1)} = 2^{u_1(i-1)}.$$

Остается разобраться со случаем центрального отрезка $T(i-1)$. Нужно доказать следующее неравенство:

$$u_1(i) - u_1(i-1) =$$
$$= \log\left(\frac{1 - x(i-1)\beta - y(i-1)\alpha}{y(i-1)\alpha}\right) - \log\left(\frac{1 - x(i-1) - y(i-1)}{x(i-1)}\right) \geqslant \log\beta.$$

Это неравенство эквивалентно следующим двум:

$$(1 - x(i-1)\beta - y(i-1)\alpha)x(i-1) \geqslant (1 - x(i-1) - y(i-1))y(i-1)\alpha\beta,$$
$$x(i-1)\frac{\beta - \alpha}{2}(y(i-1) - x(i-1)) \geqslant (1 - x(i-1) - y(i-1))(y(i-1)\alpha\beta - x(i-1)).$$

Так как левая часть этого выражения всегда положительна, то достаточно рассмотреть только ситуацию, когда и правая часть положительна. Так как $u_1(i-1) < u_1'$, то для завершения доказательства достаточно показать, что

$$y(i-1) - x(i-1) > 2(y(i-1)\alpha\beta - x(i-1)),$$
$$x(i-1) > y(i-1)(2\alpha\beta - 1).$$

Так как произошел переход, то согласно п. 3 предложения 6 имеем

$$x(i-1) > y(i-1)\frac{\alpha}{\beta}.$$

Заметим, что $2\alpha\beta - 1 < \frac{\alpha}{\beta}$. Действительно, это неравенство эквивалентно следующим:

$$2\alpha\beta - 1 < \frac{\alpha}{\beta},$$
$$2\alpha\beta^2 - \beta < \alpha = 2 - \beta,$$
$$\alpha\beta^2 < 1.$$

Последнее неравенство следует из условия (3). Доказательство первой части предложения завершено.

Теперь предположим, что $u_1(i-1) \geqslant u_1'$, и покажем, что в этом случае $u_1(i) \geqslant u_1'$. Сначала рассмотрим случай передачи с ошибкой. Тогда верно неравенство

$$\frac{t(i)}{x(i)} \geqslant \frac{\alpha t(i-1)}{\alpha x(i-1)} = \frac{t(i-1)}{x(i-1)}.$$

Таким образом, теперь мы можем считать, что передача была безошибочной. Сначала рассмотрим случай нецентрального отрезка $T(i-1)$. Верно следующее:

$$\frac{t(i)}{x(i)} \geqslant \frac{t(i-1)\beta}{x(i-1)\beta} = \frac{t(i-1)}{x(i-1)},$$

где неравенство следует из условия $\beta x(i-1) \geqslant x(i)$. При этом равенство достигается в случае отсутствия перехода, а если же переход произошел, то неравенство строгое. Таким образом, случай нецентрального отрезка $T(i-1)$ полностью разобран.

Теперь приступим к анализу случая центрального отрезка $T(i-1)$. Предположим, что перехода не было. Тогда верно следующее:

$$\frac{t(i)}{x(i)} \geqslant \frac{t(i-1)}{\beta x(i-1)} \geqslant \frac{\beta-\alpha}{\alpha\beta} > \beta-\alpha.$$

Действительно, первое неравенство верно, так как безошибочная передача не может уменьшить длину истинного отрезка. Второе выполнено, так как в противном случае из п. 4 предложения 6 следовало бы наличие перехода. Последнее неравенство следует из того, что $\alpha\beta < 1$.

Осталось рассмотреть случай, когда произошел переход. Заметим, что так как произошел переход, то из п. 3 предложения 6 следует сбалансированность отрезка $T(i-1)$. Следовательно, $y(i-1)/x(i-1) < \beta/\alpha$. Так как произошел переход, то верна цепочка равенств

$$\frac{t(i)}{x(i)} = \frac{1-x(i-1)\beta-y(i-1)\alpha}{y(i-1)\alpha} = \frac{1-x(i-1)\beta}{y(i-1)\alpha} - 1. \tag{11}$$

Из условия $u_1(i-1) \geqslant u_1'$ получаем

$$\begin{aligned}
&\frac{1-x(i-1)-y(i-1)}{x(i-1)} \geqslant \beta-\alpha,\\
&1-x(i-1)-y(i-1) \geqslant (\beta-\alpha)x(i-1),\\
&1-x(i-1)-\frac{\beta-\alpha}{2}x(i-1)-y(i-1) \geqslant \frac{\beta-\alpha}{2}x(i-1),\\
&1-\beta x(i-1)-y(i-1) \geqslant \frac{\beta-\alpha}{2}x(i-1).
\end{aligned}$$

Используя последнее неравенство, выводим

$$\frac{1-x(i-1)\beta}{y(i-1)\alpha} \geqslant \frac{\frac{\beta-\alpha}{2}x(i-1)+y(i-1)}{y(i-1)\alpha} = \frac{1}{\alpha} + \frac{x(i-1)\frac{\beta-\alpha}{2}}{y(i-1)\alpha}. \tag{12}$$

Из сбалансированности отрезка $T(i-1)$ получаем

$$\begin{aligned}
&\frac{1}{\alpha} + \frac{x(i-1)\frac{\beta-\alpha}{2}}{y(i-1)\alpha} > \frac{1}{\alpha} + \frac{\beta-\alpha}{2\beta} = \frac{2\beta+\alpha\beta-\alpha^2}{2\alpha\beta} =\\
&= \frac{(\alpha+\beta)\beta+\alpha\beta+\alpha^2-2\alpha^2}{2\alpha\beta},
\end{aligned} \tag{13}$$

что можно упростить до

$$\frac{(\alpha+\beta)^2-2\alpha^2}{2\alpha\beta} = \frac{2-\alpha^2}{\alpha\beta} = 1 + \frac{2-2\alpha}{\alpha\beta} = 1 + \frac{\beta-\alpha}{\alpha\beta} > \beta-\alpha+1, \tag{14}$$

где последнее неравенство верно, так как $1 > \alpha\beta$. Применяя равенство (11) и неравенства (12)–(14), получаем требуемое.

Теперь покажем, что из неравенства $u_1(i-1) \geqslant u_1''$ следует $u_1(i) \geqslant u_1''$. Доказывая от противного, предположим, что $u_1(i-1) \geqslant u_1''$ и $u_1(i) < u_1''$, т.е. u_1 убывает. Случаи нецентрального положения отрезка $T(i-1)$ или ошибочной передачи разбираются так же, как и в предыдущей части доказательства. Поэтому будем считать, что $T(i-1)$ центральный и передача произошла без ошибки. Следующие неравенства

эквивалентны друг другу:

$$u_1(i-1) \geqslant u_1'' = \log(\beta/\alpha - 1),$$
$$\frac{1 - x(i-1) - y(i-1)}{x(i-1)} \geqslant \beta/\alpha - 1,$$
$$\alpha \geqslant \alpha y(i-1) + \beta x(i-1).$$

Рассуждая похожим образом, можно показать эквивалентность

$$u_1(i) < u_1'',$$
$$\alpha < \alpha y(i) + \beta x(i).$$

Так как отрезок $T(i-1)$ – центральный и передача произошла без ошибки, то выполнено равенство $\alpha y(i-1) + \beta x(i-1) = x(i) + y(i)$. Это равенство и приводит к искомому противоречию

$$\alpha \geqslant \alpha y(i-1) + \beta x(i-1) = y(i) + x(i) \geqslant \alpha y(i) + \beta x(i) > \alpha. \quad \blacktriangle$$

Доказательство предложения 8. Сначала рассмотрим случай, когда оба отрезка $T(i-1)$ и $T(i)$ являются центральными. Вне зависимости от того, происходит ли ошибка, верно следующее:

$$u_2(i) - u_2(i-1) = -\log(4x(i-1)\beta y(i-1)\alpha) + \log(4x(i-1)y(i-1)) =$$
$$= -\log(\alpha\beta) > 0,$$

где последнее неравенство следует из условия $\alpha\beta < 1$.

Перейдем к случаю, когда отрезок $T(i-1)$ центральный, а $T(i)$ – нет. Пункт 5 предложения 6 позволяет утверждать, что перехода не происходило, а из п. 3 предложения 5 мы понимаем, что при передаче произошла ошибка. Преобразуем выражение:

$$u_2(i) - u_2(i-1) =$$
$$= \log\left(\frac{1-y(i)}{x(i)}\right) + \log(4x(i-1)y(i-1)) = \log\left(\frac{4(1-y(i))y(i)}{\alpha\beta}\right).$$

Поскольку $T(i)$ – не центральный, то $1/2 \leqslant y(i) \leqslant \dfrac{\beta}{2}$. Значение $y(i) = \dfrac{\beta}{2}$ минимизирует логарифм в правой части последнего равенства, поэтому

$$\log\left(\frac{4(1-y(i))y(i)}{\alpha\beta}\right) \geqslant \log 1 = 0.$$

Теперь рассмотрим случай, когда отрезок $T(i-1)$ не является центральным, причем произошел переход. Без ограничения общности полагаем $\ell(i-1) < r(i-1)$. Из п. 5 предложения 6 заключаем, что $T(i)$ является центральным. Поэтому верна цепочка преобразований

$$u_2(i) - u_2(i-1) = -\log(4\ell(i)r(i)) - \log\left(\frac{1-r(i-1)}{\ell(i-1)}\right) =$$
$$= -\log\left(\frac{4(1-r(i-1))\ell(i)r(i)}{\ell(i-1)}\right) = -\log\big(4\beta(1-r(i-1))r(i)\big).$$

Используя равенство

$$r(i) = 1 - \beta(t(i-1) + \ell(i-1)) = \frac{\alpha}{2} + \beta\left(r(i-1) - \frac{1}{2}\right),$$

получаем

$$-\log(4\beta(1-r(i-1))r(i)) = -\log\left(4\beta(1-r(i-1))\left(\frac{\alpha}{2}+\beta\left(r(i-1)-\frac{1}{2}\right)\right)\right).$$

Выражение под логарифмом является квадратичной функцией, достигающей максимума при $r(i-1) = \dfrac{3\beta-\alpha}{4\beta}$. Поэтому

$$-\log\left(4\beta(1-r(i-1))\left(\frac{\alpha}{2}+\beta\left(r(i-1)-\frac{1}{2}\right)\right)\right) \geqslant$$
$$\geqslant -\log\left((\beta+\alpha)\left(\frac{\alpha}{2}+\frac{\beta-\alpha}{4}\right)\right) = -\log 1 = 0.$$

Осталось рассмотреть случай, когда отрезок $T(i-1)$ не является центральным и отсутствует переход. Поскольку $\dfrac{1}{4x(i)y(i)} \geqslant \dfrac{1-y(i)}{x(i)}$, можно вывести, что

$$u_2(i) - u_2(i-1) \geqslant \log\left(\frac{1-y(i)}{x(i)}\right) - \log\left(\frac{1-y(i-1)}{x(i-1)}\right) = 0.$$

Равенство нулю выполнено, поскольку нет перехода, откуда следует

$$1-y(i) = \begin{cases}(1-y(i-1))\beta, & \text{если передача безошибочна,}\\ (1-y(i-1))\alpha, & \text{если произошла ошибка,}\end{cases}$$

и

$$x(i) = \begin{cases}x(i-1)\beta, & \text{если передача безошибочна,}\\ x(i-1)\alpha, & \text{если произошла ошибка,}\end{cases}$$

что завершает доказательство. ▲

Доказательство предложения 9. Если произошел переход, то согласно п. 1 предложения 5 прием должен быть безошибочным. В силу п. 2 предложения 5 мы знаем, что в этом случае истинный отрезок увеличивается, что доказывает первое утверждение.

Для доказательства второго утверждения мы сначала рассмотрим случай нецентрального отрезка $T(i-1)$. В этом случае верны равенства

$$v_1(i) - v_1(i-1) = \log t(i) - \log t(i-1) = \begin{cases}\log\beta, & \text{если передача безошибочна,}\\ \log\alpha, & \text{если произошла ошибка.}\end{cases}$$

Теперь рассмотрим случай центрального отрезка $T(i-1)$, причем перехода не было. В этом случае при передаче должна была произойти ошибка, так как иначе мы получим противоречие с п. 4 предложения 6. Получается, что нам надо доказать

$$\log t(i) - \log t(i-1) \geqslant \log\alpha,$$

а это верно в силу п. 4 предложения 5. ▲

Доказательство предложения 4. Сначала рассмотрим случай $u_1(i-1) \geqslant$ $\geqslant u_1''$, причем происходил переход. Из п. 6 предложения 6 мы знаем, что как $T(i-1)$, так и $T(i)$ являются центральными отрезками. Отсюда следует, что

$$v_2(i) - v_2(i-1) = -\log(2y(i)) + \log(2y(i-1)) =$$
$$= -\log(2x(i-1)\beta) + \log(2y(i-1)) \geqslant -\log\beta.$$

Теперь рассмотрим случаи без перехода. Пусть отрезки $T(i-1)$ и $T(i)$ являются центральными. Получаем, что

$$v_2(i) - v_2(i-1) = \log\left(\frac{y(i-1)}{y(i)}\right) =$$
$$= \begin{cases} -\log\alpha \geqslant \log\beta, & \text{если передача безошибочна,} \\ -\log\beta \geqslant \log\alpha, & \text{если произошла ошибка.} \end{cases}$$

Теперь перейдем к случаю, когда отрезок $T(i-1)$ является центральным, а $T(i)$ – нет. Это может произойти только в случае передачи с ошибкой (п. 3 предложения 5), а значит,

$$v_2(i) - v_2(i-1) = \log(2(1-y(i))) + \log(2y(i-1)) = \log(4(1-y(i))y(i)) - \log\beta,$$

но так как $y(i-1) \leqslant 1/2$, то $y(i) \leqslant \dfrac{\beta}{2}$. Таким образом, получаем

$$v_2(i) - v_2(i-1) = \log(4(1-y(i))y(i)) - \log\beta \geqslant \log(\alpha\beta) - \log\beta = \log\alpha.$$

В рассматриваемом случае также выполнено неравенство $v_2(i-1) < \log\beta$, которое следует из цепочки эквивалентных преобразований

$$v_2(i-1) < \log\beta,$$
$$-\log(2y(i-1)) < \log\beta,$$
$$\frac{1}{2} < \beta y(i-1) = y(i).$$

Здесь последнее неравенство следует из того, что отрезок $T(i)$ не является центральным.

Рассмотрим случай, когда оба отрезка $T(i-1)$ и $T(i)$ не являются центральными, причем перехода не происходило. В этом случае выполнена цепочка равенств

$$v_2(i) - v_2(i-1) = \log(2(1-y(i))) - \log(2(1-y(i-1))) =$$
$$= \log\left(\frac{t(i)+x(i)}{t(i-1)+x(i-1)}\right) = \begin{cases} \log\beta, & \text{если передача безошибочна,} \\ \log\alpha, & \text{если произошла ошибка.} \end{cases}$$

Мы покажем, что верно неравенство $v_2(i-1) < \log\beta$, если произошла ошибка. Заметим, что верно неравенство

$$\log(2(1-y(i-1))) \leqslant -\log(2y(i-1)).$$

Тогда неравенство $v_2(i-1) < \log\beta$ следует из цепочки эквивалентных преобразований

$$-\log(2y(i-1)) < \log\beta,$$
$$\frac{1}{2y(i-1)} < \beta,$$
$$\frac{1}{2} < \beta y(i-1) = y(i),$$

где последнее неравенство верно, так как отрезок $T(i)$ не является центральным и произошла ошибка.

Наконец, рассмотрим случай, когда отрезок $T(i)$ является центральным, а отрезок $T(i-1)$ – нет. В этом случае ошибок при передаче не происходило, так как

в противном случае выполнено $y(i) \geqslant y(i-1) \geqslant 1/2$. Поскольку

$$-\log(2y(i)) \geqslant \log(2(1-y(i))),$$

то

$$v_2(i) - v_2(i-1) = -\log(2y(i)) - \log(2(1-y(i-1))) \geqslant \log\left(\frac{1-y(i)}{1-y(i-1)}\right) = \log\beta,$$

что завершает доказательство. ▲

Доказательство предложения 10. Из п. 2 предложения 6 и неравенства (3) можно вывести

$$v_2(i_0) > \frac{u_2(i_0)}{2} - \frac{\log(\beta/\alpha)}{2} \geqslant -\frac{1}{2}\log(\alpha^4\beta^6) \geqslant -\frac{1}{2}\log\beta^{-2} = \log\beta. \tag{15}$$

Поскольку $v_2(i_0) = -\log(2y(i_0)) > \log\beta$, то $y(i_0) < 1/(2\beta)$. Из $y(i_0) < 1/(2\beta)$ получаем

$$u_1(i_0) = \log\left(\frac{1-y(i_0)-x(i_0)}{x(i_0)}\right) \geqslant \log\left(\frac{1-2y(i_0)}{y(i_0)}\right) > u_1'.$$

Из предложения 7 следует, что $u_1(i) \geqslant u_1'$ для всех $i \geqslant i_0$. Предложение 8 влечет неравенство $u_2(i) \geqslant c$ для всех $i \geqslant i_0$. Используя те же рассуждения, что и в (15), получаем неравенство $v_2(i) > \log\beta$ для всех i, для которых отрезок $T(i)$ является центральным и сбалансированным. Отметим, что так как $c \geqslant u_2'$, то выполняется неравенство $u_2(i) \geqslant u_2'$.

Пусть в моменты $j_1, \ldots, j_k$, $i_0 \leqslant j_1 < \ldots < j_k < i_1$, происходили переходы. Представим полуинтервал $[i_0, i_1)$ в виде объединения непересекающихся полуинтервалов $[i_0, j_1) \cup [j_1, j_1+1) \cup [j_1+1, j_2) \cup \ldots \cup [j_k, i_1)$. Из пп. 3 и 7 предложения 6 нам известно, что в моменты, соответствующие концам полуинтервалов, истинный отрезок является центральным и сбалансированным, а значит, все условия для применения предложения 10 выполнены. Более того, выполнение неравенства (4) для этих полуинтервалов влечет выполнение неравенства и для их объединения. Таким образом, достаточно рассмотреть только два случая:

1. $i_1 = i_0 + 1$ и произошел переход;
2. Между моментами i_0 и i_1 переходов не было.

Первый случай тривиален. Так как между моментами i_0 и $i_1 = i_0+1$ был переход, то согласно п. 1 предложения 5 передача произошла без ошибки. В этом случае $g(i_0, i_1) = \log\beta$. С другой стороны, $u_2(i_1) - u_2(i_0) = -\log(\alpha\beta)$ по предложению 8, что не меньше чем $\log\beta$ благодаря неравенству (3).

Перейдем ко второму случаю, в котором между моментами i_0 и i_1 нет переходов. Минимальный отрезок все время остается минимальным, следовательно,

$$g(i_0, i_1) = e(i_0, i_1)\log\alpha + f(i_0, i_1)\log\beta = \log\left(\frac{x(i_1)}{x(i_0)}\right).$$

Обозначим $v_2(i_1) - v_2(i_0)$ и $u_2(i_1) - u_2(i_0)$ через Δv_2 и Δu_2 соответственно. Так как $T(i_0)$ и $T(i_1)$ являются центральными, то

$$\Delta u_2 = -\log(x(i_1)y(i_1)) + \log(x(i_0)y(i_0)) = -\log y(i_1) + \log y(i_0) - \log\left(\frac{x(i_1)}{x(i_0)}\right) = \\ = \Delta v_2 - g(i_0, i_1),$$

что эквивалентно следующему:

$$g(i_0, i_1) = \Delta v_2 - \Delta u_2.$$

Используя пп. 1 и 2 предложения 6 для моментов i_1 и i_0 соответственно, получаем

$$g(i_0, i_1) = \Delta v_2 - \Delta u_2 \leqslant \frac{\Delta u_2 + \log\left(\frac{\beta}{\alpha}\right)}{2} - \Delta u_2 = \frac{-\Delta u_2 + \log\left(\frac{\beta}{\alpha}\right)}{2}.$$

Вспомним неравенство $y(i_0) < 1/(2\beta)$, откуда следует $y(i_0 + 1) < 1/2$, а это значит, что $T(i_0 + 1)$ является центральным. Используя предложение 8, получаем

$$\Delta u_2 = u_2(i_1) - u_2(i_0 + 1) + u_2(i_0 + 1) - u_2(i_0) \geqslant -\log(\alpha\beta).$$

Используя это неравенство вместе с неравенством (3), получаем

$$g(i_0, i_1) \leqslant \frac{-\Delta u_2 + \log\left(\frac{\beta}{\alpha}\right)}{2} \leqslant \frac{\log(\alpha\beta) + \log\left(\frac{\beta}{\alpha}\right)}{2} = \log\beta \leqslant -\log(\alpha\beta) \leqslant \Delta u_2. \quad \blacktriangle$$

Доказательство предложения 3. Вспомним, что согласно п. 5 предложения 5 неравенство $u_2(i) \geqslant u_1(i)$ выполняется для всех $x(i)$, $y(i)$ и $t(i)$.

Если $u_1(i-1) < u_1'$, то $u_1(i) \geqslant u_1(i-1)$ по предложению 7, следовательно, $u(i) \geqslant u_1(i) \geqslant u_1(i-1) = u(i-1)$. Если $T(i-1)$ центральный и произошел переход, то $u_1(i) - u_1(i-1) \geqslant \log\beta$ по предложению 7, значит, $u(i) - u(i-1) \geqslant \log\beta$.

Если же $u_1(i-1) \geqslant u_1'$, то $u_1(i) \geqslant u_1'$ по предложению 7 и $u_2(i) \geqslant u_2(i-1)$ по предложению 8. Это означает, что $u(i) = u_2(i) \geqslant u_2(i-1) = u(i-1)$. Если отрезок $T(i-1)$ центральный и произошел переход, то $T(i)$ тоже центральный согласно п. 5 предложения 6. Поэтому можно использовать предложение 8, что дает $u_2(i) - u_2(i-1) \geqslant -\log(\alpha\beta)$. Так как $\alpha\beta^2 \leqslant 1$ (см. (3)), то это выражение не меньше $\log\beta$. $\blacktriangle$

Доказательство предложения 2. Напомним, что $v_2(i) \geqslant v_1(i)$ для любых $x(i)$, $y(i)$ и $t(i)$ согласно п. 6 предложения 5.

Предположим, что $u_1(i-1) < u_1''$. Если перехода не происходило, то утверждение верно благодаря предложению 9. Если переход был, то предложение 9 дает неравенство $v_1(i) > v_1(i-1)$, следовательно, $v(i) \geqslant v_1(i) > v_1(i-1) = v(i-1)$, т.е. $v(i) - v(i-1) > 0$. Для центрального отрезка $T(i-1)$ эта оценка достаточно хороша. Остается рассмотреть случай, когда $T(i-1)$ не центральный. Из условия $u_1(i-1) < u_1''$ следует $v(i) - v(i-1) \geqslant v_1(i) - v(i-1) = v_1(i) - v_1(i-1)$. Так как произошел переход, то передача произошла без ошибки; условие нецентральности отрезка $T(i-1)$ влечет соотношения $t(i) = \beta t(i-1)$ и $v(i) - v(i-1) \geqslant v_1(i) - v_1(i-1) = \log\beta$.

Предположим, что $u_1(i-1) \geqslant u_1''$. В этом случае $u_1(i) \geqslant u_1''$ по предложению 7. Если не было перехода, то доказываемое предложение следует из предложения 4. Если переход был, то отрезок $T(i-1)$ является центральным согласно п. 6 предложения 6. Тогда $v(i) - v(i-1) = v_2(i) - v_2(i-1) \geqslant -\log\beta$ по предложению 4. $\blacktriangle$

§ 5. Заключение

В данной статье мы рассмотрели новую задачу передачи информации по комбинаторному каналу со вставками и выпадениями и обратной связью. Мы показали, как эта задача может быть сведена к задаче передачи информации по комбинаторному каналу с замещениями. Таким образом, была установлена максимальная асимптотическая скорость кодов для комбинаторного канала со вставками и выпадениями. В частности, скорость положительна, если доля ошибок меньше $1/2$. Мы также напомнили алгоритм Хорштейна [10] для комбинаторного канала с замещениями и привели более подробную версию анализа Зигангирова [9] этого алгоритма.

Подчеркнем, что все результаты, рассматриваемые в данной статье, касаются лишь двоичного канала. Возникает естественный вопрос: можно ли обобщить эти ре-

зультаты для недвоичного случая? Наилучшие на текущий момент результаты для q-ичного канала с обратной связью и замещениями описаны в работе [12]. В частности, если доля ошибок находится в интервале $0 < \tau < 1/q$, то максимальная асимптотическая скорость установлена лишь для счетного числа значений τ. Отметим также, что перенести рассуждения Зигангирова с двоичного случая на q-ичный случай достаточно сложно, поскольку не очевидны аналоги для функций u и v.

Авторы благодарны Цзылинь Цзяну за полезные замечания и комментарии при обсуждении доказательства Зигангирова.

СПИСОК ЛИТЕРАТУРЫ

1. *Левенштейн В.И.* Двоичные коды с исправлением выпадений, вставок и замещений символов // ДАН СССР. 1965. Т. 163. № 4. С. 845–848. http://mi.mathnet.ru/dan31411
2. *Варшамов Р.Р., Тенengольц Г.М.* Код, исправляющий одиночные несимметрические ошибки // АиТ. 1965. Т. 26. № 2. С. 288–292. http://mi.mathnet.ru/at11293
3. *Cheraghchi M., Ribeiro J.* An Overview of Capacity Results for Synchronization Channels // IEEE Trans. Inform. Theory. 2020. V. 67. № 6. P. 3207–3232. https://doi.org/10.1109/TIT.2020.2997329
4. *Schulman L.J., Zuckerman D.* Asymptotically Good Codes Correcting Insertions, Deletions, and Transpositions // IEEE Trans. Inform. Theory. 1999. V. 45. № 7. P. 2552–2557. https://doi.org/10.1109/18.796406
5. *Bukh B., Guruswami V.* An Improved Bound on the Fraction of Correctable Deletions // Proc. 27th Annu. ACM–SIAM Symp. on Discrete Algorithms (SODA'2016). Arlington, VA, USA. Jan. 10–12, 2016. P. 1893–1901. https://doi.org/10.1137/1.9781611974331.ch133
6. *Bukh B., Guruswami V., Håstad, J.* An Improved Bound on the Fraction of Correctable Deletions // IEEE Trans. Inform. Theory. 2016. V. 63. № 1. P. 93–103. https://doi.org/10.1109/TIT.2016.2621044
7. *Plotkin M.* Binary Codes with Specified Minimum Distance // IRE Trans. Inform. Theory. 1960. V. 6. № 4. P. 445–450. https://doi.org/10.1109/TIT.1960.1057584
8. *Berlekamp E.R.* Block Coding with Noiseless Feedback. PhD Thesis. MIT, Cambridge, USA, 1964.
9. *Зигангиров К.Ш.* О числе исправляемых ошибок при передаче по ДСК с обратной связью // Пробл. передачи информ. 1976. Т. 12. № 2. С. 3–19. http://mi.mathnet.ru/ppi1683
10. *Horstein M.* Sequential Transmission Using Noiseless Feedback // IEEE Trans. Inform. Theory. 1963. V. 9. № 3. P. 136–143. https://doi.org/10.1109/TIT.1963.1057832
11. *Schalkwijk J.* A Class of Simple and Optimal Strategies for Block Coding on the Binary Symmetric Channel with Noiseless Feedback // IEEE Trans. Inform. Theory. 1971. V. 17. № 3. P. 283–287. https://doi.org/10.1109/TIT.1971.1054625
12. *Лебедев В.С.* Кодирование при наличии бесшумной обратной связи // Пробл. передачи информ. 2016. Т. 52. № 2. С. 3–14. http://mi.mathnet.ru/ppi2200

Маринггер Георг
Технический университет Мюнхена, Германия
georg.maringer@tum.de
Полянский Никита Андреевич
Технический университет Мюнхена, Германия
Сколковский институт науки и технологий (Сколтех)
nikita.polyansky@gmail.com
Воробьев Илья Викторович
Сколковский институт науки и технологий (Сколтех)
vorobyev.i.v@yandex.ru
Вельтер Лоренц
Технический университет Мюнхена, Германия
lorenz.welter@tum.de

Поступила в редакцию
14.01.2021
После доработки
16.02.2021
Принята к публикации
20.06.2021

ПРОБЛЕМЫ ПЕРЕДАЧИ ИНФОРМАЦИИ

Том 57 **2021** **Вып. 3**

УДК 621.391 : 519.725

© 2021 г. **Э.М. Габидулин, Н.И. Пилипчук, О.В. Трушина**

ГРАНИЦЫ МОЩНОСТИ ПОДПРОСТРАНСТВЕННЫХ КОДОВ С НЕМАКСИМАЛЬНЫМ КОДОВЫМ РАССТОЯНИЕМ

Изучаются подпространственные коды с немаксимальным кодовым расстоянием. В отличие от кодов-спредов, т.е. кодов с максимальным подпространственным расстоянием, здесь они названы неспредами. Рассмотрены семейства неспредов с использованием конструкций подпространственных кодов Силвы – Кёттера – Кшишанга (SKK) и многокомпонентных кодов с нулевым префиксом (МНП) Габидулина – Боссерта. Оценены мощности неспредов для большого числа параметров. Показано, что при больших размерностях мощности практически достигают верхней границы, соответствующей неравенству Джонсона.

Ключевые слова: конечное поле, код, спреды, декодирование, пространство, подпространство, мощность кода, ранговая метрика.

DOI: 10.31857/S0555292321030037

§ 1. Введение

Интерес к подпространственным кодам возрос в связи с развитием идей случайного сетевого кодирования [1] и распределенных систем хранения [2], где они нашли свое применение.

Подпространственный код – это некоторое множество подпространств из заданного пространства. Зададим конечное n-мерное пространство $W = GF(q)^n$ над конечным полем $GF(q)$. Пусть $W(n, m)$ – множество всех m-мерных подпространств пространства W, называемое *m-грассманианом*. Размер грассманиана определяется с помощью гауссовских коэффициентов:

$$|W(n,m)| = \begin{bmatrix} n \\ m \end{bmatrix} = \frac{(q^n-1)(q^n-q)\dots(q^n-q^{m-1})}{(q^m-1)(q^m-q)\dots(q^m-q^{m-1})}.$$

Между двумя подпространствами $U, V \in W$ можно определить *подпространственное расстояние* в виде

$$d_{\mathrm{sub}}(U,V) = \dim(U \uplus V) - \dim(U \cap V) = \\ = \dim(U) + \dim(V) - 2\dim(U \cap V),$$

где $U \uplus V$ означает минимальное подпространство, содержащее оба подпространства U и V. Если U и V имеют одну и ту же размерность m, то подпространственное расстояние равно

$$d_{\mathrm{sub}}(U,V) = 2(m - \dim(U \cap V)) = 2\delta,$$

где $\delta = m - \dim(U \cap V)$. Это расстояние называется также *грассмановой метрикой*.

Если код состоит из элементов m-грассманиана $W(n,m)$ с числом кодовых подпространств M, минимальным расстоянием d_{sub} и размерностью m, то он называется кодом постоянной размерности и обозначается (M,n,d_{sub},m). Код с максимальным расстоянием $d_{\text{sub}}=2m$ называется *спредом*. Подпространственные коды с расстоянием, отличным от максимального, будем называть *неспредами*.

Выберем некоторый спред с параметрами (n,d_{sub},m). Код с параметрами $(n+1,$ $d_{\text{sub}},m+1)$ является неспредом. Мощности этих двух кодов связаны неравенством Джонсона [3]

$$M(n+1,d_{\text{sub}},m+1)\leqslant\frac{q^{n+1}-1}{q^{m+1}-1}M(n,d_{\text{sub}},m). \tag{1}$$

Величину $K_J=\dfrac{q^{n+1}-1}{q^{m+1}-1}$ будем далее навывать *коэффициентом Джонсона*. Этот коэффициент показывает, что при переходе от спреда к неспреду мощность кода для выбранных параметров может быть максимально увеличена не более чем в K_J раз.

Рассмотрим спред с длиной $n=mt+s$, где t и s – целые числа, причем $0\leqslant s\leqslant$ $\leqslant m-1$. Если $s=0$, то спред называется *полным*, а при $s>0$ – *частичным*. Известно, что существуют оптимальные полные спреды и для некоторых параметров оптимальные частичные спреды. Так, в работе [4] рассматривались МНП-спреды, т.е. спреды, построенные по принципу многокомпонентных кодов с нулевым префиксом (МНП), предложенных в [5]. Было показано, что полные МНП-спреды и частичные МНП-спреды достигают верхней границы мощности.

Мощность полного МНП-спреда равна [4]

$$M_{\text{МНП-спред}}=\frac{q^n-1}{q^m-1}. \tag{2}$$

Используя соотношения (1) и (2), можно получить верхнюю границу мощности неспреда

$$M_{\max}=K_J\frac{q^n-1}{q^m-1}=\frac{q^{n+1}-1}{q^{m+1}-1}\frac{q^n-1}{q^m-1}. \tag{3}$$

§ 2. Неспреды из SKK-спредов

В работах [6, 7] Силва, Кёттер и Кшишанг подробно описали лифтинговую конструкцию своего подпространственного SKK-кода, предназначенного для передачи по сети с помощью случайных линейных преобразований [1]. Этот код состоит из множества матриц вида

$$\mathcal{M}_{\text{SKK}}=\left\{\begin{pmatrix}\boldsymbol{I}_m & \boldsymbol{M}_{m\times(n-m)}\end{pmatrix}\right\},$$

где $\boldsymbol{I}_m$ – единичная матрица порядка m, а $\boldsymbol{M}_{m\times(n-m)}$ – кодовая матрица размера $m\times(n-m)$ из матричного рангового кода $\mathcal{M}_{\text{rank}}$ с ранговым расстоянием $d_{\text{rank}}=\delta$ (см. [8]). Подпространственное расстояние кода $\mathcal{M}_{\text{SKK}}$ равно удвоенному ранговому расстоянию матричного кода.

Мощность SKK-кода равна числу кодовых слов рангового кода с ранговым расстоянием $d_{\text{rank}}=\delta$ и длиной кодовых слов $(n-m)$:

$$M_{\text{SKK}}=|\mathcal{M}_{\text{SKK}}|=|\mathcal{M}_{\text{rank}}|=q^{(n-m)k},$$

где $k=m-\delta+1$, $\delta\leqslant m$. Для SKK-спреда с параметрами $(n,d_{\text{sub}}=2m,m)$, где $\delta=m$, $n=tm$, $t\in\mathbb{N}$, построим неспред, увеличив размерность и длину на 1, т.е. неспред

Таблица 1

Доля мощности SKK-неспреда от максимальной при $q = 2$

m	t								
	2	3	4	5	6	7	8	9	10
2	0,7225806	0,6719160	0,6601128	0,6572124	0,6564904	0,6563101	0,6562650	0,6562538	0,6562509
3	0,8398950	0,8227212	0,8206130	0,8203501	0,8203172	0,8203131	0,8203126	0,8203125	0,8203125
4	0,9135490	0,9085358	0,9082239	0,9082044	0,9082032	0,9082031	0,9082031	0,9082031	0,9082031
5	0,9550118	0,9536569	0,9536146	0,9536133	0,9536133	0,9536133	0,9536133	0,9536133	0,9536133
6	0,9770423	0,9766902	0,9766847	0,9766846	0,9766846	0,9766846	0,9766846	0,9766846	0,9766846
7	0,9884023	0,9883125	0,9883118	0,9883118	0,9883118	0,9883118	0,9883118	0,9883118	0,9883118
8	0,9941710	0,9941483	0,9941483	0,9941483	0,9941483	0,9941483	0,9941483	0,9941483	0,9941483
9	0,9970779	0,9970722	0,9970722	0,9970722	0,9970722	0,9970722	0,9970722	0,9970722	0,9970722
10	0,9985371	0,9985356	0,9985356	0,9985356	0,9985356	0,9985356	0,9985356	0,9985356	0,9985356
...	...	...	...	...	...	...	...	...	...
30	0,9999999	0,9999999	0,9999999	0,9999999	0,9999999	0,9999999	0,9999999	0,9999999	0,9999999

с параметрами $(n+1, d_{\text{sub}} = 2m, m+1)$. Для этих параметров $k = (m+1)-m+1 = 2$. Определим, какую долю составляет мощность SKK-неспреда от максимальной (3):

$$K_{\text{SKK}} = \frac{q^{2(n-m)}}{M_{\max}} = q^{2(n-m)} \frac{q^{m+1}-1}{q^{n+1}-1} \frac{q^m-1}{q^n-1}.$$

Она довольно быстро увеличивается при увеличении m. Значения для разных m и t при $q = 2$ приведены в табл. 1.

§ 3. Неспреды из МНП-спредов

Рассмотрим МНП-спред [4], т.е. $(n, d_{\text{sub}} = 2m, m)$-подпространственный код, где $n = tm$, $t \in \mathbb{N}$, и построим код размерности $m+1$ с длиной кодовых матриц $n+1 = tm+1$, $t \in \mathbb{N}$, и подпространственным расстоянием $d_{\text{sub}} = 2m$. Первой компонентой этого кода является SKK-код, а каждая следующая компонента – сдвиг предыдущей компоненты на нулевой префикс. Таким образом,

1-я компонента $\left(\boldsymbol{I}_{m+1} \quad \boldsymbol{M}^1_{(m+1)\times(n-m)}\right)$,

2-я компонента $\left(\boldsymbol{0}_{(m+1)\times m} \quad \boldsymbol{I}_{m+1} \quad \boldsymbol{M}^2_{(m+1)\times(n-2m)}\right)$,

$(t-2)$-я компонента $\left(\underbrace{\boldsymbol{0}_{(m+1)\times m} \quad \boldsymbol{0}_{(m+1)\times m} \quad \dots \quad \boldsymbol{0}_{(m+1)\times m}}_{t-3} \quad \boldsymbol{I}_{m+1} \quad \boldsymbol{M}^{t-2}_{(m+1)\times 2m}\right)$,

$(t-1)$-я компонента $\left(\underbrace{\boldsymbol{0}_{(m+1)\times m} \quad \boldsymbol{0}_{(m+1)\times m} \quad \dots \quad \boldsymbol{0}_{(m+1)\times m}}_{t-2} \quad \boldsymbol{I}_{m+1} \quad \boldsymbol{M}^{t-1}_{(m+1)\times m}\right)$,

t-я компонента $\left(\underbrace{\boldsymbol{0}_{(m+1)\times m} \quad \boldsymbol{0}_{(m+1)\times m} \quad \dots \quad \boldsymbol{0}_{(m+1)\times m}}_{t-1} \quad \boldsymbol{I}_{m+1}\right)$,

здесь $\boldsymbol{M}$ – матрица рангового кода, где верхний индекс указывает номер компоненты, а нижний – размер матрицы. Для заданных параметров $k = (m+1)-m+1 = 2$.

Мощность этого МНП-неспреда равна сумме мощностей компонент:

$$M_{\text{МНП}} = q^{k(n-m)} + \sum_{i=1}^{t-2} q^{kmi} + 1 = \frac{q^{kn}-1}{q^{km}-1}. \tag{4}$$

Как и для случая SKK-кода, определим, какую долю составляет мощность МНП-неспреда от максимальной $K_{\text{МНП}} = \dfrac{M_{\text{МНП}}}{M_{\max}}$. Значения для разных m и t при $q = 2$ представлены в табл. 2.

Таблица 2

Доля мощности МНП-неспреда от максимальной при $q = 2$

m	t								
	2	3	4	5	6	7	8	9	10
2	0,7677419	0,7165354	0,7041096	0,7010259	0,7002564	0,7000641	0,7000160	0,7000040	0,7000010
3	0,8530184	0,8357771	0,8336385	0,8333715	0,8333381	0,8333339	0,8333334	0,8333333	0,8333333
4	0,9171175	0,9120986	0,9117856	0,9117660	0,9117648	0,9117647	0,9117647	0,9117647	0,9117647
5	0,9559444	0,9545892	0,9545468	0,9545455	0,9545455	0,9545455	0,9545455	0,9545455	0,9545455
6	0,9772809	0,9769287	0,9769232	0,9769231	0,9769231	0,9769231	0,9769231	0,9769231	0,9769231
7	0,9884626	0,9883728	0,9883721	0,9883721	0,9883721	0,9883721	0,9883721	0,9883721	0,9883721
8	0,9941862	0,9941635	0,9941634	0,9941634	0,9941634	0,9941634	0,9941634	0,9941634	0,9941634
9	0,9970817	0,9970760	0,9970760	0,9970760	0,9970760	0,9970760	0,9970760	0,9970760	0,9970760
10	0,9985380	0,9985366	0,9985366	0,9985366	0,9985366	0,9985366	0,9985366	0,9985366	0,9985366
...	...	...	...	...	...	...	...	...	...
30	0,9999999	0,9999999	0,9999999	0,9999999	0,9999999	0,9999999	0,9999999	0,9999999	0,9999999

Рассмотрим отношение

$$\frac{K_{\text{МНП}}}{K_{SKK}} = \frac{q^{kn}-1}{q^{km}-1}\frac{1}{q^{k(n-m)}} = \frac{q^{kn}-1}{q^{kn}-q^{k(n-m)}}.$$

Величина $\dfrac{K_{\text{МНП}}}{K_{SKK}}$ всегда больше единицы, следовательно, МНП-неспреды быстрее приближаются к максимальной мощности с ростом размерности m.

§ 4. Неспреды из частичных МНП-спредов

Рассмотрим частичный МНП-спред $(n, d_{\text{sub}} = 2m, m)$ [4], где $n = tm + s$, $t \in \mathbb{N}$, $1 \leqslant s < m$, и построим из него неспред, увеличив длину и размерность на единицу:

1-я компонента $\quad \left(\boldsymbol{I}_{m+1} \quad \boldsymbol{M}^1_{(m+1)\times(n-m)}\right),$

2-я компонента $\quad \left(\boldsymbol{0}_{(m+1)\times m} \quad \boldsymbol{I}_{m+1} \quad \boldsymbol{M}^2_{(m+1)\times(n-2m)}\right),$

$(t-2)$-я компонента $\quad \left(\underbrace{\boldsymbol{0}_{(m+1)\times m} \quad \boldsymbol{0}_{(m+1)\times m} \quad \dots \quad \boldsymbol{0}_{(m+1)\times m}}_{t-3} \quad \boldsymbol{I}_{m+1} \quad \boldsymbol{M}^{t-2}_{(m+1)\times(2m+s)}\right),$

$(t-1)$-я компонента $\quad \left(\underbrace{\boldsymbol{0}_{(m+1)\times m} \quad \boldsymbol{0}_{(m+1)\times m} \quad \dots \quad \boldsymbol{0}_{(m+1)\times m}}_{t-2} \quad \boldsymbol{I}_{m+1} \quad \boldsymbol{M}^{t-1}_{(m+1)\times(m+s)}\right),$

t-я компонента $\quad \left(\underbrace{\boldsymbol{0}_{(m+1)\times m} \quad \boldsymbol{0}_{(m+1)\times m} \quad \dots \quad \boldsymbol{0}_{(m+1)\times m}}_{t-1} \quad \boldsymbol{I}_{m+1} \quad \boldsymbol{M}^{t}_{(m+1)\times s}\right).$

Подсчитаем мощность частичного МНП-неспреда:

$$M_{\text{част. МНП}} = \frac{q^{kn}-q^{k(m+s)}}{q^{km}-1} + 1. \tag{5}$$

Как и в предыдущих случаях, определим долю мощности частичного МНП-неспреда от максимальной. В табл. 3 для примера приведем некоторые расчетные значения.

§ 5. Известные работы по неспредам

Приведем примеры неспредов с наилучшими значениями мощности по отношению к ранее известным подпростанственным кодам с теми же параметрами. В работе [9] с помощью систем Штейнера построен подпространственный код с параметрами $(n = 13, d_{\text{sub}} = 4, m = 3)$ и мощностью $M = 1597245$. Полученное значение мощности совпадает с верхней границей Джонсона (3).

Таблица 3

Доля мощности частичного МНП-неспреда от максимальной при $q = 2$

m	t									s
	2	3	4	5	6	7	8	9	10	
2	0,8024691	0,7291248	0,7074621	0,7018763	0,7004697	0,7001175	0,7000294	0,7000073	0,7000018	1
3	0,8892734	0,8409784	0,8342984	0,8334541	0,8333484	0,8333352	0,8333336	0,8333334	0,8333333	1
	0,9117595	0,8436442	0,8346293	0,8334954	0,8333536	0,8333359	0,8333336	0,8333334	0,8333333	2
4	0,9412305	0,9137060	0,9118864	0,9117723	0,9117652	0,9117647	0,9117647	0,9117647	0,9117647	1
	0,9545451	0,9145145	0,9119369	0,9117755	0,9117654	0,9117647	0,9117647	0,9117647	0,9117647	2
	0,9615337	0,9149200	0,9119621	0,9117770	0,9117655	0,9117648	0,9117647	0,9117647	0,9117647	3
5	0,9697041	0,9550330	0,9545607	0,9545459	0,9545455	0,9545455	0,9545455	0,9545455	0,9545455	1
	0,9769231	0,9552552	0,9545676	0,9545461	0,9545455	0,9545455	0,9545455	0,9545455	0,9545455	2
	0,9806196	0,9553664	0,9545711	0,9545463	0,9545455	0,9545455	0,9545455	0,9545455	0,9545455	3
	0,9824899	0,9554220	0,9545728	0,9545463	0,9545455	0,9545455	0,9545455	0,9545455	0,9545455	4
6	0,9846163	0,9770451	0,9769250	0,9769231	0,9769231	0,9769231	0,9769231	0,9769231	0,9769231	1
	0,9883721	0,9771033	0,9769259	0,9769231	0,9769231	0,9769231	0,9769231	0,9769231	0,9769231	2
	0,9902723	0,9771325	0,9769263	0,9769231	0,9769231	0,9769231	0,9769231	0,9769231	0,9769231	3
	0,9912280	0,9771470	0,9769266	0,9769231	0,9769231	0,9769231	0,9769231	0,9769231	0,9769231	4
	0,9917073	0,9771543	0,9769267	0,9769231	0,9769231	0,9769231	0,9769231	0,9769231	0,9769231	5
7	0,9922482	0,9884026	0,9883723	0,9883721	0,9883721	0,9883721	0,9883721	0,9883721	0,9883721	1
	0,9941634	0,9884175	0,9883724	0,9883721	0,9883721	0,9883721	0,9883721	0,9883721	0,9883721	2
	0,9951267	0,9884250	0,9883725	0,9883721	0,9883721	0,9883721	0,9883721	0,9883721	0,9883721	3
	0,9956097	0,9884287	0,9883725	0,9883721	0,9883721	0,9883721	0,9883721	0,9883721	0,9883721	4
	0,9958516	0,9884306	0,9883725	0,9883721	0,9883721	0,9883721	0,9883721	0,9883721	0,9883721	5
	0,9959727	0,9884315	0,9883726	0,9883721	0,9883721	0,9883721	0,9883721	0,9883721	0,9883721	6
20	0,9999990	0,9999986	0,9999986	0,9999986	0,9999986	0,9999986	0,9999986	0,9999986	0,9999986	1
	0,9999993	0,9999986	0,9999986	0,9999986	0,9999986	0,9999986	0,9999986	0,9999986	0,9999986	2
	0,9999994	0,9999986	0,9999986	0,9999986	0,9999986	0,9999986	0,9999986	0,9999986	0,9999986	3
	0,9999995	0,9999986	0,9999986	0,9999986	0,9999986	0,9999986	0,9999986	0,9999986	0,9999986	4
	...	...	...	...	...	...	...	...	...	...
	0,9999995	0,9999986	0,9999986	0,9999986	0,9999986	0,9999986	0,9999986	0,9999986	0,9999986	19

В работах [10, 11] были построены подпространственные коды для параметров $n = 6$, $m = 3$, $d_{\text{sub}} = 4$ мощности 77 и $n = 7$, $m = 3$, $d_{\text{sub}} = 4$ мощности 329 соответственно.

Для кода с параметрами $(n = 6, d_{\text{sub}} = 4, m = 3)$ и мощностью $M = 77$ в [10] был сначала использован компьютерный поиск. Затем полученные коды были проанализированы, и в новой конструкции использованы ранговые коды с параметрами $(n = 3, d_{\text{rank}} = 3, m = 2)$ и мощностью $M = 64$. Затем построен SKK-код с параметрами $(n = 6, d_{\text{sub}} = 4, m = 3)$. Дополнительная обработка привела к построению кода мощности $M = 77$. После этого значение мощности 77 считается наилучшим для кода с параметрами $(n = 6, d_{\text{sub}} = 4, m = 3)$, в то время как неравенство Джонсона определяет верхнюю границу мощности $M(6, 4, 3) = 81$.

Другой код с параметрами $(n = 7, d_{\text{sub}} = 4, m = 3)$ и мощностью $M = 329$ был построен примерно таким же способом в [11]. Предварительно был построен ранговый код, присоединена единичная матрица и получен SKK-код. Кроме того, использованы идеи конечной геометрии. Верхняя граница мощности для этого кода из неравенства Джонсона равна 381.

В работе [12] построен код с параметрами $(n = 8, d_{\text{sub}} = 6, m = 4)$ и мощностью 272. Граница Джонсона дает для этого кода значение 281.

Таким образом, кроме первой работы [9], остальные не дают значения, близкого к максимальной мощности по Джонсону. Однако, насколько нам известно, эта работа не имела продолжения для построения кодов с другими параметрами.

Известно также несколько работ [13, 14] одной направленности. Они используют SKK-коды в параллельных конструкциях и достигают существенного увеличения числа кодовых слов для определенных параметров. Сравним мощности для параметров, рассмотренных в работах [13, 14].

В [13] была определена нижняя граница мощности кода с параметрами $(16, 8, 8)$: $A(16, 8, 8) = 1099562828461$. Код с параметрами $(16, 8, 8)$ может быть построен как

неспред полного спреда. Действительно,

$$d = 2\delta = 8 \implies \delta = 4 = m,$$
$$k = m + 1 - \delta + 1 = 8 - 4 + 1 = 5,$$
$$n = 3 \times 4 \implies t = 3,$$

и по формуле (4) при $q = 2$ получаем $M(16, 8, 8) = \dfrac{2^{60} - 1}{2^{20} - 1}$. Тогда

$$\frac{M(16, 8, 8)}{A(16, 8, 8)} = 0{,}999954.$$

Аналогично можно получить

$$\frac{M(19, 4, 6)}{A(19, 4, 6)} = 0{,}8828625, \quad \frac{M(18, 8, 9)}{A(18, 8, 9)} = 0{,}999955, \quad \frac{M(18, 6, 9)}{A(18, 6, 9)} = 0{,}9948,$$
$$\frac{M(18, 4, 5)}{A(18, 4, 5)} = 0{,}89316, \quad \frac{M(14, 4, 7)}{A(14, 4, 7)} = 0{,}884, \quad \frac{M(12, 6, 6)}{A(12, 6, 6)} = 0{,}995.$$

Конструкции, предложенные в [13, 14], достаточно сложны и используют нелинейные условия, из-за которых не удается построить коды для любых параметров. Как видно из приведенных выше соотношений, с помощью простых конструкций, предложенных в настоящей статье, можно построить коды, содержащие более 88% возможных кодовых слов, а при иных параметрах и более 99%. Более того, эти конструкции позволяют строить коды для любых параметров, и полученные коды могут быть эффективно декодированы [15].

§ 6. Заключение

Рассмотрены семейства подпространственных кодов большой мощности с немаксимальным кодовым расстоянием. Алгоритм построения основан на неравенстве Джонсона: задаем код-спред, у которого подпространственное расстояние равно удвоенной размерности, и сохраняя расстояние, увеличиваем размерность и длину на единицу. В качестве исходного спреда рассмотрено три варианта: SKK-спред, МНП-спред и частичный МНП-спред. Показано, что при больших размерностях во всех трех кодах практически достигается максимальная граница мощности в соответствии с неравенством Джонсона.

СПИСОК ЛИТЕРАТУРЫ

1. *Ahlswede R., Cai N., Li S.-Y.R., Yeung R.W.* Network Information Flow // IEEE Trans. Inform. Theory. 2000. V. 46. № 4. P. 1204–1216. https://doi.org/10.1109/18.850663
2. *Ghemawat S., Gobioff H., Leung S.-T.* The Google File System // ACM SIGOPS Oper. Syst. Rev. 2003. V. 37. № 5. P. 29–43. https://doi.org/10.1145/1165389.945450
3. *Xia S.-T., Fu F.-W.* Johnson Type Bounds on Constant Dimension Codes // Des. Codes Cryptography. 2009. V. 50. № 2. P. 163–172. https://doi.org/10.1007/s10623-008-9221-7
4. *Габидулин Э.М., Пилипчук Н.И.* Многокомпонентные коды с максимальным кодовым расстоянием // Пробл. передачи информ. 2016. Т. 52. № 3. С. 85–92. http://mi.mathnet.ru/ppi2213
5. *Gabidulin E., Bossert M.* Codes for Network Coding // Proc. 2008 IEEE Int. Symp. on Information Theory (ISIT'2008). Toronto, Canada. July 6–11, 2008. P. 867–870. https://doi.org/10.1109/ISIT.2008.4595110
6. *Koetter R., Kschischang F.R.* Coding for Errors and Erasures in Random Network Coding // IEEE Trans. Inform. Theory. 2008. V. 54. № 8. P. 3579–3591. https://doi.org/10.1109/TIT.2008.926449

7. *Silva D., Koetter R., Kschischang F.R.* A Rank-Metric Approach to Error Control in Random Network Coding // IEEE Trans. Inform. Theory. 2008. V. 54. № 9. P. 3951–3967. https://doi.org/10.1109/TIT.2008.928291
8. *Габидулин Э.М.* Теория кодов с максимальным ранговым расстоянием// Пробл. передачи информ. 1985. Т. 21. № 1. С. 3–16. http://mi.mathnet.ru/ppi967
9. *Braun M., Etzion T., Östergård P.R.G., Vardy A., Wassermann A.* Existence of q-Analogs of Steiner Systems // Forum Math. Pi. 2016. V. 4. Research Paper e7 (14 pp.). https://doi.org/10.1017/fmp.2016.5
10. *Honold T., Kiermaier M., Kurz S.* Optimal Binary Subspace Codes of Length 6, Constant Dimension 3 and Minimum Distance 4, arXiv:1311.0464v2 [math.CO], 2014.
11. *Liu H., Honold T.* A New Approach to the Main Problem of Subspace Coding, arXiv:1408.1181 [math.CO], 2014.
12. *Heinlein D, Kurz S.* An Upper Bound for Binary Subspace Codes of Length 8, Constant Dimension 4 and Minimum Distance 6 // Proc. 10th Int. Workshop on Coding and Cryptography (WCC'2017). St. Petersburg, Russia. Sept. 18–22, 2017.
13. *Xu L., Chen H.* New Constant-Dimension Subspace Codes from Maximum Rank Distance Codes // IEEE Trans. Inform. Theory. 2018. V. 64. № 9. P. 6315–6319. https://doi.org/10.1109/TIT.2018.2839596
14. *He X., Chen Y.* Constructions of Const Dimension Codes from Serval Parallel Lift MRD Code, arXiv:1911.00154 [cs.IT], 2019.
15. *Габидулин Э.М., Пилипчук Н.И., Боссерт М.* Декодирование случайных сетевых кодов // Пробл. передачи информ. 2010. Т. 46. № 4. С. 33–55. http://mi.mathnet.ru/ppi2025

Габидулин Эрнст Мухамедович
(04.06.1937 – 22.06.2021)
Пилипчук Нина Ивановна
Трушина Оксана Вячеславовна
Московский физико-технический институт
(государственный университет)
pilipchuk.nina@gmail.com
oksana.trushina@gmail.com

Поступила в редакцию
03.02.2021
После доработки
11.06.2021
Принята к публикации
23.06.2021

ПРОБЛЕМЫ ПЕРЕДАЧИ ИНФОРМАЦИИ

Том 57 **2021** **Вып. 3**

УДК 621.391 : 519.23

© 2021 г. **М.В. Бурнашев**

О МИНИМАКСНОМ ОБНАРУЖЕНИИ ГАУССОВСКИХ СТОХАСТИЧЕСКИХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ И ГАУССОВСКИХ СТАЦИОНАРНЫХ СИГНАЛОВ[1]

Рассматривается задача обнаружения гауссовских стохастических последовательностей (сигналов) с неизвестными ковариационными матрицами на фоне белого гауссовского шума. При заданной вероятности "*ложной тревоги*" (вероятности ошибки 1-го рода) качество минимаксного обнаружения определяется наилучшей экспонентой "*вероятности пропуска*" (вероятности ошибки 2-го рода) при растущем интервале наблюдений. Целью является нахождение максимального множества ковариационных матриц (сложная гипотеза), такого что его минимаксную проверку можно заменить проверкой одной конкретной ковариационной матрицы (простая гипотеза) без ухудшения экспоненты обнаружения. В статье полностью описывается это максимальное множество ковариационных матриц. Рассматриваются также некоторые следствия, касающиеся минимаксного обнаружения гауссовских стохастических сигналов в белом гауссовском шуме и обнаружения гауссовских стационарных сигналов.

Ключевые слова: минимаксная проверка гипотез, вероятности ошибки, экспонента вероятности ошибки, лемма Стейна.

DOI: 10.31857/S0555292321030049

§ 1. Введение

1.1. Введение и определения. Рассматривается задача минимаксной проверки простой гипотезы $\mathcal{H}_0$ против сложной альтернативы $\mathcal{H}_1$ [1–3] по наблюдениям $\boldsymbol{y}_n^T = \boldsymbol{y}_n' = (y_1, \ldots, y_n) \in \mathbb{R}^n$:

$$\begin{aligned} &\mathcal{H}_0 \colon \boldsymbol{y}_n = \boldsymbol{\xi}_n, \quad \boldsymbol{\xi}_n \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n), \\ &\mathcal{H}_1 \colon \boldsymbol{y}_n = \boldsymbol{\eta}_n, \quad \boldsymbol{\eta}_n \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{M}_n), \quad \boldsymbol{M}_n \in \mathcal{M}_n, \end{aligned} \tag{1}$$

где выборка $\boldsymbol{\xi}_n^T = (\xi_1, \ldots, \xi_n)$ представляет "шум" и состоит из независимых одинаково распределенных гауссовских случайных величин с нулевыми средними и дисперсиями 1, а $\boldsymbol{I}_n$ – единичная ковариационная матрица. Стохастический "сигнал" $\boldsymbol{\eta}_n$ является гауссовским случайным вектором с нулевым средним и неизвестной ковариационной матрицей $\boldsymbol{M}_n$. $\mathcal{M}_n$ – заданное множество возможных ковариационных матриц $\boldsymbol{M}_n$.

Без ограничения общности можно предполагать матрицу $\boldsymbol{M}_n$ положительно определенной, т.е. $|\boldsymbol{M}_n| = \det \boldsymbol{M}_n > 0$. Действительно, если $|\boldsymbol{M}_n| = 0$, то тогда меры $\mathbf{P}_{\boldsymbol{\xi}_n}$ и $\mathbf{P}_{\boldsymbol{\eta}_n}$ ортогональны, и поэтому гипотезы $\mathcal{H}_0$ и $\mathcal{H}_1$ можно различить без ошибок.

[1] Исследование выполнено при частичной финансовой поддержке Российского фонда фундаментальных исследований (номер проекта 19-01-00364).

Для принятия решения при различении гипотез $\mathcal{H}_0$ и $\mathcal{H}_1$ выбирается область $\mathcal{D} \in \mathbb{R}^n$, такая что

$$\boldsymbol{y}_n \in \mathcal{D} \Rightarrow \mathcal{H}_0, \quad \boldsymbol{y}_n \notin \mathcal{D} \Rightarrow \mathcal{H}_1. \tag{2}$$

Тогда вероятности ошибки 1-го рода ("*ложной тревоги*") $\alpha(\mathcal{D})$ и 2-го рода ("*вероятность пропуска*") $\beta(\mathcal{D}, \mathcal{M}_n)$ определяются, соответственно, формулами

$$\alpha(\mathcal{D}) = \mathbf{P}(\boldsymbol{y}_n \notin \mathcal{D} \,|\, \mathcal{H}_0)$$

и

$$\beta(\mathcal{D}, \mathcal{M}_n) = \mathbf{P}(\boldsymbol{y}_n \in \mathcal{D} \,|\, \mathcal{H}_1) = \sup_{\boldsymbol{M}_n \in \mathcal{M}_n} \mathbf{P}(\boldsymbol{y}_n \in \mathcal{D} \,|\, \boldsymbol{M}_n).$$

При заданной вероятности ошибки 1-го рода α, $0 < \alpha < 1$, мы исследуем минимально возможную вероятность ошибки 2-го рода

$$\beta(\alpha, \mathcal{M}_n) = \inf_{\mathcal{D}:\alpha(\mathcal{D}) \leqslant \alpha} \beta(\mathcal{D}, \mathcal{M}_n) \tag{3}$$

и соответствующую оптимальную область принятия решения $\mathcal{D}(\alpha)$ (см. (2)).

В статье рассматривается случай, когда величина α фиксирована (или медленно убывает при $n \to \infty$). Этот случай иногда называют задачей Неймана – Пирсона минимаксного различения гипотез. В этом случае ошибки 1-го и 2-го рода влекут очень разные потери для статистика и он, в основном, хочет минимизировать ошибку 2-го рода $\beta = \mathbf{P}\{\mathcal{H}_0 \,|\, \mathcal{H}_1\}$. Этот случай очень популярен в разных приложениях (см., например, работу [4] и библиографию в ней).

Для заданной матрицы $\boldsymbol{M}_n$ и фиксированного α обозначим через $\beta(\boldsymbol{M}_n) = \beta(\alpha, \boldsymbol{M}_n)$ минимально возможную вероятность ошибки 2-го рода. Аналогично для заданного множества $\mathcal{M}_n = \{\boldsymbol{M}_n\}$ и фиксированного α обозначим через $\beta(\mathcal{M}_n) = \beta(\alpha, \mathcal{M}_n)$ минимально возможную минимаксную вероятность ошибки 2-го рода (см. (3)). Ясно, что тогда

$$\sup_{\boldsymbol{M}_n \in \mathcal{M}_n} \beta(\boldsymbol{M}_n) \leqslant \beta(\mathcal{M}_n), \tag{4}$$

или, эквивалентным образом,

$$\sup_{\boldsymbol{M}_n \in \mathcal{M}_n} \inf_{\mathcal{D}} \beta(\mathcal{D}, \boldsymbol{M}_n) \leqslant \inf_{\mathcal{D}} \sup_{\boldsymbol{M}_n \in \mathcal{M}_n} \beta(\mathcal{D}, \mathcal{M}_n).$$

Во многих практических случаях величина $\beta(\boldsymbol{M}_n)$ убывает экспоненциально по $n \to \infty$. Поэтому естественно (во всяком случае, проще и продуктивнее) исследовать соответствующие экспоненты $n^{-1} \ln \beta(\boldsymbol{M}_n)$ и $n^{-1} \ln \beta(\mathcal{M}_n)$ при $n \to \infty$ (некоторые результаты относительно равенства в (4) имеются в [5]).

В статье для фиксированного α и заданной последовательности матриц $\boldsymbol{M}_n$ исследуется последовательность множеств $\mathcal{M}_n$, таких что $\boldsymbol{M}_n \in \mathcal{M}_n$, $n = 1, 2, \ldots$, и выполняется следующее равенство:

$$\lim_{n \to \infty} \frac{1}{n} \ln \beta(\boldsymbol{M}_n) = \lim_{n \to \infty} \frac{1}{n} \ln \beta(\mathcal{M}_n(\boldsymbol{M}_n)). \tag{5}$$

Другими словами, при заданной вероятности ошибки 1-го рода α множество $\mathcal{M}_n$ есть множество ковариационных матриц, которое можно заменить одной матрицей $\boldsymbol{M}_n$ без потери для экспоненты вероятности ошибки 2-го рода. Далее будет описано максимальное из таких множеств $\mathcal{M}_n(\boldsymbol{M}_n)$ для модели (1), а также некоторые оценки для него.

Мотивация рассмотрения задачи минимаксной проверки гипотез (обнаружения сигналов) подробно описана в [1–3]. Если для заданного множества матриц $\mathcal{M}_n$ выполняется соотношение (5), то тогда можно заменить (без асимптотических потерь) все множество $\mathcal{M}_n$ одной конкретной матрицей $\boldsymbol{M}_n$. Напомним, что оптимальный тест для конкретной матрицы $\boldsymbol{M}_n$ описывается леммой Неймана – Пирсона и сводится к простому LR-тесту (likelihood ratio test detector). В противном случае (без соотношения (5)) оптимальным минимаксным тестом является значительно более сложный байессовский тест относительно *наименее благоприятного* априорного распределения на множестве $\mathcal{M}_n$. Поэтому естественно исследовать, когда заданное множество матриц $\mathcal{M}_n$ можно заменить конкретной матрицей $\boldsymbol{M}_n$. Однако технически удобнее рассмотреть эквивалентную задачу: для заданной матрицы $\boldsymbol{M}_n$ найти максимальное множество матриц $\mathcal{M}_n(\boldsymbol{M}_n)$, которое может быть заменено матрицей $\boldsymbol{M}_n$. Эта задача в основном и рассматривается в настоящей статье.

О п р е д е л е н и е 1. Для фиксированного α и заданной последовательности матриц $\boldsymbol{M}_n$, $n = 1, 2, \ldots$, обозначим через $\mathcal{M}_n(\boldsymbol{M}_n)$ последовательность максимальных множеств, таких что $\boldsymbol{M}_n \in \mathcal{M}_n(\boldsymbol{M}_n)$, и выполняется следующее равенство:

$$\lim_{n\to\infty} \frac{1}{n} \ln \beta(\mathcal{M}_n(\boldsymbol{M}_n)) = \lim_{n\to\infty} \frac{1}{n} \ln \beta(\boldsymbol{M}_n).$$

Обозначим в модели (1) через $\mathbf{P}_{\boldsymbol{I}_n}$ распределение величины $\boldsymbol{y}_n = \boldsymbol{\xi}_n$, где $\boldsymbol{\xi}_n \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_n)$. Аналогично обозначим через $\mathbf{Q}_{\boldsymbol{M}_n}$ распределение величины $\boldsymbol{y}_n = \boldsymbol{\eta}_n$, где $\boldsymbol{\eta}_n \sim \mathcal{N}(\mathbf{0}, \boldsymbol{M}_n)$. Обозначим также через $p_{\boldsymbol{I}_n}(\boldsymbol{y}_n)$ и $p_{\boldsymbol{M}_n}(\boldsymbol{y}_n)$, $\boldsymbol{y}_n \in \mathbb{R}^n$, соответствующие плотности распределения вероятностей.

Заметим, что если $|\boldsymbol{V}_n| \neq 0$ и $|\boldsymbol{M}_n| \neq 0$, то тогда

$$\ln \frac{p_{\boldsymbol{V}_n}}{p_{\boldsymbol{M}_n}}(\boldsymbol{y}_n) = \frac{1}{2}\left[\ln \frac{|\boldsymbol{M}_n|}{|\boldsymbol{V}_n|} + \big(\boldsymbol{y}_n, (\boldsymbol{M}_n^{-1} - \boldsymbol{V}_n^{-1})\boldsymbol{y}_n\big)\right]. \tag{6}$$

Введем также аналогичные $\mathcal{M}_n(\boldsymbol{M}_n)$ максимальные множества $\mathcal{M}_n^{LR}(\boldsymbol{M}_n)$, возникающие, если используются LR-тесты (см. определение 2 ниже). Будет показано, что при естественных предположениях $\mathcal{M}_n(\boldsymbol{M}_n) = \mathcal{M}_n^{LR}(\boldsymbol{M}_n)$, т.е. LR-тесты являются асимптотически оптимальными.

Для этого введем логарифм отношения правдоподобия (см. (6))

$$f_{\boldsymbol{M}_n}(\boldsymbol{y}_n) = \ln \frac{p_{\boldsymbol{I}_n}}{p_{\boldsymbol{M}_n}}(\boldsymbol{y}_n) = \frac{1}{2}\Big[\ln |\boldsymbol{M}_n| + \big(\boldsymbol{y}_n, (\boldsymbol{M}_n^{-1} - \boldsymbol{I}_n)\boldsymbol{y}_n\big)\Big]. \tag{7}$$

Область принятия решения $\mathcal{D}_{LR}(\boldsymbol{M}_n, \alpha)$ в пользу $\boldsymbol{I}_n$ при проверке матриц $\boldsymbol{I}_n$ и $\boldsymbol{M}_n$ имеет вид

$$\mathcal{D}_{LR}(\boldsymbol{M}_n, \alpha) = \{\boldsymbol{y}_n \in \mathbb{R}^n : \, f_{\boldsymbol{M}_n}(\boldsymbol{y}_n) \geqslant \gamma\}, \tag{8}$$

где γ – такое, что

$$\begin{aligned} \alpha &= \mathbf{P}_{\boldsymbol{I}_n}\{\mathcal{D}_{LR}^c(\boldsymbol{M}_n, \alpha)\} = \mathbf{P}_{\boldsymbol{I}_n}\{f_{\boldsymbol{M}_n}(\boldsymbol{y}_n) \leqslant \gamma\} = \\ &= \mathbf{P}_{\boldsymbol{I}_n}\big\{\big[(\boldsymbol{\xi}_n, (\boldsymbol{M}_n^{-1} - \boldsymbol{I}_n)\boldsymbol{\xi}_n) + \ln |\boldsymbol{M}_n|\big] \leqslant 2\gamma\big\}. \end{aligned} \tag{9}$$

Предположим, что в модели (1) для проверки матриц $\boldsymbol{I}_n$ и $\boldsymbol{M}_n$ (т.е. простых гипотез) используется оптимальный детектор (т.е. LR-тест) с областью принятия решения $\mathcal{D}_{LR}(\boldsymbol{M}_n, \alpha)$ (см. (8), (9)) в пользу $\boldsymbol{I}_n$. Для каких матриц $\boldsymbol{V}_n$ вместо $\boldsymbol{M}_n$ область принятия решения $\mathcal{D}_{LR}(\boldsymbol{M}_n, \alpha)$ не ухудшит вероятность ошибки 2-го рода $\beta(\alpha, \boldsymbol{M}_n)$? Для ответа на этот вопрос введем следующее

О п р е д е л е н и е 2. Для фиксированного α и заданной последовательности матриц $\boldsymbol{M}_n$, $n = 1, 2, \ldots$, обозначим через $\mathcal{M}_n^{LR}(\boldsymbol{M}_n)$ последовательность максималь-

ных множеств матриц $\boldsymbol{V}_n$, таких что выполняется следующее равенство

$$\lim_{n\to\infty}\frac{1}{n}\ln\sup_{\boldsymbol{V}_n\in\mathcal{M}_n^{LR}(\boldsymbol{M}_n)}\beta(\boldsymbol{V}_n)=\lim_{n\to\infty}\frac{1}{n}\ln\beta(\boldsymbol{M}_n)$$

при условии, что используются области принятия решения $\mathcal{D}_{LR}(\alpha,\boldsymbol{M}_n)$.

1.2. Расстояние Кульбака – Лейблера. Для случайных элементов $\boldsymbol{x}$ и $\boldsymbol{y}$ на измеримом пространстве $(\Omega,\mathcal{B})$ с вероятностными распределениями, соответственно, $\mathbf{P}_{\boldsymbol{x}}$ и $\mathbf{Q}_{\boldsymbol{y}}$ введем функцию (*расстояние* или *дивергенция Кульбака – Лейблера* для мер $\mathbf{P}_{\boldsymbol{x}}$ и $\mathbf{Q}_{\boldsymbol{y}}$)

$$D(\mathbf{P}_{\boldsymbol{x}}\,\|\,\mathbf{Q}_{\boldsymbol{y}})=\mathbf{E}_{\mathbf{P}_{\boldsymbol{x}}}\ln\frac{d\mathbf{P}_{\boldsymbol{x}}}{d\mathbf{Q}_{\boldsymbol{y}}}(\boldsymbol{u}). \tag{10}$$

В частности, если $\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^n$ и $\boldsymbol{x}\sim\mathcal{N}(\mathbf{0},\boldsymbol{V}_n)$, $\boldsymbol{y}\sim\mathcal{N}(\mathbf{0},\boldsymbol{M}_n)$, то тогда [6, гл. 9.1] (через tr обозначен след матрицы)

$$D(\mathbf{P}_{\boldsymbol{x}}\,\|\,\mathbf{Q}_{\boldsymbol{y}})=\frac{1}{2}\ln\frac{|\boldsymbol{M}_n|}{|\boldsymbol{V}_n|}+\frac{1}{2}\operatorname{tr}\bigl(\boldsymbol{V}_n\boldsymbol{M}_n^{-1}\bigr)-\frac{n}{2}.$$

Если $\boldsymbol{V}_n=\boldsymbol{I}_n$, то

$$D(\boldsymbol{I}_n\,\|\,\boldsymbol{M}_n)=D(\mathbf{P}_{\boldsymbol{x}}\,\|\,\mathbf{Q}_{\boldsymbol{y}})=\frac{1}{2}\sum_{i=1}^{n}\Bigl(\ln\lambda_i+\frac{1}{\lambda_i}-1\Bigr), \tag{11}$$

где $\lambda_1,\ldots,\lambda_n$ – собственные значения матрицы $\boldsymbol{M}_n$ (собственные значения матрицы $\boldsymbol{M}_n^{-1}$ суть $\lambda_1^{-1},\ldots,\lambda_n^{-1}$).

Расстояние Кульбака – Лейблера играет важную роль в проверке гипотез. Предположим, например, что в модели (1) для проверки гипотез $\boldsymbol{I}_n$ и $\boldsymbol{M}_n$ (т.е. простых гипотез) используется оптимальный тест, которым является LR-тест с областью принятия решения $\mathcal{D}_{LR}(\alpha,\boldsymbol{M}_n)$ (см. (8), (9)) в пользу $\boldsymbol{I}_n$. Тогда при некоторых естественных предположениях справедлива следующая формула:

$$\lim_{\alpha\to0}\lim_{n\to\infty}\frac{1}{n}\ln\beta(\alpha)=-\lim_{n\to\infty}\frac{1}{n}D(\boldsymbol{I}_n\,\|\,\boldsymbol{M}_n). \tag{12}$$

Формула (12) называется леммой Стейна [7, 8]. В случае независимых одинаково распределенных случайных величин ее доказательство можно найти в [6, теорема 3.3; 9, теорема 12.8.1]. Естественно ожидать, что формула (12) справедлива не только при проверке простых гипотез, но и в более общих случаях проверки сложных гипотез. Некоторые частные аналоги формулы (12) уже появились для гауссовских стационарных [4] и пуассоновских [10] случайных процессов. В статье будут получены аналоги формулы (12) для моделей (1) и (68) (см. ниже).

§ 2. Предположения и основные результаты

2.1. Предположения. Пусть $\mathcal{C}_n$ – выпуклое множество всех ковариационных (т.е. симметричных и положительно определенных) $(n\times n)$-матриц в $\mathbb{R}^n$. Рассмотрим последовательность множеств $\mathcal{M}_i\in\mathcal{C}_i$ ковариационных матриц $\boldsymbol{M}_i\in\mathcal{M}_i$, $i=1,2,\ldots$, в "схеме серий", т.е. множество $\mathcal{M}_{i+1}$ не обязательно является "продолжением" множества $\mathcal{M}_i$. Обозначим через $\lambda_1(\boldsymbol{M}_n),\ldots,\lambda_n(\boldsymbol{M}_n)$ собственные значения (все они положительны) матрицы $\boldsymbol{M}_n$. Далее нам понадобятся следующие предположения:

I. Для последовательности множеств $\mathcal{M}_n$ существует положительный предел (см. (11))

$$\lim_{n\to\infty}\frac{1}{n}\inf_{\boldsymbol{M}_n\in\mathcal{M}_n}\sum_{i=1}^{n}\left(\ln\lambda_i(\boldsymbol{M}_n)+\frac{1}{\lambda_i(\boldsymbol{M}_n)}-1\right)>0 \tag{13}$$

(заметим, что $\ln z\geqslant 1-1/z$, $z>0$).

II. Для какого-нибудь $\delta>0$ выполнено

$$\lim_{n\to\infty}\frac{1}{n}\sup_{\boldsymbol{M}_n\in\mathcal{M}_n}\sum_{i=1}^{n}\lambda_i^{-1-\delta}(\boldsymbol{M}_n)<\infty. \tag{14}$$

2.2. Основные результаты. Для $(n\times n)$-матрицы $\boldsymbol{A}_n$ обозначим $|\boldsymbol{A}_n|=\det\boldsymbol{A}_n$. Пусть $\boldsymbol{A}_n>0$ означает, что матрица $\boldsymbol{A}_n$ положительно определена.

Для матриц $\boldsymbol{M}_n,\boldsymbol{V}_n\in\mathcal{C}_n$, таких что $\boldsymbol{I}_n+\boldsymbol{V}_n^{-1}-\boldsymbol{M}_n^{-1}>0$, определим функцию

$$f(\boldsymbol{M}_n,\boldsymbol{V}_n)=\frac{|\boldsymbol{I}_n+\boldsymbol{V}_n(\boldsymbol{I}_n-\boldsymbol{M}_n^{-1})|}{|\boldsymbol{M}_n|}=\frac{|\boldsymbol{M}_n+(\boldsymbol{V}_n-\boldsymbol{M}_n)(\boldsymbol{I}_n-\boldsymbol{M}_n^{-1})|}{|\boldsymbol{M}_n|}, \tag{15}$$

где использовалось полезное тождество

$$\boldsymbol{I}_n+\boldsymbol{V}_n(\boldsymbol{I}_n-\boldsymbol{M}_n^{-1})=\boldsymbol{M}_n+(\boldsymbol{V}_n-\boldsymbol{M}_n)(\boldsymbol{I}_n-\boldsymbol{M}_n^{-1}). \tag{16}$$

Основной результат статьи описывает множества $\mathcal{M}_n(\boldsymbol{M}_n)$ и $\mathcal{M}_n^{LR}(\boldsymbol{M}_n)$.

Теорема 1. *Если выполнены предположения* (13), (14), *то при* $n\to\infty$

$$\begin{aligned}
&\mathcal{M}_n(\boldsymbol{M}_n)=\mathcal{M}_n^{LR}(\boldsymbol{M}_n)=\\
&=\left\{\boldsymbol{V}_n\in\mathcal{C}_n:\ \lim_{n\to\infty}\sup_{\boldsymbol{V}_n\in\mathcal{M}_n(\boldsymbol{M}_n)}\frac{1}{n}\ln\mathbf{E}_{\boldsymbol{I}_n}\frac{p_{\boldsymbol{V}_n}}{p_{\boldsymbol{M}_n}}(\boldsymbol{x})\leqslant 0\right\}=\\
&=\left\{\boldsymbol{V}_n\in\mathcal{C}_n:\ \boldsymbol{I}_n+\boldsymbol{V}_n^{-1}-\boldsymbol{M}_n^{-1}>0,\ \lim_{n\to\infty}\inf_{\boldsymbol{V}_n\in\mathcal{M}_n(\boldsymbol{M}_n)}\frac{1}{n}\ln f(\boldsymbol{M}_n,\boldsymbol{V}_n)\geqslant 0\right\},
\end{aligned} \tag{17}$$

где функция $f(\boldsymbol{M}_n,\boldsymbol{V}_n)$ *определена в* (15).

Ясно, что множества $\mathcal{M}_n(\boldsymbol{M}_n)$и $\mathcal{M}_n^{LR}(\boldsymbol{M}_n)$ являются выпуклыми.

Замечание 1. Известно [11, гл. 8.5, теорема 4], [12, теорема 7.6.7], что функция $f(\boldsymbol{A}_n)=\ln|\boldsymbol{A}_n|$ строго вогнута на выпуклом множестве $\mathcal{C}_n$ симметричных положительно определенных матриц в $\mathbb{R}^n$. Отсюда также следует выпуклость множества $\mathcal{M}_n(\boldsymbol{M}_n)$, т.е. если $\boldsymbol{V}_n^{(1)}\in\mathcal{M}_n(\boldsymbol{M}_n)$ и $\boldsymbol{V}_n^{(2)}\in\mathcal{M}_n(\boldsymbol{M}_n)$, то $a\boldsymbol{V}_n^{(1)}+(1-a)\boldsymbol{V}_n^{(2)}\in\in\mathcal{M}_n(\boldsymbol{M}_n)$ для любого $0\leqslant a\leqslant 1$.

Упростим теорему 1, ограничиваясь в (17) только матрицами $\boldsymbol{V}_n$, коммутирующими с $\boldsymbol{M}_n$. Для ковариационной матрицы $\boldsymbol{M}_n\in\mathcal{C}_n$ введем выпуклое множество $\mathcal{C}_{\boldsymbol{M}_n}$ ковариационных матриц $\boldsymbol{V}_n$, коммутирующих с $\boldsymbol{M}_n$:

$$\mathcal{C}_{\boldsymbol{M}_n}=\{\boldsymbol{V}_n\in\mathcal{C}_n:\ \boldsymbol{M}_n\boldsymbol{V}_n=\boldsymbol{V}_n\boldsymbol{M}_n\}. \tag{18}$$

Обозначим через $\{\lambda_i\}$ собственные значения матрицы $\boldsymbol{M}_n$, а через $\{\nu_i\}$ – собственные значения матрицы $\boldsymbol{V}_n$. Тогда функция $f(\boldsymbol{M}_n,\boldsymbol{V}_n)$ из (15), (16) принимает вид

$$f(\boldsymbol{M}_n,\boldsymbol{V}_n)=\prod_{i=1}^{n}\frac{[\lambda_i+\nu_i(\lambda_i-1)]}{\lambda_i^2}=\prod_{i=1}^{n}\left[1+\frac{(\nu_i-\lambda_i)(\lambda_i-1)}{\lambda_i^2}\right] \tag{19}$$

при условии $\lambda_i+\nu_i(\lambda_i-1)>0$, $i=1,\ldots,n$.

Введем следующее подмножество множества $\mathcal{C}_{\boldsymbol{M}_n}$ при $n \to \infty$:

$$\mathcal{V}_n^{(0)}(\boldsymbol{M}_n) = \Big\{ \boldsymbol{V}_n : \inf_{\boldsymbol{V}_n \in \mathcal{V}_n^{(0)}(\boldsymbol{M}_n)} \ln f(\boldsymbol{M}_n, \boldsymbol{V}_n) \geqslant o(n) \Big\}, \tag{20}$$

где функция $f(\boldsymbol{M}_n, \boldsymbol{V}_n)$ определена в (19). Тогда справедлива следующая "внутренняя граница" $\mathcal{V}_n^{(0)}(\boldsymbol{M}_n)$ для $\mathcal{M}_n(\boldsymbol{M}_n)$.

Теорема 2. *Если выполнены предположения* (13), (14), *то множество* $\mathcal{M}_n(\boldsymbol{M}_n)$ *содержит множество* $\mathcal{V}_n^{(0)}(\boldsymbol{M}_n)$:

$$\mathcal{V}_n^{(0)}(\boldsymbol{M}_n) \subseteq \mathcal{M}_n(\boldsymbol{M}_n). \tag{21}$$

Множество $\mathcal{V}_n^{(0)}(\boldsymbol{M}_n)$ является выпуклым по $\boldsymbol{V}_n$ (см. замечание 1).

Замечание 2. Наряду с предельной формой неравенств в (17) далее некоторые подобные условия записываются также в допредельном виде с использованием остаточных членов типа $o(1)$, $o(n)$ и т.п. Таким образом автор указывает, что эти остаточные члены можно оценить, что дает оценки скорости сходимости в соответствующих формулах.

Замечание 3. Заменяя $o(n)$ величиной 0 в правой части формулы (20), рассмотрим множество

$$\mathcal{V}_n^{(1)}(\boldsymbol{M}_n) = \Big\{ \boldsymbol{V}_n : \inf_{\boldsymbol{V}_n \in \mathcal{V}_n^{(1)}(\boldsymbol{M}_n)} \ln f(\boldsymbol{M}_n, \boldsymbol{V}_n) \geqslant 0 \Big\}. \tag{22}$$

Ясно, что $\mathcal{V}_n^{(1)}(\boldsymbol{M}_n) \in \mathcal{V}_n^{(0)}(\boldsymbol{M}_n)$. В определенном смысле $\mathcal{V}_n^{(0)}(\boldsymbol{M}_n)$ – это множество $\mathcal{V}_n^{(1)}(\boldsymbol{M}_n)$, расширенное на "тонкий слой" с шириной порядка $o(n)$. Другими словами, $\mathcal{V}_n^{(1)}(\boldsymbol{M}_n)$ из (22) можно рассматривать как "ядро" множества $\mathcal{V}_n^{(0)}(\boldsymbol{M}_n)$.

Эта статья была написана под влиянием работы [4], где рассматривалась аналогичная задача для стохастических стационарных сигналов. В данной статье рассматривается более общий случай гауссовских случайных векторов с неизвестной ковариационной матрицей. При этом оказывается возможным провести, по существу, неасимптотическое исследование задачи. Стохастические стационарные сигналы являются тогда определенным предельным частным случаем задачи. Ранее подобный подход использовался в случае пуассоновских процессов [10]. Некоторые другие случаи с дополнительными ограничениями на вероятности ошибки α, β рассматривались в [13, 14].

2.3. Обратная задача. Следуя работе [4], рассмотрим также обратную задачу: когда проверку заданного множества $\mathcal{M}_n$ можно заменить проверкой некоторой матрицы $\boldsymbol{M}_n^{(0)}$?

Достаточное условие для этого следует из теоремы 1. Пусть задана последовательность множеств $\{\mathcal{M}_1, \mathcal{M}_2, \ldots\}$, $\mathcal{M}_i \subset \mathcal{C}_i$, $i = 1, 2, \ldots$, ковариационных матриц $\boldsymbol{M}_i \in \mathcal{M}_i$. Обозначим через $\{\boldsymbol{M}_i^{(0)}\}$ последовательность ковариационных матриц $\boldsymbol{M}_i^{(0)}$ (если таковая существует), удовлетворяющая следующему аналогу формулы (5):

$$\lim_{n\to\infty} \frac{1}{n} \ln \beta(\boldsymbol{M}_n^{(0)}) = \lim_{n\to\infty} \frac{1}{n} \ln \beta(\mathcal{M}_n) = \lim_{n\to\infty} \frac{1}{n} \ln \beta\big(\{\mathcal{M}_n, \boldsymbol{M}_n^{(0)}\}\big), \tag{23}$$

где $\{\mathcal{M}_n, \boldsymbol{M}_n^{(0)}\} = \mathcal{M}_n \cup \boldsymbol{M}_n^{(0)}$ – "расширение" множества $\mathcal{M}_n$ на матрицу $\boldsymbol{M}_n^{(0)}$. Мы не требуем, чтобы $\boldsymbol{M}_i^{(0)} \in \mathcal{M}_i$. Когда существует последовательность $\{\boldsymbol{M}_i^{(0)}\}$, удовлетворяющая (23)?

Из теоремы 1 получаем следующее достаточное условие для выполнения равенства (23).

Предложение 1. *Пусть* $\{\mathcal{M}_i\}$ *– какая-либо последовательность множеств, удовлетворяющая предположениям* (13), (14). *Если для последовательности матриц* $\{\boldsymbol{M}_i^{(0)}\}$ *выполняются следующие условия*:

1) $\boldsymbol{I}_n + \boldsymbol{V}_n^{-1} - (\boldsymbol{M}_n^{(0)})^{-1} > 0$ *для всех* $\boldsymbol{V}_n \in \mathcal{M}_n$;

2) *имеет место неравенство*

$$\inf_{\boldsymbol{V}_n \in \mathcal{M}_n} \ln f(\boldsymbol{M}_n^{(0)}, \boldsymbol{V}_n) \geqslant o(n), \quad n \to \infty, \tag{24}$$

то последовательность $\{\boldsymbol{M}_i^{(0)}\}$ *с LR-тестами удовлетворяет условию* (23).

Замечание 4. Если последовательность $\{\boldsymbol{M}_i^{(0)}\}$ удовлетворяет условию (24) для последовательности множеств $\{\mathcal{M}_i\}$, то она также удовлетворяет этому условию для последовательности множеств $\{\operatorname{conv}\mathcal{M}_i\}$, где $\operatorname{conv}\mathcal{M}_i$ – минимальное выпуклое множество матриц, содержащее множество $\mathcal{M}_i$. Ясно, что $\mathcal{M}_i \subseteq \operatorname{conv}\mathcal{M}_i$. Множество $\operatorname{conv}\mathcal{M}_i$ иногда называют "выпуклой оболочкой" множества $\mathcal{M}_i$.

Предложение 1 обобщает аналогичный результат в [4, теорема 1] (см. предложение 4 в п. 5.2).

Далее в § 3 приводится важная для нас вспомогательная теорема 3. В § 4 содержатся доказательства теорем 1 и 2, а также некоторые связанные с ними результаты. В § 5 рассматривается аналогичная (1) модель "сигнал + шум", а также случай стационарных стохастических сигналов. Все результаты § 5 являются следствиями теоремы 2. Некоторые примеры приводятся в § 6.

§ 3. Вспомогательная теорема 3

3.1. Простые гипотезы и теорема 3. Для исследования модели (1) рассмотрим сначала более простую модель

$$\begin{aligned} &\mathcal{H}_0 \colon \boldsymbol{y}_n = \boldsymbol{\xi}_n, \quad \boldsymbol{\xi}_n \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n), \\ &\mathcal{H}_1 \colon \boldsymbol{y}_n = \boldsymbol{\eta}_n, \quad \boldsymbol{\eta}_n \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{M}_n), \end{aligned} \tag{25}$$

соответствующую проверке простых гипотез: $\boldsymbol{I}_n$ против $\boldsymbol{M}_n$. Анализ модели (1) далее будет развитием результатов для модели (25).

Заметим, что модель (25) можно свести к эквивалентной модели с диагональной матрицей $\boldsymbol{M}_n$. Действительно, так как $\boldsymbol{M}_n$ – ковариационная матрица (т.е. симметричная и положительно определенная), то существует ортогональная матрица $\boldsymbol{T}_n$ и диагональная матрица $\boldsymbol{\Lambda}_n$, такие что $\boldsymbol{M}_n = \boldsymbol{T}_n \boldsymbol{\Lambda}_n \boldsymbol{T}_n'$ [11, гл. 4.7-4.9; 12, теорема 4.1.5]. При этом диагональная матрица $\boldsymbol{\Lambda}_n = \boldsymbol{T}_n' \boldsymbol{M}_n \boldsymbol{T}_n$ состоит из собственных значений $\{\lambda_i\}$ матрицы $\boldsymbol{M}_n$.

Отметим также, что для любой ортогональной матрицы $\boldsymbol{T}_n$ вектор $\boldsymbol{T}_n' \boldsymbol{\xi}_n$ имеет то же самое распределение, что и сам вектор $\boldsymbol{\xi}_n$ (для простой гипотезы $\mathcal{H}_0$ в (25)). Поэтому, умножая обе части (25) на $\boldsymbol{T}_n'$, мы можем свести модель (25) к эквивалентной модели с диагональной матрицей $\boldsymbol{M}_n$.

Обозначим $D(\boldsymbol{I}_n \,\|\, \boldsymbol{M}_n) = D(\mathbf{P}_{\boldsymbol{I}_n} \,\|\, \mathbf{Q}_{\boldsymbol{M}_n})$ (см. (10)). Пусть $\beta(\alpha, \boldsymbol{M}_n)$ – минимально возможная вероятность ошибки 2-го рода при заданной вероятности ошибки 1-го рода α. Следующая теорема является основным вспомогательным результатом статьи.

Теорема 3. *Величина* $\beta(\alpha, \boldsymbol{M}_n)$, $0 < \alpha < 1$, *удовлетворяет следующим неравенствам*:

$$\ln \beta(\alpha, \boldsymbol{M}_n) \geqslant -\frac{D(\boldsymbol{I}_n \,\|\, \boldsymbol{M}_n) + h(\alpha)}{1-\alpha}, \quad h(\alpha) = -\alpha \ln \alpha - (1-\alpha)\ln(1-\alpha), \tag{26}$$

и

$$\ln \beta(\alpha, \boldsymbol{M}_n) \leqslant -D(\boldsymbol{I}_n \,\|\, \boldsymbol{M}_n) + \mu_0(\alpha, \boldsymbol{M}_n), \tag{27}$$

где $\mu_0(\alpha, \boldsymbol{M}_n)$ *определяется соотношением*

$$\mathbf{P}_{\boldsymbol{I}_n}\Big\{\ln \frac{p_{\boldsymbol{I}_n}}{p_{\boldsymbol{M}_n}}(\boldsymbol{x}) \leqslant D(\boldsymbol{I}_n \,\|\, \boldsymbol{M}_n) - \mu_0\Big\} = \alpha. \tag{28}$$

Отметим, что обе границы (26) и (27) являются чисто аналитическими соотношениями, без каких-либо предельных операций. Нижняя граница (26) и верхняя граница (27) близки друг к другу, если величина $\mu_0(\alpha, \boldsymbol{M}_n)$ значительно меньше, чем $D(\boldsymbol{I}_n \,\|\, \boldsymbol{M}_n)$ (которая обычно имеет порядок n).

Следующий результат дает оценку сверху для величины $\mu_0(\alpha, \boldsymbol{M}_n)$ порядка $n^{1/p}$, $p > 1$ (см. доказательство в § 4).

Лемма 1. *Пусть для какого-либо* $1 < p \leqslant 2$ *выполнено условие*

$$\sup_n \frac{1}{n} \sum_{i=1}^{n} \lambda_i^{-p}(\boldsymbol{M}_n) \leqslant C_p < \infty. \tag{29}$$

Тогда для $\mu_0(\alpha, \boldsymbol{M}_n)$ *из* (27) *справедлива оценка сверху*

$$\mu_0(\alpha, \boldsymbol{M}_n) \leqslant \left[\frac{2(C_p+1)n}{\alpha}\right]^{1/p}. \tag{30}$$

В частности, если условие (29) выполняется при $p = 2$, то тогда (30) дает для $\mu_0(\alpha, \boldsymbol{M}_n)$ оценку сверху порядка $\sqrt{n/\alpha}$. Далее лемма 1 будет использоваться совместно с предположением (14).

3.2. Доказательство теоремы 3. Докажем сначала оценку снизу (26). Пусть $\mathcal{D} \in \mathbb{R}^n$ – область принятия решения в пользу $\boldsymbol{I}_n$, и пусть $\beta = \beta(\mathcal{D})$, $\alpha = \alpha(\mathcal{D})$ – соответствующие вероятности ошибок. Обозначая $p = p_{\boldsymbol{I}_n}$ и $q = p_{\boldsymbol{M}_n}$, имеем

$$\beta = \mathbf{Q}_{\boldsymbol{M}_n}(\mathcal{D}) = \int\limits_{\mathcal{D}} p(\boldsymbol{x}) \frac{q}{p}(\boldsymbol{x})\, d\boldsymbol{x}, \quad \alpha = \mathbf{P}_{\boldsymbol{I}_n}(\mathcal{D}^c), \tag{31}$$

где $\mathcal{D}^c = \mathbb{R}^n \setminus \mathcal{D}$. Так как $\mathbf{P}_{\boldsymbol{I}_n}(\mathcal{D}) = 1 - \alpha$, то рассматривая $\mathbf{P}_{\boldsymbol{I}_n}/(1-\alpha)$ как вероятностное распределение на $\mathcal{D}$ и используя неравенство $\ln \mathbf{E}\,\xi \geqslant \mathbf{E} \ln \xi$, имеем

$$\ln \frac{\beta}{1-\alpha} = \ln\left[\frac{1}{(1-\alpha)} \int\limits_{\mathcal{D}} p(\boldsymbol{x}) \frac{q}{p}(\boldsymbol{x})\, d\boldsymbol{x}\right] \geqslant \frac{1}{(1-\alpha)} \int\limits_{\mathcal{D}} p(\boldsymbol{x}) \ln \frac{q}{p}(\boldsymbol{x})\, d\boldsymbol{x} =$$
$$= -\frac{D(\boldsymbol{I}_n \,\|\, \boldsymbol{M}_n)}{1-\alpha} - \frac{1}{(1-\alpha)} \int\limits_{\mathcal{D}^c} p(\boldsymbol{x}) \ln \frac{q}{p}(\boldsymbol{x})\, d\boldsymbol{x}. \tag{32}$$

Так как $\mathbf{P}_{\boldsymbol{I}_n}(\mathcal{D}^c) = \alpha$, то аналогично неравенству в (32) последний член в правой части (32) дает

$$\int\limits_{\mathcal{D}^c} p(\boldsymbol{x}) \ln \frac{q}{p}(\boldsymbol{x})\, d\boldsymbol{x} \leqslant \alpha \ln \left[\frac{1}{\alpha} \int\limits_{\mathcal{D}^c} q(\boldsymbol{x})\, d\boldsymbol{x} \right] = \alpha \ln \frac{1-\beta}{\alpha} \leqslant \alpha \ln \frac{1}{\alpha}. \tag{33}$$

Поэтому из (32) и (33) получаем

$$\ln \frac{\beta}{1-\alpha} \geqslant -\frac{D(\boldsymbol{I}_n \,\|\, \boldsymbol{M}_n)}{1-\alpha} - \frac{\alpha}{(1-\alpha)} \ln \frac{1}{\alpha},$$

откуда следует оценка (26).

Для того чтобы доказать оценку сверху (27), выберем величину $\mu > 0$ и определим область принятия решения в пользу $\boldsymbol{I}_n$ как

$$\mathcal{A}_\mu = \Big\{ \boldsymbol{x} \in \mathcal{X} : \ln \frac{p}{q}(\boldsymbol{x}) \geqslant D(\boldsymbol{I}_n \,\|\, \boldsymbol{M}_n) - \mu \Big\}. \tag{34}$$

Обозначим через α_μ и β_μ вероятности ошибки, соответственно, 1-го и 2-го рода для области принятия решения $\mathcal{A}_\mu$. Тогда в силу (34)

$$\beta_\mu = \int\limits_{\mathcal{A}_\mu} p(\boldsymbol{x}) \frac{q}{p}(\boldsymbol{x})\, d\boldsymbol{x} = e^{-D(\boldsymbol{I}_n \,\|\, \boldsymbol{M}_n) + \mu_1}, \tag{35}$$

где $0 \leqslant \mu_1 \leqslant \mu$. Также имеем

$$\alpha_\mu = \mathbf{P}_{\boldsymbol{I}_n} \Big\{ \ln \frac{p}{q}(\boldsymbol{x}) \leqslant D(\boldsymbol{I}_n \,\|\, \boldsymbol{M}_n) - \mu \Big\} = \mathbf{P}(\eta \geqslant \mu), \tag{36}$$

где

$$\eta = D(\boldsymbol{I}_n \,\|\, \boldsymbol{M}_n) - \ln \frac{p}{q}(\boldsymbol{x}), \quad \mathbf{E}_{\boldsymbol{I}_n} \eta = 0.$$

Наилучшим является выбрать μ так, что $\alpha_\mu = \alpha$. Поэтому определим $\mu = \mu_0(\alpha, \boldsymbol{M}_n)$ формулой (28). Тогда

$$\alpha_{\mu_0(\alpha, \boldsymbol{M}_n)} = \alpha, \tag{37}$$

и в силу (35)

$$\ln \beta_{\mu_0(\alpha, \boldsymbol{M}_n)} \leqslant -D(\boldsymbol{I}_n \,\|\, \boldsymbol{M}_n) + \mu_0(\alpha, \boldsymbol{M}_n),$$

что доказывает оценку сверху (27). ▲

§ 4. Доказательства теорем 1 и 2

Так как $\mathcal{M}_n^{LR}(\boldsymbol{M}_n) \subseteq \mathcal{M}_n(\boldsymbol{M}_n)$, то для доказательства теоремы 1 достаточно получить "внутреннюю границу" для $\mathcal{M}_n^{LR}(\boldsymbol{M}_n)$ и аналогичную "внешнюю границу" для $\mathcal{M}_n(\boldsymbol{M}_n)$.

4.1. "Внутренняя граница" для $\boldsymbol{\mathcal{M}_n^{LR}(M_n)}$. Рассмотрим проверку простой гипотезы $\boldsymbol{I}_n$ против сложной альтернативы $\mathcal{M}_n$. Используем для этого оптимальный LR-тест для матрицы $\boldsymbol{M}_n \in \mathcal{M}_n$ с областью принятия решения $\mathcal{D}_{LR}(\boldsymbol{M}_n, \alpha) = \mathcal{A}_{\mu_0}$ в пользу $\boldsymbol{I}_n$ (см. (8), (9) и (34)), где $\mu_0 = \mu_0(\alpha, \boldsymbol{M}_n) > 0$ определено в (28). Рассмотрим какую-либо другую матрицу $\boldsymbol{V}_n \in \mathcal{M}_n$ и оценим вероятность ошибки 2-го

рода $\beta(\alpha, \boldsymbol{V}_n)$ при условии, что используется область принятия решения $\mathcal{A}_{\mu_0}$. Тогда с помощью (34), (35) имеем

$$\beta(\alpha, \boldsymbol{V}_n) = \mathbf{Q}_{\boldsymbol{V}_n}(\mathcal{A}_{\mu_0}) = \int\limits_{\mathcal{A}_{\mu_0}} p_{\boldsymbol{V}_n}(\boldsymbol{x})\, d\boldsymbol{x} = \int\limits_{\mathcal{A}_{\mu_0}} \frac{p_{\boldsymbol{V}_n}}{p_{\boldsymbol{M}_n}}(\boldsymbol{x}) \frac{p_{\boldsymbol{M}_n}}{p_{\boldsymbol{I}_n}}(\boldsymbol{x}) p_{\boldsymbol{I}_n}(\boldsymbol{x}) d\,\boldsymbol{x} =$$

$$= e^{-D(\boldsymbol{I}_n \| \boldsymbol{M}_n)+\mu_2} \int\limits_{\mathcal{A}_{\mu_0}} \frac{p_{\boldsymbol{V}_n}}{p_{\boldsymbol{M}_n}}(\boldsymbol{x}) p_{\boldsymbol{I}_n}(\boldsymbol{x})\, d\boldsymbol{x} \leqslant \beta(\alpha, \boldsymbol{M}_n) e^{\mu_2-\mu_1}\, \mathbf{E}_{\boldsymbol{I}_n} \frac{p_{\boldsymbol{V}_n}}{p_{\boldsymbol{M}_n}}(\boldsymbol{x}), \tag{38}$$

где $0 \leqslant \max\{\mu_1, \mu_2\} \leqslant \mu_0(\alpha, \boldsymbol{M}_n)$. В силу предположения (14) и (30) имеем также

$$\mu_0(\alpha, \boldsymbol{M}_n) = O(n^{1/(1+\delta)}) = o(n), \quad n \to \infty. \tag{39}$$

Поэтому если

$$\sup_{\boldsymbol{V}_n \in \mathcal{M}_n^{LR}(\boldsymbol{M}_n)} \mathbf{E}_{\boldsymbol{I}_n} \frac{p_{\boldsymbol{V}_n}}{p_{\boldsymbol{M}_n}}(\boldsymbol{x}) \leqslant e^{o(n)}, \quad n \to \infty, \tag{40}$$

то из (38)–(40) при $n \to \infty$

$$\sup_{\boldsymbol{V}_n \in \mathcal{M}_n^{LR}(\boldsymbol{M}_n)} \ln \beta(\alpha, \boldsymbol{V}_n) \leqslant \ln \beta(\alpha, \boldsymbol{M}_n) + o(n). \tag{41}$$

4.2. "Внешняя граница" для $\boldsymbol{\mathcal{M}_n(M_n)}$. Пусть $\mathcal{D}_{LR}(\boldsymbol{M}_n, \alpha) = \mathcal{A}_{\mu_0}$ в пользу $\boldsymbol{I}_n$, и пусть $\beta_{\boldsymbol{M}_n} = \beta_{\boldsymbol{M}_n}(\mathcal{D})$, $\alpha = \alpha(\mathcal{D})$ – соответствующие вероятности ошибок. Тогда, обозначая $p = p_{\boldsymbol{I}_n}$ и $q = p_{\boldsymbol{M}_n}$, аналогично (31) имеем

$$\beta_{\boldsymbol{M}_n} = \mathbf{Q}_{\boldsymbol{M}_n}(\mathcal{D}) = \int\limits_{\mathcal{D}} q(\boldsymbol{x})\, d\boldsymbol{x}, \quad \alpha = \mathbf{P}_{\boldsymbol{I}_n}(\mathcal{D}^c). \tag{42}$$

Рассмотрим какую-либо другую матрицу $\boldsymbol{V}_n \in \mathcal{M}_n(\boldsymbol{M}_n)$. Обозначая $q_1 = p_{\boldsymbol{V}_n}$, для вероятности ошибки 2-го рода $\beta_{\boldsymbol{V}_n} = \beta_{\boldsymbol{V}_n}(\mathcal{D})$ необходимо иметь

$$\beta_{\boldsymbol{V}_n} = \mathbf{Q}_{\boldsymbol{V}_n}(\mathcal{D}) = \int\limits_{\mathcal{D}} q_1(\boldsymbol{x})\, d\boldsymbol{x} \leqslant \beta_{\boldsymbol{M}_n} e^{o(n)}. \tag{43}$$

Для какого-нибудь δ, $0 \leqslant \delta \leqslant 1$, рассмотрим также функцию плотности распределения вероятностей

$$q_\delta(\boldsymbol{x}) = (1-\delta) q(\boldsymbol{x}) + \delta q_1(\boldsymbol{x}) \tag{44}$$

и связанную с ней величину

$$\beta_\delta = \int\limits_{\mathcal{D}} q_\delta(\boldsymbol{x})\, d\boldsymbol{x} = (1-\delta)\beta_{\boldsymbol{M}_n} + \delta \beta_{\boldsymbol{V}_n}.$$

В силу (42), (43) имеем

$$\beta_\delta \leqslant \beta_{\boldsymbol{M}_n}(1 - \delta + \delta e^{o(n)}). \tag{45}$$

Отметим, что плотность распределения вероятностей $q_\delta(\boldsymbol{x})$ соответствует байессовской постановке задачи, в которой альтернативная гипотеза $\mathcal{H}_1$ с вероятностью $1-\delta$ совпадает с $\boldsymbol{M}_n$, а с вероятностью δ – с $\boldsymbol{V}_n$. Соответственно, величина β_δ есть вероятность ошибки 2-го рода.

Аналогично (32), (33) оценим снизу величину β_δ. Сперва аналогично (32) имеем

$$\ln\frac{\beta_\delta}{1-\alpha} = \ln\left[\frac{1}{(1-\alpha)}\int\limits_{\mathcal{D}} p(\boldsymbol{x})\frac{q_\delta}{p}(\boldsymbol{x})\,d\boldsymbol{x}\right] \geqslant \frac{1}{(1-\alpha)}\int\limits_{\mathcal{D}} p(\boldsymbol{x})\ln\frac{q_\delta}{p}(\boldsymbol{x})\,d\boldsymbol{x} =$$

$$= -\frac{D(p(\boldsymbol{x})\,\|\,q_\delta(\boldsymbol{x}))}{1-\alpha} - \frac{1}{(1-\alpha)}\int\limits_{\mathcal{D}^c} p(\boldsymbol{x})\ln\frac{q_\delta}{p}(\boldsymbol{x})\,d\boldsymbol{x}. \tag{46}$$

Для последнего члена в правой части (46) аналогично (33) имеем

$$\int\limits_{\mathcal{D}^c} p(\boldsymbol{x})\ln\frac{q_\delta}{p}(\boldsymbol{x})\,d\boldsymbol{x} \leqslant \alpha\ln\left[\frac{1}{\alpha}\int\limits_{\mathcal{D}^c} q_\delta(\boldsymbol{x})\,d\boldsymbol{x}\right] = \alpha\ln\frac{1-\beta_\delta}{\alpha} \leqslant \alpha\ln\frac{1}{\alpha}.$$

Поэтому аналогично (26) получаем

$$\ln\beta_\delta \geqslant -\frac{D(p(\boldsymbol{x})\,\|\,q_\delta(\boldsymbol{x})) + h(\alpha)}{1-\alpha}. \tag{47}$$

Рассмотрим величину $D(p(\boldsymbol{x})\,\|\,q_\delta(\boldsymbol{x}))$ из правой части (47). Обозначая

$$r(\boldsymbol{x}) = \frac{q_1(\boldsymbol{x})}{q(\boldsymbol{x})},$$

в силу (44) имеем

$$\frac{q_\delta(\boldsymbol{x})}{q(\boldsymbol{x})} = 1-\delta+\delta\frac{q_1(\boldsymbol{x})}{q(\boldsymbol{x})} = 1-\delta+\delta r(\boldsymbol{x}).$$

Поэтому

$$D(p(\boldsymbol{x})\,\|\,q_\delta(\boldsymbol{x})) = -\int\limits_{\mathbb{R}^n} p(\boldsymbol{x})\ln\frac{q_\delta}{p}(\boldsymbol{x})\,d\boldsymbol{x} = D(p(\boldsymbol{x})\,\|\,q(\boldsymbol{x})) + g(\delta), \tag{48}$$

где

$$g(\delta) = -\int\limits_{\mathbb{R}^n} p(\boldsymbol{x})\ln[1-\delta+\delta r(\boldsymbol{x})]\,d\boldsymbol{x}. \tag{49}$$

В силу (45) и (48), (49) для любой функции $o(n)$ необходимо иметь

$$g(\delta) \geqslant -\ln(1-\delta+\delta e^{o(n)}) \quad \text{для всех } 0<\delta\leqslant 1. \tag{50}$$

Заметим, что в силу неравенства $\ln\mathbf{E}\,\xi \geqslant \mathbf{E}\ln\xi$ из (49) имеем

$$g(\delta) \leqslant \ln\int\limits_{\mathbb{R}^n}\frac{p(\boldsymbol{x})}{1-\delta+\delta r(\boldsymbol{x})}\,d\boldsymbol{x} \quad \text{для всех } 0<\delta\leqslant 1.$$

Поэтому для того чтобы выполнялось (50), необходимо иметь

$$\int\limits_{\mathbb{R}^n}\frac{p(\boldsymbol{x})}{1-\delta+\delta r(\boldsymbol{x})}\,d\boldsymbol{x} \geqslant \frac{1}{1-\delta+\delta e^{o(n)}}, \quad 0<\delta\leqslant 1. \tag{51}$$

Так как $\int p(\boldsymbol{x})\,d\boldsymbol{x} = 1$, то соотношение (51) эквивалентно условию

$$\int\limits_{\mathbb{R}^n} \frac{p(\boldsymbol{x})(r(\boldsymbol{x})-1)}{1-\delta+\delta r(\boldsymbol{x})}\,d\boldsymbol{x} \leqslant \frac{e^{o(n)}-1}{1-\delta+\delta e^{o(n)}}, \quad 0<\delta\leqslant 1. \tag{52}$$

Заметим, что

$$\int\limits_{\mathbb{R}^n} \frac{p(\boldsymbol{x})}{1-\delta+\delta r(\boldsymbol{x})}\,d\boldsymbol{x} \leqslant \frac{1}{1-\delta}, \quad 0<\delta\leqslant 1.$$

Поэтому для того чтобы выполнялось (52), необходимо иметь

$$\int\limits_{\mathbb{R}^n} \frac{p(\boldsymbol{x})r(\boldsymbol{x})}{1+\delta r(\boldsymbol{x})}\,d\boldsymbol{x} \leqslant \frac{e^{o(n)}}{(1-\delta)(1-\delta+\delta e^{o(n)})} \quad \text{для всех } 0<\delta\leqslant 1. \tag{53}$$

Полагая $\delta \downarrow 0$, из (53) получаем необходимое условие

$$\int\limits_{\mathbb{R}^n} p(\boldsymbol{x})r(\boldsymbol{x})\,d\boldsymbol{x} = \mathbf{E}_{\boldsymbol{I}_n}\frac{p_{\boldsymbol{V}_n}}{p_{\boldsymbol{M}_n}}(\boldsymbol{x}) \leqslant e^{o(n)}, \tag{54}$$

которое совпадает с (40) для матрицы $\boldsymbol{M}_n$. В случае множества $\mathcal{M}_n(\boldsymbol{M}_n)$ условие (54) должно выполняться для всех $\boldsymbol{V}_n \in \mathcal{M}_n(\boldsymbol{M}_n)$, т.е. необходимо иметь

$$\sup_{\boldsymbol{V}_n\in\mathcal{M}_n(\boldsymbol{M}_n)} \mathbf{E}_{\boldsymbol{I}_n}\frac{p_{\boldsymbol{V}_n}}{p_{\boldsymbol{M}_n}}(\boldsymbol{x}) \leqslant e^{o(n)}, \tag{55}$$

откуда следует "внешняя граница" для $\mathcal{M}_n(\boldsymbol{M}_n)$ (см. (17)).

Нам остается выразить условие (54) в терминах матриц $\boldsymbol{M}_n$ и $\boldsymbol{V}_n$.

Лемма 2. *Если $\boldsymbol{I}_n + \boldsymbol{V}_n^{-1} - \boldsymbol{M}_n^{-1} > 0$, то справедлива формула* (см. (15))

$$\mathbf{E}_{\boldsymbol{I}_n}\frac{p_{\boldsymbol{V}_n}}{p_{\boldsymbol{M}_n}}(\boldsymbol{x}) = \frac{|\boldsymbol{M}_n|^{1/2}}{|\boldsymbol{I}_n+\boldsymbol{V}_n(\boldsymbol{I}_n-\boldsymbol{M}_n^{-1})|^{1/2}} = f^{-1/2}(\boldsymbol{M}_n,\boldsymbol{V}_n). \tag{56}$$

Если матрица $\boldsymbol{I}_n + \boldsymbol{V}_n^{-1} - \boldsymbol{M}_n^{-1}$ не является положительно определенной, то

$$\mathbf{E}_{\boldsymbol{I}_n}\frac{p_{\boldsymbol{V}_n}}{p_{\boldsymbol{M}_n}}(\boldsymbol{x}) = \infty. \tag{57}$$

Доказательство. В силу (6)

$$\mathbf{E}_{\boldsymbol{I}_n}\frac{p_{\boldsymbol{V}_n}}{p_{\boldsymbol{M}_n}}(\boldsymbol{x}) = \mathbf{E}_{\boldsymbol{\xi}_n}\frac{p_{\boldsymbol{V}_n}}{p_{\boldsymbol{M}_n}}(\boldsymbol{\xi}_n) = \frac{|\boldsymbol{M}_n|^{1/2}}{|\boldsymbol{V}_n|^{1/2}}\,\mathbf{E}_{\boldsymbol{\xi}_n}\,e^{-(\boldsymbol{\xi}_n,(\boldsymbol{V}_n^{-1}-\boldsymbol{M}_n^{-1})\boldsymbol{\xi}_n)/2}. \tag{58}$$

Заметим, что если $\boldsymbol{I}_n + \boldsymbol{A}_n > 0$, то для $\boldsymbol{\xi}_n \sim \mathcal{N}(\boldsymbol{0}_n, \boldsymbol{I}_n)$ из [11, гл. 6.9, теорема 3] имеем

$$\mathbf{E}_{\boldsymbol{\xi}_n}\,e^{-(\boldsymbol{\xi}_n,\boldsymbol{A}_n\boldsymbol{\xi}_n)/2} = \frac{1}{|\boldsymbol{I}_n+\boldsymbol{A}_n|^{1/2}}. \tag{59}$$

В противном случае

$$\mathbf{E}_{\boldsymbol{\xi}_n}\,e^{-(\boldsymbol{\xi}_n,\boldsymbol{A}_n\boldsymbol{\xi}_n)/2} = \infty. \tag{60}$$

Предположим сначала, что $\boldsymbol{I}_n + \boldsymbol{V}_n^{-1} - \boldsymbol{M}_n^{-1} > 0$. Тогда из (58) и (59) имеем

$$\mathbf{E}_{\boldsymbol{\xi}_n} \frac{p_{\boldsymbol{V}_n}}{p_{\boldsymbol{M}_n}}(\boldsymbol{\xi}_n) = \frac{|\boldsymbol{M}_n|^{1/2}}{|\boldsymbol{V}_n|^{1/2}|\boldsymbol{I}_n + \boldsymbol{V}_n^{-1} - \boldsymbol{M}_n^{-1}|^{1/2}} = \frac{|\boldsymbol{M}_n|^{1/2}}{|\boldsymbol{I}_n + \boldsymbol{V}_n(\boldsymbol{I}_n - \boldsymbol{M}_n^{-1})|^{1/2}}. \tag{61}$$

Если же матрица $\boldsymbol{I}_n + \boldsymbol{V}_n^{-1} - \boldsymbol{M}_n^{-1}$ не является положительно определенной, то в силу (58) и (60)

$$\mathbf{E}_{\boldsymbol{\xi}_n} \frac{p_{\boldsymbol{V}_n}}{p_{\boldsymbol{M}_n}}(\boldsymbol{\xi}_n) = \infty. \tag{62}$$

Из (58), (61) и (62) следует лемма 1. ▲

Поэтому в силу (56) условие (55) эквивалентно соотношению

$$\inf_{\boldsymbol{V}_n \in \mathcal{M}_n(\boldsymbol{M}_n)} \frac{|\boldsymbol{I}_n + \boldsymbol{V}_n(\boldsymbol{I}_n - \boldsymbol{M}_n^{-1})|}{|\boldsymbol{M}_n|} \geqslant e^{o(n)} \tag{63}$$

при условии, что матрица $\boldsymbol{I}_n + \boldsymbol{V}_n^{-1} - \boldsymbol{M}_n^{-1}$ положительно определена.

Если же матрица $\boldsymbol{I}_n + \boldsymbol{V}_n^{-1} - \boldsymbol{M}_n^{-1}$ не является положительно определенной, то в силу (57) и (62)

$$\mathbf{E}_{\boldsymbol{I}_n} \frac{p_{\boldsymbol{V}_n}}{p_{\boldsymbol{M}_n}}(\boldsymbol{x}) = \infty,$$

и поэтому условие (55) не может быть выполнено.

Определим $\mathcal{M}_n(\boldsymbol{M}_n)$ как максимальное множество, удовлетворяющее условию (63). Это множество совпадает с определением (15)–(17). Поэтому из (41), (55) и (56) следует теорема 1. ▲

4.3. Доказательство теоремы 2. Левую часть формулы (63) можно выразить через собственные значения матриц $\boldsymbol{M}_n$ и $\boldsymbol{V}_n$ следующим образом. Для ковариационной матрицы $\boldsymbol{M}_n$ с собственными значениями $\{\lambda_i\}$ рассмотрим ковариационные матрицы $\boldsymbol{V}_n$, коммутирующие с $\boldsymbol{M}_n$, т.е. $\boldsymbol{M}_n\boldsymbol{V}_n = \boldsymbol{V}_n\boldsymbol{M}_n$. Тогда каждая пара $\boldsymbol{M}_n, \boldsymbol{V}_n$ имеет одно и то же множество собственных векторов $\{\boldsymbol{x}_i\}$ [11, гл. 4.11, теорема 5]. Обозначим через $\{\nu_i\}$ собственные значения матрицы $\boldsymbol{V}_n$. Тогда матрица $\boldsymbol{B}_n = \boldsymbol{I}_n + \boldsymbol{V}_n(\boldsymbol{I}_n - \boldsymbol{M}_n^{-1})$ имеет собственные значения

$$1 + \nu_i - \nu_i/\lambda_i, \quad i = 1, \ldots, n.$$

Поэтому функция $f(\boldsymbol{M}_n, \boldsymbol{V}_n)$ из (15) принимает вид (19), откуда следует результат теоремы 2. ▲

4.4. Доказательство леммы 1. Пусть $\boldsymbol{\xi}_n$ – гауссовский случайный вектор с распределением $\boldsymbol{\xi}_n \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n)$, а $\boldsymbol{A}_n$ – симметричная $(n \times n)$-матрица с собственными значениями $\{a_i\}$. Рассмотрим квадратичную форму $(\boldsymbol{\xi}_n, \boldsymbol{A}_n\boldsymbol{\xi}_n)$. Существует ортогональная матрица $\boldsymbol{T}_n$, такая что $\boldsymbol{T}_n'\boldsymbol{A}_n\boldsymbol{T}_n = \boldsymbol{B}_n$, где $\boldsymbol{B}_n$ – диагональная матрица с диагональными элементами $\{a_i\}$ [11, гл. 4.7]. Так как $\boldsymbol{T}\boldsymbol{\xi}_n \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n)$, то квадратичные формы $(\boldsymbol{\xi}_n, \boldsymbol{A}\boldsymbol{\xi}_n)$ и $(\boldsymbol{\xi}_n, \boldsymbol{B}\boldsymbol{\xi}_n)$ имеют одинаковые распределения. Поэтому из формул (7) и (11) имеем

$$\ln \frac{p_{\boldsymbol{I}_n}}{p_{\boldsymbol{M}_n}}(\boldsymbol{y}_n) \stackrel{d}{=} \frac{1}{2}\big[\ln|\boldsymbol{M}_n| + \zeta_n\big], \tag{64}$$

где

$$\zeta_n = \big(\boldsymbol{y}_n, [\boldsymbol{M}_n^{-1} - \boldsymbol{I}_n]\boldsymbol{y}_n\big). \tag{65}$$

Используем следующий результат [15, гл. III.5.15]: пусть $\zeta_1, \ldots, \zeta_n$ – независимые случайные величины с $\mathbf{E}\,\zeta_i = 0$, $i = 1, \ldots, n$. Тогда для любого $1 \leqslant p \leqslant 2$

$$\mathbf{E}\left|\sum_{i=1}^{n} \zeta_i\right|^p \leqslant 2\sum_{i=1}^{n} \mathbf{E}\,|\zeta_i|^p \tag{66}$$

при условии, что правая часть (66) конечна.

Для оценки величины α_μ из (36) используем неравенство Чебышева и формулы (11) и (64)–(66). Тогда при $1 \leqslant p \leqslant 2$ получаем

$$\begin{aligned}
\alpha_\mu &= \mathbf{P}_{\boldsymbol{I}_n}\Big\{\ln \frac{p_{\boldsymbol{I}_n}}{p_{\boldsymbol{M}_n}}(\boldsymbol{x}) \leqslant D(\boldsymbol{I}_n \,\|\, \boldsymbol{M}_n) - \mu\Big\} \leqslant \\
&\leqslant \mathbf{P}_{\boldsymbol{I}_n}\Big\{\Big|\ln \frac{p_{\boldsymbol{I}_n}}{p_{\boldsymbol{M}_n}}(\boldsymbol{x}) - D(\boldsymbol{I}_n \,\|\, \boldsymbol{M}_n)\Big| > \mu\Big\} = \\
&= \mathbf{P}_{\boldsymbol{\xi}_n}\left\{\left|\zeta_n - \sum_{i=1}^{n}\Big(\frac{1}{\lambda_i} - 1\Big)\right| > 2\mu\right\} = \mathbf{P}_{\boldsymbol{\xi}_n}\left\{\left|\sum_{i=1}^{n}\Big(\frac{1}{\lambda_i} - 1\Big)(\xi_i^2 - 1)\right| > 2\mu\right\} \leqslant \\
&\leqslant (2\mu)^{-p}\,\mathbf{E}\left|\sum_{i=1}^{n}\Big(\frac{1}{\lambda_i} - 1\Big)(\xi_i^2 - 1)\right|^p \leqslant 2(2\mu)^{-p}\sum_{i=1}^{n}\mathbf{E}\left|\Big(\frac{1}{\lambda_i} - 1\Big)(\xi_i^2 - 1)\right|^p \leqslant \\
&\leqslant 2\mu^{-p}\sum_{i=1}^{n}\left|\frac{1}{\lambda_i} - 1\right|^p \leqslant 2(C_p + 1)\mu^{-p} n,
\end{aligned} \tag{67}$$

так как $\mathbf{E}\,|\xi_i^2 - 1|^p \leqslant [\mathbf{E}(\xi_i^2 - 1)^2]^{p/2} = 2^{p/2}$, $0 < p \leqslant 2$. Формула (30) следует из (67) и условия $\alpha_\mu \leqslant \alpha$. ▲

§ 5. Модель "сигнал + шум". Стационарные стохастические сигналы

Наряду с моделью (1) рассмотрим также следующую популярную модель "сигнал + шум":

$$\begin{aligned}
&\mathcal{H}_0\colon \boldsymbol{y}_n = \boldsymbol{\xi}_n, \quad \boldsymbol{\xi}_n \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n), \\
&\mathcal{H}_1\colon \boldsymbol{y}_n = \boldsymbol{\xi}_n + \boldsymbol{s}_n, \quad \boldsymbol{s}_n \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{S}_n), \quad \boldsymbol{S}_n \in \mathcal{S}_n,
\end{aligned} \tag{68}$$

где $\boldsymbol{s}_n$ – не зависящий от $\boldsymbol{\xi}_n$ гауссовский случайный вектор ("стохастический сигнал") с $\boldsymbol{s}_n \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{S}_n)$, а $\mathcal{S}_n$ – заданное множество ковариационных матриц $\boldsymbol{S}_n$.

5.1. Модель (68). Так как модель (68) является частным случаем модели (1), то к ней можно применить теоремы 1 и 2, в которых следует положить $\boldsymbol{M}_n = \boldsymbol{S}_n + \boldsymbol{I}_n$. Обозначим через $\{\mu_i(\boldsymbol{S}_n)\}$ собственные значения (все они положительны) ковариационной матрицы $\boldsymbol{S}_n$. Тогда $\mu_i(\boldsymbol{S}_n) = \lambda_i(\boldsymbol{M}_n) - 1$, $i = 1, \ldots, n$. Так как $\lambda_i(\boldsymbol{M}_n) > 1$, $i = 1, \ldots, n$, то предположение (14) выполняется автоматически. Предположение (13) принимает следующий вид:

III. Для последовательности множеств $\mathcal{S}_n$ существует положительный предел

$$\lim_{n\to\infty} \frac{1}{n} \inf_{\boldsymbol{S}_n \in \mathcal{S}_n} \sum_{i=1}^{n}\left[\ln(\mu_i(\boldsymbol{S}_n) + 1) + \frac{1}{\mu_i(\boldsymbol{S}_n) + 1} - 1\right] > 0. \tag{69}$$

Вместо функции $f(\boldsymbol{M}_n, \boldsymbol{V}_n)$ из (15) введем ее аналог $t(\boldsymbol{S}_n, \boldsymbol{V}_n)$. Для любых $\boldsymbol{S}_n, \boldsymbol{V}_n \in \mathcal{C}_n$, таких что $\boldsymbol{I}_n + (\boldsymbol{I}_n + \boldsymbol{V}_n)^{-1} - (\boldsymbol{I}_n + \boldsymbol{S}_n)^{-1} > 0$, определим функцию

$$t(\boldsymbol{S}_n, \boldsymbol{V}_n) = \frac{|\boldsymbol{I}_n + \boldsymbol{S}_n + (\boldsymbol{V}_n - \boldsymbol{S}_n)\boldsymbol{S}_n(\boldsymbol{I}_n + \boldsymbol{S}_n)^{-1}|}{|\boldsymbol{I}_n + \boldsymbol{S}_n|}. \tag{70}$$

В определении (70) использовалась простая формула

$$\boldsymbol{I}_n - (\boldsymbol{I}_n + \boldsymbol{S}_n)^{-1} = \boldsymbol{S}_n(\boldsymbol{I}_n + \boldsymbol{S}_n)^{-1}.$$

В качестве прямого следствия из теоремы 1 получаем

Предложение 2. *Если для модели* (68) *выполнено предположение* (69), *то наибольшее множество* $\mathcal{S}_n(\boldsymbol{S}_n)$, *удовлетворяющее асимптотическому равенству*

$$\lim_{n\to\infty} \frac{1}{n} \ln \beta(\boldsymbol{S}_n) = \lim_{n\to\infty} \frac{1}{n} \ln \beta(\mathcal{S}_n), \tag{71}$$

имеет вид ($n \to \infty$)

$$\mathcal{S}_n(\boldsymbol{S}_n) = \Big\{ \boldsymbol{V}_n : \boldsymbol{I}_n + (\boldsymbol{I}_n + \boldsymbol{V}_n)^{-1} - (\boldsymbol{I}_n + \boldsymbol{S}_n)^{-1} > 0,$$
$$\inf_{\boldsymbol{V}_n \in \mathcal{S}_n(\boldsymbol{S}_n)} \ln t(\boldsymbol{S}_n, \boldsymbol{V}_n) \geqslant o(n) \Big\}. \tag{72}$$

Ясно, что множество $\mathcal{S}_n(\boldsymbol{S}_n)$ является выпуклым по $\boldsymbol{V}_n$.

Упростим предложение 2 с помощью теоремы 2 следующим образом. Для матрицы $\boldsymbol{S}_n$ с собственными значениями $\{\mu_i\}$ рассмотрим в (72) только матрицы $\boldsymbol{V}_n$, коммутирующие с $\boldsymbol{S}_n$. Обозначим через $\{\nu_i\}$ собственные значения матрицы $\boldsymbol{V}_n$. Тогда аналогично (19) получаем

$$t(\boldsymbol{S}_n, \boldsymbol{V}_n) = \prod_{i=1}^{n} \left[1 + \frac{(\nu_i - \mu_i)\mu_i}{(1+\mu_i)^2} \right]. \tag{73}$$

Аналогично (18) для матрицы $\boldsymbol{S}_n$ введем выпуклое множество $\mathcal{C}_{\boldsymbol{S}_n}$ ковариационных матриц $\boldsymbol{V}_n$, коммутирующих с $\boldsymbol{S}_n$:

$$\mathcal{C}_{\boldsymbol{S}_n} = \{ \boldsymbol{V}_n : \boldsymbol{S}_n \boldsymbol{V}_n = \boldsymbol{V}_n \boldsymbol{S}_n \}. \tag{74}$$

Введем также следующее подмножество этого множества (см. (73)):

$$\mathcal{V}_n^{(2)}(\boldsymbol{S}_n) = \Big\{ \boldsymbol{V}_n \in \mathcal{C}_{\boldsymbol{S}_n} : \inf_{\boldsymbol{V}_n \in \mathcal{V}_n^{(2)}(\boldsymbol{S}_n)} \ln t(\boldsymbol{S}_n, \boldsymbol{V}_n) \geqslant o(n) \Big\}. \tag{75}$$

Множество $\mathcal{V}_n^{(2)}(\boldsymbol{S}_n)$ является выпуклым по $\boldsymbol{V}_n$, так как функция $\ln z$ выпукла по $z > 0$. Тогда аналогично теореме 2 получаем следующую "внутреннюю границу" $\mathcal{V}_n^{(2)}(\boldsymbol{S}_n)$ для $\mathcal{S}_n(\boldsymbol{S}_n)$.

Предложение 3. *Если для модели* (68) *выполнено предположение* (69), *то для наибольшего множества* $\mathcal{S}_n(\boldsymbol{S}_n)$, *удовлетворяющего формуле* (71), *имеем*

$$\mathcal{V}_n^{(2)}(\boldsymbol{S}_n) \subseteq \mathcal{S}_n(\boldsymbol{S}_n),$$

где множество $\mathcal{V}_n^{(2)}(\boldsymbol{S}_n)$ *определено в* (74), (75).

5.2. Стационарные стохастические сигналы. Рассмотрим модель (68), где $\boldsymbol{s}_n \sim \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{S}_n)$ – стационарный в широком смысле гауссовский случайный процесс с нулевым средним и спектральной плотностью $f_{\boldsymbol{S}_n}(\omega)$, $\omega \in [-\pi, \pi]$. Эта модель рассматривалась в работе [4]. Покажем, что основной результат [4, теорема 1] следует из предложения 3. Для этого рассмотрим какое-либо множество $\mathcal{K}_n$ ковариационных матриц $\boldsymbol{K}_n$, содержащее $\boldsymbol{S}_n$. Ограничимся только случаем, когда ковариационные матрицы $\boldsymbol{K}_n$ со спектральными плотностями $f_{\boldsymbol{K}_n}(\omega)$ коммутируют с $\boldsymbol{S}_n$. Тогда можно применить предложение 3, выраженное через спектральные плотности.

С помощью теоремы Сегё [16, гл. 5.2, теорема; 17, теорема 4.1] заменим предположение (69) его аналогом:

IV. Для последовательности множеств $\mathcal{K}_n$ существует положительный предел

$$\lim_{n\to\infty} \inf_{\boldsymbol{K}_n\in\mathcal{K}_n} \int\limits_{-\pi}^{\pi} \left[\ln(f_{\boldsymbol{K}_n}(\omega)+1) + \frac{1}{f_{\boldsymbol{K}_n}(\omega)+1} - 1\right] d\omega > 0. \tag{76}$$

Для заданной спектральной плотности $f_{\boldsymbol{S}_n}(\omega)$ обозначим через $\mathcal{F}_n(f_{\boldsymbol{S}_n})$ наибольшее множество спектральных плотностей, удовлетворяющее следующему аналогу равенства (5):

$$\lim_{n\to\infty} \frac{1}{n} \ln \beta(f_{\boldsymbol{S}_n}) = \lim_{n\to\infty} \frac{1}{n} \ln \beta(\mathcal{F}_n(f_{\boldsymbol{S}_n})). \tag{77}$$

Другими словами, при заданной вероятности ошибки 1-го рода α множество $\mathcal{F}_n(f_{\boldsymbol{S}_n})$ есть максимальное множество спектральных плотностей, которое можно заменить одной спектральной плотностью $f_{\boldsymbol{S}_n}$ (без потери для экспоненты вероятности ошибки 2-го рода $\beta(\mathcal{F}_n(f_{\boldsymbol{S}_n}))$).

Для того чтобы описать "внутреннюю границу" для множества $\mathcal{F}_n(f_{\boldsymbol{S}_n})$, введем следующий функционал:

$$f(\boldsymbol{S}_n, \boldsymbol{K}_n) = \int\limits_{-\pi}^{\pi} \ln\left[1 + \frac{f_{\boldsymbol{S}_n}(\omega)[f_{\boldsymbol{K}_n}(\omega) - f_{\boldsymbol{S}_n}(\omega)]}{(1+f_{\boldsymbol{S}_n}(\omega))^2}\right] d\omega. \tag{78}$$

Функционал $f(\boldsymbol{S}_n, \boldsymbol{K}_n)$ является аналогом функционала $\ln t(\boldsymbol{S}_n, \boldsymbol{V}_n)$ из (73), который следует в силу теоремы Сегё. Введем также множество (при $n\to\infty$)

$$\mathcal{F}_n^{(1)}(f_{\boldsymbol{S}_n}) = \Big\{ f_{\boldsymbol{K}_n}(\omega) : \inf_{f_{\boldsymbol{K}_n}(\omega)\in\mathcal{F}_n^{(1)}(f_{\boldsymbol{S}_n})} f(\boldsymbol{S}_n, \boldsymbol{K}_n) \geqslant o(1) \Big\}. \tag{79}$$

Множество $\mathcal{F}_n^{(1)}(f_{\boldsymbol{S}_n})$ является аналогом множества $\mathcal{V}_n^{(2)}(\boldsymbol{S}_n)$ из (75).

Как прямое следствие предложения 3 получаем

Предложение 4. *Если для модели* (68) *выполнено предположение* (76), *то для наибольшего множества* $\mathcal{F}_n(f_{\boldsymbol{S}_n})$, *для которого выполняется формула* (77), *имеем*

$$\mathcal{F}_n^{(1)}(\boldsymbol{S}_n) \subseteq \mathcal{F}_n(f_{\boldsymbol{S}_n}), \tag{80}$$

где множество $\mathcal{F}_n^{(1)}(\boldsymbol{S}_n)$ *определено в* (78), (79).

Множество $\mathcal{F}_n^{(1)}(\boldsymbol{S}_n)$ является выпуклым по $f_{\boldsymbol{K}_n}(\omega)$, так как функция $\ln z$ выпукла по $z>0$. Другими словами, если $f_{\boldsymbol{K}_n^{(0)}}(\omega), f_{\boldsymbol{K}_n^{(1)}}(\omega) \in \mathcal{F}_n^{(1)}(\boldsymbol{S}_n)$, то тогда

$$cf_{\boldsymbol{K}_n^{(0)}}(\omega) + (1-c) f_{\boldsymbol{K}_n^{(1)}}(\omega) \in \mathcal{F}_n^{(1)}(\boldsymbol{S}_n) \quad \text{для любого } 0 \leqslant c \leqslant 1.$$

Замечание 5. Множество $\mathcal{F}_n^{(1)}(\boldsymbol{S}_n)$ из (78), (79) (при $n\to\infty$) использовалось в [4, теорема 1].

§ 6. Примеры

Пример 1. В некоторых симметричных случаях множества $\mathcal{M}_n(\boldsymbol{M}_n)$ из (17) и $\mathcal{V}_n^{(0)}(\boldsymbol{M}_n)$ из (20) могут совпадать. Например, пусть в модели (1) $\boldsymbol{M}_n = c\boldsymbol{I}_n$, где

$c > 0$. Тогда для любой $\boldsymbol{V}_n$ матрицы $\boldsymbol{M}_n$ и $\boldsymbol{V}_n$ коммутируют, т.е. $\boldsymbol{M}_n\boldsymbol{V}_n = \boldsymbol{V}_n\boldsymbol{M}_n$. В этом случае теоремы 1 и 2 дают

$$\mathcal{M}_n(\boldsymbol{M}_n) = \mathcal{V}_n^{(0)}(\boldsymbol{M}_n).$$

Пример 2. Пусть в модели (68) сигнальный процесс $\{s_i\}$ имеет инвариантную во времени структуру

$$\begin{aligned} &s_{i+1} = as_i + \sqrt{1-a^2}\, u_i, \quad i = 1, \dots, n, \\ &s_1 \sim \mathcal{N}(0,1), \quad u_i \sim \mathcal{N}(0,1), \end{aligned} \tag{81}$$

где a – известная величина (параметр скорости обновления), такая что $0 \leqslant a < 1$. Предположим, что шумовой процесс $\{u_i\}$ не зависит от зашумленных наблюдений $\{\xi_i\}$, а начальное состояние s_1 не зависит от $\{u_i\}$ для всех i. Сигнальная последовательность $\{s_i\}$ есть процесс авторегрессии первого порядка $AR(1)$ и является стационарным процессом. Этот пример является главным в [4, 18]. Основной результат для этого примера также следует из предложения 3. Действительно, спектр наблюдаемого процесса (81) при гипотезах $\mathcal{H}_0$ и $\mathcal{H}_1$ описывается формулами, соответственно,

$$f^{(0)}(\omega) = 1, \quad f^{(1)}(\omega) = 1 + f_{\boldsymbol{S}_n}(\omega), \quad \omega \in [-\pi, \pi],$$

где спектр сигнала дается пуассоновским ядром

$$f_{\boldsymbol{S}_n}(\omega) = \frac{1-a^2}{1 - 2a\cos\omega + a^2}.$$

Тогда множество $\mathcal{F}_n^{(1)}(\boldsymbol{S}_n)$ описывается формулами (78), (79). Аналогичные (но более громоздкие результаты) можно получить и для процессов авторегрессии $AR(k)$ порядка $k > 1$.

Автор благодарит рецензента за конструктивные критические замечания, улучшившие статью.

СПИСОК ЛИТЕРАТУРЫ

1. *Wald A.* Statistical Decision Functions. New York: Wiley, 1950.
2. *Леман Э.* Проверка статистических гипотез. М.: Наука, 1979.
3. *Poor H.V.* An Introduction to Signal Detection and Estimation. New York: Springer-Verlag, 1994.
4. *Zhang W., Poor H.V.* On Minimax Robust Detection of Stationary Gaussian Signals in White Gaussian Noise // IEEE Trans. Inform. Theory. 2011. V. 57. № 6. P. 3915–3924. https://doi.org/10.1109/TIT.2011.2136210
5. *Бурнашев М.В.* Об обнаружении гауссовских стохастических последовательностей // Пробл. передачи информ. 2017. Т. 53. № 4. С. 49–68. http://mi.mathnet.ru/ppi2252
6. *Кульбак С.* Теория информации и статистика. М.: Наука, 1967.
7. *Stein C.* Information and Comparison of Experiments. Unpublished.
8. *Chernoff H.* Large-Sample Theory: Parametric Case // Ann. Math. Statist. 1956. V. 27. № 1. P. 1–22. https://doi.org/10.1214/aoms/1177728347
9. *Cover T.M., Thomas J.A.* Elements of Information Theory. New York: Wiley, 1991.
10. *Burnashev M.V.* On Neyman-Pearson Minimax Detection of Poisson Process Intensity // Stat. Inference Stoch. Process. 2021. V. 21. № 1. P. 211–221. https://doi.org/10.1007/s11203-020-09230-4
11. *Беллман Р.* Введение в теорию матриц. М.: Наука, 1976.
12. *Horn R.A., Johnson C.R.* Matrix Analysis. Cambridge: Cabmridge Univ. Press, 1985.

13. *Бурнашев М.В.* О минимаксном обнаружении неточно известного сигнала на фоне белого гауссовского шума // Теория вероятн. и ее примен. 1979. Т. 24. № 1. С. 106–118. http://mi.mathnet.ru/tvp957
14. *Бурнашев М.В.* Две теоремы сравнения распределений гауссовских квадратичных форм // Пробл. передачи информ. 2017. Т. 53. № 3. С. 3–15. http://mi.mathnet.ru/ppi2238
15. *Петров В.В.* Суммы независимых случайных величин. М.: Наука, 1972.
16. *Гренандер У., Сегё Г.* Теплицевы формы и их приложения. М.: Изд-во иностр. лит., 1961.
17. *Gray R.M.* On the Asymptotic Eigenvalue Distribution of Toeplitz Matrices // IEEE Trans. Inform. Theory. 1972. V. 18. № 6. P. 725–730. https://doi.org/10.1109/TIT.1972.1054924
18. *Sung Y., Tong L., Poor H.V.* Neyman–Pearson Detection of Gauss–Markov Signals in Noise: Closed-Form Error Exponent and Properties // IEEE Trans. Inform. Theory. 2006. V. 52. № 4. P. 1354–1365. https://doi.org/10.1109/TIT.2006.871599

Бурнашев Марат Валиевич
Институт проблем передачи информации
им. А.А. Харкевича РАН, Москва
burn@iitp.ru

Поступила в редакцию
15.04.2021
После доработки
16.06.2021
Принята к публикации
29.06.2021

ПРОБЛЕМЫ ПЕРЕДАЧИ ИНФОРМАЦИИ

Том 57 **2021** **Вып. 3**

УДК 519.651 : 517.589

© 2021 г. **Е.А. Карацуба**

О МЕТОДЕ ВЫЧИСЛЕНИЯ ДЗЕТА-КОНСТАНТ, ОСНОВАННОМ НА ОДНОМ ТЕОРЕТИКО-ЧИСЛОВОМ ПОДХОДЕ[1]

Новые формулы для дзета-констант получены на основе теоретико-числового подхода, применяемого для доказательства иррациональности некоторых классических констант. Используя эти формулы, можно приблизить дзета-константы и их комбинации рациональными дробями и построить новый эффективный метод их вычисления.

Ключевые слова: дзета-константы, приближения, рациональные дроби, подход Эрмита – Бёйкерса, многочлены Лежандра, обобщенные гармонические числа, алгоритмы.

DOI: 10.31857/S0555292321030050

§ 1. Введение

Проблема построения эффективных методов вычисления значений дзета-функции Римана в точках разной арифметической природы рассматривалась многими авторами (см., например, [1–6]).

В [7] был построен новый метод быстрого приближения дзета-констант, т.е. значений дзета-функции Римана $\zeta(n)$, $n \geqslant 2$, n – целое число, рациональными дробями. Этот метод возник на основе подхода Эрмита – Бёйкерса (см. [8–10]), который был применен последним для доказательства иррациональности дзета-констант $\zeta(2)$ и $\zeta(3)$ с использованием двух специально подобранных полиномов $P_n(x)$ и $Q_n(x)$, $n \geqslant 1$:

$$P_n(x) = \frac{1}{n!}\left(\frac{d}{dx}\right)^n (x^n(1-x)^n), \quad Q_n(x) = (1-x)^n. \tag{1}$$

Различные варианты метода Бёйкерса, его модификаций и обобщений широко применяются в последнее время при исследовании значений различных функций на иррациональность (см., например, [9–14], а также [4]).

В [7] был представлен метод, который позволяет приблизить дзета-константы и некоторые их комбинации достаточно простыми выражениями из рациональных дробей, в которых участвуют коэффициенты многочленов (1), и затем эффективно их вычислить.

В настоящей статье продолжены исследования, начатые в [7]. Построен метод с участием трех многочленов (а не двух, как в [7]) и описан алгоритм аппроксимации и вычисления дзета-констант. Следует упомянуть, что метод из [7] предполагает, что оба многочлена необходимы для обеспечения достаточно хорошего приближения (как в методе Бёйкерса, см. [9,10]). В представленном методе с тремя многочленами

[1] Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (номер проекта 19-07-00750).

можно сделать выбор и пойти двумя путями:

1) все три многочлена обеспечат скорость сходимости; или
2) два многочлена обеспечат скорость сходимости, а подбирая коэффициенты третьего многочлена, можно уменьшить количество вычисляемых рациональных дробей.

Итак, представленный метод с тремя многочленами является первым, в котором возможны манипуляции с коэффициентами (третьего многочлена) для понижения сложности вычисления.

Далее мы используем следующие стандартные обозначения. Обобщенное гармоническое число $H_n^{(m)}$ порядка m представляет собой следующую сумму для натуральных чисел n и m:

$$H_n^{(m)} = \sum_{k=1}^{n} \frac{1}{k^m}, \quad H_n^{(1)} = H_n, \quad H_0^{(m)} = 0. \tag{2}$$

Свойства и методы вычисления гармонических чисел хорошо изучены: см., например, [15–17].

§ 2. Вспомогательные формулы

При $|t| < 1$ имеем

$$\frac{1}{1-t} = 1 + t + t^2 + \ldots = \sum_{k=0}^{\infty} t^k.$$

Отсюда

$$I_3 = \int_0^1\int_0^1\int_0^1 \frac{dx\,dy\,dz}{1-xyz} = \int_0^1\int_0^1\int_0^1 \sum_{k=0}^{\infty} x^k y^k z^k\,dx\,dy\,dz =$$

$$= \sum_{k=0}^{\infty}\left(\int_0^1 x^k\,dx\right)\left(\int_0^1 y^k\,dy\right)\left(\int_0^1 z^k dz\right) = \sum_{k=0}^{\infty}\frac{1}{(k+1)^3} = \zeta(3).$$

Лемма 1. *При любых $r_1, r_2, r_3 \geqslant 0$, $s \geqslant 3$ справедливо соотношение*

$$I(r_1,r_2,r_3) = \int_0^1 \ldots \int_0^1 \frac{x_1^{r_1}x_2^{r_2}x_3^{r_3}}{1-x_1x_2x_3\ldots x_s}\,dx_1\,dx_2\,dx_3\ldots dx_s =$$

$$= \sum_{k=0}^{\infty}\frac{1}{(r_1+k+1)(r_2+k+1)(r_3+k+1)(k+1)^{s-3}}. \tag{3}$$

Доказательство. При $0 < x_1, \ldots, x_s < 1$, $0 < x_1x_2\ldots x_s < 1$, имеем

$$\frac{1}{1-x_1\ldots x_s} = \sum_{k=0}^{\infty}(x_1\ldots x_s)^k,$$

$$I(r_1,r_2,r_3) = \int_0^1 \ldots \int_0^1 \frac{x_1^{r_1}x_2^{r_2}x_3^{r_3}}{1-x_1x_2x_3\ldots x_s}\,dx_1\,dx_2\,dx_3\ldots dx_s =$$

$$= \sum_{k=0}^{\infty}\int_0^1 x_1^{k+r_1}\,dx_1\int_0^1 x_2^{k+r_2}\,dx_2\int_0^1 x_3^{k+r_3}\,dx_3\int_0^1 x_4^k\,dx_4\,\ldots\int_0^1 x_s^k\,dx_s =$$

$$= \sum_{k=0}^{\infty} \frac{1}{r_1+k+1} \cdot \frac{1}{r_2+k+1} \cdot \frac{1}{r_3+k+1} \cdot \frac{1}{(k+1)^{s-3}},$$

что и требовалось доказать. ▲

Лемма 2. *Для любого целого $s \geqslant 1$ и любых $r \geqslant 1$, $k \geqslant 0$ справедливы тождества*

$$\frac{1}{(r+k+1)(k+1)^s} = \sum_{j=1}^{s} (-1)^{j-1} \frac{1}{r^j (k+1)^{s+1-j}} + (-1)^s \frac{1}{r^s (r+k+1)}, \tag{4}$$

$$\frac{1}{(r+k+1)^2(k+1)^s} = \sum_{j=1}^{s} (-1)^{j-1} \frac{j}{r^{j+1}(k+1)^{s+1-j}} +$$
$$+ (-1)^s \frac{s}{r^{s+1}(r+k+1)} + (-1)^s \frac{1}{r^s (r+k+1)^2}, \tag{5}$$

$$\frac{1}{(r+k+1)^3(k+1)^s} = \sum_{j=1}^{s} (-1)^{j-1} \frac{j(j+1)}{2r^{j+2}(k+1)^{s+1-j}} +$$
$$+ (-1)^s \frac{s(s+1)}{2r^{s+2}(r+k+1)} + (-1)^s \frac{s}{r^{s+1}(r+k+1)^2} + (-1)^s \frac{1}{r^s (r+k+1)^3}. \tag{6}$$

Доказательство. Формулы (4)–(6) доказываются индукцией по s. Докажем, например, (6). При $s = 1$ справедливость равенства

$$\frac{1}{(r+k+1)^3(k+1)} =$$
$$= \frac{1}{r^3(k+1)} - \frac{1}{r^3(r+k+1)} - \frac{1}{r^2(r+k+1)^2} - \frac{1}{r(r+k+1)^3} \tag{7}$$

проверяется непосредственно. Пусть (6) справедливо при $s \leqslant n$, и пусть

$$\frac{1}{(r+k+1)^3(k+1)^n} = \sum_{j=1}^{n} (-1)^{j-1} \frac{j(j+1)}{2r^{j+2}(k+1)^{n+1-j}} +$$
$$+ (-1)^n \frac{n(n+1)}{2r^{n+2}(r+k+1)} + (-1)^n \frac{n}{r^{n+1}(r+k+1)^2} + (-1)^n \frac{1}{r^n (r+k+1)^3}. \tag{8}$$

Докажем, что (6) справедливо также при $s = n+1$. Из (8) имеем

$$\frac{1}{(r+k+1)^3(k+1)^{n+1}} = \frac{1}{k+1} \sum_{j=1}^{n} \frac{(-1)^{j-1} j(j+1)}{2r^{j+2}(k+1)^{n+1-j}} +$$
$$+ \frac{(-1)^n n(n+1)}{2r^{n+2}(k+1)(r+k+1)} + \frac{(-1)^n n}{r^{n+1}(k+1)(r+k+1)^2} + \frac{(-1)^n}{r^n (k+1)(r+k+1)^3}.$$

Подставляя в последнее выражение (4), (5) при $s = 1$ и формулу (7), находим

$$\frac{1}{(r+k+1)^3(k+1)^{n+1}} =$$
$$= \sum_{j=1}^{n} \frac{(-1)^{j-1} j(j+1)}{2r^{j+2}(k+1)^{n+2-j}} + \frac{(-1)^n n(n+1)}{2r^{n+2}} \left(\frac{1}{r(k+1)} - \frac{1}{r(r+k+1)} \right) +$$

$$+\frac{(-1)^n n}{r^{n+1}}\left(\frac{1}{r^2(k+1)}-\frac{1}{r^2(r+k+1)}-\frac{1}{r(r+k+1)^2}\right)+$$
$$+\frac{(-1)^n}{r^n}\left(\frac{1}{r^3(k+1)}-\frac{1}{r^3(r+k+1)}-\frac{1}{r^2(r+k+1)^2}-\frac{1}{r(r+k+1)^3}\right)=$$
$$=\sum_{j=1}^{n}\frac{(-1)^{j-1}j(j+1)}{2r^{j+2}(k+1)^{n+2-j}}+\frac{(-1)^n\left(\frac{n(n+1)}{2}+n+1\right)}{r^{n+3}(k+1)}+$$
$$+\frac{(-1)^{n+1}\left(\frac{n(n+1)}{2}+n+1\right)}{r^{n+3}(r+k+1)}+\frac{(-1)^{n+1}(n+1)}{r^{n+2}(r+k+1)^2}+\frac{(-1)^{n+1}}{r^{n+1}(r+k+1)^3}=$$
$$=\sum_{j=1}^{n+1}\frac{(-1)^{j-1}j(j+1)}{2r^{j+2}(k+1)^{n+2-j}}+\frac{(-1)^{n+1}(n+1)(n+2)}{2r^{n+3}(r+k+1)}+$$
$$+\frac{(-1)^{n+1}(n+1)}{r^{n+2}(r+k+1)^2}+\frac{(-1)^{n+1}}{r^{n+1}(r+k+1)^3},$$

т.е. получаем формулу (6) при $s=n+1$. ▲

§ 3. Основная формула

Пусть $a_0,a_1,\ldots,a_n$, $b_0,b_1,\ldots,b_n$, $c_0,c_1,\ldots,c_n$ – произвольные числа; рассмотрим многочлены $P_n(x)$, $Q_n(x)$ и $T_n(x)$, $n\geqslant 1$:

$$P_n(x)=a_0+a_1x+\ldots+a_nx^n,$$
$$Q_n(x)=b_0+b_1x+\ldots+b_nx^n,$$
$$T_n(x)=c_0+c_1x+\ldots+c_nx^n.$$

Умножим левую и правую части соотношения (3) на $a_{r_1}b_{r_2}c_{r_3}$ и просуммируем по $r_1,r_2,r_3=0,1,\ldots,n$; $s\geqslant 3$:

$$I_s=I_s(n)=\sum_{r_1=0}^{n}\sum_{r_2=0}^{n}\sum_{r_3=0}^{n}a_{r_1}b_{r_2}c_{r_3}I(r_1,r_2,r_3)=$$
$$=\int_0^1\ldots\int_0^1\frac{P_n(x_1)Q_n(x_2)T_n(x_3)}{1-x_1x_2x_3\ldots x_s}\,dx_1\,dx_2\,dx_3\ldots dx_s= \tag{9}$$
$$=\sum_{k=0}^{\infty}\frac{1}{(k+1)^{s-3}}\sum_{r_1=0}^{n}\sum_{r_2=0}^{n}\sum_{r_3=0}^{n}\frac{a_{r_1}b_{r_2}c_{r_3}}{(r_1+k+1)(r_2+k+1)(r_3+k+1)}.$$

Выделяя в (9) члены с $r_1=r_2=r_3=0$, затем с $r_1=r_2=0$, $r_3\neq 0$, $r_2=r_3=0$, $r_1\neq 0$, $r_1=r_3=0$, $r_2\neq 0$, и наконец, с $r_1=0$, $r_2,r_3\neq 0$, $r_2=0$, $r_1,r_3\neq 0$, $r_3=0$, $r_1,r_2\neq 0$, получаем

$$I_s=a_0b_0c_0\zeta(s)+a_0b_0\sum_{k=0}^{\infty}\frac{1}{(k+1)^{s-1}}\sum_{r_3=1}^{n}\frac{c_{r_3}}{r_3+k+1}+$$
$$+a_0c_0\sum_{k=0}^{\infty}\frac{1}{(k+1)^{s-1}}\sum_{r_2=1}^{n}\frac{b_{r_2}}{r_2+k+1}+b_0c_0\sum_{k=0}^{\infty}\frac{1}{(k+1)^{s-1}}\sum_{r_1=1}^{n}\frac{a_{r_1}}{r_1+k+1}+$$
$$+a_0\sum_{k=0}^{\infty}\frac{1}{(k+1)^{s-2}}\sum_{r_2=1}^{n}\sum_{r_3=1}^{n}\frac{b_{r_2}c_{r_3}}{(r_2+k+1)(r_3+k+1)}+$$

$$+ b_0 \sum_{k=0}^{\infty} \frac{1}{(k+1)^{s-2}} \sum_{r_1=1}^{n} \sum_{r_3=1}^{n} \frac{a_{r_1} c_{r_3}}{(r_1+k+1)(r_3+k+1)} +$$
$$+ c_0 \sum_{k=0}^{\infty} \frac{1}{(k+1)^{s-2}} \sum_{r_1=1}^{n} \sum_{r_2=1}^{n} \frac{a_{r_1} b_{r_2}}{(r_1+k+1)(r_2+k+1)} +$$
$$+ \sum_{k=0}^{\infty} \frac{1}{(k+1)^{s-3}} \sum_{r_1=1}^{n} \sum_{r_2=1}^{n} \sum_{r_3=1}^{n} \frac{a_{r_1} b_{r_2} c_{r_3}}{(r_1+k+1)(r_2+k+1)(r_3+k+1)}. \tag{10}$$

Разобьем двойные суммы по r_μ, $\mu = 1, 2, 3$, в правой части (10) на две суммы: сумму, в которой индексы суммирования равны, и сумму, в которой они не равны, т.е., например,

$$\sum_{r_1=1}^{n} \sum_{r_2=1}^{n} \frac{a_{r_1} b_{r_2}}{(r_1+k+1)(r_2+k+1)} =$$
$$= \sum_{r_1=1}^{n} \frac{a_{r_1} b_{r_1}}{(r_1+k+1)^2} + \sum_{\substack{r_1, r_2=1 \\ r_1 \neq r_2}}^{n} \frac{a_{r_1} b_{r_2}}{r_2 - r_1} \left(\frac{1}{r_1+k+1} - \frac{1}{r_2+k+1} \right). \tag{11}$$

В тройной сумме по r_ξ из (10) также выделим слагаемые с одинаковыми индексами, воспользовавшись соотношениями

$$\sum_{r_1=1}^{n} \sum_{r_2=1}^{n} \sum_{r_3=1}^{n} \frac{a_{r_1} b_{r_2} c_{r_3}}{(r_1+k+1)(r_2+k+1)(r_3+k+1)} = \sum_{r_1=1}^{n} \frac{a_{r_1} b_{r_1} c_{r_1}}{(r_1+k+1)^3} +$$
$$+ \sum_{\substack{r_1, r_2, r_3=1 \\ r_1 = r_2 \neq r_3}}^{n} \frac{a_{r_1} b_{r_1} c_{r_3}}{(r_1+k+1)^2 (r_3+k+1)} + \sum_{\substack{r_1, r_2, r_3=1 \\ r_1 = r_3 \neq r_2}}^{n} \frac{a_{r_1} b_{r_2} c_{r_1}}{(r_1+k+1)^2 (r_2+k+1)} +$$
$$+ \sum_{\substack{r_1, r_2, r_3=1 \\ r_2 = r_3 \neq r_1}}^{n} \frac{a_{r_1} b_{r_2} c_{r_2}}{(r_2+k+1)^2 (r_1+k+1)} + \sum_{\substack{r_1, r_2, r_3=1 \\ r_1 \neq r_2 \neq r_3}}^{n} \frac{a_{r_1} b_{r_2} c_{r_3}}{(r_1+k+1)(r_2+k+1)(r_3+k+1)}.$$

Таким образом,

$$\sum_{r_1=1}^{n} \sum_{r_2=1}^{n} \sum_{r_3=1}^{n} \frac{a_{r_1} b_{r_2} c_{r_3}}{(r_1+k+1)(r_2+k+1)(r_3+k+1)} =$$
$$= \sum_{r_1=1}^{n} \frac{a_{r_1} b_{r_1} c_{r_1}}{(r_1+k+1)^3} + \sum_{\substack{r_1, r_2=1 \\ r_1 \neq r_2}}^{n} \frac{a_{r_1} b_{r_1} c_{r_2} + a_{r_1} b_{r_2} c_{r_1} + a_{r_2} b_{r_1} c_{r_1}}{(r_1+k+1)^2 (r_2+k+1)} -$$
$$- \sum_{\substack{r_1, r_2, r_3=1 \\ r_1 \neq r_2 \neq r_3}}^{n} \frac{a_{r_1} b_{r_2} c_{r_3}}{(r_1+k+1)(r_3-r_2)} \left(\frac{1}{r_3+k+1} - \frac{1}{r_2+k+1} \right).$$

Отсюда находим

$$\sum_{r_1=1}^{n} \sum_{r_2=1}^{n} \sum_{r_3=1}^{n} \frac{a_{r_1} b_{r_2} c_{r_3}}{(r_1+k+1)(r_2+k+1)(r_3+k+1)} =$$

$$= \sum_{r_1=1}^{n} \frac{a_{r_1} b_{r_1} c_{r_1}}{(r_1+k+1)^3} + \sum_{\substack{r_1,r_2=1 \\ r_1 \neq r_2}}^{n} \frac{a_{r_1} b_{r_1} c_{r_2} + a_{r_1} b_{r_2} c_{r_1} + a_{r_2} b_{r_1} c_{r_1}}{(r_1+k+1)^2 (r_2+k+1)} -$$

$$- \sum_{\substack{r_1,r_2,r_3=1 \\ r_1 \neq r_3 \neq r_2}}^{n} \frac{a_{r_1} b_{r_2} c_{r_3}}{(r_3-r_2)(r_2-r_1)} \left(\frac{1}{r_2+k+1} - \frac{1}{r_1+k+1} \right) +$$

$$+ \sum_{\substack{r_1,r_2,r_3=1 \\ r_1 \neq r_2 \neq r_3}}^{n} \frac{a_{r_1} b_{r_2} c_{r_3}}{(r_3-r_2)(r_3-r_1)} \left(\frac{1}{r_3+k+1} - \frac{1}{r_1+k+1} \right). \tag{12}$$

Подставляя (11), (12) в (10), находим:

$$I_s = a_0 b_0 c_0 \zeta(s) + \sum_{k=0}^{\infty} \sum_{r_1=1}^{n} \frac{a_0 b_0 c_{r_1} + a_0 c_0 b_{r_1} + b_0 c_0 a_{r_1}}{(k+1)^{s-1}(r_1+k+1)} +$$

$$+ \sum_{k=0}^{\infty} \sum_{r_1=1}^{n} \frac{a_0 b_{r_1} c_{r_1} + b_0 a_{r_1} c_{r_1} + c_0 a_{r_1} b_{r_1}}{(k+1)^{s-2}(r_1+k+1)^2} -$$

$$- \sum_{k=0}^{\infty} \sum_{\substack{r_1,r_2=1 \\ r_1 \neq r_2}}^{n} \frac{a_0 b_{r_1} c_{r_2} + b_0 a_{r_2} c_{r_1} + c_0 a_{r_1} b_{r_2}}{(k+1)^{s-2}(r_2-r_1)} \left(\frac{1}{r_2+k+1} - \frac{1}{r_1+k+1} \right) +$$

$$+ \sum_{k=0}^{\infty} \sum_{r_1=1}^{n} \frac{a_{r_1} b_{r_1} c_{r_1}}{(k+1)^{s-3}(r_1+k+1)^3} +$$

$$+ \sum_{k=0}^{\infty} \sum_{\substack{r_1,r_2=1 \\ r_1 \neq r_2}}^{n} \frac{a_{r_1} b_{r_1} c_{r_2} + a_{r_1} b_{r_2} c_{r_1} + a_{r_2} b_{r_1} c_{r_1}}{(k+1)^{s-3}(r_1+k+1)^2 (r_2-r_1)} +$$

$$+ \sum_{k=0}^{\infty} \sum_{\substack{r_1,r_2=1 \\ r_1 \neq r_2}}^{n} \frac{a_{r_1} b_{r_1} c_{r_2} + a_{r_1} b_{r_2} c_{r_1} + a_{r_2} b_{r_1} c_{r_1}}{(k+1)^{s-3}(r_2-r_1)^2} \left(\frac{1}{r_2+k+1} - \frac{1}{r_1+k+1} \right) +$$

$$+ \sum_{k=0}^{\infty} \sum_{\substack{r_1,r_2,r_3=1 \\ r_1 \neq r_2 \neq r_3}}^{n} \frac{a_{r_1} b_{r_2} c_{r_3}}{(k+1)^{s-3}} \left(\frac{1}{(r_3-r_2)(r_3-r_1)(r_3+k+1)} + \right.$$

$$\left. + \frac{1}{(r_2-r_3)(r_2-r_1)(r_2+k+1)} + \frac{1}{(r_1-r_3)(r_1-r_2)(r_1+k+1)} \right). \tag{13}$$

Из (13) при $s = 3$ имеем:

$$I_3 = a_0 b_0 c_0 \zeta(3) + \sum_{k=0}^{\infty} \sum_{r_1=1}^{n} \frac{a_0 b_0 c_{r_1} + a_0 c_0 b_{r_1} + b_0 c_0 a_{r_1}}{(k+1)^2 (r_1+k+1)} +$$

$$+ \sum_{k=0}^{\infty} \sum_{r_1=1}^{n} \frac{a_0 b_{r_1} c_{r_1} + b_0 a_{r_1} c_{r_1} + c_0 a_{r_1} b_{r_1}}{(k+1)(r_1+k+1)^2} -$$

$$- \sum_{k=0}^{\infty} \sum_{\substack{r_1,r_2=1 \\ r_1 \neq r_2}}^{n} \frac{a_0 b_{r_1} c_{r_2} + b_0 a_{r_2} c_{r_1} + c_0 a_{r_1} b_{r_2}}{(k+1)(r_2-r_1)} \left(\frac{1}{r_2+k+1} - \frac{1}{r_1+k+1} \right) +$$

$$+\sum_{k=0}^{\infty}\sum_{r_1=1}^{n}\frac{a_{r_1}b_{r_1}c_{r_1}}{(r_1+k+1)^3}+\sum_{k=0}^{\infty}\sum_{\substack{r_1,r_2=1\\ r_1\neq r_2}}^{n}\frac{a_{r_1}b_{r_1}c_{r_2}+a_{r_1}c_{r_1}b_{r_2}+b_{r_1}c_{r_1}a_{r_2}}{(r_1+k+1)^2(r_2-r_1)}+$$

$$+\sum_{k=0}^{\infty}\sum_{\substack{r_1,r_2=1\\ r_1\neq r_2}}^{n}\frac{a_{r_1}b_{r_1}c_{r_2}+a_{r_1}c_{r_1}b_{r_2}+b_{r_1}c_{r_1}a_{r_2}}{(r_2-r_1)^2}\left(\frac{1}{r_2+k+1}-\frac{1}{r_1+k+1}\right)+$$

$$+\sum_{k=0}^{\infty}\sum_{\substack{r_1,r_2,r_3=1\\ r_1\neq r_2\neq r_3}}^{n}\frac{a_{r_1}b_{r_2}c_{r_3}}{(r_3-r_2)(r_3-r_1)}\left(\frac{1}{r_3+k+1}-\frac{1}{r_1+k+1}\right)+$$

$$+\sum_{k=0}^{\infty}\sum_{\substack{r_1,r_2,r_3=1\\ r_1\neq r_2\neq r_3}}^{n}\frac{a_{r_1}b_{r_2}c_{r_3}}{(r_3-r_2)(r_1-r_2)}\left(\frac{1}{r_2+k+1}-\frac{1}{r_1+k+1}\right). \tag{14}$$

Для большей компактности представления формул введем обозначения

$$S_{00r_1}=a_0b_0c_{r_1}+a_0c_0b_{r_1}+b_0c_0a_{r_1},$$
$$S_{0r_1r_1}=a_0b_{r_1}c_{r_1}+b_0a_{r_1}c_{r_1}+c_0a_{r_1}b_{r_1},$$
$$S_{0r_1r_2}=a_0b_{r_1}c_{r_2}+b_0a_{r_2}c_{r_1}+c_0a_{r_1}b_{r_2},$$
$$S_{r_1r_1r_2}=a_{r_1}b_{r_1}c_{r_2}+a_{r_1}c_{r_1}b_{r_2}+b_{r_1}c_{r_1}a_{r_2}.$$

Преобразуя суммы из (14) на основе леммы 2, легко видеть следующее:

1) для второго слагаемого в правой части (14) справедливо соотношение

$$\sum_{k=0}^{\infty}\sum_{r_1=1}^{n}\frac{S_{00r_1}}{(k+1)^2(r_1+k+1)}=\zeta(2)\sum_{r_1=1}^{n}\frac{S_{00r_1}}{r_1}-\sum_{r_1=1}^{n}\frac{S_{00r_1}H_{r_1}}{r_1^2}; \tag{15}$$

2) для третьего слагаемого в правой части (14) справедливо равенство

$$\sum_{k=0}^{\infty}\sum_{r_1=1}^{n}\frac{S_{0r_1r_1}}{(k+1)(r_1+k+1)^2}=\sum_{r_1=1}^{n}\frac{S_{0r_1r_1}H_{r_1}}{r_1^2}-\sum_{r_1=1}^{n}\frac{S_{0r_1r_1}}{r_1}\left(\zeta(2)-H_{r_1}^{(2)}\right); \tag{16}$$

3) далее имеем

$$\sum_{k=0}^{\infty}\sum_{r_1=1}^{n}\frac{a_{r_1}b_{r_1}c_{r_1}}{(r_1+k+1)^3}=\sum_{r_1=1}^{n}a_{r_1}b_{r_1}c_{r_1}\left(\zeta(3)-H_{r_1}^{(3)}\right); \tag{17}$$

4) далее находим

$$\sum_{k=0}^{\infty}\sum_{\substack{r_1,r_2=1\\ r_1\neq r_2}}^{n}\frac{S_{0r_1r_2}}{r_2-r_1}\left(\frac{1}{(k+1)(r_2+k+1)}-\frac{1}{(k+1)(r_1+k+1)}\right)=$$
$$=\sum_{\substack{r_1,r_2=1\\ r_1\neq r_2}}^{n}\frac{S_{0r_1r_2}}{r_2-r_1}\left(\frac{H_{r_2}}{r_2}-\frac{H_{r_1}}{r_1}\right); \tag{18}$$

5) и наконец, замечаем, что

$$\sum_{k=0}^{\infty}\sum_{\substack{r_1,r_2=1\\ r_1\neq r_2}}^{n}\frac{S_{r_1r_1r_2}}{(r_1+k+1)^2(r_2-r_1)}=\sum_{\substack{r_1,r_2=1\\ r_1\neq r_2}}^{n}\frac{S_{r_1r_1r_2}}{r_2-r_1}\left(\zeta(2)-H_{r_1}^{(2)}\right). \tag{19}$$

Подставляя (15)–(19) в (14) и выполняя очевидные преобразования, получаем

$$I_3 = \zeta(3)\sum_{r_1=0}^{n} a_{r_1}b_{r_1}c_{r_1} + \zeta(2)\left(\sum_{\substack{r_1,r_2=1\\ r_1\neq r_2}}^{n}\left(\frac{S_{00r_1}-S_{0r_1r_1}}{r_1}+\frac{S_{r_1r_1r_2}}{r_2-r_1}\right)\right)+$$
$$+\sum_{\substack{r_1,r_2,r_3=1\\ r_1\neq r_2\neq r_3}}^{n}\left(-a_{r_1}b_{r_1}c_{r_1}H^{(3)}_{r_1}+\left(\frac{S_{0r_1r_1}}{r_1}-\frac{S_{r_1r_1r_2}}{r_2-r_1}\right)H^{(2)}_{r_1}+\frac{S_{0r_1r_1}-S_{00r_1}}{r_1^2}H_{r_1}-\right.$$
$$-\frac{S_{0r_1r_2}}{r_2-r_1}\left(\frac{H_{r_2}}{r_2}-\frac{H_{r_1}}{r_1}\right)-\frac{S_{r_1r_1r_2}}{(r_2-r_1)^2}\left(H_{r_2}-H_{r_1}\right)-a_{r_1}b_{r_2}c_{r_3}\times$$
$$\left.\times\left(\frac{H_{r_1}}{(r_1-r_3)(r_1-r_2)}+\frac{H_{r_2}}{(r_2-r_1)(r_2-r_3)}+\frac{H_{r_3}}{(r_3-r_1)(r_3-r_2)}\right)\right), \tag{20}$$

где $H^m_{r_\xi}$, $\xi = 1,2,3$, $m = 1,2,3$ суть гармонические числа, т.е. суммы, определяемые соотношениями (2). Аналогично из (13) при $s = 4$ находим

$$I_4 = a_0b_0c_0\zeta(4)+\sum_{k=0}^{\infty}\sum_{r_1=1}^{n}\frac{S_{00r_1}}{(k+1)^3(r_1+k+1)}+\sum_{k=0}^{\infty}\sum_{r_1=1}^{n}\frac{S_{0r_1r_1}}{(k+1)^2(r_1+k+1)^2}-$$
$$-\sum_{k=0}^{\infty}\sum_{\substack{r_1,r_2=1\\ r_1\neq r_2}}^{n}\frac{S_{0r_1r_2}}{(k+1)^2(r_2-r_1)}\left(\frac{1}{r_2+k+1}-\frac{1}{r_1+k+1}\right)+$$
$$+\sum_{k=0}^{\infty}\sum_{r_1=1}^{n}\frac{a_{r_1}b_{r_1}c_{r_1}}{(k+1)(r_1+k+1)^3}+\sum_{k=0}^{\infty}\sum_{\substack{r_1,r_2=1\\ r_1\neq r_2}}^{n}\frac{S_{r_1r_1r_2}}{(k+1)(r_1+k+1)^2(r_2-r_1)}+$$
$$+\sum_{k=0}^{\infty}\sum_{\substack{r_1,r_2=1\\ r_1\neq r_2}}^{n}\frac{S_{r_1r_1r_2}}{(r_2-r_1)^2(k+1)}\left(\frac{1}{r_2+k+1}-\frac{1}{r_1+k+1}\right)+$$
$$+\sum_{k=0}^{\infty}\sum_{\substack{r_1,r_2,r_3=1\\ r_1\neq r_2\neq r_3}}^{n}\frac{a_{r_1}b_{r_2}c_{r_3}}{k+1}\left(\frac{1}{(r_3-r_2)(r_3-r_1)(r_3+k+1)}+\right.$$
$$\left.+\frac{1}{(r_3-r_2)(r_1-r_2)(r_2+k+1)}+\frac{1}{(r_1-r_3)(r_1-r_2)(r_1+k+1)}\right). \tag{21}$$

Преобразуя суммы из (21) на основании леммы 2, получаем:

$$I_4 = a_0b_0c_0\zeta(4)+\zeta(3)\sum_{r_1=1}^{n}\frac{S_{00r_1}-a_{r_1}b_{r_1}c_{r_1}}{r_1}+$$
$$+\zeta(2)\sum_{\substack{r_1,r_2=1\\ r_1\neq r_2}}^{n}\left(\frac{2S_{0r_1r_1}-S_{00r_1}-a_{r_1}b_{r_1}c_{r_1}}{r_1^2}-\frac{S_{0r_1r_2}}{r_2-r_1}\left(\frac{1}{r_2}-\frac{1}{r_1}\right)+\frac{S_{r_1r_1r_2}}{r_1(r_2-r_1)}\right)+$$
$$+\sum_{r_1=1}^{n}\frac{S_{00r_1}-2S_{0r_1r_1}+a_{r_1}b_{r_1}c_{r_1}}{r_1^3}H_{r_1}+\sum_{r_1=1}^{n}\frac{a_{r_1}b_{r_1}c_{r_1}-S_{0r_1r_1}}{r_1^2}H^{(2)}_{r_1}+$$

$$+\sum_{r_1=1}^{n}\frac{a_{r_1}b_{r_1}c_{r_1}}{r_1}H_{r_1}^{(3)}+\sum_{\substack{r_1,r_2=1\\ r_1\neq r_2}}^{n}\frac{S_{0r_1r_2}}{r_2-r_1}\left(\frac{H_{r_2}}{r_2^2}-\frac{H_{r_1}}{r_1^2}\right)+$$

$$+\sum_{\substack{r_1,r_2=1\\ r_1\neq r_2}}^{n}\frac{S_{r_1r_1r_2}}{(r_2-r_1)r_1^2}H_{r_1}+\sum_{\substack{r_1,r_2=1\\ r_1\neq r_2}}^{n}\frac{S_{r_1r_1r_2}}{(r_2-r_1)r_1}H_{r_1}^{(2)}+$$

$$+\sum_{\substack{r_1,r_2=1\\ r_1\neq r_2}}^{n}\frac{S_{r_1r_1r_2}}{(r_2-r_1)^2}\left(\frac{H_{r_2}}{r_2}-\frac{H_{r_1}}{r_1}\right)+\sum_{\substack{r_1,r_2,r_3=1\\ r_1\neq r_2\neq r_3}}^{n}a_{r_1}b_{r_2}c_{r_3}\times$$

$$\times\left(\frac{H_{r_1}}{r_1(r_1-r_2)(r_1-r_3)}+\frac{H_{r_2}}{r_2(r_2-r_1)(r_2-r_3)}+\frac{H_{r_3}}{r_3(r_3-r_1)(r_3-r_2)}\right). \tag{22}$$

Для любого $s\geqslant 5$ на основании леммы 2 преобразуем слагаемые (13). Выполняя алгебраические вычисления, аналогичные приведенным выше, для $s\geqslant 5$ получаем

$$I_s=a_0b_0c_0\zeta(s)+\left(\sum_{r_1=1}^{n}\frac{S_{00r_1}}{r_1}\right)\zeta(s-1)+$$

$$+\left(\sum_{\substack{r_1,r_2=1\\ r_1\neq r_2}}^{n}\frac{S_{0r_1r_1}-S_{00r_1}}{r_1^2}-\frac{S_{0r_1r_2}}{r_2-r_1}\left(\frac{1}{r_2}-\frac{1}{r_1}\right)\right)\zeta(s-2)+$$

$$+\sum_{j=3}^{s-4}(-1)^{j-1}\left(\sum_{\substack{r_1,r_2,r_3=1\\ r_1\neq r_2\neq r_3}}^{n}\left(\frac{S_{00r_1}-(j-1)S_{0r_1r_1}+\dfrac{(j-2)(j-1)}{2}a_{r_1}b_{r_1}c_{r_1}}{r_1^j}+\right.\right.$$

$$+\frac{S_{r_1r_1r_2}}{r_2-r_1}\left(\frac{j-2}{r_1^{j-1}}+\frac{1}{r_2-r_1}\left(\frac{1}{r_2^{j-2}}-\frac{1}{r_1^{j-2}}\right)\right)-\frac{S_{0r_1r_2}}{r_2-r_1}\left(\frac{1}{r_2^{j-1}}-\frac{1}{r_1^{j-1}}\right)+$$

$$+a_{r_1}b_{r_2}c_{r_3}\left(\frac{1}{(r_3-r_2)(r_3-r_1)r_3^{j-2}}+\frac{1}{(r_2-r_3)(r_2-r_1)r_2^{j-2}}+\right.$$

$$\left.\left.\left.+\frac{1}{(r_1-r_3)(r_1-r_2)r_1^{j-2}}\right)\right)\right)\zeta(s-j)+$$

$$+(-1)^{s-3}\left(\sum_{\substack{r_1,r_2,r_3=1\\ r_1\neq r_2\neq r_3}}^{n}\left(\frac{a_{r_1}b_{r_1}c_{r_1}\left(1-\dfrac{(s-5)(s-4)}{2}\right)-S_{00r_1}+(s-4)S_{0r_1r_1}}{r_1^{s-3}}+\right.\right.$$

$$+\frac{S_{0r_1r_2}}{r_2-r_1}\left(\frac{1}{r_2^{s-4}}-\frac{1}{r_1^{s-4}}\right)-\frac{S_{r_1r_1r_2}}{r_2-r_1}\left(\frac{s-5}{r_1^{s-4}}+\frac{1}{r_2-r_1}\left(\frac{1}{r_2^{s-5}}-\frac{1}{r_1^{s-5}}\right)\right)-$$

$$-a_{r_1}b_{r_2}c_{r_3}\left(\frac{1}{(r_3-r_2)(r_3-r_1)r_3^{s-5}}+\frac{1}{(r_2-r_3)(r_2-r_1)r_2^{s-5}}+\right.$$

$$\left.\left.\left.+\frac{1}{(r_1-r_3)(r_1-r_2)r_1^{s-5}}\right)\right)\right)\zeta(3)+$$

$$+(-1)^{s-2}\left(\sum_{\substack{r_1,r_2,r_3=1\\ r_1\neq r_2\neq r_3}}^{n}\left(\frac{(s-2)S_{0r_1r_1}-S_{00r_1}-\dfrac{(s-2)(s-3)}{2}a_{r_1}b_{r_1}c_{r_1}}{r_1^{s-2}}+\right.\right.$$

$$+\frac{S_{0r_1r_2}}{r_2-r_1}\left(\frac{1}{r_2^{s-3}}-\frac{1}{r_1^{s-3}}\right)-\frac{S_{r_1r_1r_2}}{r_2-r_1}\left(\frac{s-3}{r_1^{s-3}}+\frac{1}{r_2-r_1}\left(\frac{1}{r_2^{s-4}}-\frac{1}{r_1^{s-4}}\right)\right)+$$
$$+a_{r_1}b_{r_2}c_{r_3}\left(\frac{1}{(r_3-r_2)(r_3-r_1)r_3^{s-4}}+\frac{1}{(r_2-r_3)(r_2-r_1)r_2^{s-4}}+\right.$$
$$\left.+\frac{1}{(r_1-r_3)(r_1-r_2)r_1^{s-4}}\right)\Bigg)\Bigg)\zeta(2)+$$
$$+(-1)^{s-2}\sum_{\substack{r_1,r_2,r_3=1\\ r_1\neq r_2\neq r_3}}^{n}\Bigg(\Bigg(\frac{-(s-2)S_{0r_1r_1}+S_{00r_1}+\dfrac{(s-2)(s-3)}{2}a_{r_1}b_{r_1}c_{r_1}}{r_1^{s-1}}+$$
$$+\frac{(s-3)S_{r_1r_1r_2}}{(r_2-r_1)r_1^{s-2}}\Bigg)H_{r_1}-\frac{S_{0r_1r_2}}{r_2-r_1}\left(\frac{H_{r_2}}{r_2^{s-2}}-\frac{H_{r_1}}{r_1^{s-2}}\right)+$$
$$+\frac{S_{r_1r_1r_2}}{(r_2-r_1)^2}\left(\frac{H_{r_2}}{r_2^{s-3}}-\frac{H_{r_1}}{r_1^{s-3}}\right)+a_{r_1}b_{r_2}c_{r_3}\left(\frac{H_{r_1}}{(r_1-r_2)(r_1-r_3)r_1^{s-3}}+\right.$$
$$\left.+\frac{H_{r_2}}{(r_2-r_1)(r_2-r_3)r_2^{s-3}}+\frac{H_{r_3}}{(r_3-r_2)(r_3-r_1)r_3^{s-3}}\right)\Bigg)+$$
$$+(-1)^{s-3}\sum_{\substack{r_1,r_2=1\\ r_1\neq r_2}}^{n}\left(\frac{S_{0r_1r_1}-(s-3)a_{r_1}b_{r_1}c_{r_1}}{r_1^{s-2}}-\frac{S_{r_1r_1r_2}}{(r_2-r_1)r_1^{s-3}}\right)H_{r_1}^{(2)}+$$
$$+(-1)^{s-4}\sum_{r_1=1}^{n}a_{r_1}b_{r_1}c_{r_1}\frac{H_{r_1}^{(3)}}{r_1^{s-3}}. \tag{23}$$

Формулы (20), (22), (23) можно представить несколько в более удобном для вычислений виде, группируя множители при одинаковых дзета-константах и заменяя суммы по несовпадающим индексам двойными и тройными суммами. Таким образом, мы доказали следующее утверждение.

Теорема 1. *Пусть $P_n(x)$, $Q_n(x)$ и $T_n(x)$ – три многочлена степени n, $n\geqslant 1$,*

$$P_n(x)=a_0+a_1x+\ldots+a_nx^n, \tag{24}$$
$$Q_n(x)=b_0+b_1x+\ldots+b_nx^n, \tag{25}$$
$$T_n(x)=c_0+c_1x+\ldots+c_nx^n, \tag{26}$$

$a_0,a_1,\ldots,a_n$, $b_0,b_1,\ldots,b_n$, $c_0,c_1,\ldots,c_n$ – произвольные числа. Пусть интеграл I_s, $s\geqslant 3$, определяется как

$$I_s=\int\limits_0^1\ldots\int\limits_0^1\frac{P_n(x_1)Q_n(x_2)T_n(x_3)}{1-x_1x_2x_3\ldots x_s}\,dx_1\,dx_2\,dx_3\ldots dx_s. \tag{27}$$

Тогда при $s\geqslant 5$ справедливы следующие соотношения:

$$\begin{aligned}
&I_3=A_{s-2,3}\zeta(3)-A_{s-2,2}\zeta(2)-A_{s-2},\\
&I_4=A_{s-3,4}\zeta(4)+A_{s-3,3}\zeta(3)-A_{s-3,2}\zeta(2)-A_{s-3},\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
&I_s=A_{1,s}\zeta(s)+A_{1,s-1}\zeta(s-1)+\ldots+A_{1,3}\zeta(3)-A_{1,2}\zeta(2)-A_1,
\end{aligned} \tag{28}$$

где при $s = 5, 6, 7, \ldots$

$$A_{s-2,3} = \sum_{r=1}^{n} a_r b_r c_r, \quad A_{s-2,2} = \sum_{r=1}^{n} \sum_{\ell=0}^{r-1} \frac{S_{rr\ell} - S_{\ell\ell r}}{r-\ell}, \tag{29}$$

$$A_{s-2} = \sum_{r=1}^{n} a_r b_r c_r H_r^{(3)} - \sum_{r=1}^{n} \sum_{\ell=0}^{r-1} (S_{rr\ell} - S_{\ell\ell r}) \left(\frac{H_r^{(2)} - H_\ell^{(2)}}{r-\ell} + \frac{H_r - H_\ell}{(r-\ell)^2} \right) +$$
$$+ \sum_{r=2}^{n} \sum_{\ell=1}^{r-1} \sum_{i=0}^{\ell-1} (S_{ir\ell} + S_{i\ell r}) \left(\frac{H_i}{(i-r)(i-\ell)} + \frac{H_r}{(r-i)(r-\ell)} + \frac{H_\ell}{(\ell-i)(\ell-r)} \right), \tag{30}$$

$$A_{s-3,4} = a_0 b_0 c_0, \quad A_{s-3,3} = \sum_{r=1}^{n} \frac{S_{00r} - a_r b_r c_r}{r}, \tag{31}$$

$$A_{s-3,2} = \sum_{r=1}^{n} \frac{a_r b_r c_r + S_{00r} - 2S_{0rr}}{r^2} +$$
$$+ \sum_{r=2}^{n} \sum_{\ell=1}^{r-1} \left(\frac{S_{0r\ell} + S_{0\ell r}}{r-\ell} \left(\frac{1}{r} - \frac{1}{\ell} \right) - \frac{1}{r-\ell} \left(\frac{S_{rr\ell}}{r} - \frac{S_{\ell\ell r}}{\ell} \right) \right), \tag{32}$$

$$A_{s-3} = \sum_{r=1}^{n} \left(\frac{2S_{0rr} - a_r b_r c_r - S_{00r}}{r^3} H_r + \frac{S_{0rr} - a_r b_r c_r}{r^2} H_r^{(2)} - \frac{a_r b_r c_r}{r} H_r^{(3)} \right) -$$
$$- \sum_{r=2}^{n} \sum_{\ell=1}^{r-1} \left(\frac{S_{rr\ell} - S_{\ell\ell r}}{(r-\ell)^2} \left(\frac{H_r}{r} - \frac{H_\ell}{\ell} \right) + \frac{1}{r-\ell} \left(S_{rr\ell} \frac{H_r}{r^2} - S_{\ell\ell r} \frac{H_\ell}{\ell^2} \right) + \right.$$
$$\left. + \frac{1}{r-\ell} \left(S_{rr\ell} \frac{H_r^{(2)}}{r} - S_{\ell\ell r} \frac{H_\ell^{(2)}}{\ell} \right) \right) - \sum_{r=2}^{n} \sum_{\ell=1}^{r-1} \sum_{i=0}^{\ell-1} (S_{ir\ell} + S_{i\ell r}) \left(\frac{H_i}{i(i-r)(i-\ell)} + \right.$$
$$\left. + \frac{H_r}{r(r-i)(r-\ell)} + \frac{H_\ell}{\ell(\ell-i)(\ell-r)} \right), \tag{33}$$

$$A_{1,s} = a_0 b_0 c_0, \quad A_{1,s-1} = \sum_{r=1}^{n} \frac{S_{00r}}{r}, \tag{34}$$

$$A_{1,s-2} = \sum_{r=1}^{n} \frac{S_{0rr} - S_{00r}}{r^2} - \sum_{r=2}^{n} \sum_{\ell=1}^{r-1} \frac{S_{0\ell r} + S_{0r\ell}}{r-\ell} \left(\frac{1}{r} - \frac{1}{\ell} \right), \tag{35}$$

а при $j = 3, 4, \ldots, s-2$

$$A_{1,s-j} = (-1)^{j-1} \left(\sum_{r=1}^{n} \frac{S_{00r} - (j-1)S_{0rr} + \frac{(j-2)(j-1)}{2} a_r b_r c_r}{r^j} - \right.$$
$$- \sum_{r=2}^{n} \sum_{\ell=1}^{r-1} \left(\frac{S_{\ell\ell r} - S_{rr\ell}}{(r-\ell)^2} \left(\frac{1}{r^{j-2}} - \frac{1}{\ell^{j-2}} \right) + \frac{j-2}{r-\ell} \left(\frac{S_{\ell\ell r}}{\ell^{j-1}} - \frac{S_{rr\ell}}{r^{j-1}} \right) \right) -$$
$$- \sum_{r=2}^{n} \sum_{\ell=1}^{r-1} \sum_{i=0}^{\ell-1} (S_{ir\ell} + S_{i\ell r}) \left(\frac{H_i}{i^{j-2}(i-r)(i-\ell)} + \right.$$
$$\left. \left. + \frac{H_r}{r^{j-2}(r-i)(r-\ell)} + \frac{H_\ell}{\ell^{j-2}(\ell-i)(\ell-r)} \right) \right), \tag{36}$$

$$A_1 = (-1)^{s-1}\Bigg(\sum_{r=1}^{n}\Bigg(\frac{-S_{00r}+(s-2)S_{0rr}-\dfrac{(s-2)(s-3)}{2}a_rb_rc_r}{r^{s-1}}H_r+$$
$$+\frac{(s-3)a_rb_rc_r-S_{0rr}}{r^{s-2}}H_r^{(2)}+\frac{a_rb_rc_r}{r^{s-3}}H_r^{(3)}\Bigg)-$$
$$-\Bigg(\sum_{r=2}^{n}\sum_{\ell=1}^{r-1}\frac{1}{r-\ell}\Bigg(\frac{H_\ell^{(2)}S_{\ell\ell r}}{\ell^{s-3}}-\frac{H_r^{(2)}S_{rr\ell}}{r^{s-3}}\Bigg)+$$
$$+\frac{s-3}{r-\ell}\left(\frac{H_\ell S_{\ell\ell r}}{\ell^{s-2}}-\frac{H_rS_{rr\ell}}{r^{s-2}}\right)+\frac{S_{\ell\ell r}-S_{rr\ell}}{(r-\ell)^2}\left(\frac{H_r}{r^{s-3}}-\frac{H_\ell}{\ell^{s-3}}\right)\Bigg)-$$
$$-\sum_{r=2}^{n}\sum_{\ell=1}^{r-1}\sum_{i=0}^{\ell-1}(S_{ir\ell}+S_{i\ell r})\Bigg(\frac{H_i}{i^{s-3}(i-r)(i-\ell)}+$$
$$+\frac{H_r}{r^{s-3}(r-i)(r-\ell)}+\frac{H_\ell}{\ell^{s-3}(\ell-i)(\ell-r)}\Bigg)\Bigg), \tag{37}$$

где для $0\leqslant\mu\leqslant n$, $0\leqslant\nu\leqslant n$, $0\leqslant\lambda\leqslant n$

$$S_{\mu\nu\lambda}=a_\mu b_\nu c_\lambda+b_\mu c_\nu a_\lambda+c_\mu a_\nu b_\lambda, \tag{38}$$

а $H_n^{(m)}$ – *гармонические числа, определяемые соотношениями* (2).

§ 4. Новая формула для вычисления дзета-констант

Отметим, что формулы теоремы 1 имеют самый общий вид, ведь коэффициенты $a_0, a_1, \ldots, a_n$, $b_0, b_1, \ldots, b_n$, $c_0, c_1, \ldots, c_n$ – произвольные числа.

Чтобы эффективно вычислить какую-либо дзета-константу $\zeta(s)$, $s\geqslant 3$, на основе формул (28)–(38), нужно подобрать такие многочлены (24)–(26), чтобы интеграл (27) был мал, например порядка 2^{-Cn}, $C=\mathrm{const}\geqslant 1$, и в то же время вычисление коэффициентов при дзета-константах и свободных членов в (29)–(37) не было бы слишком затратным.

Допустим, что выбранные нами многочлены (24)–(26) таковы, что I_s мало,

$$I_s=I_s(n)=\theta_s(n)=\theta_s. \tag{39}$$

Тогда получаем систему

$$A_{1,s}\zeta(s)+A_{1,s-1}\zeta(s-1)+\ldots+A_{1,3}\zeta(3)=A_{1,2}\zeta(2)+A_1+\theta_s,$$
$$A_{2,s-1}\zeta(s-1)+\ldots+A_{2,3}\zeta(3)=A_{2,2}\zeta(2)+A_2+\theta_{s-1},$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$A_{s-2,3}\zeta(3)=A_{s-2,2}\zeta(2)+A_{s-2}+\theta_3,$$

где коэффициенты при дзета-константах определяются формулами (29)–(38) для $s-2$ значений I_r, $r=3,4,\ldots,s$. Значение $\zeta(s)$ выражается отношением определителей:

$$\Delta=A_{1,s}A_{2,s-1}\ldots A_{s-2,3},\quad \zeta(s)=\frac{\Delta_s}{\Delta},$$

при этом

$$\Delta_s = \begin{vmatrix} A_{1,2}\zeta(2) + A_1 + \theta_s & A_{1,s-1} & \dots & A_{1,3} \\ A_{2,2}\zeta(2) + A_2 + \theta_{s-1} & A_{2,s-1} & \dots & A_{2,3} \\ \dots & \dots & \dots & \dots \\ A_{s-2,2}\zeta(2) + A_{s-2} + \theta_3 & 0 & \dots & A_{s-2,3} \end{vmatrix} = \Delta_s^{(1)} + \Delta_s^{(2)} + \Delta_s^{(3)},$$

где

$$\Delta_s^{(1)} = \begin{vmatrix} A_{1,2}\zeta(2) & A_{1,s-1} & \dots & A_{1,3} \\ A_{2,2}\zeta(2) & A_{2,s-1} & \dots & A_{2,3} \\ \dots & \dots & \dots & \dots \\ A_{s-2,2}\zeta(2) & 0 & \dots & A_{s-2,3} \end{vmatrix},$$

$$\Delta_s^{(2)} = \begin{vmatrix} A_1 & A_{1,s-1} & \dots & A_{1,3} \\ A_2 & A_{2,s-1} & \dots & A_{2,3} \\ \dots & \dots & \dots & \dots \\ A_{s-2} & 0 & \dots & A_{s-2,3} \end{vmatrix}, \quad \Delta_s^{(3)} = \begin{vmatrix} \theta_s & A_{1,s-1} & \dots & A_{1,3} \\ \theta_{s-1} & A_{2,s-1} & \dots & A_{2,3} \\ \dots & \dots & \dots & \dots \\ \theta_3 & 0 & \dots & A_{s-2,3} \end{vmatrix}.$$

Таким образом,

$$\zeta(s) = \frac{\Delta_s}{\Delta} = \frac{1}{\Delta} \sum_{\nu=1}^{s-2} \left(A_{\nu,2}\zeta(2) + A_\nu + \theta_{s+1-\nu}\right) \Delta_{\nu s},$$

где $\Delta_{\nu s}$ – соответствующие алгебраические дополнения, т.е.

$$\zeta(s) = \frac{\zeta(2)}{\Delta} \sum_{\nu=1}^{s-2} A_{\nu,2}\Delta_{\nu s} + \frac{1}{\Delta} \sum_{\nu=1}^{s-2} A_\nu \Delta_{\nu s} + \frac{1}{\Delta} \sum_{\nu=1}^{s-2} \theta_{s+1-\nu}\Delta_{\nu s}. \tag{40}$$

§ 5. Аппроксимация

Получим из (39), (40) оценку значений θ_s, $s \geqslant 3$, для специально подобранных многочленов. Как и в [7], в качестве основных, обеспечивающих точность приближения, выберем многочлены (1): в качестве $P_n(x)$ – смещенный многочлен Лежандра, в качестве $Q_n(x)$ – биномиальный многочлен. Третий многочлен $T_n(x)$ будет записан в каноническом виде, поскольку может использоваться для понижения сложности вычислений. Имеет место

Л е м м а 3. *Пусть*

$$P_n(x) = \frac{1}{n!} \left(\frac{d}{dx}\right)^n (x^n(1-x)^n) = a_0 + a_1 x + \ldots + a_n x^n, \tag{41}$$

$$a_r = \frac{(-1)^r (n+r)!}{(r!)^2 (n-r)!}, \tag{42}$$

$$Q_n(x) = (1-x)^n = b_0 + b_1 x + \ldots + b_n x^n, \tag{43}$$

$$b_r = (-1)^r \frac{n!}{r!\,(n-r)!}. \tag{44}$$

Пусть

$$T_n(x) = c_0 + c_1 x + \ldots + c_n x^n, \quad c^* = \max_{0 \leqslant j \leqslant n} |c_j|. \tag{45}$$

Тогда при любом $s \geqslant 3$ для интеграла

$$I_s = I_s(n) = \int_0^1 \dots \int_0^1 \frac{P_n(x_1)Q_n(x_2)T_n(x_3)}{1 - x_1x_2x_3 \dots x_s}\, dx_1\, dx_2\, dx_3 \dots dx_s, \quad s \geqslant 3,$$

справедлива оценка

$$|I_s| \leqslant \frac{c^*}{2^{2n}}. \tag{46}$$

Доказательство. Из (43), (44) имеем для $k \geqslant 1$

$$\int_0^1 x^k Q_n(x)\, dx = \int_0^1 x^k (1-x)^n\, dx = B(k+1, n+1) = \frac{k!\, n!}{(k+n+1)!},$$

где $B(x, y)$ – бета функция Эйлера (см., например, [18, 19]), $B(x, y) = \dfrac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$, где $\Gamma(x+1) = x\Gamma(x)$ – гамма-функция Эйлера. В то же время из (41), (42) интегрированием по частям находим

$$\int_0^1 x^k P_n(x)\, dx = \frac{1}{n!} \int_0^1 x^k \frac{d}{dx} \left(\frac{d^{n-1}}{dx^{n-1}} \left(x^n (1-x)^n\right) \right) dx =$$
$$= (-1)^n \frac{k(k-1)\dots(k-n+1)}{n!} B(k+1, n+1) = (-1)^n \frac{k(k-1)\dots(k-n+1)}{(k+1)\dots(k+n)(k+n+1)}$$

при $k \geqslant n$. Если $k < n$, то

$$\int_0^1 x^k P_n(x)\, dx = 0,$$

так как при интегрировании по частям все время участвует член вида

$$x^\nu x^n (1-x)^n \big|_0^1 = 0.$$

Таким образом, при выбранных $P_n(x)$ и $Q_n(x)$ имеем

$$\int_0^1 x^k Q_n(x)\, dx = B(k+1, n+1),$$
$$\int_0^1 x^k P_n(x)\, dx = \begin{cases} 0, & \text{если } k \leqslant n-1, \\ (-1)^n \dfrac{k(k-1)\dots(k-n+1)}{n!} B(k+1, n+1), & \text{если } k \geqslant n. \end{cases}$$

В то же время из (45)

$$\int_0^1 x^k T_n(x)\, dx = \sum_{i=0}^n \frac{c_i}{k+1+i} = T(k, n). \tag{47}$$

Следовательно,

$$I_s = I_s(n) = \int_0^1 \dots \int_0^1 \frac{P_n(x_1)Q_n(x_2)T_n(x_3)}{1 - x_1x_2x_3\dots x_s}\,dx_1\,dx_2\,dx_3\dots dx_s =$$

$$= \sum_{k=0}^{\infty} \int_0^1 \dots \int_0^1 (x_1x_2\dots x_s)^k P_n(x_1)Q_n(x_2)T_n(x_3)\,dx_1\,dx_2\,dx_3\dots dx_s =$$

$$= (-1)^n \sum_{k=n}^{\infty} \frac{k(k-1)\dots(k-n+1)}{n!}\,\frac{B^2(k+1,n+1)T(k,n)}{(k+1)^{s-3}}. \tag{48}$$

Из (47), (48) находим:

$$|I_s| \leqslant c^* \sum_{j=0}^{\infty} \frac{(n+j)!}{j!n!}\,\frac{B^2(n+1+j,n+1)}{(n+1+j)^{s-3}}$$

$$\leqslant c^* \sum_{j=0}^{\infty} \frac{B(n+1+j,n+1)}{(2n+j+1)(n+1+j)^{s-3}}\,\frac{\Gamma^2(n+j+1)}{\Gamma(2n+j+1)\Gamma(j+1)}. \tag{49}$$

Поскольку (см., например, [18])

$$\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\gamma)\Gamma(\beta-\gamma)} = \prod_{\nu=0}^{\infty} \left(1 + \frac{\gamma}{\alpha+\nu}\right)\left(1 - \frac{\gamma}{\beta+\nu}\right),$$

то

$$\frac{\Gamma^2(n+j+1)}{\Gamma(2n+j+1)\Gamma(j+1)} = \prod_{\nu=0}^{\infty} \left(1 - \left(\frac{n}{n+1+j+\nu}\right)^2\right) < 1$$

для любого $n \geqslant 1$. Следовательно, в силу (49) для любого $s \geqslant 3$ выполняется оценка

$$|I_s| \leqslant c^* \sum_{j=0}^{\infty} \frac{B(n+1+j,n+1)}{(2n+j+1)(n+1+j)^{s-3}} \leqslant \frac{c^*}{2} \sum_{j=0}^{\infty} B(n+1+j,n+1).$$

Принимая во внимание (см. [19]), что $\sum\limits_{j=0}^{\infty} B(x,y+j) = B(x-1,y)$, и учитывая соответствующие оценки для бета-функций (см., например, [18], а также [7]), отсюда находим

$$|I_s| \leqslant \frac{c^*}{2}B(n,n+1) \leqslant \frac{c^*}{4}B(n,n) \leqslant \frac{c^*}{4}2^{1-2n}B\Big(\frac{1}{2},n\Big) \leqslant c^*2^{-2n},$$

т.е. оценка (46) справедлива. ▲

§ 6. Заключение

На основе новых формул для дзета-констант построен метод их аппроксимации и вычисления, с контролем скорости аппроксимации посредством выбора коэффициентов трех многочленов.

Возникает естественный вопрос: возможно ли, по крайней мере гипотетически, на основе этих новых формул построить быстрый алгоритм? Если бы три многочлена $P_n(x)$, $Q_n(x)$ и $T_n(x)$ могли бы быть выбраны так, чтобы последняя сумма в (40)

была мала, а остальные две суммы могли бы быть вычислены с низкой оценкой сложности, то найденные формулы могли бы стать основой для такого алгоритма. Отметим, что существует множество быстрых алгоритмов вычисления константы $\zeta(2) = \pi^2/6$ (см., например, [5]).

С другой стороны, могут появиться новые возможности, если обобщить представленный метод, построив метод вычисления дзета-констант с участием $s = 2k+1$ многочленов, $k = 2, 3 \ldots$, учитывая, что коэффициенты, подобные коэффициентам из (29)–(38), относительно легко оценить.

В [20] было доказано следующее утверждение.

Лемма 4. *Для любых целых s и m, $s \geqslant 1$, $m \geqslant 1$, и любых $r \geqslant 1$, $k \geqslant 0$ справедливо тождество*

$$\frac{1}{(r+k+1)^m (k+1)^s} = \\ = \sum_{j=1}^{s} (-1)^{j-1} \frac{\binom{m+j-2}{m-1}}{r^{j+m-1}(k+1)^{s+1-j}} + (-1)^s \sum_{j=1}^{m} \frac{\binom{s+m-j-1}{m-j}}{r^{s+m-j}(r+k+1)^j}. \tag{50}$$

Пользуясь (50), можно построить обобщенный метод с участием s многочленов. Однако из-за того, что многочлены записываются в канонической форме с произвольными коэффициентами, выведенные формулы будут настолько громоздкими, что описание такого метода представляется затруднительным. По-видимому, дальнейшая работа по созданию нового метода вычисления дзета-констант на основе этого подхода должна заключаться в определении ограничений на коэффициенты, которые сделают метод более эффективным, а в долгосрочной перспективе – быстрым.

СПИСОК ЛИТЕРАТУРЫ

1. *Карацуба Е.А.* Быстрое вычисление дзета-функции Римана $\zeta(s)$ при целых значениях аргумента s // Пробл. передачи информ. 1995. Т. 31. № 4. С. 69–80. http://mi.mathnet.ru/ppi294
2. *Borwein J.M., Bradley D.M., Crandall R.E.* Computational Strategies for the Riemann Zeta Function // J. Comput. Appl. Math. 2000. V. 121. № 1–2. 247–296. https://doi.org/10.1016/S0377-0427(00)00336-8
3. *Скороходов С.Л.* Аппроксимации Паде и численный анализ дзета-функции Римана // Ж. вычисл. матем. и матем. физ. 2003. V. 43. № 9. С. 1330–1352. http://mi.mathnet.ru/zvmmf961
4. *Зудилин В.В.* О биномиальных суммах, связанных с рациональными приближениями к $\zeta(4)$ // Матем. заметки. 2004. Т. 75. № 4. С. 637–640. https://doi.org/10.4213/mzm555
5. *Карацуба Е.А.* Об одном методе быстрого приближения дзета-констант рациональными дробями // Пробл. передачи информ. 2014. Т. 50. № 2. С. 77–95. http://mi.mathnet.ru/ppi2140
6. *Матиясевич Ю.В.* Дзета-функция Римана и конечные ряды Дирихле // Алгебра и анализ. 2015. V. 27. № 6. С. 174–198. http://mi.mathnet.ru/aa1472
7. *Карацуба Е.А.* Об одном методе построения семейства аппроксимаций дзета-констант рациональными дробями // Пробл. передачи информ. 2015. V. 51. № 4. С. 78–91. http://mi.mathnet.ru/ppi2189
8. *Hermite C.* Sur la fonction exponentielle // C. R. Acad. Sci. Paris. 1873. V. 77. P. 18–24.
9. *Beukers F.* A Note on the Irrationality of $\zeta(2)$ and $\zeta(3)$ // Bull. London Math. Soc. 1979. V. 11. № 3. P. 268–272. https://doi.org/10.1112/blms/11.3.268
10. *Beukers F.* Legendre Polynomials in Irrationality Proofs // Bull. Austral. Math. Soc. 1980. V. 22. № 3. P. 431–438. https://doi.org/10.1017/S0004972700006742

11. *Dvornicich R., Viola C.* Some Remarks on Beukers' Integrals // Number Theory, Vol. II (Budapest, 1987). Colloq. Math. Soc. János Bolyai. V. 51. Amsterdam: North-Holland, 1990. P. 637–657.
12. *Hata M.* A Note on Beukers' Integral // J. Austral. Math. Soc. Ser. A. 1995. V. 58. № 2. P. 143–153. https://doi.org/10.1017/S1446788700038192
13. *Hessami Pilehrood K., Hessami Pilehrood T., Tauraso R.* Congruences Concerning Jacobi Polynomials and Apéry-like Formulae // Int. J. Number Theory. 2012. V. 8. № 7. P. 1789–1811. https://doi.org/10.1142/S1793042112501035
14. *Зудилин В.В.* Об иррациональности значений дзета-функции Римана // Изв. РАН. Сер. матем. 2002. Т. 66. № 3. С. 49–102. https://doi.org/10.4213/im387
15. *Chu W., De Donno L.* Hypergeometric Series and Harmonic Number Identities // Adv. in Appl. Math. 2005. V. 34. № 1. P. 123–137. https://doi.org/10.1016/j.aam.2004.05.003
16. *Spieß J.* Some Identities Involving Harmonic Numbers // Math. Comp. 1990. V. 55. № 192. P. 839–863. https://doi.org/10.2307/2008451
17. *Wang W., Jia C.* Harmonic Number Identities via the Newton–Andrews Method // Ramanujan J. 2014. V. 35. № 2. P. 263–285. https://doi.org/10.1007/s11139-013-9511-1
18. *Бейтмен Г., Эрдейи А.* Высшие трансцендентные функции. Т. 1. Гипергеометрическая функция. Функции Лежандра. М.: Наука, 1965.
19. *Уиттекер Э.Т., Ватсон Дж.Н.* Курс современного анализа. Ч. 2. Трансцендентные функции. М.: Физматлит, 1963.
20. *Карацуба Е.А.* Об одном тождестве с биномиальными коэффициентами // Матем. заметки. 2019. V. 105. № 1. С. 149–152. https://doi.org/10.4213/mzm11883

Карацуба Екатерина Анатольевна
Вычислительный центр им. А.А. Дородницына
Федерального исследовательского центра
"Информатика и управление" РАН, Москва
ekar@ccas.ru, ekaratsuba@gmail.com

Поступила в редакцию
06.11.2020
После доработки
03.03.2021
Принята к публикации
13.05.2021

ПРОБЛЕМЫ ПЕРЕДАЧИ ИНФОРМАЦИИ

Том 57 **2021** **Вып. 3**

УДК 621.391 : 517.938 : 519.722

© 2021 г. **Г.Д. Дворкин**

ГЕОМЕТРИЧЕСКАЯ ИНТЕРПРЕТАЦИЯ ЭНТРОПИИ: НОВЫЕ РЕЗУЛЬТАТЫ

Рассматривается связь метрической энтропии с локальной скоростью деформации границ (ЛСДГ) в символическом случае. Доказывается равенство ЛСДГ как предела в среднем и энтропии для систем, содержащих существенно синхронизованный подсдвиг полной меры. Получен пример данной связи и при отсутствии такого подсдвига. Впервые показано, что если понимать ЛСДГ как предел почти всюду, то равенства между этой величиной и энтропией может не быть.

Ключевые слова: метрическая энтропия, локальная скорость деформации границ, символическая система, синхронизованная система, инвариантная эргодическая мера, правильная скобочная последовательность, система Дика.

DOI: 10.31857/S0555292321030062

§ 1. Введение

Будем рассматривать энтропию Колмогорова – Синая как предел некоторой величины (скорости деформации границ), имеющей простую геометрическую интерпретацию.

Такой подход был предложен физиком Г.М. Заславским [1], который в 1984 г. выдвинул гипотезу (точнее, сформулировал утверждение), что объем границы множества в фазовом пространстве растет экспоненциально по времени с показателем, равным метрической энтропии. Первый математический результат в этом направлении был получен Б.М. Гуревичем [2] для символических марковских систем и совместно с С.А. Комечем [3, 4] обобщен на существенно более широкий класс символических систем, а также на некоторые гладкие (системы Аносова). В этой статье содержится ряд новых результатов в данной области.

§ 2. Определения, обозначения и обзор основных результатов

Пусть A – конечное множество, и пусть $A^{\mathbb{Z}}$ – пространство бесконечных двусторонних последовательностей символов из A.

Множество A будем называть алфавитом, а любую конечную последовательность символов (букв) из A – (конечным) словом этого алфавита. Количество букв в слове w называют его длиной и обозначают $|w|$. Слово нулевой длины называют пустым.

Элементы пространства $A^{\mathbb{Z}}$ будем называть бесконечными словами, при этом в данной терминологии бесконечные слова словами не являются.

Определение 1. Назовем *полным сдвигом* пару $(A^{\mathbb{Z}}, T)$, где A – конечное множество, а T – сдвиг на шаг вправо в пространстве последовательностей $A^{\mathbb{Z}}$.

Для $x \in A^{\mathbb{Z}}$ и целых a и b, таких что $a \leqslant b$, обозначим слово $(x(a), x(a+1), \ldots, x(b))$ через $x_{[a;b]}$. Если слово w равно $x_{[a;b]}$ для некоторых a и b, то будем говорить, что w является *подсловом* бесконечного слова x (аналогично определяется подслово конечного слова). Вместо $x_{[a;a]}$ будем писать $x_{[a]}$.

Введем метрику d на $A^{\mathbb{Z}}$: примем $d(x,x) = 0$, а для несовпадающих точек x и y положим $d(x,y) = (1/2)^m$, где $m = m(x,y)$ – целое неотрицательное число, равное нулю, если $x_{[0]} \neq y_{[0]}$, и удовлетворяющее условиям $x_{[-(m-1);m-1]} = y_{[-(m-1);m-1]}$, $x_{[-m;m]} \neq y_{[-m;m]}$ в противном случае. Эта метрика порождает топологию, которая далее считается фиксированной.

Определение 2. Будем называть *символической системой* пару (X,T), где X – замкнутое T-инвариантное подмножество $A^{\mathbb{Z}}$ с индуцированной топологией, а также первый элемент этой пары, считая сдвиг и топологию заданными по умолчанию.

Пусть далее в этом параграфе X – символическая система.

Множество всех подслов всех $x \in X$ называется *языком* символической системы X и обозначается $W(X)$. *Префикс* слова – любое его подслово, начинающееся с начала этого слова. *Суффикс* слова – любое его подслово, заканчивающееся в конце этого слова. Через $\ldots aaaaaaaaaa \ldots$ обозначается бесконечное слово, состоящее из букв "a", через a^ℓ – конкатенация ℓ букв или слов a.

Определение 3. Множество

$$\{w\}_c^X := \big\{x \in X : \, x_{[c;c+|w|-1]} = w\big\}$$

называется *цилиндром* в X с *основанием* $w \in W(X)$. Всюду далее

$$\{w\}_c := \{w\}_c^X, \quad \big\{x_{[a;b]}\big\} := \big\{x_{[a;b]}\big\}_a.$$

Рассмотрим T-инвариантную борелевскую вероятностную меру μ на X. Под мерой $\mu(w)$ слова $w \in W(X)$ понимаем меру цилиндра $\{w\}_0$ (из T-инвариантности меры очевидно, что $\mu(\{w\}_a) = \mu(\{w\}_0)$ для любого целого a).

Введем следующие обозначения:

$k(\varepsilon) := \max\{\ell \in \mathbb{Z} : \, (1/2)^{(\ell+1)} \leqslant \varepsilon\}$ (часто будем писать просто k);

$O_\varepsilon(x)$ – открытая ε-окрестность точки x в рассматриваемом пространстве;

$O_\varepsilon(C)$ – открытая ε-окрестность множества C в рассматриваемом пространстве.

Определение 4. *Метрическая энтропия* (энтропия Колмогорова – Синая) символической динамической системы (X,T) с инвариантной мерой μ может быть определена как средняя условная энтропия настоящего при условии прошлого (общее определение для произвольной динамической системы см. в [5, 6]):

$$h_\mu(X,T) = \lim_{n\to\infty} H_\mu(x_0 \,|\, x_{-1}, x_{-2}, \ldots, x_{-n}).$$

Определение 5. Для $x \in X$ определим *локальную скорость деформации границ* (ЛСДГ):

$$P_{X,T,\mu}(x) := \lim_{\varepsilon\to 0} \frac{1}{n(\varepsilon)} \log\left(\frac{\mu\big(O_\varepsilon\big(T^{n(\varepsilon)}(O_\varepsilon(x))\big)\big)}{\mu(O_\varepsilon(x))}\right),$$

где $n(\varepsilon)$ – положительная целочисленная функция со свойствами

$$n(\varepsilon) \to \infty \quad \text{и} \quad n(\varepsilon) = o(|\log(\varepsilon)|) = o(k(\varepsilon)) \quad \text{при } \varepsilon \to 0$$

(часто будем писать просто n). Такие функции n будем называть *медленными*. Допредельное выражение будем обозначать через $P^\varepsilon_{X,T,\mu}(x)$ и называть *ε-ЛСДГ*.

Определение 6. Система X *транзитивна*, если в ней есть точка со всюду плотной орбитой.

Определение 7. Символическая система X *синхронизована*, если она транзитивна и существует $w \in W(X)$, такое что для любых $u, v \in W(X)$, если $uw, wv \in \in W(X)$, то и $uwv \in W(X)$. Такое слово w называется *волшебным*.

Определение 8. *Полная правильная скобочная последовательность* (расстановка) – конечная последовательность скобок, приводящаяся к пустому слову путем сокращения стоящих подряд открывающей и закрывающей скобок одного вида. Например, в случае двух типов скобок сокращать можно только "( )" и "[ ]". Здесь и далее тип скобки – открывающая или закрывающая, вид – круглая, квадратная и т.д.

Определение 9. *Правильная скобочная последовательность* (расстановка) – конечная последовательность скобок, являющаяся подсловом некоторой полной правильной скобочной расстановки.

Определение 10. *m-язык Дика* – язык, состоящий из правильных скобочных последовательностей скобок m видов. Будем рассматривать только 2-язык и называть его просто языком Дика (но аналогичные рассуждения можно провести для любого m); его алфавит – "(,), [,]".

Определение 11. *Сдвиг Дика* – символическая система, языком которой является язык Дика.

Эта система транзитивна, но несинхронизована (см. [7]).

Пусть w – слово языка Дика. Скобка, сокращающаяся вместе с данной при описанном выше процессе сокращения, называется *парной* к ней в этом слове. Скобка, для которой нет парной в w, называется *непарной* в w. Положим

- $n_1 = n_1(w)$ – количество открывающих скобок в слове w, для которых в этом слове есть парная закрывающая;
- $n_2 = n_2(w)$ – количество непарных скобок в w.

Очевидно, $|w| = 2n_1 + n_2$.

Пусть $M(X)$ – некоторый подкласс класса борелевских вероятностных инвариантных эргодических мер на (X, T) (весь этот класс будем обозначать $M_0(X)$). Во всех случаях, когда из контекста будет ясно, что такое X, будем писать просто M и M_0.

Рассмотрим следующие утверждения.

Точечная гипотеза (ТГ). $P_{X,T,\mu}(x) = h(\mu, X, T)$ *для всех* $\mu \in M(X)$, *любой медленной функции* $n(\varepsilon)$, *любого* $x \in X$.

Обобщенная точечная гипотеза (ОТГ). $P_{X,T,\mu}(x) = h(\mu, X, T)$ *для всех* $\mu \in M(X)$, *любой медленной функции* $n(\varepsilon)$, μ*-почти любого* $x \in X$.

Основная гипотеза (ОГ). $P_{X,T,\mu}(x) = h(\mu, X, T)$ *для всех* $\mu \in M(X)$, *любой медленной функции* $n(\varepsilon)$, *где выражение в левой части понимается как предел в* $L_1(X, \mu)$.

ТГ очевидным образом влечет ОТГ и ОГ, при этом ОТГ и ОГ, вообще говоря, друг из друга не вытекают и не влекут ТГ.

Точечная и обобщенная точечная гипотезы верны для многих важных систем (ТГ, например, выполняется для автоморфизмов тора с мерой Лебега, а утверждения, близкие к ОТГ, верны для существенно более общих диффеоморфизмов Аносова с мерой Синая – Рюэлля – Боуэна; подробнее см. в [4]), но не доказаны и, как будет замечено ниже, не верны даже для класса бернуллиевских мер на полном сдвиге (контрпримером может служить любая несимметричная бернуллиевская мера при достаточно медленном росте функции $n(\varepsilon)$). Мы покажем неверность этих

гипотез для класса M_0 на сдвиге Дика методом, который применим и в бернуллиевском случае.

Основная гипотеза, в отличие от двух других, подтверждается для многих символических систем. Центральный результат в этом направлении получен Комечем в [3]. Будем называть этот результат основной теоремой (ОТ).

Основная теорема (Комеч). *Пусть $M(X)$ – класс всех мер из $M_0(X)$ на синхронизованной символической системе X, для которых мера хотя бы одного волшебного в X слова положительна. Тогда для $M(X)$ верна основная гипотеза.*

В настоящей статье получены следующие результаты:

1) Основная теорема обобщена на системы с синхронизованным подсдвигом полной меры (теорема 1). В частности, показана верность основной гипотезы для мер с синхронизованным носителем (следствие 2).
2) Построен пример меры, не отвечающей условиям теоремы 1, для которой основная гипотеза все равно верна (теорема 3).
3) Для меры из пункта 2) опровергнуты точечная и обобщенная точечная гипотезы (утверждение 1 и теорема 2).
4) Замечена общая несостоятельность ТГ и ОТГ в символическом случае (замечание 4).

Ввиду того, что ОГ не влечет ТГ и ОТГ, результаты 1)–4) не противоречат друг другу.

§ 3. Обобщение основной теоремы

Сначала приведем мотивационный пример.

Пример 1. Пусть $X \subset \{a, b, c\}^{\mathbb{Z}}$ определяется тем, что в словах $W(X)$ между двумя "a" запрещено разное количество "b" и "c", а мера сосредоточена на $L = \{b, c\}^{\mathbb{Z}}$.

Множество волшебных слов этой системы – это слова, содержащие "a". Значит, ОТ к этому примеру не применима. Тем не менее мера сосредоточена на простом синхронизованном множестве – полном сдвиге, и естественно предположить, что ОГ верна и для этого случая.

Теорема 1 (обобщенная основная теорема (ООТ)). *Пусть символическая динамическая система (X, T) и борелевская T-инвариантная эргодическая вероятностная мера μ на X удовлетворяют следующему условию: в X имеется синхронизованный подсдвиг L (T-инвариантное замкнутое подмножество) полной меры, обладающий хотя бы одним волшебным словом положительной меры. Тогда для X и μ верна основная гипотеза.*

Доказательство. Пусть μ и L – мера и подсдвиг из условия, μ_L – ограничение μ на L. Тогда, применяя основную теорему к (L, μ_L), получаем сходимость в среднем:

$$\lim_{\varepsilon \to 0} P^{\varepsilon}_{L,T,\mu_L}(x) = h(\mu_L, L, T). \tag{1}$$

Энтропия – метрический инвариант, поэтому

$$h(\mu_L, L, T) = h(\mu, X, T). \tag{2}$$

Основная сложность ситуации состоит в том, что метрическая инвариантность скорости деформации границ в общем случае не доказана. Тем не менее для рассматриваемого частного случая L проблема решается методом, близким к доказательству основной теоремы.

Заметим, что для любого $x \in L$ множество

$$O_1 := O_\varepsilon\big(T^{n(\varepsilon)}(O_\varepsilon(x))\big),$$

где окрестности понимаются как окрестности в L, входит в

$$O_2 := O_\varepsilon\big(T^{n(\varepsilon)}(O_\varepsilon(x))\big),$$

где окрестности понимаются как окрестности в X, и значит, $\mu_L(O_1) \leqslant \mu(O_2)$, откуда заключаем, что

$$P^\varepsilon_{L,T,\mu_L}(x) \leqslant P^\varepsilon_{X,T,\mu}(x). \tag{3}$$

Кроме того, $O_2 \subseteq \{T^{n(\varepsilon)}x_{[-k+n;k]}\}$, а значит,

$$\mu_L(O_1) \leqslant \mu(O_2) \leqslant \mu\big(\{T^{n(\varepsilon)}x_{[-k+n;k]}\}\big). \tag{4}$$

Далее нам понадобится лемма из [3] и ее простое следствие (тоже получено в [3], но также сразу следует из леммы и непрерывности меры).

Лемма 1 (Комеч). *Если X – синхронизованная система, а μ – мера из условия OT, то для почти любого $x \in X$ найдется такое $\delta(x)$, что для всех $\varepsilon < \delta(x)$ выполнено*

$$O_\varepsilon\big(T^{n(\varepsilon)}(O_\varepsilon(x))\big) = \{T^{n(\varepsilon)}x_{[-k+n;k]}\}.$$

Следствие 1. *В условиях леммы*

$$\mu\Big(x \in X :\ O_\varepsilon\big(T^{n(\varepsilon)}(O_\varepsilon(x))\big) = \{T^{n(\varepsilon)}x_{[-k+n;k]}\}\Big) \xrightarrow[\varepsilon\to 0]{} 1.$$

Теперь сделаем главный шаг: для подсдвигов L, удовлетворяющих условиям теоремы, выполнена лемма 1, а значит,

$$\mu\Big(x \in L :\ \mu_L(O_1) = \mu\big(\{T^{n(\varepsilon)}x_{[-k+n;k]}\}\big)\Big) \xrightarrow[\varepsilon\to 0]{} 1. \tag{5}$$

Обозначим множество под знаком меры в (5) через E_ε. Очевидно, что E_ε является подмножеством L. Из (4) и определения E_ε в (5) заключаем, что для $x \in E_\varepsilon$

$$\mu_L(O_1) = \mu(O_2) = \mu\big(\{T^{n(\varepsilon)}x_{[-k+n;k]}\}\big). \tag{6}$$

А значит, и

$$P^\varepsilon_{L,T,\mu_L}(x) = P^\varepsilon_{X,T,\mu}(x). \tag{7}$$

Завершим доказательство:

$$\big\|P^\varepsilon_{X,T,\mu}(x) - h(\mu,X,T)\big\|_{L_1(X,\mu)} =$$
$$= \int\limits_{E_\varepsilon} \big|P^\varepsilon_{X,T,\mu}(x) - h(\mu,X,T)\big| + \int\limits_{X\setminus E_\varepsilon} \big|P^\varepsilon_{X,T,\mu}(x) - h(\mu,X,T)\big|. \tag{8}$$

Из (1), (2) и (7) получаем

$$\int\limits_{E_\varepsilon} \big|P^\varepsilon_{X,T,\mu}(x) - h(\mu,X,T)\big| = \int\limits_{E_\varepsilon} \big|P^\varepsilon_{L,T,\mu_L}(x) - h(\mu_L,L,T)\big| \leqslant$$
$$\leqslant \int\limits_{L} \big|P^\varepsilon_{L,T,\mu_L}(x) - h(\mu_L,L,T)\big| \xrightarrow[\varepsilon\to 0]{} 0. \tag{9}$$

Для оценки второго слагаемого в (8) рассмотрим меру $\mu_{A^{\mathbb{Z}}}$ на $A^{\mathbb{Z}}$, определяемую на борелевских множествах $C \subseteq A^{\mathbb{Z}}$ равенством $\mu_{A^{\mathbb{Z}}}(C) = \mu(C \cap X)$. Для меры $\mu_{A^{\mathbb{Z}}}$ выполняются условия ОТ. Кроме того, понятно, что измеримые динамические системы (X, T, μ) и $(A^{\mathbb{Z}}, T, \mu_{A^{\mathbb{Z}}})$ изоморфны, а это влечет равенство их энтропий. Также заметим, что для фиксированных $x \in X$ и $\varepsilon > 0$ аналогично неравенству (3) выводится

$$P^{\varepsilon}_{X,T,\mu}(x) \leqslant P^{\varepsilon}_{A^{\mathbb{Z}},T,\mu_{A^{\mathbb{Z}}}}(x).$$

С учетом вышесказанного получаем:

$$\int\limits_{X\setminus E_\varepsilon} \left|P^{\varepsilon}_{X,T,\mu}(x) - h(\mu, X, T)\right| \leqslant \int\limits_{X\setminus E_\varepsilon} \left|P^{\varepsilon}_{X,T,\mu}(x)\right| + \int\limits_{X\setminus E_\varepsilon} h(\mu, X, T) \leqslant$$

$$\leqslant \int\limits_{X\setminus E_\varepsilon} \left|P^{\varepsilon}_{A^{\mathbb{Z}},T,\mu_{A^{\mathbb{Z}}}}(x)\right| + h(\mu, X, T)\mu(X \setminus E_\varepsilon) \leqslant$$

$$\leqslant \int\limits_{X\setminus E_\varepsilon} \left|P^{\varepsilon}_{A^{\mathbb{Z}},T,\mu_{A^{\mathbb{Z}}}}(x) - h(\mu_{A^{\mathbb{Z}}}, A^{\mathbb{Z}}, T)\right| + 2h(\mu, X, T)\mu(X \setminus E_\varepsilon) \xrightarrow[\varepsilon\to 0]{} 0.$$

Итак, оба слагаемых в правой части (8) стремятся к 0, а значит, и левая часть стремится к 0, что и завершает доказательство. ▲

Следствие 2. *Достаточным условием верности ОГ является синхронизованность носителя меры.*

Доказательство. Действительно, если L – носитель, то все его слова имеют положительную меру, а значит, если он синхронизован, то и все его волшебные слова имеют положительную меру. ▲

Теперь приведем примеры применения доказанной теоремы.

Пример 2. Пусть $X \subset \{a, (,), [,]\}^{\mathbb{Z}}$ определяется тем, что в словах $W(X)$ между двумя "a" запрещено разное количество открывающих и закрывающих скобок, а носитель меры – сдвиг Дика.

Эта система синхронизована, но мера каждого волшебного слова равна нулю, поэтому ОТ неприменима. Кроме того, носитель несинхронизован. Но взяв $L = \{(,), [,]\}^{\mathbb{Z}}$, мы можем применить ООТ.

Пример 3. X – система с алфавитом из четырех видов скобок $(,), [,], \{, \}, \langle, \rangle$ и языком, включающим в себя слова языка Дика на $\{, \}, \langle, \rangle$ и слова, преобразующиеся в них после удаления всех круглых и квадратных скобок. Например, "$\{\langle(\langle\rangle]\rangle\}$" – слово этого языка, а "$(\{\langle\rangle\rangle)$" – нет. Носитель меры – сдвиг Дика на круглых и квадратных скобках. Система X не синхронизована, как и обычная система Дика, носитель тоже не синхронизован, но взяв $L = \{(,), [,]\}^{\mathbb{Z}}$, мы можем применить ООТ.

§ 4. Контрпример к точечной гипотезе для системы Дика

В этом и следующем параграфе X – система Дика. Рассмотрим меру, определенную на цилиндрах следующим образом:

$$\mu(w_c) = (1/2)^{n_1+n_2+|w|}$$

(для любого номера c). Эта мера введена в работе [7], и там же показано, что она борелевская, вероятностная, инвариантная и эргодическая.

Замечание 1. В работе [7] также показано, что мера μ может быть определена и по-другому, а именно $\mu = (\mu_1 \times \mu_2) \circ F^{-1}$, где μ_1 и μ_2 – симметричные 2-бернуллиевские меры, а отображение F из $\{0,1\}^{\mathbb{Z}} \times \{1,2\}^{\mathbb{Z}}$ в X действует следующим образом:

1) ℓ-й член первой 2-бернуллиевской последовательности кодирует тип ℓ-й скобки (1 – открывающая, 0 – закрывающая);
2) ℓ-й член второй 2-бернуллиевской последовательности кодирует вид ℓ-й открывающей скобки.

Это отображение корректно и взаимно однозначно на множестве полной меры.

У т в е р ж д е н и е 1. *Возьмем* $x = \ldots [[[[[[[[[[. \ldots$ *Тогда* $P_{X,T,\mu}(x) \neq h(\mu, X, T)$.

Д о к а з а т е л ь с т в о. Энтропия получена Мееровичем в [7]:

$$h(\mu, X, T) = \frac{3}{2} \log 2.$$

Заметим, что для достаточно малых ε

$$O_\varepsilon\big(T^{n(\varepsilon)}(O_\varepsilon(x))\big) = \big\{T^{n(\varepsilon)} x_{[-k+n;k]}\big\},$$

откуда следует, что

$$\mu\big(O_\varepsilon\big(T^{n(\varepsilon)}(O_\varepsilon(x))\big)\big) = \mu\big(\big\{T^{n(\varepsilon)} x_{[-k+n;k]}\big\}\big) = \mu\big([^{2k-n+1}\big) = (1/2)^{2(2k-n+1)},$$
$$\mu(O_\varepsilon(x)) = \mu\big([^{2k+1}\big) = (1/2)^{2(2k+1)}.$$

А значит,

$$P_{X,T,\mu}(x) = 2 \log 2 \neq h(\mu, X, T). \quad \blacktriangle$$

Замечание 2. Зафиксируем $x \in X$ и $\varepsilon > 0$. Заметим, что если $x_{[k-n+1;k]}$ состоит только из открывающих скобок (любого вида), то ε-ЛСДГ для этих x и ε тоже будет равняться $2 \log 2$ (так как множества, стоящие под знаком меры в числителе и знаменателе, будут цилиндрами, основания которых отличаются друг от друга на $n(\varepsilon)$ непарных скобок) и не совпадать с энтропией.

§ 5. Об обобщенной точечной и основной гипотезах для системы Дика

Т е о р е м а 2. *Для системы Дика с мерой из* § 4 *обобщенная точечная гипотеза неверна. Более того, при усилении условия на функцию* $n(\varepsilon)$ *до* $2^{n(\varepsilon)} = o(\log \varepsilon)$ *предел* $P_{X,T,\mu}(x)$ *вообще не существует п.н.*

Д о к а з а т е л ь с т в о. Пусть $n(\varepsilon)$ отвечает усиленному условию. Будем также считать, что $n(\varepsilon)$ не убывает и является однозначной функцией k: $n = n(k)$ (например, $n(k) = \lfloor \log \log k \rfloor$). В противном случае можно выбрать стремящуюся к 0 последовательность ε (и, возможно, соответствующую ей подпоследовательность k), для которой эти условия выполнены.

Рассмотрим случайные величины $\Sigma(k)$, равные времени ожидания первого появления $n(k)$ открывающих скобок подряд (наблюдение начинается в нулевой момент) и события $A_k = \{\Sigma(k) \leqslant k\}$. Обратимся к определению меры μ из замечания 1. Заметим, что событие A_k инвариантно относительно корректного (т.е. не выводящего за пределы системы Дика) изменения вида скобок без изменения их типа. Такая инвариантность влечет инвариантность прообраза $F^{-1}(A_k)$ относительно изменения второй компоненты. Значит, $F^{-1}(A_k)$ будет иметь вид $B_k \times \{1,2\}^{\mathbb{Z}}$, где B_k– эквивалентные A_k события, но на множестве 2-бернуллиевских последовательностей.

Таким образом, $B_k = \{\Gamma(k) \leqslant k\}$, где $\Gamma(k)$ – случайные величины, равные времени ожидания первого появления $n(k)$ единиц подряд, начиная с нулевого момента. Понятно, что $\mu(A_k) = \mu_1(B_k)$. Далее вместо μ_1 используем обозначение P.

Покажем, что почти всюду выполняется бесконечно много событий A_k, что эквивалентно выполнению бесконечного числа событий B_k, т.е. нужно показать, что

$$P\left(\bigcap_{\ell=1}^{\infty}\bigcup_{k=\ell}^{\infty} B_k\right) = 1,$$

или (что то же самое)

$$P\left(\bigcup_{\ell=1}^{\infty}\bigcap_{k=\ell}^{\infty} \overline{B}_k\right) = 0,$$

где $\overline{B}_k$ – события, дополняющие B_k: $\overline{B}_k = \{\Gamma(k) > k\}$.

Известно (см. [8]), что математическое ожидание имеет вид

$$\mathbf{E}\,\Gamma(k) = 2^{n+1} - 2 = o(k),$$

а среднеквадратическое отклонение –

$$\sigma\Gamma(k) = 2^{n+1} + o(2^{n+1}) = o(k).$$

Зафиксируем $\delta > 0$ и натуральное L, такое что $1/L^2 < \delta$. Найдем достаточно малое $\varepsilon_0 > 0$ (и соответствующее ему $K = k(\varepsilon_0)$), такое что для всех $\varepsilon \leqslant \varepsilon_0$ (и соответственно, $k \geqslant K$) выполнено

$$\mathbf{E}\,\Gamma(k) + L\sigma\Gamma(k) < k.$$

По неравенству Чебышева

$$P(|\Gamma(k) - \mathbf{E}\,\Gamma(k)| \geqslant L\sigma\Gamma(k)) \leqslant 1/L^2,$$

и для $k \geqslant K$ имеем

$$P\big(\overline{B}_k\big) = P\big(\Gamma(k) > k\big) \leqslant P\Big(\Gamma(k) \geqslant \mathbf{E}\,\Gamma(k) + L\sigma\Gamma(k)\Big) \leqslant$$
$$\leqslant P\Big(|\Gamma(k) - \mathbf{E}\,\Gamma(k)| \geqslant L\sigma\Gamma(k)\Big) \leqslant 1/L^2 < \delta.$$

Тогда, используя непрерывность меры, получаем

$$P\left(\bigcup_{\ell=1}^{\infty}\bigcap_{k=\ell}^{\infty} \overline{B}_k\right) = \lim_{\ell\to\infty} P\left(\bigcap_{k=\ell}^{\infty} \overline{B}_k\right) \leqslant \delta.$$

Отсюда ввиду произвольности δ заключаем, что

$$P\left(\bigcup_{\ell=1}^{\infty}\bigcap_{k=\ell}^{\infty} \overline{B}_k\right) = 0,$$

и по вышесказанному

$$P\left(\bigcap_{\ell=1}^{\infty}\bigcup_{k=\ell}^{\infty} A_k\right) = 1,$$

т.е. почти наверное выполняется бесконечное количество событий A_k.

Зафиксируем $x \in X$, для которого выполняется бесконечное количество A_k. Обозначим последовательность k, для которых выполняется A_k, через k_t, и пусть $i_t := \Sigma(k_t)(x) \leqslant k_t$. Так как $n(k)$ – неубывающая функция, то $n(i_t) \leqslant n(k_t)$, и последние $n(i_t)$ скобок в момент i_t – открывающие, а это по замечанию 2 дает одинаковые (равные $2\log 2$) и не совпадающие с энтропией ε-ЛСДГ в моменты i_t (более точно, этим ε-ЛСДГ соответствуют, например, $\varepsilon_t = 1/2^{i_t+1}$). Наконец, i_t возрастает к бесконечности, ведь $i_t \geqslant n(k_t)$ (так как это просто время ожидания первого появления $n(k_t)$ открывающих скобок подряд, начиная с нулевого момента), а $n(k_t)$ возрастает к бесконечности. Так как i_t возрастает к бесконечности, то ε_t стремится к 0. Итак, мы получили стремящуюся к нулю последовательность ε_t с одинаковыми ε_t-ЛСДГ, не совпадающими с энтропией, но тогда и ЛСДГ для выбранного x как предел ε-ЛСДГ не совпадает с энтропией. Ввиду того что x – почти любое, получаем, что ЛСДГ почти нигде не совпадает с энтропией. ▲

Замечание 3. Вообще говоря, выше не доказано, что ЛСДГ не может поточечно равняться построенным ε_t-ЛСДГ, т.е. $2\log 2$. Это можно показать непосредственно тем же способом, построив последовательность с одинаковым, но другим значением ε-ЛСДГ. Но также это следует из нижеследующей теоремы о том, что для рассматриваемой системы верна ОГ, и следующего факта: если имеет место сходимость в среднем к некоторой константе, то можно выделить подпоследовательность, сходящуюся к этой константе почти наверное, что исключает поточечную сходимость к другой константе на множестве положительной меры.

Замечание 4. Для несимметричной 2-бернуллиевской меры на полном сдвиге эквивалентное утверждение показывается дословным повторением данного доказательства с момента перехода к бернуллиевскому случаю (экспоненциальный рост математического ожидания и дисперсии рассматриваемых случайных величин сохраняется и в новой ситуации, как и отличие построенных ε_t-ЛСДГ от энтропии). Таким образом, ОТГ неверна даже для бернуллиевских мер.

Теорема 3. *Для системы Дика и меры из предыдущего параграфа верна основная гипотеза.*

Доказательство. В [3] доказано, что для справедливости основной гипотезы достаточно, чтобы выполнялось условие из следствия 1, т.е., чтобы для $k = k(\varepsilon)$ и любой медленной функции $n = n(\varepsilon)$ было выполнено

$$\mu(E_\varepsilon) \xrightarrow[\varepsilon\to 0]{} 1, \tag{10}$$

где

$$E_\varepsilon = \Big\{x \in X :\, O_\varepsilon\big(T^n(O_\varepsilon(x))\big) = \big\{T^n x_{[-k+n;k]}\big\}\Big\}.$$

Это мы и покажем. Далее для простоты будем называть множества, для которых справедливо (10), "множествами большой меры".

Пусть

$$O_1 := O_\varepsilon\big(T^{n(\varepsilon)}(O_\varepsilon(x))\big), \quad O_2 := \big\{T^{n(\varepsilon)} x_{[-k+n;k]}\big\}$$

(здесь есть зависимость от x и ε, которую мы в записи опускаем); очевидно, что $O_1 \subseteq O_2$.

Значительно сложнее показать, что на множествах большой меры имеет место и обратное включение. Фактически $O_2 \subseteq O_1$ для конкретного x означает, что для любого слова w длины n, если $wx_{[-k;k-n]}$ разрешено, то и $wx_{[-k;k]}$ разрешено (здесь имеются в виду конкатенации w с соответствующими отрезками x), неформально говоря, $x_{[k-n+1;k]}$ "не влияет" на разрешенность $wx_{[-k;k]}$. Заметим, что для выполнения

данного условия достаточно, чтобы максимум разности открывающих и закрывающих скобок по суффиксам слова $x_{[-k;k-n]}$ был не меньше, чем максимум разности закрывающих и открывающих скобок по префиксам слова $x_{[k-n+1;k]}$. Действительно, первый максимум равен количеству непарных открывающих скобок в $x_{[-k;k-n]}$, а второй – количеству непарных закрывающих скобок в $x_{[k-n+1;k]}$, и при выполнении указанного выше условия все закрывающие скобки в $x_{[k-n+1;k]}$ имеют пару в $x_{[-k;k]}$, а значит, $x_{[k-n+1;k]}$ "не влияет" на разрешенность $wx_{[-k;k]}$. Таким образом, наша задача сводится к тому, чтобы показать, что одна случайная величина (первый максимум) не меньше другой (второй максимум) на множестве большой меры. Обозначим это множество через S_ε (при фиксированном ε будем писать просто S), из вышесказанного имеем $S_\varepsilon \subseteq E_\varepsilon$.

Обратимся к определению меры μ из замечания 1. Заметим, что множество S инвариантно относительно корректного (т.е. не выводящего за пределы системы Дика) изменения вида скобок без изменения их типа, ведь рассматриваемые максимумы зависят лишь от того, открывающая или закрывающая скобка стоит на конкретном месте. Такая инвариантность влечет инвариантность прообраза $F^{-1}(S)$ относительно изменения второй компоненты. Значит, $F^{-1}(S)$ будет иметь вид $R\times\{1,2\}^{\mathbb{Z}}$, где R – некоторое подмножество $\{0,1\}^{\mathbb{Z}}$. Отсюда $\mu(S) = \mu_1(R)$. Из определения F очевидно, что R (более точно, R_ε) – это множество бернуллиевских последовательностей y, для которых максимум разности числа единиц и числа нулей по суффиксам $y_{[-k;k-n]}$ не меньше максимума разности числа нулей и числа единиц по префиксам $y_{[k-n+1;k]}$, но это просто максимумы независимых симметричных случайных блужданий на отрезках $[0;2k-n+1]$ и $[0;n]$. Задача сведена к простейшей!

Осталось показать, что несмотря на независимость рассматриваемых максимумов, один из них не меньше другого на множестве, мера которого стремится к 1 при $\varepsilon \to 0$. Как известно (см., например, [9]), максимум симметричного случайного блуждания на отрезке $[0;m]$ имеет следующее распределение:

$$P(M_m \geqslant r) = 2P(C_m \geqslant r) - P(C_m = r),$$

где r – целое неотрицательное, M_m – данный максимум, C_m – величина случайного блуждания на m-м шаге. Отсюда

$$P(M_m \geqslant r) \leqslant 2P(C_m \geqslant r), \tag{11}$$
$$P(M_m \geqslant r) \geqslant 2P(C_m \geqslant r+1). \tag{12}$$

Здесь и далее для симметричной 2-бернуллиевской меры используем обозначение P. Напомним также, что $n = o(k)$.

Пусть $r = \left\lfloor \sqrt[4]{n(2k-n+1)} \right\rfloor$. Так как n мало относительно k, то $\sqrt{n}$ мал относительно r, но r мало относительно $\sqrt{2k-n+1}$. Тогда для любых $K > 0$, $\ell > 0$ и для достаточно малых ε выполняется следующее: $r \geqslant K\sqrt{n}$, но $r+1 \leqslant \ell\sqrt{2k-n+1}$.

Рассмотрим $m = n$ и зафиксируем $K > 0$. Так как симметричное случайное блуждание – это сумма независимых одинаково распределенных величин с нулевым среднем и единичной дисперсией, то $\mathbf{E}\, C_n = 0$, $\sigma C_n = \sqrt{n}$ (σ – среднеквадратическое отклонение), и из (11) по неравенству Чебышева для достаточно малых ε получаем

$$P(M_n \geqslant r) \leqslant 2P(C_n \geqslant r) \leqslant 2P(C_n \geqslant K\sqrt{n}) \leqslant 2/K^2,$$

т.е. для любого $K > 0$ можно найти правую окрестность нуля, такую что для любого ε из этой окрестности $P(M_n \geqslant r) \leqslant 2/K^2$, т.е. $P(M_n \geqslant r) \xrightarrow[\varepsilon\to 0]{} 0$, откуда

$$P(M_n < r) \xrightarrow[\varepsilon\to 0]{} 1.$$

Напомним центральную предельную теорему (ЦПТ) для симметричного случайного блуждания.

Теорема 4. *Имеет место сходимость по распределению*

$$\frac{C_m}{\sqrt{m}} \xrightarrow[m\to\infty]{d} N,$$

где N – стандартная нормальная величина.

Пусть теперь $m = 2k - n + 1$ и $\ell > 0$. Из (12) для достаточно малых ε имеем

$$P\big(M_{2k-n+1} \geqslant r\big) \geqslant 2P\big(C_{2k-n+1} \geqslant r+1\big) \geqslant 2P\big(C_{2k-n+1} \geqslant \ell\sqrt{2k-n+1}\big).$$

Заметим, что $2k - n + 1 \xrightarrow[\varepsilon\to 0]{} \infty$, и воспользуемся ЦПТ для симметричного случайного блуждания:

$$P\big(C_{2k-n+1} \geqslant \ell\sqrt{2k-n+1}\big) = P\big(C_{2k-n+1}/\sqrt{2k-n+1} \geqslant \ell\big) \xrightarrow[\varepsilon\to 0]{} P(N \geqslant \ell),$$

в частности, для достаточно малых ε

$$P\big(C_{2k-n+1} \geqslant \ell\sqrt{2k-n+1}\big) \geqslant P(N \geqslant 2\ell),$$

откуда в правой окрестности нуля (зависящей от ℓ):

$$P(M_{2k-n+1} \geqslant r) \geqslant 2P(N \geqslant 2\ell).$$

Учитывая произвольность ℓ и непрерывность функции нормального распределения (и то, что вероятность, очевидно, не больше $1 = 2P(N \geqslant 0)$), получаем

$$P(M_{2k-n+1} \geqslant r) \xrightarrow[\varepsilon\to 0]{} 2P(N \geqslant 0) = 1.$$

Итак,

$$P(M_n < r) \xrightarrow[\varepsilon\to 0]{} 1,$$
$$P(M_{2k-n+1} \geqslant r) \xrightarrow[\varepsilon\to 0]{} 1.$$

Теперь, вспоминая, что в нашей ситуации M_n и M_{2k-n+1} – независимые величины, находим

$$P(M_n \leqslant M_{2k-n+1}) \geqslant P(M_n \leqslant r \leqslant M_{2k-n+1}) =$$
$$= P(M_n < r)P(M_{2k-n+1} \geqslant r) \xrightarrow[\varepsilon\to 0]{} 1.$$

Возвращаясь к первоначальным обозначениям, тем сам мы показали, что один максимум не меньше другого на множестве R_ε большой меры, но мера S_ε равна мере R_ε, и значит, тоже стремится к единице. Наконец, $S_\varepsilon \subseteq E_\varepsilon$, откуда $\mu(E_\varepsilon) \to 1$. ▲

СПИСОК ЛИТЕРАТУРЫ

1. *Заславский Г.М.* Стохастичность динамических систем. М.: Наука, 1984.
2. *Gurevich B.M.* Geometric Interpretation of Entropy for Random Processes // Sinai's Moscow Seminar on Dynamical Systems. Providence, RI: Amer. Math. Soc., 1996. P. 81–87.
3. *Комеч С.А.* Скорость искажения границы в синхронизованных системах: геометрический смысл энтропии // Пробл. передачи информ. 2012. Т. 48. № 1. С. 15–25. http://mi.mathnet.ru/ppi2065

4. *Гуревич Б.М., Комеч С.А.* Скорость деформации границ в системах Аносова и близких к ним // Тр. МИАН. 2017. Т. 297. С. 211–223. https://doi.org/10.1134/S03719685170 2011X
5. *Синай Я.Г.* О понятии энтропии динамической системы // ДАН СССР. 1959. Т. 124. С. 768–771.
6. *Корнфельд И.П., Синай Я.Г., Фомин С.В.* Эргодическая теория. М.: Наука, 1980.
7. *Meyerovitch T.* Tail Invariant Measures of the Dyck-Shift and Non-sofic Systems. Master Thesis. Tel-Aviv Univ., Israel, 2004.
8. *Perng C., Yamoah G,, Ali A.* Computing Probability Generating Functions of Coin Flipping and Its Generalizations // Appl. Math. Sci. (Ruse). 2013. Vol. 7. № 115. P. 5693–5710. https://doi.org/10.12988/ams.2013.38460
9. Ширяев А.Н. Вероятность. Т. 1. М.: МЦНМО, 2007.

Дворкин Григорий Дмитриевич
Московский государственный университет
им. М.В. Ломоносова, механико-математический факультет,
кафедра математической статистики и случайных процессов
grisha230531415@gmail.com

Поступила в редакцию
25.03.2021
После доработки
17.06.2021
Принята к публикации
27.06.2021

ПРОБЛЕМЫ ПЕРЕДАЧИ ИНФОРМАЦИИ

Том 57 **2021** **Вып. 3**

УДК 621.391 : 004.932

© 2021 г. **С.М. Карпенко, Е.И. Ершов**

ИССЛЕДОВАНИЕ СВОЙСТВ ДИАДИЧЕСКОГО ПАТТЕРНА БЫСТРОГО ПРЕОБРАЗОВАНИЯ ХАФА[1]

Получена оценка максимального отклонения от геометрической прямой аппроксимирующего ее дискретного (диадического) паттерна, используемого при вычислении быстрого преобразования Хафа (дискретного преобразования Радона) для квадратного изображения с размером стороны $n = 2^p$, $p \in \mathbb{N}$. Для четных p максимальное отклонение составляет $p/6$. Важную роль в доказательстве играет анализ тонких свойств простого комбинаторного объекта – таблицы циклических сдвигов произвольного двоичного числа.

Ключевые слова: быстрое преобразование Хафа, быстрое преобразование Радона, диадический паттерн, анализ ошибки, комбинаторная оптимизация, двоичные слова.

DOI: 10.31857/S0555292321030074

§ 1. Введение

Преобразованием Радона (ПР) на евклидовой плоскости называется интегральное преобразование, относящее функции f ее интегралы по всевозможным прямым [1]. В полярной параметризации преобразование Радона функции $f(x, y)$ задается следующим равенством:

$$\mathcal{R}f(\varphi, p) = \int_{-\infty}^{\infty} f(-t\sin\varphi + p\cos\varphi, t\cos\varphi + p\sin\varphi)\, dt.$$

В анализе изображений используются различные вариации дискретного преобразования Радона, в частности быстрое преобразование Хафа (БПХ) [2]. Интересно, что алгоритм изобретался как минимум четыре раза [2–5]. О быстром дискретном преобразовании Радона и его обращении замечательно написано в работе У. Пресса [6]. БПХ вычисляет суммы по набору дискретных паттернов специального вида, так называемых диадических, и используется для поиска прямолинейных или составленных из прямых объектов на изображении.

В работе [7] приведена (в соответствии с вычислительными экспериментами) оценка максимального отклонения диадического паттерна от соответствующей геометрической прямой, и поставлена задача доказать эту оценку. В настоящей статье мы доказываем и уточняем оценку $\frac{p}{6}$ из работы [7].

[1] Исследование выполнено при частичной финансовой поддержке Российского фонда фундаментальных исследований в рамках научного проекта № 18-29-26020.

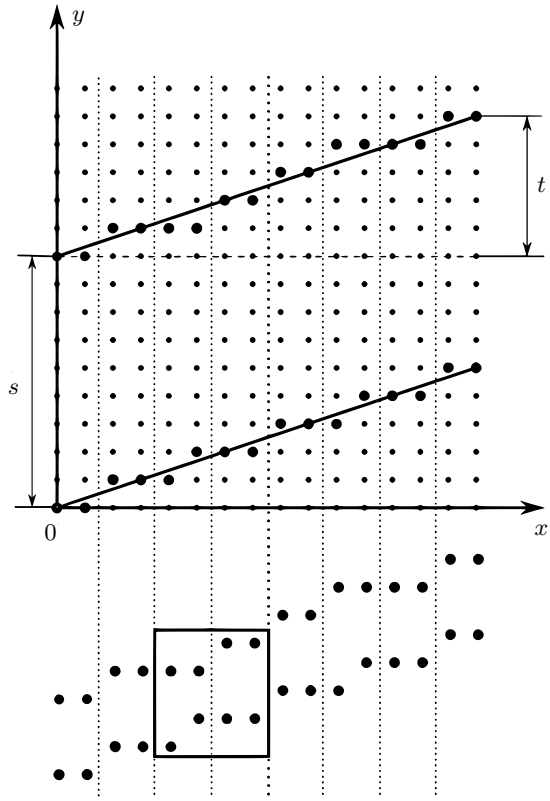

Рис. 1. Изображение размера 16×16 с заданной системой координат. Иллюстрация структур ближайшего (нижний) и диадического (верхний) паттернов, аппроксимирующих прямую с наклоном $t = 5$

§ 2. Параметризация, структура и ошибка аппроксимации диадического паттерна

Изображением будем называть векторнозначную функцию, заданную на дискретном множестве Ω вида $[0, \ldots, n-1] \times [0, \ldots, m-1]$, где $n, m \in \mathbb{Z}$. В статье будем считать, что $n = m = 2^p$, где p – некоторое натуральное число.

Паттерном будем называть любое подмножество множества Ω, а *прямолинейным паттерном* или *дискретной прямой* будем называть подмножество, аппроксимирующее непрерывную прямую. Например, *ближайшим* прямолинейным паттерном будем называть подмножество, задаваемое уравнением

$$y = \big[kx + b\big],$$

здесь и далее $[\,\cdot\,]$ обозначает округление до ближайшего целого, $k, b \in \mathbb{R}$ – параметры аппроксимируемой прямой.

Без ограничения общности далее в статье рассмотрим только прямые с наклонами $0 \leqslant k \leqslant 1$, будем задавать их парой точек $P_1(0, s)$ и $P_2(n-1, s+t)$ на границах изображения, где $s \in [0, n-1]$ – целочисленный сдвиг паттерна, а $t \in [0, n-1]$ – целочисленный наклон (см. рис. 1). Далее будем рассматривать случай $s = 0$.

В работе [8] предложено аналитическое определение диадического паттерна с использованием базисных диадических паттернов.

Определение 1. *Базисным диадическим паттерном* называется дискретная прямая вида

$$y = D_r(x) = \left[\frac{tx}{2^p - 1}\right] = \left[\frac{2^r x}{2^p - 1}\right],$$

где $x, y \in \mathbb{Z}^+$ – абсцисса и ордината графика прямой, $x, y < n$, а $t = 2^r$ – параметр наклона, $r < p$.

Определение 2. Пусть $t = t_{p-1}t_{p-2}\dots t_0$ – двоичная запись параметра наклона. Тогда дискретная прямая вида

$$D(x,t) = \sum_{r=0}^{p-1} t_r D_r(x) \tag{1}$$

называется *диадическим паттерном* с наклоном $t \in [0, 2^p - 1]$ (см. [8]).

Используя определение (1), запишем ошибку аппроксимации диадическим паттерном непрерывной прямой вида $L(x,t) = \frac{tx}{2^p - 1}$:

$$E(x,t) = D(x,t) - L(x,t) = \sum_{r=0}^{p-1} t_r \left[\frac{2^r x}{2^p - 1}\right] - \frac{tx}{2^p - 1} =$$

$$= \sum_{r=0}^{p-1} t_r \left(\left[\frac{2^r x}{2^p - 1}\right] - \frac{2^r x}{2^p - 1} \right). \tag{2}$$

Другим существенным свойством диадического паттерна является знакопеременность ошибки аппроксимации относительно центра паттерна.

Утверждение 1. *Для любого* $x \in [0, 2^{p-1} - 1]$ *и любого* $t \in [0, 2^p - 1]$ *верно, что* $E(2^{p-1} - 1 - x, t) = -E(2^{p-1} + x, t)$.

Доказательство. Запишем утверждение согласно формуле (2):

$$\frac{(2^{p-1} - 1 - x)2^r}{2^p - 1} - \left[\frac{(2^{p-1} - 1 - x)2^r}{2^p - 1}\right] = \left[\frac{(2^{p-1} + x)2^r}{2^p - 1}\right] - \frac{(2^{p-1} + x)2^r}{2^p - 1}.$$

Преобразуем данное выражение:

$$\left[\frac{(2^{p-1} + x)2^r}{2^p - 1}\right] + \left[\frac{(2^{p-1} - 1 - x)2^r}{2^p - 1}\right] = 2^r.$$

Покажем, что это уравнение является тождеством:

$$\left[\frac{2^{p-1}2^r}{2^p - 1} + \frac{2^r x}{2^p - 1}\right] + \left[\frac{2^{p-1}2^r}{2^p - 1} - \frac{(x+1)2^r}{2^p - 1}\right] = 2^r,$$

$$\left[2^{r-1} + \frac{2^{r-1}}{2^p - 1} + \frac{2^r x}{2^p - 1}\right] + \left[2^{r-1} + \frac{2^{r-1}}{2^p - 1} - \frac{(y+1)2^r}{2^p - 1}\right] = 2^r.$$

В результате получаем

$$\left[\frac{2^{r-1} + 2^r x}{2^p - 1}\right] + \left[\frac{-2^{r-1} - 2^r x}{2^p - 1}\right] = 0. \quad \blacktriangle$$

Утверждение 2. *Значение функции* $E(x,t)$ *не меняется при перестановке ее аргументов, т.е. координаты вдоль прямой и наклона прямой.*

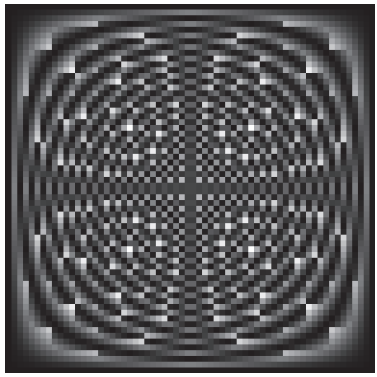

Рис. 2. Симметрия модуля ошибки аппроксимации относительно отражения (утверждение 1) и перестановки аргументов – абсциссы x и наклона прямой t (утверждение 2)

Доказательство. Итак, пусть $x = \sum\limits_{s=0}^{p-1} 2^s x_s$, $t = \sum\limits_{r=0}^{p-1} 2^r t_r$. Мы хотим продемонстрировать, что $E(x,t) = E(t,x)$. В силу равенства (2) достаточно показать, что

$$\left[\frac{2^r x}{2^p-1}\right] = \sum_{s=0}^{p-1} x_s \left[\frac{2^r 2^s}{2^p-1}\right].$$

Заметим, что $\left[\frac{2^{p-1}}{2^p-1}\right] = \left[\frac{2^p}{2^p-1}\right] = 1$; вообще, при $0 \leqslant d \leqslant 2p-2$ имеем

$$\left[\frac{2^d}{2^p-1}\right] = \begin{cases} 0, & \text{если } d < p-1, \\ 1, & \text{если } d = p-1, \\ 2^{d-p}, & \text{если } p-1 < d \leqslant 2p-2. \end{cases} \tag{3}$$

Поскольку для $x \in [0, 2^{p-1}]$ верно, что

$$\left[\frac{x}{2^p-1}\right] = 0{,}(x_{p-1}\ldots x_0),$$

где через $0{,}(x_{p-1}\ldots x_0)$ обозначена бесконечная двоичная периодическая дробь, то для любого $r \leqslant p-1$ справедливо

$$\left[\frac{2^r x}{2^p-1}\right] = [2^r \cdot 0{,}(x_{p-1}\ldots x_0)] = [x_{p-1}\ldots x_{p-r}{,}(x_{p-r-1}\ldots x_{p-r})] =$$
$$= x_{p-1}\ldots x_{p-r} + x_{p-r-1} = x_{p-r-1} + \sum_{s=p-r}^{s=p-1} 2^{r+s-p} x_s.$$

Полагая $d = r+s$ в (3), получаем требуемое. ▲

§ 3. Вычисление ошибки аппроксимации

3.1. Двоичное представление ошибки аппроксимации. Запишем выражение ошибки аппроксимации диадическим паттерном наклона t в точке x согласно выраже-

нию (2):

$$E(x,t) = \sum_{r=0}^{p-1} t_r E_r(x), \tag{4}$$

где $E_r(x) = \left(\left[\frac{2^r x}{2^p-1}\right] - \frac{2^r x}{2^p-1}\right)$ – ошибка аппроксимации базисным диадическим паттерном с наклоном 2^r.

Нетрудно выписать двоичное представление $E_r(x)$, а именно:

1. $\frac{1}{2^p-1} = 0,\underbrace{(0\dots1)}_{p};$
2. $\frac{x}{2^p-1} = 0,(x_{p-1}x_{p-2}\dots x_0)$ для $x = x_{p-1}x_{p-2}\dots x_0;$
3. $E_r(x) = [\dots x_{p-r-1},(x_{p-r-2}\dots x_0\dots x_{p-r-1})] - \dots x_{p-r-1},(x_{p-r-2}\dots x_0\dots x_{p-r-1});$
4. $E_r(x) = \begin{cases} 1-0,(x_{p-r-2}\dots x_0\dots x_{p-r-1}), & x_{p-r-2}=1, \\ -0,(x_{p-r-2}\dots x_0\dots x_{p-r-1}), & x_{p-r-2}=0. \end{cases}$

Заметим, что для фиксированного x модуль ошибки $E(x,t)$ будет максимален, когда $t_r = 1$ только для слагаемых одного знака. С учетом утверждения 1 для поиска максимальной по модулю ошибки аппроксимации мы можем ограничиться только такими (r,x), для которых, например, $E_r(x) < 0$.

Таким образом, мы перешли от исходной задачи оценивания ошибки аппроксимации к комбинаторной, описанной в следующем пункте.

3.2. Постановка задачи про таблицу циклических сдвигов. Пусть $T(x)$ – таблица всевозможных циклических сдвигов двоичной записи x:

x_{p-1}	x_{p-2}	$\dots$	x_0
x_{p-2}	x_{p-3}	$\dots$	x_{p-1}
x_{p-3}	x_{p-4}	$\dots$	x_{p-2}
$\dots$	$\dots$	$\dots$	$\dots$
x_0	x_{p-1}	$\dots$	x_1.

Определим функцию от таблицы $S(T(x))$ как арифметическую сумму чисел, соответствующих строкам с нулевым старшим битом. Первый столбец и строка в таблице отделены в иллюстративных целях: в столбце располагаются “управляющие” биты, в строке – слагаемые.

Задача: найти $\max\limits_x S(T(x))$.

3.3. Анализ свойств таблицы циклических сдвигов. Символами $\boldsymbol{v}$ и $\boldsymbol{w}$ обозначим произвольную двоичную последовательность, а символами $\boldsymbol{v}^*$ и $\boldsymbol{w}^*$ – их инверсию, записанную в обратном порядке. Например, если $\boldsymbol{v} = 010000$, то $\boldsymbol{v}^* = 111101$.

По построению для любого числа x таблица $T(x)$ будет иметь следующую структуру, где $\boldsymbol{v} = x_{p-2},\dots,x_0$:

x_{p-1}	$\boldsymbol{v}$		
$\boldsymbol{v}$	x_{p-3}	$\dots$	x_{p-1}
	x_{p-4}	$\dots$	x_{p-2}
	$\dots$	$\dots$	$\dots$
	x_{p-1}	$\dots$	x_1.

Заметим, что битовое представление x содержится как в первой строке, так и в первом столбце.

Лемма 1 (о замене). *Пусть $T_\varepsilon(x)$ обозначает таблицу размера $p \times p$ вида*

ε	$\boldsymbol{v}$		
	x_{p-3}	$\dots$	ε
	x_{p-4}	$\dots$	x_{p-2}
$\boldsymbol{v}$	$\dots$	$\dots$	$\dots$
	ε	$\dots$	x_1,

где $x_{p-1} = \varepsilon$. Тогда $S(T_0(x)) > S(T_1(x))$ эквивалентно $\boldsymbol{v} > \boldsymbol{v}^*$.

Доказательство. Для доказательства леммы достаточно показать, что

$$S(T_0(x)) - S(T_1(x)) = \boldsymbol{v} - \boldsymbol{v}^*.$$

Вычислим разность $S(T_0(x)) - S(T_1(x))$. Так как суммирование производится только по строкам с нулевым старшим битом, то разность первых строк равна $\boldsymbol{v}$. Далее, для пары строк с номером $r > 1$ в таблице $T_0(x)$ значение в разряде $p - r$ равно нулю, в то время как в правой таблице $T_1(x)$ – единице, а значения в остальных разрядах строк совпадают, т.е. вкладом такой пары строк является отрицательное число с единственной единицей в $p-r$ разряде. Тогда сумма вкладов всех таких пар строк равна $-\boldsymbol{v}^*$. ▲

Лемма 2 (о перестановке). *Пусть символ 1_p обозначает битовую последовательность единиц длинны p, а $T_{\varepsilon\delta}(x)$ – таблицу размера $p \times p$ вида*

ε	δ	$\boldsymbol{v}$	
δ	x_{p-3}	$\dots$	ε
	x_{p-4}	$\dots$	δ
$\boldsymbol{v}$	$\dots$	$\dots$	$\dots$
	ε	$\dots$	x_1.

Тогда $S(T_{01}(x)) > S(T_{10}(x))$ эквивалентно $\boldsymbol{v} + \boldsymbol{v}^* < 1_{p-2}$.

Доказательство. Аналогично доказательству леммы 1 запишем разность сумм по таблицам:

$$S(T_{01}(x)) - S(T_{10}(x)) = 01\boldsymbol{v} - 0\boldsymbol{v}1 - \boldsymbol{v}^*,$$

здесь первое слагаемое есть вклад пары строк с $r = 0$, второе – вклад пары строк с $r = 1$, а третье – вклад оставшихся пар строк, полученный аналогичным лемме 1 способом. После преобразования получим

$$S(T_{01}(x)) - S(T_{10}(x)) = 1_{p-2} - \boldsymbol{v} - \boldsymbol{v}^*,$$

что и требовалось. ▲

Назовем *максимайзером* любой $\widehat{x} = \arg\max\limits_{x \in [0, 2^p-1]} S(T(x))$. Будем теперь последовательно сужать множество возможных максимайзеров.

Обозначим символом $\overset{m}{\sim}$ последовательность чередующихся нулей и единиц длины m, а символом $^{\varepsilon}\overset{m}{\sim}{}^{\delta}$ – последовательность чередующихся нулей и единиц длины m со значениями ε и δ в старшем и младшем разрядах соответственно, например, $^1\overset{5}{\sim}{}^1 = 10101$.

Предложение 1. *Если в двоичной записи числа x встречаются три подряд идущих равных бита (далее* – битовая тройка), *то такой x не может быть максимайзером.*

Доказательство. Без ограничения общности рассмотрим число, в котором есть как минимум одна битовая тройка единиц $111\boldsymbol{w}$. Пусть $\widehat{x} = 11\boldsymbol{w}1$. Так как $T(x)$

содержит все циклические сдвиги, то $S(T(x)) = S(T(\widehat{x}))$. Следовательно, достаточно показать, что

$$S(T(01\boldsymbol{w}1)) > S(T(11\boldsymbol{w}1)).$$

Обозначим $\boldsymbol{v} = 1\boldsymbol{w}1$, тогда $\boldsymbol{v}^* = 0\boldsymbol{w}^*0$. Видно, что $\boldsymbol{v} > \boldsymbol{v}^*$. Следовательно, согласно лемме 1 имеем $S(T(11\boldsymbol{w}1)) > S(T(01\boldsymbol{w}1))$, и $111\boldsymbol{w}$ не может быть максимайзером. ▲

Предложение 2. *Число вида* $x = 11^0 \overset{\ell}{\sim} 0 11^0 \overset{m}{\sim} 0 11\boldsymbol{w}$ *не может быть максимайзером.*

Доказательство. Пусть $\widehat{x}$ получен в результате циклического сдвига x:

$$\widehat{x} = 11^0 \overset{m}{\sim} 0 11\boldsymbol{w}11^0 \overset{\ell}{\sim} 0.$$

Обозначим

$$\boldsymbol{v} = 1^0 \overset{m}{\sim} 0 11\boldsymbol{w}11^0 \overset{\ell}{\sim} 0 = {}^1 \overset{m+2}{\sim} {}^1 1\boldsymbol{w}11^0 \overset{\ell}{\sim} 0,$$

тогда

$$\boldsymbol{v}^* = {}^1 \overset{\ell}{\sim} {}^1 00\boldsymbol{w}^*00^1 \overset{m}{\sim} {}^1 0,$$

Для сравнения $\boldsymbol{v}$ и $\boldsymbol{v}^*$ необходимо рассмотреть три случая: $m+2 < \ell$, $m+2 = \ell$ и $m+2 > \ell$. Рассмотрим, например, случай, когда $m+2 < \ell$. Будем побитово сравнивать $\boldsymbol{v}$ и $\boldsymbol{v}^*$, начиная со старшего бита. Видно, что в первом несовпавшем разряде $\boldsymbol{v}$ содержит единицу, а $\boldsymbol{v}^*$ – ноль, следовательно, $\boldsymbol{v} > \boldsymbol{v}^*$. Аналогичным образом рассматриваются остальные два случая. Следовательно, согласно лемме 1 имеем

$$S(T(01^0 \overset{\ell}{\sim} 0 01^0 \overset{m}{\sim} 0 11\boldsymbol{w})) > S(T(11^0 \overset{\ell}{\sim} 0 11^0 \overset{m}{\sim} 0 11\boldsymbol{w})),$$

т.е. x не является максимайзером. ▲

Предложение 3. *Если в битовом представлении* x *содержится подпоследовательность вида* 1100, *то* x *не может быть максимайзером.*

Доказательство. Пусть $x = 100\boldsymbol{w}1$, а $\widehat{x} = 010\boldsymbol{w}1$.

Пусть $\boldsymbol{v} = 0\boldsymbol{w}1$, тогда $\boldsymbol{v}^* = 0\boldsymbol{w}^*1$. Очевидно, что $\boldsymbol{v} + \boldsymbol{v}^* < 1_{p-2}$, так как у обоих слагаемых старший разряд имеет нулевое значение.

Тогда согласно лемме 2 имеем $S(T(x)) > S(T(\widehat{x}))$, и следовательно, x не может быть максимайзером. ▲

Предложение 4. *Число вида* $x = 11^0 \overset{m}{\sim} 0 11^0 \overset{n}{\sim} {}^1 00\boldsymbol{w}$ *не может быть максимайзером.*

Доказательство. Пусть $\widehat{x}$ – циклический сдвиг числа x:

$$\widehat{x} = 10^1 \overset{n-1}{\sim} {}^1 00\boldsymbol{w}11^0 \overset{m}{\sim} {}^0 1.$$

Докажем данное утверждение для $\widehat{x}$. Заметим, что $\widehat{x} = 10\boldsymbol{v}$, где

$$\boldsymbol{v} = 1^0 \overset{n-2}{\sim} {}^1 00\boldsymbol{w}11^0 \overset{m}{\sim} {}^0 1,$$

тогда

$$\boldsymbol{v}^* = 0^1 \overset{m}{\sim} {}^1 00\boldsymbol{w}^*11^0 \overset{n-2}{\sim} {}^1 0.$$

Независимо от соотношения величин $n-2$ и m видно, что $\boldsymbol{v}+\boldsymbol{v}^* < 1_{p-2}$. Следовательно, согласно лемме 2 имеем

$$S(T(10^1 \overbrace{\sim}^{n-1} 1 00\boldsymbol{w}11^0 \overbrace{\sim}^{m} 0 1)) < S(T(00^1 \overbrace{\sim}^{n-1} 1 00\boldsymbol{w}11^0 \overbrace{\sim}^{m} 0 1)),$$

и $\widehat{x}$ – не максимайзер. ▲

Предложение 5. *Число вида* $x = 11^0 \overbrace{\sim}^{m+k} 1 00^1 \overbrace{\sim}^{m} 0 11\boldsymbol{w}$, $m > 1$, *не может быть максимайзером.*

Доказательство. Пусть $\widehat{x}$ – циклический сдвиг числа x:

$$\widehat{x} = 01^0 \overbrace{\sim}^{m-1} 0 11\boldsymbol{w}11^0 \overbrace{\sim}^{m+k} 1 0,$$

докажем данное утверждение для $\widehat{x}$, так как $S(T(x)) = S(T(\widehat{x}))$. Пусть

$$\boldsymbol{v} = {}^0 \overbrace{\sim}^{m-1} 0 11\boldsymbol{w}11^0 \overbrace{\sim}^{m+k} 1 0,$$

тогда

$$\boldsymbol{v}^* = {}^1 \overbrace{\sim}^{m+k+1} 1 00\boldsymbol{w}^* 00^1 \overbrace{\sim}^{m-1} 1.$$

Поскольку для любых положительных m, k справедливо $m+k+1 > m-1$, то $\boldsymbol{v}+\boldsymbol{v}^* > 1_p$. Следовательно, согласно лемме 2 такое x не может быть максимайзером. ▲

3.4. Точное значение ошибки для четного и асимптотика для нечетного случаев.

Теорема 1. *Максимальная ошибка аппроксимации диадическим паттерном достигается при* $x = \left[\frac{2^p}{3}\right]$ *и* $x = \left[\frac{2^{p+1}}{3}\right]$ *при наклонах* $t = \left[\frac{2^p}{3}\right]$ *и* $t = \left[\frac{2^{p+1}}{3}\right]$, *и ее модуль равен* $\frac{p}{6}$ *для четных* p.

Доказательство. Ясно, что $x_1 = 0 \overbrace{\sim}^{p} 1 = \left[\frac{2^p}{3}\right]$, $x_2 = 1 \overbrace{\sim}^{p} 0 = \left[\frac{2^{p+1}}{3}\right]$. Любые другие x не могут быть максимайзерами согласно предложениям 1–5. Заметим, что случай $x = 11\overbrace{\sim}^{n}11\overbrace{\sim}^{m}$ соответствует предложению 2, а $x = 11\overbrace{\sim}^{n}00\overbrace{\sim}^{m}$ – предложению 4.

В силу перестановочности наклона и координаты (см. утверждение 2 в § 2) наклонами для данных максимайзеров являются $t_1 = \left[\frac{2^p}{3}\right]$, $t_2 = \left[\frac{2^{p+1}}{3}\right]$, и для каждого из этих наклонов максимальная ошибка достигается при $x_1 = \left[\frac{2^p}{3}\right]$, $x_2 = \left[\frac{2^{p+1}}{3}\right]$.

Рассчитаем теперь значение максимальной ошибки, к примеру, для случая $x_1 = \left[\frac{2^p}{3}\right]$. Для этого согласно формуле (2) достаточно вычислить сумму по таблице $S(T(x_1))$ и поделить результат на $2^p - 1$. Сумма по одной значимой строке в таблице есть сумма геометрической прогрессии из $\frac{p}{2}$ элементов со знаменателем $q = 4$ и первым членом, равным 1. Нетрудно вычислить, что сумма по одной строке равна $s = \frac{4^{p/2}-1}{3} = \frac{2^p-1}{3}$, тогда для всех $\frac{p}{2}$ значимых строк сумма по таблице равна $\frac{p(2^p-1)}{6}$. В результате максимальная ошибка аппроксимации согласно формуле (4) равна $\frac{p}{6}$.

Аналогичный расчет применим для остальных пар (x, t), где достигается максимальное значение ошибки. ▲

Теорема 2. *Модуль максимальной ошибки аппроксимации диадическим паттерном для нечетных p достигается при $x = 00^1\overbrace{\ldots}^{p-2}1$ или $x = 11^0\overbrace{\ldots}^{p-2}0$, и при $p \to \infty$ сходится к $\dfrac{p}{6} - \dfrac{1}{18}$.*

Доказательство. Согласно предложениям 1–5 для нечетных p максимайзером может быть только число, в битовом представлении которого содержится только одна пара соседствующих единичных или нулевых битов.

Для определения величины этой ошибки при $p \to \infty$ аналогично теореме 1 подсчитаем сумму по таблице. Пусть $x_1 = {}^0\overbrace{\ldots}^{p-2}011$ и $t_1 = {}^0\overbrace{\ldots}^{p-2}011$. Видно, что такое число содержит $m = \dfrac{p-1}{2}$ нулей и $n = \dfrac{p-1}{2} + 1$ единиц.

Для начала вычислим сумму S_1 по первой строке таблицы. Ясно, что ${}^0\overbrace{\ldots}^{p-2}011 = {}^0\overbrace{\ldots}^{p}0 + 1$. Заметим, что сумма по ${}^0\overbrace{\ldots}^{p}0$ является суммой геометрической прогрессии с первым элементом, равным $a_1 = 2$, и знаменателем $q = 4$, и равна $\dfrac{2}{3}(2^{p-1} - 1)$. Тогда $S^1 = \dfrac{2}{3}(2^{p-1} - 1) + 1$.

Заметим, что разность соседних значимых строк Δ_i в таблице образует геометрическую прогрессию, $\Delta_1 = T_3(x_1) - T_1(x_1) = 10$, $\Delta_2 = T_5(x_1) - T_3(x_1) = 1000$. Обобщая, получаем $\Delta_i = 2 \cdot 4^{i-1}$, где $i \in \left[1, \dfrac{p-1}{2} - 1\right]$.

Тогда значение суммы по каждой значимой строке k (индекс пробегает по строкам с нулевым старшим битом) можно представить как сумму по первой строке и добавки к каждой строке, равной соответствующей сумме $S_k = \sum\limits_{i=1}^{k-1} \Delta_i$. Таким образом, сумма по таблице равна

$$S = \left[\frac{p-1}{2} \cdot S^1 + \sum_{k=1}^{\frac{p-1}{2}} S_k\right].$$

Преобразовав данное выражение, получим

$$S = \frac{1}{18}\big(-3p + 2^p(3p-1) - 1\big). \tag{5}$$

Тогда, поделив выражение (5) на $2^p - 1$ согласно формуле (4) и устремляя $p \to \infty$, получим

$$E(x_1, x_2) = \frac{p}{6} - \frac{1}{18},$$

что и требовалось доказать.

Аналогичным образом производится расчет и для всех остальных пар (x, t), где достигается максимальное значение ошибки. ▲

Авторы благодарят Вячеслава Васильева за педантичную вычитку текста статьи.

СПИСОК ЛИТЕРАТУРЫ

1. *Гельфанд И.М., Гиндикин С.Г., Граев М.И.* Избранные задачи интегральной геометрии. М.: Добросвет, 2012.
2. *Brady M.L., Yong W.* Fast Parallel Discrete Approximation Algorithms for the Radon Transform // Proc. 4th Annu. ACM Symp. on Parallel Algorithms and Architectures (SPAA'92). San Diego, CA, USA. June 29–July 1, 1992. P. 91–99. https://doi.org/10.1145/140901.140911

3. *Götz W.A.* Eine Schnelle Diskrete Radon Transformation basierend auf rekursiv definierten Digitalen Geraden. PhD Thesis. Univ. of Innsbruck, Austria, 1993.
4. *Vuillemin J.E.* Fast Linear Hough Transform // Proc. IEEE Int. Conf. on Application Specific Array Processors (ASSAP'94). San Francisco, CA, USA. Aug. 22–24, 1994. P. 1–9. https://doi.org/10.1109/ASAP.1994.331821
5. *Карпенко С.М., Николаев Д.П., Николаев П.П., Постников В.В.* Быстрое преобразование Хафа с управляемой робастностью // Труды Международных конференций "Искусственные интеллектуальные системы" (AIS'04) и "Интеллектуальные САПР" (CAD-2004). Дивноморское, Краснодарский край. 3–10 сентября 2004 г. М.: Физматлит, 2004. Т. 2. С. 303–309.
6. *Press W.H.* Discrete Radon Transform Has an Exact, Fast Inverse and Generalizes to Operations Other than Sums along Lines // Proc. Natl. Acad. Sci. 2006. V. 103. № 51. P. 19249–19254. https://doi.org/10.1073/pnas.0609228103
7. *Götz W.A., Druckmüller H.J.* A Fast Digital Radon Transform—An Efficient Means for Evaluating the Hough Transform // Pattern Recognit. 1996. V. 29. № 4. P. 711–718. https://doi.org/10.1016/0031-3203(96)00015-5
8. *Ershov E., Terekhin A., Nikolaev D., Postnikov V., Karpenko S.* Fast Hough Transform Analysis: Pattern Deviation from Line Segment // Eighth International Conference on Machine Vision (ICMV 2015). Barcelona, Spain. Nov. 19–21, 2015. Proc. SPIE. V. 9875. Article ID 987509 (5 pp.). https://doi.org/10.1117/12.2228852

Карпенко Семен Михайлович
Ершов Егор Иванович
Институт проблем передачи информации
им. А.А. Харкевича РАН, Москва
Московский физико-технический институт
(государственный университет)
simon.karpenko@gmail.com
e.i.ershov@gmail.com

Поступила в редакцию
04.07.2017
После доработки
30.07.2021
Принята к публикации
07.08.2021

Адрес редакции: 127051, Москва, Б. Каретный пер., 19, стр. 1, тел. (495) 650-47-39

Оригинал-макет подготовил *Д.Ю. Ногин*
по контракту с ООО «ИКЦ«АКАДЕМКНИГА»

Москва
ООО «Объединённая редакция»