ISSN 2949-5598

ДИСКРЕТНЫЙ АНАЛИЗ И ИССЛЕДОВАНИЕ ОПЕРАЦИЙ

Том 31 № 1 2024

Новосибирск
Издательство Института математики

Учредители журнала Сибирское отделение РАН
Институт математики им. С. Л. Соболева СО РАН

Журнал включён в базу данных Russian Science Citation Index (RSCI) на платформе Web of Science. Переводы статей на английский язык публикуются в *Journal of Applied and Industrial Mathematics* и доступны по ссылке `www.springer.com/mathematics/journal/11754`.

СИБИРСКОЕ ОТДЕЛЕНИЕ РОССИЙСКОЙ АКАДЕМИИ НАУК
ИНСТИТУТ МАТЕМАТИКИ им. С. Л. СОБОЛЕВА СО РАН

ДИСКРЕТНЫЙ АНАЛИЗ И ИССЛЕДОВАНИЕ ОПЕРАЦИЙ

Выпускается с 1994 г. Научный журнал 4 номера в год

Том 31, № 1 (159) **Январь–март 2024**

СОДЕРЖАНИЕ

НОВОСИБИРСК
ИЗДАТЕЛЬСТВО ИНСТИТУТА МАТЕМАТИКИ

В журнале публикуются оригинальные научные статьи и обзоры теоретической и прикладной направленности по следующим разделам дискретного анализа, исследования операций и информатики:

- дискретная оптимизация
- комбинаторика
- контроль и надёжность дискретных устройств
- математические модели и методы принятия решений
- математическое программирование
- модели экономики
- моделирование процессов управления
- построение и анализ алгоритмов
- синтез и сложность управляющих систем
- теория автоматов
- теория графов
- теория игр и её приложения
- теория кодирования
- теория расписаний и размещений

Адрес редакции:
Институт математики
им. С. Л. Соболева СО РАН,
пр. Академика Коптюга, 4,
630090 Новосибирск, Россия
Телефон: +7 (383) 329–75–79
E-mail: discopr@math.nsc.ru

SIBERIAN BRANCH OF THE RUSSIAN ACADEMY OF SCIENCES
SOBOLEV INSTITUTE OF MATHEMATICS

DISKRETNYI ANALIZ I ISSLEDOVANIE OPERATSII

/DISCRETE ANALYSIS AND OPERATIONS RESEARCH/

Published since 1994 Scientific journal 4 issues per year

Vol. 31, No. 1 (159) **January–March, 2024**

CONTENTS

NOVOSIBIRSK
SOBOLEV INSTITUTE PRESS

In this journal we publish original research papers and survey papers of both theoretical and practical importance on the following topics of discrete analysis, operations research and informatics:

- discrete optimization
- combinatorics
- control and reliability of discrete devices
- decision making models and methods
- mathematical programming
- economic models
- management modeling
- design and analysis of algorithms
- synthesis and complexity of control systems
- automata theory
- graph theory
- game theory and its applications
- coding theory
- theory of scheduling and facility location

Editorial office address:
Sobolev Institute of Mathematics,
4 Acad. Koptyug Avenue,
630090 Novosibirsk, Russia
Phone: +7 (383) 329–75–79
E-mail: discopr@math.nsc.ru

ДИСКРЕТНЫЙ АНАЛИЗ И ИССЛЕДОВАНИЕ ОПЕРАЦИЙ
Январь–март 2024. Т. 31, № 1. С. 5–18

УДК 519.85+517.518.244 DOI: 10.33048/daio.2024.31.779

ВЫПУКЛОЕ ПРОДОЛЖЕНИЕ БУЛЕВОЙ ФУНКЦИИ И ЕГО ПРИЛОЖЕНИЯ

Д. Н. Баротов

Финансовый университет при Правительстве Российской Федерации,
4-й Вешняковский пр-д, 4, 109456 Москва, Россия

E-mail: dnbarotov@fa.ru

Аннотация. Строится выпуклое продолжение произвольной булевой функции на множество $[0,1]^n$. Более того, доказывается, что для любой булевой функции $f(x_1, x_2, \ldots, x_n)$, не имеющей соседних точек на множестве $\operatorname{supp} f$, построенная функция $f_C(x_1, x_2, \ldots, x_n)$ является единственным суммарно максимально выпуклым продолжением на $[0,1]^n$. На базе этого, в частности, конструктивно утверждается, что задача решения произвольной системы булевых уравнений может быть сведена к задаче минимизации функции, любой локальный минимум которой в искомой области является глобальным минимумом, и тем самым для этой задачи проблема локальных минимумов полностью решается. Библиогр. 15.

Ключевые слова: выпуклое продолжение функции, система булевых уравнений, SAT, безусловная оптимизация, булева функция, локальный минимум.

Введение

Системы булевых уравнений являются важным объектом в математике, компьютерных и прикладных науках, появляющимся повсеместно в различном виде [1–5]. В связи с этим, с одной стороны, для таких систем разрабатываются новые направления, методы и алгоритмы, а с другой — совершенствуются существующие направления, методы и алгоритмы решения таких систем. Одно из направлений заключается в том, что, во-первых, система булевых уравнений, заданная над кольцом булевых полиномов, трансформируется в систему уравнений над полем действительных чисел, а во-вторых, трансформированная система сводится либо к задаче численной минимизации соответствующей целевой функции [6–8], либо к задаче MILP или QUBO [9], либо к системе полиномиальных

уравнений, решаемой на множестве целых чисел [1], либо к эквивалентной системе полиномиальных уравнений, решаемой символьными методами [10]. Имеется много способов, позволяющих трансформировать систему булевых уравнений в задачу непрерывной минимизации, поскольку принципиальное отличие таких методов от «переборных» алгоритмов локального поиска — на каждой итерации алгоритма сдвиг по антиградиенту производится по всем переменным одновременно [2, 3, 6–8, 11–14]. Однако, одна из основных проблем, возникающая при применении этих способов, состоит в том, что минимизируемая целевая функция в искомой области может иметь множество локальных минимумов, что значительно усложняет их практическое использование [2, 3, 6–8, 11, 12]. По теореме Д. Н. Баротова полилинейное продолжение булевой функции тоже играет важную роль в том числе и для уменьшения числа локальных минимумов целевой функции [3, 11]. По данной тематике недавно в [11] были найдены явные формы полилинейных продолжений для произвольных функций, определённых на множестве вершин n-мерного единичного куба, произвольного куба и параллелепипеда, и в каждом конкретном случае была доказана единственность соответствующего полилинейного продолжения. С учётом этого в данной работе конструируется выпуклое продолжение произвольной булевой функции на множество $[0,1]^n$. Доказывается, что для любой булевой функции f, не имеющей соседних точек на множестве $\operatorname{supp} f$, построенная функция f_C является не только выпуклым продолжением f на $[0,1]^n$, но и её единственным суммарно максимально выпуклым продолжением на $[0,1]^n$. На основе этого конструктивно утверждается, что задача решения произвольной системы булевых уравнений может быть сведена к задаче минимизации функции, любой локальный минимум которой в искомой области является глобальным минимумом, и тем самым для этой задачи проблема локальных минимумов полностью решается.

1. Определения и обозначения

Пусть $\mathbb{B}^n = \{b = (b_1, b_2, \ldots, b_n) \in \{0,1\}^n\}$ — множество всех двоичных слов (булевых векторов) длины n, $\mathbb{K}^n = \{x = (x_1, x_2, \ldots, x_n) \in [0,1]^n\}$ — n-мерный куб, натянутый на булевы вектора длины n.

Пусть $\mathbb{P}_b^n = \left\{x \in \mathbb{K}^n \mid \sum\limits_{k=1}^{n} (2b_k - 1)x_k > b_1 + b_2 + \cdots + b_n - 1\right\}$ — n-мерная пирамида, прилежащая к вершине $b = (b_1, b_2, \ldots, b_n) \in \mathbb{B}^n$.

Определение 1. Функцию вида $f\colon \mathbb{B}^n \to \mathbb{B}$ назовём *булевой функцией.*

Пусть $\operatorname{supp} f = \{b \in \mathbb{B}^n \mid f(b) = 1\}$ — носитель f, т. е. множество всех булевых векторов, на которых булева функция f принимает значение 1.

Определение 2. Функцию вида $f\colon \mathbb{K}^n \to \mathbb{R}$ назовём *выпуклой* на $\mathbb{K}^n$, если для любых $x, y \in \mathbb{K}^n$ и любого $\alpha \in [0,1]$ выполняется

$$f(\alpha x + (1-\alpha)y) \leqslant \alpha f(x) + (1-\alpha)f(y).$$

Определение 3. Непрерывную функцию вида $f_C\colon \mathbb{K}^n \to \mathbb{R}$ назовём *выпуклым продолжением* булевой функции f на $\mathbb{K}^n$, если f_C на $\mathbb{K}^n$ выпуклая и для любого $(b_1, b_2, \ldots, b_n) \in \mathbb{B}^n$

$$f_C(b_1, b_2, \ldots, b_n) = f(b_1, b_2, \ldots, b_n).$$

Определение 4. Функцию вида $f_{CM}\colon \mathbb{K}^n \to \mathbb{R}$ назовём *суммарно максимально выпуклым продолжением* булевой функции f на $\mathbb{K}^n$, если

1) f_{CM} является выпуклым продолжением f на $\mathbb{K}^n$,

2) имеет место равенство

$$\max_{f_C} \iint\limits_{\mathbb{K}^n} \cdots \int f_C(x_1, x_2, \ldots, x_n)\, dx_1 dx_2 \ldots dx_n = \iint\limits_{\mathbb{K}^n} \cdots \int f_{CM}(x_1, x_2, \ldots, x_n)\, dx_1 dx_2 \ldots dx_n.$$

2. Конструирование выпуклых и суммарно максимально выпуклых продолжений булевых функций

Лемма 1. *Пусть $f_b(x_1, x_2, \ldots, x_n) = x_1^{b_1} \wedge x_2^{b_2} \wedge \cdots \wedge x_n^{b_n}$, $b \in \mathbb{B}^n$, — булева функция. Тогда существует единственная функция $f_C^b(x_1, x_2, \ldots, x_n)$, являющаяся суммарно максимально выпуклым продолжением f_b на $\mathbb{K}^n$.*

Доказательство. Существование. Пусть g_C — произвольное выпуклое продолжение булевой функции f_b на $\mathbb{K}^n$. В этом пункте на основании выпуклости функции g_C оцениваем её значение сверху с помощью неравенства Йенсена [15]. По оценке сверху конструируем новое — суммарно максимально выпуклое — продолжение f_{CM} булевой функции f_b.

Сначала покажем, что для любого $x^* \in \mathbb{K}^n \setminus \mathbb{P}_b^n$ имеем

$$g_C(x_1^*, x_2^*, \ldots, x_n^*) \leqslant 0. \tag{1}$$

В силу выпуклости множества $\mathbb{K}^n \setminus \mathbb{P}_b^n$ условие $x^* \in \mathbb{K}^n \setminus \mathbb{P}_b^n$ означает, что

$$(x_1^*, x_2^*, \ldots, x_n^*) = \sum_{a \in \mathbb{B}^n \setminus \{b\}} \lambda_a^* \cdot (a_1, a_2, \ldots, a_n),$$

где $\lambda_a^* \geqslant 0$ и $\sum\limits_{a \in \mathbb{B}^n \setminus \{b\}} \lambda_a^* = 1$. Тогда в силу неравенства Йенсена [15] имеем

$$g_C(x_1^*, x_2^*, \ldots, x_n^*) = g_C\Big(\sum_{a \in \mathbb{B}^n \setminus \{b\}} \lambda_a^* \cdot (a_1, a_2, \ldots, a_n) \Big) \leqslant$$

$$\leqslant \sum_{a\in\mathbb{B}^n\setminus\{b\}} \lambda_a^*\cdot g_C(a_1,a_2,\dots,a_n) = \sum_{a\in\mathbb{B}^n\setminus\{b\}} \lambda_a^*\cdot f_b(a_1,a_2,\dots,a_n) =$$

$$= \sum_{a\in\mathbb{B}^n\setminus\{b\}} \lambda_a^*\cdot 0 = 0.$$

Далее покажем, что для любого $x^*\in\mathbb{P}_b^n$

$$g_C(x_1^*,x_2^*,\dots,x_n^*) \leqslant 1+\sum_{k=1}^{n}((2b_k-1)x_k^*-b_k). \tag{2}$$

В силу выпуклости множества $\mathbb{P}_b^n$ условие $x^*\in\mathbb{P}_b^n$ означает, что

$$(x_1^*,x_2^*,\dots,x_n^*) = \sum_{\substack{a\in\mathbb{B}^n:\\ \rho(a,b)\leqslant 1}} \lambda_a^*\cdot(a_1,a_2,\dots,a_n),$$

где $\lambda_a^*\geqslant 0$ и $\sum\limits_{\substack{a\in\mathbb{B}^n:\\ \rho(a,b)\leqslant 1}} \lambda_a^* = 1$. Приравняв покоординатно, заметим, что $\lambda_b^* = 1+\sum\limits_{k=1}^{n}((2b_k-1)x_k^*-b_k)$. Тогда в силу неравенства Йенсена [15]

$$g_C(x_1^*,x_2^*,\dots,x_n^*) =$$

$$= g_C\Big(\sum_{\substack{a\in\mathbb{B}^n:\\ \rho(a,b)\leqslant 1}} \lambda_a^*\cdot(a_1,a_2,\dots,a_n)\Big) \leqslant \sum_{\substack{a\in\mathbb{B}^n:\\ \rho(a,b)\leqslant 1}} \lambda_a^*\cdot g_C(a_1,a_2,\dots,a_n) =$$

$$= \sum_{\substack{a\in\mathbb{B}^n:\\ \rho(a,b)=0}} \lambda_a^*\cdot g_C(a_1,a_2,\dots,a_n) + \sum_{\substack{a\in\mathbb{B}^n:\\ \rho(a,b)=1}} \lambda_a^*\cdot g_C(a_1,a_2,\dots,a_n) =$$

$$= \lambda_b^*\cdot g_C(b) + \sum_{\substack{a\in\mathbb{B}^n:\\ \rho(a,b)=1}} \lambda_a^*\cdot 0 = \lambda_b^* = 1+\sum_{k=1}^{n}((2b_k-1)x_k^*-b_k).$$

Объединив (1) и (2), для любого $x\in\mathbb{K}^n$ получим

$$g_C(x_1,x_2,\dots,x_n) \leqslant$$

$$\leqslant \left\{\begin{array}{ll} 0, & \text{если } 1+\sum\limits_{k=1}^{n}((2b_k-1)x_k-b_k)\leqslant 0,\\ 1+\sum\limits_{k=1}^{n}((2b_k-1)x_k-b_k), & \text{если } 1+\sum\limits_{k=1}^{n}((2b_k-1)x_k-b_k)\geqslant 0 \end{array}\right\} =$$

$$= \frac{1}{2}\left(1+\sum_{k=1}^{n}((2b_k-1)x_k-b_k)+\left|1+\sum_{k=1}^{n}((2b_k-1)x_k-b_k)\right|\right). \tag{3}$$

Докажем, что сконструированная оценка наиболее точная, т. е.

$$f_C^b(x_1,x_2,\ldots,x_n)=\frac{1}{2}\left(1+\sum_{k=1}^{n}((2b_k-1)x_k-b_k)+\left|1+\sum_{k=1}^{n}((2b_k-1)x_k-b_k)\right|\right).$$

Для этого осталось проверить справедливость следующих свойств:

1) $f_C^b(a_1,a_2,\ldots,a_n)=f_b(a_1,a_2,\ldots,a_n)$ для любого $a\in\mathbb{B}^n$;
2) функция f_C^b на множестве $\mathbb{K}^n$ непрерывна и выпукла;
3) имеет место равенство

$$\max_{g_C}\iint\limits_{\mathbb{K}^n}\cdots\int g_C(x_1,x_2,\ldots,x_n)\,dx_1dx_2\ldots dx_n=$$
$$=\iint\limits_{\mathbb{K}^n}\cdots\int f_C^b(x_1,x_2,\ldots,x_n)\,dx_1dx_2\ldots dx_n. \quad (4)$$

1. Действительно, для любого $a\in\mathbb{B}^n$

$$f_C^b(a_1,a_2,\ldots,a_n)=\begin{Bmatrix}1, & \text{если } a=b,\\ 0, & \text{если } a\neq b\end{Bmatrix}=f_b(a_1,a_2,\ldots,a_n).$$

2. Непрерывность очевидна, поэтому проверим только выпуклость. Пусть $x,y\in\mathbb{K}^n$, $\alpha\in[0,1]$. Тогда

$$f_C^b(\alpha x+(1-\alpha)y)=\frac{1}{2}\left(1+\sum_{k=1}^{n}((2b_k-1)(\alpha x_k+(1-\alpha)y_k)-b_k)+\right.$$
$$\left.+\left|1+\sum_{k=1}^{n}((2b_k-1)(\alpha x_k+(1-\alpha)y_k)-b_k)\right|\right)=$$
$$=\frac{1}{2}\left(\alpha\left(1+\sum_{k=1}^{n}((2b_k-1)x_k-b_k)\right)+(1-\alpha)\left(1+\sum_{k=1}^{n}((2b_k-1)y_k-b_k)\right)+\right.$$
$$\left.+\left|\alpha\left(1+\sum_{k=1}^{n}((2b_k-1)x_k-b_k)\right)+(1-\alpha)\left(1+\sum_{k=1}^{n}((2b_k-1)y_k-b_k)\right)\right|\right)\leqslant$$
$$\leqslant\frac{1}{2}\left(\alpha\left(1+\sum_{k=1}^{n}((2b_k-1)x_k-b_k)\right)+(1-\alpha)\left(1+\sum_{k=1}^{n}((2b_k-1)y_k-b_k)\right)+\right.$$
$$\left.+\alpha\left|1+\sum_{k=1}^{n}((2b_k-1)x_k-b_k)\right|+(1-\alpha)\left|1+\sum_{k=1}^{n}((2b_k-1)y_k-b_k)\right|\right)=$$
$$=\alpha\cdot\frac{1}{2}\left(1+\sum_{k=1}^{n}((2b_k-1)x_k-b_k)+\left|1+\sum_{k=1}^{n}((2b_k-1)x_k-b_k)\right|\right)+$$

$$+(1-\alpha)\cdot\frac{1}{2}\left(1+\sum_{k=1}^{n}((2b_k-1)y_k-b_k)+\left|1+\sum_{k=1}^{n}((2b_k-1)y_k-b_k)\right|\right)=$$
$$=\alpha f_C^b(x)+(1-\alpha)f_C^b(y).$$

3. Действительно, с одной стороны, из пп. 1 и 2 следует, что f_C^b является одним из выпуклых продолжений функции f_b и тем самым

$$\iint\limits_{\mathbb{K}^n}\cdots\int f_C^b(x_1,x_2,\ldots,x_n)\,dx_1dx_2\ldots dx_n\leqslant$$
$$\leqslant\max_{g_C}\iint\limits_{\mathbb{K}^n}\cdots\int g_C(x_1,x_2,\ldots,x_n)\,dx_1dx_2\ldots dx_n,$$

с другой стороны, из (3) следует, что

$$\max_{g_C}\iint\limits_{\mathbb{K}^n}\cdots\int g_C(x_1,x_2,\ldots,x_n)\,dx_1dx_2\ldots dx_n\leqslant$$
$$\leqslant\iint\limits_{\mathbb{K}^n}\cdots\int f_C^b(x_1,x_2,\ldots,x_n)\,dx_1dx_2\ldots dx_n.$$

Из этого рассуждения получим требуемое равенство (4).

Единственность. Докажем от противного. Пусть существует функция $r\colon\mathbb{K}^n\to\mathbb{R}$, которая также является суммарно максимально выпуклым продолжением функции f_b на $\mathbb{K}^n$, причём $r(x_1^*,x_2^*,\ldots,x_n^*)\neq f_C^b(x_1^*,x_2^*,\ldots,x_n^*)$ для некоторого $(x_1^*,x_2^*,\ldots,x_n^*)\in\mathbb{K}^n$. Тогда, с одной стороны, в силу (4)

$$\iint\limits_{\mathbb{K}^n}\cdots\int(r(x_1,x_2,\ldots,x_n)-f_C^b(x_1,x_2,\ldots,x_n))\,dx_1dx_2\ldots dx_n=0,\quad(5)$$

с другой стороны, в силу (3) для любого $x\in\mathbb{K}^n$

$$r(x_1,x_2,\ldots,x_n)-f_C^b(x_1,x_2,\ldots,x_n)\leqslant 0.\quad(6)$$

Тогда, во-первых, из (6) следует, что

$$d=r(x_1^*,x_2^*,\ldots,x_n^*)-f_C^b(x_1^*,x_2^*,\ldots,x_n^*)<0,$$

во-вторых, в силу непрерывности функции $r-f_C^b$ существует δ-окрестность

$$O_\delta=\{x\in\mathbb{K}^n\mid||(x_1,x_2,\ldots,x_n)-(x_1^*,x_2^*,\ldots,x_n^*)||<\delta\},\quad\delta>0,$$

такая, что для любого $x\in O_\delta$

$$r(x_1,x_2,\ldots,x_n)-f_C^b(x_1,x_2,\ldots,x_n)<\frac{d}{2},$$

а значит,

$$\iint\limits_{\mathbb{K}^n}\cdots\int \big(r(x_1,x_2,\ldots,x_n)-f_C^b(x_1,x_2,\ldots,x_n)\big)\,dx_1dx_2\ldots dx_n=$$
$$=\iint\limits_{\mathbb{K}^n\setminus O_\delta}\cdots\int \big(r(x_1,x_2,\ldots,x_n)-f_C^b(x_1,x_2,\ldots,x_n)\big)\,dx_1dx_2\ldots dx_n+$$
$$+\iint\limits_{O_\delta}\cdots\int \big(r(x_1,x_2,\ldots,x_n)-f_C^b(x_1,x_2,\ldots,x_n)\big)\,dx_1dx_2\ldots dx_n\leqslant$$
$$\leqslant\iint\limits_{\mathbb{K}^n\setminus O_\delta}\cdots\int 0\cdot dx_1dx_2\ldots dx_n+\iint\limits_{O_\delta}\cdots\int\frac{d}{2}\cdot dx_1dx_2\ldots dx_n=$$
$$=0+\frac{d}{2}\iint\limits_{O_\delta}\cdots\int dx_1dx_2\ldots dx_n\leqslant\frac{d}{2}\cdot\frac{1}{2^n}\cdot\frac{\pi^n}{\Gamma(\frac{n}{2}+1)}\cdot\delta^n<0;$$

противоречие с равенством (5). Лемма 1 доказана.

Теперь на основе леммы 1, теоремы 2 из [11] и формы СДНФ конструируем выпуклое продолжение произвольной булевой функции f в следующем виде:

$$f_C(x_1,x_2,\ldots,x_n)=\sum_{b\in\mathbb{B}^n}f(b)f_C^b(x_1,x_2,\ldots,x_n).$$

Далее докажем, что для любой булевой функции $f(x_1,x_2,\ldots,x_n)$, не имеющей соседних точек на множестве $\operatorname{supp}f$, функция $f_C(x_1,x_2,\ldots,x_n)$ является не только выпуклым продолжением функции $f(x_1,x_2,\ldots,x_n)$ на $\mathbb{K}^n$, но и единственным суммарно максимально выпуклым продолжением этой функции $f(x_1,x_2,\ldots,x_n)$ на $\mathbb{K}^n$.

Теорема 1. *Если булева функция f не имеет соседних вершин на множестве* $\operatorname{supp}f$, *то*

$$f_C(x_1,x_2,\ldots,x_n)=\sum_{b\in\operatorname{supp}f}f_C^b(x_1,x_2,\ldots,x_n)$$

является единственным суммарно максимально выпуклым продолжением f на $\mathbb{K}^n$.

Доказательство. 1. Сначала заметим, что функция f_C выпукла по построению, как сумма некоторых непрерывно выпуклых функций на множестве $\mathbb{K}^n$. Действительно, пусть $x,y\in\mathbb{K}^n$, $\alpha\in[0,1]$. Тогда в силу леммы 1 имеем

$$f_C(\alpha x+(1-\alpha)y)=\sum_{b\in\mathbb{B}^n}f(b)f_C^b(\alpha\cdot x+(1-\alpha)\cdot y)\leqslant$$

$$\leqslant \sum_{b\in\mathbb{B}^n} f(b)\big(\alpha f_C^b(x)+(1-\alpha)f_C^b(y)\big) = \alpha\sum_{b\in\mathbb{B}^n} f(b)f_C^b(x) + \\ + (1-\alpha)\sum_{b\in\mathbb{B}^n} f(b)f_C^b(y) = \alpha f_C(x)+(1-\alpha)f_C(y).$$

Проверим, что $f_C = f$ на $\mathbb{B}^n$. Действительно, в силу леммы 1 для любого $a\in\mathbb{B}^n$ имеем

$$f_C(a_1,a_2,\ldots,a_n) = \sum_{b\in\mathbb{B}^n} f(b)f_C^b(a_1,a_2,\ldots,a_n) = \\ = f(a)f_C^a(a_1,a_2,\ldots,a_n) + \sum_{b\in\mathbb{B}^n\setminus\{a\}} f(b)f_C^b(a_1,a_2,\ldots,a_n) = \\ = f(a)\cdot 1 + \sum_{b\in\mathbb{B}^n\setminus\{a\}} f(b)\cdot 0 = f(a_1,a_2,\ldots,a_n).$$

2. Пусть $p_C\colon \mathbb{K}^n\to\mathbb{R}$ — произвольное выпуклое продолжение булевой функции f на $\mathbb{K}^n$. Докажем оценку

$$p_C(x_1,x_2,\ldots,x_n)\leqslant f_C(x_1,x_2,\ldots,x_n),\quad x\in\mathbb{K}^n.$$

С этой целью рассмотрим два случая.

Случай 1. Пусть $\operatorname{supp} f=\varnothing$. Тогда для любого $x^*\in\mathbb{K}^n$ имеем

$$p_C(x_1^*,x_2^*,\ldots,x_n^*) = p_C((1-x_1^*)\cdot 0 + x_1^*\cdot 1, x_2^*,\ldots,x_n^*) = \\ = p_C((1-x_1^*)\cdot 0 + x_1^*\cdot 1, (1-x_1^*)x_2^*+x_1^*x_2^*,\ldots,(1-x_1^*)x_n^*+x_1^*x_n^*) \leqslant \\ \leqslant (1-x_1^*)p_C(0,x_2^*,\ldots,x_n^*) + x_1^*p_C(1,x_2^*,\ldots,x_n^*)\leqslant \\ \leqslant (1-x_1^*)(1-x_2^*)p_C(0,0,\ldots,x_n^*) + (1-x_1^*)x_2^*p_C(0,1,\ldots,x_n^*) + \\ + x_1^*(1-x_2^*)p_C(1,0,\ldots,x_n^*) + x_1^*x_2^*p_C(1,1,\ldots,x_n^*)\leqslant \\ \leqslant\ldots\leqslant \sum_{b\in\mathbb{B}^n} p_C(b)\prod_{k=1}^n((2b_k-1)x_k^*+1-b_k) = \\ = \sum_{b\in\mathbb{B}^n} f(b)\prod_{k=1}^n((2b_k-1)x_k^*+1-b_k) = \sum_{b\in\mathbb{B}^n} 0\cdot\prod_{k=1}^n((2b_k-1)x_k^*+1-b_k) = \\ = 0 = \sum_{b\in\mathbb{B}^n} f(b)f_C^b(x_1^*,x_2^*,\ldots,x_n^*) = f_C(x_1^*,x_2^*,\ldots,x_n^*).$$

Случай 2. Пусть $\operatorname{supp} f=\{s_k=(s_{k1},s_{k2},\ldots,s_{kn})\}_{k=1}^m\neq\varnothing$. Относительно $x^*\in\mathbb{K}^n$ рассмотрим ещё два случая.

Случай 2.1. Пусть $x^*\in\mathbb{P}^n_{s_1}\cup\mathbb{P}^n_{s_2}\cup\cdots\cup\mathbb{P}^n_{s_m}$, т. е. $x^*\in\mathbb{P}^n_{s_k}$ для некоторого единственного $k\in\{1,2,\ldots,m\}$, так как элементы $\operatorname{supp} f$ не соседние

и, следовательно, множества $\mathbb{P}^n_{s_1}, \mathbb{P}^n_{s_2}, \ldots, \mathbb{P}^n_{s_m}$ попарно не пересекаются. В силу выпуклости $\mathbb{P}^n_{s_k}$ условие $x^* \in \mathbb{P}^n_{s_k}$ означает, что

$$(x_1^*, x_2^*, \ldots, x_n^*) = \sum_{\substack{a \in \mathbb{B}^n : \\ \rho(a, s_k) \leqslant 1}} \lambda_a^* \cdot (a_1, a_2, \ldots, a_n),$$

где $\lambda_a^* \geqslant 0$ и $\sum\limits_{\substack{a \in \mathbb{B}^n : \\ \rho(a, s_k) \leqslant 1}} \lambda_a^* = 1$. Приравняв покоординатно, заметим, что $\lambda_{s_k}^* = 1 + \sum\limits_{i=1}^{n} (2s_{ki} - 1)x_i^* - s_{ki}$. В силу неравенства Йенсена [15] имеем

$$\begin{aligned}
&p_C(x_1^*, x_2^*, \ldots, x_n^*) = \\
&= p_C\Big(\sum_{\substack{a \in \mathbb{B}^n : \\ \rho(a, s_k) \leqslant 1}} \lambda_a^* \cdot (a_1, a_2, \ldots, a_n) \Big) \leqslant \sum_{\substack{a \in \mathbb{B}^n : \\ \rho(a, s_k) \leqslant 1}} \lambda_a^* \cdot p_C(a_1, a_2, \ldots, a_n) = \\
&= \sum_{\substack{a \in \mathbb{B}^n : \\ \rho(a, s_k) = 0}} \lambda_a^* \cdot p_C(a_1, a_2, \ldots, a_n) + \sum_{\substack{a \in \mathbb{B}^n : \\ \rho(a, s_k) = 1}} \lambda_a^* \cdot p_C(a_1, a_2, \ldots, a_n) = \\
&\qquad = \lambda_{s_k}^* \cdot p_C(s_k) + \sum_{\substack{a \in \mathbb{B}^n : \\ \rho(a, s_k) = 1}} \lambda_a^* \cdot 0 = \lambda_{s_k}^* = 1 + \sum_{i=1}^{n} (2s_{ki} - 1)x_i^* - s_{ki}.
\end{aligned}$$

Случай 2.2. Пусть $x^* \in \mathbb{K}^n \setminus \big(\mathbb{P}^n_{s_1} \cup \mathbb{P}^n_{s_2} \cup \cdots \cup \mathbb{P}^n_{s_m}\big)$. В силу выпуклости множества $\mathbb{K}^n \setminus \big(\mathbb{P}^n_{s_1} \cup \mathbb{P}^n_{s_2} \cup \cdots \cup \mathbb{P}^n_{s_m}\big)$ это означает, что

$$(x_1^*, x_2^*, \ldots, x_n^*) = \sum_{a \in \mathbb{B}^n \setminus \operatorname{supp} f} \lambda_a^* \cdot (a_1, a_2, \ldots, a_n),$$

где $\lambda_a^* \geqslant 0$ и $\sum\limits_{a \in \mathbb{B}^n \setminus \operatorname{supp} f} \lambda_a^* = 1$. В силу неравенства Йенсена [15] имеем

$$\begin{aligned}
&p_C(x_1^*, x_2^*, \ldots, x_n^*) = \\
&= p_C\Big(\sum_{a \in \mathbb{B}^n \setminus \operatorname{supp} f} \lambda_a^* \cdot (a_1, a_2, \ldots, a_n) \Big) \leqslant \sum_{a \in \mathbb{B}^n \setminus \operatorname{supp} f} \lambda_a^* \cdot p_C(a_1, a_2, \ldots, a_n) = \\
&\qquad = \sum_{a \in \mathbb{B}^n \setminus \operatorname{supp} f} \lambda_a^* \cdot f(a_1, a_2, \ldots, a_n) = \sum_{a \in \mathbb{B}^n \setminus \operatorname{supp} f} \lambda_a^* \cdot 0 = 0.
\end{aligned}$$

Таким образом, для любого $x \in \mathbb{K}^n$ получаем

$$\begin{aligned}
&p_C(x_1, x_2, \ldots, x_n) \leqslant \\
&\leqslant \left\{ \begin{array}{ll} 1 + \sum\limits_{i=1}^{n} ((2s_{ki} - 1)x_i - s_{ki}), & \text{если } x \in \mathbb{P}^n_{s_k},\ k = 1, 2, \ldots, m, \\ 0 & \text{иначе} \end{array} \right\} =
\end{aligned}$$

$$= \begin{Bmatrix} f_C^{s_k}(x_1, x_2, \ldots, x_n), & \text{если } x \in \mathbb{P}^n_{s_k},\ k = 1, 2, \ldots, m, \\ 0 & \text{иначе} \end{Bmatrix} =$$

$$= \sum_{k=1}^{m} f_C^{s_k}(x_1, x_2, \ldots, x_n) = \sum_{b \in \operatorname{supp} f} f_C^b(x_1, x_2, \ldots, x_n) =$$

$$= f_C(x_1, x_2, \ldots, x_n),$$

так как множества $\mathbb{P}^n_{s_1}, \mathbb{P}^n_{s_2}, \ldots, \mathbb{P}^n_{s_m}$ попарно не пересекаются и $\{x \in \mathbb{K}^n \mid f_C^{s_k}(x) \neq 0\} = \mathbb{P}^n_{s_k}$ для любого $k \in \{1, 2, \ldots, m\}$ в силу леммы 1. Исходя из доказанного в п. 1 и рассуждения, аналогичного рассуждению п. 3 в доказательстве леммы 1, получим

$$\max_{p_C} \iint\limits_{\mathbb{K}^n} \cdots \int p_C(x_1, x_2, \ldots, x_n)\, dx_1 dx_2 \ldots dx_n =$$

$$= \iint\limits_{\mathbb{K}^n} \cdots \int f_C(x_1, x_2, \ldots, x_n)\, dx_1 dx_2 \ldots dx_n.$$

Единственность суммарно максимально выпуклого продолжения булевой функции f доказывается аналогично тому, как это сделано в доказательстве леммы 1. Теорема 1 доказана.

Замечание. На основе полученных результатов задача решения системы булевых уравнений вида

$$p_k(x_1, x_2, \ldots, x_n) = 0, \quad k = 1, 2, \ldots, m,$$

может быть трансформирована в соответствующую систему выпуклых уравнений вида

$$p_{Ck}(x_1, x_2, \ldots, x_n) = 0, \quad k = 1, 2, \ldots, m, \tag{7}$$

где $p_{Ck}(x) = \sum\limits_{b \in \mathbb{B}^n} p_k(b) f_C^b(x)$ — неотрицательное выпуклое продолжение булевой функции p_k. В свою очередь, задача решения системы (7) может быть сведена к задаче минимизации выпуклой функции

$$t f_C(x_1, x_2, \ldots, x_n) = \sum_{k=1}^{m} p_{Ci}(x_1, x_2, \ldots, x_n).$$

Заключение

В настоящей работе построено выпуклое продолжение произвольной булевой функции на множество $[0, 1]^n$. Доказано, что для любой булевой функции f, не имеющей соседних точек в своём носителе $\operatorname{supp} f$, построенная функция f_C является не просто выпуклым продолжением f на $[0, 1]^n$, но и её единственным суммарно максимально выпуклым продолжением на $[0, 1]^n$. На основе этого, в частности, показано, что задача

решения произвольной системы булевых уравнений может быть сведена к задаче минимизации функции, любой локальный минимум которой в искомой области является глобальным минимумом, и тем самым упомянутая выше проблема локальных минимумов полностью решена.

Финансирование работы

Исследование выполнено за счёт бюджета Финансового университета при Правительстве Российской Федерации. Дополнительных грантов на проведение или руководство этим исследованием получено не было.

Конфликт интересов

Автор заявляет, что у него нет конфликта интересов.

Литература

1. **Abdel-Gawad A. H., Atiya A. F., Darwish N. M.** Solution of systems of Boolean equations via the integer domain // Inf. Sci. 2010. V. 180, No. 2. P. 288–300. DOI: 10.1016/j.ins.2009.09.010.
2. **Barotov D. N., Barotov R. N.** Polylinear transformation method for solving systems of logical equations // Mathematics. 2022. V. 10, No. 6. Paper ID 918. 10 p. DOI: 10.3390/math10060918.
3. **Barotov D. N.** Target function without local minimum for systems of logical equations with a unique solution // Mathematics. 2022. V. 10, No. 12. Paper ID 2097. 8 p. DOI: 10.3390/math10122097.
4. **Armario J. A.** Boolean functions and permanents of Sylvester Hadamard matrices // Mathematics. 2021. V. 9, No. 2. Paper ID 177. 8 p. DOI: 10.3390/math9020177.
5. **Valiant L. G.** The complexity of computing the permanent // Theor. Comput. Sci. 1979. V. 8, No. 2. P. 189–201. DOI: 10.1016/0304-3975(79)90044-6.
6. **Файзуллин Р. Т., Дулькейт В. И., Огородников Ю. Ю.** Гибридный метод поиска приближенного решения задачи 3-выполнимость, ассоциированной с задачей факторизации // Тр. Ин-та математики и механики. 2013. Т. 19, № 2. С. 285–294.
7. **Gu J.** Global optimization for satisfiability (SAT) problem // IEEE Trans. Knowl. Data Eng. 1994. V. 6, No. 3. P. 361–381. DOI: 10.1109/69.334864.
8. **Gu J., Gu Q., Du D.** On optimizing the satisfiability (SAT) problem // J. Comput. Sci. Technol. 1999. V. 14, No. 1. P. 1–17. DOI: 10.1007/BF02952482.
9. **Pakhomchik A. I., Voloshinov V. V., Vinokur V. M., Lesovik G. B.** Converting of Boolean expression to linear equations, inequalities and QUBO penalties for cryptanalysis // Algorithms. 2022. V. 15, No. 2. Paper ID 33. 22 p. DOI: 10.3390/a15020033.
10. **Barotov D. N., Barotov R. N., Soloviev V., Feklin V., Muzafarov D., Ergashboev T., Egamov Kh.** The development of suitable inequalities and their application to systems of logical equations // Mathematics. 2022. V. 10, No. 11. Paper ID 1851. 9 p. DOI: 10.3390/math10111851.

11. **Баротов Д. Н., Баротов Р. Н.** Полилинейные продолжения некоторых дискретных функций и алгоритм их нахождения // Вычисл. методы и программирование. 2023. Т. 24, вып. 1. С. 10–23. DOI: 10.26089/NumMet.v24r102.
12. **Barotov D. N., Osipov A., Korchagin S., Pleshakova E., Muzafarov D., Barotov R. N., Serdechnyi D.** Transformation method for solving system of Boolean algebraic equations // Mathematics. 2021. V. 9, No. 24. Paper ID 3299. 12 p. DOI: 10.3390/math9243299.
13. **Owen G.** Multilinear extensions of games // Manage. Sci. 1972. V. 18, No. 5-2. P. 64–79. DOI: 10.1287/mnsc.18.5.64.
14. **Wittmann D. M., Krumsiek J., Saez-Rodriguez J., Lauffenburger D. A., Klamt S., Theis F. J.** Transforming Boolean models to continuous models: Methodology and application to T-cell receptor signaling // BMC Syst. Biol. 2009. V. 3, No. 1. Paper ID 98. 21 p. DOI: 10.1186/1752-0509-3-98.
15. **Jensen J. L. W. V.** Sur les fonctions convexes et les inégalités entre les valeurs moyennes // Acta Math. 1906. V. 30. P. 175–193. [French]. DOI: 10.1007/BF02418571.

Баротов Достонжон Нумонжонович

Статья поступила
17 июля 2023 г.
После доработки —
4 августа 2023 г.
Принята к публикации
22 сентября 2023 г.

DISKRETNYI ANALIZ I ISSLEDOVANIE OPERATSII
/DISCRETE ANALYSIS AND OPERATIONS RESEARCH/
January–March 2024. Vol. 31, No. 1. P. 5–18

UDC 519.85+517.518.244 DOI: 10.33048/daio.2024.31.779

CONVEX CONTINUATION OF A BOOLEAN FUNCTION AND ITS APPLICATIONS

D. N. Barotov

Financial University under the Government of the Russian Federation,
4 Chetvyortyi Veshnyakovskii Passage, 109456 Moscow, Russia
E-mail: dnbarotov@fa.ru

Abstract. A convex continuation of an arbitrary Boolean function to the set $[0,1]^n$ is constructed. Moreover, it is proved that for any Boolean function $f(x_1, x_2, \ldots, x_n)$ that has no neighboring points on the set $\operatorname{supp} f$, the constructed function $f_C(x_1, x_2, \ldots, x_n)$ is the only totally maximally convex continuation to $[0,1]^n$. Based on this, in particular, it is constructively stated that the problem of solving an arbitrary system of Boolean equations can be reduced to the problem of minimizing a function any local minimum of which in the desired region is a global minimum, and thus for this problem the problem of local minima is completely resolved. Bibliogr. 15.

Keywords: convex continuation of a function, system of Boolean equations, SAT, global optimization, Boolean function, local minimum.

References

1. **A. H. Abdel-Gawad, A. F. Atiya,** and **N. M. Darwish,** Solution of systems of Boolean equations via the integer domain, *Inf. Sci.* **180** (2), 288–300 (2010), DOI: 10.1016/j.ins.2009.09.010.
2. **D. N. Barotov** and **R. N. Barotov,** Polylinear transformation method for solving systems of logical equations, *Mathematics* **10** (6), ID 918 (2022), DOI: 10.3390/math10060918.
3. **D. N. Barotov,** Target function without local minimum for systems of logical equations with a unique solution, *Mathematics* **10** (12), ID 2097 (2022), DOI: 10.3390/math10122097.
4. **J. A. Armario,** Boolean functions and permanents of Sylvester Hadamard matrices, *Mathematics* **9** (2), ID 177 (2021), DOI: 10.3390/math9020177.

English transl.: *Journal of Applied and Industrial Mathematics* **18** (1), 1–9 (2024), DOI: 10.1134/S1990478924010010.

5. **L. G. Valiant,** The complexity of computing the permanent, *Theor. Comput. Sci.* **8** (2), 189–201 (1979), DOI: 10.1016/0304-3975(79)90044-6.
6. **R. T. Faizullin, V. I. Dul'keit,** and **Yu. Yu. Ogorodnikov,** Hybrid method for the approximate solution of the 3-satisfiability problem associated with the factorization problem, *Tr. Inst. Mat. Mekh.* **19** (2), 285–294 (2013) [Russian].
7. **J. Gu,** Global optimization for satisfiability (SAT) problem, *IEEE Trans. Knowl. Data Eng.* **6** (3), 361–381 (1994), DOI: 10.1109/69.334864.
8. **J. Gu, Q. Gu,** and **D. Du,** On optimizing the satisfiability (SAT) problem, *J. Comput. Sci. Technol.* **14** (1), 1–17 (1999), DOI: 10.1007/BF02952482.
9. **A. I. Pakhomchik, V. V. Voloshinov, V. M. Vinokur,** and **G. B. Lesovik,** Converting of Boolean expression to linear equations, inequalities and QUBO penalties for cryptanalysis, *Algorithms* **15** (2), ID 33 (2022), DOI: 10.3390/a15020033.
10. **D. N. Barotov, R. N. Barotov, V. Soloviev, V. Feklin, D. Muzafarov, T. Ergashboev,** and **Kh. Egamov,** The development of suitable inequalities and their application to systems of logical equations, *Mathematics* **10** (11), ID 1851 (2022), DOI: 10.3390/math10111851.
11. **D. N. Barotov** and **R. N. Barotov,** Polylinear continuations of some discrete functions and an algorithm for finding them, *Vychisl. Metody Program.* **24** (1), 10–23 (2023) [Russian], DOI: 10.26089/NumMet.v24r102.
12. **D. N. Barotov, A. Osipov, S. Korchagin, E. Pleshakova, D. Muzafarov, R. N. Barotov,** and **D. Serdechnyi,** Transformation method for solving system of Boolean algebraic equations, *Mathematics* **9** (24), ID 3299 (2021), DOI: 10.3390/math9243299.
13. **G. Owen,** Multilinear extensions of games, *Manage. Sci.* **18** (5-2), 64–79 (1972), DOI: 10.1287/mnsc.18.5.64.
14. **D. M. Wittmann, J. Krumsiek, J. Saez-Rodriguez, D. A. Lauffenburger, S. Klamt,** and **F. J. Theis,** Transforming Boolean models to continuous models: Methodology and application to T-cell receptor signaling, *BMC Syst. Biol.* **3** (1), ID 98 (2009), DOI: 10.1186/1752-0509-3-98.
15. **J. L. W. V. Jensen,** Sur les fonctions convexes et les inégalités entre les valeurs moyennes, *Acta Math.* **30**, 175–193 (1906) [French], DOI: 10.1007/BF02418571.

Dostonjon N. Barotov

Received July 17, 2023
Revised August 4, 2023
Accepted September 22, 2023

ДИСКРЕТНЫЙ АНАЛИЗ И ИССЛЕДОВАНИЕ ОПЕРАЦИЙ
Январь–март 2024. Т. 31, № 1. С. 19–34

УДК 512.534.1 DOI: 10.33048/daio.2024.31.783

ОПРЕДЕЛЯЕМОСТЬ ОТНОШЕНИЙ ПОЛУГРУППАМИ ИЗОТОННЫХ ПРЕОБРАЗОВАНИЙ

А. А. Клюшин[1, a], *И. Б. Кожухов*[2, 3, b], *Д. Ю. Манилов*[4, c], *А. В. Решетников*[2, d]

[1] Cadence Design Systems,
Bld. 1, Penrose Dock, Penrose Quay, Cork, T23 KW81, Ireland
[2] Московский институт электронной техники,
пл. Шокина, 1, 124498 Москва, Россия
[3] Московский гос. университет им. М. В. Ломоносова,
Ленинские горы, 1, 119991 Москва, Россия
[4] НПЦ «Электронные вычислительно-информационные системы»,
ул. Конструктора Лукина, 14, с. 14, 124460 Зеленоград, Москва, Россия
E-mail: [a] xkreed@gmail.com, [b] kozhuhov_i_b@mail.ru, [c] thdi@ro.ru,
[d] a_reshetnikov@hush.com

Аннотация. В 1961 г. Л. М. Глускин доказал, что множество X с заданным на нём нетривиальным квазипорядком ρ с точностью до изоморфизма или антиизоморфизма определяется полугруппой $T_\rho(X)$ изотонных преобразований множества X (т. е. преобразований, сохраняющих ρ). Позже Л. М. Попова доказала аналогичное утверждение для полугруппы $P_\rho(X)$ частичных изотонных преобразований, причём ρ — необязательно квазипорядок, а любое нетривиальное рефлексивное или антирефлексивное бинарное отношение на X. В настоящей работе доказано, что полугруппа $B_\rho(X)$ изотонных бинарных отношений (многозначных отображений) при тех же самых ограничениях на отношение ρ также определяет данное отношение ρ с точностью до изоморфизма или антиизоморфизма. Кроме того, для каждого из условий $T_\rho(X) = T(X)$, $P_\rho(X) = P(X)$, $B_\rho(X) = B(X)$ авторами охарактеризованы n-арные отношения ρ, удовлетворяющие данному условию. Библиогр. 8.

Ключевые слова: полугруппа бинарных отношений, изотонное преобразование.

Введение

Пусть ρ — бинарное отношение, заданное на множестве X. Отображение $\alpha\colon X \to X$ *сохраняет* отношение ρ, если $(x, y) \in \rho \to (x\alpha, y\alpha) \in \rho$ для

всех $x, y \in X$. Обозначим через $T_\rho(X)$ множество всех преобразований множества X, сохраняющих отношение $\rho \subseteq X^2$:

$$T_\rho(X) = \{\alpha \in T(X) \mid \alpha \text{ сохраняет } \rho\}.$$

Здесь через $T(X)$ обозначена *симметрическая полугруппа* на множестве X, т. е. множество всех преобразований $X \to X$.

Пусть заданы бинарные отношения $\rho \subseteq X^2$ и $\rho' \subseteq X'^2$. Предположим, что их полугруппы $T_\rho(X)$ и $T_{\rho'}(X')$ изоморфны друг другу. Как в таком случае могут быть связаны между собой ρ и ρ'? Данный вопрос, а также ряд похожих вопросов, изучались в ряде работ разных авторов.

Л. М. Глускин рассмотрел в [1] обозначенную проблему в случае, когда ρ — отношение квазипорядка, а ρ' — произвольное рефлексивное отношение. Им было доказано, что при определённых ограничениях на квазипорядок ρ его полугруппа $T_\rho(X)$ определяет ρ с точностью до изоморфизма или антиизоморфизма [1, лемма 4]. Эти ограничения очень простые: $\rho \neq \Delta_X$ и $\rho \neq \nabla_X$, где $\Delta_X = \{(x, x) \mid x \in X\}$ — отношение равенства элементов множества X, а $\nabla_X = X \times X = \{(x, y) \mid x, y \in X\}$ — универсальное отношение на X.

В случае $\rho = \Delta_X$ или $\rho = \nabla_X$ полугруппа $T_\rho(X)$, вообще говоря, не определяет отношение ρ, поскольку в обоих этих случаях $T_\rho(X) = T(X)$.

Л. М. Попова получила в [2] результаты, аналогичные результатам Глускина, для полугрупп $P_\rho(X)$ и $P_{\rho'}(X')$, состоящих из всех частичных преобразований множеств X и X', сохраняющих бинарные отношения ρ и ρ' соответственно. Из леммы 2 в [2] следует, что если ρ — рефлексивное или антирефлексивное бинарное отношение, а ρ' — произвольное отношение и выполнено условие

$$\rho \notin \{\varnothing, \Delta_X, \nabla_X\}, \tag{1}$$

то полугруппа $P_\rho(X)$ определяет данное отношение ρ с точностью до изоморфизма или антиизоморфизма.

Бинарные отношения, удовлетворяющие условию (1), будем далее называть *нетривиальными*.

Поскольку множество X с заданным на нём бинарным отношением ρ — это то же самое, что ориентированный граф $G = (X, \rho)$ (возможно, содержащий петли) с множеством вершин X и множеством рёбер ρ, вопрос о выявлении связей между отношением ρ и полугруппами преобразований, сохраняющих ρ, сводится к вопросу о связи между графом G и его производными полугруппами: полугруппой его полных эндоморфизмов $\operatorname{End} G$, частичных эндоморфизмов $\operatorname{PEnd} G$ и т. д. Результаты, описывающие связи такого рода, приведены в обзорной работе В. А. Молчанова [4].

Одним из основных результатов настоящей работы является доказательство того, что множество X с заданным на нём нетривиальным рефлексивным или антирефлексивным бинарным отношением ρ с точностью до изоморфизма или антиизоморфизма определяется полугруппой $B_\rho(X)$ бинарных отношений, сохраняющих ρ. Соответствующую теорему можно считать продолжением исследований, представленных теоремой 1 в [1] и теоремами 1 и 2 в [2].

Для заданного на множестве X n-арного отношения ρ также можно ввести полугруппы $T_\rho(X)$, $P_\rho(X)$ и $B_\rho(X)$. Их можно рассматривать как полугруппы *изотонных* преобразований — преобразований, сохраняющих информацию о структуре ориентированного гиперграфа (X, ρ). В данной работе мы находим условия выполнения каждого из равенств $T_\rho(X) = T(X)$, $P_\rho(X) = P(X)$ и $B_\rho(X) = B(X)$, интересных тем, что удовлетворяющие им n-арные отношения представляют тривиальный случай решения проблемы, о которой было сказано выше: найти условия на гиперграф (X, ρ), при которых он определяется какой-либо из своих полугрупп изотонных преобразований.

Авторы придерживаются определений и обозначений из теории полугрупп, содержащихся в монографии [5]. В доказательстве теоремы 1 используются соображения, связанные с тем, что многие значимые подмножества полугрупп $T(X)$, $T_\rho(X)$, $P_\rho(X)$ и других полугрупп преобразований могут быть заданы формулами узкого исчисления предикатов [1)]. Сведения по математической логике, связанные с проблематикой данной статьи, можно почерпнуть, например, из книги [7].

1. Основные определения и обозначения

Пусть X — множество. *Квазипорядком* на множестве X называется рефлексивное и транзитивное бинарное отношение на X. Пусть $T(X)$ обозначает полугруппу всех преобразований множества X, т. е. отображений $X \to X$ с умножением, осуществляемым слева направо: $x(\alpha\beta) = (x\alpha)\beta$ для всех $x \in X$, $\alpha, \beta \in T(X)$. Если на X задано бинарное отношение ρ, то будем называть отображение $\alpha\colon X \to X$ *изотонным* (или *сохраняющим отношение* ρ), если для любых $x, y \in X$

$$(x, y) \in \rho \to (x\alpha, y\alpha) \in \rho.$$

Множество $T_\rho(X)$ всех изотонных отображений $X \to X$ является подполугруппой полугруппы $T(X)$.

Обозначим через Δ_X отношение равенства на множестве X, а через ∇_X — универсальное отношение, т. е. $\Delta_X = \{(x, x) \mid x \in X\}$, $\nabla_X =$

[1)] В другой терминологии — формулами логики первого порядка; формулами заданной сигнатуры.

$X \times X$. Будем часто опускать индекс X и писать Δ, ∇ вместо Δ_X, ∇_X. Бинарное отношение σ на множестве X называется *тривиальным,* если $\sigma \in \{\varnothing, \Delta, \nabla\}$, в противном случае σ называется *нетривиальным.* Бинарное отношение σ называется *антирефлексивным*, если $\sigma \cap \Delta = \varnothing$. Для отношения σ полагаем $\sigma^{-1} = \{(y, x) \mid (x, y) \in \sigma\}$.

Частичным преобразованием множества X называется отображение $\alpha\colon X_1 \to X$, где $X_1 \subseteq X$. Множество X_1 — это *область определения* отображения α: $X_1 = \operatorname{dom}\alpha$. *Образом* частичного преобразования α называется множество $\operatorname{im}\alpha = \{x\alpha \mid x \in \operatorname{dom}\alpha\}$. Обозначим через $P(X)$ множество всех частичных преобразований множества X с умножением, определённым правилом

$$x(\alpha\beta) = (x\alpha)\beta, \quad \text{если } x \in \operatorname{dom}\alpha \text{ и } x\alpha \in \operatorname{dom}\beta.$$

Очевидно, что $P(X)$ — полугруппа с нулём, причём нулём является пустое отображение — отображение, у которого область определения — пустое множество.

Полугруппу бинарных отношений на X обозначим через $B(X)$. Умножение бинарных отношений обычное: $(x, y) \in \rho\sigma$ в том и только том случае, если $(x, z) \in \rho$, $(z, y) \in \sigma$ для некоторого $z \in X$. Бинарное отношение можно рассматривать как многозначное отображение: $x \mapsto y$, если $(x, y) \in \sigma$. Для $\sigma \in B(X)$ обозначим

$$\operatorname{dom}\sigma = \{x \in X \mid \exists\, y \in X\colon (x, y) \in \sigma\},$$
$$\operatorname{im}\sigma = \{y \in X \mid \exists\, x \in X\colon (x, y) \in \sigma\}.$$

Нетрудно проверить, что $T(X) \subseteq P(X) \subseteq B(X)$, причём каждая предыдущая из этих полугрупп является подполугруппой следующей.

Для частичных отображений $\alpha \in P(X)$ понятие сохранения бинарного отношения ρ может быть определено различными не эквивалентными способами (см. [3]). Мы примем следующее определение, согласующееся с определением из [2]. Будем говорить, что частичное преобразование $\alpha \in P(X)$ *сохраняет заданное на X бинарное отношение* ρ, если для любых $x, y \in \operatorname{dom}\alpha$

$$(x, y) \in \rho \to (x\alpha, y\alpha) \in \rho.$$

Легко проверяется, что множество $P_\rho(X)$ всех сохраняющих ρ частичных преобразований α образует подполугруппу полугруппы $P(X)$. Аналогично будем говорить, что бинарное отношение $\alpha \in B(X)$ *сохраняет* бинарное отношение ρ, если для любых $x, y, x', y' \in X$ имеем

$$(x, y) \in \rho \wedge (x, x'), (y, y') \in \alpha \to (x', y') \in \rho.$$

Пусть $B_\rho(X)$ — множество всех бинарных отношений, сохраняющих ρ. Непосредственно проверяется, что $B_\rho(X)$ — подполугруппа полугруппы $B(X)$, кроме того, $B_\rho(X) \cap P(X) = P_\rho(X)$ и $B_\rho(X) \cap T(X) = T_\rho(X)$.

Пару (X, ρ), где X — множество, а ρ — бинарное отношение на X, будем называть *структурой*. Пусть (X, ρ) и (X', ρ') — две структуры. Взаимно однозначное отображение $f\colon X \to X'$ называется *изоморфизмом* структур (X, ρ) и (X', ρ'), если для любых $x, y \in X$ верно

$$(x, y) \in \rho \Leftrightarrow (xf, yf) \in \rho'.$$

Отображение $f\colon X \to X'$ — *антиизоморфизм* структур (X, ρ) и (X', ρ'), если f — изоморфизм (X, ρ) и (X', ρ'^{-1}). Очевидно, что всякий изоморфизм $f\colon (X, \rho) \to (X', \rho')$ индуцирует изоморфизмы полугрупп:

$$T_\rho(X) \cong T_{\rho'}(X), \quad P_\rho(X) \cong P_{\rho'}(X), \quad B_\rho(X) \cong B_{\rho'}(X). \tag{2}$$

Например, отображение

$$B_\rho(X) \to B_{\rho'}(X)\colon \alpha \mapsto \alpha' = \{(xf, yf) \mid (x, y) \in \alpha\}$$

представляет собой изоморфизм полугрупп, индуцированный изоморфизмом f множеств X и X'. Далее, нетрудно видеть, что

$$T_\rho(X) = T_{\rho^{-1}}(X), \quad P_\rho(X) = P_{\rho^{-1}}(X), \quad B_\rho(X) = B_{\rho^{-1}}(X), \tag{3}$$

поэтому антиизоморфизм $g\colon (X, \rho) \to (X', \rho')$ также индуцирует изоморфизмы (2). Таким образом, отношение ρ на множестве X, вообще говоря, не определяется однозначно полугруппой $T_\rho(X)$, или $P_\rho(X)$, или $B_\rho(X)$. Однако, как отмечалось во введении, для многих отношений ρ имеет место определяемость с точностью до изоморфизма или антиизоморфизма.

2. Предварительные рассмотрения

Пусть S — полугруппа. Подмножество $M \subseteq S^n$ назовём L-*подмножеством*, если существует формула $\varphi(x_1, \ldots, x_n)$ логики первого порядка такая, что для всех $a_1, \ldots, a_n \in S$ соотношение $(a_1, \ldots, a_n) \in M$ выполняется тогда и только тогда, когда высказывание $\varphi(a_1, \ldots, a_n)$ истинно. L-подмножество, являющееся подполугруппой, называется L-*подполугруппой*. В настоящей работе доказано, что ряд интересных подмножеств полугрупп преобразований являются L-подмножествами, а значит, сохраняются при изоморфизмах. Если предикат $u(x_1, \ldots, x_n)$ на S эквивалентен формуле логики первого порядка $\xi(x_1, \ldots, x_n)$, то этот факт будем записывать так: $u \equiv \xi$.

Определим некоторые специальные элементы полугруппы $B(X)$: константа $c_x = X \times \{x\}$; частичное тождественное отображение (для $Y \subseteq X$) $i_Y = \{(y, y) \mid y \in Y\}$; локальная единица $i_x = \{(x, x)\}$; атомарное отношение $j_{xy} = \{(x, y)\}$. Множество $C(X)$ всех констант является правым идеалом полугруппы $T(X)$, а если ρ рефлексивно, то $C(X) \subseteq T_\rho(X)$. Положим

$$PI(X) = \{i_Y \mid Y \subseteq X\}, \quad I(X) = \{i_x \mid x \in X\}, \quad J(X) = \{j_{xy} \mid x, y \in X\}.$$

Нетрудно проверить, что

$$B_{\varnothing}(X) = B_{\nabla}(X) = B(X), \quad T_{\varnothing}(X) = T_{\Delta}(X) = T(X),$$

но при $|X| > 1$ $B_{\Delta}(X) \neq B(X)$, а $B_{\Delta}(X) = P(X)$.

Следующая лемма описывает случай, когда $\rho \in \{\varnothing, \Delta, \nabla\}$. Её доказательство является фактически небольшим видоизменением доказательства леммы 1 в [2]. Также отметим, что ниже утверждение (1) леммы 1 будет обобщено на n-арные отношения (см. теорему 3(1)).

Лемма 1. *Пусть $|X| > 1$. Тогда*
(1) *$B_\rho(X) = B(X)$ в том и только том случае, когда $\rho = \varnothing$ или $\rho = \nabla$;*
(2) *$B_\rho(X) = P(X)$ в том и только том случае, когда $\rho = \Delta$.*

Доказательство. Достаточность очевидна, докажем необходимость.

(1) Пусть $B_\rho(X) = B(X)$. Если $\rho \neq \varnothing$ и $\rho \neq \nabla$, то существуют $x, y, u, v \in X$ такие, что $(x, y) \in \rho$ и $(u, v) \notin \rho$. Рассмотрим бинарное отношение $\alpha = \{(x, u), (y, v)\}$. Нетрудно проверить, что $\alpha \in B(X) \setminus B_\rho(X)$.

(2) Пусть $B_\rho(X) = P(X)$. Если $\rho \not\subseteq \Delta$, то $(x, y) \in \rho$ для некоторых $x \neq y$. Рассмотрим бинарное отношение $\alpha = \{(x, x), (x, y)\}$. Нетрудно проверить, что $\alpha \in B_\rho(X) \setminus P(X)$. Пусть $\rho \subset \Delta$ (т. е. $\rho \subseteq \Delta$ и $\rho \neq \Delta$). Возьмём $x \in X$ такое, что $(x, x) \notin \rho$. Тогда для $\beta = \{(x, y)\}$ имеем $\beta \in P(X) \setminus B_\rho(X)$. Лемма 1 доказана.

Следующая лемма является видоизменением леммы 2 из [2].

Лемма 2. *Пусть ρ — нетривиальное рефлексивное бинарное отношение на множестве X, а ρ' — произвольное бинарное отношение на этом множестве. Тогда равенство $B_\rho(X) = B_{\rho'}(X)$ имеет место в том и только том случае, когда $\rho = \rho'$ или $\rho = \rho'^{-1}$.*

Доказательство. Достаточность очевидна ввиду равенств (3). Докажем необходимость.

Пусть $B_\rho(X) = B_{\rho'}(X)$. Если ρ' тривиально, то ρ также тривиально по лемме 1, что противоречит условию. Тем самым, ρ' нетривиально.

Докажем, что ρ' рефлексивно. Возьмём любой элемент $x \in X$. Так как ρ рефлексивно, то $(x, x) \in \rho$. Так как $\rho' \neq \varnothing$, существует пара $(u, v) \in \rho'$. Положим $\alpha = \{(u, x), (v, x)\}$. Очевидно, что $\alpha \in B_\rho(X)$. Тогда $\alpha \in B'_\rho(X)$. Это означает, что $(x, x) \in \rho'$. Таким образом, ρ' рефлексивно.

Докажем, что

$$\rho \cap \rho^{-1} = \rho' \cap \rho'^{-1}. \tag{4}$$

Ясно, что достаточно доказать включение $\rho' \cap \rho'^{-1} \subseteq \rho \cap \rho^{-1}$. Пусть $(x_1, y_1) \in \rho' \cap \rho'^{-1}$. Если $x_1 = y_1$, то $(x_1, y_1) \in \Delta$, а значит, $(x_1, y_1) \in \rho \cap \rho^{-1}$. Далее будем считать, что $x_1 \neq y_1$. Пусть $(x, y) \in \rho \cap \rho^{-1}$. Положим $\alpha =$

$\{(x_1,x),(y_1,y)\}$. Очевидно, что $\alpha \in B_\rho(X)$, следовательно, $\alpha \in B_{\rho'}(X)$, а значит, $(x_1,y_1) \in \rho \cap \rho^{-1}$.

Рассмотрим случай, когда ρ симметрично. Пусть $(x_1,y_1) \in \rho'$. Докажем, что $(x_1,y_1) \in \rho$. Если $(y_1,x_1) \in \rho'$, то $(x_1,y_1) \in \rho' \cap \rho'^{-1} = \rho \cap \rho^{-1}$, откуда $(x_1,y_1) \in \rho$. Пусть $(y_1,x_1) \notin \rho'$. Положим $\alpha = \{(x_1,y_1),(y_1,x_1)\}$. Так как ρ рефлексивно и симметрично, то $\alpha \in B_\rho(X)$. Значит, $\alpha \in B_{\rho'}(X)$, что противоречит соотношениям $(x_1,y_1) \in \rho'$, $(y_1,x_1) \notin \rho'$. Таким образом, доказано, что $\rho' \subseteq \rho$. Ввиду симметричности отношения ρ из (4) вытекает, что $\rho \subseteq \rho'$. Следовательно, $\rho = \rho'$.

Осталось рассмотреть случай, когда ρ несимметрично. В этом случае найдётся пара элементов $x, y \in X$ таких, что

$$(x,y) \in \rho, \quad (y,x) \notin \rho. \tag{5}$$

Далее доказательство разбивается на несколько случаев.

(а) $(x,y),(y,x) \notin \rho'$. Положим $\alpha = \{(x,y),(y,x)\}$. Тогда ввиду (5) $\alpha \in B_{\rho'}(X) \setminus B_\rho(X)$, что противоречит равенству $B_{\rho'}(X) = B_\rho(X)$.

(б) $(x,y),(y,x) \in \rho'$. Тогда $(x,y) \in \rho' \cap \rho'^{-1} = \rho \cap \rho^{-1}$, откуда $(y,x) \in \rho$, что противоречит условию.

(в) $(x,y) \in \rho', (y,x) \notin \rho'$. Возьмём любые $u, v \in X$ и положим $\alpha = \{(u,x),(v,y)\}$. Тогда $(u,v) \in \rho \leftrightarrow \alpha \in B_\rho(X)$. Ввиду (5) $(u,v) \in \rho' \leftrightarrow \alpha \in B_{\rho'}(X)$. Так как $B_\rho(X) = B_{\rho'}(X)$, то $(u,v) \in \rho \leftrightarrow (u,v) \in \rho'$. Следовательно, $\rho = \rho'$.

(г) $(x,y) \notin \rho', (y,x) \in \rho'$. Как и в случае (в), возьмём любые элементы $u, v \in X$ и положим $\alpha = \{(u,x),(v,y)\}$. Очевидно, что $\alpha \in B_{\rho'}(X) \leftrightarrow (u,v) \in \rho'^{-1}(X)$, а ввиду (5) $\alpha \in B_\rho(X) \leftrightarrow (u,v) \in \rho$. Так как $B_\rho(X) = B_{\rho'}(X)$, то $\rho = \rho'$. Лемма 2 доказана.

3. L-подмножества полугрупп $B(X)$ и $B_\rho(X)$

В этом разделе построим логические формулы, выделяющие определённые подмножества в полугруппе бинарных отношений. Не все из полученных здесь результатов будут использованы для доказательства основной теоремы. Мы их приводим, так как считаем, что они представляют самостоятельный интерес.

Пусть S — полугруппа с нулём. Рассмотрим её подмножества

$$\begin{aligned} \mathrm{al}(S) &= \{\sigma \in S \mid \forall\, \alpha \in S\ (\sigma\alpha = 0 \to \alpha = 0)\}, \\ \mathrm{ar}(S) &= \{\sigma \in S \mid \forall\, \alpha \in S\ (\alpha\sigma = 0 \to \alpha = 0)\}. \end{aligned} \tag{6}$$

Элементы из $\mathrm{al}(S)$ будем называть *левыми*, а элементы из $\mathrm{ar}(S)$ — *правыми неделителями нуля*. Если S — полугруппа с единицей, то эти множества непусты, так как $1 \in \mathrm{al}(S) \cap \mathrm{ar}(S)$. Нетрудно проверить, что $\mathrm{al}(S)$ и $\mathrm{ar}(S)$ — подполугруппы полугруппы S, а $S \setminus \mathrm{al}(S)$ и $S \setminus \mathrm{ar}(S)$ — соответственно левый и правый идеалы.

Утверждение 1. *Полугруппа $T(X)$ является L-подполугруппой полугруппы $P(X)$. Кроме того, для произвольного бинарного отношения ρ на множестве X полугруппа $T_\rho(X)$ является L-подполугруппой полугруппы $P_\rho(X)$.*

Доказательство. Ввиду равенства (6) достаточно доказать, что

$$T(X) = \mathrm{ar}(P(X)), \quad T_\rho(X) = \mathrm{ar}(P_\rho(X)).$$

Пусть $\alpha \in \mathrm{ar}(P(X))$. Предположим, что $\mathrm{dom}\,\alpha \neq X$. Возьмём любой элемент $x \in X \setminus \mathrm{dom}\,\alpha$. Локальная единица i_x такова, что $i_x \in P(X)$, $i_x \neq 0$ и $i_x\alpha = 0$; противоречие с тем, что $\alpha \in \mathrm{ar}(P(X)$. Следовательно, $\mathrm{dom}\,\alpha = X$, а значит, $\alpha \in T(X)$. Наоборот, если $\alpha \in T(X)$, то для любого $\beta \neq 0$ имеем $\mathrm{dom}\,\beta\alpha = \mathrm{dom}\,\beta \neq \varnothing$, а значит, $\alpha \in \mathrm{ar}(P(X))$.

Очевидно, что для любого бинарного отношения ρ на множестве X имеет место соотношение $i_x \in P_\rho(X)$, поэтому предыдущие рассуждения остаются верными для доказательства равенства $T_\rho(X) = \mathrm{ar}(P_\rho(X))$. Утверждение 1 доказано.

Рассмотрим вновь произвольную полугруппу S с нулём. Для элемента $a \in S$ определим его *правый аннулятор* $r(a)$ и *левый аннулятор* $l(a)$ равенствами

$$r(a) = \{b \in S \mid ab = 0\}, \quad l(a) = \{b \in S \mid ba = 0\}.$$

Очевидно, что $r(a)$ и $l(a)$ — левый и правый идеалы полугруппы S. Пусть $\mathcal{R} = \{r(a) \mid a \neq 0\}$, $\mathcal{L} = \{l(a) \mid a \neq 0\}$. Заметим, что $\mathcal{R}$ и $\mathcal{L}$ частично упорядочены отношением включения $\subseteq$. Условие $r(a_1) \subseteq r(a_2)$ можно выразить формулой логики первого порядка:

$$r(a_1) \subseteq r(a_2) \equiv \forall\, b \in S\ (a_1 b = 0 \to a_2 b = 0).$$

Аналогично логической формулой можно записать условие $l(a_1) \subseteq l(a_2)$. Тот факт, что $r(a)$, $a \neq 0$, — максимальный элемент в $\mathcal{R}$, также записывается логической формулой:

$$a \neq 0 \wedge \forall\, a' \neq 0\ (\forall\, b\ (ab = 0 \to a'b = 0) \to \forall\, b\ (a'b = 0 \to ab = 0)). \quad (7)$$

Аналогично формулой логики первого порядка записывается максимальность в $\mathcal{L}$ левого аннулятора $l(a)$, $a \neq 0$:

$$a \neq 0 \wedge \forall\, a' \neq 0\ ((\forall\, b\ (ba = 0 \to ba' = 0) \to (\forall\, b\ (ba' = 0 \to ba = 0)). \quad (8)$$

Пусть $S = B(X)$. Нетрудно проверить, что для $\alpha \in B(X)$ имеем

$$r(\alpha) = \{\beta \mid \mathrm{dom}\,\beta \cap \mathrm{im}\,\alpha = \varnothing\}, \quad l(\alpha) = \{\beta \mid \mathrm{im}\,\beta \cap \mathrm{dom}\,\alpha = \varnothing\},$$

поэтому для любого $\alpha \neq 0$

$$r(\alpha) \text{ максимален в } \mathcal{R} \Leftrightarrow |\,\mathrm{im}\,\alpha| = 1, \quad (9)$$

$$l(\alpha) \text{ максимален в } \mathcal{L} \Leftrightarrow |\,\mathrm{dom}\,\alpha| = 1. \quad (10)$$

Для произвольного бинарного отношения ρ на множестве X положим $J_\rho(X) = J(X) \cap B_\rho(X)$.

Лемма 3. *Для любого бинарного отношения ρ на множестве X множество $I(X)$ всех локальных единиц и множество $J_\rho(X)$ являются L-подмножествами полугруппы $B_\rho(X)$.*

Доказательство. Нетрудно видеть, что утверждения (9) и (10) верны не только при $S = B(X)$, но и при $S = B_\rho(X)$. Ввиду (9) и (10) локальные единицы i_x таковы, что правый аннулятор $r(i_x)$ максимален в $\mathcal{R}$, а левый аннулятор $l(i_x)$ максимален в $\mathcal{L}$. Максимальность правого аннулятора $r(a)$ для $a \neq 0$ отмечается формулой (7), а левого аннулятора — формулой (8). Формулы (7) и (8) являются формулами логики первого порядка. Очевидно, что формулы (7) и (8), взятые вместе, определяют множество $J_\rho(X)$, поэтому $J_\rho(X)$ — L-подмножество. Нетрудно видеть, что $i_x \in B_\rho(X)$ для любого $x \in X$ (и любого бинарного отношения ρ) и $i_x = j_{xx}$. Локальные единицы i_x выделяются среди элементов множества $J_\rho(X)$ свойством идемпотентности: $\alpha^2 = \alpha$ для $\alpha = i_x$ и $\alpha^2 \neq \alpha$ для $\alpha = j_{xy}$ при $x \neq y$. Следовательно, I — L-подмножество полугруппы $B_\rho(X)$. Лемма 3 доказана.

Утверждение 2. *Для любого бинарного отношения ρ на множестве X полугруппа $P_\rho(X)$ является L-подполугруппой полугруппы $B_\rho(X)$.*

Доказательство. Очевидно, что для отношения $\alpha \in B_\rho(X)$ имеет место $\alpha \in P_\rho(X)$ в том и только том случае, когда α не содержит одновременно двух пар вида (x, y), (x, z), где $y \neq z$. Проверим, что

$$\alpha \in P_\rho(X) \equiv \forall\, \beta \in I(X)\ (\beta\alpha = 0 \vee \beta\alpha \in J_\rho(X)). \tag{11}$$

Действительно, пусть $\alpha \in B_\rho(X) \setminus P_\rho(X)$. Тогда существуют $x, y, z \in X$ такие, что $(x, y), (x, z) \in \alpha$ и $y \neq z$. Имеем $i_x\alpha \ni (x, y), (x, z)$, поэтому $i_x\alpha \neq 0$ и $i_x\alpha \notin J_\rho(X)$, т. е. правая часть (11) ложна. Пусть $\alpha \in P_\rho(X)$, а $\beta \in I(X)$ таково, что $\beta\alpha \neq 0$. Имеем $\beta = i_x$ при некотором x, поэтому $|\operatorname{dom}(\beta\alpha)| = 1$. Если $\beta\alpha \notin J_\rho(X)$, то $|\operatorname{im}(\beta\alpha)| \geqslant 2$, т. е. $(x, y), (x, z) \in \beta\alpha$ для некоторых $y, z \in X$, $y \neq z$. Очевидно, что $\beta\alpha = (\{x\} \times X) \cap \alpha$. Отсюда $(x, y), (x, z) \in \beta\alpha$, что противоречит условию $\alpha \in P_\rho(X)$. Таким образом, доказано (11), а так как $I(X)$ и $J_\rho(X)$ — L-подмножества (по лемме 3), то $P_\rho(X)$ — L-подполугруппа. Утверждение 2 доказано.

4. Теорема Глускина для многозначных отображений

Перед тем как перейти к доказательству теоремы, сделаем

Замечание 1. Понятие бинарного отношения на множестве A, как подмножества $\sigma \subseteq A \times A$, можно обобщить, считая для произвольных

множеств A и B бинарным отношением между элементами этих множеств произвольное подмножество $\sigma \subseteq A \times B$. Умножение $\sigma\tau$ отношений $\sigma \subseteq A \times B$ и $\tau \subseteq B \times C$ осуществляется обычным образом: $\sigma\tau \subseteq A \times C$, причём $(a, c) \in \sigma\tau$ в том и только том случае, когда $(a, b) \in \sigma$ и $(b, c) \in \tau$ для некоторого $b \in B$ (см. [8]). В этом смысле всякое отображение $A \to B$, частичное отображение $A_1 \to B$ и многозначное отображение $A \to 2^B$ являются бинарными отношениями — подмножествами множества $A \times B$. При этом произведение отображений совпадает с произведением бинарных отношений.

Теорему Глускина [1, теорема 1] можно переформулировать так: если $\varphi\colon T_{\rho_1}(X_1) \to T_{\rho_2}(X_2)$ — изоморфизм полугрупп, где ρ_1 — квазипорядок на X_1, а ρ_2 — нетривиальное рефлексивное бинарное отношение на X_2, то φ имеет вид $\varphi\colon \alpha \mapsto f^{-1}\alpha f$, где $f\colon X_1 \to X_2$ — изоморфизм или антиизоморфизм структуры (X_1, ρ_1) на структуру (X_2, ρ_2):

$$\begin{array}{ccc} X_1 & \xrightarrow{\ f\ } & X_2 \\ {\scriptstyle\alpha}\downarrow & & \downarrow{\scriptstyle\alpha\varphi} \\ X_1 & \xrightarrow[\ f\ ]{} & X_2 \end{array}.$$

Отметим, что правую часть равенства $\alpha\varphi = f^{-1}\alpha f$ можно понимать как произведение бинарных отношений $f^{-1} \subseteq X_2 \times X_1$, $\alpha \subseteq X_1 \times X_1$, $f \subseteq X_1 \times X_2$, а левую часть — как бинарное отношение $\alpha\varphi \subseteq X_2 \times X_2$.

Теорема 1. *Пусть ρ_1 — нетривиальное рефлексивное или антирефлексивное отношение на множестве X_1, ρ_2 — произвольное нетривиальное отношение на множестве X_2. Полугруппы $B_{\rho_1}(X_1)$ и $B_{\rho_2}(X_2)$ изоморфны тогда и только тогда, когда структура (X_1, ρ_1) изоморфна одной из структур (X_2, ρ_2) или $\left(X_2, \rho_2^{-1}\right)$. При этом всякий изоморфизм φ полугруппы $B_{\rho_1}(X_1)$ на $B_{\rho_2}(X_2)$ имеет вид*

$$\alpha \in B_{\rho_1}(X_1) \mapsto \alpha\varphi = f^{-1}\alpha f \in B_{\rho_2}(X_2),$$

где f — изоморфизм или антиизоморфизм структуры (X_1, ρ_1) на (X_2, ρ_2).

Доказательство. Достаточность очевидна, докажем необходимость. Пусть имеется изоморфизм полугрупп $\varphi\colon B_{\rho_1}(X_1) \to B_{\rho_2}(X_2)$. Поскольку $P_{\rho_1}(X_1)$ и $P_{\rho_2}(X_2)$ являются L-полугруппами, изоморфизм φ индуцирует изоморфизм $\varphi'\colon P_{\rho_1}(X_1) \to P_{\rho_2}(X_2)$. По теоремам Поповой [2, теоремы 1, 2] существует изоморфизм или антиизоморфизм f структур (X_1, ρ_1) и (X_2, ρ_2) такой, что $\alpha\varphi' = f^{-1}\alpha f$. Проверим, что $\alpha\varphi = f^{-1}\alpha f$.

Рассмотрим изоморфизм полугрупп $\psi\colon B_{\rho_2}(X_2) \to B_{\rho_1}(X_1)$, который индуцирован изоморфизмом структур $f^{-1}\colon (X_2, \rho_2) \to (X_1, \rho_1)$. Композиция $\varphi\psi = \xi$ даёт автоморфизм полугруппы $B_{\rho_1}(X_1)$, который действует тождественно на подполугруппе $P_{\rho_1}(X_1)$. Докажем, что ξ действует тождественно на всей полугруппе $B_{\rho_1}(X_1)$.

Пусть $\sigma \in B_{\rho_1}(X_1)$ и $\sigma\xi = \tau$. Покажем, что $\sigma = \tau$, для чего проверим включение $\sigma \subseteq \tau$. Обратное включение $\tau \subseteq \sigma$ доказывается аналогично. Пусть $\sigma \not\subseteq \tau$, т. е. существует $(a, b) \in \sigma \setminus \tau$. Рассмотрим частичные константы $i_a = \{(a, a)\}$ и $i_b = \{(b, b)\}$, которые, очевидно, принадлежат $PT_{\rho_1}(X_1)$. Заметим, что бинарное отношение $i_a\sigma i_b = \{(a, b)\}$ непусто и $i_a\sigma i_b \in PT_{\rho_1}(X_1)$. Тогда $i_a\sigma i_b = i_a\sigma i_b\xi = i_a\tau i_b \neq \varnothing$, а это противоречит тому факту, что $(a, b) \notin \tau$. Следовательно, $\sigma \subseteq \tau$. Аналогично $\tau \subseteq \sigma$, откуда следует $\sigma = \tau$.

Тем самым показано, что автоморфизм ξ действует тождественно на всём $B_{\rho_1}(X_1)$. Это и означает справедливость равенства $\alpha\varphi = f^{-1}\alpha f$. Теорема 1 доказана.

5. Полугруппы, сохраняющие n-арное отношение

Перейдём к рассмотрению ситуации, когда на множестве X задано n-арное отношение ρ. *Диагональ* Δ и *универсальное отношение* ∇ определяются аналогично бинарному случаю:

$$\Delta = \{(x, x, \ldots, x) \in X^n \mid x \in X\}, \quad \nabla = X^n.$$

Элементы $\overline{x} = (x_1, \ldots, x_n) \in X^n$ будем рассматривать как отображения $\overline{x}\colon \{1, 2, \ldots, n\} \to X\colon i \mapsto x_i$ $(i = 1, 2, \ldots, n)$. Будем говорить, что отношение $\rho \subseteq X^n$ удовлетворяет условию $(*)$, если

$$\forall\, \overline{x}, \overline{y} \in X^n \ (\overline{x} \in \rho \wedge \ker \overline{y} \supseteq \ker \overline{x} \to \overline{y} \in \rho).$$

Нетрудно видеть, что в случае $n = 2$ для $\overline{x} = (x_1, x_2)$ имеем

$$\ker \overline{x} = \begin{cases} \nabla_{\{1,2\}} & \text{при } x_1 = x_2, \\ \Delta_{\{1,2\}} & \text{при } x_1 \neq x_2, \end{cases}$$

поэтому в случае $n = 2$ условию $(*)$ удовлетворяют только бинарные отношения $\varnothing$, Δ_X и ∇_X, в случае $n > 2$ имеются и другие отношения с этим свойством.

Для дальнейшего понадобится ещё одно хорошо известное утверждение, доказательство которого не приводим. Напомним, что умножение отображений осуществляем слева направо, т. е. $x(\alpha\beta) = (x\alpha)\beta$.

Лемма 4. *Пусть A, B, C — произвольные множества. Тогда*
(1) *для отображений $\alpha\colon A \to C$ и $\beta\colon A \to B$ существует $\gamma\colon B \to C$ такое, что $\alpha = \beta\gamma$, в том и только том случае, когда $\ker\alpha \supseteq \ker\beta$;*
(2) *для отображений $\alpha\colon B \to A$ и $\beta\colon C \to A$ существует $\gamma\colon B \to C$ такое, что $\alpha = \gamma\beta$, в том и только том случае, когда $\operatorname{im}\alpha \subseteq \operatorname{im}\beta$.*

Определим полугруппы $T_\rho(X)$, $P_\rho(X)$ и $B_\rho(X)$ в случае n-арного отношения ρ на множестве X. Для $\overline{x} = (x_1, \ldots, x_n) \in X^n$ и $\alpha \in T(X)$ полагаем $\overline{x}\alpha = (x_1\alpha, \ldots, x_n\alpha)$. Если $\alpha \in P(X)$ и $x_1, \ldots, x_n \in \operatorname{dom}\alpha$, также полагаем $\overline{x}\alpha = (x_1\alpha, \ldots, x_n\alpha)$. В этих обозначениях

$$T_\rho(X) = \{\alpha \in T(X) \mid \forall\, \overline{x} \in X^n\ (\overline{x} \in \rho \to \overline{x}\alpha \in \rho\},$$
$$P_\rho(X) = \{\alpha \in P(X) \mid \forall\, \overline{x} \in (\operatorname{dom} X)^n\ (\overline{x} \in \rho \to \overline{x}\alpha \in \rho\},$$
$$B_\rho(X) = \{\alpha \in B(X) \mid \forall\, x_1, \ldots, x_n, x'_1, \ldots, x'_n \in X$$
$$((x_1, \ldots, x_n) \in \rho \wedge (x_1, x'_1), \ldots, (x_n, x'_n) \in \alpha \to (x'_1, \ldots, x'_n) \in \rho)\}.$$

Следующая теорема даёт необходимые и достаточные условия выполнения равенств $T_\rho(X) = T(X)$ и $P_\rho(X) = P(X)$ в n-арном случае.

Теорема 2. *Пусть ρ — n-арное отношение на множестве X. Следующие условия эквивалентны:*
(1) $T_\rho(X) = T(X)$;
(2) $P_\rho(X) = P(X)$;
(3) *ρ удовлетворяет условию* $(*)$.

Доказательство. (1) $\Rightarrow$ (3) Пусть $\overline{x} \in \rho$, $\overline{y} \in X^n$ и $\ker\overline{y} \supseteq \ker\overline{y}$. По лемме 4(2) существует отображение $\alpha \in T(X)$ такое, что $\overline{x}\alpha = \overline{y}$. Так как $T_\rho(X) = T(X)$, то $\alpha \in T_\rho(X)$, поэтому $\overline{y} \in \rho$. Таким образом, ρ удовлетворяет условию $(*)$.

(2) $\Rightarrow$ (3) Рассуждая, как выше, предположим, что $\overline{x} \in \rho$, $\overline{y} \in X^n$ и $\ker\overline{y} \supseteq \ker\overline{x}$. По лемме 4(2) $\overline{x}\alpha = \overline{y}$ для некоторого $\alpha \in T(X)$. Так как $T(X) \subseteq P(X)$, то $\alpha \in P(X)$. По условию $P(X) = P_\rho(X)$, поэтому $\alpha \in P_\rho(X)$. Следовательно, $\overline{y} \in \rho$.

(3) $\Rightarrow$ (2) Пусть $\alpha \in P(X)$ и $\overline{x} \in \rho \cap ((\operatorname{dom}\alpha)^n$. Так как $x_1, \ldots, x_n \in \operatorname{dom}\alpha$, то $\overline{x}\alpha$ существует. Очевидно, что $\ker(\overline{x}\alpha) \supseteq \ker\overline{x}$, поэтому из условия $(*)$ получаем $\overline{\alpha} \in \rho$. Следовательно, $\alpha \in P_\rho(X)$.

(3) $\Rightarrow$ (1) Пусть $\alpha \in T(X)$. Тогда $\alpha \in P(X)$. Ввиду доказанного (3) $\Rightarrow$ (2) имеем равенство $P_\rho(X) = P(X)$. Следовательно, $\alpha \in P_\rho(X)$. Так как $\alpha \in T(X)$, то $\alpha \in T_\rho(X)$.

Теорема 3. *Пусть ρ — n-арное отношение на множестве X. Тогда*
(1) *$B_\rho(X) = B(X)$ в том и только том случае, если $\rho \in \{\varnothing, \nabla\}$;*
(2) *если $T(X) \subseteq B_\rho(X)$, то ρ удовлетворяет условию* $(*)$.

Доказательство. (1) Достаточность очевидна, докажем необходимость. Пусть $\rho \neq \varnothing$ и $\rho \neq \nabla$. Тогда существуют $\overline{x} \in \rho$, $\overline{y} \notin \rho$. Очевидно, что для бинарного отношения $\alpha = \{(x_1, y_1), \ldots, (x_n), y_n)\}$ имеет место соотношение $\alpha \notin B_\rho(X)$.

(2) Пусть ρ не удовлетворяет условию $(*)$. Тогда существуют такие $\overline{x}, \overline{y} \in X^n$, что $\overline{x} \in \rho$, $\ker \overline{y} \supseteq \ker \overline{x}$, но $\overline{y} \notin \rho$. По лемме 4(2) существует $\alpha \in T(X)$ такое, что $\overline{y} = \overline{x}\alpha$. Очевидно, что $\alpha \notin B_\rho(X)$.

Замечание 2. В бинарном случае ($n = 2$) условию $(*)$ удовлетворяют ровно три отношения — $\varnothing$, Δ и ∇. Лемма 1 показывает, что условие $(*)$ влечёт выполнение одного из равенств $B_\rho(X) = B(X)$, $B_\rho(X) = P(X)$. При $n > 2$ это неверно, как показывает приводимый ниже пример. Необходимые и достаточные условия выполнения равенства $B_\rho(X) = P(X)$ авторам неизвестны.

Изложение примера сделаем после введения некоторых обозначений. Для натурального числа k положим

$$B^{(k)}(X) = \{\sigma \in B(X) \mid |\operatorname{im} \sigma| \leqslant k\}.$$

Очевидно, что $B^{(k)}(X)$ — левый идеал полугруппы $B(X)$. Далее, пусть

$$\sigma_k = \{\overline{x} \in X^n \mid |\operatorname{im} \overline{x}| \leqslant k\}.$$

Так же ясно, что σ_k удовлетворяет условию $(*)$.

Пример. Пусть $2 \leqslant k < n$ и $|X| > k$. Тогда σ_k удовлетворяет условию $(*)$, но $B_{\sigma_k}(X) \neq B(X), P(X), T(X)$. Докажем это. Заметим вначале, что $B^{(k)}(X) \subseteq B_{\sigma_k}(X)$. Действительно, если $\alpha \in B^{(k)}(X)$ и $\overline{x} \in \sigma_k$, то $|\operatorname{im}(\overline{x}\alpha)| \leqslant |\operatorname{im} \alpha| \leqslant k$, откуда $\overline{x}\alpha \in \sigma_k$. Далее, так как $k \geqslant 2$, то $P(X) \subset B^{(k)}(X) \subseteq B_{\sigma_k}(X)$. Кроме того, ввиду неравенства $k < n$ имеем $B_{\sigma_k}(X) \neq B(X)$.

Финансирование работы

Исследование выполнено при финансовой поддержке Российского научного фонда (проект № 22–11–00052).

Конфликт интересов

Авторы заявляют, что у них нет конфликта интересов.

Литература

1. **Глускин Л. М.** Полугруппы изотонных преобразований // Успехи мат. наук. 1961. Т. 16, вып. 5. С. 157–162.
2. **Попова Л. М.** Полугруппы частичных эндоморфизмов множества с отношением // Сиб. мат. журн. 1963. Т. 4, № 2. С. 309–317.

3. Кожухов И. Б., Ярошевич В. А. Полугруппы отображений, сохраняющих бинарное отношение // Фундам. и прикл. математика. 2008. Т. 14, вып. 7. С. 129–135.
4. Molchanov V. A. Semigroups of mappings on graphs // Semigroup Forum. 1983. V. 27. P. 155–199.
5. Клиффорд А., Престон Г. Алгебраическая теория полугрупп. Т. 1. М.: Мир, 1972.
6. Ершов Ю. Л., Палютин Е. А. Математическая логика. М.: Наука, 1987. 336 с.
7. Плоткин Б. И. Группы автоморфизмов алгебраических систем. М.: Наука, 1966. 604 с.
8. Вагнер В. В. Теория отношений и алгебра частичных отображений // Теория полугрупп и её приложения. Саратов: Изд-во Саратов. ун-та, 1965. С. 3–178.

Клюшин Алексей Александрович
Кожухов Игорь Борисович
Манилов Дмитрий Юрьевич
Решетников Артём Владимирович

Статья поступила
28 августа 2023 г.
После доработки —
6 сентября 2023 г.
Принята к публикации
22 сентября 2023 г.

DISKRETNYI ANALIZ I ISSLEDOVANIE OPERATSII
/DISCRETE ANALYSIS AND OPERATIONS RESEARCH/
January–March 2024. Vol. 31, No. 1. P. 19–34

UDC 512.534.1 DOI: 10.33048/daio.2024.31.783

DEFINABILITY OF RELATIONS BY SEMIGROUPS OF ISOTONE TRANSFORMATIONS

A. A. Klyushin[1, a], *I. B. Kozhukhov*[2, 3, b], *D. Yu. Manilov*[4, c], and *A. V. Reshetnikov*[2, d]

[1] Cadence Design Systems,
Bld. 1 Penrose Dock, Penrose Quay, Cork, T23 KW81, Ireland

[2] National Research University of Electronic Technology,
1 Shokin Square, 124498 Moscow, Russia

[3] Lomonosov Moscow State University,
1 Leninskie Gory, 119991 Moscow, Russia

[4] ELVEES Research and Development Center,
14 Bld. 14 Konstruktor Lukin Street, 1244660 Zelenograd, Moscow, Russia

E-mail: [a] xkreed@gmail.com, [b] kozhuhov_i_b@mail.ru, [c] thdi@ro.ru, [d] a_reshetnikov@hush.com

Abstract. In 1961, L. M. Gluskin proved that a given set X with an arbitrary nontrivial quasiorder ρ is determined up to isomorphism or anti-isomorphism by the semigroup $T_\rho(X)$ of all isotone transformations of (X, ρ), i. e., the transformations of X preserving ρ. Subsequently, L. M. Popova proved a similar statement for the semigroup $P_\rho(X)$ of all partial isotone transformations of (X, ρ); here the relation ρ does not have to be a quasiorder but can be an arbitrary nontrivial reflexive or antireflexive binary relation on the set X. In the present paper, under the same constraints on the relation ρ, we prove that the semigroup $B_\rho(X)$ of all isotone binary relations (set-valued mappings) of (X, ρ) determines ρ up to an isomorphism or anti-isomorphism as well. In addition, for each of the conditions $T_\rho(X) = T(X)$, $P_\rho(X) = P(X)$, $B_\rho(X) = B(X)$, we enumerate all n-ary relations ρ satisfying the given condition. Bibliogr. 8.

Keywords: semigroup of binary relations, isotone transformation.

English transl.: *Journal of Applied and Industrial Mathematics* **18** (1), 60–69 (2024), DOI: 10.1134/S199047892401006X.

References

1. **L. M. Gluskin,** Semigroups of isotone transformations, *Usp. Mat. Nauk* **16** (5), 157–162 (1961) [Russian].
2. **L. M. Popova,** Semigroups of partial endomorphisms of a set with a relation, *Sib. Mat. Zh.* **4** (2), 309–317 (1963) [Russian].
3. **I. B. Kozhukhov** and **V. A. Yaroshevich,** Transformation semigroups preserving a binary relation, *Fundam. Prikl. Mat.* **14** (7), 129–135 (2008) [Russian] [*J. Math. Sci.* **164** (2), 240–244 (2010)].
4. **V. A. Molchanov,** Semigroups of mappings on graphs, *Semigroup Forum* **27**, 155–199 (1983).
5. **A. H. Clifford** and **G. B. Preston,** *The Algebraic Theory of Semigroups,* Vol. 1 (AMS, Providence, 1961; Mir, Moscow, 1972 [Russian]).
6. **Yu. L. Ershov** and **E. A. Palyutin,** *Mathematical Logic* (Nauka, Moscow, 1987) [Russian].
7. **B. I. Plotkin,** *Automorphism Groups of Algebraic Systems* (Nauka, Moscow, 1966) [Russian].
8. **V. V. Vagner,** Relation theory and algebra of partial mappings // Theory of Semigroups and Its Applications (Izd. Saratov. Univ., Saratov, 1965), pp. 3–178 [Russian].

Aleksey A. Klyushin
Igor B. Kozhukhov
Dmitry Yu. Manilov
Artyom V. Reshetnikov

Received August 28, 2023
Revised September 6, 2023
Accepted September 22, 2023

ДИСКРЕТНЫЙ АНАЛИЗ И ИССЛЕДОВАНИЕ ОПЕРАЦИЙ
Январь–март 2024. Т. 31, № 1. С. 35–51

УДК 519.83 DOI: 10.33048/daio.2024.31.781

ВЕРОЯТНОСТНЫЙ ПОДХОД К ИГРЕ В УГАДЫВАНИЕ В СЛУЧАЙНОЙ СРЕДЕ

А. П. Ковалевский

Институт математики им. С. Л. Соболева,
пр. Акад. Коптюга, 4, 630090 Новосибирск, Россия
E-mail: artyom.kovalevskii@gmail.com

Аннотация. В статье формализуется и решается следующая игра двух лиц. Некоторый вопрос задан первому игроку. Второй игрок знает правильный ответ. Кроме того, оба игрока знают все возможные варианты ответа и их априорные вероятности. Второй игрок должен выбрать подмножество заданной мощности ответов-обманок. Первый игрок выбирает один из предложенных вариантов ответа. Первый игрок выигрывает у второго игрока единицу, если он угадал правильный ответ, и нуль иначе. Эта игра сводится к матричной игре. Однако матрица игры имеет большую размерность, из-за чего классический метод, основанный на решении пары двойственных задач линейного программирования, не может быть реализован для каждой индивидуальной задачи, поэтому необходимо разработать метод радикального понижения размерности.

Всё множество таких игр разбивается на два класса. Надравномерный класс игр характеризуется тем условием, что наибольшая из априорных вероятностей больше вероятности выбора ответа наудачу, а подравномерный класс соответствует противоположному неравенству: каждая из априорных вероятностей при умножении на общее число предъявляемых первому игроку ответов не превосходит единицы. Для каждого из этих двух классов решение расширенной матричной игры сводится к решению задачи линейного программирования существенно меньшей размерности. Для подравномерного класса игра переформулируется в терминах теории вероятностей. Условие на оптимальность смешанной стратегии формулируется с помощью теоремы Байеса. Для надравномерного класса решение игры использует вспомогательную задачу, относящуюся к подравномерному классу. Для обоих классов доказаны результаты о вероятностях угадывания правильного ответа при использовании оптимальных смешанных стратегий обоими игроками, а также разработаны алгоритмы получения этих стратегий. В подравномерном классе оптимальная смешанная стратегия первого игрока —

выбирать ответ наудачу, а в надравномерном — выбирать наиболее вероятный ответ. Оптимальные смешанные стратегии второго игрока имеют значительно более сложную структуру. Библиогр. 7.

Ключевые слова: матричная игра, теорема Байеса, равновероятное распределение, вероятность угадывания, решение в чистых стратегиях, решение в смешанных стратегиях.

Введение

Пусть игроку $\mathcal{A}$ задан вопрос, на который существует $n > 2$ возможных вариантов ответа $O_1, \ldots, O_n$. В каждом раунде игры правильный ответ на вопрос выбирается случайно и не зависит от предыстории. Вероятность для каждого возможного варианта ответа O_i быть правильным постоянна и равна r_i. Ответы упорядочены таким образом, что $r_1 \geqslant r_2 \geqslant \ldots \geqslant r_n > 0$. Мы считаем этот порядок фиксированным и будем употреблять выражение «наиболее вероятный ответ» (из некоторого множества ответов) в значении «ответ с наименьшим номером».

Игрок $\mathcal{A}$ знает вопрос и вероятности $r_1, \ldots, r_n$. Игрок $\mathcal{B}$, помимо этого, знает правильный ответ. Игрок $\mathcal{B}$ предлагает игроку $\mathcal{A}$ угадать, какой ответ правильный, предоставляя выбор из k вариантов, $1 < k < n$. Задача игрока $\mathcal{B}$ — по правильному ответу подобрать $k - 1$ вариантов ложных ответов так, чтобы вероятность того, что $\mathcal{A}$ угадает, была наименьшей. Задача игрока $\mathcal{A}$ — использовать алгоритм выбора из предложенных k вариантов, при котором вероятность угадывания будет наибольшей.

Эта игра является антагонистической игрой двух лиц и после усреднения по случайной среде, сформированной последовательностью номеров правильных ответов, формализуется как матричная игра (игра двух лиц с нулевой суммой [6, гл. 13]). Игрок $\mathcal{B}$ выбирает, какие k из n возможных ответов предложить. Более точно, игрок $\mathcal{B}$ в зависимости от варианта ответа, который верен для данного вопроса, выбирает другие варианты ответа, которые будут показаны игроку $\mathcal{A}$ вместе с верным вариантом в случайном порядке. Игрок $\mathcal{A}$ выбирает один из k предложенных ответов. Средний выигрыш игрока $\mathcal{A}$ при каждом выборе стратегий игроками равен вероятности угадывания верного ответа.

Теорема 1 даёт верхнюю и нижнюю цену игры и показывает, что верхняя цена игры больше нижней, т. е. нет решения в чистых стратегиях. Однако решение этой игры в смешанных стратегиях неприемлемо с вычислительной точки зрения. Как будет показано ниже, платёжная матрица имеет размерность $\left(C_{n-1}^{k-1}\right)^n \times k^{C_n^k}$, где C_n^k — биномиальный коэффициент. Классический подход [6, гл. 17] предполагает решение задачи линейного программирования с системой ограничений, определяемых платёжной матрицей. Так как матрица имеет скорость роста много выше

экспоненциальной (по количеству вариантов ответа), решение получить невозможно даже для малых n и k.

В [4, 5] представлены общие методы понижения размерности для матричных игр. Они не подходят для изучаемой задачи, так как представленные в работах алгоритмы имеют псевдополиномиальную по максимальному из размерностей платёжной матрицы временную сложность, что делает их неприменимыми. Необходимо понизить размерность задачи на основании другого подхода.

Разрабатываемый в данной работе подход использует связи между теорией игр, теорией вероятностей и линейным программированием. Ряд полезных связей (помимо прочего и с математической статистикой) изложен в [1, гл. 4].

Отнесём каждую индивидуальную задачу к одному из двух классов в зависимости от входных параметров. *Надравномерный* класс игр характеризуется условием $r_1 > \frac{1}{k}$, а *подравномерный* класс соответствует условию $r_1 \leqslant \frac{1}{k}$. Для этих двух классов существенно отличаются оптимальные смешанные стратегии игрока $\mathcal{A}$. Для каждого из этих двух классов решение расширенной матричной игры сводится к решению задачи линейного программирования существенно меньшей размерности.

Теорема 2 содержит результат для подравномерного класса. В этом случае задача формулируется в терминах теории вероятностей. Условие на оптимальность смешанной стратегии формулируется с помощью теоремы Байеса. Доказывается, что цена расширенной игры из этого класса равна $1/k$, оптимальной стратегией игрока $\mathcal{A}$ в подравномерном случае является смешанная стратегия «выбирать варианты ответа равновероятно», а оптимальная смешанная стратегия игрока $\mathcal{B}$ вычисляется решением задачи линейного программирования с матрицей размеров $n \times C_n^k$.

Теорема 3 даёт ответ для надравномерного класса. Она использует вспомогательную задачу, относящуюся к подравномерному классу. Цена расширенной игры из этого класса равна r_1, а оптимальной стратегией игрока $\mathcal{A}$ в надравномерном случае является чистая стратегия «выбирать самый популярный ответ». Размерность задачи линейного программирования для отыскания оптимальной смешанной стратегии игрока $\mathcal{B}$ остаётся той же, что и в теореме 2.

Современная теория игр изучает, в частности, марковские матричные игры, в которых предполагается, что платёжная матрица изменяется от раунда к раунду, образуя цепь Маркова. Данную задачу можно модифицировать, потребовав, чтобы последовательность правильных ответов образовывала марковскую цепь. В настоящее время теория марковских игр развивается, и нет аналитических методов получения точного решения [3]. Тем не менее существуют эффективные алгоритмы для приближённого решения, аппроксимирующего точное сколь угодно близко [7].

1. Сведéние к матричной игре

Пусть $O_1, \ldots, O_n$ — возможные варианты ответа, а событие $\{T = O_i\}$ означает, что верным является ответ с номером i. Тогда $r_i = P(T = O_i)$.

Игрок $\mathcal{B}$ для каждого варианта правильного ответа должен предложить $k-1$ ответов-ловушек. Игрок $\mathcal{A}$ для каждого варианта показанных ему ответов должен выбрать один ответ.

Чистая (*детерминированная*) стратегия игрока $\mathcal{B}$ состоит в том, что по известному верному ответу выбираются ещё $k-1$ ответов, т. е. каждому из ответов O_i сопоставляется подмножество $B_i \subseteq \{O_1, \ldots, O_n\} \setminus \{O_i\}$, содержащее $k-1$ элементов: $|B_i| = k-1$.

Введём обозначения для стратегий игрока $\mathcal{B}$:

$$\delta_{B_1 \ldots B_n} = B_i, \quad \text{если } T = O_i,\ i \in \{1, \ldots, n\}.$$

Число чистых стратегий игрока $\mathcal{B}$ — это возведённое в степень n число способов выбрать $k-1$ элементов из $n-1$ элементов:

$$\left(C_{n-1}^{k-1}\right)^n = \left(\frac{(n-1)!}{(k-1)!(n-k)!}\right)^n.$$

Игрок $\mathcal{A}$ из предложенных вариантов ответа выбирает один. *Чистая* (*детерминированная*) стратегия игрока $\mathcal{A}$ — из k-элементного множества ответов A_i выбрать один ответ D_i. Ясно, что для каждого множества этот выбор можно сделать k способами, а всего k-элементных множеств $m = C_n^k$ штук. Будем предполагать, что множества A_i упорядочены. Конкретный порядок не играет роли. Удобно считать, что множества упорядочены в лексикографическом порядке номеров их элементов.

Введём обозначения для стратегий игрока $\mathcal{A}$:

$$\theta_{D_1 \ldots D_m} = D_s, \quad \text{если множество ответов } A_s,\ s \in \{1, \ldots, m\}.$$

Таким образом, получено

Утверждение 1. *Число чистых стратегий игрока $\mathcal{A}$ равно*

$$M = k^{C_n^k} = k^{\frac{n!}{k!(n-k)!}},$$

а число чистых стратегий игрока $\mathcal{B}$ равно

$$N = \left(C_{n-1}^{k-1}\right)^n = \left(\frac{(n-1)!}{(k-1)!(n-k)!}\right)^n.$$

Пример 1. Пусть $n = 3$, $k = 2$. У каждого из игроков $2^3 = 8$ стратегий. Опишем их детально. Стратегии игрока $\mathcal{B}$:

$$\delta_{ijs} = \begin{cases} \text{показывать дополнительно ответ } i, \text{ если верный ответ 1,} \\ \text{показывать дополнительно ответ } j, \text{ если верный ответ 2,} \\ \text{показывать дополнительно ответ } s, \text{ если верный ответ 3.} \end{cases}$$

Стратегии игрока $\mathcal{A}$:

$$\theta_{ijs} = \begin{cases} \text{выбирать ответ } i, \text{ если показывают ответы 1 и 2,} \\ \text{выбирать ответ } j, \text{ если показывают ответы 1 и 3,} \\ \text{выбирать ответ } s, \text{ если показывают ответы 2 и 3.} \end{cases}$$

Выигрыш игрока $\mathcal{A}$ (равный потерям игрока $\mathcal{B}$) в каждой ситуации, т. е. при каждом сделанном игроками выборе, — это вероятность того, что игрок $\mathcal{A}$ угадает верный ответ. Тем самым соответствующий паре стратегий $(\theta_{D_1\ldots D_m}, \delta_{B_1\ldots B_n})$ элемент платёжной матрицы равен

$$\varphi_{st} = \varphi(\theta_{D_1\ldots D_m}, \delta_{B_1\ldots B_n}) = \sum_{i=1}^{n} r_i \left(\sum_{j=1}^{C_n^k} \mathbf{1}(A_j = B_i \cup \{O_i\},\ D_j = O_i) \right),$$
$$s = 1, \ldots M,\ t = 1, \ldots, N.$$

Вернёмся к примеру 1, в котором $k = 2$, $n = 3$. Имеем три вероятности $r_1 \geqslant r_2 \geqslant r_3$. Платёжная матрица имеет вид

	δ_{211}	δ_{212}	δ_{231}	δ_{232}	δ_{311}	δ_{312}	δ_{331}	δ_{332}
θ_{112}	r_1	r_1	$r_1 + r_2$	$r_1 + r_2$	r_1	r_1	$r_1 + r_2$	$r_1 + r_2$
θ_{113}	r_1	$r_1 + r_3$	r_1	$r_1 + r_3$	r_1	$r_1 + r_3$	r_1	$r_1 + r_3$
θ_{132}	$r_1 + r_3$	r_1	1	$r_1 + r_2$	r_3	0	$r_2 + r_3$	r_2
θ_{133}	$r_1 + r_3$	$r_1 + r_3$	$r_1 + r_3$	$r_1 + r_3$	r_3	r_3	r_3	r_3
θ_{212}	r_2	r_2	r_2	r_2	$r_1 + r_2$	$r_1 + r_2$	$r_1 + r_2$	$r_1 + r_2$
θ_{213}	r_2	$r_2 + r_3$	0	r_3	$r_1 + r_2$	1	r_1	$r_1 + r_3$
θ_{232}	$r_2 + r_3$	r_2	$r_2 + r_3$	r_2	$r_2 + r_3$	r_2	$r_2 + r_3$	r_2
θ_{233}	$r_2 + r_3$	$r_2 + r_3$	r_3	r_3	$r_2 + r_3$	$r_2 + r_3$	r_3	r_3

Следующая теорема даёт верхнюю и нижнюю цены игры.

Теорема 1. 1) *Нижняя цена игры равна*

$$\max_{s \leqslant M} \min_{t \leqslant N} \varphi_{st} = r_1,$$

а соответствующая максиминная стратегия игрока $\mathcal{A}$ — из предъявленных выбирать ответ с наименьшим номером (*самый вероятный ответ*).

2) *Верхняя цена игры равна*

$$\min_{t \leqslant N} \max_{s \leqslant M} \varphi_{st} = \sum_{i:\, 1+ik \leqslant n} r_{1+ik} = r_1 + r_{1+k} + \cdots > r_1,$$

а соответствующая минимаксная стратегия игрока $\mathcal{B}$ — показывать вместе первые k наиболее вероятных ответов (*ответы с номерами* $1, \ldots, k$), *вторые k наиболее вероятных ответов* (*ответы с номерами* $k+1, \ldots, 2k$), *и т. д. Если n не кратно k, то в последний набор ответов включаются дополнительно первые $k(\lfloor n/k \rfloor + 1) - n$ ответов.*

Доказательство. 1) Максиминное решение означает, что игрок $\mathcal{A}$ выбирает свою чистую стратегию, а игрок $\mathcal{B}$, зная его чистую стратегию, выбирает свою чистую стратегию так, чтобы минимизировать свой проигрыш.

У игрока $\mathcal{A}$ есть чистая стратегия, состоящая в выборе ответа с наименьшим номером (наиболее вероятного ответа) из множества предъявленных ответов. В тех случаях, когда первый ответ верный, игрок $\mathcal{B}$ обязан его предъявить. Так как вероятность первого ответа равна r_1, выигрыш первого игрока составляет не меньше r_1. В тех случаях, когда первый ответ не верен, игрок $\mathcal{B}$ показывает первый ответ в числе прочих, и игрок $\mathcal{A}$ выбирает его, при этом он выигрывает 0 (с вероятностью $1 - r_1$). Таким образом, чистая стратегия s_0, состоящая в выборе ответа с наименьшим номером, обеспечивает равенство

$$\min_{t \leqslant N} \varphi_{s_0 t} = r_1.$$

Покажем, что стратегия s_0 максиминная, т. е. её замена любой другой чистой стратегией не приводит к увеличению минимального (по всем чистым стратегиям игрока $\mathcal{B}$) среднего выигрыша игрока $\mathcal{A}$. Действительно, рассмотрим все такие множества A_j предъявляемых игроку $\mathcal{A}$ ответов, которые содержат конкретный ответ O_i. Если хотя бы на одном из этих множеств игрок $\mathcal{A}$ выбирает ответ, не равный O_i, то игрок $\mathcal{B}$ предъявляет именно это множество в том случае, когда O_i является верным ответом, и выигрыш игрока $\mathcal{A}$ в этом случае равен 0.

Значит, будем предполагать, что существует ответ O_i, который выбирается игроком $\mathcal{A}$ для всех множеств, его содержащих. Тогда игрок $\mathcal{B}$ обязан показать одно из таких множеств, если O_i является верным ответом, и это происходит с вероятностью r_i. Если O_i не является верным ответом, то игрок $\mathcal{B}$ также показывает одно из таких множеств A_j, включая в него ответ O_i наряду с верным ответом. Таким образом, в любом случае, когда O_i не является верным ответом, выигрыш игрока $\mathcal{A}$ равен 0. Итак, средний суммарный выигрыш равен r_i. Следовательно,

$$\max_{s \leqslant M} \min_{t \leqslant N} \varphi_{st} = \max_{i \leqslant n} r_i = r_1.$$

2) Минимаксное решение означает, что игрок $\mathcal{B}$ выбирает свою чистую стратегию, а игрок $\mathcal{A}$, зная его чистую стратегию, выбирает свою чистую стратегию так, чтобы максимизировать свой выигрыш. Каждому ответу $O_1, \ldots, O_n$ игрок $\mathcal{B}$ сопоставляет содержащее его множество $A_1, \ldots, A_n$. Если какие-либо из этих множеств совпадают, то игрок $\mathcal{A}$ не может различить, какой из ответов верен, и для максимизации среднего выигрыша выбирает наиболее вероятный ответ. Средний выигрыш игрока $\mathcal{A}$ равен сумме наибольших вероятностей по всем различным множествам из $A_1, \ldots, A_n$. Доказательство проведём индукцией по n.

Базис индукции: пусть $n = k + 1$ (минимально возможное значение). Множество A_1 содержит первый ответ и приносит игроку $\mathcal{A}$ выигрыш r_1. Так как $|A_1| = k$, игрок $\mathcal{B}$ выбирает $k - 1$ множеств совпадающими с A_1 (что даёт нулевые потери), а одно множество должно отличаться. Потери будут наименьшими (и равными $r_1 + r_{k+1}$), если первые k множеств совпадают, а последнее отличается. В частности, можно в множество A_{k+1} добавить ответы $O_1, \ldots, O_{k-1}$.

Шаг индукции: предположим, что утверждение теоремы выполнено для $n \geqslant k + 1$. Докажем, что оно верно для $n + 1$. Рассмотрим два варианта.

Первый вариант: n не кратно k. Тогда игрок $\mathcal{B}$ сохраняет множества

$$A_1 = \cdots = A_k = \{O_1, \ldots, O_k\},$$

$$\ldots$$

$$A_{k\lfloor n/k\rfloor - k + 1} = \cdots = A_{k\lfloor n/k\rfloor} = \{O_{k\lfloor n/k\rfloor - k + 1}, \ldots, O_{k\lfloor n/k\rfloor}\}$$

и выбирает

$$A_{k\lfloor n/k\rfloor + 1} = \cdots = A_{n+1} = \{O_{k\lfloor n/k\rfloor + 1}, \ldots, O_{n+1}\}$$

(если $n + 1$ кратно k) или

$$A_{k\lfloor n/k\rfloor + 1} = \cdots = A_{n+1} = \{O_{k\lfloor n/k\rfloor + 1}, \ldots, O_{n+1}, O_1, \ldots, O_{k-n-1+k\lfloor (n+1)/k\rfloor}\}$$

(если $n + 1$ не кратно k), что даёт нулевое увеличение его средних потерь. Так как при переходе от n к $n + 1$ его средние потери не могут уменьшиться (максимум берётся по более широкому множеству), оптимальность решения сохраняется.

Второй вариант: n делится нацело на k. Тогда игрок $\mathcal{B}$ сохраняет все предыдущие множества и выбирает

$$A_{n+1} = \{O_{n+1}, O_1, \ldots, O_{k-1}\}.$$

Такой выбор оптимален, так как увеличивает цену простой игры на минимально возможную величину r_{n+1}. Теорема 1 доказана.

Возвращаясь к примеру 1 ($k = 2$, $n = 3$), получаем, что здесь нижняя цена игры равна r_1, а верхняя цена игры равна $r_1 + r_3$.

2. Подравномерный случай

Согласно теореме 1 нижняя цена простой игры всегда строго меньше верхней, поэтому необходимо рассматривать расширенную матричную игру, в которой стратегии каждого из игроков выбираются случайно в соответствии с некоторым вероятностным распределением на множестве стратегий. Таким образом, имеются три независимых в совокупности последовательности случайных величин: одна отвечает за выбор ответа, другая — за выбор стратегии первым игроком, третья — за выбор

стратегии вторым игроком. Вероятностные распределения на соответствующих множествах стратегий называются *смешанными* стратегиями игроков $\mathcal{A}$ и $\mathcal{B}$.

Классический метод решения расширенных матричных игр подразумевает решение пары двойственных задач линейного программирования. Однако для данной игры матрица системы ограничений для задачи линейного программирования имеет размерность $M \times N = k^{C_n^k} \times \left(C_{n-1}^{k-1}\right)^n$. При $n = 6$, $k = 4$ размерность равна $2^{30} \times 10^6$. Решение задачи даже в таком частном случае требует значительных вычислительных ресурсов. Кроме того, размерность растёт со скоростью, сильно большей, чем экспоненциальная (по n). Решение задачи в такой постановке для приложений не представляется возможным. Необходимо искать методы понижения размерности системы ограничений, либо искать решение, основываясь на вероятностной постановке задачи.

Существуют алгоритмы понижения размерностей для матричных игр (например, [4], основанный на составлении новой эквивалентной задачи линейного программирования, состоящей из выпуклой оболочки множества решений), однако в случае размерностей такого порядка, как в этой задаче, они остаются бесполезны. В [4] предлагаемый алгоритм получает решение матричной игры за время $O(TN^2 + n^{3,5})$, где N — максимальная из размерностей платёжной матрицы, n — размерность редуцированной матрицы, T — некоторый параметр, зависящий от матрицы. В любом случае необходимо провести псевдополиномиальное число операций над матрицей, что для данных размерностей практически невозможно на практике, поэтому становится актуальным поиск другого метода решения.

Посмотрим на задачу с помощью формулы Байеса. Игрок $\mathcal{A}$ знает множество возможных ответов и априорные вероятности $r_1, \ldots, r_n$ того, что соответствующий ответ будет верным. Также он видит предложенное подмножество A_s из k ответов, из которого он должен выбрать правильный ответ. Тогда по формуле Байеса апостериорная вероятность того, что вариант O_i будет верным ($T = O_i$) при условии, что игрок $\mathcal{A}$ выбирает из множества A_s, равна

$$P(T = O_i \mid A_s) = \frac{P(A_s \mid T = O_i) r_i}{P(A_s)}, \quad i \in A_s.$$

Обозначим условные вероятности выбора игроком $\mathcal{A}$ из множества A_s через

$$q_{i,A_s} = P(A_s \mid T = O_i).$$

Игрок $\mathcal{B}$ не может повлиять на априорные вероятности, однако может изменять апостериорные, чтобы уменьшить вероятность угадать верный

ответ. Потребуем, чтобы условная вероятность быть верным ответом была одинаковой для любого элемента из данного подмножества. Для этого числитель дроби не должен зависеть от i, т. е. для любых $1 \leqslant s \leqslant C_n^k$, $|A_s| = k$, $O_i \in A_s$ требуется существование $z_s \geqslant 0$ такого, что

$$q_{i,A_s} = \frac{z_s}{r_i}.$$

Тем самым знания об априорном распределении становятся бесполезными, а у игрока $\mathcal{A}$ нет лучшей возможности, чем выбрать наугад из k элементов. В новых обозначениях условная вероятность того, что i–й элемент будет верным ответом, равна

$$P(T = O_i \mid A_s) = \frac{z_s}{P(A_s)}, \quad O_i \in A_s.$$

Заметим, что для всех i

$$\sum_{s\colon O_i \in A_s} P(A_s \mid T = O_i) = P\Big(\bigcup_{s\colon O_i \in A_s} A_s \mid T = O_i \Big) = 1.$$

Тогда, умножив равенство на r_i, получим

$$\sum_{s\colon O_i \in A_s} P(A_s \mid T = O_i) r_i = \sum_{s\colon O_i \in A_s} z_s = r_i, \quad i = 1, \ldots, n.$$

Кроме того, $z_s \geqslant 0$ для любого A_s.

Получили систему уравнений и неравенств для определения z_s из n уравнений с C_n^k неизвестными, а также с условиями на знак:

$$\begin{cases} \sum\limits_{s\colon O_i \in A_s} z_s = r_i, & i = 1, \ldots, n, \\ z_s \geqslant 0, & 1 \leqslant s \leqslant C_n^k. \end{cases} \tag{1}$$

Решив её, можно получить вероятность для множества ответов $A_s \ni O_i$ при условии, что O_i является верным ответом:

$$P(A_s \mid T = O_i) = q_{i,A_s} = \frac{z_s}{r_i}. \tag{2}$$

Запишем задачу линейного программирования, состоящую в максимизации суммы переменных при ограничениях в виде неравенств:

$$\begin{cases} \gamma = \sum\limits_{s=1}^{C_n^k} z_s \to \max, & \\ \sum\limits_{s\colon O_i \in A_s} z_s \leqslant r_i, & i = 1, \ldots, n, \\ z_s \geqslant 0, & 1 \leqslant s \leqslant C_n^k. \end{cases} \tag{3}$$

Здесь задача максимизации суммы всех переменных введена искусственно, она будет использоваться для доказательства существования решения у системы (1).

Теорема 2. *Если* $r_1 \leqslant \frac{1}{k}$, *то цена игры равна* $\frac{1}{k}$, *задача линейного программирования* (3) *совместна, её решение удовлетворяет ограничениям* (1) *и оптимальной смешанной стратегией игрока* $\mathcal{B}$ *является выбор вероятностей в соответствии с* (2). *При этом оптимальной является смешанная стратегия игрока* $\mathcal{A}$, *при которой варианты ответов выбираются равновероятно. Если же* $r_1 > \frac{1}{k}$, *то система* (1) *не имеет решений и равновероятная стратегия игрока* $\mathcal{A}$ *не оптимальна.*

Доказательство. Обозначим через $(b_{is})_{i=1,\dots,n}^{s=1,\dots,C_n^k}$ матрицу системы уравнений (1). Тогда $b_{is} = 1$ в том и только том случае, когда i-й элемент содержится в подмножестве A_s. Так как в любом A_s содержится ровно k элементов, сложив все уравнения, получим $k\sum\limits_{i=1}^{C_n^k} z_s = \sum\limits_{i=1}^{n} r_i = 1$, т. е.

$$\sum_{i=1}^{C_n^k} z_s = \frac{1}{k}.$$

Тогда выигрыш игрока $\mathcal{A}$ при разыгрывании любой своей стратегии равен $\frac{1}{k}$, если система уравнений и неравенств совместна.

Предположим, что $r_1 > \frac{1}{k}$ и $z_s \geqslant 0$ для всех s. Тогда

$$\sum_{s\colon O_1 \in A_s} z_s = r_1 \leqslant \sum_{s=1}^{C_n^k} z_s = \frac{1}{k},$$

чего не может быть.

Для доказательства совместности системы в подравномерном случае рассмотрим задачу, двойственную задаче (3):

$$\begin{cases} \omega = \sum\limits_{i=1}^{n} r_i x_i \to \min, & \\ \sum\limits_{j \in A_s} x_j \geqslant 1, & s = 1, \dots, C_n^k, \\ x_i \geqslant 0, & i = 1, \dots, n. \end{cases} \tag{4}$$

Из ограничений на знак переменных двойственной задачи следует, что ω ограничена снизу. Кроме того, решение

$$x_1 = x_2 = \dots = x_n = \frac{1}{k} \tag{5}$$

допустимое, и целевая функция для такого решения равна $\frac{1}{k}$. В силу теоремы двойственности существует решение прямой задачи такое, что $\gamma_{\max} = \omega_{\min}$. Покажем, что в подравномерном случае решение (5) оптимально, т. е. всегда $\omega \geqslant \frac{1}{k}$.

Для этого изучим соседние с (5) вершины симплекса. В этих вершинах одна из переменных становится базисной и принимает нулевое значение: $x_i = 0$ для некоторого $i \in \{1, \ldots, n\}$. Небазисные переменные выражаются через базисные, поэтому для всех A_s таких, что $O_i \in A_s$, выполнено

$$\sum_{j::\, O_j \in A_s \setminus \{O_i\}} x_j = 1,$$

при этом $|A_s| = k$. В силу симметрии $x_j = \frac{1}{k-1}$ для всех $j \neq i$.

Следовательно, соседние с (5) вершины симплекса имеют следующий вид: $x_i = 0$ для некоторого $i \in \{1, \ldots, n\}$ и $x_j = \frac{1}{k-1}$ для всех $j \neq i$. Однако в этих вершинах

$$\omega = \frac{1}{k-1} \sum_{j \neq i} r_j = \frac{1 - r_i}{k-1} \geqslant \frac{1}{k}$$

в силу условия подравномерности $r_i \leqslant r_1 \leqslant \frac{1}{k}$.

Итак, (5) является решением двойственной задачи, и все переменные принимают строго положительные значения. Значит, существует решение прямой задачи (3), для которого все существенные ограничения обратятся в строгие равенства (свойство дополняющей нежёсткости, см. [2, § 4.5]), что означает существование неотрицательного решения системы уравнений (1). Теорема 2 доказана.

Пример 2. Пусть $n = 3$, $k = 2$, $\frac{1}{2} \geqslant r_1 \geqslant r_2 \geqslant r_3$. Тогда $z_{12} + z_{13} = r_1$, $z_{12} + z_{23} = r_2$, $z_{13} + z_{23} = r_3$. Отсюда

$$z_{12} = \frac{r_1 + r_2 - r_3}{2}, \quad z_{13} = \frac{r_1 - r_2 + r_3}{2}, \quad z_{23} = \frac{-r_1 + r_2 + r_3}{2},$$

$$q_{23} = \frac{z_{23}}{r_2}, \quad q_{13} = \frac{z_{13}}{r_1}, \quad q_{12} = \frac{z_{12}}{r_1},$$

$$q_{21} = \frac{z_{12}}{r_2}, \quad q_{31} = \frac{z_{13}}{r_3}, \quad q_{32} = \frac{z_{23}}{r_3}.$$

Здесь q_{ij} — вероятность того, что игрок $\mathcal{B}$ добавит ответ-ловушку с номером j, если правильный ответ имеет номер i. Игрок $\mathcal{A}$ выбирает ответы равновероятно, его средний выигрыш равен $\frac{1}{2}$.

3. Надравномерный случай

Надравномерный случай характеризуется условием $r_1 > \frac{1}{k}$. Докажем, что здесь цена расширенной игры равна r_1, а у игрока $\mathcal{A}$ есть оптимальная чистая стратегия — выбирать самый вероятный ответ O_1. Также найдём оптимальную смешанную стратегию игрока $\mathcal{B}$. Таких стратегий очень много, но трудно найти конкретную стратегию, которая подходит для общего случая.

Будем сводить надравномерный случай к подравномерному. Для этого введём модифицированные (срезанные и масштабированные) вероятности

$$r'_i = L\min(r_i, c).$$

Здесь $c > 0$ — уровень срезки, а $L = \Big(\sum\limits_{i=1}^{n} \min(r_i, c)\Big)^{-1}$ — нормирующая константа, обеспечивающая выполнение условия

$$\sum_{i=1}^{n} r'_i = 1.$$

Выберем c как наименьший положительный корень уравнения

$$ck = \sum_{i=1}^{n} \min(r_i, c) =: g(c). \tag{6}$$

Такой корень $c \in \big(0, \frac{1}{k}\big)$ обязательно существует, так как левая и правая части уравнения (6) непрерывны как функции от c, равны 0 при $c = 0$, $\frac{dg}{dx}|_{x=0+} = n > k$, $g\big(\frac{1}{k}\big) < 1$. Тогда

$$r'_1 = Lc = \frac{1}{k},$$

и выполнено условие подравномерности для вероятностей r'_i.

По аналогии с (3) запишем задачу линейного программирования:

$$\begin{cases} \gamma' = \sum\limits_{s=1}^{C_n^k} z'_s \to \max, \\ \sum\limits_{A_s \ni O_i} z'_s \leqslant r'_i, & i = 1, \ldots, n,\ |A_s| = k, \\ z'_s \geqslant 0, & 1 \leqslant s \leqslant C_n^k. \end{cases} \tag{7}$$

Она имеет решение согласно теореме 2.

Теорема 3. *Если $r_1 > \frac{1}{k}$, то цена расширенной игры равна r_1. При этом оптимальна чистая стратегия игрока $\mathcal{A}$, при которой выбирается наиболее вероятный ответ. Оптимальной смешанной стратегией игрока $\mathcal{B}$ будет следующая: если ответ O_i верный, то множество ответов A_s выбирается с вероятностью $q'_{i,A_s} = z'_s/r'_i$, $O_i \in A_s$. Здесь $\{z'_s\}_{s=1}^{C_n^k}$ — решение задачи* (7).

Доказательство. Рассмотрим произвольную чистую стратегию игрока $\mathcal{A}$ и оценим сверху её средний выигрыш.

Игрок $\mathcal{A}$ из каждого возникающего в игре множества ответов A_s выбирает один из его элементов. Пусть

$$r_1 \geqslant \ldots \geqslant r_{k_0} \geqslant c \geqslant r_{k_0+1} \geqslant \ldots \geqslant r_n.$$

Из (6) получаем

$$ck = ck_0 + r_{k_0+1} + \cdots + r_n,$$

откуда

$$c = \frac{r_{k_0+1} + \cdots + r_n}{k - k_0}.$$

Следовательно,

$$k_0 = \min\left\{t \geqslant 1 \mid \frac{r_{t+1} + \cdots + r_n}{k - t} \geqslant r_{t+1}\right\}.$$

Такое $k_0 < k$ всегда существует, так как при $t = k - 1$ неравенство выполнено.

Если $i \leqslant k_0$, то $r'_i = \frac{1}{k}$. Это максимально возможная правая часть в (7), поэтому существует смешанная стратегия игрока $\mathcal{B}$, в которой любое выбираемое множество A_s содержит ответы $1, \ldots, k_0$ с вероятностью 1, при этом $r_i/r'_i = kr_i$.

Если $i > k_0$, то

$$r'_i = \frac{(k - k_0)r_i}{k(r_{k_0+1} + \cdots + r_n)},$$
$$r_i/r'_i = \frac{k(r_{k_0+1} + \cdots + r_n)}{k - k_0}.$$

Оценим сверху условную вероятность того, что O_i является верным ответом, если предложено множество ответов A_s:

$$P(T = O_i \mid A_s) = \frac{P(T = O_i, A_s)}{P(A_s)} =$$
$$= \frac{P(A_s \mid T = O_i)r_i}{\sum\limits_{j\colon O_j \in A_s} P(A_s \mid T = O_j)r_j} = \frac{z'_s r_i/r'_i}{\sum\limits_{j\colon O_j \in A_s} z'_s r_j/r'_j} = \frac{r_i/r'_i}{\sum\limits_{j\colon O_j \in A_s} r_j/r'_j} \leqslant$$
$$\leqslant \frac{kr_1}{k(r_1 + \cdots + r_{k_0}) + \sum\limits_{j\colon O_j \in A_s \cap \{O_{k_0+1}, \ldots, O_n\}} \frac{k(r_{k_0+1} + \cdots + r_n)}{k - k_0}} = r_1,$$

так как $|A_s \cap \{O_{k_0+1}, \ldots, O_n\}| = k - k_0$.

Итак, при предлагаемой смешанной стратегии игрока $\mathcal{B}$ средний выигрыш любой чистой стратегии игрока $\mathcal{A}$ не превосходит r_1. Теорема 3 доказана.

Пример 3. Пусть $n = 3$, $k = 2$, $r_1 > \frac{1}{2} > r_2 \geqslant r_3$. В силу условия нормировки $r_2 + r_3 < \frac{1}{2}$. Согласно теореме 3 $c = r_2 + r_3$,

$$r'_1 = \frac{1}{2}, \quad r'_2 = \frac{r_2}{2(r_2 + r_3)}, \quad r'_3 = \frac{r_3}{2(r_2 + r_3)}, \quad r'_2 + r'_3 = \frac{1}{2}.$$

Максимум целевой функции достигается на решении системы уравнений

$$\begin{cases} z'_{12} + z'_{13} = \frac{1}{2}, \\ z'_{12} + z'_{23} = r'_2, \\ z'_{13} + z'_{23} = r'_3. \end{cases}$$

Отсюда

$$z'_{12} = r'_2, \quad z'_{13} = r'_3, \quad z'_{23} = 0,$$
$$q'_{12} = \frac{r_2}{r_2 + r_3}, \quad q'_{13} = \frac{r_3}{r_2 + r_3}, \quad q'_{21} = q'_{31} = 1, \quad q'_{23} = q'_{32} = 0.$$

Как и в примере 2, здесь q_{ij} — вероятность того, что игрок $\mathcal{B}$ добавит ответ-ловушку с номером j, если правильный ответ имеет номер i. Игрок $\mathcal{A}$ выбирает наиболее вероятный ответ, его средний выигрыш равен r_1.

Заключение

В результате работы была полностью решена задача об игре в угадывание в случайной среде. С помощью теоретико-игровой модели удалось получить верхнюю и нижнюю цены игры. Были разработаны методы понижения размерности для данной задачи, что позволило получать численные решения для произвольной индивидуальной задачи более эффективно, чем это делают существующие общие методы понижения размерности и алгоритмы для теоретико-игровых задач.

В ходе исследования был выявлен класс индивидуальных задач, для которых существует стратегия игрока $\mathcal{B}$ такая, что информация об априорном распределении вариантов ответов не приносит дополнительного выигрыша игроку $\mathcal{A}$.

В продолжение исследования можно рассмотреть модифицированную формулировку задачи, в которой последовательность платёжных матриц образует цепь Маркова. Такая задача относится к классу марковских игр, исследования которых активно проводятся в последние годы.

Благодарности

Автор благодарит Е. Прокопенко за предложение провести исследование в этом направлении и И. Смирнова за численную реализацию разработанных алгоритмов.

Финансирование работы

Исследование выполнено в рамках государственного задания Института математики им. С. Л. Соболева СО РАН (проект № FWNF–2022–0010).

Конфликт интересов

Автор заявляет, что у него нет конфликта интересов.

Литература

1. **Borovkov A. A.** Mathematical statistics. New York: Gordon and Breach, 1998. 570 p.
2. **Bradley S. P., Hax A. C., Magnanti T. L.** Applied mathematical programming. Boston: Addison-Wesley, 1977. 716 p.
3. **Hörner J., Rosenberg D., Solan E., Vieille N.** On a Markov game with one-sided information // Oper. Res. 2010. V. 58, No. 4-2. P. 1107–1115.
4. **Li S., Chen M., Wang Y., Wu Q.** A fast algorithm to solve large-scale matrix games based on dimensionality reduction and its application in multiple unmanned combat air vehicles attack-defense decision-making // Inf. Sci. 2022. V. 594. P. 305–321.
5. **Lipton R. J., Young N. E.** Simple strategies for large zero-sum games with applications to complexity theory // Proc. 26th Annu. ACM Symp. Theory of Computing (Montreal, Canada, May 23–25, 1994). New York: ACM, 1994. P. 734–740.
6. **Neumann J., Morgenstern O.** Theory of games and economic behavior. Princeton: Princeton Univ. Press, 2007. 776 p.
7. **Wei Ch.-Y., Lee Ch.-W., Zhang M., Luo H.** Last-iterate convergence of decentralized optimistic gradient descent-ascent in infinite-horizon competitive Markov games // Proc. Mach. Learn. Res. 2021. V. 134. P. 4259–4299.

Ковалевский Артём Павлович

Статья поступила
9 августа 2023 г.
После доработки —
5 сентября 2023 г.
Принята к публикации
22 сентября 2023 г.

DISKRETNYI ANALIZ I ISSLEDOVANIE OPERATSII
/DISCRETE ANALYSIS AND OPERATIONS RESEARCH/
January–March 2024. Vol. 31, No. 1. P. 35–51

UDC 519.83 DOI: 10.33048/daio.2024.31.781

A PROBABILISTIC APPROACH TO THE GAME OF GUESSING IN A RANDOM ENVIRONMENT

A. P. Kovalevskii

Sobolev Institute of Mathematics,
4 Acad. Koptyug Avenue, 630090 Novosibirsk, Russia
E-mail: artyom.kovalevskii@gmail.com

Abstract. The following game of two persons is formalized and solved in the paper. Player 1 is asked a question. Player 2 knows the correct answer. Moreover, both players know all possible answers and their a priori probabilities. Player 2 must choose a subset of the given cardinality of deception answers. Player 1 chooses one of the proposed answers. Player 1 wins one from Player 2 if he/she guessed the correct answer and zero otherwise. This game is reduced to a matrix game. However, the game matrix is of large dimension, so the classical method based on solving a pair of dual linear programming problems cannot be implemented for each individual problem. Therefore, it is necessary to develop a method to radically reduce the dimension.

The whole set of such games is divided into two classes. The superuniform class of games is characterized by the condition that the largest of the a priori probabilities is greater than the probability of choosing an answer at random, and the subuniform class corresponds to the opposite inequality — each of the a priori probabilities when multiplied by the total number of answers presented to Player 1 does not exceed one. For each of these two classes, the solving the extended matrix game is reduced to solving a linear programming problem of a much smaller dimension. For the subuniform class, the game is reformulated in terms of probability theory. The condition for the optimality of a mixed strategy is formulated using the Bayes theorem. For the superuniform class, the solution of the game uses an auxiliary problem related to the subuniform class. For both classes, we prove results on the probabilities of guessing the correct answer when using optimal mixed strategies by both players. We present algorithms for obtaining these strategies. The optimal mixed strategy of Player 1 is to choose

English transl.: *Journal of Applied and Industrial Mathematics* **18** (1), 70–80 (2024), DOI: 10.1134/S1990478924010071.

an answer at random in the subuniform class and to choose the most probable answer in the superuniform class. Optimal mixed strategies of Player 2 have much more complex structure. Bibliogr. 7.

Keywords: matrix game, Bayes' theorem, equiprobable distribution, guessing probability, solution in pure strategies, solution in mixed strategies.

References

1. **A. A. Borovkov,** *Mathematical Statistics* (Gordon and Breach, New York, 1998).
2. **S. P. Bradley, A. C. Hax,** and **T. L. Magnanti,** *Applied Mathematical Programming* (Addison-Wesley, Boston, 1977).
3. **J. Hörner, D. Rosenberg, E. Solan,** and **N. Vieille,** On a Markov game with one-sided information, *Oper. Res.* **58** (4-2), 1107–1115 (2010).
4. **S. Li, M. Chen, Y. Wang,** and **Q. Wu,** A fast algorithm to solve large-scale matrix games based on dimensionality reduction and its application in multiple unmanned combat air vehicles attack-defense decision-making, *Inf. Sci.* **594**, 305–321 (2022).
5. **R. J. Lipton** and **N. E. Young,** Simple strategies for large zero-sum games with applications to complexity theory, in *Proc. 26th Annu. ACM Symp. Theory of Computing (Montreal, Canada, May 23–25, 1994)* (ACM, New York, 1994), pp. 734–740.
6. **J. Neumann** and **O. Morgenstern,** *Theory of Games and Economic Behavior* (Princeton Univ. Press, Princeton, 2007).
7. **Ch.-Y. Wei, Ch.-W. Lee, M. Zhang,** and **H. Luo,** Last-iterate convergence of decentralized optimistic gradient descent-ascent in infinite-horizon competitive Markov games, *Proc. Mach. Learn. Res.* **134**, 4259–4299 (2021).

Artyom P. Kovalevskii

Received August 9, 2023
Revised September 5, 2023
Accepted September 22, 2023

ДИСКРЕТНЫЙ АНАЛИЗ И ИССЛЕДОВАНИЕ ОПЕРАЦИЙ
Январь–март 2024. Т. 31, № 1. С. 52–84

УДК 519.7 DOI: 10.33048/daio.2024.31.772

ПОСТКВАНТОВЫЕ КРИПТОСИСТЕМЫ: ОТКРЫТЫЕ ВОПРОСЫ И СУЩЕСТВУЮЩИЕ РЕШЕНИЯ. КРИПТОСИСТЕМЫ НА ИЗОГЕНИЯХ И КОДАХ, ИСПРАВЛЯЮЩИХ ОШИБКИ

Е. С. Малыгина[1,2,a], *А. В. Куценко*[2,b], *С. А. Новосёлов*[1,c], *Н. С. Колесников*[1,d], *А. О. Бахарев*[2,e], *И. С. Хильчук*[2,f], *А. С. Шапоренко*[2,g], *Н. Н. Токарева*[2,1,h]

[1] Балтийский федеральный университет им. И. Канта,
ул. Александра Невского, 14, 236041 Калининград, Россия
[2] Новосибирский гос. университет,
ул. Пирогова, 2, 630090 Новосибирск, Россия
E-mail: [a] emalygina@kantiana.ru, [b] alexandrkutsenko@bk.ru,
[c] novsem@gmail.com, [d] nikolesnikov100@gmail.com, [e] a.bakharev@g.nsu.ru,
[f] irina.khilchuk@gmail.com, [g] shaporenko.alexandr@gmail.com,
[h] crypto1127@mail.ru

Аннотация. Представлен обзор основных постквантовых криптографических схем на основе кодов, исправляющих ошибки, и изогений эллиптических кривых, а также вычислительно трудных задач, лежащих в их основе. Особое внимание уделено описанию атак на представленные схемы. В частности, для кодовых криптосистем описаны атаки на основе информационных совокупностей и расщепления носителя, для криптосистем на изогениях дано подробное описание атаки Кастрика — Декру на схему SIDH/SIKE. Табл. 2, библиогр. 43.

Ключевые слова: постквантовая криптография, код, исправляющий ошибки, эллиптическая кривая, изогения.

Введение

Настоящая статья является продолжением работы [1], в которой рассмотрены криптосистемы, основанные на теории решёток, и задачи, лежащие в основе их стойкости. Как отмечено в [1], помимо схем на основе решёток, выбранных в качестве нового стандарта постквантовой криптографии, существуют альтернативные кандидаты, а именно — криптосистемы на основе кодов, исправляющих ошибки, и изогений суперсингулярных эллиптических кривых. Данная работа посвящена описанию

таких криптосистем, а также вычислительно трудных задач, на которых базируется их стойкость.

В отличие от криптографии на решётках, в которой безопасность ряда протоколов основывается на сложности аппроксимации случайного вектора далёко за пределами ближайшего вектора решётки, в криптографии на кодах безопасность протоколов зависит от нахождения кодовых слов малого веса ниже границы Варшамова — Гилберта, поскольку алгоритм декодирования за пределами этой границы на данный момент не известен.

К основным преимуществам и недостаткам криптосистем на кодах (в частности, схем цифровой подписи) стоит отнести следующие:

- использование алгоритма Куртуа — Финьяза — Сондриера (CFS) [2] приводит к малому размеру самой подписи при сравнительно медленной скорости работы и большому размеру открытого ключа;
- использование эвристики Фиата — Шамира [3], наоборот, приводит к малому размеру открытого ключа (как правило, несколько сотен бит), обеспечивает достаточно высокую скорость работы, однако размер подписи сравнительно велик (составляет около 10^5 бит).

Криптография на изогениях существенно отличается от предыдущих двух типов, поскольку основана на естественно возникающей в теории чисел задаче вычисления изогений между эллиптическими кривыми. Таким образом, криптосистемы на изогениях составляют одно из немногих семейств, на данный момент устойчивых к атакам с использованием квантового компьютера и основанных на теоретико-числовых задачах (если рассматривать задачу решения нелинейных систем уравнений от многих переменных как теоретико-числовую). В некотором смысле задачу нахождения изогении можно рассматривать как аналог задачи дискретного логарифмирования, в рамках которого вместо абелевой группы точек кривой используют граф изогений. В то время как квантовый компьютер решает задачу дискретного логарифма в группе точек эллиптической кривой за полиномиальное время, для вычисления изогении существующими алгоритмами требуется субэкспоненциальное время при использовании обычных эллиптических кривых и экспоненциальное время при использовании суперсингулярных эллиптических кривых [4].

К основным преимуществам и недостаткам криптосистем на изогениях стоит отнести небольшие размеры ключей при относительно медленной скорости работы. Несмотря на то, что криптографические системы данного типа выглядят многообещающе, они нуждаются в более глубоком изучении. Это обусловлено появлением эффективных атак на некоторые варианты подобных криптосистем, например, атаки Кастрика — Декру (см. разд. 2.5), опубликованной в 2022 г.

1. Коды, исправляющие ошибки

1.1. Предварительные сведения. Коды, исправляющие ошибки, используются для передачи информации в каналах связи, в которых информация искажается. Определённая избыточность в передаваемых кодовых словах (блоках) позволяет обнаруживать ошибки в принятых словах (блоках) и исправлять их, выбирая ближайшие кодовые слова. Особое распространение получили линейные коды в силу более эффективных алгоритмов кодирования и декодирования.

Пусть $\mathbb{F}_q$ — конечное поле, состоящее из q элементов, $\mathbb{F}_q^n$ — векторное пространство над $\mathbb{F}_q$. *Линейным* $[n,k]$*-кодом* $\mathcal{C}$ называется k-мерное векторное подпространство в $\mathbb{F}_q^n$. При этом вектор $(c_1, c_2, \ldots, c_n) \in \mathcal{C}$ называется *кодовым словом* $\mathcal{C}$.

Важным качеством кода является возможность исправления приобретённых ошибок в ходе передачи информации по зашумлённому каналу. Для определения кодового расстояния введём метрику на векторном пространстве $\mathbb{F}_q^n$.

Расстоянием Хэмминга между векторами $x, y \in \mathbb{F}_q^n$ называется число координат, в которых векторы различаются:

$$d_H(x,y) = |\{i \mid x_i \neq y_i\}|.$$

Весом Хэмминга вектора $x \in \mathbb{F}_q^n$ называется число его ненулевых координат:

$$wt_H(x) = |\{i \mid x_i \neq 0\}| = d_H(x,0).$$

Кодовым расстоянием кода $\mathcal{C}$ называется минимальное расстояние Хэмминга между его различными кодовыми словами:

$$d = \min\{d_H(x,y) \mid x,y \in \mathcal{C},\ x \neq y\}.$$

В случае линейных кодов имеем

$$d = \min\{d_H(x,y) \mid x,y \in \mathcal{C},\ x \neq y\} = \min\{wt_H(x) \mid x \in \mathcal{C},\ x \neq 0\}.$$

При этом число исправляемых кодом $\mathcal{C}$ ошибок равно $t = \left\lfloor \frac{d-1}{2} \right\rfloor$.

Чтобы задать линейный код, можно либо задать его базис, либо задать систему линейных уравнений, решением которой являются координаты кодовых слов. Формально это можно сделать с помощью матриц.

Порождающей матрицей линейного $[n,k]$-кода $\mathcal{C}$ называется матрица G над $\mathbb{F}_q$ размера $k \times n$ такая, что

$$\mathcal{C} = \left\{xG \mid x \in \mathbb{F}_q^k\right\}.$$

Проверочной матрицей линейного $[n,k]$-кода $\mathcal{C}$ называется матрица H над $\mathbb{F}_q$ размера $(n-k) \times n$ такая, что

$$\mathcal{C} = \{y \in \mathbb{F}_q^n \mid Hy^\top = 0\}.$$

Одним из главных в теории кодирования является вопрос, как использовать коды для исправления ошибок. Ответ на него даёт процедура декодирования. *Декодированием* кода $\mathcal{C}$ называется отображение $D_{\mathcal{C}}\colon \mathbb{F}_q^n \to \mathcal{C}$. Код *исправляет* t *ошибок*, если $D_{\mathcal{C}}(c+e) = c$ для любых $c \in \mathcal{C}$ и $e \in \mathbb{F}_q^n$, $wt_H(e) \leqslant t$.

1.2. Базовые кодовые криптосистемы.

Схема Мак-Элиса. Первой криптосистемой на кодах была схема шифрования с открытым ключом, предложенная в 1978 г. Мак-Элисом [5]. Однако практически все асимметричные модификации на базе кодов, предложенные позже, имеют общий недостаток — большие требования к памяти.

Рассмотрим оригинальную криптосистему Мак-Элиса. Её секретным ключом является классический код Гоппы.

Пусть m и t — положительные целые числа. Определим *многочлен Гоппы*

$$g(X) = \sum_{i=0}^{t} g_i X^i \in \mathbb{F}_{2^m}[X]$$

и множество

$$\mathcal{L} = \{\alpha_0, \ldots, \alpha_{n-1}\} \in \mathbb{F}_{2^m}^n,$$

где $g(\alpha_j) \neq 0$, $j = 0, \ldots, n-1$. *Синдром* представляет собой многочлен вида

$$\mathcal{S}_c(X) = \left(-\sum_{i=0}^{n-1} \frac{c_i}{g(\alpha_i)} \cdot \frac{g(X) - g(\alpha_i)}{X - \alpha_i} \right) \bmod g(X),$$

где $c = (c_0, \ldots, c_{n-1}) \in \mathbb{F}_{2^m}^n$. *Двоичным кодом Гоппы* над $\mathbb{F}_{2^m}$ называется линейный код

$$\mathcal{C}(\mathcal{L}, g(X)) = \left\{ c \in \mathbb{F}_{2^m}^n \mid \mathcal{S}_c(X) = \sum_{i=0}^{n-1} \frac{c_i}{X - \alpha_i} \equiv 0 \pmod{g(X)} \right\}.$$

При этом проверочная матрица кода Гоппы $\mathcal{C}(\mathcal{L}, g(X))$ имеет вид

$$H = \begin{pmatrix} \frac{g_s}{g(\alpha_0)} & \frac{g_s}{g(\alpha_1)} & \cdots & \frac{g_s}{g(\alpha_{n-1})} \\ \frac{g_{s-1}+g_s\alpha_0}{g(\alpha_0)} & \frac{g_{s-1}+g_s\alpha_0}{g(\alpha_1)} & \cdots & \frac{g_{s-1}+g_s\alpha_0}{g(\alpha_{n-1})} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{g_1+g_2\alpha_0+\cdots+g_s\alpha_0^{s-1}}{g(\alpha_0)} & \frac{g_1+g_2\alpha_0+\cdots+g_s\alpha_0^{s-1}}{g(\alpha_1)} & \cdots & \frac{g_1+g_2\alpha_0+\cdots+g_s\alpha_0^{s-1}}{g(\alpha_{n-1})} \end{pmatrix},$$

или

$$H = \begin{pmatrix} g_s & 0 & \dots & 0 \\ g_{s-1} & g_s & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ g_1 & g_2 & \dots & g_s \end{pmatrix} \times \begin{pmatrix} \frac{1}{g(\alpha_0)} & \frac{1}{g(\alpha_1)} & \dots & \frac{1}{g(\alpha_{n-1})} \\ \frac{\alpha_0}{g(\alpha_0)} & \frac{\alpha_1}{g(\alpha_1)} & \dots & \frac{\alpha_{n-1}}{g(\alpha_{n-1})} \\ \vdots & \vdots & \dots & \vdots \\ \frac{\alpha_0^{s-1}}{g(\alpha_0)} & \frac{\alpha_1^{s-1}}{g(\alpha_1)} & \dots & \frac{\alpha_{n-1}^{s-1}}{g(\alpha_{n-1})} \end{pmatrix},$$

где $g(X) = g_s X^s + g_{s-1} X^{s-1} + \dots + g_0$.

Генерация ключей. Алгоритм генерации ключей начинается с выбора двоичного кода Гоппы, исправляющего до t ошибок. Для этого случайным образом выбирается неприводимый многочлен Гоппы степени t. Затем вычисляется соответствующая порождающая матрица G. Поскольку злоумышленник, зная G, сможет определить структуру используемого кода и эффективно его декодировать, алгебраическая структура матрицы G должна быть скрыта. Для этой цели используются обратимая матрица S и перестановочная матрица P, которые генерируются случайным образом и умножаются на G слева и справа соответственно, чтобы сформировать $\widetilde{G} = SGP$. Таким образом, $\widetilde{G}$ является порождающей матрицей для кода, эквивалентного $\mathcal{C}$.

Будем считать, что $q = 2^m$. Обобщая вышесказанное, заметим, что для генерации ключей необходимо следующее.

1. Выбрать двоичный $[n, k, d]$-код Гоппы $\mathcal{C}$, исправляющий $t = \lfloor \frac{d-1}{2} \rfloor$ ошибок.
2. Вычислить порождающую матрицу кода G над $\mathbb{F}_q$ размера $k \times n$.
3. Выбрать случайную обратимую матрицу $S \in GL_k(\mathbb{F}_q)$.
4. Выбрать случайную перестановочную матрицу $P \in S_n(\mathbb{F}_q)$.
5. Вычислить $\widetilde{G} = SGP$.

Секретным ключом является тройка $K_{\text{priv}} = \{G, S, P\}$, открытым — $K_{\text{pub}} = \widetilde{G}$.

Зашифрование представляет собой простое векторно-матричное умножение k-битного сообщения m на порождающую матрицу $\widetilde{G}$ и добавление случайного вектора ошибки e веса Хэмминга, не превосходящего t.

1. Представить исходное сообщение в виде двоичной строки m длины k.
2. Выбрать случайный вектор ошибки $e \in \mathbb{F}_q^n$ такой, что $wt_H(e) \leqslant t$.
3. Вычислить $c = m\widetilde{G} + e$.

Расшифрование представляет собой исправление ошибок в полученном сообщении с использованием известного алгоритма декодирования $D_{\mathcal{C}}$ для кода $\mathcal{C}$. Декодирование является наиболее трудоёмкой частью

процесса расшифрования, что делает его более медленным, чем шифрование.

1. Вычислить $\widetilde{c} = cP^{-1}$.
2. Применяя алгоритм декодирования $D_{\mathcal{C}}$ к $\widetilde{c}$, вычислить вектор $\widetilde{m}$ длины k.
3. Вычислить $m = (\widetilde{m}G^{-1})S^{-1}$.

Корректность расшифрования. Имеем

$$\widetilde{c} = cP^{-1} = (mSGP + e)P^{-1} = (mS)G + eP^{-1}.$$

Поскольку $wt_H(eP^{-1}) \leqslant t$, то $\widetilde{m} = D_{\mathcal{C}}(\widetilde{c}) = D_{\mathcal{C}}(cP^{-1}) = mSG$. Восстанавливаем исходное сообщение m:

$$(\widetilde{m}G^{-1})S^{-1} = ((mSG)G^{-1})S^{-1} = m.$$

Схема Нидеррайтера. В 1986 г. Нидеррайтер предложил ещё одну кодовую криптосистему [6]. В этой схеме сообщение полностью кодируется в вектор ошибок, что позволяет избежать утечки информации из битов открытого текста, в отличие от схемы Мак-Элиса. В качестве открытого ключа для вычисления синдрома используется проверочная матрица. Преимуществом этой схемы является меньший размер открытого ключа.

Генерация ключей. Как и ранее, $q = 2^m$. Чтобы сгенерировать ключи, необходимо следующее.

1. Выбрать двоичный $[n, k, d]$-код Гоппы $\mathcal{C}$, исправляющий $t = \lfloor \frac{d-1}{2} \rfloor$ ошибок.
2. Вычислить проверочную матрицу кода H над $\mathbb{F}_q$ размера $(n-k)\times n$.
3. Выбрать случайную обратимую матрицу $S \in GL_{n-k}(\mathbb{F}_q)$.
4. Выбрать случайную перестановочную матрицу $P \in S_n(\mathbb{F}_q)$.
5. Вычислить $\widetilde{H} = SHP$.

Секретным ключом является тройка $K_{\text{priv}} = \{H, S, P\}$, открытым — $K_{\text{pub}} = \widetilde{H}$.

Зашифрование осуществляется следующим образом.

1. Представить исходное сообщение в виде двоичной строки e длины n веса $wt_H(e) = t$.
2. Вычислить $c = \widetilde{H}e^{\top}$.

Расшифрование выполняется с применением алгоритма декодирования кода Гоппы, лежащего в основе криптосистемы.

1. Вычислить $\widetilde{c} = S^{-1}c$.
2. Применяя алгоритм декодирования $D_{\mathcal{C}}$ к $\widetilde{c}$, вычислить $\widetilde{e}$.
3. Вычислить вектор $e = P^{-1}\widetilde{e}$ длины n веса t.

Корректность расшифрования заключается в том, что исходное сообщение имеет вес t. Это значит, что к зашифрованному сообщению можно применить алгоритм декодирования.

Позже было показано, что схема Нидеррайтера имеет такой же уровень стойкости, что и схема Мак-Элиса [7].

1.3. Задачи, лежащие в основе безопасности. Согласно [8] два основных аспекта, на которых основывается безопасность схемы Мак-Элиса, следующие:

(1) сложность задачи декодирования общего неизвестного кода, которая NP-трудна [9];

(2) сложность атак, восстанавливающих структуру базового кода.

Ключевым действием, обеспечивающим безопасность кодовых криптосистем, является сокрытие структуры используемого кода. Как и ранее, пусть G — порождающая матрицей приватного кода $\mathcal{C}$, а $\widetilde{\mathcal{C}}$ — открытый код, полученный из $\mathcal{C}$ с помощью одного или нескольких секретных преобразований. Приведём наиболее распространённые преобразования.

• Умножение G справа на случайную обратимую матрицу S над $\mathbb{F}_q$ размера $k \times k$.

• Умножение G слева на случайную обратимую матрицу T над $\mathbb{F}_q$ размера $n \times n$.

• Умножение G справа на случайную матрицу полного ранга S над $\mathbb{F}_q$ размера $\ell \times k$, $\ell < k$. В этом случае имеем подкод $\mathcal{C}_{SG} \subseteq \mathcal{C}$ с порождающей матрицей SG, для которого всё так же можно использовать соответствующий алгоритм декодирования, исправляющий заданное число ошибок.

• Использование подполевых подкодов $\mathcal{C} \cap \mathbb{F}_p$ при условии, что исходный код $\mathcal{C}$ определён над расширением конечного поля $\mathbb{F}_p$.

• Использование кода $\mathcal{C}_{[G|SG]}$, чья порождающая матрица получена с помощью конкатенации $[\,G \mid SG\,]$, где S определена над $\mathbb{F}_q$, имеет размер $k \times k$ и обратима.

• Добавление ℓ случайных столбцов слева к матрице G.

• Укорачивание и прокалывание кода $\mathcal{C}$.

Под *прокалыванием* и *укорачиванием* кода будем понимать следующее. *Проколотый* (*единожды*) код $\mathcal{C}^*$ получается из $\mathcal{C}$ удалением одной и той же i-й координаты в каждом кодовом слове. Если G — порождающая матрица кода $\mathcal{C}$, то порождающая матрица G^* кода $\mathcal{C}^*$ получается удалением i-го столбца из G. Пусть T — множество индексов, $|T| = s$. *Проколотый* (*многократно*) код $\mathcal{C}^T$ получается из $\mathcal{C}$ удалением позиций с индексами из T в каждом кодовом слове $\mathcal{C}$. Отметим, что $\mathcal{C}^T$ — код с параметрами $[n - s, \geqslant k - s, \geqslant d - s]$. Пусть $\mathcal{C}(T) \subseteq \mathcal{C}$ — подкод с нулевыми координатами при индексах из T. *Укороченный* код $\mathcal{C}_T$ длины $n - s$ получается прокалыванием кода $\mathcal{C}(T)$ в позициях с индексами из T.

Рассмотрим некоторые известные задачи, сложность решения которых лежит в основе безопасности схем типа Мак-Элиса и Нидеррайтера.

Задача Мак-Элиса. *Для заданных открытого ключа* $\widetilde{G}$ *и шифртекста* c *найти единственное сообщение* m *такое, что* $wt_H(m\widetilde{G} - c) = t$.

Поскольку код $\widetilde{\mathcal{C}}$ с порождающей матрицей $\widetilde{G}$ эквивалентен исходному коду $\mathcal{C}$, нельзя предполагать, что задача Мак-Элиса NP-трудна в отличие от общей задачи декодирования. Однако решение данной задачи позволило бы найти решение общей задачи декодирования лишь для некоторых классов кодов, но не для всех.

В схеме Нидеррайтера процесс зашифрования можно записать иначе, а именно: $c = e\widetilde{H}$. Тогда задача декодирования сводится к нахождению кодового слова $x \in \mathcal{C}$, близкого к e относительно расстояния Хэмминга. На практике трудно проверить, действительно ли вектор ошибки, лежащий в смежном классе $e + \mathcal{C}$, имеет минимальный вес, поэтому рассматриваемая задача декодирования не NP-трудна. Рассмотрим задачу синдромного декодирования.

Синдромом вектора $y \in \mathbb{F}_q^n$ относительно проверочной матрицы H кода $\mathcal{C}$ называется вектор $S_H(y) = yH \in \mathbb{F}_q^{n-k}$. Отметим, что два вектора из $\mathbb{F}_q^n$ имеют один и тот же синдром тогда и только тогда, когда лежат в одном смежном классе по $\mathcal{C}$.

Задача синдромного декодирования. *Для матрицы* H *над* $\mathbb{F}_2$ *размера* $r \times n$, *вектора* $c \in \mathbb{F}_2^r$ *и целого* $t > 0$ *найти слово* x *в смежном классе* $S_H^{-1}(c) = e + \mathcal{C}$ *такое, что* $d(x) \leqslant t$.

Значение параметра t существенно влияет на сложность решения задачи синдромного декодирования. Задача имеет решение тогда и только тогда, когда t таково, что с высокой вероятностью обеспечивает существование единственного решения (т. е. t не больше границы Варшамова — Гилберта).

Задача нахождения ненулевых слов малого веса Хэмминга в заданном линейном коде схожа с задачей синдромного декодирования.

Задача нахождения кодового слова малого веса Хэмминга. *Для матрицы* H *над* $\mathbb{F}_2$ *размера* $r \times n$ *и целого* $t > 0$ *найти ненулевое слово* x *в* $S_H^{-1}(0)$ *такое, что* $d(x) \leqslant t$.

Отметим, что если $\mathcal{C}$ — линейный код с проверочной матрицей H, то любое решение задачи синдромного декодирования с входными параметрами H, t, $eH^\top$ также является решением задачи нахождения кодового слова малого веса Хэмминга с параметрами H', t, где H' — проверочная матрица кода $\mathcal{C}' = \mathcal{C} \cup (y + \mathcal{C})$. Обратное верно только в том случае, если $t < d$, где d — кодовое расстояние кода $\mathcal{C}$, которое обычно

неизвестно. Однако большинство двоичных линейных кодов длины n и коразмерности r (под коразмерностью имеем в виду число $r = n - \dim \mathcal{C}$ при условии, что $\mathcal{C} \subseteq \mathbb{F}_q^n$) имеют кодовое расстояние, очень близкое к расстоянию Варшамова — Гилберта $d_0(n, r)$, которое является максимально возможным значением, удовлетворяющим условию

$$\sum_{i=0}^{d_0(n,r)-1} \binom{n}{i} \leqslant 2^r.$$

В трёх рассматриваемых задачах вес t является входным значением. Особый интерес представляют те случаи, когда вес t будет зависеть от длины n и размерности k кода. Такие случаи имеют место в задачах полного и ограниченного декодирования. Расшифрование будет заключаться в нахождении слова минимального веса. Если синдром случайный, то зачастую решение будет иметь вес, равный расстоянию Варшамова — Гилберта.

Задача полного декодирования. *Для матрицы H над $\mathbb{F}_2$ размера $r \times n$ и $c \in \mathbb{F}_2^r$ найти слово x в классе $S_H^{-1}(c)$ такое, что $d(x) \leqslant d_0(n, r)$.*

В действительности задача полного декодирования является самой сложной вычислительной задачей для заданных параметров n и r. В кодовых криптосистемах типа Мак-Элиса или Нидеррайтера вес t равен числу ошибок, исправляемых кодом.

Код Гоппы длины $n = 2m$, исправляющий t ошибок, имеет коразмерность $r = tm$. В этом случае сложность решения следующей вычислительной задачи напрямую зависит от стойкости схемы.

Задача ограниченного декодирования. *Для матрицы H над $\mathbb{F}_2$ размера $r \times n$ и вектора $c \in \mathbb{F}_2^r$ найти слово x в классе $S_H^{-1}(c)$ такое, что $d(x) \leqslant \frac{r}{\log_2 n}$.*

1.4. Атаки. В качестве основных атак можно выделить две:

1) дешифрование конкретного зашифрованного сообщения;

2) структурная атака, направленная на определение структуры кода, а значит, секретного ключа.

Подбор параметров, обеспечивающих безопасность схем типа Мак-Элиса, должен осуществляться с учётом известных атак. Несмотря на то, что в основе кодовых криптосистем лежит задача декодирования, структурная атака на схемы типа Мак-Элиса отличается от общей задачи декодирования.

Задача декодирования на основе информационных совокупностей лежит в основе первого типа атак. Представим обобщённый вариант такого декодирования, отмечая, что Ли и Брикелл были первыми, кто использовал его для анализа криптосистемы Мак-Элиса [10].

Алгоритм 1. Декодирование на основе информационных совокупностей для параметра α

Вход: Матрица G над $\mathbb{F}_q$ размера $k \times n$, $w \in \mathbb{Z}$.
Выход: Кодовое слово $x \neq 0$, $wt(x) \leqslant w$.

1: **repeat**
2: $x = 0 \in \mathbb{F}_q^n$
3: Выбрать перестановочную матрицу P над $\mathbb{F}_q$ размера $n \times n$.
4: Вычислить $G' = UGP = [\mathbb{I} \mid R]$, полагая, что первые k позиций информационные, U — обратимая матрица над $\mathbb{F}_q$ размера $k \times k$, $\mathbb{I}$ — единичная матрица размера $k \times k$.
5: Вычислить все суммы α строк матрицы G'.
6: **if** wt(одна из сумм) $\leqslant w$ **then** $x \leftarrow$ эта строка
7: **until** $x \neq 0$
8: **return** x

Перейдём к рассмотрению структурной атаки. Коды, исправляющие ошибки и имеющие эффективный алгоритм декодирования, как правило, либо обладают алгебраической структурой, либо строятся специальным образом. Зная порождающую матрицу кода, можно эффективно исправить приобретённые ошибки. Это справедливо для всех кодов, имеющих большую размерность, с целью рассмотрения их в криптографических приложениях.

Рассмотрим, как можно восстановить алгебраическую структуру кода при условии, что проверочная матрица не разреженная.

В кодовых криптосистемах, как было сказано ранее, секретный код $\mathcal{C}$ стараются скрыть. Один из способов — применить к коду изометрию f. Тогда открытым ключом будет порождающая или проверочная матрица эквивалентного кода $\mathcal{C}' = f(\mathcal{C})$. В двоичном случае изометрией является любая перестановка носителя. Если метрическое пространство образует векторное пространство, то рассматривают полулинейные изометрии

$$f\colon \mathbb{F}_q^n \to \mathbb{F}_q^n\colon (x_1, \ldots, x_n) \mapsto (v_1\pi(x_{\sigma^{-1}(1)}), \ldots, v_n\pi(x_{\sigma^{-1}(n)})).$$

Здесь $(v_1, \ldots, v_n) \in \mathbb{F}_q^n$ — вектор с ненулевыми компонентами, π — автоморфизм поля $\mathbb{F}_q$, σ — перестановка носителя исходного кода, т. е. множества $\{1, \ldots, n\}$.

Два линейных кода $\mathcal{C}$ и $\mathcal{C}'$ назовём *эквивалентными,* если $\mathcal{C}' = f(\mathcal{C})$ для некоторой полулинейной изометрии f. В случае произвольного поля $\mathbb{F}_q$ эквивалентность отличается от перестановочной эквивалентности: коды $\mathcal{C}$ и $\mathcal{C}'$ *перестановочно эквивалентны,* если для некоторой перестановки $\sigma \in S_n$ имеем $\mathcal{C}' = \{(x_{\sigma^{-1}(1)}, \ldots, x_{\sigma^{-1}(n)}) \in \mathbb{F}_q^n \mid (x_1, \ldots, x_n) \in \mathcal{C}\}$. В двоичном случае эти понятия совпадают.

Алгоритм расщепления носителя решает задачу эквивалентности кодов, которая заключается в следующем. Необходимо определить, существуют ли перестановочно эквивалентные линейные коды $\mathcal{C}_1$ и $\mathcal{C}_2$ с порождающими матрицами G_1 и G_2, определёнными над некоторым конечным полем. Эта задача была предложена Петранком и Ротом [11], которые показали, что она достаточно сложная однако не NP-полна.

Прежде чем описать алгоритм расщепления носителя, приведём ряд необходимых определений. Более детально с терминологией, относящейся к алгоритму, можно ознакомиться в [12].

Для некоторого множества $J \subseteq I = \{1, \ldots, n\}$ через $\mathcal{C}_J$ обозначим множество векторов, которые получены из кодовых слов кода $\mathcal{C}$ путём зануления координат, имеющих номера из J. Через $\mathcal{L}_n$ обозначим множество всех линейных кодов длины n. Тогда множество $\mathcal{L} = \bigcup\limits_{n \geqslant 1} \mathcal{L}_n$ является множеством всех линейных кодов. Отображение $\nu\colon \mathcal{L} \to E$ называется *инвариантом* над некоторым множеством E, если для любых двух перестановочно эквивалентных кодов $\mathcal{C}$ и $\mathcal{C}'$ имеет место равенство $\nu(\mathcal{C}) = \nu(\mathcal{C}')$. Отображение $\Sigma\colon \mathcal{L}_n \times I \to E$ называется *сигнатурой* над E, если $\Sigma(\mathcal{C}, i) = \Sigma(\sigma(\mathcal{C}), \sigma(i))$ для любых $\mathcal{C} \in \mathcal{L}_n$, $i \in I$ и $\sigma \in S_n$. Далее будем рассматривать сигнатуры, удовлетворяющие условию $\Sigma(\mathcal{C}, i) = \nu(\mathcal{C}_{\{i\}})$. Здесь $\mathcal{C}_{\{i\}} = \mathcal{C}_J$ для $J = \{i\}$ в соответствии с обозначением, введённым выше.

Имея сигнатуру Σ, гораздо проще ответить на вопрос, являются ли коды $\mathcal{C}$ и $\mathcal{C}'$ перестановочно эквивалентными. Для этого следует вычислить $\Sigma(\mathcal{C}, I)$ и $\Sigma(\mathcal{C}', I)$. Если коды $\mathcal{C}$ и $\mathcal{C}'$ перестановочно эквивалентны, то $\Sigma(\mathcal{C}, I) = \Sigma(\mathcal{C}', I)$. Сигнатура Σ называется *дискриминантом* кода $\mathcal{C}$, если существуют $i, j \in I$ такие, что $\Sigma(\mathcal{C}, i) \neq \Sigma(\mathcal{C}, j)$. Сигнатура Σ называется *полным дискриминантом* кода $\mathcal{C}$, если для любых $i \neq j$ из I выполняется $\Sigma(\mathcal{C}, i) \neq \Sigma(\mathcal{C}, j)$.

Отметим, что если $\mathcal{C}' = \sigma(\mathcal{C})$ и Σ — полный дискриминант кода $\mathcal{C}$, то для любого $i \in I$ найдётся единственный элемент $j \in I$ такой, что $\Sigma(\mathcal{C}, i) = \Sigma(\mathcal{C}', j)$. Равенства $\sigma(i) = j$, $i \in I$, определяют перестановку σ.

Алгоритм 2. Алгоритм расщепления носителя

Вход: $G_1, G_2 \in \mathbb{F}_q^{k \times n}$.
Выход: Перестановка σ такая, что $\mathcal{C}' = \sigma(\mathcal{C})$.

1: Вычислить сигнатуру Σ, являющуюся полным дискриминантом.
2: **for** $i, j \in \{1, \ldots, n\}$ **do**
3: **if** $\Sigma(G_1, i) = \Sigma(G_2, j)$ **then** $\sigma(i) = j$
4: **return** σ

1.5. Схемы, актуальные на сегодняшний день.

Инкапсуляция ключа переключением битов. BIKE (bit flipping key encapsulation) — схема шифрования на основе двоичных линейных QC-MDPC-кодов [13].

Под QC-MDPC-кодом, ассоциированным с тройкой (n, r, w), будем понимать код, проверочная матрица H которого определена над $\mathbb{F}_2$, имеет размер $r \times n$ и строку веса w:

$$H = [H_0 \mid H_1 \mid \cdots \mid H_{n_0-1}],$$

где r простое, $n = n_0 r$ и H_i — циркулянтный блок над $\mathbb{F}_2$ размера $r \times r$.

Обозначим $\mathcal{R} = \mathbb{F}_2[X]/(X^r - 1)$. Секретным ключом схемы BIKE является проверочная матрица $H = (H_0 \mid H_1)$ над $\mathbb{F}_2$ размера 1×2 двоичного линейного $[n, k]$-QC-MDPC-кода $\mathcal{C}$, причём $|H_0| = |H_1| = \frac{w}{2}$. Благодаря изоморфизму между кольцом циркулянтных матриц размера $r \times r$ и кольцом многочленов $\mathcal{R}$ все матричные операции могут быть рассмотрены как операции над многочленами. Открытым ключом является секретный ключ в каноническом виде, а именно: $H_{\text{pub}} = \left(\mathbb{I} \mid H_0^{-1} H_1\right)$.

В основе процедуры шифрования лежит схема Нидеррайтера. Исходное сообщение представляет собой вектор e веса t, а соответствующим шифртекстом является $H_{\text{pub}} e^\top$. Расшифрование выполняется с помощью умножения шифртекста на блок H_0, чтобы получить синдром $He^\top$, а далее для восстановления e используется black-grey-flip-декодер с переключением битов [14].

Безопасность IND-CPA, лежащая в основе схемы BIKE, базируется на сложности решения следующих задач.

Задача квазициклического синдромного декодирования (QCSD)**.** *Для заданного $h \in \mathcal{R}$, вектора $y \in \mathcal{R}$ и параметра $t > 0$ определить, существует ли пара $(e_0, e_1) \in \mathcal{R}^2$ такая, что $wt_H(e_0) + wt_H(e_1) = t$ и $e_0 + e_1 h = y$.*

Задача поиска квазициклического кодового слова (QCCF)**.** *Для заданного $h \in \mathcal{R}$ и параметра $v > 0$ определить, существует ли пара $(c_0, c_1) \in \mathcal{R}^2$ такая, что $wt_H(c_0) + wt_H(c_1) = v$ и $c_0 + c_1 h = 0$.*

Считается, что вес $wt_H(h)$ нечётный, а $wt_H(y) = t$. Наиболее известными алгоритмами для решения этих задач являются декодирование на основе информационных совокупностей и его вариации. Чтобы обеспечить λ бит защиты в смысле IND-CPA, сложность обеих задач QCSD и QCCF должна превышать 2λ. В [15] показано, что параметры BIKE для каждого уровня стойкости выбираются согласно условию

$$\lambda \approx t - \frac{1}{2}\log_2 r \approx w - \log_2 r.$$

Безопасность IND-CCA обеспечивается с помощью использования преобразования Фуджисаки — Окомото [16]

Классическая схема Мак-Элиса. Эта схема на основе двоичных кодов Гоппы использует стандартные методы для достижения надёжности в смысле IND-CCA.

Использование классической схемы Мак-Элиса гарантирует доказательство безопасности IND-CCA2 в модели квантового оракула [17], основанной на предположении, что схема обеспечивает одностороннюю защиту при атаках на выбранный открытый текст. В качестве альтернативы стойкость схемы может быть обеспечена благодаря предположениям, что проверочная матрица двоичного кода Гоппы неотличима от проверочной матрицы случайного линейного кода такой же размерности, а задача синдромного декодирования сложна для случайных линейных кодов такой же размерности, что и используемый код.

Наиболее эффективной из известных атак на классическую схему Мак-Элиса является декодирование на основе информационных совокупностей. Попытки найти секретный ключ с помощью алгебраического криптоанализа или перебором являются более дорогостоящими.

Квазициклическая схема Хэмминга. HQC (Hamming quasi-cyclic) — схема на основе QC-MDPC-кодов без скрытой структуры [18].

Основная идея заключается в извлечении выгоды из квазициклической структуры наряду со снижением стойкости при декодировании случайного линейного кода. В частности, трудно свести обеспечение стойкости кодовой схемы к сложности решения общей задачи декодирования, если открытым ключом является замаскированный секретный ключ с применением скремблирования или перестановки.

Рассмотрим кольцо $\mathcal{R} = \mathbb{F}_2[X]/(X^p - 1)$, где p простое. Секретный ключ представляет собой случайно выбранную пару $(x, y) \in \mathcal{R}^2$, а открытый ключ — пару $(h, s = x + hy)$, где $h \in \mathcal{R}$ выбирается случайным образом и используется для построения порождающей матрицы $G \in \mathbb{F}_2^{k \times n}$ кода. Поскольку секретный ключ генерируется независимо от кода, в самом коде нет скрытой структуры. Проверочная матрица открыта, что позволяет редуцировать уровень безопасности независимо от алгоритма декодирования, используемого в расшифровании.

Чтобы зашифровать сообщение $m \in \mathbb{F}_2^k$, необходимо выбрать случайным образом три элемента $e, r_1, r_2 \in \mathcal{R}$ подходящего веса. Тогда зашифрованный текст представляет собой пару $(u, v) = (r_1 + hr_2, mG + sr_2 + e)$. Для расшифрования шифртекста необходимо применить алгоритм декодирования для вектора $v - uy$. Декодер, лежащий в основе схемы HQC, представляет собой конкатенацию кодов Рида — Соломона и Рида — Маллера [19].

Безопасность схемы HQC в смысле IND-CPA зависит от сложности решения задачи QCSD. Декодер, используемый в схеме HQC, имеет корректно определённое минимальное расстояние d и, следовательно, может исправить до $t = \lfloor \frac{d-1}{2} \rfloor$ ошибок. Вероятность того, что зашифрованный текст содержит вектор ошибки e и $wt_H(e) > t$, используется для получения верхней границы частоты сбоев процедуры расшифрования. Как и в случае с другими кодовыми схемами, наиболее известные атаки на HQC представляют собой декодирование на основе информационных совокупностей и его модификации. Безопасность IND-CCA обеспечивается с помощью использования преобразования Фуджисаки — Окомото.

В табл. 1 приведён сравнительный анализ актуальных криптосистем: классической схемы Мак-Элиса, BIKE и HQC; все представлены с уровнями стойкости 1, 3 и 5.

Таблица 1

Размеры ключа и зашифрованного текста для актуальных KEM схем (в байтах)

Схема	Уровень стойкости	Открытый ключ	Закрытый ключ	Шифр-текст
McEliece348864	1	261 120	6492	128
McEliece460896	3	524 160	13 608	188
McEliece6688128	5	104 992	13 932	240
McEliece6960119	5	1 047 319	13 948	226
McEliece8192128	5	1 357 824	14 120	240
BIKE	1	1540	280	1572
	3	3082	418	3114
	5	5122	580	5154
HQC-128	1	2249	40	4481
HQC-192	3	4522	40	9026
HQC-256	5	7245	40	14 469

2. Изогении

2.1. Предварительные сведения. *Эллиптической кривой* над полем $\mathbb{F}$ называется гладкая кривая E, заданная уравнением

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6,$$

где $a_1, a_2, a_3, a_4, a_6 \in \mathbb{F}$. Условие гладкости означает, что кривая не имеет сингулярных точек, т. е. точек, в которых обе частные производные функции $y^2 + a_1xy + a_3y - (x^3 + a_2x^2 + a_4x + a_6)$ равны нулю. Если характеристика поля $\mathbb{F}$ не равна 2 или 3, то кривую можно привести (изоморфным преобразованием) к *краткой форме Вейерштрасса*

$$y^2 = x^3 + a\ x + b.$$

Множество точек кривой E вместе с бесконечно удалённой точкой $\mathcal{O}$ обозначается через $E(\mathbb{F})$. Заметим, что кривая может быть задана над одним полем $\mathbb{F}$ (т. е. коэффициенты a_1, a_2, a_3, a_4, a_6 или a, b лежат в $\mathbb{F}$), но при этом точки кривой можно брать над некоторым его расширением, например, алгебраическим замыканием $\overline{\mathbb{F}}$. Такие точки получаются из решений уравнения кривой над алгебраическим замыканием поля или — в общем случае — над расширением поля. Множество точек кривой E (заданной над $\mathbb{F}$), которые имеют координаты из поля $\overline{\mathbb{F}}$, обозначается через $E(\overline{\mathbb{F}})$. На множестве $E(\mathbb{F})$ по известным формулам задаётся групповая операция, называемая сложением (см., например, [20, § 13.1] или [21, § 2.2.2]). При этом бесконечно удалённая точка является нейтральным элементом группы точек. Операцию сложения точки с самой собой ℓ раз будем обозначать через $[\ell]$, а результат её применения к точке $P \in E(\mathbb{F})$ — через $[\ell]P$.

Порядком точки $P \in E(\mathbb{F})$ называется наименьшее натуральное число $\operatorname{ord} P$, при котором $[\operatorname{ord} P]P = \mathcal{O}$. Множество точек

$$E[\ell] = \{P \in E(\overline{\mathbb{F}}) \mid [\ell]P = \mathcal{O}\}$$

образует подгруппу $E[\ell] \subset E(\overline{\mathbb{F}})$, которая называется *подгруппой ℓ-кручения кривой*. Это множество всех точек кривой (с координатами из алгебраического замыкания поля), чей порядок делит ℓ.

Эллиптическая кривая E над конечным полем $\mathbb{F}_q$, где $q = p^n$ — степень простого p, называется *суперсингулярной*, если $|E(\mathbb{F}_q)| \equiv 1 \pmod{p}$.

Изогенией двух эллиптических кривых E_1, E_2 над одним и тем же полем $\mathbb{F}$ называется ненулевой гомоморфизм эллиптических кривых, задаваемый рациональными отображениями. Заметим, что в литературе также встречаются другие эквивалентные определения:

1) гомоморфизм кривых E_1 и E_2, который над замыканием поля $\mathbb{F}$ сюръективен и имеет конечное ядро;

2) сюръективный морфизм кривых, отображающий единицу группы точек E_1 в единицу группы точек E_2.

Данные определения получаются из соответствующих определений для более общих объектов — проективных кривых и абелевых многообразий, поэтому они сильно упрощаются в частном случае — случае эллиптических кривых.

Явно изогении задаются в виде рациональных функций:

$$\varphi\colon (x, y) \mapsto \left(\frac{f_1(x, y)}{f_2(x, y)}, \frac{g_1(x, y)}{g_2(x, y)}\right)$$

для некоторых $f_1, f_2, g_1, g_2 \in \mathbb{F}[x, y]$. Используя замену $y^2 \mapsto x^3 + a\,x + b$, изогению можно привести [22, лемма 4.26] к виду

$$\varphi\colon (x,y) \mapsto \left(\frac{u(x)}{v(x)}, \frac{s(x)y}{t(x)}\right),$$

где $u, v, s, t \in \mathbb{F}[x]$. Такая форма изогении называется стандартной. *Степень* изогении определяется как $\deg\varphi = \max(\deg u, \deg v)$. Изогения называется *сепарабельной*, если производная $\frac{u}{v}$ по x не равна 0, и *несепарабельной* в противном случае. Для сепарабельных изогений выполняется условие $\deg\varphi = |\ker\varphi|$, где $\ker\varphi = \{P \in E_1(\overline{\mathbb{F}}) \mid \varphi(P) = \mathcal{O}_{E_2}\}$.

Кривые, между которыми существует изогения степени ℓ, называются *ℓ-изогенными*. Имея подгруппу $G \subseteq E_1(\overline{\mathbb{F}}_q)$, можно построить изогению степени $\ell = |G|$ и уравнение соответствующей изогенной кривой E_2, используя формулы Велу [23]. Кривая E_2, построенная по подгруппе G группы точек кривой E_1, обозначается также через E_1/G и называется *фактор-кривой* E_1 по модулю G. Для каждой изогении $\varphi\colon E_1 \to E_2$ существует изогения $\widehat{\varphi}\colon E_2 \to E_1$ такая, что $\widehat{\varphi} \circ \varphi = [\deg\varphi]$, называемая *дуальной*.

Для описания атаки Кастрика и Декру на схему SIDH в последующих разделах нам потребуется выйти за пределы эллиптических кривых к более общим объектам — абелевым многообразиям.

Абелевым многообразием называется группа, образованная множеством решений системы уравнений, составленной из однородных многочленов от нескольких переменных, при выполнении условий:

1) групповой закон задаётся рациональными отображениями, определёнными во всех точках;

2) данное множество является неприводимым многообразием, т. е. оно не может быть представлено в виде объединения двух множеств, которые являются множествами решений систем уравнений, составленных из однородных многочленов.

Размерность абелева многообразия определяется как размерность соответствующей системы уравнений. Эллиптическая кривая представляет собой абелево многообразие размерности 1 (строго говоря, абелевым многообразием является множество $E(\mathbb{F})$, которое содержит в себе бесконечно удалённую точку; для её явного определения требуется переход к проективным координатам и задание кривой однородным многочленом). Абелево многообразие размерности 2 называется *абелевой поверхностью*. Есть два типа абелевых поверхностей — произведение двух эллиптических кривых и якобианы кривых рода 2 (определение якобиана кривой рода 2 и его свойства можно найти в [24]). Первый случай часто можно свести ко второму методом склейки двух эллиптических кривых в якобиан кривой рода 2. Случаи, когда склейку невозможно осуществить, относительно редки, их классификация приведена в работе [25]. Соответствующие теоремы лежат в основе атаки Кастрика — Декру. Так же, как и для случая эллиптических кривых, подгруппе G

абелева многообразия A можно поставить в соответствие некоторую изогению с ядром G в другое абелево многообразие A/G, которое называется *фактор-многообразием* по подгруппе G. Однако в общем случае формулы и алгоритмы для вычисления изогений и фактор-многообразий (аналоги формул Велу) достаточно громоздки, поэтому, как правило, на практике ограничиваются якобианами кривых, что накладывает дополнительные ограничения на выбор подгруппы G (необходимо выбирать максимальные изотропические подгруппы, подробнее см. в [26]). В случае якобианов задача построения фактор-многообразия и изогении сильно упрощается: например, в случае изогений степени 2 есть явные формулы Ришело [27, утверждение 1], которых достаточно для получения секретного ключа одного из участников в протоколе SIDH/SIKE.

2.2. Построение криптосистем на действиях групп. Пусть X — некоторое множество, а G — группа. Будем говорить, что G действует на множестве X, если задано отображение $* : G \times X \to X$ такое, что для любых $g_1, g_2 \in G$ и $x \in X$ выполняется $g_1 * (g_2 * x) = (g_1 g_2) * x$.

В терминах действия группы можно описать многие схемы шифрования с открытым ключом, протоколы распределения и инкапсуляции ключа (key encapsulation mechanism, KEM) [28]. При этом действие группы должно обладать некоторым криптографическим («трудновычислимым») свойством, как например, следующие:

• группа G действует как односторонняя функция, т. е. для любых $x_1, x_2 \in X$ нахождение элемента $g \in G$ такого, что $x_1 = g * x_2$, является трудной задачей (даже в предположении, что такой элемент существует);

• действие группы обладает свойством псевдослучайного генератора, т. е. для случайно выбранного элемента $g \in G$ злоумышленник не может отличить множество принятых векторов $\{(x_i, g * x_i)\}_{i \in I}$ от множества векторов вида $\{(x_i, u_i)\}_{i \in I}$, где u_i, $i \in I$, — равномерно распределённые на X случайные величины.

Например, опишем в терминах действия группы на множестве широко известный алгоритм обмена ключами (выработки общего секрета) Диффи — Хеллмана. В качестве пространства открытых ключей выберем некоторую группу большого простого порядка $X = \mathbb{Z}_p = \langle x \rangle$. На множестве X определим действие мультипликативной группы $G = \mathbb{Z}_p^*$ как отображение $*\colon G \times X \to X\colon z * h \mapsto h^z$. Генерация общего секретного ключа происходит следующим образом.

1. Пользователь A выбирает секретный ключ $a \in G$ и вычисляет его действие на образующем элементе $x \in X$. Полученное значение $a * x = x^a$ является открытым ключом и отправляется пользователю B.

2. Пользователь B выбирает секретный ключ $b \in G$, вычисляет с помощью действия группы открытый ключ $b * x = x^b$ и отправляет его пользователю A.

После этого каждый пользователь действует своим секретным ключом $a \in G$ или $b \in G$ на полученное значение $b*x$ или $a*x$ соответственно, получая общий секрет $a*(b*x) = b*(a*x) = x^{ab} \in X$.

В данном примере группа G действует как односторонняя функция. В самом деле, злоумышленник знает пары значений $(x, a*x)$, $(x, b*x) \in X \times X$, однако нахождение по ним элементов $a \in G$, $b \in G$ является трудной задачей, так как по сути это задача дискретного логарифмирования в $\mathbb{Z}_p^*$ (discrete logarithm problem, DLP).

2.3. Схема SIDH/SIKE. Схема SIDH (supersingular isogeny Diffie — Hellman), предложенная в 2011 г. де Фео, Яо и Плутом [29], представляет собой протокол обмена ключами, аналогичный протоколу Диффи — Хеллмана, где в качестве X используется множество суперсингулярных эллиптических кривых над конечным полем, а элементы $\varphi_A, \varphi_B \in G$ — изогении суперсингулярных кривых. Схема уязвима к атаке Кастрика — Декру, описанной в п. 2.5, и вследствие этого небезопасна. Кратко опишем саму схему обмена ключами SIDH.

Публичные параметры схемы:

• простое число $p = \ell_A^{e_A} \ell_B^{e_B} c \pm 1$, где ℓ_A, ℓ_B — малые простые, c — фиксированный дополнительный множитель, как правило, малый;

• E — суперсингулярная кривая над $\mathbb{F}_{p^2}$ с числом рациональных точек, равным $|E(\mathbb{F}_{p^2})| = \left(\ell_A^{e_A} \ell_B^{e_B} c\right)^2$.

Открытый ключ:

A: (P_A, Q_A) — базис подгруппы точек $E\left[\ell_A^{e_A}\right] \subseteq E(\mathbb{F}_{p^2})$;

B: (P_B, Q_B) — базис подгруппы $E\left[\ell_B^{e_B}\right] \subseteq E(\mathbb{F}_{p^2})$.

Секретный ключ:

A: $m_A, n_A \in \mathbb{Z}/\ell_A^{e_A}\mathbb{Z}$; изогения $\varphi_A \colon E \to E_A$, заданная своим ядром $\ker \varphi_A = \langle [m_A]P_A + [n_A]Q_A \rangle$.

B: $m_B, n_B \in \mathbb{Z}/\ell_B^{e_B}\mathbb{Z}$; изогения $\varphi_B \colon E \to E_B$, заданная своим ядром $\ker \varphi_B = \langle [m_B]P_B + [n_B]Q_B \rangle$.

Схема обмена ключами SIDH следующая.

1. Пользователь A выбирает случайным образом секретные параметры $m_A, n_A \in \mathbb{Z}/\ell_A^{e_A}\mathbb{Z}$ (не должны делиться на ℓ_A одновременно), которые определяют изогению φ_A. Затем вычисляет образы базисных точек $\varphi_A(P_B), \varphi_A(Q_B)$ и отправляет их пользователю B вместе с кривой E_A.

2. Пользователь B выбирает случайным образом секретные параметры $m_B, n_B \in \mathbb{Z}/\ell_B^{e_B}\mathbb{Z}$ (не должны делиться на ℓ_B одновременно), которые определяют изогению φ_B. Затем вычисляет образы базисных точек $\varphi_B(P_A), \varphi_B(Q_A)$ и отправляет их пользователю A вместе с кривой E_B.

3. Пользователь A вычисляет изогению $\varphi'_A \colon E_B \to E_{AB}$, которая определяется своим ядром $\ker \varphi'_A = \langle [m_A]\varphi_B(P_A) + [n_A]\varphi_B(Q_A) \rangle$.

4. Пользователь B вычисляет аналогично изогению $\varphi'_B\colon E_A \to E_{BA}$.

В итоге пользователи A и B имеют кривые E_{AB} и E_{BA} соответственно, связанные изогенией $E_{AB} = \varphi'_B(\varphi_A(E)) \cong \varphi'_A(\varphi_B(E)) = E_{BA}$. В качестве общего секретного ключа можно использовать j-инварианты [20, § 13.1] этих кривых, так как $j(E_{AB}) = j(E_{BA})$.

SIKE — протокол инкапсуляции ключа (KEM), предложенный к стандартизации в NIST в качестве алгоритма, устойчивого к квантовым атакам. Конструкция протокола вполне повторяет схему SIDH, дополняя её выбором конкретных «стандартных» параметров и рядом технических модификаций, позволяющих использовать SIDH для инкапсуляции ключа. После публикации атаки [30] протокол считается небезопасным.

2.4. Схема CSIDH. Схема CSIDH (commutative supersingular isogeny Diffie — Hellman) — ещё один протокол обмена ключами, безопасность которого основана на сложности нахождения изогении между двумя суперсингулярными кривыми. Конструкция схемы основана на криптосистеме Ростовцева — Столбунова, однако вместо обычных эллиптических кривых используются суперсингулярные эллиптические кривые. Кроме того, в отличие от схемы SIDH в CSIDH используется действие коммутативной группы. Впервые протокол CSIDH описан в 2018 г. Кастриком, Ланге и др. [31], впоследствии опубликовано множество его технических оптимизаций [32].

Теорема (форма Монтгомери [31, утверждение 8]). *Пусть $\mathbb{F}_p$ — конечное поле характеристики $p \equiv 3 \pmod 8$, E — суперсингулярная кривая над $\mathbb{F}_p$. Тогда кольцо эндоморфизмов кривой имеет вид* $\operatorname{End}(E) = \mathbb{Z}[\pi_p]$ *в том и только том случае, когда существует единственный элемент $A \in \mathbb{F}_p$ такой, что $E \simeq \widetilde{E}\colon y^2 = x^3 + Ax^2 + x$, где $\pi_p(x, y) = (x^p, y^p)$ — эндоморфизм Фробениуса.*

Кривая $\widetilde{E}$, удовлетворяющая условиям теоремы, называется *формой Монтгомери* эллиптической кривой, а элемент $A \in \mathbb{F}_p$ — *коэффициентом Монтгомери*. Обозначим через $\mathcal{E}\ell\ell_p(\mathbb{Z}[\pi_p])$ множество эллиптических кривых, удовлетворяющих вышеперечисленным условиям, т. е. допускающих представление в форме Монтгомери.

Конструкция схемы CSIDH основана на действии *группы классов идеалов* $G = \operatorname{Cl}(\mathbb{Z}[\pi_p])$ на множестве $X = \mathcal{E}\ell\ell_p(\mathbb{Z}[\pi_p])$ эллиптических кривых, определённых над полем $\mathbb{F}_p$ и имеющих кольцо эндоморфизмов, изоморфное $\mathbb{Z}[\pi_p]$. Определение и базовые свойства группы классов идеалов описаны в [33, разд. 4.9].

Действие класса идеалов $\bigl[\prod_i \mathfrak{l}_i^{e_i}\bigr] \in G$ на кривую $E \in X$ вычисляется следующим образом. Так как $\pi^2 = -p \equiv 1 \pmod{\ell_i}$, собственные

значения действия эндоморфизма Фробениуса на все подгруппы ℓ_i-кручения равны $\lambda_i = \pm 1$. Следовательно, для вычисления действия каждого из классов $[\mathfrak{l}_i]$ можно найти все $\mathbb{F}_p$-рациональные (или $\mathbb{F}_{p^2}$-рациональные в случае собственного значения $\lambda_i = -1$) точки порядка ℓ_i и применить к ним формулы Велу. Более подробно этот алгоритм описан в [31, § 3]. Кратко опишем схему обмена ключами CSIDH.

Публичные параметры схемы:

• простое число $p = 4\ell_1 \cdots \ell_n - 1$, где $\ell_1, \ldots, \ell_n$ — попарно различные малые нечётные простые;

• суперсингулярная эллиптическая кривая $E_0\colon y^2 = x^3 + x$ над $\mathbb{F}_p$.

Схема обмена ключами CSIDH следующая.

1. Пользователь A формирует целочисленный вектор $(e_1, \ldots, e_n) \in \{-m, \ldots, m\}^n$, после чего определяет класс идеалов $[\mathfrak{a}] = \left[\mathfrak{l}_1^{e_1} \cdots \mathfrak{l}_n^{e_n}\right] \in \mathrm{Cl}(\mathbb{Z}[\pi_p])$, где $m > 0$ — наименьшее целое число такое, что $2m + 1 \geqslant \sqrt[n]{|\mathrm{Cl}(\mathbb{Z}[\pi_p])|}$,

$$[\mathfrak{l}_i] = [(\ell_i, \pi_p - 1)], \quad [\mathfrak{l}_i]^{-1} = [(\ell_i, \pi_p + 1)].$$

Далее пользователь вычисляет действие $[\mathfrak{a}] * E_0$, приводит уравнение полученной кривой к форме Монтгомери $E_A\colon y^2 = x^3 + Ax^2 + x$. Полученный коэффициент Монтгомери $A \in \mathbb{F}_p$ является открытым ключом пользователя A, а исходный случайный вектор $(e_1, \ldots, e_n)$ — секретным ключом.

2. Пользователь B формирует целочисленный вектор $(e_1, \ldots, e_n) \in \{-m, \ldots, m\}^n$, затем определяет класс $[\mathfrak{b}] = \left[\mathfrak{l}_1^{e_1} \cdots \mathfrak{l}_n^{e_n}\right] \in \mathrm{Cl}(\mathbb{Z}[\pi_p])$. Пользователь вычисляет действие $[\mathfrak{b}] * E_0$, приводит уравнение полученной кривой к форме Монтгомери $E_B\colon y^2 = x^3 + Bx^2 + x$. Полученный коэффициент Монтгомери $B \in \mathbb{F}_p$ является открытым ключом пользователя B, а исходный случайный вектор $(e_1, \ldots, e_n)$ — секретным ключом.

В итоге пользователи A и B имеют ключевые пары $([\mathfrak{a}], A)$ и $([\mathfrak{b}], B)$ соответственно. Пользователь A действует своим секретным ключом $[\mathfrak{a}]$ на принятую кривую E_B и получает $[\mathfrak{a}] * E_B = [\mathfrak{a}][\mathfrak{b}]E_0$. Пользователь B действует симметрично. В качестве общего секретного ключа может выступать коэффициент Монтгомери общей кривой $[\mathfrak{a}][\mathfrak{b}]E_0$.

В отличие от схемы SIDH схема CSIDH не использует значения секретных изогений в точках, и поэтому атака Кастрика — Декру её не затрагивает.

2.5. Атака Кастрика — Декру. Протокол обмена ключами SIDH использует в своей работе образы $\varphi_B(P_A)$, $\varphi_B(Q_A)$, $\varphi_A(P_B)$, $\varphi_A(Q_B)$ открытых ключей участников протокола под действием секретных изогений Алисы φ_A и Боба φ_B. Открытые ключи (P_A, Q_A) и (P_B, Q_B) являются образующими групп $E\left[\ell_A^{e_A}\right] = \langle P_A, Q_A \rangle$ и $E\left[\ell_B^{e_B}\right] = \langle P_B, Q_B \rangle$.

Долгое время считалось, что эта дополнительная информация не позволяет взломать криптосистему. Однако в августе 2022 г. Кастрик и Декру опубликовали препринт [30], в котором описывается полиномиальная атака на криптосистему SIKE — версию SIDH-кандидата на стандартизацию NIST. Атака была рассчитана на использование начальной кривой из параметров SIKE. В [34] параллельно работающие в том же направлении Майно и Мартиндейл представили вариант атаки для любой начальной кривой. Изначально доказательство полиномиальности атаки в указанных работах основывалось на эвристиках. Этот недостаток был устранён Робером в [35], где было доказано, что атака занимает классическое детерминированное полиномиальное время. В этом пункте приведём описание атаки Кастрика — Декру и её последствий.

Пусть $B = \langle [m_B]P_A + [n_B]Q_A \rangle$ — секретное ядро изогении Боба, по которому с помощью формул Велу можно построить секретную изогению φ_B и получить открытые параметры Боба: уравнение кривой E/B и пару $(\varphi_B(P_A),\ \varphi_B(Q_A))$. Тогда в строгом виде задача восстановления ключа в SIDH формулируется следующим образом: по известной четвёрке $(E, E/B, \varphi_B(P_A),\ \varphi_B(Q_A))$ восстановить φ_B.

Изогения φ_B имеет степень $\ell_B^{e_B}$ и представляет собой композицию изогений степени ℓ_B, т. е. $\varphi_B = \varphi_{e_B} \circ \cdots \circ \varphi_1$, тем самым имеется цепочка изогений

$$E \xrightarrow{\varphi_1} E_1 \xrightarrow{\varphi_2} E_2 \xrightarrow{\varphi_3} \ldots \xrightarrow{\varphi_{e_B}} E/B.$$

Изогении φ_i имеют степень ℓ_B, причём для криптосистем это число ℓ_B выбирается малым ($\ell_B = 3$ в SIKE), поэтому и количество возможных вариантов для φ_i мало. Точнее, так как $\ker \varphi_i$ — подгруппа $E_{i-1}[\ell_{e_B}] \simeq (\mathbb{Z}/\ell\mathbb{Z})^2$, всего имеется ℓ_B^2 вариантов для выбора φ_i. Имея эффективно (за полиномиальное время) вычисляемый критерий для определения правильного варианта, можно последовательно перебирать все варианты для определения φ_1, затем, найдя φ_1, перейти к перебору вариантов для изогении φ_2 и т. д., пока не получим φ_{e_B}. До работы Кастрика и Декру такого эффективного критерия известно не было. Кастрик и Декру предложили использовать в качестве такого критерия теорему Кани [25, теорема 2.8] и изогении абелевых поверхностей для проверки её выполнения. Для формулировки теоремы и её применения в качестве критерия нам потребуется ввести несколько дополнительных определений.

Определение 1. *Алмазной изогенной конфигурацией степени N* называется тройка (φ, G_1, G_2) такая, что

(1) $\varphi\colon E' \to E''$ — изогения;
(2) $G_1, G_2 \subseteq \ker \varphi$;
(3) $\deg \varphi = |G_1| \cdot |G_2|$;
(4) $N = |G_1| + |G_2|$ и $G_1 \cap G_2 = \{0\}$.

Подгруппа $\langle R_1, R_2\rangle$ группы $E' \times E''$ такая, что $|R_1| = |R_2| = N$ (другими словами, (N, N)-подгруппа), называется *разложимой*, если фактор-поверхность $(E' \times E'')/\langle R_1, R_2\rangle$ изоморфна декартову произведению двух эллиптических кривых. В противном случае подгруппа называется *неразложимой*, и фактор-поверхность $(E' \times E'')/\langle R_1, R_2\rangle$ в этом случае изоморфна якобиану кривой рода 2. Атака Кастрика — Декру основана на том факте, что первый случай очень редко встречается по сравнению со вторым. Теорема Кани описывает данный редкий случай, она утверждает (в упрощённом виде), что (N, N)-подгруппа разложима тогда и только тогда, когда она получается из некоторой алмазной изогенной конфигурации степени N. Точнее,

$$\langle R_1, R_2\rangle = \langle (P, [x]\varphi(P)), (Q, [x]\varphi(Q))\rangle$$

для некоторого целого x и некоторых P, Q таких, что $E'[N] = \langle P, Q\rangle$.

Теорема Кани может быть применена для нахождения φ_1 методом «вписывания» изогении $\varphi'_B = \varphi_{e_B} \circ \cdots \circ \varphi_2$ в некоторую алмазную конфигурацию степени ℓ^{e_A} (напомним, что общий секретный ключ в SIDH получается из композиции секретных изогений степени $\ell_A^{e_A}$ и $\ell_B^{e_B}$), чтобы гарантировать, что при правильном выборе изогении φ_1 теорема Кани выполняется, т. е. соответствующая фактор-группа разложима, а при всех остальных выборах φ_1 получаются (в подавляющем большинстве случаев) неразложимые группы.

Предположим, что φ'_1 — один из возможных вариантов для φ_1 и $E'_1 = E/\ker\varphi_1$ (если $\varphi'_1 = \varphi_1$, то $E_1 = E'_1$). Для построения алмазной изогенной конфигурации подбирается вспомогательная изогения γ степени $\ell^{e_A} - \ell^{e_B - 1}$ (подойдёт любая изогения такой степени) из кривой E'_1 в некоторую эллиптическую кривую C (в качестве γ можно взять подходящий эндоморфизм, тогда $C \simeq E'_1$). При правильном выборе φ'_1, т. е. когда $\varphi'_1 = \varphi_1$, тройка $(\varphi'_B \circ \widehat{\gamma}, \ker\widehat{\gamma}, \gamma(B))$, где $\varphi'_B = \varphi_{e_B} \circ \cdots \circ \varphi_2$ — алмазная конфигурация степени $\ell_A^{e_A}$ и подгруппа

$$G = \langle (\gamma(\varphi'_1(P_A)), \varphi_B(P_A)), (\gamma(\varphi'_1(Q_A)), \varphi_B(Q_A))\rangle \subseteq C \times E/B$$

разложима по теореме Кани. Тем самым при неправильном выборе φ'_1 подгруппа неразложима и соответствующая фактор-поверхность является якобианом кривой рода 2.

Отсюда получается следующий метод для определения φ_1.

1. Выбрать φ'_1.

2. Построить для некоторой эллиптической кривой C вспомогательную изогению $\gamma\colon E'_1 \to C$ степени $\ell_A^{e_A} - \ell_B^{e_B - 1}$.

3. Шаг склейки. Построить фактор-поверхность A произведения $C \times E/B$ по подгруппе

$$G = \langle (\gamma(\varphi'_1(P_A)), \varphi_B(P_A)), (\gamma(\varphi'_1(Q_A)), \varphi_B(Q_A))\rangle.$$

4. Шаг разложения. Определить, изоморфна фактор-поверхность произведению двух эллиптических кривых или нет. Если изоморфна, то $\varphi_1' = \varphi_1$, в противном случае перейти к шагу 1.

После нахождения φ_1 данный метод можно применить и для нахождения φ_2, выбрав в качестве γ изогению степени $\ell_A^{e_A} - \ell_B^{e_B - 2}$, а затем аналогичным образом — для нахождения всех остальных φ_i, и получить на выходе секретную изогению φ_B. Рассмотрим подробнее шаги метода.

Выбор φ_1'. Выбирается подгруппа $E[\ell_B]$, затем по формулам Велу строятся изогения φ_1' и уравнение для кривой $E/\ker\varphi_1'$. Всего возможных выборов ℓ_B^2, так как $E[\ell_B] \simeq (\mathbb{Z}/\ell_B\mathbb{Z})^2$.

Построение вспомогательной изогении γ. Предположим, что надо найти φ_i. Тогда на этом шаге необходимо построить произвольную изогению γ степени $c = \ell_A^{e_A} - \ell_B^{e_B - i}$. В случае, если c гладкое, т. е. раскладывается на малые простые числа размера $(\log p)^{O(1)}$, это можно сделать с помощью формул Велу. В случае наличия у кривой эндоморфизма η малой нормы можно также использовать факторизацию c в $\mathbb{Z}[\eta]$.

Например, кривая $y^2 = x^3 + x$ имеет автоморфизм $\eta\colon (x, y) \mapsto (-x, iy)$. В этом случае можем найти с помощью алгоритма Корначчи [33, алгоритм 1.5.2] целые числа u, v такие, что $c = u^2 + v^2$ и $c = (u + iv)(u - iv)$. Тогда в качестве γ можно взять $u + iv\colon P \mapsto [u]P + [v]\eta(P)$. Аналогично можно построить γ для кривой $y^2 = x^3 + 6x^2 + x$ из SIKE, найдя представление $c = u^2 + 4v^2 = (u + 2iv)(u - 2iv)$.

Шаг склейки. Подгруппа G изоморфна $(\mathbb{Z}/\ell^{e_A}\mathbb{Z})^2$, поэтому построение фактор-поверхности $(C \times E/B)/G$ можно свести к построению цепочки изогений степени ℓ_A с ядрами, изоморфными $(\mathbb{Z}/\ell^{e_A}\mathbb{Z})^2$ (т. е. (ℓ_A, ℓ_A)-изогений), которая ведёт в целевую фактор-поверхность $(C \times E/B)/G$. Чтобы построить такую цепочку, кривые C и E/B сначала «склеиваются» в якобиан J_H кривой H рода 2. Другими словами, строится такая кривая H, что $J_H \sim C \times E/B$. Для $\ell_A = 2$, как в SIKE, это можно сделать с помощью формул Хау — Лепревоста — Пунена [36]. Затем после нахождения J_H итоговая фактор-поверхность A строится по цепочке изогений якобианов кривых рода 2 степени ℓ_A^2. Для случая $\ell_A = 2$ такие изогении и соответствующие им уравнения кривых рода 2 можно построить с помощью формул Ришело. Для $\ell_A = 3$ можно использовать формулы [37]. Для общего случая можно использовать алгоритмы из работ [38, 39].

Шаг разложения. Данный шаг можно объединить с шагом склейки, так как определение, является ли итоговая абелева поверхность произведением эллиптических кривых, осуществляется при попытке построения якобиана кривой на последнем шаге цепочки изогений степени ℓ_A.

В случае произведения эллиптических кривых кривой с таким якобианом не существует и будет получена ошибка. Например, для $\ell_A = 2$ появится деление на нуль при вычислении формул Ришело.

2.6. Перспективы. В атаке Кастрика — Декру используется информация о действии секретной изогении на базисы групп кручения, и применить её напрямую к общей задаче вычисления изогении между двумя заданными эллиптическими кривыми не получится, поэтому схемы из [31, 40], не использующие данную информацию, остаются стойкими к атаке.

В табл. 2 представлены параметры для актуальных схем на изогениях. Для схемы CRS (Кувейна — Ростовцева — Столбунова) указаны размеры параметров, предложенные в [41, § 4, табл. 3]. Для схемы OSIDH указаны размеры ключей, предложенные в недавней статье [42, § 5.2] по криптоанализу данной схемы.

Таблица 2

Размеры ключей для актуальных схем обмена ключами на изогениях (в байтах)

Схема	Уровень стойкости	Открытый ключ	Закрытый ключ	Общий ключ
CRS [41]	128/56	64	8	64
OSIDH [43]	128/128	36	31	36
CSIDH-512	128/62	64	32	64

Заметим, что указанные схемы на изогениях, несмотря на небольшие размеры ключей, достаточно медленные и для использования на практике требуют оптимизации.

Переход к кривым рода 2. Для обхода атаки Кастрика — Декру можно было бы рассмотреть кривые рода 2. Например, в [26] предложен соответствующий вариант схемы SIDH. Однако в этом случае используются суперсингулярные кривые рода 2, которые изогенны декартовому произведению суперсингулярных эллиптических кривых, значит, можно ожидать адаптации атаки Кастрика — Декру и к этому случаю. Использование же несупресингулярных кривых ведёт к субэкспоненциальным квантовым атакам из-за коммутативности кольца эндоморфизмов якобиана кривой.

Заключение

Постквантовая криптография на изогениях и кодах представляет собой перспективную область исследований с большим числом открытых проблем. В представленной статье приведён обзор основных подходов к построению постквантовых криптосистем на основе кодов и изогений,

а также вычислительно трудных задач, лежащих в основе их стойкости. В случае кодов с учётом размеров открытого и закрытого ключей классическая схема Мак-Элиса существенно проигрывает и схеме BIKE, и схеме HQC. Несмотря на то, что схема HQC обеспечивает надёжные гарантии безопасности, а также разумную частоту отказов при расшифровании, она проигрывает BIKE относительно размеров открытого ключа и зашифрованного текста. Тем самым схема BIKE показала себя вполне конкурентно способной. В случае изогений, несмотря на малый размер ключа, главной проблемой остаётся медленная скорость работы схем. Атака Кастрика — Декру вывела из рассмотрения наиболее перспективную с точки зрения практики схему SIDH/SIKE и все схемы, для оптимизации которых использовались значения изогении в точках кручения. Таким образом, наиболее важным направлением исследований в области изогений является исследование вопросов оптимизации имеющихся схем.

Финансирование работы

Работа первого, третьего и четвёртого авторов выполнена при поддержке Северо-западного центра математических исследований им. С. Ковалевской (Балтийский федеральный университет им. И. Канта) в рамках соглашения с Министерством науки и высшего образования Российской Федерации (соглашение № 075–02–2023–934). Работа второго, пятого, шестого, седьмого и восьмого авторов выполнена при поддержке Математического центра в Академгородке в рамках соглашения с Министерством науки и высшего образования Российской Федерации (соглашение № 075–15–2022–282).

Конфликт интересов

Авторы заявляют, что у них нет конфликта интересов.

Литература

1. **Малыгина Е. С., Токарева Н. Н., Куценко А. В.** [и др.]. Постквантовые криптосистемы: открытые вопросы и существующие решения. Криптосистемы на решётках // Дискрет. анализ и исслед. операций. 2023. Т. 30, № 4. С. 46–90.
2. **Courtois N. T., Finiasz M., Sendrier N.** How to achieve a McEliece-based digital signature scheme // Advances in cryptology — ASIACRYPT'01. Proc. Int. Conf. Theory and Application of Cryptology (Gold Coast, Australia, Dec. 9–13, 2001). Heidelberg: Springer, 2001. P. 157–174. (Lect. Notes Comput. Sci.; V. 2248).
3. **Stern J.** A new paradigm for public key identification // IEEE Trans. Inf. Theory. 1996. V. 42, No. 6. P. 1757–1768.
4. **Childs A., Jao D., Soukharev V.** Constructing elliptic curve isogenies in quantum subexponential time // J. Math. Cryptology. 2014. V. 8, No. 1. P. 1–29. DOI: 10.1515/jmc-2012-0016.

5. **McEliece R. J.** A public-key cryptosystem based on algebraic coding theory // The deep space network. Progress report 42-44. Pasadena, CA: California Inst. Technol., 1978. P. 114–116.
6. **Niederreiter H.** Knapsack-type cryptosystems and algebraic coding theory // J. Prob. Contr. Inform. Theory. 1986. V. 15, No. 2. P. 159–166.
7. **Niederreiter H., Xing C.** Algebraic geometry in coding theory and cryptography. Princeton, NJ: Princeton Univ. Press, 2009. 272 p.
8. **Minder L., Shokrollahi A.** Cryptanalysis of the Sidelnikov cryptosystem // Advances in cryptology — EUROCRYPT'07. Proc. Int. Conf. Theory and Applications of Cryptographic Techniques (Barcelona, Spain, May 20–24, 2007). Heidelberg: Springer, 2007. P. 347–360. (Lect. Notes Comput. Sci.; V. 4515). DOI: 10.1007/978-3-540-72540-4_20.
9. **Berlekamp E., McEliece R. J., van Tilborg H.** On the inherent intractability of certain coding problems // IEEE Trans. Inf. Theory. 1978. V. 24, No. 3. P. 384–386.
10. **Lee P. J., Brickell E. F.** An observation on the security of McEliece's public-key cryptosystem // Advances in cryptology — EUROCRYPT'88. Proc. Workshop Theory and Application of Cryptographic Techniques (Davos, Switzerland, May 25–27, 1988). Heidelberg: Springer, 1988. P. 275–280. (Lect. Notes Comput. Sci.; V. 330).
11. **Petrank E., Roth R.** Is code equivalence easy to decide? // IEEE Trans. Inf. Theory. 1997. V. 43, No. 5. P. 1602–1604.
12. **Sendrier N.** Finding the permutation between equivalent linear codes: The support splitting algorithm // IEEE Trans. Inf. Theory. 2000. V. 46, No. 4. P. 1193–1203. DOI: 10.1109/18.850662.
13. **Misoczki R., Tillich J.-P., Sendrier N., Barreto P.** MDPC-McEliece: New McEliece variants from moderate density parity-check codes // Proc. IEEE Int. Symp. Information Theory (Istanbul, Turkey, Jul. 7–12, 2013). Los Alamitos, CA: IEEE Comput. Soc., 2013. P. 2069–2073. DOI: 10.1109/ISIT.2013.6620590.
14. **Drucker N., Gueron S., Kostic D.** QC-MDPC decoders with several shades of gray // Post-quantum cryptography. Proc. Int. Conf. (Paris, France, Apr. 15–17, 2020). Cham: Springer, 2020. P. 35–50. (Lect. Notes Comput. Sci.; V. 12100).
15. **Torres R. C., Sendrier N.** Analysis of information set decoding for a sublinear error weight // Post-quantum cryptography. Proc. Int. Conf. (Fukuoka, Japan, Feb. 24–26, 2016). Cham: Springer, 2016. P. 144–161. (Lect. Notes Comput. Sci.; V. 9606).
16. **Fujisaki E., Okamoto T.** Secure integration of asymmetric and symmetric encryption schemes // J. Cryptology. 2013. V. 26. P. 80–101.
17. **Bindel N., Hamburg M., Hövelmanns K., Hülsing A., Persichetti E.** Tighter proofs of CCA security in the quantum random oracle model // Theory of cryptography. Proc. Int. Conf. (Nuremberg, Germany, Dec. 1–5, 2019). Cham: Springer, 2019. P. 61–90. (Lect. Notes Comput. Sci.; V. 11892). DOI: 10.1007/978-3-030-36033-7_3.

18. **Aguilar-Melchor C., Blazy O., Deneuville J.-C., Gaborit P., Zémor G.** Efficient encryption from random quasi-cyclic codes // IEEE Trans. Inf. Theory. 2018. V. 64, No. 5. P. 3927–3943.
19. **Aragon N., Gaborit P., Zémor G.** HQC-RMRS, an instantiation of the HQC encryption framework with a more efficient auxiliary error-correcting code. Ithaca, NY: Cornell Univ., 2005. (Cornell Univ. Libr. e-Print Archive; arXiv:2005.10741)
20. **Doche C., Lange T.** Arithmetic of elliptic curves // Handbook of elliptic and hyperelliptic curve cryptography. Boca Raton, FL: Chapman & Hall/CRC Press, 2006. P. 267–302.
21. **Болотов А. А., Гашков С. Б., Фролов А. Б., Часовских А. А.** Элементарное введение в эллиптическую криптографию: Алгебраические и алгоритмические основы. М.: КомКнига, 2006.
22. **Sutherland A.** Ellipic curves. Isogenies. Lecture notes. Cambridge, MA: MIT, 2022. Available at `math.mit.edu/classes/18.783/2022/LectureNotes4.pdf` (accessed Dec. 22, 2023).
23. **Vélu J.** Isogénies entre courbes elliptiques // C. R. Acad. Sci. Paris. Sér. A. 1971. V. 273. P. 238-241.
24. **Duquesne S., Lange T.** Arithmetic of hyperelliptic curves //Handbook of elliptic and hyperelliptic curve cryptography. Boca Raton, FL: Chapman & Hall/CRC Press, 2006. P. 303–353.
25. **Kani E.** The number of curves of genus two with elliptic differentials // J. Reine Angew. Math. 1997. V. 485. P. 93-122.
26. **Flynn E. V., Ti Y. B.** Genus two isogeny cryptography // Post-quantum cryptography. Proc. Int. Conf. (Chongquin, China, May 10–12, 2019). Cham: Springer, 2019. P. 286–306. (Lect. Notes Comput. Sci.; V. 11505).
27. **Castryck W., Decru T., Smith B.** Hash functions from superspecial genus-2 curves using Richelot isogenies // J. Math. Cryptology. 2020. V. 14, No. 1. P. 268–292.
28. **Alamati N., de Feo L., Montgomery H., Patranabis S.** Cryptographic group actions and applications. San Diego: Univ. California, 2020. (Cryptology ePrint Archive, Paper ID 2020/1188). Available at `eprint.iacr.org/2020/1188.pdf` (accessed Dec. 22, 2023).
29. **De Feo L., Jao D., Plût J.** Towards quantum-resistant cryptosystems from supersingular elliptic curve isogenies. San Diego: Univ. California, 2011. (Cryptology ePrint Archive, Paper ID 2011/506). Available at `eprint.iacr.org/2011/506.pdf` (accessed Dec. 22, 2023).
30. **Castryck W., Decru T.** An efficient key recovery attack on SIDH. San Diego: Univ. California, 2022. (Cryptology ePrint Archive, Paper ID 2022/975). Available at `eprint.iacr.org/2022/975.pdf` (accessed Dec. 22, 2023).
31. **Castryck W., Lange T., Martindale C., Panny L., Renes J.** CSIDH: An efficient post-quantum commutative group action. San Diego: Univ. California, 2018. (Cryptology ePrint Archive, Paper ID 2018/383). Available at `eprint.iacr.org/2018/383.pdf` (accessed Dec. 22, 2023).

32. Chi-Domínguez J.-J., Rodríguez-Henríquez F. Optimal strategies for CSIDH // J. Adv. Math. Commun. 2022. V. 16, No. 2. P. 383–411.
33. Cohen H. A course in computational algebraic number theory. Berlin: Springer, 1993. 536 p.
34. Maino L., Martindale C. An attack on SIDH with arbitrary starting curve. San Diego: Univ. California, 2022. (Cryptology ePrint Archive, Paper ID 2022/1026). Available at `eprint.iacr.org/2022/1026.pdf` (accessed Dec. 22, 2023).
35. Robert D. Breaking SIDH in polynomial time. San Diego: Univ. California, 2022. (Cryptology ePrint Archive, Paper ID 2022/1038). Available at `eprint.iacr.org/2022/1038.pdf` (accessed Dec. 22, 2023).
36. Howe E. W., Leprevost F., Poonen B. Large torsion subgroups of split Jacobians of curves of genus two or three // J. Forum Math. 2000. V. 12, No. 3. P. 315–364.
37. Bruin N., Flynn E. V., Testa D. Descent via $(3, 3)$-isogeny on Jacobians of genus 2 curves // J. Acta Arithmetica. 2014. V. 165, No. 3. P. 201–223.
38. Cosset R., Robert D. Computing (ℓ, ℓ)-isogenies in polynomial time on Jacobians of genus 2 curves // Math. Comput. 2015. V. 84, No. 294. P. 1953–1975.
39. Milio E. Computing isogenies between Jacobians of curves of genus 2 and 3 // Math. Comput. 2020. V. 89, No. 323. P. 1331–1364.
40. De Feo L., Dobson S., Galbraith S. D., Zobernig L. SIDH proof of knowledge. San Diego: Univ. California, 2021. (Cryptology ePrint Archive, Paper ID 2021/1023). Available at `eprint.iacr.org/2021/1023.pdf` (accessed Dec. 22, 2023).
41. De Feo L., Kieffer J., Smith B. Towards practical key exchange from ordinary isogeny graphs // Advances in cryptology — ASIACRYPT'18. Proc. Int. Conf. Theory and Application of Cryptology (Brisbane, Australia, Dec. 2–6, 2018). Cham: Springer, 2018. P. 365–394. (Lect. Notes Comput. Sci.; V. 11274).
42. Dartois P., de Feo L. On the security of OSIDH // Public-key cryptography — PKC 2022. Proc. 25th IACR Int. Conf. Practice and Theory of Public-Key Cryptography (Yokohama, Japan, Mar. 8–11, 2022). Pt. I. Cham: Springer, 2022. P. 52–81. (Lect. Notes Comput. Sci.; V. 13177).
43. Colò L., Kohel D. Orienting supersingular isogeny graphs // J. Math. Cryptology. 2020. V. 14, No. 1. P. 414–437.

Малыгина Екатерина Сергеевна
Куценко Александр Владимирович
Новосёлов Семён Александрович
Колесников Никита Сергеевич
Бахарев Александр Олегович
Хильчук Ирина Сергеевна
Шапоренко Александр Сергеевич
Токарева Наталья Николаевна

Статья поступила
11 мая 2023 г.
После доработки —
7 августа 2023 г.
Принята к публикации
22 сентября 2023 г.

DISKRETNYI ANALIZ I ISSLEDOVANIE OPERATSII
/DISCRETE ANALYSIS AND OPERATIONS RESEARCH/
January–March 2024. Vol. 31, No. 1. P. 52–84

UDC 519.7 DOI: 10.33048/daio.2024.31.772

POST-QUANTUM CRYPTOSYSTEMS: OPEN PROBLEMS AND CURRENT SOLUTIONS. ISOGENY-BASED AND CODE-BASED CRYPTOSYSTEMS

E. S. Malygina[1, 2, a], *A. V. Kutsenko*[2, b], *S. A. Novoselov*[1, c], *N. S. Kolesnikov*[1, d], *A. O. Bakharev*[2, e], *I. S. Khilchuk*[2, f], *A. S. Shaporenko*[2, g], and *N. N. Tokareva*[2, 1, h]

[1] Immanuel Kant Baltic Federal University,
14 Aleksandr Nevskii Street, 236041 Kaliningrad, Russia

[2] Novosibirsk State University,
2 Pirogov Street, 630090 Novosibirsk, Russia

E-mail: [a] emalygina@kantiana.ru, [b] alexandrkutsenko@bk.ru, [c] novsem@gmail.com, [d] nikolesnikov100@gmail.com, [e] a.bakharev@g.nsu.ru, [f] irina.khilchuk@gmail.com, [g] shaporenko.alexandr@gmail.com, [h] crypto1127@mail.ru

Abstract. This paper is a survey of modern post-quantum cryptographic schemes based on codes and isogenies. Special attention is paid to cryptanalysis of these schemes. In particular, for code-based cryptosystems we describe the information set decoding and the support splitting algorithm as main attacks, and for cryptosystems based on isogenies we describe in detail the Castryck — Decru attack on SIDH/SIKE. Tab. 2, bibliogr. 43.

Keywords: post-quantum cryptography, error-correcting code, elliptic curve, isogeny.

References

1. **E. S. Malygina, N. N. Tokareva, A. V. Kutsenko** [et al.]. Post-quantum cryptosystems: Open questions and solutions. Lattice-based cryptosystems, *Diskretn. Anal. Issled. Oper.* **30** (4), 46–90 (2023) [Russian] [*J. Appl. Ind. Math.* **17** (4), 767–790 (2023)].

English transl.: *Journal of Applied and Industrial Mathematics* **18** (1), 103–121 (2024), DOI: 10.1134/S1990478924010101.

2. **N. T. Courtois, M. Finiasz,** and **N. Sendrier,** How to achieve a McEliece-based digital signature scheme, in *Advances in Cryptology — ASIACRYPT'01,* Proc. Int. Conf. Theory and Application of Cryptology (Gold Coast, Australia, Dec. 9–13, 2001) (Springer, Heidelberg, 2001), pp. 157–174 (Lect. Notes Comput. Sci., Vol. 2248).
3. **J. Stern,** A new paradigm for public key identification, *IEEE Trans. Inf. Theory* **42** (6), 1757–1768 (1996).
4. **A. Childs, D. Jao,** and **V. Soukharev,** Constructing elliptic curve isogenies in quantum subexponential time, *J. Math. Cryptology* **8** (1), 1–29 (2014), DOI: 10.1515/jmc-2012-0016.
5. **R. J. McEliece,** A public-key cryptosystem based on algebraic coding theory, in *The Deep Space Network, Prog. Rep. 42-44* (California Inst. Tech., Pasadena, CA, 1978), pp. 114–116.
6. **H. Niederreiter,** Knapsack-type cryptosystems and algebraic coding theory, *J. Prob. Contr. Inform. Theory* **15** (2), 159–166 (1986).
7. **H. Niederreiter** and **C. Xing,** *Algebraic Geometry in Coding Theory and Cryptography* (Princeton Univ. Press, Princeton, NJ, 2009).
8. **L. Minder** and **A. Shokrollahi,** Cryptanalysis of the Sidelnikov cryptosystem, in *Advances in Cryptology — EUROCRYPT'07,* Proc. Int. Conf. Theory and Applications of Cryptographic Techniques (Barcelona, Spain, May 20–24, 2007) (Springer, Heidelberg, 2007), pp. 347–360 (Lect. Notes Comput. Sci., Vol. 4515), DOI: 10.1007/978-3-540-72540-4_20.
9. **E. Berlekamp, R. J. McEliece,** and **H. van Tilborg,** On the inherent intractability of certain coding problems, *IEEE Trans. Inf. Theory* **24** (3), 384–386 (1978).
10. **P. J. Lee** and **E. F. Brickell,** An observation on the security of McEliece's public-key cryptosystem, in *Advances in Cryptology — EUROCRYPT'88,* Proc. Workshop Theory and Application of Cryptographic Techniques (Davos, Switzerland, May 25–27, 1988) (Springer, Heidelberg, 1988), pp. 275–280 (Lect. Notes Comput. Sci., Vol. 330).
11. **E. Petrank** and **R. Roth,** Is code equivalence easy to decide?, *IEEE Trans. Inf. Theory* **43** (5), 1602–1604 (1997).
12. **N. Sendrier,** Finding the permutation between equivalent linear codes: The support splitting algorithm, *IEEE Trans. Inf. Theory* **46** (4), 1193–1203 (2000), DOI: 10.1109/18.850662.
13. **R. Misoczki, J.-P. Tillich, N. Sendrier,** and **P. Barreto,** MDPC-McEliece: New McEliece variants from moderate density parity-check codes, in *Proc. IEEE Int. Symp. Information Theory (Istanbul, Turkey, Jul. 7–12, 2013)* (IEEE Comput. Soc., Los Alamitos, CA, 2013), pp. 2069–2073, DOI: 10.1109/ISIT.2013.6620590.
14. **N. Drucker, S. Gueron,** and **D. Kostic,** QC-MDPC decoders with several shades of gray, in *Post-Quantum Cryptography,* Proc. Int. Conf. (Paris, France, Apr. 15–17, 2020) (Springer, Cham, 2020), pp. 35–50 (Lect. Notes Comput. Sci., Vol. 12100).

15. **R. C. Torres** and **N. Sendrier,** Analysis of information set decoding for a sublinear error weight, in *Post-Quantum Cryptography,* Proc. Int. Conf. (Fukuoka, Japan, Feb. 24–26, 2016) (Springer, Cham, 2016), pp. 144–161 (Lect. Notes Comput. Sci., Vol. 9606).
16. **E. Fujisaki** and **T. Okamoto,** Secure integration of asymmetric and symmetric encryption schemes, *J. Cryptology* **26**, 80–101 (2013).
17. **N. Bindel, M. Hamburg, K. Hövelmanns, A. Hülsing,** and **E. Persichetti,** Tighter proofs of CCA security in the quantum random oracle model, in *Theory of Cryptography,* Proc. Int. Conf. (Nuremberg, Germany, Dec. 1–5, 2019) (Springer, Cham, 2019), pp. 61–90 (Lect. Notes Comput. Sci., Vol. 11892), DOI: 10.1007/978-3-030-36033-7_3.
18. **C. Aguilar-Melchor, O. Blazy, J.-C. Deneuville, P. Gaborit,** and **G. Zémor,** Efficient encryption from random quasi-cyclic codes, *IEEE Trans. Inf. Theory* **64** (5), 3927–3943 (2018).
19. **N. Aragon, P. Gaborit,** and **G. Zémor,** HQC-RMRS, an instantiation of the HQC encryption framework with a more efficient auxiliary error-correcting code (Cornell Univ., Ithaca, NY, 2005) (Cornell Univ. Libr. e-Print Archive; arXiv:2005.10741)
20. **C. Doche** and **T. Lange,** Arithmetic of elliptic curves, in *Handbook of Elliptic and Hyperelliptic Curve Cryptography* (Chapman & Hall/CRC Press, Boca Raton, FL, 2006), pp. 267–302.
21. **A. A. Bolotov, S. B. Gashkov, A. B. Frolov,** and **A. A. Chasovskikh,** *An Elementary Introduction to Ellipic Cryptography: Algebraic and Algorithmic Basics* (KomKniga, Moscow, 2006) [Russian].
22. **A. Sutherland,** Ellipic Curves. Isogenies, *Lecture Notes* (MIT, Cambridge, MA, 2022). Available at `math.mit.edu/classes/18.783/2022/LectureNotes4.pdf` (accessed Dec. 22, 2023).
23. **J. Vélu,** Isogénies entre courbes elliptiques, *C. R. Acad. Sci., Paris, Sér. A,* **273**, 238–241 (1971).
24. **S. Duquesne** and **T. Lange,** Arithmetic of hyperelliptic curves, in *Handbook of Elliptic and Hyperelliptic Curve Cryptography* (Chapman & Hall/CRC Press, Boca Raton, FL, 2006), pp. 303–353.
25. **E. Kani,** The number of curves of genus two with elliptic differentials, *J. Reine Angew. Math.* **485**, 93-122 (1997).
26. **E. V. Flynn** and **Y. B. Ti,** Genus two isogeny cryptography, in *Post-Quantum Cryptography,* Proc. Int. Conf. (Chongquin, China, May 10–12, 2019) (Springer, Cham, 2019), pp. 286–306 (Lect. Notes Comput. Sci., Vol. 11505).
27. **W. Castryck, T. Decru,** and **B. Smith,** Hash functions from superspecial genus-2 curves using Richelot isogenies, *J. Math. Cryptology* **14** (1), 268–292 (2020).
28. **N. Alamati, L. de Feo, H. Montgomery,** and **S. Patranabis,** Cryptographic group actions and applications (Univ. California, San Diego, 2020) (Cryptology ePrint Archive, ID 2020/1188). Available at `eprint.iacr.org/2020/1188.pdf` (accessed Dec. 22, 2023).

29. L. De Feo, D. Jao, PlĴ. ut, Towards quantum-resistant cryptosystems from supersingular elliptic curve isogenies (Univ. California, San Diego, 2011) (Cryptology ePrint Archive, ID 2011/506). Available at `eprint.iacr.org/2011/506.pdf` (accessed Dec. 22, 2023).

30. W. Castryck and **T. Decru,** An efficient key recovery attack on SIDH (Univ. California, San Diego, 2022) (Cryptology ePrint Archive, ID 2022/975). Available at `eprint.iacr.org/2022/975.pdf` (accessed Dec. 22, 2023).

31. W. Castryck, T. Lange, C. Martindale, L. Panny, and **J. Renes,** CSIDH: An efficient post-quantum commutative group action (Univ. California, San Diego, 2018) (Cryptology ePrint Archive, ID 2018/383). Available at `eprint.iacr.org/2018/383.pdf` (accessed Dec. 22, 2023).

32. J.-J. Chi-Domínguez and **F. Rodríguez-Henríquez,** Optimal strategies for CSIDH, *J. Adv. Math. Commun.* **16** (2), 383–411 (2022).

33. H. Cohen, *A Course in Computational Algebraic Number Theory* (Springer, Berlin, 1993).

34. L. Maino and **C. Martindale,** An attack on SIDH with arbitrary starting curve (Univ. California, San Diego, 2022) (Cryptology ePrint Archive, ID 2022/1026). Available at `eprint.iacr.org/2022/1026.pdf` (accessed Dec. 22, 2023).

35. D. Robert, Breaking SIDH in polynomial time (Univ. California, San Diego, 2022) (Cryptology ePrint Archive, ID 2022/1038). Available at `eprint.iacr.org/2022/1038.pdf` (accessed Dec. 22, 2023).

36. E. W. Howe, F. Leprevost, and **B. Poonen,** Large torsion subgroups of split Jacobians of curves of genus two or three, *J. Forum Math.* **12** (3), 315–364 (2000).

37. N. Bruin, E. V. Flynn, and **D. Testa,** Descent via $(3,3)$-isogeny on Jacobians of genus 2 curves, *J. Acta Arithmetica* **165** (3), 201–223 (2014).

38. R. Cosset and **D. Robert,** Computing (ℓ,ℓ)-isogenies in polynomial time on Jacobians of genus 2 curves, *Math. Comput.* **84** (294), 1953–1975 (2015).

39. E. Milio, Computing isogenies between Jacobians of curves of genus 2 and 3, *Math. Comput.* **89** (323), 1331–1364 (2020).

40. L. De Feo, S. Dobson, S. D. Galbraith, and **L. Zobernig,** SIDH proof of knowledge (Univ. California, San Diego, 2021) (Cryptology ePrint Archive, ID 2021/1023). Available at `eprint.iacr.org/2021/1023.pdf` (accessed Dec. 22, 2023).

41. L. De Feo, J. Kieffer, and **B. Smith,** Towards practical key exchange from ordinary isogeny graphs, in *Advances in Cryptology — ASIACRYPT'18,* Proc. Int. Conf. Theory and Application of Cryptology (Brisbane, Australia, Dec. 2–6, 2018) (Springer, Cham, 2018), pp. 365–394 (Lect. Notes Comput. Sci., Vol. 11274).

42. P. Dartois and **L. de Feo,** On the security of OSIDH, in *Public-Key Cryptography — PKC 2022,* Proc. 25th IACR Int. Conf. Practice and Theory of Public-Key Cryptography (Yokohama, Japan, Mar. 8–11, 2022), Pt. I (Springer, Cham, 2022), pp. 52–81 (Lect. Notes Comput. Sci., Vol. 13177).

43. L. Colò and **D. Kohel,** Orienting supersingular isogeny graphs, *J. Math. Cryptology* **14** (1), 414–437 (2020).

Ekaterina S. Malygina
Aleksandr V. Kutsenko
Semyon A. Novoselov
Nikita S. Kolesnikov
Aleksandr O. Bakharev
Irina S. Khilchuk
Aleksandr S. Shaporenko
Natalia N. Tokareva

Received May 11, 2023
Revised August 7, 2023
Accepted September 22, 2023

ДИСКРЕТНЫЙ АНАЛИЗ И ИССЛЕДОВАНИЕ ОПЕРАЦИЙ
Январь–март 2024. Т. 31, № 1. С. 85–108

УДК 519.176 DOI: 10.33048/daio.2024.31.780

МЕТОД АВТОМАТИЧЕСКОГО ПОИСКА СЕМЕЙСТВ ОПТИМАЛЬНЫХ ХОРДАЛЬНЫХ КОЛЬЦЕВЫХ СЕТЕЙ

Э. А. Монахова [a], *О. Г. Монахов* [b]

Институт вычислительной математики и математической геофизики,
пр. Акад. Лаврентьева, 6, 630090 Новосибирск, Россия

E-mail: [a] emilia@rav.sscc.ru, [b] monakhov@rav.sscc.ru

Аннотация. Арден и Ли предложили класс хордальных кольцевых сетей степени три в качестве сетей связи мультикомпьютерных систем, получили формулу для вычисления диаметра и алгоритм поиска кратчайших путей для них. В настоящей работе показано, что найденные ими формула диаметра и алгоритм поиска кратчайших путей являются неточными. Нами построен большой массив экспериментальных данных (датасет), содержащий параметры описаний оптимальных по диаметру хордальных колец с числом узлов до 60 тысяч, и найдена точная нижняя граница диаметра хордальных колец. На основе анализа полученного массива данных предложен новый метод и реализованы алгоритмы автоматического поиска аналитических описаний семейств оптимальных хордальных колец, с помощью которых найдены аналитические описания более 500 новых семейств оптимальных хордальных кольцевых сетей для многих значений числа узлов (до этого в литературе были известны шесть семейств). Табл. 5, ил. 6, библиогр. 26.

Ключевые слова: оптимальная хордальная сеть степени три, диаметр, экстремальный хордальный граф, датасет оптимальных хордальных сетей.

Введение

В работе [1] Арден и Ли ввели класс хордальных кольцевых графов (chordal ring networks) степени три в качестве возможной топологии сетей связи мультикомпьютерных систем и исследовали новый класс по трём направлениям: определение диаметра графов, поиск кратчайших путей для них и определение максимального числа вершин при заданном диаметре. Эта работа вызвала большой интерес (более 450 ссылок к настоящему времени) по исследованию свойств новых структур

сетей связи и построению расширений предложенной топологии (см. например, [2–15]). Исследования проводились по следующим направлениям: 1) расширение области значений образующих графа; 2) построение алгоритмов маршрутизации для них; 3) увеличение степени вершин графа и степени его симметрии; 4) построение иерархических структур на их основе; 5) рассмотрение ориентированных хордальных кольцевых графов, в том числе с исследованием их надёжности при отказах элементов, и др.

Неориентированный *хордальный* граф $C_N(1, -1, s)$, где N — чётное, а $3 \leqslant s \leqslant N/2$ — нечётное числа, имеет множество вершин $V = \mathbb{Z}_N = \{0, 1, \ldots, N-1\}$, в котором вершины i и $(i+1) \bmod N$ смежны для любого i, а каждая нечётная вершина i смежна также с $(i+s) \bmod N$. Параметры описания хордального графа: s — образующая (длина хорды), N — его порядок. Степень вершин равна трём. *Диаметр* графа — длина максимального кратчайшего пути на множестве всевозможных пар вершин. При фиксированном значении N назовём *оптимальным* хордальный граф $C_N(1, -1, s)$ с минимально возможным диаметром для данного N. Недавно проведённые исследования структурной надёжности неориентированных хордальных колец показали, что минимизация диаметра при одном и том же порядке графа путём выбора соответствующей образующей даёт большую надёжность структуры связей графа при отказах элементов.

На рис. 1а изображён хордальный граф $C_{20}(1, -1, 5)$ диаметра $d = 4$. Граф Хивуда $C_{14}(1, -1, 5)$ также входит в рассматриваемый класс. Следует отметить, что хордальные кольца являются двудольными графами и входят в класс PRS-графов [16], как графы с двумя типами вершин, в отличие от другого широко изучаемого класса циркулянтных графов

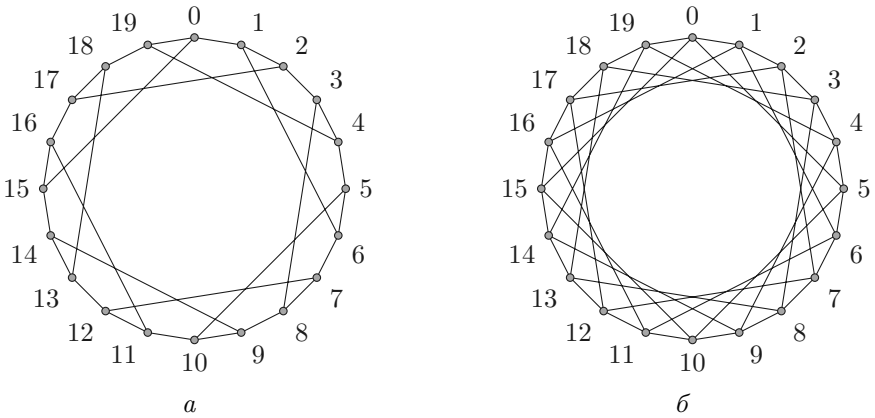

Рис. 1. а) Хордальный граф $C_{20}(1, -1, 5)$; б) циркулянтный граф $C(20; 1, 5)$

степени четыре [4–6], входящих также в класс PRS-графов, но вершинно-симметричных и имеющих один тип вершин. Можно сказать, что хордальные графы вида $C_N(1,-1,s)$ являются подмножеством класса циркулянтов вида $C(N;1,s)$ с чётным числом вершин и упорядоченным удалением рёбер s или $-s$, но в отличие от циркулянтов имеют обхват 6, а не 4. На рис. 1b для сравнения изображён циркулянт $C(20;1,5)$. В настоящей работе исследуется решение двух оптимизационных проблем поиска оптимальных хордальных графов: 1) графов с минимальным диаметром при заданном числе вершин (порядке) графа и 2) графов с максимальным порядком при заданном диаметре (так называемых экстремальных графов). В общей постановке — это поиск аналитических описаний бесконечных семейств (серий) графов с минимальным диаметром при заданных порядках графов.

1. Известные результаты по оптимизации хордальных кольцевых графов

Первыми решение второй проблемы предложили Арден и Ли [1, теорема 3]. В силу того, что они рассмотрели не все возможные типы путей в хордальных графах [5], найденные ими формулы для максимальных порядков при заданных диаметрах оказались неточными. Авторы [2, 3] смогли улучшить их результат для максимального N и получили следующие экстремальные графы. Представляем их результат в виде теоремы, доказательство которой можно найти в [2].

Теорема 1. *Пусть $d \geqslant 3$ — целое число. Максимальный порядок N_d хордального графа $C_N(1,-1,s)$ диаметра d равен*

$$N = N_d = \begin{cases} (3d^2+1)/2 & \text{при нечётном } d, \\ 3d^2/2 - d & \text{при чётном } d, \end{cases} \tag{1}$$

образующая графа равна

$$s = \begin{cases} 3d & \text{при нечётном } d, \\ 3d+1 & \text{при чётном } d. \end{cases} \tag{2}$$

Отметим, что это оптимальное решение найдено авторами [2, 3] методом плотной укладки (tessellation) на плоскости введённого ими обобщённого класса графов Ардена и Ли, а именно в их обозначении графов $CR_N(a,b,c)$ степени 3, где порядок N — чётное, а образующие a,b,c — нечётные числа.

Проблема получения минимально возможного диаметра для заданного значения N является более сложной задачей, чем получение максимального N, поскольку диаметр не всегда монотонно увеличивается

Таблица 1

Семейства оптимальных хордальных колец [2]

Семейство	$2N$	s	$d \bmod 2$	$N_d - N$
f_0	$3d^2 - 2d$	$3d + 1$	0	0
f_0	$3d^2 + 1$	$3d$	1	0
f_1	$3d^2 - 4d$	$3d - 1$	0	d
f_2	$3d^2 - 2d - 1$	$3d - 2$	1	$d + 1$
f_3	$3d^2 - 4d - 3$	$3d - 4$	1	$2(d + 1)$
f_4	$3d^2 - 4d + 1$	$3d - 4$	1	$2d$
f_5	$3d^2 - 6d - 1$	$3d - 6$	1	$3d + 1$

с ростом порядка. Кроме получения семейства с максимальным N, авторы [2] нашли ещё пять семейств графов с минимально возможным диаметром, используя плотную укладку графов на плоскости. В табл. 1 приведены все ранее известные семейства оптимальных графов (семейство экстремальных графов обозначено как f_0) с описаниями параметров в виде функций (полиномов) от диаметра d. В последнем столбце показана разница между значением N_d и порядками графов соответствующих семейств. В настоящей работе с помощью анализа построенного датасета оптимальных хордальных кольцевых графов были найдены новые закономерности в появлении бесконечных семейств таких графов, предложен и реализован метод автоматического поиска и открытия новых семейств и на его основе расширен список семейств оптимальных хордальных колец: получено дополнительное описание образующих для бесконечного экстремального семейства из теоремы 1 и построено более 500 новых аналитически задаваемых семейств оптимальных графов.

2. Формула для диаметра хордальных кольцевых графов

Проведённый нами анализ формулы для вычисления диаметра хордальных графов из [1] показал, что она верна для ряда значений N и s, но для большинства графов наблюдается рост (при росте числа вершин) разницы между реальным диаметром и значением диаметра, вычисляемым по формуле из [1].

Из теоремы 1, как следствие, получаем вид зависимости диаметра экстремальных графов от их порядка.

Лемма 1. *Пусть для любого целого $d \geqslant 3$ порядок и образующая хордального графа $C_N(1, -1, s)$ определяются формулами* (1) *и* (2) *соответственно. Тогда диаметр экстремального графа равен $\lfloor \sqrt{2N/3} \rfloor$ при нечётном d и $\lceil \sqrt{2N/3} \rceil$ при чётном d.*

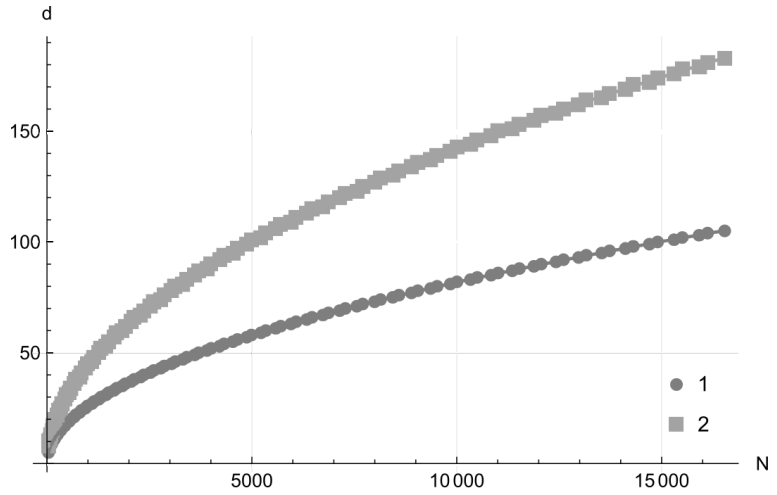

Рис. 2. Диаметр экстремальных графов:
1 — реальный, 2 — вычисленный по формуле из [1]

На рис. 2 разница между реальным диаметром хордальных графов и диаметром, получаемым по формуле из [1], наглядно показана на примере семейства экстремальных графов. Данная ошибка авторов [1] произошла в силу того, что они рассмотрели не все возможные виды путей в хордальных графах, а ограничились тремя видами вместо шести [5]. Таким образом, формула вычисления диаметра хордальных колец, полученная в [1], в общем случае даёт только его грубую верхнюю оценку. Следует отметить, что в 2009 г. в работе [17] автор частично скорректировал формулу из [1], но нахождение точной формулы для диаметра хордальных кольцевых графов при любых N и s по-прежнему остаётся открытой проблемой.

Имеет место аналогичная история поиска общей формулы для вычисления диаметра циркулянтных графов степени четыре вида $C(N; 1, s)$: в течение 30 лет появляется ряд работ, которые либо не полностью учитывают все возможные условия для соотношений между N и s, либо дают неточные результаты. Последний по времени пример неверного в ряде случаев определения диаметра — работа [18].

В связи с выявлением неточности вычисления диаметра в [1] нами была проверена также справедливость теоремы 2 из [1], где дана следующая оценка диаметра хордальных колец с числом вершин $N = m^2$ при чётном $m \geqslant 8$: диаметр графа $d(C_{m^2}(1, -1, s)) \geqslant m - 1$. Она оказалась верной только для значений $m = 8, 10, 12, 14$, при $16 \leqslant m \leqslant 244$ $d(C_{m^2}(1, -1, s)) < m - 1$.

3. Построение датасета оптимальных хордальных графов

Используя при поиске оптимальных хордальных кольцевых графов сокращённый вариант переборного алгоритма и его параллельные реализации на кластере Kunpeng 920 (подробное описание алгоритмов и их оценки можно найти в [19]), мы впервые построили большой массив данных (датасет) оптимальных хордальных графов. Для заданного значения N перебирались все образующие $3 \leqslant s \leqslant N/2$ и определялся граф (графы) с минимальным диаметром. Полная версия полученного датасета для $N \leqslant 60\,000$ доступна по ссылке `github.com/mila0411/ChordalRing`. Даны значения порядка N, всех образующих s, задающих оптимальный граф и минимальный диаметр d графа. На рис. 3 показаны графики зависимостей образующих s от N для оптимальных хордальных графов. Нижний ряд точек соответствует образующим с линейной зависимостью от диаметра, верхний ряд — максимальные по величине образующие квадратичного вида от диаметра. Прослеживаются некоторые упорядоченности в расположении образующих при росте N, которые указывают на возможность существования многих аналитически задаваемых семейств графов. Анализ осложняется тем, что у разных порядков графов число оптимальных образующих меняется от одного до заранее не определённого количества (меньшего чем $N/2 - 3$). Анализ датасета

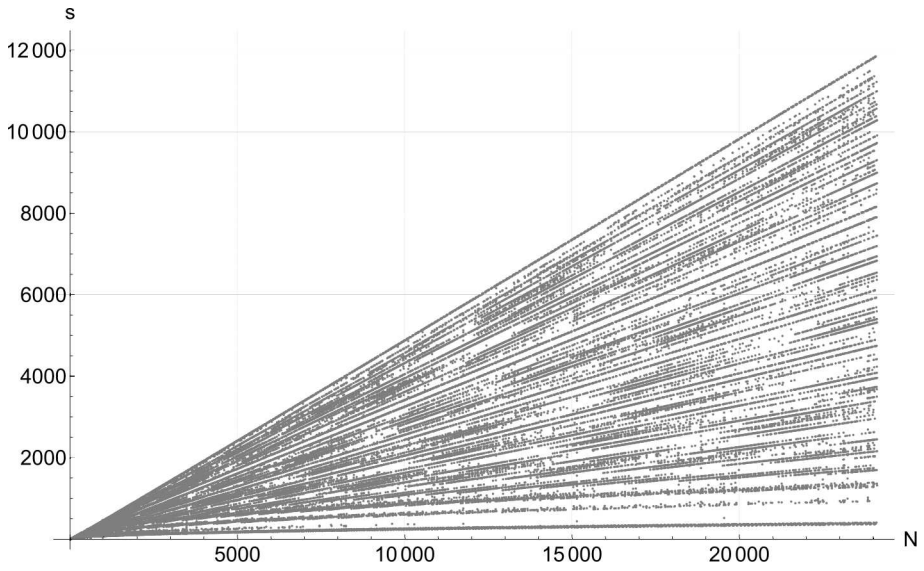

Рис. 3. График зависимости образующих s от порядка N для оптимальных хордальных графов

показал, что есть много оптимальных графов с единственной образующей, которая не является линейным полиномом от d.

Авторы хотели бы отметить, что при поиске аналитических зависимостей параметров на множестве графического отображения точек $\{N, s, d\}$, представленного на рис. 3 как новый геометрический объект, прослеживаются некоторые аналогии с постановкой задач для таких математических объектов как спираль Улама [20], рекуррентные графики динамических систем [21] и других подобных множеств.

4. Получение новых образующих для экстремального семейства графов

При анализе датасета для семейства с максимальным $N = N_d$ было найдено описание образующих квадратичного вида в отличие от линейных образующих (2) при диаметрах d, кратных четырём.

Теорема 2. *Пусть $d \geqslant 4$ — целое число и $d \equiv 0 \pmod 4$. Существует оптимальный хордальный граф $C_N(1, -1, s)$ с максимальным порядком и диаметром d, где*

$$N = N_d = 3d^2/2 - d,$$
$$s = 3d^2/4 - 2d + 1. \tag{3}$$

Доказательство. Покажем, что при диаметре $d \geqslant 4$, кратном четырём, существует изоморфизм между графами, задаваемыми (1) и (2), и графами, задаваемыми (1) и (3). Обозначим образующую (2), равную $3d + 1$, через s_1, а образующую (3) — через s_2. Тогда имеем следующее соотношение между N и образующими s_1 и s_2: $N = 2s_2 + s_1 - 3$.

Для доказательства будем использовать лемму 10 из [22], которая утверждает, что хордальные кольцевые графы $C_N(a, b, c)$ и $C_N(a', b', c')$, где $a, b, c, a', b', c' \in \overline{1, N-1}$, изоморфны, если и только если существует взаимно простое с $N/2$ число l, для которого выполняются следующие сравнения:

$$a' - b' \equiv l(a - b) \pmod N, \tag{4}$$
$$b' - c' \equiv l(b - c) \pmod N. \tag{5}$$

Положим $a = 1$, $b = N - 1$, $c = s_1$, $a' = s_2$, $b' = N - 1$, $c' = 1$. Тогда $a' - b' = -N + s_2 + 1$, отсюда $(a' - b') \bmod N = s_2 + 1$. Аналогично $(b' - c') \bmod N = N - 2$.

Возьмём $l = (s_2 + 1)/2$. При $d = 4t$, $t \geqslant 1$, имеем НОД$(l, N/2) = 1$. Подставив значение l в (4), имеем $l(a - b) = -(s_2 + 1)/2 \cdot N + s_2 + 1$ и $l(a - b) \bmod N = s_2 + 1$.

Подставив значение l в (5) и проведя соответствующие замены, имеем $l(b-c) = (3d^2/8 - 7d/4 + 1)N + N - 2$, отсюда $l(b-c) \bmod N = N - 2$. Теорема 2 доказана.

Отметим, что изображённый на рис. 1а хордальный граф оптимальный и принадлежит семейству из теоремы 2.

5. Новые семейства оптимальных хордальных колец с образующими $s = s(d)$

При анализе датасета дополнительно был получен ряд новых семейств оптимальных хордальных графов с образующими, линейно зависящими от диаметра: $s = s(d)$. Новые семейства представлены в табл. 2 (здесь и далее $d_{\min}$ — значение минимального диаметра, начиная с которого данное семейство существует).

Таблица 2

Новые семейства оптимальных хордальных колец с образующей $s = s(d)$

Семейство	$2N$	s	$d \bmod 2$	$d_{\min}$	$N_d - N$
k_1	$3d^2 - 8d$	$3d+1$	0	10	$3d$
k_2	$3d^2 - 8d - 4$	$3d+1$	0	12	$3d+2$
k_3	$3d^2 - 14d$	$3d+1$	0	30	$6d$
k_4	$3d^2 - 14d - 8$	$3d+1$	0	34	$6d+4$
k_5	$3d^2 - 6d + 3$	$3d$	1	7	$3d-1$
k_6	$3d^2 - 6d - 1$	$3d$	1	7	$3d+1$
k_7	$3d^2 - 12d + 5$	$3d$	1	15	$6d-2$
k_8	$3d^2 - 12d - 3$	$3d$	1	19	$6d+2$
k_9	$3d^2 - 10d$	$3d-1$	0	14	$4d$
k_{10}	$3d^2 - 16d$	$3d-1$	0	38	$7d$

Докажем для примера существование семейства k_5.

Теорема 3. *Пусть $d \geqslant 7$ — нечётное число. Существует оптимальный хордальный граф $C_N(1, -1, s)$ с диаметром d, где*

$$N = 3d^2/2 - 3d + 3/2, \quad s = 3d.$$

Доказательство. Для доказательства применим метод построения плотной укладки графов семейства на плоскости Z^2, предложенный авторами [2] для доказательства существования при любых диаметрах семейств из табл. 1. Укладка графов семейства k_5 на плоскости Z^2 может

быть получена из плотной укладки (см. детали в [2]), соответствующей экстремальному графу нечётного диаметра $d \geqslant 7$ с порядком N_d, путём удаления требуемого числа вершин, в данном случае $3d-1$ вершин, с внешних слоёв вершин (треугольников). На рис. 4 показан пример плотной укладки графа $C_{54}(1,-1,21)$ диаметра $d=7$. Число окрашенных треугольников равно разности между порядком экстремального графа того же диаметра и порядком графа семейства k_5.

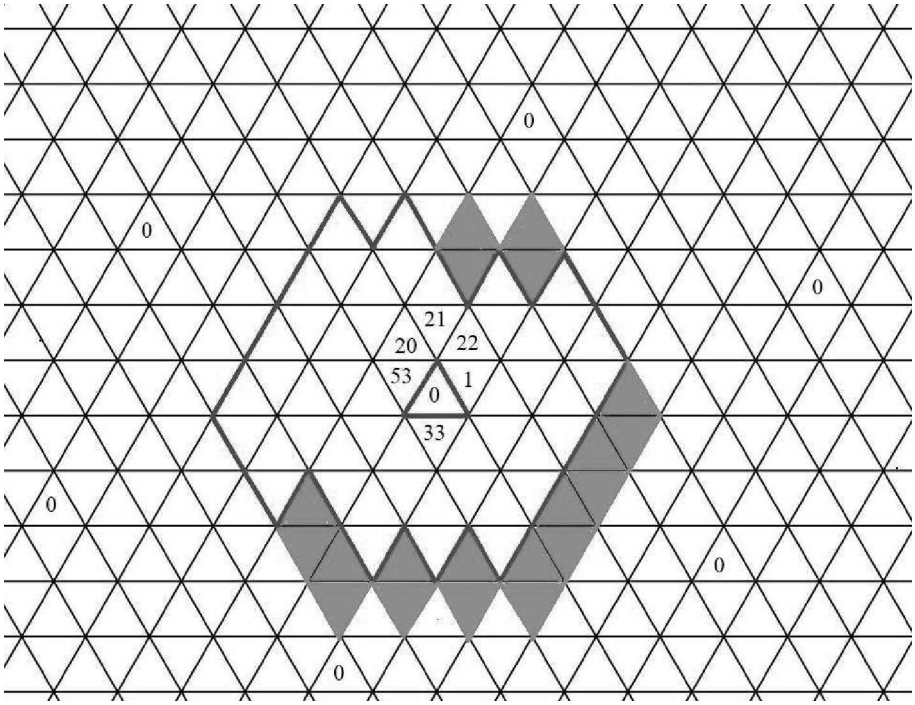

Рис. 4. Плотная укладка графа $C_{54}(1,-1,21)$ на плоскости Z^2

Оставшийся граф по построению имеет тот же минимальный диаметр и образует плотную укладку на плоскости, определяемую кратчайшими путями между центральным нулём и соседними нулями на плоскости. Для достижения левого верхнего нуля требуется пройти $(d-1)/2$ шагов по образующей s, 0 шагов по образующей $s=1$ и $3(d-1)/2$ шагов по образующей $s=-1$. Чтобы достичь правый верхний нуль, требуется пройти $d-2$ шагов по s, $(d+1)/2$ шагов по $s=1$ и $(d-5)/2$ шагов по $s=-1$. Для достижения следующего по часовой стрелке нуля требуется пройти $(d-3)/2$ шагов по s, 0 шагов по $s=-1$ и $3(d+1)/2$ шагов по $s=1$. Остальные три соседних нуля расположены симметрично по отношению

к указанным. Таким образом, получаем следующие три сравнения для распределения соседних нулей на плоскости:

$$(d-1)/2 \cdot s - 3(d-1)/2 \equiv 0 \pmod{N},$$
$$(d-2)s + 3 \equiv 0 \pmod{N},$$
$$(d-3)/2 \cdot s + 3(d+1)/2 \equiv 0 \pmod{N}.$$

Подстановка $s = 3d$ даёт решение всех трёх сравнений. Теорема 3 доказана.

При анализе датасета оптимальных хордальных графов были замечены следующие интересные особенности: существуют тетрады (quadruple) семейств графов, идущие по четыре семейства подряд с одинаковой образующей при одинаковом диаметре. Существуют также такие же триплеты (triplet) семейств, идущие по три семейства подряд, и дуплеты (double) семейств, идущие по два семейства подряд. Все перечисленные семейства графов показаны в табл. 3. Отметим, что шесть семейств из табл. 3 являются новыми по сравнению с [2].

Таблица 3

Семейства оптимальных графов

Семейство	$N = \bigcup N_i$	s	$d \bmod 2$	$d_{\min}$
Тетрада	$N_1 = (3d^2+1)/2 - 3d + 3$, $N_2 = N_1 - 2$, $N_3 = N_1 - 4$, $N_4 = N_1 - 6$	$3d-6$	1	7
Триплет	$N_1 = (3d^2+1)/2 - 2d - 2$, $N_2 = N_1 + 2$, $N_3 = N_1 + 4$	$3d-4$	1	7
Дуплет	$N_1 = (3d^2+1)/2 - d - 1$, $N_2 = N_1 + 2$	$3d-2$	1	5
Дуплет	$N_1 = 3d^2/2 - 2d + 2$, $N_2 = N_1 - 2$	$3d-1$	0	6

Существование всех семейств графов из табл. 2 и 3 было проверено и подтверждено экспериментально для всех диаметров $d = d_{\min} \div 154$ и порядков графов $N \leqslant 35000$. Полученные экспериментальные результаты говорят в пользу того, что найденные семейства графов существуют при любых диаметрах, больших $d_{\min}$, т. е. являются бесконечными семействами. Теоретическое подтверждение этого свойства (или его опровержения) составляет предмет будущей работы. Также дополнительного исследования требует задача определения условий возникновения новых семейств оптимальных графов и определения нижних границ для $d_{\min}$.

6. Точная нижняя граница диаметра хордальных колец

До сих пор мы рассматривали оптимальные графы как графы с минимально возможным диаметром при заданном порядке графа. Возникает вопрос: совпадает ли их диаметр с точной нижней границей?

Для этого надо знать точную нижнюю границу диаметра хордальных графов для любого N. Определяем её, исходя из значений диаметров, полученных для графов с максимальным числом вершин при заданном диаметре.

Теорема 4. *Пусть $d \geqslant 2$ — целое число. Точная нижняя граница для диаметра хордального кольцевого графа порядка N равна $lb(N) = d + 1$ при нечётном d и $(3d^2+1)/2+2 \leqslant N \leqslant (3d^2+1)/2+2d$ либо при чётном d и $3d^2/2 - d + 2 \leqslant N \leqslant 3d^2/2 + 3d + 2$.*

Доказательство. Точная нижняя граница диаметра хордальных кольцевых графов $lb(N)$ представляет собой кусочно постоянную функцию величины $d + 1$ на интервалах $N_d + 2 \leqslant N \leqslant N_{d+1}$, где $d \geqslant 2$ — диаметр графа. При переходе к очередному интервалу величина функции $lb(N)$ увеличивается на единицу. Подставляя значения функции N_d из теоремы 1, получаем искомый результат. Теорема 4 доказана.

На рис. 5 показан вид полученной кривой для точной нижней границы диаметра хордальных кольцевых графов при $N \leqslant 24000$. Отметим, что

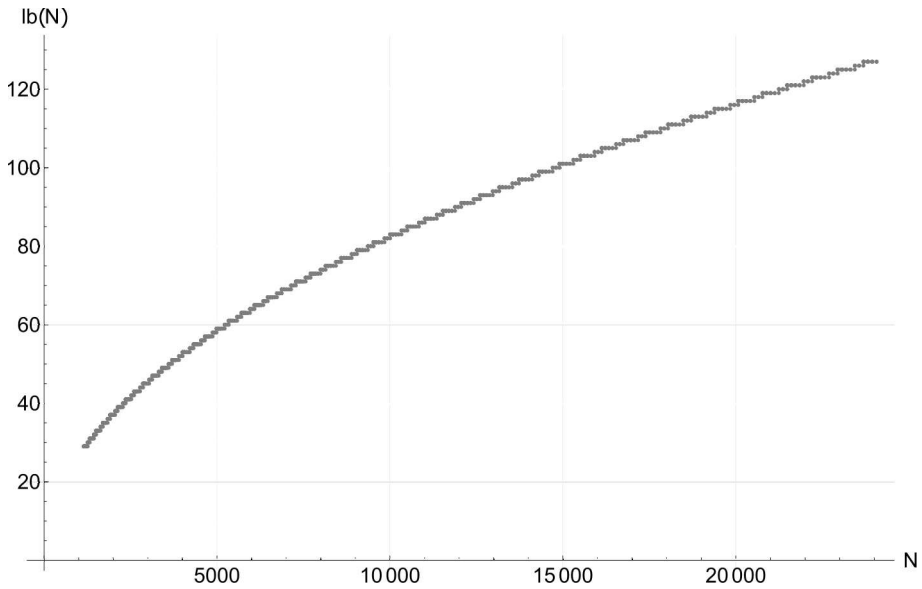

Рис. 5. Точная нижняя граница диаметра хордального кольцевого графа

все графы из табл. 3 имеют диаметры, равные $lb(N)$. В табл. 2 только графы семейств k_5 и k_6 имеют диаметры, совпадающие с точной нижней границей. Графы семейства k_1 имеют диаметр $d = lb(N) + 1$.

7. Семейства оптимальных графов, достигающих точной нижней границы диаметра

Для дальнейшего исследования из общего массива экспериментальных данных оптимальных графов выделим параметры $\{N, s, d\}$ только тех графов, диаметры которых совпадают с $lb(N)$. Назовём их предельно оптимальными или п-оптимальными. На рис. 6 в координатах N и s показан фрагмент датасета для графов с порядками $1200 \leqslant N \leqslant 24000$, относящийся к п-оптимальным графам.

Сначала рассмотрим множество образующих квадратичного вида, относящихся к п-оптимальным графам. Нашей целью является поиск семейств п-оптимальных графов с общими аналитическими описаниями параметров N и s, представленными в виде полиномов от диаметра d.

Отметим следующие особенности в расположении порядков и образующих квадратичного вида для п-оптимальных графов.

1. Члены семейств располагаются вдоль линий общего угла наклона, определяемого значением $\lfloor N/s \rfloor$.

2. Повторяемость членов семейств при росте диаметра d реализуется через чётное число $p \geqslant 4$ значений, зависящее от $\lfloor N/s \rfloor$. Назовём p

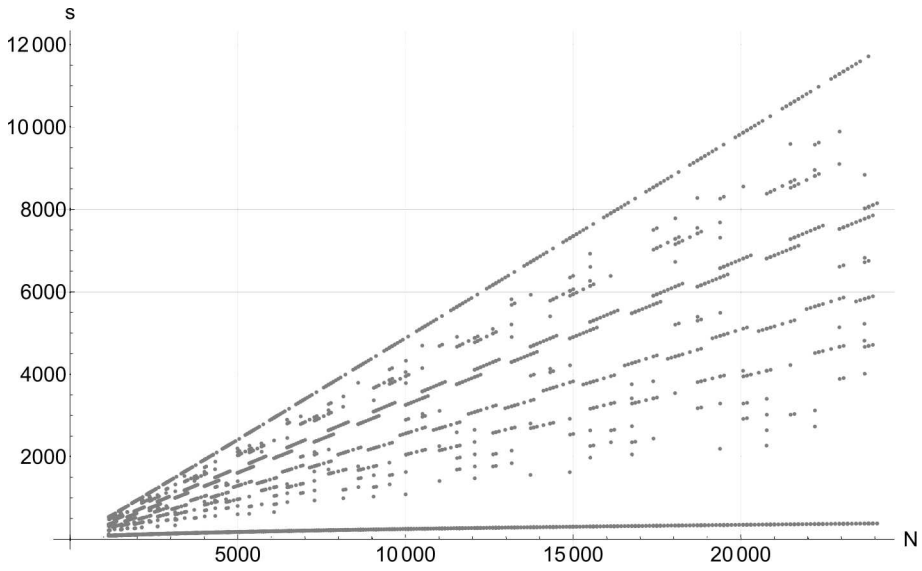

Рис. 6. П-оптимальные хордальные графы

периодом повторяемости аналитического описания. Начальное значение диаметра, при котором описание семейства приобретает повторяемость, может быть достаточно большим.

3. Семейства на каждой линии дублируются с некоторым сдвигом по диаметру на ближайшей линии снизу следующим образом: на верхней линии — $N_1 = \lfloor N_1/s_1 \rfloor s_1 - N_1/s_1 \pmod{s_1}$, на нижней линии — $N_2 = \lfloor N_2/s_2 \rfloor s_2 + N_2/s_2 \pmod{s_2}$, где $N_2/s_2 \pmod{s_2} = N_1/s_1 \pmod{s_1} + 3p$. Исключение составляет самая верхняя линия графика, соответствующая $p = 4$, поскольку по определению хордальных графов значения образующих s ограничены величиной $N/2$.

4. Так как нас интересуют семейства графов с общим аналитическим описанием, которые имеют устойчивую повторяемость на больших диапазонах изменения диаметра, целесообразно рассматривать для поиска семейств достаточно большие начальные значения диаметров.

8. Алгоритм поиска семейств оптимальных графов с квадратичными образующими

С учётом найденных особенностей расположения семейств n-оптимальных графов для автоматизации их поиска был разработан и реализован эвристический алгоритм. При записи алгоритма использованы следующие обозначения (здесь q_i — целая часть, r_i — остаток):

$$q_i = \lfloor N_i/s_i \rfloor, \quad r_i = N_i \bmod s_i, \quad i = 1, 2,$$
$$q'_i = \lfloor r_i/d_i \rfloor, \quad r'_i = r_i \bmod d_i, \quad i = 1, 2,$$
$$q''_i = \lfloor s_i/d_i \rfloor, \quad r''_i = s_i \bmod d_i, \quad i = 1, 2.$$

Алгоритм 1 (автоматический поиск семейств оптимальных графов)

Шаг 1. Выбирается начальное значение периода $p \in \{4, 6, \dots\}$ и диаметра $d = d_1 \equiv 0 \pmod{p}$. Перебираются все точки в массиве данных $\{N_1, s_1, d_1\}$ и среди всех точек вида $\{N_2, s_2, d_2\}$ выбираются точки, где $d_2 = d_1 + p$ и $\lfloor N_2/s_2 \rfloor = \lfloor N_1/s_1 \rfloor$. Для точки $\{N_1, s_1, d_1\}$ и удовлетворяющей условиям очередной точки $\{N_2, s_2, d_2\}$ создаётся общая формула возможного семейства графов с s и N в виде полиномов от диаметра:

$$N = ad^2 + bd + c, \quad s = ed^2 + fd + g.$$

Для этого используется свойство повторяемости вида членов семейства через $k = (d - d_1)/p$ при создании образующих и порядков с помощью формул, сохраняющих пропорциональность наращивания коэффициентов при степенях d:

$$s = (q''_1 + k(q''_2 - q''_1))d + r''_1 + k(r''_2 - r''_1), \tag{6}$$
$$N = q_1 s + (q'_1 + k(q'_2 - q'_1))d + r'_1 + k(r'_2 - r'_1). \tag{7}$$

Подстановками $k = (d - d_1)/p$ в (6), значений k и s в (7) и приведением подобных слагаемых получаются коэффициенты для полиномов N и s:

$$a = (q_2' - q_1')/p + q_1(q_2'' - q_1'')/p,$$
$$b = q_1' + q_1 q_1'' + (r_2' - r_1')/p + q_1(r_2'' - r_1'')/p - d_1 a,$$
$$c = r_1' + q_1 r_1'' - d_1(r_2' - r_1')/p - d_1 q_1(r_2'' - r_1'')/p,$$
$$e = (q_2'' - q_1'')/p,$$
$$f = q_1'' - d_1(q_2'' - q_1'')/p + (r_2'' - r_1'')/p,$$
$$g = r_1'' - d_1(r_2'' - r_1'')/p.$$

Таким образом, для точек $\{N_1, s_1, d_1\}$ и $\{N_2, s_2, d_2\}$ аналитический вид графов возможного семейства сформирован.

Шаг 2. Проверяется наличие членов сформированного семейства в массиве данных при увеличении диаметра на величину p. Если на выбранном фрагменте данных существование следующих членов семейства подтверждено, то считается, что семейство прошло тестирование и новое семейство найдено, и оно записывается в список новых семейств.

Таблица 4

Семейства оптимальных хордальных кольцевых графов

$2N$	s	$2N$	s
$3d^2-6d-5$	$\frac{1}{10}(3d^2-17)$	$3d^2-6d+19$	$\frac{1}{8}(3d^2+13)$
$3d^2-6d-5$	$\frac{1}{10}(3d^2-12d+7)$	$3d^2-6d+19$	$\frac{1}{8}(3d^2-12d+25)$
$3d^2-6d-5$	$\frac{1}{5}(3d^2-9d+1)$	$3d^2-4d-3$	$\frac{1}{6}(3d^2-10d+5)$
$3d^2-6d-5$	$\frac{1}{5}(3d^2-3d-11)$	$3d^2-4d-3$	$\frac{1}{6}(3d^2+2d-11)$
$3d^2-6d-1$	$\frac{1}{4}(3d^2-12d+5)$	$3d^2-4d+1$	$\frac{1}{4}(3d^2-10d+3)$
$3d^2-6d+7$	$\frac{1}{4}(3d^2-12d+13)$	$3d^2-4d+5$	$\frac{1}{6}(3d^2-10d+1)$
$3d^2-6d+7$	$\frac{1}{14}(3d^2-5)$	$3d^2-4d+5$	$\frac{1}{10}(3d^2-10d+13)$
$3d^2-6d+7$	$\frac{1}{8}(3d^2+13)$	$3d^2-4d+5$	$\frac{1}{5}(3d^2-7d+9)$
$3d^2-6d+7$	$\frac{1}{8}(3d^2-12d+1)$	$3d^2-4d+5$	$\frac{1}{5}(3d^2-d+1)$
$3d^2-6d+7$	$\frac{1}{14}(3d^2-12d+19)$	$3d^2-4d+5$	$\frac{1}{10}(3d^2+2d-3)$
$3d^2-6d+7$	$\frac{1}{7}(3d^2-9d+13)$	$3d^2-4d+5$	$\frac{1}{6}(3d^2+2d+9)$
$3d^2-6d+7$	$\frac{3}{14}(3d^2-8d+11)$	$3d^2-4d+9$	$\frac{1}{8}(3d^2-10d+11)$
$3d^2-6d+7$	$\frac{3}{14}(3d^2-4d+3)$	$3d^2-4d+9$	$\frac{1}{4}(3d^2-10d+11)$
$3d^2-6d+7$	$\frac{1}{7}(3d^2-3d+1)$	$3d^2-4d+9$	$\frac{1}{6}(3d^2-10d+17)$

Шаг 3. Поиск семейств продолжается до тех пор, пока не рассмотрены сначала диаметры для всех значений вычетов по модулю p, а затем весь заданный диапазон значений параметра p.

Шаг 4. После этого происходит валидация найденных семейств, для чего используется часть массива данных, не участвующая в ранее проведённом тестировании на шаге 2. При положительном результате считается, что новое семейство найдено.

Шаг 5. На выходе алгоритма генерируется множество аналитических описаний (формул $N(d)$ и $s(d)$) семейств п-оптимальных графов с квадратичными образующими, ограниченное заданным диапазоном изменения параметра p.

Алгоритм реализован в системе Wolfram Mathematica 10. Реализовав алгоритм для значений $36 \leqslant d \leqslant 156$ и всех чётных значений $4 \leqslant p \leqslant 18$, получили на выходе множество семейств п-оптимальных графов с квадратичными образующими, которое совпало полностью со всеми точками выделенного фрагмента массива данных на рис. 6. Общее количество полученных аналитически описываемых семейств п-оптимальных графов с квадратичными образующими равно 455. В табл. 4 приведён фрагмент описаний оптимальных семейств, полученных при реализации алгоритма 1. В табл. 5 приведён фрагмент полученных оптимальных семейств вместе с их периодом повторяемости и видом диаметров.

Для того, чтобы продвигаться дальше, увеличивая порядки графов, необходимо увеличивать значения параметра p, тем самым выявлять новые возникающие семейства.

Описанный алгоритм открытия семейств п-оптимальных хордальных графов позволяет также получать аналитические описания множеств семейств оптимальных графов с квадратичными образующими, диаметры которых превосходят точную нижнюю границу $lb(N)$, так как положенные в его основу принципы построения семейств могут быть использованы для других оптимальных хордальных колец. Все найденные семейства оптимальных графов также включены в отдельный раздел датасета.

9. Алгоритм поиска семейств оптимальных графов с линейными образующими

Для автоматизации поиска семейств оптимальных хордальных колец с линейными образующими вида $s = 3d - \alpha$, где α принимает любые целые значения, в том числе и отрицательные, был разработан эвристический алгоритм 2. Анализируя известные семейства оптимальных хордальных колец с линейными образующими, мы пришли к выводу, что условия для включения точек $\{N, s, d\}$ массива данных в искомое семейство, как и вид формул для N и s графов семейства, одинаковы

Таблица 5

Семейства оптимальных графов с указанием периода повторяемости и вида диаметров

$2N/3$	s	p	$d \bmod p$	$2N/3$	s	p	$d \bmod p$
$(d-1)^2$	$\frac{1}{2}(d^2-3)$	6	1	$(d-1)^2$	$\frac{1}{2}(d^2+1)$	12	9
$(d-1)^2$	$\frac{1}{2}(d^2-3)$	12	1	$(d-1)^2$	$\frac{1}{2}(d^2+1)$	18	3
$(d-1)^2$	$\frac{1}{2}(d^2-3)$	12	7	$(d-1)^2$	$\frac{1}{2}(d^2+1)$	18	9
$(d-1)^2$	$\frac{1}{2}(d^2-3)$	18	1	$(d-1)^2$	$\frac{1}{2}(d^2+1)$	18	15
$(d-1)^2$	$\frac{1}{2}(d^2-3)$	18	7	$(d-1)^2$	$\frac{1}{2}(d^2-4d+1)$	6	1
$(d-1)^2$	$\frac{1}{2}(d^2-3)$	18	13	$(d-1)^2$	$\frac{1}{2}(d^2-4d+1)$	12	1
$(d-1)^2$	$\frac{3}{8}(d^2-1)$	8	3	$(d-1)^2$	$\frac{1}{2}(d^2-4d+1)$	12	7
$(d-1)^2$	$\frac{3}{8}(d^2-1)$	16	3	$(d-1)^2$	$\frac{1}{2}(d^2-4d+1)$	18	1
$(d-1)^2$	$\frac{3}{8}(d^2-1)$	16	11	$(d-1)^2$	$\frac{1}{2}(d^2-4d+1)$	18	7
$(d-1)^2$	$\frac{1}{2}(d^2+1)$	6	3	$(d-1)^2$	$\frac{1}{2}(d^2-4d+1)$	18	13
$(d-1)^2$	$\frac{1}{2}(d^2+1)$	12	3	$(d-1)^2$	$\frac{3}{8}(d^2-4d+3)$	8	7

для чётных и нечётных диаметров d. Виды семейств графов с нечётным (чётным) диаметром повторяются через $p = 2$. Алгоритм 2 строится аналогично алгоритму 1 с использованием тех же обозначений. Алгоритм 2 применяется для поиска семейств графов с любыми диаметрами.

Алгоритм 2 (автоматический поиск семейств оптимальных графов)

Шаг 1. Выбирается начальное значение чётного (нечётного) диаметра $d = d_1$. Из массива данных сначала в отдельный список выбираются точки вида $\{N_1, s_1, d_1\}$, а затем в другой список — все точки вида $\{N_2, s_2, d_2\}$ с $d_2 = d_1 + 2$, для которых выполняются условия

$$\lfloor N_2/s_2 \rfloor = \lfloor N_1/s_1 \rfloor + (d_2 - d_1)/2,$$
$$N_2 \bmod s_2 = N_1 \bmod s_1 + 3(d_2 - d_1)/2.$$

Последовательно образуются все возможные сочетания точек $\{N_1, s_1, d_1\}$ и $\{N_2, s_2, d_2\}$. Для очередной пары $\{N_1, s_1, d_1\}$ и $\{N_2, s_2, d_2\}$ создаётся общая формула возможного семейства графов с N и s в виде полиномов от диаметра:

$$N = ad^2 + bd + c, \quad s = 3d - 3d_1 + s_1. \tag{8}$$

Для этого используется свойство повторяемости вида членов семейства через $k = (d - d_1)/2$ шагов при образовании порядков с помощью формулы, сохраняющей пропорциональность наращивания коэффициентов при степенях d:

$$N = (q_1 + k)s + r_1 + 3k. \tag{9}$$

Аналогично алгоритму 1 подстановкой $k = (d - d_1)/2$ и s из (8) в (9) и сборкой коэффициентов при степенях d получаются коэффициенты для полинома N:

$$a = 3/2, \quad b = 3q_1 - 3d_1 + (s_1 + 3)/2,$$
$$c = q_1(s_1 - 3d_1) + r_1 + d_1(3d_1 - s_1 - 3)/2.$$

Таким образом, для точек $\{N_1, s_1, d_1\}$ и $\{N_2, s_2, d_2\}$ аналитический вид графов возможного семейства сформирован.

Шаг 2. Проверяется наличие членов сформированного семейства в массиве данных при увеличении диаметра на величину $p = 2$. Если на выбранном фрагменте данных существование следующих членов семейства подтверждено, то считается, что семейство прошло тестирование и новое семейство найдено, и оно записывается в список новых семейств.

Шаг 3. Поиск семейств продолжается до тех пор, пока не рассмотрены все имеющиеся паросочетания графов с диаметрами $d = d_1$ и $d = d_2$.

Шаг 4. Проводится валидация полученных семейств, для чего используется часть массива данных, не участвовавшая в ранее проведённом тестировании на шаге 2. При положительном результате считается, что новое семейство найдено, и его аналитическое описание пополняет список найденных семейств.

Шаг 5. На выходе алгоритма генерируется множество аналитических описаний (формул $N(d)$ и $s(d)$) семейств оптимальных хордальных графов с линейными образующими вида $s = 3d - \alpha$.

Алгоритм реализован в системе Wolfram Mathematica 10. Реализовав алгоритм для чётных и нечётных значений диаметров, получили на выходе множество семейств оптимальных графов с линейными образующими. Общее количество полученных аналитически описываемых семейств оптимальных графов с линейными образующими равно 80. Описания всех оптимальных семейств, полученных при реализации алгоритма 2, приведены в разделе датасета.

10. Проблема маршрутизации в хордальных кольцах степени три

В работе [1] авторы показали, что способность эффективной маршрутизации в хордальных кольцах зависит от диаметра и соответственно от длины хорды. Они предложили аналитический алгоритм вычисления кратчайшего пути по номерам вершин источника и приёмника. Алгоритм рассмотрен относительно источника $u = 0$, и суть его состоит в том, что для расчёта требуемого числа шагов по образующим в кратчайшем пути в вершину v из $u = 0$ найдены некоторые соотношения между s

и значениями $\lfloor v/w \rfloor$ и $v \bmod w$, где $w = s + 1$. Проверить правильность изложенного алгоритма невозможно, так как в статье он представлен не полностью (со ссылкой [23] на диссертацию, которая нам не доступна). Но учитывая, что авторы рассмотрели не все возможные типы движений по образующим в графе, можно сказать, что в общем их алгоритм даёт путь из источника в приёмник, но не всегда кратчайший, и оценить его длину не представляется возможным.

В [8] для хордальных графов представлены два квазиоптимальных алгоритма маршрутизации с гарантированными границами порядка $s-3$ дополнительной длины пути, превышающей длину кратчайшего пути.

В [24, 25] для упрощения расчёта маршрутизации в хордальных кольцевых графах также рассмотрено частично ограниченное множество путей в приёмник, разработаны два квазиоптимальных алгоритма маршрутизации и определён процент и частота превышения длины кратчайшего пути при экспериментальной реализации этих алгоритмов маршрутизации.

Таким образом, проблема разработки оптимального (аналитического) алгоритма маршрутизации в хордальных кольцевых графах остаётся не решённой до настоящего времени, а в связи с возможностью реализации этих структур в сетях на кристалле [26] по-прежнему актуальна.

Заключение

В настоящей работе рассмотрена проблема построения семейств оптимальных по диаметру хордальных колец с аналитическим описанием в виде полиномов от диаметра сети. Построен большой массив экспериментальных данных, содержащий параметры оптимальных по диаметру хордальных колец для числа вершин $12 \leqslant N \leqslant 60000$. Разработан новый метод, и реализованы два алгоритма автоматического поиска аналитических описаний семейств оптимальных хордальных колец. Найдены новое аналитическое описание образующих экстремальных хордальных колец с максимальным числом вершин при заданном диаметре, а также аналитические описания более 500 новых семейств оптимальных графов с разного типа образующими на больших диапазонах изменения диаметра. Как показал предварительный анализ, найденный метод автоматического получения аналитических описаний новых оптимальных семейств может быть обобщён и использован для открытия новых семейств в классе оптимальных циркулянтных сетей вида $C(N; 1, s)$.

Построенный массив данных оптимальных хордальных колец является эффективным инструментом для дальнейшего изучения топологических и коммуникативных свойств хордальных сетей — среднего расстояния, надёжности, пропускной способности, алгоритмов маршрутизации, а также для открытия новых математических закономерностей

при поиске описаний оптимальных графов и условий их существования. Оптимальные графы уже полученных семейств могут быть рассмотрены в качестве основы для построения моделей надёжных иерархических структур связи сетей на кристалле и суперкомпьютерных систем.

Финансирование работы

Исследование выполнено в рамках государственного задания Института вычислительной математики и математической геофизики СО РАН (соглашение № FWNM–2022–0005).

Конфликт интересов

Авторы заявляют, что у них нет конфликта интересов.

Литература

1. **Arden B. W., Lee H.** Analysis of chordal ring network // IEEE Trans. Comput. 1981. V. C-30, No. 4. P. 291–295.
2. **Morillo P., Comellas F., Fiol M. A.** The optimization of chordal ring networks // Communication technology. Singapore: World Scientific, 1987. P. 295–299.
3. **Yebra J. L. A., Fiol M. A., Morillo P., Alegre I.** The diameter of undirected graphs associated to plane tessellations // Ars Comb. 1985. V. 20-B. P. 159–171.
4. **Bermond J. C., Comellas F., Hsu D. F.** Distributed loop computer-networks: A survey // J. Parallel Distrib. Comput. 1995. V. 24, No. 1. P. 2–10.
5. **Hwang F. K.** A survey on multi-loop networks // Theor. Comput. Sci. 2003. V. 299, No. 1–3. P. 107–121.
6. **Monakhova E. A.** A survey on undirected circulant graphs // Discrete Math. Algorithms Appl. 2012. V. 4, No. 1. Paper ID 1250002. 30 p.
7. **Pedersen J. M., Riaz M. T., Madsen O. B.** Distances in generalized double rings and degree three chordal rings // Parallel and distributed computing and networks. Proc. IASTED Int. Conf. (Innsbruck, Austria, Feb. 15–17, 2005). Calgary: ACTA Press, 2005. P. 153–158.
8. **Parhami B.** Periodically regular chordal rings are preferable to double-ring networks // J. Interconnect. Netw. 2008. V. 9, No. 1. P. 99–126.
9. **Farah R. N., Chien S. L. E., Othman M.** Optimum free-table routing in the optimised degree six 3-modified chordal ring network // J. Commun. 2017. V. 12, No. 12. P. 677–682.
10. **Hwang F. K., Wright P. E.** Survival reliability of some double-loop networks and chordal rings // IEEE Trans. Comput. 1995. V. 44, No. 12. P. 1468–1471.
11. **Chen S. K., Hwang F. K., Liu Y. C.** Some combinatorial properties of mixed chordal rings // J. Interconnect. Netw. 2003. V. 4, No. 1. P. 3–16.

12. **Stojmenović I.** Honeycomb networks: Topological properties and communication algorithms // IEEE Trans. Parallel Distrib. Syst. 1997. V. 8. P. 281–305.
13. **Ahmad M., Zahid Z., Zavaid M., Bonyah E.** Studies of chordal ring networks via double metric dimensions // Math. Probl. Eng. 2022. V. 98. Paper ID 8303242. 7 p.
14. **Bujnowski S., Marciniak B., Oyerinde O. O., Lutowski Z., Flizikowski A., Galan S. G.** The possibility of equalising the transmission properties of networks described by chordal rings // Proc. 15th Int. Conf. Signal Processing and Communication Systems (Sydney, Australia, Dec. 13–15, 2021). Piscataway: IEEE, 2021. 8 p.
15. **Irwan N., Othman M., Farah R. N.** Network properties for classes of chordal ring network degree three topologies // 2010 Proc. Int. Conf. Intelligent Network and Computing (Kuala Lumpur, Malaysia, Nov. 26–28, 2010). Piscataway: IEEE, 2010. P. 16–20.
16. **Monakhov O. G., Monakhova E. A.** A class of parametric regular networks for multicomputer architectures // Comput. Sist. J. 2000. V. 4, No. 2. P. 85–93.
17. **Huang H.-C.** The diameter-edge invariant property of chordal ring networks: Master Thes. Hsinchu: Natl. Chiao Tung Univ., 2009. 30 p.
18. **Loudiki L., Kchkech M., Essaky E. H.** A new approach for computing the distance and the diameter in circulant graphs. Ithaca, NY: Cornell Univ., 2022. 14 p. (Cornell Univ. Libr. e-Print Archive; arXiv:2210.11116).
19. **Monakhov O. G., Monakhova E. A., Kireev S. E.** Parallel generation and analysis of optimal chordal ring networks using Python tools on Kunpeng prosessors // Parallel computing technologies. Proc. 17th Int. Conf. (Astana, Kazakhstan, Aug. 21–25, 2023). Cham: Springer, 2023. P. 126–135. (Lect. Notes Comput. Sci.; V. 14098).
20. **Mollin R. A.** Quadratic polynomials producing consecutive, distinct primes and class groups of complex quadratic fields // Acta Arithmetica. 1996. V. 74. P. 17–30.
21. **Marwan N., Romano M. C., Thiel M., Kurths J.** Recurrence plots for the analysis of complex systems // Phys. Rep. 2007. V. 438, No. 5–6. P. 237–329.
22. **Barriere L.** Symmetry properties of chordal rings of degree 3 // Discrete Appl. Math. 2003. V. 129. P. 211–232.
23. **Lee H.** Modeling of multi-microcomputer networks: PhD Thes. Princeton, NJ: Princeton Univ., 1979.
24. **Gutierrez J., Riaz T., Pedersen J., Labeaga S., Madsen O.** Degree 3 networks topological routing // Image Process. Commun. 2009. V. 14, No. 4. P. 35–48.
25. **Gutierrez J., Riaz T., Pedersen J., Labeaga S., Madsen O.** On topological routing on degree 3 chordal rings // Proc. 1st Int. Conf. Image Processing & Communications (Bydgoszcz, Poland, Sept. 16–18, 2009). Warsaw: Acad. Publ. House EXIT, 2009. Paper ID 59020255. 8 p.

26. Monakhova E. A., Monakhov O. G., Romanov A. Yu. Routing algorithms in optimal degree four circulant networks based on relative addressing: comparative analysis for networks-on-chip // IEEE Trans. Netw. Sci. Eng. 2023. V. 10, No. 1. P. 413–425.

Монахова Эмилия Анатольевна
Монахов Олег Геннадьевич

Статья поступила
9 августа 2023 г.
После доработки —
30 августа 2023 г.
Принята к публикации
22 сентября 2023 г.

DISKRETNYI ANALIZ I ISSLEDOVANIE OPERATSII
/DISCRETE ANALYSIS AND OPERATIONS RESEARCH/
January–March 2024. Vol. 31, No. 1. P. 85–108

UDC 519.176 DOI: 10.33048/daio.2024.31.780

A METHOD FOR AUTOMATIC SEARCH FOR FAMILIES OF OPTIMAL CHORDAL RING NETWORKS

E. A. Monakhova[a] and *O. G. Monakhov*[b]

Institute of Computational Mathematics and Mathematical Geophysics,
6 Acad. Lavrentiev Avenue, 630090 Novosibirsk, Russia
E-mail: [a] emilia@rav.sscc.ru, [b] monakhov@rav.sscc.ru

Abstract. Arden and Lee proposed a class of chordal ring networks of degree three as communication networks for multicomputer systems, derived a formula for the diameter, and produced an algorithm for finding the shortest paths for them. In this paper, it is shown that the formula for the diameter and the routing algorithm presented by them are inaccurate. We have obtained a large dataset containing parameters for describing optimal diameter chord rings for all the numbers of nodes up to 60 000 and found the exact lower bound for the diameter of chordal ring networks. A new method is proposed and the algorithms for automatic search for analytical descriptions of families of optimal chordal rings are realized based on an analysis of the database. Using the latter, analytical descriptions of over 500 new families of optimal chordal ring networks for many values of the number of nodes are found (only six families have been known until now in the literature). Tab. 5, illustr. 6, bibliogr. 26.

Keywords: optimal degree-three chordal network, diameter, extremal chordal graph, optimal chordal ring dataset.

References

1. **B. W. Arden** and **H. Lee,** Analysis of chordal ring network, *IEEE Trans. Comput.* **C-30** (4), 291–295 (1981).
2. **P. Morillo, F. Comellas,** and **M. A. Fiol,** The optimization of chordal ring networks, in *Communication Technology* (World Scientific, Singapore, 1987), pp. 295–299.

English transl.: *Journal of Applied and Industrial Mathematics* **18** (1), 122–136 (2024), DOI: 10.1134/S1990478924010113.

3. **J. L. A. Yebra, M. A. Fiol, P. Morillo,** and **I. Alegre,** The diameter of undirected graphs associated to plane tessellations, *Ars Comb.* **20-B**, 159–171 (1985).
4. **J. C. Bermond, F. Comellas,** and **D. F. Hsu,** Distributed loop computer-networks: A survey, *J. Parallel Distrib. Comput.* **24** (1), 2–10 (1995).
5. **F. K. Hwang,** A survey on multi-loop networks, *Theor. Comput. Sci.* **299** (1–3), 107–121 (2003).
6. **E. A. Monakhova,** A survey on undirected circulant graphs, *Discrete Math. Algorithms Appl.* **4** (1), ID 1250002 (2012).
7. **J. M. Pedersen, M. T. Riaz,** and **O. B. Madsen,** Distances in generalized double rings and degree three chordal rings, in *Parallel and Distributed Computing and Networks,* Proc. IASTED Int. Conf. (Innsbruck, Austria, Feb. 15–17, 2005) (ACTA Press, Calgary, 2005), pp. 153–158.
8. **B. Parhami,** Periodically regular chordal rings are preferable to double-ring networks, *J. Interconnect. Netw.* **9** (1), 99–126 (2008).
9. **R. N. Farah, S. L. E. Chien,** and **M. Othman,** Optimum free-table routing in the optimised degree six 3-modified chordal ring network, *J. Commun.* **12** (12), 677–682 (2017).
10. **F. K. Hwang** and **P. E. Wright,** Survival reliability of some double-loop networks and chordal rings, *IEEE Trans. Comput.* **44** (12), 1468–1471 (1995).
11. **S. K. Chen, F. K. Hwang,** and **Y. C. Liu,** Some combinatorial properties of mixed chordal rings, *J. Interconnect. Netw.* **4** (1), 3–16 (2003).
12. **I. Stojmenović,** Honeycomb networks: Topological properties and communication algorithms, *IEEE Trans. Parallel Distrib. Syst.* **8**, 281–305 (1997).
13. **M. Ahmad, Z. Zahid, M. Zavaid,** and **E. Bonyah,** Studies of chordal ring networks via double metric dimensions, *Math. Probl. Eng.* **98**, ID 8303242 (2022).
14. **S. Bujnowski, B. Marciniak, O. O. Oyerinde, Z. Lutowski, A. Flizikowski,** and **S. G. Galan,** The possibility of equalising the transmission properties of networks described by chordal rings, in *Proc. 15th Int. Conf. Signal Processing and Communication Systems (Sydney, Australia, Dec. 13–15, 2021)* (IEEE, Piscataway, 2021).
15. **N. Irwan, M. Othman,** and **R. N. Farah,** Network properties for classes of chordal ring network degree three topologies, in *2010 Proc. Int. Conf. Intelligent Network and Computing (Kuala Lumpur, Malaysia, Nov. 26–28, 2010)* (IEEE, Piscataway, 2010), pp. 16–20.
16. **O. G. Monakhov** and **E. A. Monakhova,** A class of parametric regular networks for multicomputer architectures, *Comput. Sist. J.* **4** (2), 85–93 (2000).
17. **H.-C. Huang,** The diameter-edge invariant property of chordal ring networks, *Master Thesis* (Natl. Chiao Tung Univ., Hsinchu, 2009).
18. **L. Loudiki, M. Kchkech,** and **E. H. Essaky,** A new approach for computing the distance and the diameter in circulant graphs (Cornell Univ., Ithaca, NY, 2022) (Cornell Univ. Libr. e-Print Archive; arXiv:2210.11116).

19. O. G. Monakhov, E. A. Monakhova, and **S. E. Kireev,** Parallel generation and analysis of optimal chordal ring networks using Python tools on Kunpeng prosessors, in *Parallel Computing Technologies,* Proc. 17th Int. Conf. (Astana, Kazakhstan, Aug. 21–25, 2023) (Springer, Cham, 2023), pp. 126–135 (Lect. Notes Comput. Sci., Vol. 14098).

20. R. A. Mollin, Quadratic polynomials producing consecutive, distinct primes and class groups of complex quadratic fields, *Acta Arithmetica* **74**, 17–30 (1996).

21. N. Marwan, M. C. Romano, M. Thiel, and **J. Kurths,** Recurrence plots for the analysis of complex systems, *Phys. Rep.* **438** (5–6), 237–329 (2007).

22. L. Barriere, Symmetry properties of chordal rings of degree 3, *Discrete Appl. Math.* **129**, 211–232 (2003).

23. H. Lee, Modeling of multi-microcomputer networks, *PhD Thesis* (Princeton Univ., Princeton, NJ, 1979).

24. J. Gutierrez, T. Riaz, J. Pedersen, S. Labeaga, and **O. Madsen,** Degree 3 networks topological routing, *Image Process. Commun.* **14** (4), 35–48 (2009).

25. J. Gutierrez, T. Riaz, J. Pedersen, S. Labeaga, and **O. Madsen,** On topological routing on degree 3 chordal rings, in *Proc. 1st Int. Conf. Image Processing & Communications (Bydgoszcz, Poland, Sept. 16–18, 2009)* (Acad. Publ. House EXIT, Warsaw, 2009), ID 59020255.

26. E. A. Monakhova, O. G. Monakhov, and **A. Yu. Romanov,** Routing algorithms in optimal degree four circulant networks based on relative addressing: comparative analysis for networks-on-chip, *IEEE Trans. Netw. Sci. Eng.* **10** (1), 413–425 (2023).

Emilia A. Monakhova
Oleg G. Monakhov

Received August 9, 2023
Revised August 30, 2023
Accepted September 22, 2023

ДИСКРЕТНЫЙ АНАЛИЗ И ИССЛЕДОВАНИЕ ОПЕРАЦИЙ
Январь–март 2024. Т. 31, № 1. С. 109–128

УДК 519.17 DOI: 10.33048/daio.2024.31.770

О КОЛИЧЕСТВЕ k-ДОМИНИРУЮЩИХ НЕЗАВИСИМЫХ МНОЖЕСТВ В ПЛАНАРНЫХ ГРАФАХ

Д. С. Талецкий[1, 2]

[1] Национальный исследовательский университет «Высшая школа экономики», ул. Большая Печёрская, 25/12, 603155 Нижний Новгород, Россия

[2] Санкт-Петербургский гос. университет, Университетская наб., 7/9, 199034 Санкт-Петербург, Россия

E-mail: dmitalmail@gmail.com

Аннотация. Множество J_k вершин графа называется k-доминирующим независимым ($k \geqslant 1$), если его вершины попарно не смежны и каждая вершина не из J_k смежна хотя бы с k вершинами из J_k. В этой статье получены новые оценки количества k-доминирующих независимых множеств при различных значениях $k \geqslant 2$ в некоторых классах планарных графов. Ил. 7, библиогр. 15.

Ключевые слова: независимое множество, доминирующее множество, k-доминирующее независимое множество, планарный граф.

Введение

Граф называется *планарным*, если он может быть уложен на плоскости без пересечений рёбер не по вершинам, и *внешнепланарным*, если существует его плоская укладка такая, что все вершины графа принадлежат внешней грани. Планарный (внешнепланарный) граф называется *максимальным*, если в него нельзя добавить ребро, не нарушая свойства планарности (внешнепланарности). Обозначим через $\mathcal{P}$, $\mathcal{MP}$, $\mathcal{OP}$ и $\mathcal{MOP}$ классы всех планарных, максимальных планарных, внешнепланарных и максимальных внешнепланарных графов соответственно.

Независимым множеством графа называется произвольное подмножество попарно не смежных его вершин. Множество вершин D_k называется *k-доминирующим* ($k \geqslant 1$), если каждая вершина не из D_k смежна хотя бы с k вершинами из D_k. В настоящей работе рассматриваются k-доминирующие независимые множества (сокращённо k-ДНМ). Следуя [1, 2], будем обозначать через $\mathrm{mi}_k(n, \mathcal{F})$ максимально возможное количество k-ДНМ, которое может содержать n-вершинный граф из класса $\mathcal{F}$. Если $\mathcal{F}$ совпадает с классом всех графов, то будем использовать обозначение $\mathrm{mi}_k(n)$.

Как известно, каждое 1-ДНМ графа является его максимальным по включению независимым множеством (и наоборот). В [3] получено точное значение величины $\mathrm{mi}_1(n)$ при всех $n \geqslant 1$. Позднее в [4] было предложено значительно более простое доказательство этого результата. На сегодняшний день существует большое число работ, посвящённых перечислению 1-ДНМ в графах из тех или иных классов. Так, известны точные значения величины $\mathrm{mi}_1(n, \mathcal{F})$ в случае, когда $\mathcal{F}$ является классом связных графов [5], деревьев [6], двудольных графов [7], графов без треугольников [8], уницикличеких [9] и r-циклических [10] графов, уницикличеких связных графов [11].

Известно несколько работ, в которых исследуются свойства 2-ДНМ. В статье [12] описаны некоторые классы графов, содержащих хотя бы одно 2-ДНМ. В [13, 14] сформулированы достаточные условия существования 2-ДНМ в декартовых и тензорных произведениях графов соответственно. В [15] исследуются обобщения графа Петерсена, содержащие хотя бы одно 2-ДНМ.

Автору известны лишь две работы [1, 2], связанные с перечислением k-ДНМ при $k \geqslant 3$. В [1] доказано существование констант $c, c', c_k, c'_k > 0$ таких, что $1{,}22^n c \leqslant \mathrm{mi}_2(n) \leqslant 1{,}246^n c'$ и при любом $k \geqslant 3$ верно неравенство $(\sqrt[2k]{2})^n c_k \leqslant \mathrm{mi}_k(n) \leqslant (\sqrt[k+1]{2})^n c'_k$. Кроме того, в [1] получен ряд других результатов: например, для всех $n, k \geqslant 2$ и класса деревьев $\mathcal{T}$ доказано неравенство $\mathrm{mi}_k(n, \mathcal{T}) \leqslant 1$. Позднее, в [2] были получены новые верхние и нижние оценки величины $\mathrm{mi}_k(n)$, в частности, доказано неравенство $\mathrm{mi}_k(n) < (\sqrt[k]{1{,}98})^n$ для всех $k \geqslant 3$.

В настоящей работе получены новые оценки величины $\mathrm{mi}_k(n, \mathcal{F})$, где $\mathcal{F} \in \{\mathcal{P}, \mathcal{MP}, \mathcal{OP}, \mathcal{MOP}\}$ и $k \geqslant 2$. Показано, что при $k \geqslant 3$ каждый внешнепланарный граф содержит не более одного k-ДНМ. При этом каждый максимальный внешнепланарный граф содержит не более одного 2-ДНМ. С другой стороны, при $k \geqslant 4$ каждый планарный граф содержит не более одного k-ДНМ, но при всех $n \geqslant 12$ существуют максимальные планарные графы, содержащие не менее чем $2^{[n/50]-1}$ 3-ДНМ.

1. Терминология и обозначения

Грань плоского графа называется *внутренней*, если она имеет конечную площадь, и *внешней* в противном случае. Назовём ребро плоского графа *внешним*, если оно лежит на его внешней грани. Всюду в работе будем использовать термин «грань» вместо термина «внутренняя грань». Кроме того, будем предполагать, что все рассматриваемые планарные графы уложены на плоскости. Будем говорить, что грани f_1 и f_2 *смежны*, если они имеют общее ребро. *Треугольником* называется грань, содержащая три вершины.

Назовём грань максимального внешнепланарного графа *крайней* или 1-*крайней*, если она содержит вершину степени 2. При $s \geqslant 2$ назовём грань s-*крайней*, если она не $(s-1)$-крайняя и все смежные с ней грани, кроме, быть может, одной, s'-крайние для некоторого $1 \leqslant s' < s$. Будем говорить, что грань $\{1,2,3\}$-*крайняя*, если она s-крайняя для некоторого $s \in \{1,2,3\}$.

Под *добавлением вершины* v *в грань* f планарного графа G будем понимать добавление в G вершины v и соединение её со всеми вершинами f. Отметим, что если G является максимальным планарным графом, то он останется таковым при добавлении вершины в любую его грань. Как известно, каждая внутренняя грань максимального (внешне)планарного графа является треугольником.

Обозначим через $T(G)$ *слабо двойственный граф* графа $G \in \mathcal{P}$, вершинами которого являются внутренние грани G. Две вершины $T(G)$ соединены ребром, если и только если соответствующие им грани G смежны. Как известно, если $G \in \mathcal{OP}$, то $T(G)$ является лесом, если при этом G двусвязен, то $T(G)$ является деревом, если же $G \in \mathcal{MOP}$, то $T(G)$ является субкубическим деревом.

При $k \geqslant 1$ вершину графа G назовём k-*универсальной* (соответственно k-*пустой*), если она содержится в каждом k-ДНМ графа G (соответственно не содержится ни в одном k-ДНМ графа G).

Как обычно, через $V(G)$ и $E(G)$ обозначаются множества вершин и рёбер графа G соответственно. Пусть $A \subset V(G)$. Через $G \setminus A$ обозначается порождённый подграф G с множеством вершин $V(G) \setminus A$. Граф, полученный в результате стягивания ребра $uv \in E(G)$ графа G, обозначается через G/uv. *Объединением* $G \cup H$ графов G и H называется граф с множеством вершин $V(G) \cup V(H)$ и множеством рёбер $E(G) \cup E(H)$. Через kG обозначается объединение $k \geqslant 2$ копий графа G. Через $\deg(v), N(v)$ и $N[v]$ обозначаются степень вершины v, открытая окрестность v и замкнутая окрестность v соответственно. Через $\delta(G)$ обозначается наименьшая степень вершины в G.

Деревом называется граф без циклов, а *листом* дерева — его вершина степени 1. *Диаметром* связного графа называется наибольшее попарное расстояние между его вершинами, а *диаметральным путём* графа — некоторый простой путь, длина которого равна его диаметру. Отметим, что в любом дереве, содержащем не менее двух вершин, концы каждого его диаметрального пути являются листьями.

Через K_n, C_n и P_n обозначаются n-вершинный полный граф, простой цикл и простой путь соответственно. Обозначим через W_n n-вершинный граф, полученный путём добавления в цикл C_{n-1} новой вершины, смежной со всеми вершинами цикла.

Точкой сочленения связного графа называется вершина, при удалении которой граф перестаёт быть связным. *Блоком* B графа G называется его максимальный по включению двусвязный подграф. Блок B называется *крайним* в графе G, если он содержит не более одной точки сочленения G.

2. Предварительные результаты

В этом разделе приводятся некоторые простые факты, которые будут использованы при доказательстве основных результатов работы.

Лемма 1. *Для любого графа G, содержащего хотя бы 4 вершины, имеют место следующие утверждения.*

1. *Если $G \in \mathcal{P}$, то $\delta(G) \leqslant 5$. Если же $G \in \mathcal{MP}$, то $\delta(G) \in \{3, 4, 5\}$.*
2. *Если $G \in \mathcal{OP}$, то $\delta(G) \leqslant 2$. Если же $G \in \mathcal{MOP}$, то $\delta(G) = 2$.*

Доказательство. П. 1 следует из формулы Эйлера. Докажем п. 2. Если $|V(G)| \leqslant 3$, то доказывать нечего; предположим, что $|V(G)| \geqslant 4$. Если $G \in \mathcal{MOP}$, то граф $T(G)$ является деревом и содержит лист, который соответствует грани G, содержащей вершину степени 2, откуда $\delta(G) \leqslant 2$. Поскольку каждая внутренняя грань G является треугольником, то $\delta(G) = 2$. Если же $G \notin \mathcal{MOP}$, то существует граф $G^* \in \mathcal{MOP}$ такой, что G является остовным подграфом G^*, откуда $\delta(G) \leqslant \delta(G^*) = 2$. Лемма 1 доказана.

Лемма 2. *Для любого графа $G \in \mathcal{MOP}$, вершины $w \in V(G)$ и ребра $uv \in E(G)$ имеют место следующие утверждения.*

1. *Если $\deg(w) = 2$, то $G \setminus \{w\} \in \mathcal{MOP}$.*
2. *Если uv — внешнее ребро G, то $G/uv \in \mathcal{MOP}$.*

Доказательство. Поскольку каждая внутренняя грань максимального внешнепланарного графа является треугольником, п. 1 очевиден. Докажем п. 2. Обозначим через uvw единственную внутреннюю грань G, содержащую ребро uv. Очевидно, что все внутренние грани G, кроме uvw, являются внутренними гранями графа G/uv, следовательно, каждая внутренняя грань G/uv является треугольником. Кроме того, поскольку G не содержит точек сочленения и $G \setminus \{u, v\}$ связен, граф G/uv также не содержит точек сочленения, следовательно, $G/uv \in \mathcal{MOP}$. Лемма 2 доказана.

Лемма 3. *Для любого $k \geqslant 1$, графа $G \in \mathcal{P}$ и его вершин $u, v \in V(G)$ имеют место следующие утверждения.*

1. *Если u k-пустая в G, то $\mathrm{mi}_k(G \setminus u) \geqslant \mathrm{mi}_k(G)$.*
2. *Если v k-универсальная в G, то $\mathrm{mi}_k(G \setminus N[v]) \geqslant \mathrm{mi}_k(G)$.*

Доказательство. Покажем, что любое k-ДНМ J графа G будет являться k-ДНМ и для графа $G \setminus u$. Пусть $w \in V(G) \setminus u$ и $w \notin J$. Тогда w имеет хотя бы k соседей в G, принадлежащих J. Поскольку все эти соседи не k-пусты в G, они отличны от u и принадлежат графу $G \setminus u$. Таким образом, каждая вершина графа $G \setminus u$ либо принадлежит независимому множеству J, либо имеет в нём не менее k соседей, что и требовалось.

Докажем п. 2. Из определения k-универсальной вершины следует, что все вершины окрестности $N(v)$ k-пусты. Тогда для каждого k-ДНМ J графа G множество $J \setminus v$ является k-ДНМ графа $G \setminus N[v]$. Лемма 3 доказана.

3. Внешнепланарные графы

3.1. Класс $\mathcal{OP}$. В упомянутой ранее работе [3] доказана

Теорема 1. *Для всех $n \geqslant 2$ имеет место равенство*

$$\mathrm{mi}_1(n) = \begin{cases} 3^m, & \text{если } n = 3m, \\ 4 \cdot 3^{m-1}, & \text{если } n = 3m+1, \\ 2 \cdot 3^m, & \text{если } n = 3m+2. \end{cases}$$

Как показано в [3], если $n = 3m$, то единственным экстремальным графом является mK_3; если $n = 3m+1$, то существуют два экстремальных графа: $(m-1)K_3 \cup 2K_2$ и $(m-1)K_3 \cup K_4$; если же $n = 3s+2$, то единственным экстремальным графом является $mK_3 \cup K_2$. Поскольку при любом n хотя бы один из экстремальных графов внешнепланарный, то $\mathrm{mi}_1(n, \mathcal{OP}) = \mathrm{mi}_1(n, \mathcal{P}) = \mathrm{mi}_1(n)$.

В этом пункте установлены точные значения величины $\mathrm{mi}_k(n, \mathcal{OP})$ при $n \geqslant 1$ и $k \geqslant 2$.

Теорема 2. *Для всех $n \geqslant 1$ верно равенство $\mathrm{mi}_2(n, \mathcal{OP}) = 2^{[n/4]}$. При этом если $n = 4m$, то экстремальный граф единствен и изоморфен mC_4.*

Доказательство. Отметим, что $\mathrm{mi}_2(mC_4 \cup rK_1) = 2^m$ при любом $n = 4m+r \geqslant 1$. Предположим, что найдётся n-вершинный граф G такой, что либо $\mathrm{mi}_2(G) > 2^{[n/4]}$, либо $\mathrm{mi}_2(G) = 2^{[n/4]}$, $n = 4m$ и G не изоморфен mC_4; при этом будем полагать, что при всех $n' < n$ графов с таким свойством не существует. Тогда в G найдётся некоторая компонента связности H, не изоморфная C_4. Если H содержит хотя бы одну 2-пустую вершину u, то удалим её из графа, тогда $\mathrm{mi}_2(G) \leqslant \mathrm{mi}_2(G \setminus u)$ по лемме 3. При этом если $n = 4m$, то $\mathrm{mi}_2(G) < 2^{[n/4]}$, если же $n = 4m+r$, то $\mathrm{mi}_2(G) \leqslant 2^{[(n-1)/4]}$, что противоречит предположению. Если H содержит хотя бы одну 2-универсальную вершину v, то проведём аналогичные рассуждения для графа $G \setminus N[v]$.

Таким образом, предполагаем, что H не содержит 2-универсальных и 2-пустых вершин. Тогда $|V(H)| \geqslant 4$ и $\delta(H) = 2$, при этом ни одна вершина степени 2 графа H не принадлежит треугольнику (иначе эта вершина была бы 2-пустой). Пусть в H найдётся пара смежных вершин u и v степени 2. Обозначим через u' и v' вторых соседей вершин u и v соответственно. Каждое 2-ДНМ графа G либо содержит вершины u и v' и не содержит ни одного соседа v', либо содержит вершины v и u' и не содержит ни одного соседа u'. Поскольку H не совпадает с C_4, множество $N[u'] \cup N[v']$ содержит не менее 5 вершин. Тогда имеет место неравенство $\mathrm{mi}_2(G) \leqslant 2^{[(n-4)/2]} + 2^{[(n-5)/2]}$, причём если $n = 4m$, то неравенство строгое, что и требовалось.

Осталось рассмотреть случай, когда в H не найдётся двух смежных вершин степени 2. Обозначим через B один из крайних блоков H (если H двусвязен, то считаем, что $B \cong H$). Поскольку блок B крайний, он содержит не более одной точки сочленения H. Кроме того, так как $\delta(H) = 2$, блок B содержит хотя бы 3 вершины и, следовательно, является двусвязным внешнепланарным графом. Тогда слабо двойственный граф $T(B)$ является деревом. Если при этом $T(B)$ состоит из одной вершины, то $B \cong C_s$ для некоторого $s \geqslant 4$, при этом хотя бы $s - 1$ вершин цикла не являются точками сочленения G и, следовательно, имеют степень 2 в H; противоречие. Если же дерево $T(B)$ содержит не менее двух вершин, то оно содержит не менее двух листьев x и x', которым соответствуют некоторые грани f и f' графа H. Поскольку эти грани не являются треугольниками, обе они содержат хотя бы одну пару смежных вершин, имеющих степень 2 в B, причём хотя бы в одной из этих пар обе вершины не являются точками сочленения H и, следовательно, имеют степень 2 в H; противоречие. Теорема 2 доказана.

При $k \geqslant 3$ каждая вершина степени 2 внешнепланарного графа k-универсальна. Кроме того, любой подграф внешнепланарного графа также будет внешнепланарным. Отсюда вытекает следующий простой факт.

Теорема 3. *Для всех $k \geqslant 3$ и $n \geqslant 1$ верно $\mathrm{mi}_k(n, \mathcal{OP}) = 1$.*

Доказательство. Ясно, что при любых $n, k \geqslant 1$ пустой граф nK_1 содержит единственное k-ДНМ, откуда $\mathrm{mi}_k(n, \mathcal{OP}) \geqslant 1$. Предположим, что для некоторых $k \geqslant 3$ и $n \geqslant 1$ найдётся внешнепланарный граф G такой, что $\mathrm{mi}_k(G) > 1$, а для всех графов G', содержащих менее $|V(G)|$ вершин, верно $\mathrm{mi}_k(G') \leqslant 1$. По лемме 1 существует вершина $v \in V(G)$ степени 2, которая будет k-универсальной. Обозначим через J_1 и J_2 два различных k-ДНМ G. Тогда множества $J_1 \setminus \{v\}$ и $J_2 \setminus \{v\}$ являются различными k-ДНМ графа $G \setminus N[v] \in \mathcal{OP}$, откуда $\mathrm{mi}_2(G \setminus N[v]) \geqslant 2$; противоречие. Теорема 3 доказана.

3.2. Класс $\mathcal{MOP}$. Назовём граф $G \in \mathcal{MOP}$ *критическим*, если $\mathrm{mi}_2(G) > 1$ и при этом для любого графа $G' \in \mathcal{MOP}$, содержащего менее $|V(G)|$ вершин, верно $\mathrm{mi}_2(G') \leqslant 1$. Цель рассуждений этого пункта — доказать, что критических графов не существует.

Будем говорить, что граф $G' \in \mathcal{MOP}$ *соответствует* графу $G \in \mathcal{MOP}$, если $\varnothing \subsetneq V(G') \subsetneq V(G)$ и найдётся множество $\varnothing \subseteq A \subsetneq V(G)$ такое, что для любого 2-ДНМ J графа G множество $J \setminus A$ является 2-ДНМ графа G'. Очевидно, что имеет место неравенство $\mathrm{mi}_2(G') \geqslant \mathrm{mi}_2(G)$. Таким образом, для того чтобы показать, что граф G не критический, достаточно привести пример соответствующего ему графа G'.

Напомним, что если $G \in \mathcal{MOP}$, то его слабо двойственный граф $T(G)$ является субкубическим деревом. Рассмотрим некоторый диаметральный путь $X = x_1x_2\dots x_p$ в $T(G)$. Всюду в этом пункте будем обозначать через x_i вершины $T(G)$, лежащие на X, а через x'_j — вершины $T(G)$, не лежащие на X. Кроме того, через f_i и f'_j будем обозначать грани G, соответствующие вершинам x_i и x'_j соответственно. При $p \leqslant 4$ граф G содержит не более 6 внутренних граней и не более 8 вершин. Легко проверить, что в этом случае $\mathrm{mi}_2(G) \leqslant 1$ и G не критический. Таким образом, предполагаем, что $p \geqslant 5$.

Лемма 4. *Если в графе $G \in \mathcal{MOP}$ найдётся грань f, смежная с 1-крайними гранями f' и f'', то G не критический.*

Доказательство. Обозначим грани f, f', f'' через uvw, $uu'w$, $vv'w$ соответственно. Тогда вершины u', v' 2-универсальны, а смежные с ними вершины u, v, w 2-пусты в G. Рассмотрим граф $G_2 = G\setminus\{u', v'\}$. Очевидно, что вершина w 2-универсальна, а вершины u, v 2-пусты в G_2. Тогда для каждого 2-ДНМ J графа G множество $(J \setminus \{u', v'\}) \cup \{w\}$ является 2-ДНМ графа G_2, откуда $\mathrm{mi}_2(G) \leqslant \mathrm{mi}_2(G_2)$. Лемма 4 доказана.

Лемма 5. *Пусть граф $G \in \mathcal{MOP}$ критический. Тогда любая вершина $v \in V(G)$, принадлежащая некоторой $\{1,2,3\}$-крайней грани G, будет либо 2-универсальной, либо 2-пустой.*

Доказательство. Покажем, что $\{1,2,3\}$-крайняя грань графа G содержит 2-универсальную вершину, тогда две другие вершины этой грани 2-пусты. Поскольку все 1-крайние грани G содержат 2-универсальную вершину степени 2, условие леммы для них выполнено.

Рассмотрим некоторую 2-крайнюю грань abc графа G. Из определения 2-крайней грани следует, что abc смежна с некоторой 1-крайней гранью (например abd). Поскольку вершина d 2-универсальна, вершины a и b 2-пусты. Если $\min(\deg(a), \deg(b)) = 3$, то вершина c 2-универсальна. Если же $\min(\deg(a), \deg(b)) \geqslant 4$, то грань abc смежна с тремя другими

гранями, хотя бы две из которых 1-крайние, тогда G не является критическим графом по лемме 4; противоречие.

Наконец, рассмотрим некоторую 3-крайнюю грань uvw графа G. По определению 3-крайней грани найдётся некоторая 2-крайняя грань (например uva), смежная с uvw. Если $\deg(a) \geqslant 4$, то uva смежна с двумя 1-крайними гранями, что невозможно по лемме 4. Если же $\deg(a) = 3$, то найдётся 1-крайняя грань, содержащая вершину a и одну из вершин u и v (например u), а также некоторую 2-универсальную вершину. Тогда вершины u и a 2-пусты, а вершина v 2-универсальна, что и требовалось. Лемма 5 доказана.

Назовём граф $G \in \mathcal{MOP}$ *особым*, если в нём найдутся вершины $a_1, a_2, \ldots, a_5, b_1, b_2$ и подграф $G_0 = G \setminus \{a_1, a_2, \ldots, a_5\} \in \mathcal{MOP}$ с внешним ребром b_1b_2, расположенные таким образом, что грани $a_1a_2b_1$ и $a_4a_5b_2$ 1-крайние, а грани $a_2a_3b_1$ и $a_3a_4b_2$ 2-крайние в G. Кроме того, грань $a_1a_2b_1$ соответствует концу некоторого диаметрального пути X дерева $T(G)$ (рис. 1). При этом предполагаем, что второй конец X можно выбрать таким образом, чтобы соответствующая ему грань G была отлична от $a_4a_5b_2$ (в противном случае G содержит не более 8 вершин и, как легко проверить, не будет критическим).

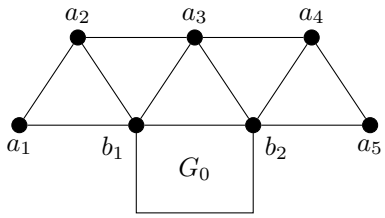

Рис. 1. Структура особого графа

Лемма 6. *Если граф G особый, то он не критический.*

Доказательство. Обозначим грани f_1, f_2, f_3, f_4 графа G через $a_1a_2b_1$, $a_2a_3b_1$, $b_1b_2a_3$, $b_1b_2c_1$ соответственно. В зависимости от значения величин $\deg(x_4)$ и $\deg(x_5)$ возможны три случая.

Случай 1: $\deg(x_4) = 3$. Обозначим через f_3' грань, смежную с f_4 и отличную от f_3 и f_5. В силу симметрии можем считать, что f_3' содержит ребро b_2c_1. Обозначим через v вершину, отличную от a_5 и такую, что b_2v является внешним ребром в G. Поскольку в G существует два внешних ребра, инцидентных b_2, вершина v единственна. Нетрудно видеть, что она принадлежит некоторой $\{1,2,3\}$-крайней грани G, а значит, будет либо 2-универсальной, либо 2-пустой по лемме 5. Если v 2-универсальна в G, то граф $G \setminus \{a_4, a_5\}$ соответствует G, при этом $A = \{a_5\}$. Если

вершина v 2-пуста в G, то удалим вершину b_2 и соединим с v все вершины, смежные с b_2 в G. Ясно, что $G' \cong G/b_2v$, а значит, $G' \in \mathcal{MOP}$ по лемме 2. Кроме того, G' соответствует G (здесь $A = \varnothing$), что и требовалось.

При рассмотрении случаев 2 и 3 будем предполагать, что $\deg(x_4) = 2$ и грань f_5 содержит ребро c_1b_2. Обозначим через c_2 третью вершину f_5. Считаем, что вершина c_1 ни 2-универсальна, ни 2-пуста, поскольку если она 2-универсальна, то граф $G\backslash\{a_4, a_5\}$ соответствует G (здесь $A = \{a_5\}$), а если эта вершина 2-пуста, то граф G', полученный из графа $G \setminus \{b_1\}$ добавлением рёбер c_1a_1, c_1a_2, c_1a_3 соответствует G (здесь $A = \varnothing$).

Случай 2: $\deg(x_5) \leqslant 2$.

Вариант 2а: $\deg(x_5) = 1$. Вершина c_2 2-универсальна, тогда смежная с ней вершина c_1 2-пуста; противоречие.

Вариант 2в: $\deg(x_5) = 2$ и $\deg(c_1) = 3$. Вершина c_1 2-универсальна, так как она смежна с двумя 2-пустыми вершинами b_1 и b_2; противоречие.

Вариант 2с: $\deg(x_5) = 2$ и $\deg(c_1) \geqslant 4$. Граф $G \setminus \{a_1, \ldots, a_5, b_1, b_2\} \in \mathcal{MOP}$ соответствует G (здесь $A = \{a_1, a_3, a_5\}$).

Случай 3: $\deg(x_5) = 3$. Обозначим через f_4' грань, смежную с f_5 и отличную от f_4 и f_6.

Вариант 3а: грань f_4' содержит ребро c_1c_2. Обозначим через d_1 третью вершину f_4'. Если $\deg(x_4') = 1$, то вершина d_1 2-универсальна и вершина c_1 2-пуста; противоречие. Если $\deg(x_4') \geqslant 2$, то все грани, смежные с f_4' и отличные от f_5, будут $\{1, 2, 3\}$-крайними. По предположению вершина c_1 не принадлежит ни одной такой грани. Тогда $\deg(c_1) = 4$, причём вершины b_1 и b_2, смежные с c_1, 2-пусты, а оставшиеся два соседа c_1 смежны, а тогда c_1 2-универсальна; противоречие.

Вариант 3в: грань f_4' содержит ребро b_2c_2. Обозначим через v вершину, отличную от a_5 и такую, что b_2v является внешним ребром в G. Аналогично случаю 1 вершина v единственна, принадлежит некоторой $\{1, 2, 3\}$-внешней грани и, следовательно, либо 2-универсальна, либо 2-пуста по лемме 5. Если v 2-универсальна, то граф $G \setminus \{a_4, a_5\}$ соответствует G (здесь $A = \{a_5\}$). Если же v 2-пуста, то удалим вершину b_2 и соединим с вершиной v все вершины, смежные с b_2 в G. Тогда для полученного графа G' верно $G' \cong G/b_2v \in \mathcal{MOP}$. Кроме того, G' соответствует G (здесь $A = \varnothing$), что и требовалось. Лемма 6 доказана.

Лемма 7. *Если* $\deg(x_3) = 3$, *то граф* G *не критический.*

Доказательство. Обозначим через $a_1a_2b_1$ грань f_1, через $a_2b_1b_2$ грань f_2, через $b_1b_2c_2$ грань f_3. Поскольку $\deg(x_3) = 3$, у вершин b_1 и c_2 найдётся единственный сосед c_1, отличный от b_2, а у вершин b_2 и c_2 найдётся единственный сосед c_3, отличный от b_1 (поскольку $G \in \mathcal{OP}$, вершины c_1 и c_3 не совпадают). Отметим, что вершины a_1 и b_2 2-универсальны, а тогда смежные с ними вершины a_2, b_1, c_2 и c_3 2-пусты.

Обозначим через f_2' грань, смежную с f_3 и отличную от f_2 и f_4. По лемме 4 грань f_2' либо 1-крайняя, либо смежна с единственной 1-крайней гранью f_1'. В зависимости от расположения грани f_2' возможны два случая.

Случай 1: грань f_2' является треугольником $b_2c_2c_3$. Если $\deg(x_2') = 1$, то вершина c_3 2-универсальна и смежна с b_2, что невозможно. Если же $\deg(x_2') = 2$, то либо b_2 принадлежит грани f_1' и смежна с 2-универсальной вершиной степени 2, либо G особый, что также невозможно.

Случай 2: грань f_2' является треугольником $b_1c_1c_2$. Если $\deg(x_2') = 1$, то вершина c_1 универсальна и смежна с b_1, тогда граф $G \setminus \{a_1, a_2\}$ соответствует G (здесь $A = \{a_1\}$). Если же $\deg(x_2') = 2$, то вне зависимости от расположения грани f_1' вершина c_1 имеет степень 3 и смежна с 2-пустыми вершинами b_1 и c_2, а также с 2-универсальной вершиной степени 2 грани f_1', что невозможно. Лемма 7 доказана.

При доказательстве лемм 8 и 9 предполагаем, что $\deg(x_3) = 2$ и попрежнему обозначаем грани f_1, f_2 и f_3 через $a_1a_2b_1$, $a_2b_1b_2$ и $b_1b_2c_2$ соответственно.

Лемма 8. *Если* $\deg(x_4) = 3$, *то граф* G *не критический.*

Доказательство. Если грань f_4 не содержит b_1, то $\deg(b_1) = 4$ и, как нетрудно проверить, граф $G_3 = G\setminus\{a_1, a_2, b_1\}$ соответствует G (здесь $A = \{a_1\}$), поэтому предполагаем, что $\deg(b_1) \geqslant 5$.

Обозначим через f_4 грань $b_1c_1c_2$, а через f_3' — грань, смежную с f_4 и отличную от граней f_3 и f_5. Если $\deg(x_3') = 3$, то грань f_3' либо смежна с двумя 1-крайними гранями, что невозможно по лемме 4, либо смежна хотя бы с одной 2-крайней гранью, что невозможно по лемме 7. Таким образом, предполагаем, что $\deg(x_3') \in \{1, 2\}$. Если $\deg(x_3') = 2$, то обозначим через f_2' грань, смежную с f_3' и отличную от f_4. Если при этом $\deg(x_2') = 2$, то обозначим через f_1' грань, смежную с f_2' и отличную от f_3' (случай $\deg(x_2') = 3$ невозможен по лемме 4).

Если хотя бы одна вершина, смежная с b_1 и отличная от вершин a_1, b_2, 2-универсальна, то граф $G_2 = G\setminus\{a_1, a_2\}$ соответствует G, поэтому предполагаем, что все такие вершины, включая c_1, не 2-универсальны. При этом c_1 принадлежит $\{1, 2, 3\}$-крайней грани f_3', а тогда по лемме 5 она 2-пуста. Далее возможны два случая в зависимости от расположения грани f_3'.

Случай 1. Грань f_3' содержит ребро b_1c_1. Обозначим через b_0 третью вершину f_3'.

Вариант 1а: $\deg(x_3') = 1$. Поскольку $\deg(b_0) = 2$, вершина b_0 универсальна и смежна с b_1, что противоречит предположению.

Вариант 1в: $\deg(x'_3) = 2$ и $\deg(x'_2) = 1$. В этом случае вершина b_0 степени 3 смежна с двумя 2-пустыми вершинами b_1 и c_1, а также с 2-универсальной вершиной степени 2 грани f'_2, что невозможно.

Вариант 1с: $\deg(x'_3) = \deg(x'_2) = 2$. Предполагаем, что вершина b_0 не 2-универсальна, тогда по лемме 5 она 2-пуста. Легко проверить, что при любом из четырёх вариантов расположения граней f'_2 и f'_1 хотя бы $\deg(b_0) - 1$ соседей b_0 также 2-пусты, что невозможно.

Случай 2. Грань f'_3 содержит ребро c_1c_2. Обозначим через d_1 третью вершину f'_3.

Вариант 2а: $\deg(x'_3) = 1$. Рассмотрим граф $G'_2 \in \mathcal{MOP}$, полученный из графа $G \setminus \{c_2, d_1\}$ добавлением ребра b_2c_1. Поскольку c_1 2-пуста в G, полученный граф G'_2 соответствует G (здесь $A = \{d_1\}$).

Вариант 2в: $\deg(x'_3) = 2$ и $\deg(x'_2) = 1$. Обозначим через d_2 вершину грани f'_2, отличную от c_1, c_2 и d_1. Если d_2 смежна с c_2, то вершина d_1 степени 3 смежна с двумя 2-пустыми вершинами и одной 2-универсальной вершиной d_2, что невозможно. Если же d_2 смежна с c_1, то вершина c_2 степени 4 смежна с тремя 2-пустыми вершинами b_1, c_1, d_1 и одной 2-универсальной вершиной b_2, что невозможно.

Вариант 2с: $\deg(x'_3) = \deg(x'_2) = 2$. Обозначим через d_3 вершину степени 2 грани f'_1. Возможны четыре варианта расположения граней f'_2 и f'_1, изображённые на рис. 2. Если d_1 принадлежит f'_1, то она 2-пуста и все смежные с ней вершины, кроме одной, 2-пусты, что невозможно. Если же d_1 не принадлежит f'_1, то подграф G'_2, полученный из графа $G \setminus \{d_2, d_3\}$ заменой ребра b_1c_2 ребром c_1b_2, соответствует G (здесь $A = \{d_3\}$). Лемма 8 доказана.

Лемма 9. *Если* $\deg(x_4) = 2$, *то граф* G *не критический.*

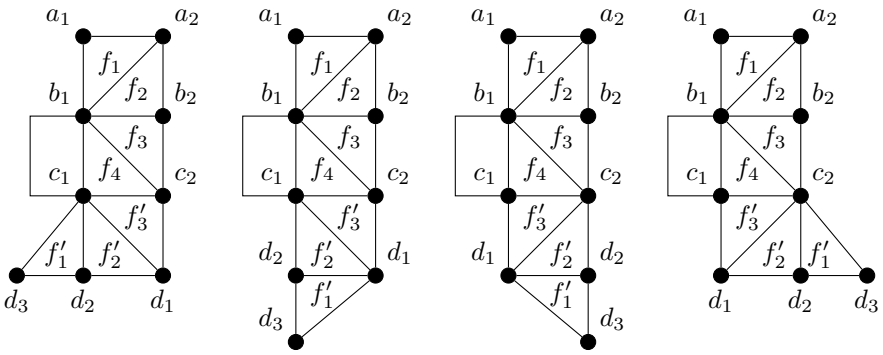

Рис. 2. Возможные конфигурации в варианте 2с леммы 8

Доказательство. Как и в доказательстве леммы 8, предполагаем, что $\deg(b_1) \geqslant 5$, иначе граф $G \setminus \{a_1, a_2, b_1\}$ соответствует G. Обозначим грань f_4 через $b_1c_1c_2$.

Случай 1: $\deg(b_1) \geqslant 6$. В этом случае $\deg(b_2) = \deg(c_2) = 3$. Вершина c_2 2-пуста и смежна с 2-пустой вершиной b_1, а также с 2-универсальной вершиной b_2. Значит, смежная с ней вершина c_1 2-универсальна и граф $G \setminus \{a_1, a_2\}$ соответствует G.

При рассмотрении случаев 2–4 предполагаем, что грань f_5 содержит ребро c_1c_2, тогда $\deg(b_1) = 5$ и $\deg(b_2) = 3$. Обозначим через d_2 третью вершину f_5.

Случай 2: $\deg(x_5) \leqslant 2$. Если $\deg(x_5) = 1$, то вершина c_1 смежна с одной 2-универсальной и двумя 2-пустыми вершинами, что невозможно. Если $\deg(x_5) = 2$ и грань f_6 не содержит c_2, то поскольку вершины b_1 и c_2 пусты, граф $G \setminus \{a_1, a_2, b_1, b_2, c_2\} \in \mathcal{MOP}$ соответствует G (здесь $A = \{a_1, b_2\}$). Если же $\deg(x_5) = 2$ и грань f_6 не содержит c_1, то вершина c_1 имеет степень 3 и смежна с 2-пустыми вершинами b_1 и c_2, тогда она 2-универсальна и граф $G \setminus \{a_1, a_2\}$ соответствует G.

Случай 3: $\deg(x_5) = 3$, грань f_4' содержит ребро c_1d_2. Обозначим через d_1 третью вершину f_4'. Если вершина c_1 2-универсальна, то граф $G \setminus \{a_1, a_2\}$ соответствует G. Если вершина c_1 2-пуста, то граф $G_3' \in \mathcal{MOP}$, полученный добавлением в граф $G \setminus \{a_1, a_2, b_1\}$ ребра c_1b_2, соответствует G (здесь $A = \{a_1\}$). Если вершина c_1 не универсальна и не пуста, то она смежна хотя бы с двумя вершинами, которые не смежны между собой и не 2-пусты. Поскольку вершины b_1 и c_2 пусты, то $\deg(c_1) \geqslant 5$. Значит, c_1 принадлежит некоторой грани f_3', которая является $\{1, 2, 3\}$-крайней; по лемме 5 получили противоречие.

Случай 4: $\deg(x_5) = 3$, грань f_4' содержит ребро c_2d_2. Обозначим через d_3 третью вершину f_4'. Обозначим через v вершину, смежную с c_2, отличную от b_2 и такую, что ребро c_2v принадлежит внешней грани G (поскольку вершина c_2 инцидентна двум внешним рёбрам G, вершина v единственна). Из рассуждений случая 1 леммы 6 следует, что v принадлежит некоторой $\{1, 2, 3\}$-крайней грани G и либо 2-универсальна, либо 2-пуста. Если v 2-универсальна, то граф $G_2' \in \mathcal{MOP}$, полученный из графа $G \setminus \{a_1, a_2\}$ заменой ребра b_1c_1 ребром b_1v, соответствует G (здесь $A = \{a_1\}$). Если же v 2-пуста, то граф $G_3' \in \mathcal{MOP}$, полученный из графа $G \setminus \{a_1, a_2, b_1\}$ добавлением ребра b_2v, соответствует G (здесь также $A = \{a_1\}$). Лемма 9 доказана.

Теорема 4. *Для всех* $n \geqslant 6$ *верно* $\mathrm{mi}_2(n, \mathcal{MOP}) = 1$.

Доказательство. Из лемм 4–9 следует, что критических графов не существует, тем самым $\mathrm{mi}_2(n, \mathcal{MOP}) \leqslant 1$. Докажем, что при всех

$n \geqslant 6$ найдётся максимальный внешнепланарный граф с единственным 2-ДНМ.

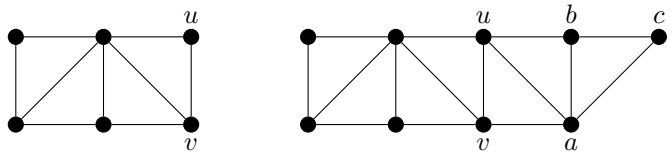

Рис. 3. Граф W'_6 и полученный из него граф G_9

Обозначим через W'_{2k} $2k$-вершинный внешнепланарный граф, полученный из пути P_{2k-1} добавлением новой вершины степени $2k-1$. При всех $2k \geqslant 4$ верно $\mathrm{mi}_2(W'_{2k}) = \mathrm{mi}_2(P_{2k-1}) = 1$. Теперь для всех $n \geqslant 4$ и графа $G_n \in \mathcal{MOP}$ с единственным 2-ДНМ построим граф $G_{n+3} \in \mathcal{MOP}$ с единственным 2-ДНМ. Выберем в G_n некоторую 2-универсальную вершину u и смежную с ней вершину v. Обозначим через G_{n+3} результат добавления в G вершин a, b, c и рёбер ab, bc, ac, au, bu, av (рис. 3). Ясно, что $G_{n+3} \in \mathcal{MOP}$ и для 2-ДНМ J графа G множество $J \cup \{c\}$ является 2-ДНМ графа G_{n+3}, что и требовалось. Теорема 4 доказана.

4. Планарные графы

Теорема 5. *Для всех $n \geqslant 1$ и $k \geqslant 4$ верно $\mathrm{mi}_k(n, \mathcal{P}) = 1$.*

Доказательство. При любом $n \geqslant 1$ пустой граф nK_1 содержит единственное k-ДНМ, откуда $\mathrm{mi}_k(n, \mathcal{P}) \geqslant 1$. Предположим, что при некотором $k \geqslant 4$ неравенство $\mathrm{mi}_k(n, \mathcal{P}) \leqslant 1$ не выполнено, и рассмотрим минимальный по числу вершин граф $G \in \mathcal{P}$, содержащий хотя бы два различных k-ДНМ J_1 и J_2. Покажем, что имеют место следующие свойства.

Свойство 1: $J_1 \cup J_2 = V(G)$. Если это не так, то в подграфе G, порождённом множеством $J_1 \cup J_2$ и содержащем менее $|V(G)|$ вершин, как J_1, так и J_2 являются k-ДНМ, что противоречит предположению о минимальности G.

Свойство 2: $J_1 \cap J_2 = \varnothing$. Если это не так, то для некоторой вершины $v \in V(G)$ верно $v \in J_1 \cap J_2$. Тогда вершина v изолированная в графе G, поскольку она не смежна ни с одной из вершин множества $J_1 \cup J_2$, а множество $V(G) \setminus (J_1 \cup J_2)$ пусто по предыдущему свойству. Следовательно, граф $G \setminus \{v\}$ содержит два различных k-ДНМ $J_1 \setminus v$ и $J_2 \setminus v$; снова получили противоречие с минимальностью G.

Таким образом, G является двудольным графом с долями J_1 и J_2. При этом $\delta(G) \geqslant k \geqslant 4$, поскольку G не содержит k-универсальных вершин, но тогда $G \notin \mathcal{P}$ по формуле Эйлера; противоречие. Теорема 5 доказана.

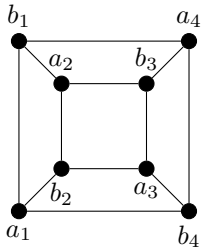

Рис. 4. Граф Q_3 и его 3-ДНМ $\{a_1, a_2, a_3, a_4\}$ и $\{b_1, b_2, b_3, b_4\}$

На рис. 4 изображён граф трёхмерного куба Q_3. Легко проверить, что $\mathrm{mi}_3(Q_3) = 2$. Поскольку при любых $m, r \geqslant 0$ для $n = 8m + r$ верно $\mathrm{mi}_3(mQ_3 \cup rK_1) = 2^m$, то $\mathrm{mi}_3(n, \mathcal{PL}) \geqslant 2^{[n/8]}$. Из теоремы 2 следует неравенство $\mathrm{mi}_2(n, \mathcal{PL}) \geqslant 2^{[n/4]}$. Вопрос о том, точны ли данные оценки, остаётся открытым.

Цель оставшейся части раздела — для каждого $m \geqslant 1$ построить максимальный планарный граф, содержащий подграф mQ_3 и 2^m 3-ДНМ. Сначала покажем, что в каждый максимальный планарный граф можно добавить 5 вершин, не уменьшая числа 3-ДНМ в нём.

Лемма 10. *Для любого n-вершинного графа $G \in \mathcal{MP}$ существует $(n+5)$-вершинный граф $G' \in \mathcal{MP}$ такой, что* $\mathrm{mi}_3(G) \leqslant \mathrm{mi}_3(G')$.

Доказательство. Если $\mathrm{mi}_3(G) = 0$, то доказывать нечего. Предположим, что $\mathrm{mi}_3(G) \geqslant 1$. Выберем вершину $w \in V(G)$ степени $\delta(G) \leqslant 5$. Поскольку $G \in \mathcal{MP}$ и все грани G являются треугольниками, в нём найдётся простой цикл, на котором лежат все вершины открытой окрестности $N(w)$. Значит, не более двух вершин $N(w)$ могут одновременно входить в 3-ДНМ G, тем самым w 3-универсальна в G.

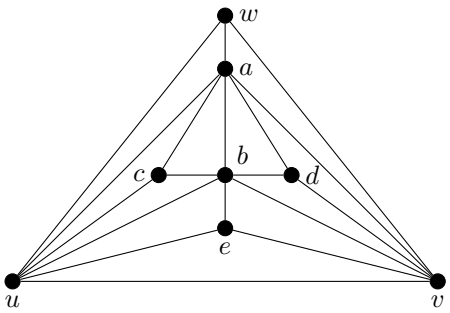

Рис. 5. Добавление 5 вершин в граф G

Рассмотрим одну из треугольных граней, содержащих w, и обозначим через u и v две другие вершины этой грани, они 3-пусты. Добавим вершину a в грань uvw, после этого добавим вершину b в грань uav и вершины c, d, e в грани uab, abv, uvb соответственно (рис. 5). Полученный граф обозначим через G'. Поскольку для любого 3-ДНМ J графа G множество $J \cup \{c, d, e\}$ является 3-ДНМ графа G', то $\mathrm{mi}_3(G) \leqslant \mathrm{mi}_3(G')$, что и требовалось. Лемма 10 доказана.

Теорема 6. *Для всех* $n \geqslant 12$ *верно* $\mathrm{mi}_3(n, \mathcal{MP}) \geqslant 2^{[n/50]-1}$. *При этом* $\mathrm{mi}_3(n, \mathcal{MP}) \geqslant 2^{[n/50]}$, *если* $n = 50m + r$, *где* $m \geqslant 0$ *и* $r \in \{0, 2\} \cup \{4, 5, 6, \ldots, 49\}$.

Доказательство. Сначала покажем, что при всех $n \geqslant 12$ существует n-вершинный граф $G_n \in \mathcal{MP}$, содержащий хотя бы одно 3-ДНМ. Напомним, что через W_n обозначается n-вершинный граф, полученный путём добавления в цикл C_{n-1} новой вершины, смежной со всеми вершинами цикла. Для всех $s \geqslant 4$ построим граф G_{2s}, полученный из графа W_s добавлением вершины степени 3 в каждую из его $s-1$ треугольных граней и добавлением вершины степени $s-1$ во внешнюю грань. Ясно, что граф $G_{2s} \in \mathcal{MP}$ определён однозначно и его единственное 3-ДНМ содержит в точности те вершины, которые были добавлены в W_s (поскольку вершины степени 3 3-универсальны в G_{2s}). Таким образом, искомый граф G_n построен для всех чётных $n \geqslant 8$. По лемме 10 граф G_n существует для всех нечётных $n \geqslant 13$, что и требовалось.

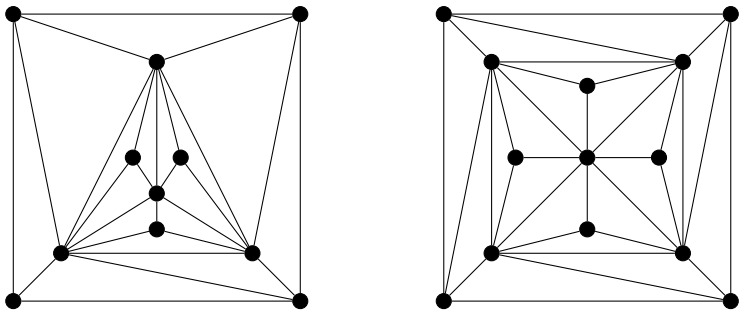

Рис. 6. Графы H и H'

Для всех $m \geqslant 1$ и $0 \leqslant s \leqslant 4$ построим непустой класс $\mathcal{G}_{m,s}$ максимальных планарных графов, содержащих $50m+2s$ вершин и 2^m 3-ДНМ. Класс $\mathcal{G}_{1,s}$ состоит в точности из тех $(50 + 2s)$-вершинных графов, которые могут быть получены из графа Q_3 в результате замены $6 - s$ его граней 11-вершинным графом H и s оставшихся граней — 13-вершинным графом H' (рис. 6). Здесь под заменой грани подграфом понимаем замену вершин грани вершинами внешней грани подграфа и добавление

в граф внутренних вершин подграфа вместе со всеми инцидентными им рёбрами. Несмотря на то, что результат такой замены не определён однозначно, все добавленные вершины, не принадлежавшие внешним граням, являются либо 3-универсальными степени 3, либо смежными с ними 3-пустыми. Таким образом, для любого графа $G_{1,s} \in \mathcal{G}_{1,s}$ верно $\mathrm{mi}_3(G_{1,s}) = \mathrm{mi}_3(Q_3) = 2$.

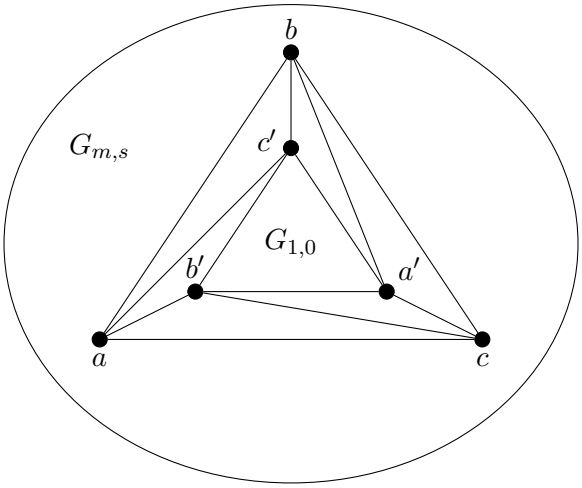

Рис. 7. Структура графа $G_{m+1,s}$

Рассмотрим граф $G'_{m+1,s} = G_{m,s} \cup G_{1,0}$, где $G_{m,s} \in \mathcal{G}_{m,s}$ и $G_{1,0} \in \mathcal{G}_{1,0}$. Можно считать, что в подграфе $G_{m,s}$ найдётся треугольник abc с 3-универсальной вершиной a и пустой внутренностью, а внешней гранью подграфа $G_{1,0}$ является треугольник $a'b'c'$ с 3-универсальной вершиной a'. Добавим в граф $G'_{m+1,s}$ рёбра ab', ac', ba', bc', ca', cb' и обозначим получившийся граф через $G_{m+1,s}$ (рис. 7). По построению $G_{m+1,s} \in \mathcal{MP}$. Поскольку $G_{m+1,s}$ получен из $G'_{m+1,s}$ соединением рёбрами нескольких пар 3-пустых вершин, $\mathrm{mi}_3(G_{m+1,s}) = \mathrm{mi}_3(G'_{m+1,s}) = 2^{m+1}$. Для всех $m \geqslant 1$ определим класс $\mathcal{G}_{m+1,s}$ как совокупность графов $G_{m+1,s}$, которые могут быть получены применением описанной процедуры.

Пусть $n = 50m+r$, где $m \geqslant 0$ и $0 \leqslant r \leqslant 49$. Если $r \notin \{1, 3\}$, то найдётся целое число $0 \leqslant s \leqslant 4$ такое, что $r - 2s = 5p \geqslant 0$. По лемме 10 найдётся n-вершинный граф G, содержащий не менее 2^m k-ДНМ (при $p = 0$ подойдёт любой граф из класса $\mathcal{G}_{m,s}$). Если же $r \in \{1, 3\}$, то найдётся целое число $0 \leqslant s \leqslant 4$ такое, что $50 + r - 2s = 5q$, а тогда по лемме 10 найдётся n-вершинный граф G, содержащий не менее 2^{m-1} k-ДНМ. Теорема 6 доказана.

Финансирование работы

Исследование выполнено в Санкт-Петербургском международном математическом институте им. Леонарда Эйлера при финансовой поддержке Министерства науки и высшего образования Российской Федерации (соглашение № 075–15–2022–287).

Конфликт интересов

Автор заявляет, что у него нет конфликта интересов.

Литература

1. **Nagy Z. L.** On the number of k-dominating independent sets // J. Graph Theory. 2017. V. 84, No. 4. P. 566–580.
2. **Gerbner D., Keszegh B., Methuku A., Patkós B., Vizer M.** An improvement on the maximum number of k-dominating independent sets // J. Graph Theory. 2019. V. 91, No. 1. P. 88–97.
3. **Moon J., Moser L.** On cliques in graphs // Israel J. Math. 1965. V. 3, No. 1. P. 23–28.
4. **Wood D. R.** On the number of maximal independent sets in a graph // Discrete Math. Theor. Comput. Sci. 2011. V. 13, No. 3. P. 17–19.
5. **Griggs J., Grinstead C., Guichard D.** The number of maximal independent sets in a connected graph // Discrete Math. 1988. V. 68. P. 211–220.
6. **Wilf H.** The number of maximal independent sets in a tree // SIAM J. Algebr. Discrete Methods. 1986. V. 7, No. 1. P. 125–130.
7. **Liu J.** Maximal independent sets in bipartite graphs // J. Graph Theory. 1993. V. 17, No. 4. P. 495–507.
8. **Hujter M., Tuza Z.** The number of maximal independent sets in triangle-free graphs // SIAM J. Discrete Math. 1993. V. 6, No. 2. P. 284–288.
9. **Jou M. J., Chang G.** Maximal independent sets in graphs with at most one cycle // Discrete Appl. Math. 1997. V. 79. P. 67–73.
10. **Ying G. C., Meng K. K., Sagan B. E., Vatter V. E.** Maximal independent sets in graphs with at most r cycles // J. Graph Theory. 2006. V. 53, No. 4. P. 270–282.
11. **Koh K. M., Goh C. Y., Dong F. M.** The maximum number of maximal independent sets in unicyclic connected graphs // Discrete Math. 2008. V. 308, No. 17. P. 3761–3769.
12. **Włoch A.** On 2-dominating kernels of graphs // Australas. J. Comb. 2012. V. 52. P. 273–284.
13. **Bednarz P., Włoch I.** On (2-d)-kernels in the Cartesian product of graphs // Ann. Univ. Mariae Curie-Skłodowska. Sect. A. 2016. V. 70. P. 1–8.
14. **Bednarz P.** On (2-d)-kernels in the tensor product of graphs // Symmetry. 2021. V. 13. Paper ID 230. 9 p.

15. Bednarz P. On (2-d)-kernels in two generalizations of the Petersen graph // Symmetry. 2021. V. 13. Paper ID 1948. 10 p.

Талецкий Дмитрий Сергеевич

Статья поступила
4 мая 2023 г.
После доработки —
3 июля 2023 г.
Принята к публикации
22 сентября 2023 г.

DISKRETNYI ANALIZ I ISSLEDOVANIE OPERATSII
/DISCRETE ANALYSIS AND OPERATIONS RESEARCH/
January–March 2024. Vol. 31, No. 1. P. 109–128

UDC 519.17 DOI: 10.33048/daio.2024.31.770

ON THE NUMBER OF k-DOMINATING INDEPENDENT SETS IN PLANAR GRAPHS

D. S. Taletskii[1, 2]

[1] National Research University "Higher School of Economics",
25/12 Bolshaya Pechyorskaya Street, 603155 Nizhny Novgorod, Russia
[2] Saint Petersburg State University,
7/9 Universitetskaya Embankment, 199034 Saint Petersburg, Russia
E-mail: dmitalmail@gmail.com

Abstract. A set J_k of graph vertices is said to be k-dominating independent ($k \geqslant 1$) if its vertices are pairwise adjacent and every vertex not in J_k is adjacent to at least k vertices in J_k. In the present paper, we obtain new upper bounds for the number of k-dominating independent sets for $k \geqslant 2$ in some planar graph classes. Illustr. 7, bibliogr. 15.

Keywords: independent set, dominating set, k-dominating independent set, planar graph.

References

1. **Z. L. Nagy,** On the number of k-dominating independent sets, *J. Graph Theory* **84** (4), 566–580 (2017).
2. **D. Gerbner, B. Keszegh, A. Methuku, B. Patkós,** and **M. Vizer,** An improvement on the maximum number of k-dominating independent sets, *J. Graph Theory* **91** (1), 88–97 (2019).
3. **J. Moon** and **L. Moser,** On cliques in graphs, *Israel J. Math.* **3** (1), 23–28 (1965).
4. **D. R. Wood,** On the number of maximal independent sets in a graph, *Discrete Math. Theor. Comput. Sci.* **13** (3), 17–19 (2011).
5. **J. Griggs, C. Grinstead,** and **D. Guichard,** The number of maximal independent sets in a connected graph, *Discrete Math.* **68**, 211–220 (1988).
6. **H. Wilf,** The number of maximal independent sets in a tree, *SIAM J. Algebr. Discrete Methods* **7** (1), 125–130 (1986).
7. **J. Liu,** Maximal independent sets in bipartite graphs, *J. Graph Theory* **17** (4), 495–507 (1993).

English transl.: *Journal of Applied and Industrial Mathematics* **18** (1), 167–178 (2024), DOI: 10.1134/S1990478924010149.

8. **M. Hujter** and **Z. Tuza,** The number of maximal independent sets in triangle-free graphs, *SIAM J. Discrete Math.* **6** (2), 284–288 (1993).
9. **M. J. Jou** and **G. Chang,** Maximal independent sets in graphs with at most one cycle, *Discrete Appl. Math.* **79**, 67–73 (1997).
10. **G. C. Ying, K. K. Meng, B. E. Sagan,** and **V. E. Vatter,** Maximal independent sets in graphs with at most r cycles, *J. Graph Theory* **53** (4), 270–282 (2006).
11. **K. M. Koh, C. Y. Goh,** and **F. M. Dong,** The maximum number of maximal independent sets in unicyclic connected graphs, *Discrete Math.* **308** (17), 3761–3769 (2008).
12. **A. Włoch,** On 2-dominating kernels of graphs, *Australas. J. Comb.* **52**, 273–284 (2012).
13. **P. Bednarz** and **I. Włoch,** On (2-d)-kernels in the Cartesian product of graphs, *Ann. Univ. Mariae Curie-Skłodowska, Sect. A,* **70**, 1–8 (2016).
14. **P. Bednarz,** On (2-d)-kernels in the tensor product of graphs, *Symmetry* **13**, ID 230 (2021).
15. **P. Bednarz,** On (2-d)-kernels in two generalizations of the Petersen graph, *Symmetry* **13**, ID 1948 (2021).

Dmitrii S. Taletskii

Received May 4, 2023
Revised July 3, 2023
Accepted September 22, 2023

ДИСКРЕТНЫЙ АНАЛИЗ
И ИССЛЕДОВАНИЕ ОПЕРАЦИЙ

2024. Том 31, № 1

Зав. редакцией Ю. В. Шамардин

Журнал подготовлен с использованием макропакета LaTeX 2_ε.
The present publication has been typeset using LaTeX 2_ε.

Журнал зарегистрирован в Федеральной службе по надзору
в сфере связи, информационных технологий и массовых коммуникаций.
Свидетельство о регистрации ЭЛ № ФС77-85978 от 26.09.2023 г.
Размещение в сети Интернет: `math-sobolev.ru`.

Дата размещения в сети Интернет 10.04.2024 г.
Формат $70 \times 100\,1/16$. Усл. печ. л. 10,4. Объём 1,43 МБ.

Издательство Института математики,
пр. Академика Коптюга, 4, 630090 Новосибирск, Россия