МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

ПРИКЛАДНАЯ ДИСКРЕТНАЯ МАТЕМАТИКА

Научный журнал

2019 № 45

Зарегистрирован в Федеральной службе по надзору
в сфере связи и массовых коммуникаций

Свидетельство о регистрации ПИ № ФС 77-33762 от 16 октября 2008 г.

Подписной индекс в объединённом каталоге «Пресса России» 38696

УЧРЕДИТЕЛЬ
Томский государственный университет

Адрес редакции и издателя: 634050, г. Томск, пр. Ленина, 36
E-mail: vestnik_pdm@mail.tsu.ru

В журнале публикуются результаты фундаментальных и прикладных научных исследований отечественных и зарубежных ученых, включая студентов и аспирантов, в области дискретной математики и её приложений в криптографии, компьютерной безопасности, кибернетике, информатике, программировании, теории надёжности, интеллектуальных системах.

Периодичность выхода журнала: 4 номера в год.

Редактор *Н. И. Шидловская*
Верстка *И. А. Панкратовой*

Подписано к печати 16.09.2019. Формат $60 \times 84\frac{1}{8}$. Усл. п. л. 14,78. Тираж 300 экз.
Заказ № 3947. Цена свободная. Дата выхода в свет 20.09.2019.

Отпечатано на оборудовании
Издательского Дома Томского государственного университета
634050, г. Томск, пр. Ленина, 36
Тел.: 8(3822)53-15-28, 52-98-49

СОДЕРЖАНИЕ

CONTENTS

ПАМЯТИ ВАЛЕНТИНЫ ВЛАДИМИРОВНЫ БЫКОВОЙ

К величайшему сожалению, 18 июля с.г. на 72-м году жизни скончалась Валентина Владимировна Быкова — выпускница механико-математического факультета ТГУ, крупный учёный в области прикладной дискретной математики и информатики, доктор физико-математических наук, профессор Сибирского федерального университета. Отечественные наука и высшее образование понесли невосполнимую утрату. Ушла из жизни выдающийся Педагог и настоящий Учёный, чьи преподавательские и интеллектуальные способности, чьи чисто человеческие качества высочайшей чистоты и преданности делу стали для нас и наших учеников неисчерпаемым источником и образцом настоящей математической и образовательной культуры. От этого чистого истока под её председательством в ГАК начали свой профессиональный путь последние выпускники кафедры защиты информации и криптографии ТГУ. Развитие Валентиной Владимировной теоретических основ анализа и синтеза параметризированных алгоритмов сделало её ведущим специалистом с мировым именем. Редакция журнала «Прикладная дискретная математика», одним из наиболее полезных членов которой она была, Оргкомитет Всероссийской научной конференции Sibecrypt по компьютерной безопасности и криптографии, в составе которого она была одним из наиболее ценных организаторов, лаборатория Компьютерной криптографии Томского государственного университета, которая в дни своего зарождения получила её почти божественное благословение и которая благотворно вдохновляется теперь её научными трудами, горько скорбят в связи с её утратой. Мы приносим глубочайшие соболезнования всем нашим коллегам, знавшим Валентину Владимировну и сотрудничавшим с ней в науке и образовании, и вместе с ними разделяем огромное горе её родных и близких. Да будет вечной и светлой память о Вас, дорогая Валентина Владимировна!

19.07.2019 Г. П. Агибалов, И. А. Панкратова

ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ПРИКЛАДНОЙ ДИСКРЕТНОЙ МАТЕМАТИКИ

УДК 510.665

РАЗРЕШИМОСТЬ ОГРАНИЧЕННЫХ ТЕОРИЙ КЛАССА ЧАСТИЧНЫХ ПОРЯДКОВ[1]

А. Ю. Никитин

Институт математики им. С. Л. Соболева СО РАН, г. Омск, Россия

Классическая алгебраическая геометрия изучает множества решений алгебраических уравнений над полями вещественных и комплексных чисел. В последние 20 лет активно развивается так называемая универсальная алгебраическая геометрия, в которой изучаются системы уравнений над произвольными алгебраическими системами. При этом особое значение имеют универсальные и экзистенциальные теории — от их сложности зависят перспективы построения «хорошей» алгебраической геометрии над той или иной алгебраической системой. В работе доказывается, что экзистенциальная и универсальная теории класса всех частичных порядков являются разрешимыми.

Ключевые слова: *частично упорядоченное множество, частичные порядки, разрешимость универсальной теории, разрешимость экзистенциальной теории, классы.*

DOI 10.17223/20710410/45/1

DECIDABILITY OF THE RESTRICTED THEORIES OF A CLASS OF PARTIAL ORDERS

A. Yu. Nikitin

Sobolev Institute of Mathematics, Omsk, Russia

E-mail: nikitinlexey@gmail.com

Classical algebraic geometry studies the solution sets of algebraic equations over the fields of real and complex numbers. In the past 20 years, the so-called universal algebraic geometry, which studies systems of equations over arbitrary algebraic structures, has been actively developed. In this frameworks, universal and existential theories are very important, the prospect for constructing good algebraic geometry over algebraic systems depends on their complexity. In this paper, we prove that the existential and universal theories of the class of all finite orders are decidable.

Keywords: *partially ordered set, poset, decidability of univarsal theory, decidability of existential theory, classes.*

[1]Работа поддержана грантом РНФ № 18-71-10028.

Введение

Решение уравнений и систем уравнений над вещественными, комплексными, рациональными, целыми числами является классической темой исследований в различных областях математики в течение многих сотен лет. Классическая алгебраическая геометрия изучает множества решений алгебраических уравнений над полями вещественных и комплексных чисел. В рамках диофантовой геометрии и диофантова анализа изучаются решения алгебраических уравнений над целыми и рациональными числами. В последние 20 лет активно развивается так называемая универсальная алгебраическая геометрия [1], в которой изучаются системы уравнений над произвольными алгебраическими системами. Под уравнениями понимаются атомарные формулы языка алгебраической системы. Многие понятия и алгоритмы классической алгебраической геометрии переносятся на произвольные алгебраические системы. При этом особую роль играют универсальные и экзистенциальные теории — от их сложности зависят перспективы построения «хорошей» алгебраической геометрии над той или иной алгебраической системой. Как правило, эти теории оказываются либо неразрешимыми, либо разрешимыми, но вычислительно трудными. Например, экзистенциальная теория кольца целых чисел неразрешима [2]. Вопрос о разрешимости экзистенциальной теории поля рациональных чисел до сих пор открыт. Известные на сегодняшний день алгоритмы для экзистенциальной теории полей комплексных и вещественных чисел имеют дважды экспоненциальную сложность [3]. Экзистенциальная теория любого конечного поля, эквивалентная проблеме выполнимости булевых формул, является NP-полной [4].

В XX веке, в связи с бурным развитием компьютерной техники и прикладной математики, на первый план вышли исследования различных конечных комбинаторных и алгебраических объектов. Прежде всего, это конечные графы, конечные поля, конечные порядки (частично упорядоченные множества). Классическими подходами к изучению конечных алгебраических систем являются алгебраический и комбинаторный. Новый подход к изучению этих объектов — логический и теоретико-модельный — родился в рамках универсальной алгебраической геометрии [1]. Многие практически важные задачи о конечных графах, конечных полях и конечных порядках можно формулировать как задачи, связанные с решением систем уравнений над этими системами, что приводит к необходимости развития алгебраической геометрии. Алгебраическая геометрия над этими объектами тесным образом связана со свойствами экзистенциальных и универсальных теорий. С практической точки зрения важнейшими являются вопросы разрешимости и вычислительной сложности этих теорий, хотя некоторые задачи алгебраической геометрии над данными объектами могут быть алгоритмически сложны [5]. Что касается классических алгебраических структур, то ещё в 1949 г. А. Тарский [6] установил, что элементарная теория класса конечных порядков неразрешима. Дж. Акс в 60-х годах прошлого века получил серьёзный результат о разрешимости элементарной теории класса всех конечных полей [7]. Из этого следует разрешимость экзистенциальной и универсальной теорий класса конечных полей. И. А. Лавров [8] доказал, что элементарная теория класса конечных графов неразрешима. Но экзистенциальная и универсальная теории класса конечных графов уже разрешимы, что показано А. В. Ильевым [9].

В данной работе доказывается, что экзистенциальная и универсальная теории класса всех частичных порядков являются разрешимыми и, как следствие, теории конечных частичных порядков тоже.

1. Предварительные сведения

Напомним базовые определения из теории частичных порядков [10], теории графов [11] и теории моделей [12].

Частично упорядоченным множеством (*частичным порядком*) называется алгебраическая система $\mathcal{P} = \langle P| \leqslant^{(2)}\rangle$, где $\leqslant$ — предикатный символ отношения порядка, на которой выполнены следующие три аксиомы:

1) $\forall p \in P\ (p \leqslant p)$ (рефлексивность);
2) $\forall p_1, p_2 \in P\ (p_1 \leqslant p_2 \wedge p_2 \leqslant p_1 \Rightarrow p_1 = p_2)$ (антисимметричность);
3) $\forall p_1, p_2, p_3 \in P\ (p_1 \leqslant p_2 \wedge p_2 \leqslant p_3 \Rightarrow p_1 \leqslant p_3)$ (транзитивность).

Данный частичный порядок определён в языке без констант. Будем обозначать такой язык через L, а множество переменных — через X.

Элементы x и y частичного порядка $\mathcal{P}$ называются *сравнимыми*, если либо $x \leqslant y$, либо $y \leqslant x$ верно в $\mathcal{P}$. Если оба неравенства не верны, то элементы *несравнимы*.

Атомарной формулой языка L от переменных X называется выражение одного из следующих типов:

1) $x_i = x_j$, где $x_i, x_j \in X$;
2) $x_i \leqslant x_j$, где $x_i, x_j \in X$.

Формулой языка L называется выражение, определённое рекурсивно следующим образом:

1) Атомарная формула — это формула.
2) Если φ — формула, то $\neg\varphi$ — тоже формула.
3) Если φ, ψ — формулы, то $\varphi \vee \psi$ и $\varphi \wedge \psi$ — тоже формулы.
4) Если φ — формула, то $\exists x_i\ \varphi$ и $\forall x_i\ \varphi$ — тоже формулы.

Определим следующие типы формул языка L. *Предложением* языка L называется формула без свободных переменных (т. е. все переменные находятся под действием кванторов). Предложение имеет общий вид $Q_1x_1 \ldots Q_nx_n\Psi(x_1, \ldots, x_n)$, где Q_i для $i \in \{1, \ldots, n\}$ — это квантор существования или всеобщности, а Ψ — формула языка L со свободными переменными $x_1, \ldots, x_n$. *Универсальным*, или $\forall$-предложением, называется предложение, где все переменные находятся под действием кванторов всеобщности. Аналогично, *экзистенциальным*, или $\exists$-предложением, называется предложение, в котором все переменные находятся под действием кванторов существования.

Для краткости формулу $\neg(x = y)$ будем обозначать $x \neq y$. Определим формулу строгого порядка следующим образом: $x < y \Leftrightarrow x \leqslant y \wedge x \neq y$. Формула, показывающая, что элементы x и y несравнимы, определяется как $x \not\sim y \Leftrightarrow x \neq y \wedge \neg(x < y) \wedge$ $\wedge\ \neg(y < x)$. Отрицание неравенства определяется следующим образом: $\neg(x \leqslant y) \Leftrightarrow$ $\Leftrightarrow (y < x) \vee (x \not\sim y)$. При отрицании неравенства между переменными они могут оказаться либо сравнимыми обратным строгим неравенством (первый конъюнкт), либо несравнимыми (второй конъюнкт).

Теорией частичного порядка $\mathcal{P}$ языка L называется всё множество предложений языка L, верное над $\mathcal{P}$. *Универсальной* теорией (или $\forall$-теорией) частичного порядка $\mathcal{P}$ языка L называется подмножество $\forall$-предложений теории частичного порядка $\mathcal{P}$. По аналогии, *экзистенциальной* теорией ($\exists$-теорией) называется подмножество $\exists$-предложений теории частичного порядка $\mathcal{P}$. Строго говоря, здесь дано определение *полной* теории частичного порядка. В данной работе рассматриваются именно такие теории.

Теория T частичного порядка $\mathcal{P}$ языка L называется *разрешимой*, если существует алгоритм, который для любого предложения φ языка L проверяет, принадлежит ли предложение φ теории T.

Определим теории класса частичных порядков. Под классом частичных порядков $\mathbf{P}$ языка L будем понимать семейство частичных порядков языка L, которое вместе с любым частичным порядком содержит все изоморфные ему частичные порядки языка L. *Элементарной* теорий или просто теорией класса $\mathbf{P}$ называется множество $Th(\mathbf{P})$ всех предложений языка L, выполненных на всех частичных порядках из $\mathbf{P}$. Универсальной теорией класса $\mathbf{P}$ называется подмножество $Th_\forall(\mathbf{P}) \subseteq Th(\mathbf{P})$ универсальных предложений теории класса $\mathbf{P}$. Аналогично, экзистенциальной теорией класса $\mathbf{P}$ называется подмножество $Th_\exists(\mathbf{P}) \subseteq Th(\mathbf{P})$ экзистенциальных предложений теории класса $\mathbf{P}$.

Частичные порядки тесно связаны с графами, а именно: каждому конечному частичному порядку можно сопоставить взаимно однозначно соответствующий конечный граф. Эта связь хорошо известна [11], но для понимания опишем алгоритм перехода от конечного частичного порядка $\mathcal{P} = \langle P| \leqslant\rangle$ к конечному графу $\Gamma = \langle V|E\rangle$. Вершинам графа Γ соответствуют элементы носителя $\mathcal{P}$. В графе Γ есть дуга (p_1, p_2), если в $\mathcal{P}$ верно $p_1 \geqslant p_2$. Полученный граф Γ является транзитивно замкнутым с петлями в каждой вершине. Если убрать все петли из Γ, то получится ациклический ориентированный граф. Назовём такие графы, соответствующие частичным порядкам, *p-графами*.

2. Разрешимость теории

Теорема 1. Универсальная теория класса всех частичных порядков в языке L без констант разрешима.

Доказательство. В качестве доказательства построим алгоритм проверки вхождения произвольного универсального предложения φ в $\forall$-теорию класса частичных порядков $\mathbf{P}$. Заметим, что для того, чтобы предложение φ принадлежало универсальной теории класса, необходимо, чтобы её отрицание $\neg\varphi$ было ложно на всех частичных порядках из $\mathbf{P}$. Если же существует частичный порядок из $\mathbf{P}$, на котором $\neg\varphi$ верно, то само предложение φ не принадлежит универсальной теории класса $\mathbf{P}$.

Краткое описание алгоритма таково. Пусть $Th_\forall(\mathbf{P})$ — универсальная теория класса всех частичных порядков. На вход алгоритма подаётся произвольное универсальное предложение φ. От этого предложения берётся его отрицание $\neg\varphi$ и приводится к предварённой дизъюнктивной форме. Далее алгоритм строит графы по конъюнктам из $\neg\varphi$ и определяет наличие циклов в этих графах. Если хотя бы один граф является ациклическим, то предложение φ не принадлежит $Th_\forall(\mathbf{P})$. Если не удалось построить ни одного графа либо все построенные графы содержат цикл, то $\varphi \in Th_\forall(\mathbf{P})$ (алгоритм 1).

Алгоритм 1.

1: От универсального предложения φ берём его отрицание. Полученное предложение $\neg\varphi = \exists p_1 \ldots \exists p_n\, \psi$ является экзистенциальным, ψ — его бесквантворная часть. Далее $\neg\varphi$ преобразуется в эквивалентное предложение $\neg\varphi_1 = \exists p_1 \ldots \exists p_n(\psi_1 \vee \ldots \vee \vee\psi_m)$, находящееся в предварённой дизъюнктивной форме, где ψ_i, $i \in \{1, \ldots, m\}$, — конъюнкты, имеющие вид $\psi_i = \bigwedge\limits_j \xi_j$, ξ_j — либо атомарные формулы, либо их отрицания.

2: Алгоритм последовательно просматривает все конъюнкты ψ_i в предложении $\neg\varphi_1$. Если нашёлся конъюнкт $\psi_i = \xi_i \wedge \neg(x \leqslant y)$, где ξ_i — предложение, то ψ_i заменяется

на дизъюнкцию $(\xi_i \wedge y < x) \vee (\xi_i \wedge x \not< y)$. В конце процедуры получится новое предложение $\neg\varphi_2$.

3: Алгоритм последовательно просматривает все конъюнкты ψ_i' в предложении $\neg\varphi_2$. Если нашёлся конъюнкт $\psi_i' = \xi_i \wedge x \leqslant y$, где ξ_i — предложение, то ψ_i' заменяется на дизъюнкцию $(\xi_i \wedge x < y) \vee (\xi_i \wedge x = y)$. В конце процедуры получится новое предложение $\neg\varphi_3$.

4: Алгоритм последовательно просматривает все конъюнкты ψ_i'' в предложении $\neg\varphi_3$. Если в каком-то конъюнкте ψ_i'' нет равенства или отрицания равенства между переменными x и y, то ψ_i'' заменяется на дизъюнкцию $(\psi_i'' \wedge x = y) \vee (\psi_i'' \wedge x \neq y)$. В конце процедуры получится новое предложение $\neg\varphi_4$.

5: В каждом конъюнкте ψ_i''' предложения $\neg\varphi_4$, содержащем равенства переменных $x = y$, это равенство удаляется из конъюнкта и вхождение переменной y в ψ_i''' заменяется на переменную x. В конце процедуры получится новое предложение $\neg\varphi_5$.

6: Предложение $\neg\varphi$ истинно над частичным порядком $\mathcal{P}$ из $\mathbf{P}$, если над $\mathcal{P}$ выполнен хотя бы один конъюнкт из предложения $\neg\varphi_5$. Алгоритм последовательно просматривает все конъюнкты $\widehat{\psi}_i$ из $\neg\varphi_5$ и удаляются те из них, которые ложны на всех частичных порядках. Это конъюнкты следующего вида:

1) конъюнкт содержит уравнение типа $x \neq x$;
2) конъюнкт содержит и $x \not< y$, и цепь между этими элементами;
3) конъюнкт содержит циклическую цепь типа $x_1 < x_2 < \ldots < x_n < x_1$.

Если в конце процедуры удалены все конъюнкты, то предложение φ истинно для всех частичных порядков и, следовательно, принадлежит универсальной теории $Th_\forall(\mathbf{P})$. Иначе не принадлежит.

Для окончания доказательства остаётся обосновать алгоритм.

Первый шаг алгоритма приводит к предварённой дизъюнктивной форме формулу $\neg\varphi$. Шаги со второго по пятый приводят формулу $\neg\varphi_1$ к такому эквивалентному виду, где в каждом конъюнкте присутствуют только формулы типа $x < y$, $x \neq y$ и $x \not< y$. На втором шаге происходит ветвление конъюнкта, содержащего формулу $\neg(x \leqslant y)$. Так как отрицание отношения порядка двух элементов — это либо $y < x$, либо $x \not< y$, то этим объясняется ветвление на данном шаге. Аналогично происходит деление на третьем шаге, где $x \leqslant y \Leftrightarrow x < y \vee x = y$. Далее, если в конъюнкте для переменных x и y нет уравнений $x = y$ и $x \neq y$, то опять производится ветвление конъюнкта на пару: конъюнкт с равенством переменных x и y и конъюнкт с отрицанием равенства. Наконец, на пятом шаге происходит замена всех групп равных переменных на представителей. При этом могут получиться противоречивые конъюнкты. Например, конъюнкт $\psi = (x \neq z) \wedge (y = z) \wedge (y = x)$, состоящий из трёх формул, после пятого шага преобразуется в $y \neq y$.

Далее по каждому конъюнкту ψ_i в формуле $\neg\varphi_5$ можно построить граф Γ_i таким образом, что переменные в конъюнкте — это вершины графа Γ_i, а уравнение $x < y$ задаёт дугу (v_y, v_x) в графе Γ. Естественно, в конъюнкте не должно содержаться отрицаний равенств вида $x \neq x$, иначе конъюнкт несовместен и построение графа по нему невозможно. Конъюнкт также несовместен, если в нём содержится $x \not< y$ и одновременно с этим есть цепь между x и y. Это означает, что элементы в конъюнкте одновременно сравнимы и несравнимы. Теперь, если граф Γ_i не содержит циклов, то обязательно найдётся частичный порядок $\mathcal{P} \in \mathbf{P}$, p-граф которого содержит граф Γ_i. p-Граф является ациклическим (исключая петли) и транзитивно замкнутым, поэтому в такой граф может вкладываться граф Γ_i. Поэтому, если в конъюнкте содержится

цикл $x_1 < x_2 < \cdots < x_n < x_1$, то для такого графа Γ_i не существует соответствующего p-графа и, следовательно, конъюнкт не верен для любых значений переменных. Иначе конъюнкт верен.

Если хотя бы один конъюнкт формулы $\neg\varphi_5$ верен, то вся формула $\neg\varphi$ верна и, следовательно, формула φ не принадлежит универсальной теории $Th_\forall(\mathbf{P})$. ■

Поскольку универсальная теория класса всех частичных порядков в языке L разрешима, то и универсальная теория всех *конечных* частичных порядков в языке L также является разрешимой. Сформулируем следствие.

Теорема 2. Экзистенциальная теория класса всех частичных порядков в языке L без констант разрешима.

Доказательство. Класс $\mathbf{P}$ содержит частичный порядок, состоящий ровно из одного элемента. Обозначим такой частичный порядок через ε. Это значит, что для того чтобы предложение φ принадлежало экзистенциальной теории частичных порядков, предложение φ должно быть выполнено над ε. Предваренная дизъюнктивная форма предложения φ — это $\varphi_1 = \exists x_1 \ldots \exists x_n \bigvee\limits_j \xi_j$, где ξ_j — конъюнкт. Если φ выполнено над ε, то в φ_1 существует такой конъюнкт ξ_i, что предложение $\omega = \exists x_1 \ldots \exists x_n \xi_i$ выполнено над ε. Такое предложение ω будет выполнено над любым частичным порядком из $\mathbf{P}$. Поэтому предложение φ принадлежит теории $Th_\exists(\mathbf{P})$, если его предварённая ДФ содержит такой конъюнкт ξ_i.

Поиск такого конъюнкта можно провести с помощью алгоритма 1 по шагам 1–4. Конъюнкт будет состоять их всех возможных равенств между переменными. Если такой конъюнкт нашёлся, то предложение φ принадлежит экзистенциальной теории класса $\mathbf{P}$, иначе не принадлежит. ■

Аналогично выводу из предыдущей теоремы, экзистенциальная теория класса всех конечных частичных порядков в языке L является разрешимой.

Автор выражает благодарность рецензенту за важные замечания и предложения по улучшению текста статьи.

ЛИТЕРАТУРА

1. *Даниярова Э. Ю., Мясников А. Г., Ремесленников В. Н.* Алгебраическая геометрия над алгебраическими системами. Новосибирск: Изд-во СО РАН, 2016. 288 с.
2. *Матиясевич Ю. В.* Диофантовость перечислимых множеств // Доклады АН СССР. 1970. Т. 191. № 2. С. 279–282.
3. *Mayr E. W. and Meyer A. R.* The complexity of the word problems for commutative semigroups and polynomial ideals // Adv. Math. 1982. V. 46(3). P. 305–329.
4. *Cook S. A.* The complexity of theorem proving procedures // Proc. 3d Ann. ACM Symp. Theory of Computing. N.Y., USA, 1971. P. 151–158.
5. *Никитин А. Ю., Рыбалов А. Н.* О сложности проблемы разрешимости систем уравнений над конечными частичными порядками // Прикладная дискретная математика. 2018. № 39. С. 94–98.
6. *Tarski A.* Undecidability of the theory of lattices and projective geometries // J. Symbolic Logic. 1949. V. 15. P. 77–78.
7. *Ax J.* The elementary theory of finite fields // Ann. Math. 1968. V. 88. No. 2. P. 239–271.
8. *Лавров И.А.* Эффективная неотделимость множества тождественно истинных и конечно опровержимых формул некоторых элементарных теорий // Алгебра и логика. 1963. Т. 2. № 1. С. 5–18.

9. *Ильев А.В.* Разрешимость универсальных теорий и аксиоматизируемость наследственных классов графов // Труды Института математики и механики УрО РАН. 2016. Т. 22. № 1. С. 100–111.
10. *Gratzer G.* General Lattice Theory. Birkhauser, 1998. 663 p.
11. *Оре О.* Теория графов. М.: Наука, 1980. 336 с.
12. *Marker D* Model Theory: An Introduction. N.Y..: Springer, 2002. 342 p.

REFERENCES

1. *Daniyarova E. Y., Myasnikov A. G., and Remeslennikov V. N.* Algebraicheskaya geometria nad algebraicheskimi sistemami [Algebraic Geometry over Algebraic Systems]. Novosibirsk, SB RAS Publ., 2016. 288 p (in Russian).
2. *Matiyasevich Yu. V.* Diofantovost' perechislimykh mnozhestv [Diophantineity of enumerable sets]. Doklady Akademii Nauk USSR, 1970, vol. 191, no. 2, pp. 279–282. (in Russian)
3. *Mayr E. W. and Meyer A. R.* The complexity of the word problems for commutative semigroups and polynomial ideals. Adv. Math., 1982, vol. 46 (3), pp. 305–329.
4. *Cook S. A.* The complexity of theorem proving procedures. Proc. 3d Ann. ACM Symp. Theory of Computing, N.Y., USA, 1971, pp. 151–158.
5. *Nikitin A. Yu. and Rybalov A. N.* O slozhnosti problemy razreshimosti sistem uravneniy nad konechnymi chastichnymi poryadkami [On complexity of the satisfiability problem of systems over finite posets]. Prikladnaya Diskretnaya Matematika, 2018, no. 39, pp. 94–98. (in Russian)
6. *Tarski A.* Undecidability of the theory of lattices and projective geometries. J. Symbolic Logic, 1949, vol. 15, pp. 77–78.
7. *Ax J.* The elementary theory of finite fields. Ann. Math., 1968, vol. 88, no. 2, pp. 239–271.
8. *Lavrov I. A.* Effectivnaya neotdelimost' mnojestva tojdestvenno istinnih i conechno oproverjimih formul nekotorih elementarnih teoriy [Effective inseparability of the set of identically true and finitely refutable formulas of some elementary theories]. Algebra i Logica, 1963, vol. 2, no. 1, pp. 5–18. (in Russian)
9. *Il'ev A. V.* Razreshimost' universalnih teoriy i aksiomatiziruemost' nasledstvennih classov grafov [Decidability of universal theories and axiomatizability of hereditary classes of graphs]. Trudi Instituta Matematiki i Mehaniki UrO RAN, 2016, vol. 22, no. 1, pp. 100–111. (in Russian)
10. *Gratzer G.* General Lattice Theory. Birkhauser, 1998. 663 p.
11. *Ore O.* Theory of Graphs. American Mathematical Soc, 1962. 270 p.
12. *Marker D* Model Theory: An Introduction. N.Y., Springer, 2002. 342 p.

УДК 519.1,519.7

О СТЕПЕНИ ОГРАНИЧЕНИЙ ФУНКЦИЙ q-ЗНАЧНОЙ ЛОГИКИ НА ЛИНЕЙНЫЕ МНОГООБРАЗИЯ

В. Г. Рябов

НП «ГСТ», г. Москва, Россия

В случае конечного поля $\mathbb{F}_q$ степень ограничения функции q-значной логики от n переменных на линейное многообразие размерности r векторного пространства $\mathbb{F}_q^n$ определена как степень полинома от r переменных, представляющего данное ограничение. Для многообразий фиксированной размерности оценена вероятность появления у функции ограничений степени не выше заданной, а также получена асимптотика числа многообразий, на которых ограничения аффинны. Показано, что при $n \to \infty$ для почти всех функций q-значной логики от n переменных значение максимальной размерности линейного многообразия, на котором ограничение аффинно, принадлежит отрезку $[\lfloor \log_q n + \log_q \log_q n \rfloor, \lceil \log_q n + \log_q \log_q n \rceil]$, в то время как аналогичный параметр для случая фиксации переменных находится в пределах $[\lfloor \log_q n \rfloor, \lceil \log_q n \rceil]$.

Ключевые слова: *многозначная логика, булева функция, ограничение, линейное многообразие, степень.*

DOI 10.17223/20710410/45/2

ON THE DEGREE OF RESTRICTIONS OF q-VALUED LOGIC FUNCTIONS TO LINEAR MANIFOLDS

V. G. Ryabov

NP "GST", Moscow, Russia

E-mail: 4vryabov@gmail.com

In case of a finite field $\mathbb{F}_q$, the degree of restricting a q-valued logic function in n variables to a r-dimensional linear manifold of the vector space $\mathbb{F}_q^n$ is defined as the degree of a polynomial in r variables that represents this restriction. For manifolds of a fixed dimension, the probability of occurrence of restrictions with a degree not higher than the given one is estimated, and the asymptotics of the number of manifolds on which the restrictions are affine is obtained. It is shown that if $n \to \infty$, for almost all q-valued logic functions in n variables, the value of the maximum dimension of a linear manifold on which the restriction is affine belongs to the segment $[\lfloor \log_q n + \log_q \log_q n \rfloor, \lceil \log_q n + \log_q \log_q n \rceil]$, while the analogous parameter for the case of fixing variables is in the range $[\lfloor \log_q n \rfloor, \lceil \log_q n \rceil]$.

Keywords: *many-valued logic, Boolean function, restriction, linear manifold, degree.*

Введение

Пусть $\mathbb{F}_q$ — конечное поле, состоящее из q элементов, где $q = p^m$, p — простое число, $m \in \mathbb{N}$, и $\mathbb{F}_q^n$ — n-мерное векторное пространство над данным полем. В качестве подпространства векторного пространства $\mathbb{F}_q^n$ выступает его непустое подмножество S,

замкнутое относительно операций сложения векторов и умножения вектора на элемент поля. Любое определённое таким образом подпространство само является векторным пространством над полем $\mathbb{F}_q$, возможно, меньшей размерности. Размерность векторного подпространства, состоящего из нулевого вектора, равна нулю.

Согласно [1], определим сдвиг на вектор δ из $\mathbb{F}_q^n$ как отображение φ пространства $\mathbb{F}_q^n$ в себя вида $\varphi_\delta(\alpha) = \alpha \oplus \delta$, где α принимает всевозможные значения из $\mathbb{F}_q^n$; $\oplus$ — операция сложения в $\mathbb{F}_q^n$. Для любого подпространства S сдвиг приводит к появлению образа $\varphi(S)$, который является линейным многообразием векторного пространства $\mathbb{F}_q^n$ (в дальнейшем для линейного многообразия будем использовать термин «многообразие»; линейное многообразие также называют аффинным подпространством и плоскостью). При этом многообразие M является исходным подпространством S, если вектор δ принадлежит S, или M содержит нулевой вектор $\mathbb{F}_q^n$. Для каждого многообразия существует единственное подпространство S, сдвиг которого приводит к получению M, а вот в качестве вектора сдвига может выступать любой вектор, образованный суммой заданного вектора сдвига и произвольного вектора из S. В этом случае будем говорить, что многообразие имеет направление S и такую же размерность, как S. Непустое пересечение многообразий также является многообразием с направлением, заданным пересечением соответствующих подпространств.

Рассмотрим множество всех отображений пространства $\mathbb{F}_q^n$ в поле $\mathbb{F}_q$, или функций q-значной логики от n переменных, которое обозначим P_q^n. Известно (см., например, [2]), что любая функция f из P_q^n с точностью до перестановки слагаемых может быть представлена полиномом над полем $\mathbb{F}_q$ следующего вида:

$$f(x_1, \ldots, x_n) = \bigoplus_{i_1,\ldots,i_n} a_{i_1,\ldots,i_n} \otimes x_1^{i_1} \otimes \cdots \otimes x_n^{i_n}, \tag{1}$$

где $i_1, \ldots, i_n$ — всевозможные различные наборы n целых чисел, принимающих значения от 0 до $q-1$; $a_{i_1,\ldots,i_n}$ — элемент $\mathbb{F}_q$; $\oplus$ и $\otimes$ — операции сложения и умножения в $\mathbb{F}_q$, а степень определяется через операцию умножения в $\mathbb{F}_q$ с учётом того, что $x^0 = 1$ и $x^1 = x$. Случай $q = p$, где p — простое число, приводит к полиному с операциями сложения и умножения по модулю p, а для $q = 2$ получаем полином Жегалкина булевой функции. Степень функции f определим как степень полинома (максимальную сумму степеней в мономе с ненулевым коэффициентом) вида (1) и обозначим d_f. Для неё справедливы неравенства $0 \leqslant d_f \leqslant n(q-1)$. Определим функцию f из P_q^n как аффинную, если $d_f \leqslant 1$. Аффинная функция является линейной, если свободный член полинома $a_{0,\ldots,0}$ является нулевым элементом $\mathbb{F}_q$, и константой, если $d_f = 0$.

Для функции f из P_q^n обозначим $f|_R$ её ограничение на подмножество R пространства $\mathbb{F}_q^n$. Когда R является линейным многообразием, можно получить полиномиальное представление ограничения и определить степень ограничения. Действительно, пусть M — многообразие размерности r. В [3] показано, что M является совокупностью решений системы $n-r$ линейных уравнений с n неизвестными над полем $\mathbb{F}_q$ и рангом матрицы коэффициентов $n-r$. Случаю, когда многообразие является подпространством, соответствует система однородных уравнений. Указанная система позволяет получить аффинные выражения $n-r$ переменных через оставшиеся r переменных. Подставив их в полином, отвечающий функции f, и приведя получившийся полином от r переменных к виду (1), можно получить полиномиальное представление функции g из P_q^n, несущественно зависящей от $n-r$ переменных, для которой выполняется условие $g|_M = f|_M$. При различных способах выражения переменных степень получившихся полиномов инвариантна и удовлетворяет неравенствам $0 \leqslant d_g \leqslant r(q-1)$. Для определённости будем последовательно на каждом из $n-r$ шагов выражать переменную с наибольшим

возможным индексом через оставшиеся переменные с меньшими индексами. Последнее позволяет говорить об однозначном представлении ограничения $f|_M$ полиномом от r переменных над полем $\mathbb{F}_q$ и определить степень ограничения $d_{f|_M}$ как степень получившегося полинома от r переменных.

Наибольший практический интерес представляет случай многообразия максимальной размерности, на котором ограничение функции имеет степень, не превосходящую 1 (такие ограничения по аналогии с функциями будем называть аффинными). Он определяет минимальное число линейных комбинаций, которые можно зафиксировать так, что отвечающий данной функции полином примет аффинный вид, что является важным для решения задачи линеаризации уравнения над конечным полем.

Фиксация переменных константами поля $\mathbb{F}_q$, или переход к подфункции, является частным случаем и не всегда даёт оптимальный результат для подобной линеаризации. Вместе с тем данный случай является важным для теории и практики применения функций q-значной логики и в работе выделен особо. До настоящего времени в научной литературе в основном изучались свойства подфункций булевых функций, в том числе их степень. Одним из первых изучением асимптотических свойств подфункций-констант занимался Ю. И. Журавлев.

1. Вероятностные оценки появления у функции ограничений степени не выше заданной на линейных многообразиях

Обозначим $\mathfrak{M}_q^n(r)$ множество всех многообразий пространства $\mathbb{F}_q^n$ размерности r, где $0 \leqslant r \leqslant n$, а $\hat{\mathfrak{M}}_q^n(r)$ — его подмножество, заданное фиксацией $(n-r)$ переменных элементами поля $\mathbb{F}_q$. Для функции f из P_q^n введём характеристики: $d_f(r) = \min\limits_{M \in \mathfrak{M}_q^n(r)} d_{f|_M}$ и $\hat{d}_f(r) = \min\limits_{M \in \hat{\mathfrak{M}}_q^n(r)} d_{f|_M}$. Тогда справедливы неравенства $0 \leqslant d_f(r) \leqslant \hat{d}_f(r) \leqslant r(q-1)$, а также соотношения $0 = d_f(0) \leqslant \ldots \leqslant d_f(n) = d_f$ и $0 = \hat{d}_f(0) \leqslant \ldots \leqslant \hat{d}_f(n) = d_f$. В отличие от булевых функций, для которых ограничения на многообразия размерности 1 всегда являются аффинными, ограничения функций q-значной логики могут иметь степень от 0 до $q-1$ даже на таких многообразиях.

Пусть r и d являются целыми неотрицательными числами, удовлетворяющими условиям $0 \leqslant r \leqslant n$ и $0 \leqslant d \leqslant r(q-1)$. В рамках классического способа задания конечного вероятностного пространства на множестве P_q^n [4] рассмотрим дискретные случайные величины $d_f(r)$ и $\hat{d}_f(r)$, а также связанные с ними события $P_q^n(d_f(r) \leqslant d) = \{f \in P_q^n \mid d_f(r) \leqslant d\}$ и $P_q^n(\hat{d}_f(r) \leqslant d) = \{f \in P_q^n \mid \hat{d}_f(r) \leqslant d\}$ (в дальнейшем будем одновременно использовать краткую форму обозначения событий вида $\{d_f(r) \leqslant d\}$). При фиксированной размерности многообразий r параметр d задаёт функции распределения величин $d_f(r)$ и $\hat{d}_f(r)$ с вероятностями $\dfrac{|P_q^n(d_f(r) \leqslant d)|}{|P_q^n|}$ и $\dfrac{|P_q^n(\hat{d}_f(r) \leqslant d)|}{|P_q^n|}$ соответственно; здесь функции распределения имеют вид $F(d) = \mathsf{P}(d_f(r) \leqslant d)$ и $\hat{F}(d) = \mathsf{P}(\hat{d}_f(r) \leqslant d)$. Для оценки вероятностей предварительно докажем лемму, которая, впрочем, имеет и самостоятельное значение для характеристики функций q-значной логики.

Лемма 1. Пусть d — целое неотрицательное число, удовлетворяющее условию $0 \leqslant d \leqslant n(q-1)$, и $f_q^d(n)$ — число всех функций из P_q^n со степенью, не превышающей d, то есть выполняется равенство $f_q^d(n) = |P_q^n(d_f(n) \leqslant d)|$. Тогда справедливо

соотношение

$$f_q^d(n) = q^{\sum\limits_{j=0}^{\lfloor d/q \rfloor} (-1)^j \binom{n}{j}\binom{n+d-qj}{d-qj}}. \tag{2}$$

Доказательство. Обратимся к полиномиальному представлению функций из P_q^n. Число полиномов степени не выше d выражается равенством $f_q^d(n) = \prod\limits_{m=0}^{d} h_q^m(n)$, где $h_q^m(n)$ — число однородных компонент многочлена от n переменных, соответствующих степени m. Используя формулу числа m-выборок в коммутативном несимметричном n-базисе, для которых показатели первичной спецификации не превосходят $(q-1)$ [5], получим выражение для $h_q^m(n)$:

$$h_q^m(n) = q^{\sum\limits_{j=0}^{n} (-1)^j \binom{n}{j}\binom{n+m-qj-1}{m-qj}}. \tag{3}$$

Тогда выражение для $f_q^d(n)$ примет следующий вид:

$$f_q^d(n) = q^{\sum\limits_{m=0}^{d}\sum\limits_{j=0}^{n} (-1)^j \binom{n}{j}\binom{n+m-qj-1}{m-qj}}.$$

В завершение доказательства осталось привести цепочку равенств, вытекающую из рекуррентного соотношения для биномиальных коэффициентов:

$$\sum_{m=0}^{d}\sum_{j=0}^{n}(-1)^j\binom{n}{j}\binom{n+m-qj-1}{m-qj} = \sum_{j=0}^{n}(-1)^j\binom{n}{j}\sum_{m=0}^{d}\binom{n+m-qj-1}{m-qj} =$$
$$= \sum_{j=0}^{\lfloor d/q \rfloor}(-1)^j\binom{n}{j}\binom{n+d-qj}{d-qj}.$$

Лемма 1 доказана. ■

Следствие 1. В условиях леммы 1 имеет место оценка $f_q^d(n) \leqslant q^{\binom{n+d}{d}}$.

Данная оценка обусловлена тем, что при снятии ограничений на показатели первичной спецификации [5, с. 252] формула (3) примет вид $h_q^m(n) = q^{\binom{n+m-1}{m}}$.

Замечание 1. При малых d в формуле (2) остаются только первые слагаемые. Например, при $0 \leqslant d < q-1$ получим $f_q^d(n) = q^{\binom{n+d}{d}}$, а при $q \leqslant d < 2q$ имеем $f_q^d(n) = q^{\binom{n+d}{d} - n\binom{n+d-q}{d-q}}$. В случае булевых функций имеет место формула $f_2^d(n) = 2^{\sum\limits_{k=0}^{d}\binom{n}{k}}$.

Замечание 2. Формула (2) может быть также использована для нахождения числа всех функций из P_q^n со степенью равной d, поскольку последнее определяется как разность $f_q^d(n) - f_q^{d-1}(n)$.

Получим теперь оценки распределения случайных величин $d_f(r)$ и $\hat{d}_f(r)$ для $r \geqslant 1$.

Теорема 1. Пусть r и d — целые неотрицательные числа, удовлетворяющие условиям $1 \leqslant r \leqslant n$ и $0 \leqslant d \leqslant r(q-1)$. Тогда для вероятностей наступления событий $\{d_f(r) \leqslant d\}$ и $\{\hat{d}_f(r) \leqslant d\}$ справедливы оценки

$$\mathsf{P}(d_f(r) \leqslant d) \leqslant \frac{q^{\sum\limits_{j=0}^{\lfloor d/q \rfloor} (-1)^j \binom{r}{j}\binom{r+d-qj}{d-qj}+n-r}\binom{n}{r}_q}{q^{q^r}}; \tag{4}$$

$$\mathsf{P}(\hat{d}_f(r) \leqslant d) \leqslant \frac{q^{\sum\limits_{j=0}^{\lfloor d/q \rfloor} (-1)^j \binom{r}{j}\binom{r+d-qj}{d-qj}+n-r}\binom{n}{r}}{q^{q^r}}, \tag{5}$$

где $\binom{n}{r}_q$ и $\binom{n}{r}$ — гауссов и обыкновенный биномиальные коэффициенты соответственно.

Доказательство. При $r = n$ утверждения теоремы следуют из леммы 1. Докажем их при $r < n$.

Пусть $f \in P_q^n$ и $M \in \mathfrak{M}_q^n(r)$. Тогда ограничению $f|_M$ соответствует полином вида (1) от r переменных. Поскольку выбранный выше способ выражения $n - r$ переменных через оставшиеся r переменных зависит только от M, можно утверждать, что для всех функций из P_q^n ограничения на M будут представлены полиномами от одних и тех же r переменных. С другой стороны, каждый из q^{q^r} полиномов от данных r переменных отвечает ограничениям на M различных $q^{q^n-q^r}$ функций от n переменных, которые в совокупности покрывают все множество P_q^n. Тогда вероятность наступления события $P_q^n(d_{f|_M} \leqslant d) = \{f \in P_q^n \mid d_{f|_M} \leqslant d\} \subset P_q^n(d_f(r) \leqslant d)$ при условии случайной и равновероятной выборки может быть найдена следующим образом:

$$\mathsf{P}(d_{f|_M} \leqslant d) = \frac{|P_q^n(d_{f|_M} \leqslant d)|}{|P_q^n|} = \frac{q^{q^n-q^r} f_q^d(r)}{q^{q^n}} = \frac{f_q^d(r)}{q^{q^r}}. \tag{6}$$

Обозначим $m_q^n(r)$ число всех линейных многообразий размерности r пространства $\mathbb{F}_q^n$, т. е. $m_q^n(r) = |\mathfrak{M}_q^n(r)|$. Известно [3], что при $r \geqslant 1$ число подпространств размерности r пространства $\mathbb{F}_q^n$ определяется по формуле $\prod\limits_{i=0}^{r-1} \dfrac{q^n - q^{r-i}}{q^r - q^{r-i}}$. После сокращения соответствующих дробей придём к гауссову биномиальному коэффициенту $\binom{n}{r}_q$, где $\binom{n}{0}_q = 1$ и $\binom{n}{r}_q = \prod\limits_{i=0}^{r-1} \dfrac{q^{n-i} - 1}{q^{r-i} - 1}$ при $r = 1, \ldots, n$. С учётом сдвига получим формулу

$$m_q^n(r) = q^{n-r} \binom{n}{r}_q. \tag{7}$$

Используя свойство вероятности суммы событий, получим

$$\mathsf{P}(d_f(r) \leqslant d) \leqslant \sum_{M \in \mathfrak{M}_q^n(r)} \mathsf{P}(d_{f|_M} \leqslant d) = \frac{f_q^d(r) m_q^n(r)}{q^{q^r}}. \tag{8}$$

Для завершения доказательства формулы (4) осталось подставить в правую часть (8) значения величин $f_q^d(r)$ из (2) и $m_q^n(r)$ из (7).

Формула (5) доказывается аналогично, с той лишь разницей, что суммирование в (8) ведётся по многообразиям из множества $\hat{\mathfrak{M}}_q^n(r)$, число которых $\hat{m}_q^n(r)$ равно

$$\hat{m}_q^n(r) = q^{n-r} \binom{n}{r}. \tag{9}$$

Теорема 1 доказана. ■

Следствие 2. В условиях теоремы 1 имеют место оценки

$$|P_q^n(d_f(r) \leqslant d)| \leqslant q^{q^n - q^r + n - r + \sum\limits_{j=0}^{\lfloor d/q \rfloor} (-1)^j \binom{r}{j} \binom{r+d-qj}{d-qj}} \binom{n}{r}_q,$$

$$|P_q^n(\hat{d}_f(r) \leqslant d)| \leqslant q^{q^n - q^r + n - r + \sum\limits_{j=0}^{\lfloor d/q \rfloor} (-1)^j \binom{r}{j} \binom{r+d-qj}{d-qj}} \binom{n}{r},$$

вытекающие из определения вероятностной меры.

Теорема 1 позволяет дать верхнюю оценку вероятности появления ограничений и, в частности, подфункций степени не выше заданной на многообразиях фиксированной размерности для случайно выбранной функции q-значной логики.

Пример 1. При $q = 3$ и $n = 5$ вероятности появления ограничений и подфункций степени не выше 3 ($d = 3$) на многообразиях размерности 3 ($r = 3$) не превосходят значений 0,1844... и 0,0015... соответственно. Столь большая разница получившихся оценок обусловлена тем, что число многообразий, заданных фиксацией двух аффинных комбинаций переменных, значительно превышает число многообразий, получающихся при фиксации двух переменных, а именно: $m_3^5(3) = 10890$, в то время как $\hat{m}_3^5(3) = 90$.

Представляют также интерес менее точные, но более удобные для вычислений и дальнейшего анализа оценки.

Следствие 3. В условиях теоремы 1 справедливы оценки

$$\mathsf{P}(d_f(r) \leqslant d) < \frac{q^{\sum\limits_{j=0}^{\lfloor d/q \rfloor}(-1)^j \binom{r}{j}\binom{r+d-qj}{d-qj}+(n-r)(r+1)}\mathrm{e}^{q/(q-1)^2}}{q^{q^r}}; \tag{10}$$

$$\mathsf{P}(\hat{d}_f(r) \leqslant d) \leqslant \frac{q^{\sum\limits_{j=0}^{\lfloor d/q \rfloor}(-1)^j \binom{r}{j}\binom{r+d-qj}{d-qj}+n-r}(n(n-r+1)/r)^{r/2}}{q^{q^r}}. \tag{11}$$

Доказательство. Для доказательства неравенства (10) преобразуем $\binom{n}{r}_q$:

$$\prod_{i=0}^{r-1}\frac{q^{n-i}-1}{q^{r-i}-1} = q^{(n-r)r}\prod_{i=0}^{r-1}\frac{q^{n-i}-1}{q^{n-r}(q^{r-i}-1)} = q^{(n-r)r}\prod_{i=0}^{r-1}\frac{q^{n-i}-q^{n-r}+q^{n-r}-1}{q^{n-r}(q^{r-i}-1)} =$$

$$= q^{(n-r)r}\prod_{i=0}^{r-1}\left(1+\frac{1}{q^{r-i}-1}-\frac{1}{q^{n-r}(q^{r-i}-1)}\right) = q^{(n-r)r}\mathrm{e}^{\sum\limits_{i=0}^{r-1}\ln\left(1+\frac{1}{q^{r-i}-1}-\frac{1}{q^{n-r}(q^{r-i}-1)}\right)}.$$

В отношении показателя степени e имеют место соотношения

$$0 \leqslant \sum_{i=0}^{r-1}\ln\left(1+\frac{1}{q^{r-i}-1}-\frac{1}{q^{n-r}(q^{r-i}-1)}\right) < \sum_{i=0}^{r-1}\frac{1}{q^{r-i}-1}.$$

Таким образом, гауссов биномиальный коэффициент удовлетворяет неравенствам

$$q^{(n-r)r} \leqslant \binom{n}{r}_q < q^{(n-r)r}\mathrm{e}^{q/(q-1)^2} \leqslant q^{(n-r)r}\mathrm{e}^2, \tag{12}$$

из которых, в частности, следует переход от (4) к (10).

Нетрудно видеть, что для сомножителей в формуле биномиального коэффициента $\dfrac{n\cdot\ldots\cdot(n-r+1)}{r\cdot\ldots\cdot 1}$ для любого целого j, принимающего значения от 1 до $\lfloor r/2 \rfloor$, выполняются соотношения $\dfrac{(n-j)(n-r+j+1)}{(r-j)(j+1)} \leqslant \dfrac{(n-j+1)(n-r+j)}{(r-j+1)j}$, откуда следуют неравенства[1]

[1]Более точная нижняя оценка имеет вид $\left(\frac{(n-\lceil r/2\rceil+1)(n-\lfloor r/2\rfloor)}{(\lfloor r/2\rfloor+1)\lceil r/2\rceil}\right)^{r/2} \leqslant \binom{n}{r}$.

$$\left(\frac{2n-r}{r+2}\right)^r < \left(\frac{(2n-r+1)(2n-r)}{(r+2)(r+1)}\right)^{r/2} < \binom{n}{r} \leqslant \left(\frac{n(n-r+1)}{r}\right)^{r/2}, \qquad (13)$$

из которых, в частности, следует переход от (5) к (11). ■

В примере 1 значения оценок, полученных на основе результатов следствия 3, составляют 0,2352... и 0,0017... соответственно, что не меняет качественного характера оценок.

Замечание 3. Что касается значения параметра $f_q^d(r)$, используемого в оценках теоремы 1, из следствия 1 имеем неравенство $f_q^d(r) \leqslant q^{\binom{r+d}{d}}$. Однако использование вместо $\sum\limits_{j=0}^{\lfloor d/q \rfloor} (-1)^j \binom{r}{j}\binom{r+d-qj}{d-qj}$ значения $\binom{r+d}{d}$ существенно искажает полученные оценки при $d \geqslant q$. В случае булевых функций с учётом замечания 1 можно воспользоваться известными оценками суммы биномиальных коэффициентов.

Замечание 4. В соответствии с замечанием 2 результаты теоремы 1 и следствия 3 могут быть использованы для оценки вероятности событий $\{d_f(r) = d\}$ и $\{\hat{d}_f(r) = d\}$. Для этого достаточно заменить $f_q^d(r)$ в соответствующих формулах на разность $f_q^d(r) - f_q^{d-1}(r)$. При этом значения оценок $\mathsf{P}(d_f(r) = d)$ и $\mathsf{P}(\hat{d}_f(r) = d)$ будут незначительно отличаться от соответствующих значений $\mathsf{P}(d_f(r) \leqslant d)$ и $\mathsf{P}(\hat{d}_f(r) \leqslant d)$, поскольку справедливо соотношение $\dfrac{f_q^d(r) - f_q^{d-1}(r)}{f_q^d(r)} = 1 - \dfrac{1}{h_q^d(r)}$. Исключение составляет случай $d = r(q-1)$, при котором $h_q^d(r) = q$.

Далее рассмотрим вопросы, связанные с изучением аффинных ограничений функций q-значной логики на линейные многообразия. В связи с тем, что $f_q^1(r) = q^{r+1}$, оценки, приведённые в теореме 1, следствиях 2 и 3, существенно упрощаются.

2. Асимптотические оценки числа линейных многообразий с аффинными ограничениями функции

Для $f \in P_q^n$ обозначим через $a_f(r)$ и $\hat{a}_f(r)$ число многообразий из множеств $\mathfrak{M}_q^n(r)$ и $\hat{\mathfrak{M}}_q^n(r)$ соответственно, на которых ограничения функции f являются аффинными. Очевидно, что справедливы соотношения $a_f(0) = \hat{a}_f(0) = q^n$. Для случая булевых функций также выполняются равенства $a_f(1) = 2^{n-1}(2^{n-1}-1)$ и $\hat{a}_f(1) = 2^{n-1}n$.

Теорема 2. При $n \to \infty$ для почти всех функций из P_q^n относительно параметров $a_f(r)$ и $\hat{a}_f(r)$, где r — целое неотрицательное число, удовлетворяющее условиям $1 \leqslant r \leqslant n$, справедливы соотношения

$$a_f(r) = \begin{cases} \dfrac{q^{n+1}\binom{n}{r}_q}{q^{q^r}}(1+\mathrm{o}(1)), & \text{если } r \leqslant \lfloor \log_q n + \log_q \log_q n \rfloor, \\ 0, & \text{если } r > \lceil \log_q n + \log_q \log_q n \rceil; \end{cases} \qquad (14)$$

$$\hat{a}_f(r) = \begin{cases} \dfrac{q^{n+1}\binom{n}{r}}{q^{q^r}}(1+\mathrm{o}(1)), & \text{если } r \leqslant \lfloor \log_q n \rfloor, \\ 0, & \text{если } r > \lceil \log_q n \rceil. \end{cases} \qquad (15)$$

Доказательство. Начнём со вторых частей формул (14) и (15).

Рассмотрим дискретные случайные величины $a_f(r)$ и $\hat{a}_f(r)$ и события $P_q^n(a_f(r) = 0) = \{f \in P_q^n \mid a_f(r) = 0\}$ и $P_q^n(\hat{a}_f(r) = 0) = \{f \in P_q^n \mid \hat{a}_f(r) = 0\}$.

Для вероятности события $\{a_f(r) = 0\}$ выполняются соотношения $\mathsf{P}(a_f(r) = 0) = \mathsf{P}(d_f(r) > 1) = 1 - \mathsf{P}(d_f(r) \leqslant 1)$. Из теоремы 1 следует неравенство $\mathsf{P}(d_f(r) \leqslant 1) \leqslant \phi_q^n(r)$, где $\phi_q^n(r) = \dfrac{q^{n+1}\binom{n}{r}_q}{q^{q^r}}$.

Используем представление гауссовых биномиальных коэффициентов, введенное при доказательстве следствия 3. Задействуем вспомогательные параметры $\left\langle {n \atop r} \right\rangle_q = \dfrac{\binom{n}{r}_q}{q^{(n-r)r}}$, которые назовём нормированными гауссовыми биномиальными коэффициентами. Из (12) вытекает справедливость неравенств

$$1 \leqslant \left\langle {n \atop r} \right\rangle_q < \mathrm{e}^{q/(q-1)^2} \leqslant \mathrm{e}^2, \tag{16}$$

а выражение для оценки доли функций, имеющих аффинные ограничения на многообразиях размерности r, примет вид

$$\phi_q^n(r) = \frac{q^{(n-r+1)(r+1)}\left\langle {n \atop r} \right\rangle_q}{q^{q^r}}. \tag{17}$$

Пусть $\breve{r}$ — наименьшее целое, удовлетворяющее неравенству $\breve{r} > \lceil \log_q n + \log_q \log_q n \rceil$, и, следовательно, $\breve{r} = \log_q n + \log_q \log_q n + \lambda$, где $1 \leqslant \lambda < 2$. Подставив $\breve{r}$ в (17) и используя оценку из (16), получим

$$\phi_q^n(\breve{r}) = \frac{q^{(n-\log_q n - \log_q \log_q n - \lambda + 1)(\log_q n + \log_q \log_q n + \lambda + 1)}\mathrm{e}^2}{q^{q^\lambda n \log_q n}}.$$

Переходя к пределу при $n \to \infty$, имеем $\lim\limits_{n\to\infty} \phi_q^n(\breve{r}) = 0$, из чего следует справедливость второй части утверждения (14) для $\breve{r}$, а значит, и для всех бóльших r.

Аналогично доказывается вторая часть формулы (15). При этом в соответствии с (13) для оценки доли функций, имеющих аффинные подфункции, может быть использовано выражение

$$\hat{\phi}_q^n(r) = \frac{q^{n+1+(\log_q n + \log_q (n-r+1) - \log_q r)r/2}}{q^{q^r}}.$$

Для доказательства первых частей формул (14) и (15) найдём математические ожидания и дисперсии величин $a_f(r)$ и $\hat{a}_f(r)$. Введём вспомогательную случайную величину $a_{f|_M}$, положив

$$a_{f|_M} = \begin{cases} 1, & \text{если } d_{f|_M} \leqslant 1, \\ 0, & \text{если } d_{f|_M} > 1. \end{cases}$$

Для математического ожидания величины $a_{f|_M}$, с учётом (6), выполняются соотношения $\mathsf{E}(a_{f|_M}) = \mathsf{P}(a_{f|_M} = 1) = \mathsf{P}(d_{f|_M} \leqslant 1) = \dfrac{q^{r+1}}{q^{q^r}}$. Применив формулу математического ожидания суммы случайных величин, получим

$$\mathsf{E}(a_f(r)) = \sum_{M \in \mathfrak{M}_q^n(r)} \mathsf{E}(a_{f|_M}) = \frac{q^{n+1}\binom{n}{r}_q}{q^{q^r}}; \tag{18}$$

$$\mathsf{E}(\hat{a}_f(r)) = \sum_{M \in \hat{\mathfrak{M}}_q^n(r)} \mathsf{E}(a_{f|_M}) = \frac{q^{n+1}\binom{n}{r}}{q^{q^r}}. \tag{19}$$

Используя нижние оценки в (12) и (13), нетрудно показать, что при $n \to \infty$ в случае $r \leqslant \lfloor \log_q n + \log_q \log_q n \rfloor$ верно неравенство $\mathsf{E}(a_f(r)) > 1$, а для $r \leqslant \lfloor \log_q n \rfloor$ имеем $\mathsf{E}(\hat{a}_f(r)) > 1$.

Для определения значений дисперсии докажем вспомогательную лемму.

Лемма 2. Пусть s является целым неотрицательным числом, удовлетворяющим условию $0 < s \leqslant r$; $m_q^n(r,s)$ — число пар многообразий размерности r, пересечением которых является многообразие размерности $r-s$, и $\hat{m}_q^n(r,s)$ — аналогичный параметр для многообразий, заданных фиксацией переменных. Тогда справедливы соотношения

$$m_q^n(r,s) = \begin{cases} \dfrac{1}{2}\, q^{n-r+s+s^2} \begin{pmatrix} n \\ r \end{pmatrix}_q \begin{pmatrix} r \\ s \end{pmatrix}_q \begin{pmatrix} n-r \\ s \end{pmatrix}_q, & \text{если } s \leqslant n-r, \\ 0, & \text{если } s > n-r; \end{cases} \tag{20}$$

$$\hat{m}_q^n(r,s) = \begin{cases} \dfrac{1}{2}\, q^{n-r+s} \begin{pmatrix} n \\ r \end{pmatrix} \begin{pmatrix} r \\ s \end{pmatrix} \begin{pmatrix} n-r \\ s \end{pmatrix}, & \text{если } s \leqslant n-r, \\ 0, & \text{если } s > n-r. \end{cases} \tag{21}$$

Доказательство. Используя метод подсчёта числа подпространств [3], получим

$$m_q^n(r,s) = \begin{cases} \dfrac{1}{2}\, m_q^n(r-s) \displaystyle\prod_{i=r-s}^{r-1} \frac{q^n - q^i}{q^r - q^i} \prod_{j=r-s}^{r-1} \frac{q^n - q^{s+j}}{q^r - q^j}, & \text{если } s \leqslant n-r, \\ 0, & \text{если } s > n-r. \end{cases} \tag{22}$$

Из (7) имеем $m_q^n(r-s) = q^{n-r+s} \begin{pmatrix} n \\ r-s \end{pmatrix}_q$. Преобразуем другие сомножители, выделив гауссовы биномиальные коэффициенты, следующим образом:

$$\prod_{i=r-s}^{r-1} \frac{q^n - q^i}{q^r - q^i} = \prod_{i=0}^{s-1} \frac{q^{n-r+s-i} - 1}{q^{s-i} - 1} = \begin{pmatrix} n-r+s \\ s \end{pmatrix}_q,$$

$$\prod_{j=r-s}^{r-1} \frac{q^n - q^{s+j}}{q^r - q^j} = q^{s^2} \prod_{j=r-s}^{r-1} \frac{q^{n-s} - q^j}{q^r - q^j} = q^{s^2} \prod_{j=0}^{s-1} \frac{q^{n-r-j} - 1}{q^{s-j} - 1} = q^{s^2} \begin{pmatrix} n-r \\ s \end{pmatrix}_q.$$

Подставив полученные выражения в (22) и используя соотношение для гауссовых биномиальных коэффициентов $\begin{pmatrix} n \\ r-s \end{pmatrix}_q \begin{pmatrix} n-r+s \\ s \end{pmatrix}_q = \begin{pmatrix} n \\ r \end{pmatrix}_q \begin{pmatrix} r \\ s \end{pmatrix}_q$, придём к (20).

В свою очередь, величина $\hat{m}_q^n(r,s)$ задается системой вида

$$\hat{m}_q^n(r,s) = \begin{cases} \dfrac{1}{2}\, \hat{m}_q^n(r-s) \begin{pmatrix} n-r+s \\ s \end{pmatrix} \begin{pmatrix} n-r \\ s \end{pmatrix}, & \text{если } s \leqslant n-r, \\ 0, & \text{если } s > n-r. \end{cases} \tag{23}$$

Подставив в (23) выражение $\hat{m}_q^n(r-s)$ из (9) и используя аналогичное соотношение для обыкновенных биномиальных коэффициентов, получим (21).

Лемма 2 доказана. ■

Вернёмся к доказательству теоремы. Используя формулу дисперсии суммы случайных величин, получим выражение дисперсии для $a_f(r)$:

$$\mathsf{D}(a_f(r)) = \sum_{M \in \mathfrak{M}_q^n(r)} \mathsf{D}(a_{f|_M}) + 2 \sum_{M_i, M_j \in \mathfrak{M}_q^n(r), i<j} \operatorname{cov}(a_{f|_{M_i}}, a_{f|_{M_j}}). \tag{24}$$

Поскольку $\mathsf{E}(a^2_{f|_M}) = \mathsf{E}(a_{f|_M})$, имеем соотношение $\mathsf{D}(a_{f|_M}) = \dfrac{q^{r+1}}{q^{q^r}} - \dfrac{q^{2r+2}}{q^{2q^r}}$.

Как следует из леммы 2, пересечением многообразий M_i и M_j размерности r может быть многообразие размерности $r-s$, где $1 \leqslant s \leqslant \min\{r, n-r\}$, или пустое множество. В первом случае для ковариации справедливо $\operatorname{cov}(a_{f|_{M_i}}, a_{f|_{M_j}}) = \dfrac{q^{r+s+1}}{q^{2q^r-q^{r-s}}} - \dfrac{q^{2r+2}}{q^{2q^r}}$, в случае пустого множества имеем $\operatorname{cov}(a_{f|_{M_i}}, a_{f|_{M_j}}) = 0$. С учётом полученного значения дисперсии для $a_{f|_M}$ и леммы 2 преобразуем (24) следующим образом:

$$\begin{aligned}
\mathsf{D}(a_f(r)) &= m_q^n(r)\Big(\frac{q^{r+1}}{q^{q^r}} - \frac{q^{2r+2}}{q^{2q^r}}\Big) + 2\sum_{s=1}^{\min\{r,n-r\}} m_q^n(r,s)\Big(\frac{q^{r+s+1}}{q^{2q^r-q^{r-s}}} - \frac{q^{2r+2}}{q^{2q^r}}\Big) = \\
&= \frac{q^{n+1}\binom{n}{r}_q}{q^{2q^r}}\left[(q^{q^r} - q^{r+1}) + \sum_{s=1}^{\min\{r,n-r\}} q^{2s+s^2}(q^{q^{r-s}} - q^{r-s+1})\binom{r}{s}_q\binom{n-r}{s}_q\right] = \qquad (25)\\
&= \frac{q^{n+1}\binom{n}{r}_q}{q^{2q^r}} \sum_{s=0}^{\min\{r,n-r\}} q^{2s+s^2}(q^{q^{r-s}} - q^{r-s+1})\binom{r}{s}_q\binom{n-r}{s}_q.
\end{aligned}$$

Для оценки дисперсии вновь потребуются нормированные гауссовы биномиальные коэффициенты. После перехода к ним в (25) получим

$$\mathsf{D}(a_f(r)) = \frac{q^{(n-r+1)(r+1)}\left\langle{n \atop r}\right\rangle_q}{q^{2q^r}} \sum_{s=0}^{\min\{r,n-r\}} q^{(n-s+2)s}(q^{q^{r-s}} - q^{r-s+1})\left\langle{r \atop s}\right\rangle_q\left\langle{n-r \atop s}\right\rangle_q. \qquad (26)$$

При изменении параметра s на отрезке от 0 до r, где $1 \leqslant r \leqslant \lfloor \log_q n + \log_q \log_q n \rfloor < n/2$, исследование выражений $(n-s+2)s + q^{r-s}$ и $\left\langle{n-r \atop s}\right\rangle_q$ показывает, что они принимают максимальное значение при $s = r$. Кроме того, из рекуррентного соотношения для нормированных гауссовых биномиальных коэффициентов $\left\langle{n \atop r}\right\rangle_q = \left\langle{n-1 \atop r}\right\rangle_q + \dfrac{1}{q^{n-r}}\left\langle{n-1 \atop r-1}\right\rangle_q$ следует неравенство $\left\langle{n-r \atop r}\right\rangle_q < \left\langle{n \atop r}\right\rangle_q$, а из (16) вытекает оценка $\sum\limits_{s=0}^{r}\left\langle{r \atop s}\right\rangle_q < \mathrm{e}^2(r+1)$. С учётом перечисленных замечаний к выражению (26) имеем неравенство

$$\mathsf{D}(a_f(r)) < \frac{q^{(n-r+1)(r+1)}\left\langle{n \atop r}\right\rangle_q^2}{q^{2q^r}} q^{(n-r+2)r+1}\mathrm{e}^2(r+1).$$

Из последнего с очевидностью следует

$$\mathsf{D}(a_f(r)) \leqslant \mathsf{E}^2(a_f(r))q^{-n+2r}\mathrm{e}^2(r+1).$$

Для оценки доли функций, на которых случайная величина $a_f(r)$ отклоняется от значения своего математического ожидания на величину, большую или равную ε, воспользуемся неравенством Чебышева для случайной величины $a_f(r)$ вида

$$\mathsf{P}(|a_f(r) - \mathsf{E}(a_f(r))| \geqslant \varepsilon) \leqslant \frac{\mathsf{D}(a_f(r))}{\varepsilon^2}.$$

Положим $\varepsilon = \mathsf{E}(a_f(r))/n$. Тогда при $1 \leqslant r \leqslant \lfloor \log_q n + \log_q \log_q n \rfloor < n/2$ имеем оценку

$$\mathsf{P}(|a_f(r) - \mathsf{E}(a_f(r))| \geqslant \mathsf{E}(a_f(r))/n) \leqslant \psi_q^n(r),$$

где $\psi_q^n(r) = q^{-n+2(r+\log_q n)}\mathrm{e}^2(r+1)$. При изменении r в указанных границах, переходя к пределу при $n \to \infty$, получим $\lim\limits_{n\to\infty} \psi_q^n(r) = 0$, из чего следует справедливость второй части соотношения (14).

Первая часть формулы (15) доказывается аналогично. При этом, опираясь на результат леммы 2 относительно многообразий из множества $\hat{\mathfrak{M}}_q^n(r)$, можно получить следующее значение дисперсии для $\hat{a}_f(r)$:

$$\mathsf{D}(\hat{a}_f(r)) = \frac{q^{n+1}\binom{n}{r}}{q^{2q^r}}\left((q^{q^r} - q^{r+1}) + \sum_{s=1}^{\min\{r,n-r\}} q^{2s}(q^{q^{r-s}} - q^{r-s+1})\binom{r}{s}_q\binom{n-r}{s}_q\right).$$

При изменении параметра s на отрезке от 1 до r, где $1 \leqslant r \leqslant \lceil \log_q n \rceil < n/2$, исследование выражения $2s + q^{r-s}$ показывает, что оно принимает максимальное значение при $s = 1$. Используя формулу для биномиальных коэффициентов $\binom{n}{r} = \sum\limits_{s=0}^{r}\binom{r}{s}\binom{n-r}{s}$, придём к оценке дисперсии $\mathsf{D}(\hat{a}_f(r)) \leqslant \mathsf{E}^2(\hat{a}_f(r))\left(q^{q^r-n-1}/\binom{n}{r} + q^{q^{r-1}-n+1}\right)$, которая позволяет воспользоваться неравенством Чебышева для завершения доказательства (15).

Теорема 2 доказана. ■

Для $f \in P_q^n$ обозначим через r_f и $\hat{r}_f$ максимальную размерность многообразия из множеств $\mathfrak{M}_q^n(r)$ и $\hat{\mathfrak{M}}_q^n(r)$ соответственно, на котором ограничение функции f является аффинным.

Следствие 4. При $n \to \infty$ для почти всех функций из P_q^n справедливы оценки

$$\lfloor \log_q n + \log_q \log_q n \rfloor \leqslant r_f \leqslant \lceil \log_q n + \log_q \log_q n \rceil; \tag{27}$$

$$\lfloor \log_q n \rfloor \leqslant \hat{r}_f \leqslant \lceil \log_q n \rceil. \tag{28}$$

В случае целых значений логарифмов получим $r_f = \log_q n + \log_q \log_q n$ и $\hat{r}_f = \log_q n$.

В случае булевых функций в [6] при $n \to \infty$ для почти всех функций из P_2^n получена верхняя оценка максимальной размерности интервала в множестве единиц функции, равная $\lfloor \log_q n \rfloor + 1$ (термин «r-мерный интервал» соответствует многообразию размерности r, заданному фиксацией $n - r$ переменных). Последнее означает, что при достаточно больших значениях n для случайно выбранной булевой функции от n переменных с вероятностью близкой к 1 не найдётся единичной подфункции-константы, заданной фиксацией менее чем $n - \lfloor \log_q n \rfloor - 1$ переменных. Из данной оценки с очевидностью следует, что при $n \to \infty$ для почти всех функций из P_2^n справедлива верхняя оценка параметра $\hat{r}_f$ вида $\lfloor \log_q n \rfloor + 2$. В [7] показано, при $n \to \infty$ для почти всех функций из P_2^n имеет место нижняя оценка параметра $\hat{r}_f$ вида $\lfloor \log_2 n \rfloor - 1$. Оценка (28) настоящей работы уточняет и обобщает указанный результат на случай функций q-значной логики. Предпринятая в [8] попытка получить верхние асимптотические оценки параметра r_f для почти всех булевых функций не привела к успеху из-за ошибок в доказательстве. В [7, 8] величины $n - \hat{r}_f$ и $n - r_f$ названы уровнем аффинности и обобщённым уровнем аффинности функции f.

В работе [9] детально рассмотрено поведение параметров r_f и $\hat{r}_f$ квадратичных форм булевых функций.

Заключение

Результаты первой части работы позволяют говорить о поведении степени ограничений функций q-значной логики от n переменных на линейные многообразия векторного пространства $\mathbb{F}_q^n$ для произвольных значений n. В частности, из теоремы 1 можно получить верхние оценки максимальной размерности многообразия, на котором ограничение функции и, в частности, подфункция являются аффинными, для преобладающего числа функций q-значной логики. В таблице приведены подобные оценки для булевых функций от 4 до 16 переменных.

n	4	5	6	7	8	9	10	11	12	13	14	15	16
r_f	$\leqslant 3$	$\leqslant 3$	$\leqslant 4$	$\leqslant 4$	$\leqslant 4$	$\leqslant 5$	$\leqslant 5$	$\leqslant 5$	$\leqslant 5$	$\leqslant 5$	$\leqslant 6$	$\leqslant 6$	$\leqslant 6$
$\hat{r}_f$	$\leqslant 2$	$\leqslant 3$	$\leqslant 3$	$\leqslant 3$	$\leqslant 4$	$\leqslant 4$	$\leqslant 4$	$\leqslant 4$	$\leqslant 4$	$\leqslant 4$	$\leqslant 4$	$\leqslant 4$	$\leqslant 4$

Из таблицы следует, что преимущество использования произвольных многообразий по сравнению с фиксацией переменных при решении задач линеаризации начинает сказываться при $n \geqslant 9$.

Вторая часть работы даёт представление об асимптотике появления многообразий, на которых ограничения и, в частности, подфункции являются аффинными для почти всех функций q-значной логики при $n \to \infty$. Стоит, однако, заметить, что, как видно из таблицы, полученные в следствии 4 верхние оценки параметров r_f и $\hat{r}_f$ справедливы для преобладающего числа булевых функций при всех $n \geqslant 4$, за исключением оценки $\hat{r}_f$ при $n = 8$, который требует дополнительного изучения.

Для оценки долей функций, обладающих заданными свойствами, использован аппарат вероятностных методов в дискретной математике. Некоторые промежуточные результаты, как, например, определение числа функций q-значной логики степени не выше заданной и оценки гауссовых и обыкновенных биномиальных коэффициентов, имеют самостоятельное значение.

ЛИТЕРАТУРА

1. *Дьедонне Ж.* Линейная алгебра и элементарная геометрия. М.: Наука, 1972. 336 с.
2. *Глухов М. М., Шишков А. Б.* Математическая логика. Дискретные функции. Теория алгоритмов. СПб.: Лань, 2012. 416 с.
3. *Глухов М. М., Елизаров В. П., Нечаев А. А.* Алгебра. Т. II. М.: Гелиос АРВ, 2003. 416 с.
4. *Феллер В.* Введение в теорию вероятностей и её приложения. Т. 1. М.: Мир, 1984. 528 с.
5. *Сачков В. Н.* Комбинаторные методы дискретной математики. М.: Наука, 1977. 320 с.
6. *Журавлев Ю. И.* Теоретико-множественные методы в алгебре логики // Проблемы кибернетики. 1962. Вып. 8. С. 5–44.
7. *Логачев О. А.* О значениях уровня аффинности для почти всех булевых функций // Прикладная дискретная математика. 2010. № 3(9). С. 17–21.
8. *Буряков М. Л.* Асимптотические оценки уровня аффинности для почти всех булевых функций // Дискретная математика. 2008. Т. 20. Вып. 3. С. 73–79.
9. *Черемушкин А. В.* Об оценке уровня аффинности квадратичных форм // Дискретная математика. 2017. Т. 29. Вып. 1. С. 114–125.

REFERENCES

1. *D'edonne Zh.* Lineynaya algebra i elementarnaya geometriya [Linear Algebra and Elementary Geometry]. Moscow, Nauka Publ., 1972. 336 p. (in Russian)

2. *Glukhov M. M. and Shishkov A. B.* Matematicheskaya logika. Diskretnye funktsii. Teoriya algoritmov [Mathematical Logic. Discrete Functions. Theory of Algorithms]. St. Petersburg, Lan Publ., 2012. 416 p. (in Russian)
3. *Glukhov M. M., Elizarov V. P., and Nechaev A. A.* Algebra [Algebra]. Vol. II. Moscow, Gelios ARV Publ., 2003. 416 p. (in Russian)
4. *Feller V.* Vvedenie v teoriyu veroyatnostey i ee prilozheniya [An Introduction to Probability Theory and its Applications]. Vol. 1. Moscow, Mir Publ., 1984. 528 p. (in Russian)
5. *Sachkov V. N.* Kombinatornye metody diskretnoy matematiki [Combinatorial Methods in Discrete Mathematics]. Moscow, Nauka Publ., 1977. 320 p.. (in Russian)
6. *Zhuravlev Yu. I.* Teoretiko-mnozhestvennyye metody v algebre logiki [Set-theoretical methods in the algebra of logic]. Problemy Kibernetiki, 1962, no. 8, pp. 5–44. (in Russian)
7. *Logachev O. A.* O znacheniyakh urovnya affinnosti dlya pochti vsekh bulevykh funktsiy [On values of affinity level for almost all Boolean functions]. Prikladnaya Diskretnaya Matematika, 2010, no. 3(9), pp. 17–21. (in Russian)
8. *Buryakov M. L.* Asimptoticheskiye otsenki urovnya affinnosti dlya pochti vsekh bulevykh funktsiy [Asymptotic bounds for the affinity level for almost all Boolean functions]. Discrete Math. Appl., 2008, vol. 20, no. 3, pp. 73–79. (in Russian)
9. *Cheremushkin A. V.* Ob otsenke urovnya affinnosti kvadratichnykh form [Estimating the level of affinity of a quadratic form]. Discrete Math. Appl., 2017, vol. 29, no. 1, pp. 114–125. (in Russian)

УДК 511.528

О ДВОИЧНЫХ РЕШЕНИЯХ СИСТЕМ УРАВНЕНИЙ[1]

А. В. Селиверстов

Институт проблем передачи информации им. А. А. Харкевича Российской академии наук, г. Москва, Россия

Решение называется двоичным, если каждая переменная равна нулю или единице. Известно, что трудно найти двоичное решение системы алгебраических уравнений, коэффициенты которых являются целыми числами с малыми абсолютными значениями. Целью работы является обоснование эффективного вероятностного сведения системы к одному новому уравнению в случае, когда существует небольшая разница между числом двоичных решений первого уравнения и числом двоичных решений всей системы. Более того, если первое уравнение линейное, то существует алгоритм псевдополиномиального времени для проверки правильности такого сведения к новому уравнению в общем случае.

Ключевые слова: *алгебраическое уравнение, вероятностный алгоритм, вычислительная сложность.*

DOI 10.17223/20710410/45/3

ON BINARY SOLUTIONS TO SYSTEMS OF EQUATIONS

A. V. Seliverstov

Institute for Information Transmission Problems of the Russian Academy of Sciences (Kharkevich Institute), Moscow, Russia

E-mail: slvstv@iitp.ru

A solution is called binary if each variable is equal to either zero or one. It is well known that it is hard to find a binary solution to the system of algebraic equations in which the coefficients are integers with small absolute values. The aim of the paper is to propose an effective probabilistic reduction of the system to a new equation when there is a small difference between the number of binary solutions to the first equation and the number of binary solutions to the entire system. The proposed method is based on replacing the given system of equations with a linear combination of these equations. Coefficients are random integers that are independently and uniformly distributed over the segment from zero to some upper bound. The bound depends on the number of redundant binary solutions to the first equation that do not serve as solutions to the entire system. The proof uses the Schwartz — Zippel lemma. Moreover, if the first equation is linear, then there exists a pseudo-polynomial time algorithm to check the correctness of the reduction to the new equation in the general case.

Keywords: *algebraic equation, probabilistic algorithm, computational complexity.*

[1]Работа поддержана грантом РФФИ № 18-29-13037.

Введение

Рассмотрим поиск двоичных решений системы алгебраических уравнений с целыми коэффициентами. Иначе такие решения называют булевыми. Рассмотрим сведение исходной системы уравнений к системе с меньшим числом уравнений так, чтобы максимальная степень уравнений не возрастала, а коэффициенты новых уравнений были целыми числами, абсолютные величины которых не слишком велики по сравнению с коэффициентами в исходных уравнениях.

Ограничение на степень уравнения существенно, поскольку любая система из m уравнений вида $\ell_k(\mathbf{x}) = 0$ эквивалентна над полем вещественных чисел одному уравнению $\ell_1^2(\mathbf{x}) + \ldots + \ell_m^2(\mathbf{x}) = 0$. Здесь через $\mathbf{x}$ обозначен список переменных $x_1, \ldots, x_n$. Ограничение на абсолютную величину коэффициентов тоже существенно, поскольку при достаточно быстро возрастающей последовательности чисел γ_k эта система имеет те же двоичные решения, что и одно уравнение $\gamma_1\ell_1(\mathbf{x}) + \ldots + \gamma_m\ell_m(\mathbf{x}) = 0$. Сведение системы линейных уравнений к одному линейному уравнению, имеющему те же двоичные решения, рассмотрено в работе [1].

Для одного линейного уравнения некоторое двоичное решение можно вычислить за псевдополиномиальное время методом динамического программирования, рассмотренным в нескольких работах Р. Э. Беллмана и (независимо) Дж. Б. Данцига в 1956–1957 гг. История развития этого метода до 1986 г. описана в [2]. Позже В. В. Смолев заметил, что если модули коэффициентов при линейных членах уравнения целые и отличаются друг от друга не более чем на небольшую величину, то время работы алгоритма существенно зависит от максимума этой величины, а не от максимума модулей коэффициентов. Более того, за псевдополиномиальное время можно найти число всех его двоичных решений [3]. Недавно опубликовано много работ, в которых улучшаются оценки памяти и времени, необходимых для решения этих задач [4–8]. Известна связь вычислительной сложности поиска двоичного решения линейного уравнения со сложностью некоторых задач, разрешимых за полиномиальное время [9].

В общем случае задача распознавания существования некоторого двоичного решения у системы линейных уравнений с коэффициентами из множества $\{-1, 0, 1\}$ является NP-полной [2]. Поэтому сведение такой системы уравнений к одному линейному уравнению с малыми коэффициентами возможно лишь при выполнении некоторых дополнительных условий.

Сейчас большое внимание уделяется разработке *генерических* алгоритмов полиномиального времени [10, 11]. Их можно рассматривать как частный случай эвристических алгоритмов, когда в типичном случае результатом работы алгоритма служит правильный ответ, но на малой доле входов, стремящейся к нулю при увеличении длины входа, алгоритм может отказаться от вычислений; при этом выдаётся сообщение об отказе. Аналогично определяются недетерминированные генерические алгоритмы. При этом некоторые алгоритмически трудные задачи остаются трудными и для генерических алгоритмов [12, 13].

Одно из возможных приложений рассматриваемых алгоритмов, стимулировавшее развитие новых методов поиска двоичных решений систем уравнений, — это решение биоинформатической задачи сравнения так называемых хромосомных структур с паралогами (то есть такими генами, которые трудно различить выравниванием последовательностей) и поиска оптимального соответствия между ними [14]. К поиску двоичных решений системы уравнений легко сводится также поиск решений диофантовых уравнений в произвольной ограниченной области. В этом случае целые числа из некоторого отрезка представимы последовательностью нулей и единиц [2].

1. Результаты

Теорема 1. Даны система, состоящая из $m \geqslant 2$ алгебраических уравнений $\ell_k(\mathbf{x}) = 0$, и целое число r из отрезка от единицы до $m-1$. Пусть подсистема, состоящая из первых r уравнений $\ell_1(\mathbf{x}) = 0, \ldots, \ell_r(\mathbf{x}) = 0$, имеет не более μ избыточных двоичных решений, которые не служат решениями всей системы. Существуют такие целые числа $\gamma_{r+1}, \ldots, \gamma_m$ из отрезка от нуля до μ, что каждое двоичное решение новой системы алгебраических уравнений $\ell_1(\mathbf{x}) = 0, \ldots, \ell_r(\mathbf{x}) = 0$ и $\gamma_{r+1}\ell_{r+1}(\mathbf{x}) + \ldots + \gamma_m\ell_m(\mathbf{x}) = 0$ служит решением исходной системы. Более того, для любого $\varepsilon > 0$ если случайные целые числа $\gamma_{r+1}, \ldots, \gamma_m$ независимы и равномерно распределены на отрезке от нуля до $\lceil \mu/\varepsilon \rceil$, то указанное свойство выполнено с вероятностью большей разности $1 - \varepsilon$.

Доказательство. Обозначим через $\mathcal{M}$ множество тех двоичных решений рассматриваемой подсистемы, которые не являются решениями всей системы уравнений:

$$\mathcal{M} = \{\mathbf{x} \in \{0,1\}^n : \ell_1(\mathbf{x}) = 0 \wedge \cdots \wedge \ell_r(\mathbf{x}) = 0 \wedge \exists k \leqslant m\, (\ell_k(\mathbf{x}) \neq 0)\}.$$

Мощность $|\mathcal{M}|$ не превышает числа μ. Определим многочлен

$$f(y_{r+1}, \ldots, y_m) = \prod_{\mathbf{x} \in \mathcal{M}} \left(\sum_{k=r+1}^{m} \ell_k(\mathbf{x}) y_k \right).$$

Если множество $\mathcal{M}$ пустое, то полагаем $f = 1$. В этом случае числа $\gamma_{r+1}, \ldots, \gamma_m$ могут быть любыми, например равными нулю: $\gamma_{r+1} = \ldots = \gamma_m = 0$.

Пусть множество $\mathcal{M}$ непустое. Если некоторая последовательность целых чисел $\gamma_{r+1}, \ldots, \gamma_m$ достаточно быстро возрастает, то значение $f(\gamma_{r+1}, \ldots, \gamma_m)$ отлично от нуля. Следовательно, многочлен f не равен нулю тождественно. С другой стороны, выполнено неравенство $\deg f \leqslant \mu$. По лемме Шварца — Зиппеля [15] существуют искомые целые числа $\gamma_{r+1}, \ldots, \gamma_m$ из отрезка от нуля до μ, для которых $f(\gamma_{r+1}, \ldots, \gamma_m) \neq 0$, и доля таких наборов чисел не меньше разности $1 - \varepsilon$ при $\varepsilon > 0$. ■

Замечание 1. Теорема 1 остаётся справедливой, если рассматривать не $(0,1)$-решения, а (α, β)-решения для произвольных чисел $\alpha \neq \beta$.

Далее ограничимся случаем, когда $r = 1$. В этом случае теорема 1 позволяет свести систему уравнений к системе двух уравнений. Переход от двух уравнений к одному, равному их линейной комбинации и имеющему те же двоичные решения, приводит к относительно небольшому увеличению коэффициентов (хотя в общем случае трудно свести систему из многих уравнений к одному при таком ограничении). Далее рассмотрено сведение системы уравнений к одному, имеющему те же двоичные решения, которое для краткости называется *новым уравнением*.

Теорема 2. Даны система из $m \geqslant 2$ алгебраических уравнений $\ell_k(\mathbf{x}) = 0$ и вещественное число $\varepsilon > 0$. Пусть первое уравнение $\ell_1(\mathbf{x}) = 0$ имеет не более μ избыточных двоичных решений, которые не служат решениями всей системы. Если случайные целые числа $\gamma_2, \ldots, \gamma_m$ независимы и равномерно распределены на отрезке от нуля до $\lceil \mu/\varepsilon \rceil$, то с вероятностью большей разности $1 - \varepsilon$ каждое двоичное решение нового уравнения $\ell_1(\mathbf{x}) + \lambda\,(\gamma_2\ell_2(\mathbf{x}) + \ldots + \gamma_m\ell_m(\mathbf{x})) = 0$ служит решением исходной системы. Здесь коэффициент λ равен произвольному целому числу, превосходящему сумму модулей всех коэффициентов многочлена ℓ_1.

Доказательство. Применяя теорему 1 для случая $r = 1$, сведём исходную систему уравнений к системе двух уравнений вида

$$\begin{cases} \ell_1(\mathbf{x}) = 0, \\ \gamma_2\ell_2(\mathbf{x}) + \ldots + \gamma_m\ell_m(\mathbf{x}) = 0 \end{cases}$$

с тем же набором двоичных решений. В каждой точке с координатами из множества $\{0,1\}$ модуль значения функции ℓ_1 меньше числа λ. По модулю λ новое уравнение эквивалентно первому уравнению. Поэтому каждое двоичное решение нового уравнения служит решением первого уравнения $\ell_1 = 0$. Следовательно, оно служит решением и для второго уравнения, равного линейной комбинации (с рациональными коэффициентами) первого и нового уравнений. ■

Замечание 2. Требование, чтобы число μ было маленьким, существенно для практического применения рассмотренной сводимости, поскольку при больших значениях μ коэффициенты γ_k тоже могут быть большими. С другой стороны, число μ является лишь верхней границей; не требуется знание точного значения разности чисел двоичных решений у первого уравнения и у всей системы уравнений. Кроме того, небольшое увеличение значения μ может быть удобно для вычислений. Если $\lceil \mu/\varepsilon \rceil = 2^\nu - 1$ для целого ν, то равномерно распределённая случайная величина γ_k легко получается конкатенацией независимых симметричных бернуллиевских случайных величин, равных цифрам двоичной записи. Создание равномерного распределения на перестановках (то есть на множестве из $\nu!$ элементов) на основе бернуллиевских случайных величин обсуждается в [16].

Теорема 3. Дан многочлен $f(x_1, \ldots, x_n)$ степени d, имеющий m мономов. Обозначим через $\ell(y_1, \ldots, y_m)$ линейную форму, коэффициенты которой равны коэффициентам многочлена f, расставленным в любом порядке. Для любого числа α число двоичных решений уравнения $f(\mathbf{x}) = \alpha$ не более чем в $(2^d - 1)^m$ раз превосходит число двоичных решений линейного уравнения $\ell(\mathbf{y}) = \alpha$.

Доказательство. Каждое двоичное решение уравнения $f(\mathbf{x}) = \alpha$ соответствует двоичному решению уравнения $\ell(\mathbf{y}) = \alpha$, где значение переменной y_k равно значению соответствующего монома. При этом равенство $y_k = 0$ соответствует равенству нулю хотя бы одной из переменных, входящих в k-й моном. Каждый моном зависит самое большее от d переменных и обнуляется на самое большее $2^d - 1$ наборах двоичных значений этих переменных. Поэтому каждое решение уравнения $\ell(\mathbf{y}) = \alpha$ соответствует не более чем $(2^d - 1)^m$ двоичным решениям уравнения $f(\mathbf{x}) = \alpha$. ■

Пример 1. Рассмотрим многочлен $f = x_1x_2 + x_3x_4 + \ldots + x_{2m-1}x_{2m}$. Линейное уравнение $y_1 + \ldots + y_m = 0$ имеет только нулевое двоичное решение. По теореме 3 число двоичных решений уравнения $f(\mathbf{x}) = 0$ не превосходит числа 3^m, тогда как число всех наборов двоичных значений переменных $\mathbf{x}$ равно 4^m. Рассуждая, как при доказательстве теоремы 3, легко видеть, что эта оценка числа двоичных решений точная.

2. Обсуждение

Рассмотрим случай, когда первое уравнение системы линейное, а число его двоичных решений известно (оно может быть вычислено за псевдополиномиальное время [3]). Даже в этом случае трудно вычислить оптимальное значение μ в теореме 2. Однако если угадано некоторое значение μ и для этого значения случайно выбраны целые коэффициенты γ_k из отрезка от нуля до $\lceil \mu/\varepsilon \rceil$, то *a posteriori* можно проверить правильность выбора μ и коэффициентов γ_k, следовательно, проверить корректность вычисления. Для этого надо оценить сверху число решений полученного нового уравнения. Очевидно, любое решение системы служит решением этого уравнения. Поэтому при любом выборе коэффициентов γ_k число двоичных решений нового уравнения не меньше, чем число двоичных решений исходной системы. Если разность числа двоичных решений первого и нового уравнений превосходит число μ или отрицательна,

то сведение системы к новому уравнению должно быть признано некорректным. Причиной неудачи может быть как неправильно угаданное значение μ, так и неудачная реализация случайных коэффициентов γ_k. Если же эта разность положительна и не превосходит числа μ, то с вероятностью не меньше разности $1-\varepsilon$ значение μ корректно, следовательно, новое уравнение тоже корректное.

Если все уравнения системы линейные, то новое уравнение тоже линейное и число его двоичных решений можно вычислить за псевдополиномиальное время. Так, по аналогии с генерическими алгоритмами, можно получить вероятностный алгоритм псевдополиномиального времени, который для систем линейных уравнений на большой доле входов с большой вероятностью выдаёт корректное новое уравнение, а на оставшейся доле входов, а также с малой вероятностью на любом входе выдаёт сообщение об отказе от вычисления. Такой алгоритм может сделать ошибку, не выдавая никакого предупреждения, но вероятность такой ошибки можно ограничить сверху сколь угодно малым положительным числом.

Если (алгебраическое) уравнение нелинейное, но имеет мало мономов, то число его двоичных решений можно оценить сверху посредством теоремы 3, рассматривая мономы как независимые переменные. При этом число решений линейного уравнения снова оценивается псевдополиномиальным алгоритмом. Такая оценка будет точнее для уравнений с меньшим числом мономов. Поскольку рассматриваются только двоичные решения, такое уравнение можно считать мультилинейным. В случае, когда система имеет двоичные решения, линейность первого уравнения существенна для вычисления верхней оценки оптимального значения μ в теореме 2 за псевдополиномиальное время. С другой стороны, несмотря на грубость оценки числа двоичных решений, когда таковые существуют, теорема 3 часто позволяет точно подтвердить отсутствие двоичных решений у нового уравнения и, следовательно, их отсутствие у исходной системы нелинейных уравнений.

Приведём пример, когда система уравнений имеет столько же двоичных решений, что и первое уравнение.

Пример 2. Рассмотрим систему двух уравнений

$$\begin{cases} x_1 + \ldots + x_n = 1, \\ \sum\limits_{j<k} x_j x_k = 0. \end{cases}$$

Двоичные решения первого уравнения имеют вид $(0, \ldots, 0, 1, 0, \ldots, 0)$, где ровно одна переменная равна единице, а прочие равны нулю. Каждое из них служит решением второго уравнения.

Пример 3. Рассмотрим систему двух уравнений

$$\begin{cases} x_1 + \ldots + x_n = 2, \\ x_1 x_2 + x_2 x_3 + \ldots + x_{n-1} x_n = 0. \end{cases}$$

Двоичные решения первого уравнения таковы, что ровно две переменные равны единице. Число таких решений равно $n(n-1)/2$. Двоичное решение первого уравнения служит решением второго уравнения при условии, что равные единице переменные не оказались соседними. Поэтому число избыточных двоичных решений первого уравнения равно $n-1$.

ЛИТЕРАТУРА

1. *Seliverstov A. V.* Binary solutions to some systems of linear equations // OPTA 2018. Communications in Computer and Information Science. V. 871. Cham: Springer, 2018. P. 183–192. https://doi.org/10.1007/978-3-319-93800-4_15
2. *Схрейвер А.* Теория линейного и целочисленного программирования. Т. 2. М.: Мир, 1991. 344 с.
3. *Смолев В. В.* Об одном подходе к решению булевого линейного уравнения с целыми положительными коэффициентами // Дискретная математика. 1993. Т. 5. № 3. С. 81–89.
4. *Gál A., Jang J.-T., Limaye N., et al.* Space-efficient approximations for Subset Sum // ACM Trans. Computation Theory. 2016. V. 8. No. 4. Article 16. https://doi.org/10.1145/2894843
5. *Bringmann K.* A near-linear pseudopolynomial time algorithm for subset sum // SODA'17 Proc. Twenty-Eighth Ann. ACM-SIAM Symp. on Discrete Algorithms. Philadelphia: SIAM, 2017. P. 1073–1084.
6. *Koiliaris K. and Xu C.* A faster pseudopolynomial time algorithm for subset sum // SODA'17 Proc. Twenty-Eighth Ann. ACM-SIAM Symp. on Discrete Algorithms. Philadelphia: SIAM, 2017. P. 1062–1072.
7. *Bateni M. H., Hajiaghayi M. T., Seddighin S., and Stein C.* Fast algorithms for knapsack via convolution and prediction // Proc. 50th Ann. ACM SIGACT Symp. on the Theory of Computing (STOC'18). N.Y.: ACM, 2018. P. 1269–1282. https://doi.org/10.1145/3188745.3188876
8. *Curtis V. V. and Sanches C. A. A.* An improved balanced algorithm for the subset-sum problem // Europ. J. Operat. Res. 2019. V. 275. P. 460–466. https://doi.org/10.1016/j.ejor.2018.11.055
9. *Cygan M., Mucha M., Węgrzycki K., and Włodarczyk M.* On problems equivalent to (min,+)-convolution // ACM Trans. Algorithms. 2019. V. 15. No. 1. Article 14. https://doi.org/10.1145/3293465
10. *Kapovich I., Miasnikov A., Schupp P., and Shpilrain V.* Generic-case complexity, decision problems in group theory and random walks // J. Algebra. 2003. V. 264. No. 2. P. 665–694. https://doi.org/10.1016/S0021-8693(03)00167-4
11. *Рыбалов А. Н.* О генерической сложности проблемы разрешимости систем диофантовых уравнений в форме Сколема // Прикладная дискретная математика. 2017. № 37. С. 100–106.
12. *Рыбалов А. Н.* О генерической NP-полноте проблемы выполнимости булевых формул // Прикладная дискретная математика. 2017. № 36. С. 106–112.
13. *Рыбалов А. Н.* Релятивизованные генерические классы P и NP // Прикладная дискретная математика. 2018. № 40. С. 100–104.
14. *Lyubetsky V. A., Gershgorin R. A., and Gorbunov K. Yu.* Chromosome structures: reduction of certain problems with unequal gene content and gene paralogs to integer linear programming // BMC Bioinformatics. 2017. V. 18. No. 40. https://doi.org/10.1186/s12859-017-1944-x
15. *Schwartz J. T.* Fast probabilistic algorithms for verification of polynomial identities // J. ACM. 1980. V. 27. No. 4. P. 701–717. https://doi.org/10.1145/322217.322225
16. *Bacher A., Bodini O., Hwang H.-K., and Tsai T.-H.* Generating random permutations by coin-tossing: classical algorithms, new analysis, and modern implementation // ACM Trans. Algorithms. 2017. V. 13. No. 2. Article 24. https://doi.org/10.1145/3009909

REFERENCES

1. *Seliverstov A. V.* Binary solutions to some systems of linear equations. OPTA 2018. Communications in Computer and Information Science, vol. 871, Cham, Springer, 2018, pp. 183–192. https://doi.org/10.1007/978-3-319-93800-4_15
2. *Schrijver A.* Theory of Linear and Integer Programming. N.Y., John Wiley and Sons, 1986.
3. *Smolev V. V.* On an approach to the solution of a Boolean linear equation with positive integer coefficients. Discrete Mathematics and Applications, 1993, vol. 3, no. 5, pp. 523–530. https://doi.org/10.1515/dma.1993.3.5.523
4. *Gál A., Jang J.-T., Limaye N., et al.* Space-efficient approximations for subset sum. ACM Trans. Computation Theory, 2016, vol. 8, no. 4, article 16. https://doi.org/10.1145/2894843
5. *Bringmann K.* A near-linear pseudopolynomial time algorithm for subset sum. SODA'17 Proc. Twenty-Eighth Ann. ACM-SIAM Symp. on Discrete Algorithms, SIAM, Philadelphia, 2017, pp. 1073–1084.
6. *Koiliaris K. and Xu C.* A faster pseudopolynomial time algorithm for subset sum. SODA'17 Proc. Twenty-Eighth Ann. ACM-SIAM Symp. on Discrete Algorithms, SIAM, Philadelphia, 2017, pp. 1062–1072.
7. *Bateni M. H., Hajiaghayi M. T., Seddighin S., and Stein C.* Fast algorithms for knapsack via convolution and prediction. Proc. 50th Ann. ACM SIGACT Symp. on the Theory of Computing (STOC'18), N.Y., ACM, 2018, pp. 1269–1282. https://doi.org/10.1145/3188745.3188876
8. *Curtis V. V., and Sanches C. A. A.* An improved balanced algorithm for the subset-sum problem. Europ. J. Operat. Res., 2019, vol. 275, pp. 460–466. https://doi.org/10.1016/j.ejor.2018.11.055
9. *Cygan M., Mucha M., Węgrzycki K., and Włodarczyk M.* On problems equivalent to (min,+)-convolution. ACM Trans. on Algorithms, 2019, vol. 15, no. 1, article 14. https://doi.org/10.1145/3293465
10. *Kapovich I., Miasnikov A., Schupp P., and Shpilrain V.* Generic-case complexity, decision problems in group theory and random walks. J. Algebra, 2003, vol. 264, no. 2, pp. 665–694. https://doi.org/10.1016/S0021-8693(03)00167-4
11. *Rybalov A. N.* O genericheskoy slozhnosti problemy razreshimosti sistem diofantovykh uravneniy v forme Skolema [On generic complexity of decidability problem for Diophantine systems in the Skolem's form]. Prikladnaya Diskretnaya Matematika, 2017, no. 37, pp. 100–106. (in Russian)
12. *Rybalov A. N.* O genericheskoy NP-polnote problemy vypolnimosti bulevych formul [On generic NP-completeness of the Boolean satisfiability problem]. Prikladnaya Diskretnaya Matematika, 2017, no. 36, pp. 106–112. (in Russian)
13. *Rybalov A. N.* Relyativizovannye generichеckie klassy P i NP [Relativized generic classes P and NP]. Prikladnaya Diskretnaya Matematika, 2018, no. 40, pp. 100–104. (in Russian)
14. *Lyubetsky V. A., Gershgorin R. A., and Gorbunov K. Yu.* Chromosome structures: reduction of certain problems with unequal gene content and gene paralogs to integer linear programming. BMC Bioinformatics, 2017, vol. 18, no. 40. https://doi.org/10.1186/s12859-017-1944-x
15. *Schwartz J. T.* Fast probabilistic algorithms for verification of polynomial identities. J. ACM, 1980, vol. 27, no. 4, pp. 701–717. https://doi.org/10.1145/322217.322225
16. *Bacher A., Bodini O., Hwang H.-K., and Tsai T.-H.* Generating random permutations by coin-tossing: classical algorithms, new analysis, and modern implementation. ACM Trans. Algorithms, 2017, vol. 13, no. 2, article 24. https://doi.org/10.1145/3009909

МАТЕМАТИЧЕСКИЕ МЕТОДЫ КРИПТОГРАФИИ

UDC 621.391.7 DOI 10.17223/20710410/45/4

ON THE CONSTRUCTION OF A SEMANTICALLY SECURE MODIFICATION OF THE MCELIECE CRYPTOSYSTEM

Y. V. Kosolapov, O. Y. Turchenko

Southern Federal University, Rostov-on-Don, Russia

E-mail: itaim@mail.ru

The security of currently used asymmetric cryptosystems is based on the problems of discrete logarithm or discrete factorization. These problems can be effectively solved using Shor's algorithm on quantum computers. An alternative to such cryptosystems can be the McEliece cryptosystem. Its security is based on the problem of decoding a general linear code. In its original form, the McEliece cryptosystem is not semantically secure, from here the problem of constructing a semantically secure cryptosystem of the McEliece type is relevant. In the paper, the goal is to construct a McEliece type cryptosystem that has the IND-CPA property. Further, one can suppose that this system can be used as base cryptosystem for building the McEliece type encryption scheme with the IND-CCA2 property and an efficient information transfer rate.

Keywords: *McEliece type cryptosystems, IND-CPA, semantic security, standart model.*

Introduction

Many public-key cryptosystems are vulnerable to attacks on ciphertext: chosen plaintext attack, chosen ciphertext attack, malleability attack. The readers are referred to [1] for detailed description of these attacks. Semantically secure cryptosystems are immune to most of these attacks. Semantic security was introduced in [2] and means that the ciphertext does not give the adversary any information about the plaintext with polynomial restrictions on adversary's computing resources. One way to build such cryptosystems is to use probabilistic encryption. For example, M. Bellare and P. Rogaway in [3] proposed the optimal asymmetric encryption padding (OAEP) modification for the widely used asymmetric RSA cryptosystem. It should be noted that the security of currently used asymmetric cryptosystems is based on the problems of discrete logarithm or discrete factorization. These problems can be effectively solved using Shor's algorithm [4] on quantum computers. An alternative to such cryptosystems can be the McEliece cryptosystem [5], whose security is based on the problem of decoding a general linear code. In its original form, the McEliece cryptosystem is not semantically secure. The problem of constructing a semantically secure cryptosystem of the McEliece type is relevant. In [1] a modification has been constructed that possesses the strongest persistence property — the indistinguishability under adaptive chosen ciphertext attack (IND-CCA2). However, this property is achieved only in the random oracle model. This model was first used in [6] and means that protocol participants have access to some theoretical function (oracle). The oracle for any unique argument produces a truly random value and if the argument

repeats, the oracle repeats the corresponding output. In [7] a modification of McEliece cryptosystem is constructed that has the property of indistinguishability under chosen plaintext attack (IND-CPA) without using the random oracle model. In this case, one can say that the standard model is used. This modification was later used in [8] as a *base cryptosystem* to construct a system that has the IND-CCA2 property within the standard model. In [8] one information message is encrypted l times, which leads to a decrease in the information transfer rate by at least l times. It is important to note that l is the length of the digital signature key. To provide high security according to [9] the key length of the asymmetric cryptosystem underlying the digital signature algorithm should be at least 256 bits. From here, the rate of information transfer of the cryptosystem from [8] is essentially low. Consequently, the development of cryptosystems of the McEliece type with the IND-CCA2 property and the high information transfer rate is current of interest.

In the present paper, the goal is to construct a McEliece type cryptosystem that has the IND-CPA property. Further, using the ideas of [8], one can suppose that this system can be used as base cryptosystem for building the McEliece type encryption scheme with the IND-CCA2 property and a higher information transfer rate.

The paper has the following structure. In Section 1 2 we introduce the basic definitions. The Section 2 describes the McEliece cryptosystem [5] and its semantically secure modification [7]. Three new cryptosystems are also constructed here. Two of them are used in Section 3 to prove the semantic security of the third one. Section 4 proposes data transfer protocol using this modification.

1. Preliminaries

Let $\mathbb{F}_q$ be a Galois field of cardinality q, where q is the degree of a prime number, $\mathbf{m} = (m_1, \ldots, m_n) \in \mathbb{F}_q^n$. The support of the vector $\mathbf{m}$ is the set $\mathrm{supp}(\mathbf{m}) = \{i : m_i \neq 0\}$ and the Hamming weight of this vector is a number $\mathrm{wt}(\mathbf{m}) = |\mathrm{supp}(\mathbf{m})|$. For the vector $\mathbf{m} \in \mathbb{F}_q^n$ and the ordered set $\omega \subseteq \{1, \ldots, n\}$ we consider the projection operator $\Pi_\omega : \mathbb{F}_q^n \to \mathbb{F}_q^{|\omega|}$ acting according to the rule:

$$\Pi_\omega(\mathbf{m}) = (m_{i_1}, \ldots, m_{i_{|\omega|}}), \;\; i_j \in \omega, \; j = 1, \ldots, |\omega|.$$

Let $\mathbf{x} \in \mathbb{F}_q^{n_1}$, $\mathbf{y} \in \mathbb{F}_q^{n_2}$, $\mathbf{z} \in \mathbb{F}_q^n$, $n_1 + n_2 = n$, $\omega \subset \{1, \ldots, n\}$, $|\omega| = n_1$, then $\mathbf{z} = \mathbf{x} \parallel \mathbf{y}$ will be a concatenation of the vectors $\mathbf{x}$ and $\mathbf{y}$. Denote $\mathbf{z} = \mathbf{x} \parallel_\omega \mathbf{y}$ as merging of these vectors over an ordered set ω. In other words, $\Pi_\omega(\mathbf{z}) = \mathbf{x}$ and $\Pi_{\{1,\ldots,n\}\setminus\omega}(\mathbf{z}) = \mathbf{y}$. Further we will use the standard notations for writing algorithms and experiments described in [10]. By $y \leftarrow \mathcal{A}(x_1, x_2, \ldots)$ we mean that the algorithm $\mathcal{A}$ runs with input parameters $x_1, x_2, \ldots$ and output value y. If the algorithm $\mathcal{A}$ has access to the output of the algorithm (oracle) $\mathcal{O}$ then we write $y \leftarrow \mathcal{A}^{\mathcal{O}}(x_1, x_2, \ldots)$. If S is a finite set, then $s \in_{\mathrm{R}} S$ denotes the operation of picking an element at random and uniformly from S. To denote an asymmetric encryption scheme we will use the triplet of algorithms, i.e. $\Sigma = (\mathcal{K}, \mathcal{E}, \mathcal{D})$, where 1) $\mathcal{K}$ is a probabilistic polynomial-time key generation algorithm which takes as input a security parameter $N \in \mathbb{N}$ and outputs a public-key pk and a secret-key sk; 2) $\mathcal{E}$ is probabilistic polynomial-time encryption algorithm which receives as input a public-key pk and a message $\mathbf{m}$, and outputs a ciphertext $\mathbf{c}$. We will write $\{\mathbf{m}\}_{pk}^{\Sigma}$ as encryption of the message $\mathbf{m}$ with the key pk; 3) $\mathcal{D}$ is deterministic polynomial-time decryption algorithm which takes as input a secret-key sk and a ciphertext $\mathbf{c}$, and outputs either a message $\mathbf{m}$ or a symbol $\perp$ in the case, when ciphertext is incorrect. Decryption of the ciphertext $\mathbf{c}$ on the secret key sk we will denote $\{\mathbf{c}\}_{sk}^{\Sigma}$.

We say a function $\gamma : \mathbb{N} \to [0,1]$ is negligible in k, if $\forall c \in \mathbb{N}\ \exists k_c\ (\gamma(k) \leqslant k^{-c}$ for all $k > k_c)$.

Now we will consider the notions of the security of public key cryptosystems. The first one is the indistinguishability under chosen plaintext attack introduced in [2]. We will consider it in the same way as [8].

Let Σ be an encryption scheme and let $\mathcal{A} = (\mathcal{A}_1, \mathcal{A}_2)$ be an adversary. It should be noted that $\mathcal{A}$ is polynomial time if both probabilistic algorithm $\mathcal{A}_1$ and probabilistic algorithm $\mathcal{A}_2$ are polynomial time. Now one can consider the following experiment (Algorithm 1).

Algorithm 1. $\mathbf{Exp}^{\mathrm{cpa}}_{\Sigma,\mathcal{A}}$

1: $(pk, sk) \leftarrow \mathcal{K}(\mathbf{1}^N)$;
2: $(\mathbf{m}_0, \mathbf{m}_1, st) \leftarrow \mathcal{A}_1(pk)$;
3: $b \leftarrow \{0,1\}$;
4: $\mathbf{c} \leftarrow \{\mathbf{m}_b\}^{\Sigma}_{pk}$;
5: $B \leftarrow \mathcal{A}_2(\mathbf{c}, st)$.
6: If $B = b$, then return 1, else return 0.

The meaning of this experiment can be explained by an example. Let Σ be the basic RSA cryptosystem over ring $\mathbb{Z}_n$. The adversary selects two plaintexts using the algorithm $\mathcal{A}_1$ which generates messages randomly or by using some features of the cryptosystem. In the basic RSA cryptosystem the feature is the fact that $\{0\}^{\Sigma}_{pk} = 0 \in \mathbb{Z}_n$ for any pk. Let $\mathcal{A}_1$ always gives a pair $(0, a, st)$, where $a \neq 0$ and st is the whole state information obtained during the run of $\mathcal{A}_1$. For instance st contains a public key pk and generated messages $\mathbf{m}_0$, $\mathbf{m}_1$. Then the experimenter selects random coin b and encrypts $\mathbf{m}_b$. The adversary's task, given the encryption $\mathbf{c}$, is to determine which of the two plaintexts was encrypted. In the framework of this example, algorithm $\mathcal{A}_2$ can be trivial. In fact $\mathcal{A}_2$ checks whether the resulting cipher is a zero number. If it is, then $\mathcal{A}_2$ outputs 0 (corresponds to zero plain text), otherwise 1 (corresponds to plain text a).

The advantage of the adversary $\mathcal{A}$ is determined by the value

$$\mathbf{Adv}^{\mathrm{cpa}}_{\Sigma,\mathcal{A}}(N) = \left| \mathsf{P}[\mathbf{Exp}^{\mathrm{cpa}}_{\Sigma,\mathcal{A}} = 1] - \frac{1}{2} \right|,$$

where $\mathsf{P}[A]$ denotes probability of the event A. It is said that the cryptosystem Σ has the property IND-CPA if for any polynomial algorithm $\mathcal{A} = (\mathcal{A}_1, \mathcal{A}_2)$ the advantage of $\mathbf{Adv}^{\mathrm{cpa}}_{\Sigma,\mathcal{A}}(N)$ is a negligible function in N.

Now let the adversary $\mathcal{A}^D = (\mathcal{A}^{\mathcal{D}}_1, \mathcal{A}^{\mathcal{D}}_2)$ has access to the decryption oracle $\mathcal{D}$. By $\mathcal{A}^{\mathcal{D}\{\cdot\}}_i$ we mean that adversary $\mathcal{A}^{\mathcal{D}}_i$ has a polynomial number of queries to the oracle $\mathcal{D}$. Let us consider the following experiment (Algorithm 2).

Algorithm 2. $\mathbf{Exp}^{\mathrm{cca2}}_{\Sigma,\mathcal{A}}$

1: $(pk, sk) \leftarrow \mathcal{K}(\mathbf{1}^N)$;
2: $(\mathbf{m}_0, \mathbf{m}_1, st) \leftarrow \mathcal{A}^{\mathcal{D}\{\cdot\}}_1(pk)$;
3: $b \leftarrow \{0,1\}$;
4: $\mathbf{c}^* \leftarrow \{\mathbf{m}_b\}^{\Sigma}_{pk}$;
5: $B \leftarrow \mathcal{A}^{\mathcal{D}\{\cdot\}}_2(\mathbf{c}^*, st)$, and $\mathcal{D}\{\mathbf{c}^*\} = \perp$;
6: If $B = b$, then return 1, otherwise 0.

The principal difference from the previous experiment is that the algorithms $\mathcal{A}_1$ and $\mathcal{A}_2$ have access to decryption oracle. The decryption oracle takes as input a ciphertext and for a polynomial time outputs the corresponding plain text. The only limitation is that this oracle can not be requested by the cipher text produced by the experimenter on step 4 ($\mathcal{D}\{\mathbf{c}^*\} = \perp$). In [11], a practical attack on the RSA standard PKCS #1 was presented (the SSL protocol used that standard at that time), which was based on the idea of decryption oracle.

The advantage of adversary $\mathcal{A}^D$ is

$$\mathbf{Adv}^{\text{cca2}}_{\Sigma,\mathcal{A}}(N) = \left|\mathsf{P}[\mathbf{Exp}^{\text{cca2}}_{\Sigma,\mathcal{A}} = 1] - \frac{1}{2}\right|.$$

It is said that the cryptosystem Σ has the property IND-CCA2 if for any polynomial algorithm $\mathcal{A}^D$ an advantage $\mathbf{Adv}^{\text{cca2}}_{\Sigma,\mathcal{A}}(N)$ is negligible function in N.

Further we need some notions from [12, p. 22–26]. Let X_0 and X_1 be finite random variables with the set of values $\mathbf{D}$. Then the statistical distance is the function

$$\delta(X_0, X_1) = \frac{1}{2}\sum_{d\in\mathbf{D}} \big|\,\mathsf{P}[X_0 = d] - \mathsf{P}[X_1 = d]\big|.$$

Let $\mathbb{A}$ be a class of polynomial-time algorithms, which take a cipher text $\mathbf{c}$ and some state information st as input and output one bit. For example, within the framework of the experiment $\mathbf{Exp}^{\text{cpa}}_{\Sigma,\mathcal{A}}$ algorithm $\mathcal{A}_2$ belongs to this class.

Then we will say that ciphertexts of two different cryptosystems $\Sigma_1 = (\mathcal{K}, \mathcal{E}_1, \mathcal{D}_1)$ and $\Sigma_2 = (\mathcal{K}, \mathcal{E}_2, \mathcal{D}_2)$ are indistinguishable by the class of polynomial algorithms $\mathbb{A}$ if for any information message $\mathbf{m}$ and for all $\mathbf{A} \in \mathbb{A}$

$$\delta(\mathbf{A}(\{\mathbf{m}\}^{\Sigma_1}_{pk_1}, st_1), \mathbf{A}(\{\mathbf{m}\}^{\Sigma_2}_{pk_2}, st_2))$$

is a negligible function in N, where pk_i is generated by $\mathcal{K}(\mathbf{1}^N)$. It is not difficult to verify that for all $\mathbf{A} \in \mathbb{A}$

$$\delta(\mathbf{A}(\{\mathbf{m}\}^{\Sigma_1}_{pk_1}, st_1), \mathbf{A}(\{\mathbf{m}\}^{\Sigma_2}_{pk_2}, st_2)) = \big|\,\mathsf{P}[\mathbf{A}(\{\mathbf{m}\}^{\Sigma_1}_{pk_1}, st_1) = 0] - \mathsf{P}[\mathbf{A}(\{\mathbf{m}\}^{\Sigma_2}_{pk_2}, st_2) = 0]\big|.$$

Lemma 1. Let $\Sigma_1 = (\mathcal{K}, \mathcal{E}_1, \mathcal{D}_1)$ and $\Sigma_2 = (\mathcal{K}, \mathcal{E}_2, \mathcal{D}_2)$ are cryptosystems, Σ_1 has the IND-CPA property. If ciphertexts of two different cryptosystems are indistinguishable by the class of polynomial algorithms $\mathbb{A}$, then Σ_2 has the IND-CPA property.

Proof. Suppose that there is an adversary $\mathcal{A} = (\mathcal{A}_1, \mathcal{A}_2)$ such that $\mathbf{Adv}^{\text{cpa}}_{\Sigma_2,\mathcal{A}}(N)$ is a function ψ that is not negligible in N. Now we construct the adversary algorithm $\mathcal{B} = (\mathcal{B}_1, \mathcal{B}_2)$ on the basis of $\mathcal{A}$ and estimate $\mathbf{Adv}^{\text{cpa}}_{\Sigma_1,\mathcal{B}}(N)$. Let pk_1 is public key generated by $\mathcal{K}$. The algorithm $\mathcal{B}_1$ takes as input pk_1 and generates public key pk_2 using $\mathcal{K}$. Then $\mathcal{B}_1$ calls the algorithm $\mathcal{A}_1(pk_2)$ and outputs a triplet $(\mathbf{m}_0, \mathbf{m}_1, st_1)$. Thus, in spite of different public keys, the outputs of $\mathcal{B}_1(pk_1)$ and $\mathcal{A}_1(pk_2)$ will be identical. The algorithm $\mathcal{B}_2$ simply calls the $\mathcal{A}_2$ algorithm from its input. Since the experiments $\mathbf{Exp}^{\text{cpa}}_{\Sigma_1,\mathcal{B}}$ and $\mathbf{Exp}^{\text{cpa}}_{\Sigma_2,\mathcal{A}}$ differ on fourth step, the outputs of the algorithms $\mathcal{B}_2$ and $\mathcal{A}_2$ may differ. Consider the statistical distance between their outputs. By the condition of the lemma, the ciphers are indistinguishable by the class of polynomial algorithms $\mathbb{A}$. Since $\mathcal{A}_2$ belongs to this class, then $\big|\,\mathsf{P}[\mathcal{A}_2(\{\mathbf{m}\}^{\Sigma_1}_{pk_1}, st_1){=}0] - \mathsf{P}[\mathcal{A}_2(\{\mathbf{m}\}^{\Sigma_2}_{pk_2}, st_2) = 0]\big| = \eta$, where η is a negligible function in N. Because of $\mathcal{B}_2$ simply calls $\mathcal{A}_2$ we have

$$|\,\mathsf{P}[\mathcal{B}_2(\{\mathbf{m}\}^{\Sigma_1}_{pk_1}, st_1) = 0] - \mathsf{P}[\mathcal{A}_2(\{\mathbf{m}\}^{\Sigma_2}_{pk_2}, st_2) = 0]| = \eta.$$

It follows that $\mathbf{Adv}^{\mathrm{cpa}}_{\Sigma_1,\mathcal{B}}(N) = \psi \pm \eta$, as $\mathbf{Adv}^{\mathrm{cpa}}_{\Sigma_1,\mathcal{B}}(N)$ is directly related to the output of $\mathcal{B}_2$. But $\psi \pm \eta$ is not a negligible in N. This contradicts the fact that Σ_1 has the IND-CPA property. ■

2. McEliece type cryptosystems

Consider the McEliece cryptosystem $\mathrm{McE}(C)$ on the linear $[n, k, d]$-code $C(\subseteq \mathbb{F}_q^n)$, where n is the length, k is the code dimension, and d is the minimum code distance. Let G be the generating matrix of the code C, $t = \lfloor (d-1)/2 \rfloor$. A secret key sk is a pair (S, P), where S is a non-singular $(k \times k)$-matrix over the field $\mathbb{F}_q$, and P is a permutation $(n \times n)$-matrix. A public key pk is a pair $(\widetilde{G} = SGP, t)$. Encryption of a message $\mathbf{x} \in \mathbb{F}_q^k$ is performed according to the rule

$$\{\mathbf{x}\}^{\mathrm{McE}}_{pk} = \mathbf{x}\widetilde{G} + \mathbf{e} = \mathbf{y},\ \mathrm{wt}(\mathbf{e}) \leqslant t.$$

To decrypt the ciphertext $\mathbf{y}$ one should use an effective decoder $\mathrm{Dec}_C : \mathbb{F}_q^n \to \mathbb{F}_q^k$ of the code C and the secret key sk:

$$\{\mathbf{y}\}^{\mathrm{McE}}_{sk} = \mathrm{Dec}_C(\mathbf{y}P^{-1})S^{-1}. \tag{1}$$

For the same code C, we consider the modification $\mathrm{McE}_l(C)$ of the McEliece type cryptosystem described in [7], where encryption rule has the form

$$\{\mathbf{x}\}^{\mathrm{McE}_l}_{pk} = \{\mathbf{x} \parallel \mathbf{v}\}^{\mathrm{McE}}_{pk} = \mathbf{y},\ \mathbf{x} \in \mathbb{F}_q^l,\ \mathbf{v} \in_R \mathbb{F}_q^{k-l}. \tag{2}$$

To decrypt the ciphertext $\mathbf{y}$, it is enough to apply the rule (1) and discard the last $k - l$ symbols:

$$\{\mathbf{y}\}^{\mathrm{McE}_l}_{sk} = \{\mathbf{y}\}^{\mathrm{McE}}_{sk} (I_l \parallel O_{n-l})^\top,$$

where I_l is the unit $(l \times l)$ matrix, O_{k-l} is the zero $(k - l \times k - l)$ matrix, and $A^\top$ is the transposed matrix A.

On the basis of the cryptosystem $\mathrm{McE}_l(C)$ we construct a new cryptosystem $2\mathrm{McE}_l(C)$, in which the message of length l is encrypted twice according to the rule (2):

$$\{\mathbf{x}\}^{2\mathrm{McE}_l}_{pk} = \{\mathbf{x}\}^{\mathrm{McE}_l}_{pk} \parallel \{\mathbf{x}\}^{\mathrm{McE}_l}_{pk} = \mathbf{y},\ \mathbf{x} \in \mathbb{F}_q^l.$$

Then the decryption rule can be written in the form:

$$\{\mathbf{y}\}^{2\mathrm{McE}_l}_{sk} = \left\{\mathbf{y} (I_n \parallel O_n)^\top\right\}^{\mathrm{McE}_l}_{sk}.$$

Consider a subset $\mathcal{G}_l$ of permutations group $\mathcal{S}_k$ acting on the elements of the set $\{1, \ldots, k\}$ such that for any $\pi \in \mathcal{G}_l$ the condition $\pi(1) < \ldots < \pi(l)$ is satisfied. The set $\{\pi(1), \ldots, \pi(l)\}$ is denoted by ω_π. Note that $|\mathcal{G}_l| = \mathrm{C}_k^l (k-l)!$, since only C_k^l subsets of cardinality l are in the set of k elements, and for each such subset ω there is a class $\mathcal{G}(\omega) \subseteq \mathcal{S}_k$ permutations with cardinality $|\mathcal{G}(\omega)| = (k-l)!$. With every permutation π from $\mathcal{G}_l$ we associate a permutation $(k \times k)$-matrix R_π. Consider the cryptosystem $\omega 2\mathrm{McE}'_l$ with the encryption rule

$$\{\mathbf{x}\}^{\omega 2\mathrm{McE}'_l}_{pk} = \{(\mathbf{x} \parallel \mathbf{v}_1)R_\pi\}^{\mathrm{McE}}_{pk} \parallel \{(\mathbf{x} \parallel \mathbf{v}_2)R_\pi\}^{\mathrm{McE}}_{pk} = \mathbf{y}, \tag{3}$$

where $\mathbf{x} \in \mathbb{F}_q^l, \mathbf{v}_i \in_{\mathrm{R}} \mathbb{F}_q^{k-l}, i = 1, 2, \pi \in_R \mathcal{G}_l$. For decryption, in addition to the secret key sk, the recipient needs to know the matrix R_π. Then the decryption rule takes the form

$$\{\mathbf{y}\}^{\omega 2\mathrm{McE}'_l}_{sk,R_\pi} = (\{\mathbf{y} (I_n \parallel O_n)^\top\}^{\mathrm{McE}}_{sk} \cdot R_\pi^{-1})(I_l \parallel O_{k-l})^\top.$$

Finally, we construct a cryptosystem $\omega 2\mathrm{McE}_l$ based on previous one with the following restriction: $\mathrm{supp}(\mathbf{v}_1 - \mathbf{v}_2) = \{1, \ldots, k\} \setminus \omega_\pi$. Then, for decryption, the recipient does not need the matrix R_π. To find ω, it suffices to compute the vector

$$\mathbf{z} = \{\mathbf{y}\,(I_n \parallel O_n)^\top\}_{sk}^{\mathrm{McE}} - \{\mathbf{y}\,(O_n \parallel I_n)^\top\}_{sk}^{\mathrm{McE}}$$

and find its support $\mathrm{supp}\,(\mathbf{z})$. Then the decryption rule takes the form

$$\{\mathbf{y}\}_{sk}^{\omega 2\mathrm{McE}_l} = (\mathbf{z} \cdot R_{\pi'}^{-1})(I_l \parallel O_{k-l})^\top, \pi' \in \mathcal{G}_l(\omega), \omega = \mathrm{supp}\,(\mathbf{z})\,.$$

3. Semantic security of McEliece type cryptosystems

3.1. Security assumptions

Let $\mathrm{McE}(C)$ be the basic McEliece cryptosystem with security parameter N. The security of $\mathrm{McE}(C)$ is based on the problem of decoding a random linear code [5]. Note that, if there is no polynomial algorithm capable of distinguishing the $(k \times n)$-matrix of the public key of the $\mathrm{McE}(C)$ cryptosystem from a random $(k \times n)$-matrix with non-negligible probability in N, then the cryptosystem $\mathrm{McE}_l(C)$ has the IND-CPA property [7].

Further we will use two additional assumptions.

Assumption 1. There is no polynomial algorithm that can distinguish two random noisy codewords of the code C from random vectors with a non-negligible probability in security parameter N.

The assumption is based on the fact that at present there are no such polynomial algorithms. For example, recent algorithms [13 – 15] that solve the given problem are not polynomial.

Assumption 2. There is no polynomial algorithm that takes as input ciphertext $\mathbf{c}$ of the $\mathrm{McE}(C)$ and the number $l \in \mathbb{N}$, and outputs 0 if $\mathbf{c}$ corresponds to an information message of a weight less than l and outputs 1 if $\mathbf{c}$ corresponds to an information message of weight l with non-negligible distinguishing advantage in the N.

3.2. IND-CPA security of $2\mathrm{McE}_l(C)$

It is easy to verify that the cryptosystem $\mathrm{McE}(C)$ is not IND-CPA-secure for an arbitrary $[n, k, d]$-code C. At the same time, the cryptosystem $\mathrm{McE}_l(C)$ on the Goppa code C is IND-CPA-secure [7].

Let us consider the matrix $\widetilde{G}$ of the public key of the cryptosystem $\mathrm{McE}_l(C)$ in the form

$$\widetilde{G} = \begin{pmatrix} \widetilde{G}_1 \\ \widetilde{G}_2 \end{pmatrix},$$

where $\widetilde{G}_1$ is $(l \times n)$-matrix and $\widetilde{G}_2$ is $(kl \times n)$-matrix. To prove IND-CPA-security of cryptosystem $2\mathrm{McE}_l(C)$ consider the algorithm $\mathcal{D} = (\mathcal{D}_1, \mathcal{D}_2)$ and following experiment (Algorithm 3).

It is important to note that the algorithm $\mathcal{D}_2$ takes a decision only by two vectors and does not accumulate vectors.

Suppose that there exists a polynomial $g(N)$, a polynomial algorithm $\mathcal{D}' = (\mathcal{D}'_1, \mathcal{D}'_2)$ and an infinite subsequence of natural numbers $(N_1, N_2, \ldots)$ such that for all $i = 1, 2, \ldots$ the following inequality holds:

$$\mathsf{P}[\mathbf{Exp}_{\widetilde{G},t,\mathcal{D}'}^{\mathrm{dif}_1}(N_i) = 1] \geqslant \frac{1}{2} + \frac{1}{q(N_i)}. \tag{4}$$

Algorithm 3. $\mathbf{Exp}^{\mathrm{dif}_1}_{\widetilde{G},t,\mathcal{D}}$

1: $\mathbf{m}_0 \leftarrow \mathcal{D}_1(N)$;
2: $b \leftarrow \{0,1\}$.
3: If $b=1$, then $\mathbf{c} = \{\mathbf{m}_0\}^{2\mathrm{McE}_l}_{(\widetilde{G},t)}$, otherwise $\mathbf{c} \in_R \mathbb{F}_q^{2n}$.
4: $B \leftarrow \mathcal{D}_2(\mathbf{c}, \mathbf{m}_0)$;
5: If $B = b$, return 1, otherwise 0.

In other words, the algorithm $\mathcal{D}'$ with a non-negligible probability distinguishes one pair of ciphertexts corresponding to one information message from a pair of random vectors. Let's construct one more algorithm $\mathcal{WD}$ and experiment $\mathbf{Exp}^{\mathrm{dif}_2}_{\widetilde{G}_2,t,\mathcal{WD}}$ (Algorithm 4).

Algorithm 4. $\mathbf{Exp}^{\mathrm{dif}_2}_{\widetilde{G}_2,t,\mathcal{WD}}$

1: $b \leftarrow \{0,1\}$.
2: If $b=0$, then $\mathbf{y}_1, \mathbf{y}_2 \in_{\mathrm{R}} \mathbb{F}_q^n$, otherwise $\mathbf{y}_i = \mathbf{r}_i\widetilde{G}_2 + \mathbf{e}_i$, $\mathbf{r}_i \in_{\mathrm{R}} \mathbb{F}_q^{k-l}$, $\mathrm{wt}(\mathbf{e}_i) \leqslant t$, $i = 1, 2$.
3: $B \leftarrow \mathcal{WD}(\mathbf{y}_1, \mathbf{y}_2, \widetilde{G}_2, t)$.
4: If $B = b$, then return 1, else 0.

In the experiment $\mathbf{Exp}^{\mathrm{dif}_2}_{\widetilde{G}_2,t,\mathcal{WD}}$, given algorithm $\mathcal{WD}$ distinguishes two random noisy codewords of the code with the generator matrix $\widetilde{G}_2$ from random vectors. From here, using $\mathcal{D}'$ one can construct polynomial algorithm $\mathcal{WD}'$ to solve this distinguishing problem (Algorithm 5).

Algorithm 5. $\mathcal{WD}'(\mathbf{y}_1, \mathbf{y}_2, \widetilde{G}_2, t)$

1: $\mathbf{m}_0 \leftarrow \mathcal{D}'_1(N)$;
2: $\mathbf{c}' = (\mathbf{m}_0\widetilde{G}_1 + \mathbf{y}_1) \parallel (\mathbf{m}_0\widetilde{G}_1 + \mathbf{y}_2)$.
3: Return $\mathcal{D}'_2(\mathbf{c}', \mathbf{m}_0)$.

Given (4), we get : $\mathsf{P}[\mathbf{Exp}^{\mathrm{dif}_2}_{\widetilde{G}_2,t,\mathcal{WD}'}] \geqslant \dfrac{1}{2} + \dfrac{1}{q(N_i)}$. But it contradicts the assumption 1.

Hence we obtain that for any polynomial algorithm $\mathcal{D}'$ and any polynomial $q(N)$, the following inequality holds:

$$\left| \mathsf{P}[\mathbf{Exp}^{\mathrm{dif}_1}_{\widetilde{G},t,\mathcal{D}'}(N) = 1] - \frac{1}{2} \right| < \frac{1}{q(N)}. \tag{5}$$

Note that for the experiment $\mathbf{Exp}^{\mathrm{dif}_3}_{\widetilde{G},t,\mathcal{M}}$ (Algorithm 6) the probability of occurrence of 1 is also differs from 1/2 by a negligibly small function. Otherwise, based on corresponding algorithm, one can construct an algorithm $\mathcal{WD}''$ with not negligible $|\,\mathsf{P}[\mathbf{Exp}^{\mathrm{dif}_2}_{\widetilde{G}_2,t,\mathcal{WD}''}(N) = 1] - 1/2|$ in N.

Hence it follows that there is no polynomial algorithm $\mathcal{Q}$ that distinguishes the ciphertext $\{\mathbf{m}_0\}^{2\mathrm{McE}_l}_{(\widetilde{G},t)}$ from $[\{\mathbf{m}_1\}^{\mathrm{McE}_l}_{(\widetilde{G},t)} \parallel \{\mathbf{m}_2\}^{\mathrm{McE}_l}_{(\widetilde{G},t)}]$ with a probability that is not negligible greater than 1/2 for any $\mathbf{m}_0, \mathbf{m}_1, \mathbf{m}_2$. Otherwise, we can construct the polynomial algorithm $\widetilde{D} = (\widetilde{D}_1, \widetilde{D}_2)$ (see Algorithm 7) for the experiment $\mathbf{Exp}^{\mathrm{dif}_1}_{\widetilde{G},t,\mathcal{M}}$, which with a non-negligible probability would distinguish a pair of ciphertexts from a pair of random vectors, which contradicts the assumption.

Algorithm 6. $\mathbf{Exp}^{\mathrm{dif}_3}_{\widetilde{G},t,\mathcal{M}}$

1: $\mathbf{m}_1, \mathbf{m}_2 \leftarrow \mathcal{M}_1(N)$;
2: $b \leftarrow \{0,1\}$.
3: If $b = 1$, then $\mathbf{c} = [\{\mathbf{m}_1\}^{\mathrm{McE}_l}_{(\widetilde{G},t)} \parallel \{\mathbf{m}_2\}^{\mathrm{McE}_l}_{(\widetilde{G},t)}]$, otherwise $\mathbf{c} \in_{\mathrm{R}} \mathbb{F}_q^{2n}$.
4: $b' \leftarrow \mathcal{M}_2(\mathbf{c}, \mathbf{m}_1, \mathbf{m}_2)$.
5: If $B = b$, return 1, otherwise 0.

Algorithm 7. $\widetilde{D}_2(\mathbf{c}_0, \mathbf{m}_0)$

1: $\mathbf{m}_1, \mathbf{m}_2 \leftarrow \mathcal{M}_1(N)$;
2: $\mathbf{c}_1 = \{\mathbf{m}_1\}^{\mathrm{McE}_l}_{pk} \parallel \{\mathbf{m}_2\}^{\mathrm{McE}_l}_{pk} + \mathbf{c}$;
3: $v \leftarrow \{0,1\}$.
4: Return $\mathcal{Q}(\mathbf{c}_v, \mathbf{m}_0, \mathbf{m}_1 + \mathbf{m}_0, \mathbf{m}_2 + \mathbf{m}_0)$.

Theorem 1. If the cryptosystem $\mathrm{McE}_l(C)$ has the IND-CPA property, then the cryptosystem $2\mathrm{McE}_l(C)$ also has this property .

Proof. Suppose that the cryptosystem $2\mathrm{McE}_l(C)$ does not have the IND-CPA property. Then there exists a polynomial algorithm (adversary) $\mathcal{A}' = (\mathcal{A}'_1, \mathcal{A}'_2)$, the polynomial $p(N)$ and an infinite subsequence of natural numbers $(N_1, N_2, \ldots)$ such that for all $i = 1, 2, \ldots$ the following inequality holds:

$$\mathbf{Adv}^{\mathrm{cpa}}_{2\mathrm{McE}_l(C),\mathcal{A}'}(N_i) \geqslant \frac{1}{p(N_i)}. \tag{6}$$

On the basis of the algorithm $\mathcal{A}'$ we construct the algorithm $\mathcal{A}'' = (\mathcal{A}'_1, \mathcal{A}''_2)$ for the attack on the cryptosystem McE_l. The algorithm $\mathcal{A}''_2$ takes as input the ciphertext $\mathbf{c} = \{\mathbf{m}_b\}^{\mathrm{McE}_l}_{pk}$ of the cryptosystem $\mathrm{McE}_l(C)$ and two messages $\mathbf{m}_0$, $\mathbf{m}_1$; the algorithm $\mathcal{A}''_2$ randomly picks up the value v from $\{0,1\}$ and returns the result $\mathcal{A}'_2(\mathbf{c} \parallel \{\mathbf{m}_v\}^{\mathrm{McE}_l(C)}_{pk})$. Then

$$\mathsf{P}[\mathbf{Exp}^{\mathrm{cpa}}_{\mathrm{McE}_l,\mathcal{A}''} = 1] = \frac{1}{2}\,\mathsf{P}[\mathcal{A}'_2(\{\mathbf{M}\}^{\mathrm{McE}_l(C)}_{pk} \parallel \{\mathbf{M}'\}^{\mathrm{McE}_l(C)}_{pk}) = 1 | \mathbf{M} = \mathbf{M}'] +$$
$$+\frac{1}{2}\,\mathsf{P}[\mathcal{A}'_2(\{\mathbf{M}\}^{\mathrm{McE}_l(C)}_{pk} \parallel \{\mathbf{M}'\}^{\mathrm{McE}_l(C)}_{pk}) = 1 | \mathbf{M} \neq \mathbf{M}'].$$

From (6) we get $|\,\mathsf{P}[\mathcal{A}'_2(\{\mathbf{M}\}^{\mathrm{McE}_l(C)}_{pk} \parallel \{\mathbf{M}'\}^{\mathrm{McE}_l(C)}_{pk}) = 1 | \mathbf{M} = \mathbf{M}'] - 1/2| \geqslant 1/p(N_i)$, and from the explanation , which comes after the (5), we have

$$\left|\mathsf{P}[\mathcal{A}'_2(\{\mathbf{M}\}^{\mathrm{McE}_l(C)}_{pk} \parallel \{\mathbf{M}'\}^{\mathrm{McE}_l(C)}_{pk}) = 1 | \mathbf{M} \neq \mathbf{M}'] - \frac{1}{2}\right| < \frac{1}{q(N)}.$$

In this way,

$$\left|\mathsf{P}[\mathbf{Exp}^{\mathrm{cpa}}_{\mathrm{McE}_l,\mathcal{A}''} = 1] - \frac{1}{2}\right| > \frac{1}{p(N_i)} \pm \phi(N),$$

where $\phi(N)$ is a negligibly small function. Since $\dfrac{1}{p(N_i)} \pm \phi(N)$ is not a negligibly small function, we have obtained that the cryptosystem $\mathrm{McE}_l(C)$ does not have the property $\mathrm{IND} - \mathrm{CPA}$, which contradicts the condition. ■

3.3. IND-CPA property for $\omega 2\mathrm{McE}'_l(C)$

Lemma 2. If the cryptosystem $2\mathrm{McE}_l(C)$ has the IND-CPA property, then the cryptosystem $\omega 2\mathrm{McE}'_l(C)$ also has this property.

Proof. The encryption rule (3) can be rewritten as

$$\{\mathbf{x}\}_{pk}^{\omega 2\mathrm{McE}'_l} = ((\mathbf{x} \parallel \mathbf{v}_1)R_\pi \widetilde{G} \oplus \mathbf{e}_1) \parallel ((\mathbf{x} \parallel \mathbf{v}_2)R_\pi \widetilde{G} \oplus \mathbf{e}_2).$$

Denote $\widetilde{G}' = R_\pi \widetilde{G}$. Then we get

$$\{\mathbf{x}\}_{pk}^{\omega 2\mathrm{McE}'_l} = ((\mathbf{x} \parallel \mathbf{v}_1\widetilde{G}' \oplus \mathbf{e}_1) \parallel ((\mathbf{x} \parallel \mathbf{v}_2)\widetilde{G}' \oplus \mathbf{e}_2) = \{\mathbf{x}\}_{pk'}^{\mathrm{McE}_l} \parallel \{\mathbf{x}\}_{pk'}^{\mathrm{McE}_l} = \{\mathbf{x}\}_{pk'}^{2\mathrm{McE}_l},$$

where $pk' = (\widetilde{G}', t)$. Thus by construction $\omega 2\mathrm{McE}'_l(C)$ is the same as $2\mathrm{McE}_l(C)$ but with different pair (pk, sk). From here $\omega 2\mathrm{McE}'_l(C)$ also has the IND-CPA property. ■

Note, that adversary doesn't know the relationships between (pk, sk) and (pk', sk'). From here adding a permutation in the $2\mathrm{McE}_l(C)$ cryptosystem with the help of the set ω can only increase the security.

3.4. IND − CPA-property for $\omega 2\mathrm{McE}_l(C)$

Theorem 2. The cryptosystem $\omega 2\mathrm{McE}_l(C)$ has the IND-CPA property if the cryptosystem $\omega 2\mathrm{McE}'_l(C)$ has this property.

Proof. For the proof it is sufficiently to show that the ciphertexts of cryptosystems $\omega 2\mathrm{McE}_l(C)$ and $\omega 2\mathrm{McE}'_l(C)$, corresponding to one information message, are indistinguishable by the class of algorithms $\mathbb{A}$. We fix an arbitrary $\mathbf{m}$ and consider the ciphertexts of cryptosystems $\omega 2\mathrm{McE}_l(C)$ and $\omega 2\mathrm{McE}'_l(C)$ as a system of the form:

$$\{\mathbf{m}\}_{pk}^{\omega 2\mathrm{McE}_l} = X \parallel Y, \quad \begin{cases} X = \mathbf{m}\widetilde{G}^1_\omega \oplus \mathbf{r}_1\widetilde{G}^2_\omega \oplus \mathbf{e}_1, \\ Y = \mathbf{m}\widetilde{G}^1_\omega \oplus (\mathbf{1} \oplus \mathbf{r}_1)\widetilde{G}^2_\omega \oplus \mathbf{e}_2, \end{cases}$$

$$\{\mathbf{m}\}_{pk}^{\omega 2\mathrm{McE}'_l} = X \parallel Y', \quad \begin{cases} X = \mathbf{m}_1\widetilde{G}^1_\omega \oplus \mathbf{r}_1\widetilde{G}^2_\omega \oplus \mathbf{e}'_1, \\ Y' = \mathbf{m}_1\widetilde{G}^1_\omega \oplus \mathbf{r}_2\widetilde{G}^2_\omega \oplus \mathbf{e}'_2. \end{cases}$$

Denote $\mathbf{r}'_2 = \mathbf{r}_1 \oplus \mathbf{r}_2$. Then the systems can be rewritten:

$$\{\mathbf{m}\}_{pk}^{\omega 2\mathrm{McE}_l} = X \parallel Y, \quad \begin{cases} X = \mathbf{m}\widetilde{G}^1_\omega \oplus \mathbf{r}_1\widetilde{G}^2_\omega \oplus \mathbf{e}_1, \\ Y = \mathbf{m}\widetilde{G}^1_\omega \oplus \mathbf{r}_1\widetilde{G}^2_\omega \oplus \mathbf{1}\widetilde{G}^2_\omega \oplus \mathbf{e}_2, \end{cases}$$

$$\{\mathbf{m}\}_{pk}^{\omega 2\mathrm{McE}'_l} = X \parallel Y', \quad \begin{cases} X = \mathbf{m}\widetilde{G}^1_\omega \oplus \mathbf{r}_1\widetilde{G}^2_\omega \oplus \mathbf{e}'_1, \\ Y' = \mathbf{m}\widetilde{G}^1_\omega \oplus \mathbf{r}_1\widetilde{G}^2_\omega \oplus \mathbf{r}'_2\widetilde{G}^2_\omega \oplus \mathbf{e}'_2. \end{cases}$$

Now we consider the last parts of Y and Y': $LP = \mathbf{1}\widetilde{G}^2_\omega \oplus \mathbf{e}_2$ and $LP' = \mathbf{r}'_2\widetilde{G}^2_\omega \oplus \mathbf{e}'_2$. Denote Z as $Y = Z \oplus LP$ and Z' as $Y' = Z' \oplus LP'$. Since (5) we get that the rest $X \parallel Z'$ does not provide any information about LP. One should note that $X \parallel Z' = X \parallel Z$. From here $X \parallel Z'$ does not provide any information about LP'. Consequently, to distinguish the ciphertexts one should distinguish LP and LP'. A vector LP of the form $\mathbf{1}\widetilde{G}^2_\omega \oplus \mathbf{e}_1$ can be rewritten as $(\mathbf{0} \parallel \mathbf{1})R_\pi\widetilde{G} \oplus \mathbf{e}_1$. Thus, for a random choice of ω, LP is a ciphertext of basic McEliece cryptosystem corresponding to a random information message with a fixed weight l. The vector $LP' = (\mathbf{r}_1 \oplus \mathbf{r}_2)\widetilde{G}^2_\omega \oplus \mathbf{e}_1$ can similarly be rewritten as $(\mathbf{0} \parallel \mathbf{r}_1 \oplus \oplus\mathbf{r}_2)R_\pi\widetilde{G}\oplus\mathbf{e}_1$ and is also a ciphertext of the basic McEliece cryptosystem, but corresponding

to a random information message of arbitrary weight not exceeding l. By Assumption 2, algorithm for distinguishing vectors of this kind does not exist. Hence the ciphertexts of cryptosystems $\omega 2\mathrm{McE}_l(C)$ and $\omega 2\mathrm{McE}'_l(C)$, corresponding to one information message, are indistinguishable by the class of algorithms $\mathbb{A}$. ■

4. Implementation of ω2McE

We suppose a possible implementation of ω2McE to modify k-repetition scheme [8]. The idea of k-repetition scheme is to encrypt information message k-times using INC-CPA-secure cryptosystem Σ. Encryption of k-repetition scheme has the form $\{\mathbf{m}\}^{\Sigma}_{pk_1} \parallel \parallel \{\mathbf{m}\}^{\Sigma}_{pk_2} \parallel \ldots \parallel \{\mathbf{m}\}^{\Sigma}_{pk_k}$. Note that to encryption requires k unique key pairs. We suggest use ω2McE in k-repetition scheme with some modifications. The idea of our modification is to encrypt $k/2$ information messages using only one set ω. So encryption will take the form $\{\mathbf{m_1}\}^{\omega 2\mathrm{McE}}_{pk_1} \parallel \{\mathbf{m_2}\}^{\omega 2\mathrm{McE}}_{pk_2} \parallel \ldots \parallel \{\mathbf{m_{k/2}}\}^{\omega 2\mathrm{McE}}_{pk_{k/2}}$. In fact, it also requires to encrypt k-times. Let us remind that k is the length of signature key and should be more than 512. However, our construction transmits $k/2$ information messages. From here, with this approach, the data transfer rate will increase by $k/2$ times.

REFERENCES

1. *Kobara K. and Imai H.* Semantically secure McEliece public-key cryptosystems — conversions for McEliece PKC. LNCS, 2001, vol. 1992, pp. 19–35.
2. *Goldwasser S. and Micali S.* Probabilistic encryption. J. Computer and System Sciences, 1984, vol. 38, no. 2, pp. 270–299.
3. *Bellare M. and Rogaway P.* Optimal asymmetric encryption — how to encrypt with RSA. Advances in Cryptology — EUROCRYPT'94, Springer Verlag, 1995, pp. 92–111.
4. *Shor P.* Algorithms for quantum computation: discrete logarithms and factoring. Proc. 35th Ann. Symp. FCS, Santa Fe, USA, IEEE Publ., 1994, pp. 124–134.
5. *McEliece R. J.* A public-key cryptosystem based on algebraic coding theory. DSN Progress Report, 1978, vol. 42, no. 44, pp. 114–116.
6. *Bellare M. and Rogaway P.* Random oracles are practical: A paradigm for designing efficient protocols. CCS '93 Proc. 1st ACM conf. CCS'93, N.Y., ACM, 1993, pp. 62–73.
7. *Nojima R., Imai H., Kobara K., and Morozov K.* Semantic security for the McEliece cryptosystem without random oracles. Designs, Codes and Cryptography, 2008, vol. 49, no. 1–3, pp. 289–305.
8. *Dottling N., Dowsley R., Muller-Quade J., and Nascimento C. A. A.* A CCA2 secure variant of the McEliece cryptosystem. IEEE Trans. Inform. Theory, 2012, vol. 58, no. 10, pp. 6672–6680.
9. *Lenstra A. K. and Verheul E. R.* Selecting cryptographic key sizes. J. Cryptology, 2001, vol. 14, no. 4, pp. 255–293.
10. *Bellare M., Desai A., Pointcheval D., and Rogaway P.* Relations among notions of security for public-key encryption schemes. Advances in Cryptology — CRYPTO'98, LNCS, 1998, vol. 1462, pp. 26–45.
11. *Bleichenbacher D.* Chosen ciphertext attacks against protocols based on the RSA encryption standard PKCS#1. Advances in Cryptology — CRYPTO'98, LNCS, 1998, vol. 1462, pp. 255–293.
12. *Cramer R., Damgard I., and Nielsen J. B.* Secure Multiparty Computation and Secret Sharing. Cambridge, Cambridge University Press, 2015. 373 p.
13. *Kosolapov Y. V. and Turchenko O. Y.* Primenenie odnogo metoda raspoznavaniya koda dlya kanala s podslushivaniem [Application of one method of linear code recognition to the wire-tap channel]. Prikladnaya Diskretnaya Matematika, 2017, no. 35, pp. 76–88. (in Russian)

14. *Chabot C.* Recognition of a code in a noisy environment. Proc. IEEE ISIT, Nice, France, 2007, pp. 2211–2215.
15. *Yardi A. D. and Vijayakumaran S.* Detecting linear block codes in noise using the GLRT. IEEE Intern. Conf. Communications, Budapest, Hungary, 2013, pp. 4895–4899.

МАТЕМАТИЧЕСКИЕ ОСНОВЫ НАДЁЖНОСТИ ВЫЧИСЛИТЕЛЬНЫХ И УПРАВЛЯЮЩИХ СИСТЕМ

УДК 519.718

ДОСТАТОЧНЫЕ УСЛОВИЯ РЕАЛИЗАЦИИ БУЛЕВЫХ ФУНКЦИЙ АСИМПТОТИЧЕСКИ ОПТИМАЛЬНЫМИ ПО НАДЁЖНОСТИ СХЕМАМИ С ТРИВИАЛЬНОЙ ОЦЕНКОЙ НЕНАДЁЖНОСТИ ПРИ НЕИСПРАВНОСТЯХ ТИПА 0 НА ВЫХОДАХ ЭЛЕМЕНТОВ[1]

М. А. Алехина*, С. М. Грабовская**, Ю. С. Гусынина*

*Пензенский государственный технологический университет, г. Пенза, Россия
**Пензенский государственный университет, г. Пенза, Россия

Рассматривается реализация булевых функций схемами из ненадёжных функциональных элементов в полном конечном базисе. Предполагается, что все функциональные элементы независимо друг от друга с вероятностью $\varepsilon \in (0, 1/2)$ переходят в неисправные состояния типа 0 на выходах элементов. Найдены и дополнены ранее известные условия на функции базиса, при выполнении которых почти любую булеву функцию можно реализовать асимптотически оптимальной по надёжности схемой, функционирующей с ненадёжностью, асимптотически равной ε при $\varepsilon \to 0$.

Ключевые слова: *схема, неисправности типа 0 на выходах элементов, ненадёжность, асимптотически оптимальная по надёжности схема, булева функция.*

DOI 10.17223/20710410/45/5

SUFFICIENT CONDITIONS FOR IMPLEMENTATION OF BOOLEAN FUNCTIONS BY ASYMPTOTICALLY OPTIMAL ON RELIABILITY CIRCUITS WITH THE TRIVIAL ESTIMATE OF UNRELIABILITY IN THE CASE OF FAULTS OF TYPE 0 AT THE ELEMENT OUTPUTS

M. A. Alekhina*, S. M. Grabovskaya**, Yu. S. Gusynina*

*Penza State Technological University, Penza, Russia
**Penza State University, Penza, Russia

E-mail: alekhina@penzgtu.ru, alekhina.marina19@yandex.ru, swetazin@mail.ru, gusynina@mail.ru

The implementation of Boolean functions by circuits of unreliable functional elements is considered in a complete finite basis, containing a function of the set M, where $M = \bigcup\limits_{i=1}^{4} (M_i \cup M_i^*)$, $M_1 = \mathrm{Congr}\{x_1^{\sigma_1} x_2^{\sigma_2} \vee x_1^{\bar\sigma_1} x_2^{\bar\sigma_2} x_3^{\sigma_3} : \sigma_i \in \{0,1\}, i \in \{1,2,3\}\}$, $M_2 = \mathrm{Congr}\{x_1^{\sigma_1} x_2^{\sigma_2} x_3^{\sigma_3} \vee x_1^{\sigma_1} x_2^{\bar\sigma_2} x_3^{\bar\sigma_3} \vee x_1^{\bar\sigma_1} x_2^{\sigma_2} x_3^{\bar\sigma_3} : \sigma_i \in \{0,1\}, i \in \{1,2,3\}\}$, $M_3 = = \mathrm{Congr}\{\bar x_1 (x_2^{\sigma_2} \vee x_3^{\sigma_3}) : \sigma_i \in \{0,1\}, i \in \{1,2,3\}\}$, $M_4 = \mathrm{Congr}\{x_1^{\sigma_1} x_2^{\sigma_2} x_3^{\sigma_3} \vee x_1^{\bar\sigma_1} x_2^{\bar\sigma_2} x_3^{\bar\sigma_3} :$

[1]Работа поддержана грантом РФФИ № 17-01-00451-а.

$\sigma_i \in \{0,1\}, i \in \{1,2,3\}\}$. The set M_i^* is the set of functions, each of which is dual to some function of M_i. All functional elements independently of each other with the probability $\varepsilon \in (0, 1/2)$ are assumed to be prone to faults of type 0 at the element outputs. These faults are characterized by the fact that in good condition the functional element implements the function assigned to it, and in the faulty — constant 0. It is proved that almost any Boolean function can be implemented in a complete finite basis B, $B \cap M \neq \varnothing$, by an asymptotically optimal on reliability circuit working with unreliability asymptotically equal to ε at $\varepsilon \to 0$.

Keywords: *circuit, faults of type 0 at the element outputs, unreliability, asymptotically optimal on reliability circuit, Boolean function.*

Введение

Работа относится к одному из важнейших разделов математической кибернетики — теории синтеза, надёжности и сложности управляющих систем. Актуальность исследований в этой области обусловлена важностью многочисленных приложений, возникающих в различных разделах науки и техники.

К числу основных модельных объектов математической теории синтеза, сложности и надёжности управляющих систем относятся схемы из ненадёжных функциональных элементов, реализующие булевы функции. Проблема построения оптимальных по критериям надёжности и сложности схем из ненадёжных элементов является одной из наиболее важных и в то же время трудных в теории синтеза управляющих систем. Разработка специальных методов синтеза схем из ненадёжных функциональных элементов связана, главным образом, с выбранной математической моделью неисправностей. К основным моделям неисправностей относятся, например, инверсные и константные неисправности на выходах элементов. В работе рассматривается задача построения асимптотически оптимальных по надёжности схем в предположении, что функциональные элементы подвержены неисправностям типа 0 на выходах.

Исторически сложилось так, что сначала исследовались инверсные неисправности функциональных элементов, реализующих булевы функции. Первые существенные математические результаты, касающиеся синтеза надёжных схем из ненадёжных элементов, получил Дж. фон Нейман [1]. Он предполагал, что элементы подвержены инверсным неисправностям на выходах, когда функциональный элемент E с приписанной ему булевой функцией $e(\tilde{x})$, переходя в неисправное состояние с вероятностью ε $(0 < \varepsilon < 1/6)$, реализует функцию $\bar{e}(\tilde{x})$. С помощью итерационного метода Дж. фон Неймана произвольную булеву функцию можно реализовать схемой, вероятность ошибки на выходе которой при любом входном наборе значений переменных не превосходит $c \cdot \varepsilon$ (c — некоторая положительная, зависящая лишь от базиса константа), т. е. ненадёжность схемы сравнима с ненадёжностью одного элемента (такие схемы в теории надёжности управляющих систем принято называть надёжными). С ростом числа итераций сложность схемы при использовании метода Дж. фон Неймана увеличивается экспоненциально.

Любой метод синтеза схем из ненадёжных элементов характеризуется двумя важными параметрами: вероятностью ошибки на выходе схемы (ненадёжностью) и сложностью схемы. Именно оптимизации сложности схем, реализующих булевы функции, уделялось главное внимание в работах Р. Л. Добрушина, С. И. Ортюкова [2, 3], Д. Улига [4] и некоторых других авторов. Задача построения асимптотически оптимальных по надёжности схем из ненадёжных элементов, подверженных тем или иным

неисправностям, ни Дж. фон Нейманом, ни другими исследователями до появления работ М. А. Алехиной не рассматривалась.

Н. Пиппенджер [5] в классическом базисе $\{x_1 \,\&\, x_2, x_1 \vee x_2, \bar{x}_1\}$ построил надёжные схемы без существенного увеличения сложности в предположении, что все элементы схемы ненадёжны, подвержены инверсным неисправностям на выходах.

С. В. Яблонский [6] рассматривал задачу синтеза надёжных схем в базисе $\{x_1 \,\&\, x_2, x_1 \vee x_2, \bar{x}_1, g(x_1, x_2, x_3)\}$. Он предполагал, что элемент, реализующий функцию голосования $g(x_1, x_2, x_3) = x_1x_2 \vee x_1x_3 \vee x_2x_3$, абсолютно надёжный, а конъюнктор, дизъюнктор и инвертор ненадёжные, подвержены произвольным неисправностям, ненадёжность каждого из них не больше ε. Им доказано, что для любого p существует алгоритм, который для каждой булевой функции строит асимптотически оптимальную по сложности схему, ненадёжность которой не больше p.

В. В. Тарасов [7] рассматривал задачу построения схем сколь угодно высокой надёжности (когда ненадёжность схемы стремится к 0). Для базисов из ненадёжных функциональных элементов с двумя входами и одним выходом он нашёл необходимые и достаточные условия, при которых любую булеву функцию можно реализовать схемой сколь угодно высокой надёжности.

Позднее в работах В. В. Чугуновой, А. В. Васина, Д. М. Клянчиной и некоторых других авторов решалась задача реализации булевых функций асимптотически оптимальными по надёжности схемами при различных неисправностях элементов.

Данная работа продолжает исследования, начатые в [8] для полного базиса $\{\bar{x}_1 \vee \bar{x}_2\}$, элементы которого подвержены неисправностям типа 0 на выходах. Позднее задача синтеза асимптотически оптимальных по надёжности схем при неисправностях типа 0 была решена [9] во всех полных неприводимых базисах из двухвходовых функциональных элементов, кроме одного. Задача построения асимптотически оптимальных по надёжности схем при инверсных неисправностях на выходах элементов решена А. В. Васиным [10] во всех полных конечных базисах, содержащих функции трёх переменных. В частности, А. В. Васин нашёл необходимые и достаточные условия на функции полного конечного базиса B_1, содержащего функции трёх переменных, при которых почти любую функцию можно реализовать схемой с ненадёжностью, асимптотически равной 2ε при $\varepsilon \to 0$. Аналогичная задача при неисправностях типа 0 на выходах элементов ранее была решена в некоторых полных конечных базисах [11 – 13], а полученные результаты привели к гипотезе «В базисе B_1 почти любую функцию можно реализовать схемой с ненадёжностью, асимптотически равной ε при $\varepsilon \to 0$, если элементы подвержены неисправностям типа 0 на выходах». Доказательство гипотезы (исключая базисы, содержащие линейную функцию) приводится в этой работе.

Введём необходимые понятия и определения.

1. Необходимые понятия, определения и ранее известные результаты

Рассмотрим реализацию булевых функций схемами из ненадёжных функциональных элементов в полном конечном базисе B. Схема реализует функцию $f(x_1, \ldots, x_n)$ $(n \in \mathbb{N})$, если при поступлении на входы схемы набора $\tilde{a}^n = (a_1, \ldots, a_n)$ при отсутствии неисправностей на выходе схемы появляется значение $f(\tilde{a}^n)$. Предполагается, что все функциональные элементы независимо друг от друга с вероятностью $\varepsilon \in (0, 1/2)$ переходят в неисправные состояния типа 0 на выходах элементов. Эти неисправности характеризуются тем, что в исправном состоянии функциональный элемент реализует приписанную ему функцию, а в неисправном — константу 0.

Пусть $P_{\bar{f}(\tilde{a}^n)}(S, \tilde{a}^n)$ — вероятность появления значения $\bar{f}(\tilde{a}^n)$ на выходе схемы S, реализующей функцию $f(\tilde{x}^n)$ при входном наборе $\tilde{a}^n$. Ненадёжность схемы S равна $P(S) = \max\{P_{\bar{f}(\tilde{a}^n)}(S, \tilde{a}^n)\}$, где максимум берётся по всем наборам $\tilde{a}^n$. Надёжность схемы S равна $1 - P(S)$.

Пусть $P_\varepsilon(f) = \inf P(S)$, где инфимум берётся по всем схемам S из ненадёжных элементов, реализующим функцию $f(\tilde{x}^n)$. Схему A из ненадёжных элементов, реализующую f, назовем *асимптотически оптимальной* (*асимптотически наилучшей*) *по надёжности*, если $P(A) \sim P_\varepsilon(f)$ при $\varepsilon \to 0$, т. е. $\lim\limits_{\varepsilon\to 0} \dfrac{P_\varepsilon(f)}{P(A)} = 1$.

Булевы функции f_1 и f_2 назовем *конгруэнтными* [14], если одна из них может быть получена из другой заменой переменных (без отождествления).

Пусть $X \subseteq P_2$. Обозначим $\mathrm{Congr}(X)$ множество всех функций, каждая из которых конгруэнтна некоторой функции множества X.

Рассмотрим следующие множества булевых функций:

$M_1 = \mathrm{Congr}\{x_1^{\sigma_1}x_2^{\sigma_2} \vee x_1^{\bar\sigma_1}x_2^{\bar\sigma_2}x_3^{\sigma_3} : \sigma_i \in \{0,1\},\ i \in \{1,2,3\}\}$,
$M_2 = \mathrm{Congr}\{x_1^{\sigma_1}x_2^{\sigma_2}x_3^{\sigma_3} \vee x_1^{\sigma_1}x_2^{\bar\sigma_2}x_3^{\bar\sigma_3} \vee x_1^{\bar\sigma_1}x_2^{\sigma_2}x_3^{\bar\sigma_3} : \sigma_i \in \{0,1\},\ i \in \{1,2,3\}\}$,
$M_3 = \mathrm{Congr}\{\bar{x}_1(x_2^{\sigma_2} \vee x_3^{\sigma_3}) : \sigma_i \in \{0,1\},\ i \in \{1,2,3\}\}$,
$M_4 = \mathrm{Congr}\{x_1^{\sigma_1}x_2^{\sigma_2}x_3^{\sigma_3} \vee x_1^{\bar\sigma_1}x_2^{\bar\sigma_2}x_3^{\bar\sigma_3} : \sigma_i \in \{0,1\},\ i \in \{1,2,3\}\}$,

а также M_1^*, M_2^*, M_3^*, M_4^* — множества функций, двойственных функциям из множеств M_1, M_2, M_3, M_4 соответственно.

Обозначим $M = \bigcup\limits_{i=1}^{4} (M_i \cup M_i^*)$, $P_2(n)$ — множество булевых функций, зависящих от переменных $x_1, x_2, \ldots, x_n$.

Ранее при инверсных неисправностях на выходах элементов [10] исследовались все полные в $P_2(3)$ конечные базисы, каждый из которых содержит функцию множества M. Доказано [10] (см. также [15]), что если базис B содержит функцию из множества M, то при инверсных неисправностях на выходах элементов любую булеву функцию можно реализовать схемой, ненадёжность которой не больше $2\varepsilon + 109\varepsilon^2$ при всех $\varepsilon \in (0, 1/960]$. Следовательно, в базисе B любую булеву функцию можно реализовать схемой, функционирующей с ненадёжностью асимптотически не больше 2ε при $\varepsilon \to 0$.

При неисправностях типа 0 на выходах элементов в [11, 12] исследовались полные конечные базисы, содержащие некоторые из функций множества M. При этих неисправностях в рассмотренных базисах доказано, что любую булеву функцию можно реализовать схемой, ненадёжность которой не больше $\varepsilon + 100\varepsilon^2$ при всех $\varepsilon \in (0, 1/960]$. Следовательно, в базисе B любую булеву функцию можно реализовать схемой, функционирующей с ненадёжностью асимптотически не больше ε при $\varepsilon \to 0$.

Замечание 1. При неисправностях типа 0 на выходах элементов любая схема, содержащая хотя бы один функциональный элемент и реализующая отличную от константы 0 функцию, имеет ненадёжность не меньше ε при всех $\varepsilon \in (0, 1/2)$ [9].

Таким образом, из результатов, полученных в [11, 12, 13], с учётом замечания 1 следует, что в ряде базисов, содержащих некоторые из функций множества M, почти любую функцию можно реализовать асимптотически оптимальной по надёжности схемой, функционирующей с ненадёжностью, асимптотически равной ε при $\varepsilon \to 0$. Теперь более детально сформулируем выдвинутую во введении и доказанную далее гипотезу: «В полном конечном базисе, содержащем любую из функций множества M, почти любую булеву функцию можно реализовать асимптотически оптимальной по надёжности схемой, функционирующей с ненадёжностью, асимптотически равной ε при $\varepsilon \to 0$».

Заметим, что в некоторых из рассматриваемых базисов для повышения надёжности исходных схем можно использовать схемы из работы [15], в других — искать новые схемы или способы доказательства результатов.

Сформулируем необходимые для доказательств теорем ранее известные леммы 1–3 и теорему 1. Заметим, что леммы 1 и 3 верны в любом полном конечном базисе при произвольных неисправностях элементов.

Обозначим $G = \mathrm{Congr}\{x_1^{\sigma_1}x_2^{\sigma_2} \vee x_1^{\sigma_1}x_3^{\sigma_3} \vee x_2^{\sigma_2}x_3^{\sigma_3} : \sigma_i \in \{0,1\},\ i \in \{1,2,3\}\}$.

Лемма 1 [9]. Пусть функция f реализована схемой S с ненадёжностью не больше $p \leqslant 1/2$. Пусть схема S_g реализует функцию $g \in G$ с ненадёжностью $P(S_g) \leqslant 1/2$, причём v_0 и v_1 — вероятности ошибок схемы S_g на наборах $(\sigma_1, \sigma_2, \sigma_3)$ и $(\bar\sigma_1, \bar\sigma_2, \bar\sigma_3)$ соответственно. Тогда функцию f можно реализовать схемой $\Phi(S)$, ненадёжность которой $P(\Phi(S)) \leqslant \max\{v_0, v_1\} + 3pP(S_g) + 3p^2$.

Лемма 2 [16]. При неисправностях типа 0 на выходах элементов любую булеву функцию f можно реализовать такой схемой S, что $P(S) \leqslant 5{,}2\varepsilon$ при всех $\varepsilon \in (0, 1/960]$.

Лемма 3 [11]. Пусть схема S_φ реализует функцию $\varphi = x_1^{\sigma_1}x_2^{\sigma_2} \oplus x_3^{\sigma_3}$ ($\sigma_i \in \{0,1\}$, $i \in \{1,2,3\}$) с ненадёжностью $P(S_\varphi)$, причём w_0, w_1 — вероятности ошибок схемы S_φ на наборах $(\bar\sigma_1, \bar\sigma_2, 0), (\bar\sigma_1, \bar\sigma_2, 1)$. Тогда можно построить такую схему S_g, реализующую функцию $g(x_1, x_2, x_3) = (x_1x_2 \vee x_1x_3 \vee x_2x_3)^{\sigma_3}$, что $P(S_g) \leqslant P(S_\varphi) + 2p_\oplus$ ($p_\oplus = \max\{P(S_1), P(S_2)\}$, S_1 — любая схема, реализующая функцию $x_1 \oplus x_2$, S_2 — любая схема, реализующая функцию $x_1 \oplus x_2 \oplus 1$ в рассматриваемом базисе), а для вероятностей ошибок v_1 и v_0 схемы S_g на наборах $(0,0,0)$ и $(1,1,1)$ выполняются неравенства $v_1, v_0 \leqslant \max\{w_0, w_1\} + 2p_\oplus^2$.

Теорема 1 [12]. Если полный конечный базис B содержит функцию из множества $M_3 \cup M_3^*$, то при неисправностях типа 0 на выходах элементов любую булеву функцию в этом базисе можно реализовать схемой, ненадёжность которой не больше $\varepsilon + 100\varepsilon^2$ при всех $\varepsilon \in (0, 1/960]$.

2. Основные результаты

Теорема 2. Если полный конечный базис B содержит функцию из множества $M_1 \cup M_1^*$, то любую булеву функцию можно реализовать схемой, ненадёжность которой не больше $\varepsilon + 100\varepsilon^2$ при всех $\varepsilon \in (0, 1/960]$.

Доказательство. Пусть базис B содержит некоторую функцию множества $M_1 \cup M_1^*$, т. е. функцию, конгруэнтную одной из следующих функций:

1) $\varphi_1(x_1, x_2, x_3) = x_1x_2 \vee \bar x_1\bar x_2x_3^{\sigma_3}$ ($\sigma_3 \in \{0,1\}$);
2) $\varphi_2(x_1, x_2, x_3) = x_1\bar x_2 \vee \bar x_1x_2x_3^{\sigma_3}$;
3) $\varphi_3(x_1, x_2, x_3) = \bar x_1\bar x_2 \vee x_1x_2x_3^{\sigma_3}$;
4) $\varphi_4(x_1, x_2, x_3) = (x_1 \vee x_2)\&(\bar x_1 \vee \bar x_2 \vee x_3^{\sigma_3})$;
5) $\varphi_5(x_1, x_2, x_3) = (x_1 \vee \bar x_2)\&(\bar x_1 \vee x_2 \vee x_3^{\sigma_3})$;
6) $\varphi_6(x_1, x_2, x_3) = (\bar x_1 \vee \bar x_2)\&(x_1 \vee x_2 \vee x_3^{\sigma_3})$.

С л у ч а й 1. Для σ_3 возможны два варианта: $\sigma_3 = 0$ или $\sigma_3 = 1$. Случай $\sigma_3 = 0$ рассмотрен в [11], для него утверждение теоремы верно.

Пусть $\sigma_3 = 1$. Тогда имеем $\varphi_1(x_1, x_1, x_3) = x_1 \vee x_3$, $\varphi_1(x_1, x_2, x_2) = x_1\&x_2$ и по результатам из [9] утверждение теоремы верно.

С л у ч а й 2 рассмотрен в [11], для него утверждение теоремы верно.

С л у ч а й 3 при $\sigma_3 = 0$ рассмотрен в [11], для него утверждение теоремы верно. Пусть $\sigma_3 = 1$. Тогда имеем $\varphi_3(x_1, x_1, x_3) = \bar x_1 \vee x_3$, $\varphi_3(x_1, x_2, x_2) = x_1 \sim x_2$ и по результатам из [17] утверждение теоремы верно.

С л у ч а й 4 рассмотрен в [12], для него утверждение теоремы верно.

С л у ч а й 5. При $\sigma_3 = 0$ имеем $\varphi_5(x_1, x_2, x_2) = x_1 \vee \bar{x}_2$, $\varphi_5(x_1, x_2, x_1) = x_1 \sim x_2$ и по результатам из [17] утверждение теоремы верно.

Пусть $\sigma_3 = 1$. Тогда имеем $\varphi_5(x_1, x_2, x_1) = x_1 \vee \bar{x}_2$, $\varphi_5(x_1, x_2, x_2) = x_1 \sim x_2$ и по результатам из [17] утверждение теоремы верно.

С л у ч а й 6 рассмотрен в [12], для него утверждение теоремы верно. ■

Лемма 4. Пусть $\varphi(x_1, x_2, x_3) = x_1x_2x_3 \oplus x_1x_2 \oplus x_2x_3 \oplus x_1x_3$ или двойственная ей функция φ^* содержатся в полном конечном базисе B, тогда в этом базисе любую булеву функцию можно реализовать схемой, ненадёжность которой не больше $\varepsilon + 74\varepsilon^2$ при всех $\varepsilon \in (0, 1/960]$.

Доказательство.

1) Пусть $\varphi(x_1, x_2, x_3) = x_1x_2x_3 \oplus x_1x_2 \oplus x_2x_3 \oplus x_1x_3 \in B$. Для повышения надёжности исходных схем будем использовать схему $S_{\bar{g}}$ (рис. 1), реализующую функцию $\bar{g}(x_1, x_2, x_3) = \overline{x_1x_2 \vee x_2x_3 \vee x_1x_3} \in G$. Эта схема использовалась также в [15], но теперь её элементы подвержены неисправностям типа 0 на выходах. Оценим её вероятностные характеристики и покажем, как их улучшить.

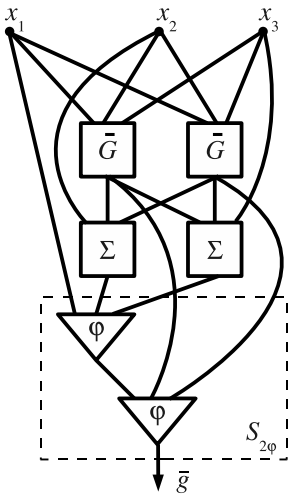


Рис. 1. Схема $S_{\bar{g}}$

Э т а п 1. По лемме 2 для функций $x_1 \oplus x_2 \oplus x_3$ и $\bar{g}(x_1, x_2, x_3) = \overline{x_1x_2 \vee x_1x_3 \vee x_2x_3}$ можно построить схемы Σ и $\tilde{G}$ соответственно, функционирующие с ненадёжностями $P(\Sigma)$ и $P(\tilde{G})$, не превосходящими величины $5{,}2\varepsilon$. Возьмём два элемента, реализующих функцию φ, и по два экземпляра схем Σ и $\tilde{G}$ и соединим их, как показано на рис. 1. Построенная схема $S_{\bar{g}}$ реализует функцию $\bar{g} \in G$. Ненадёжность схемы $S_{\bar{g}}$ удовлетворяет неравенству $P(S_{\bar{g}}) \leqslant 2P(\tilde{G}) + 2P(\tilde{\Sigma}) + 2\varepsilon \leqslant 23\varepsilon$.

Оценим вероятность ошибки v_0 схемы $S_{\bar{g}}$ на наборе $(0, 0, 0)$. Нетрудно видеть, что ошибка в точности одной из подсхем $\tilde{G}$ или Σ на входном наборе $(0, 0, 0)$, а также ошибка невыходного элемента подсхемы $S_{2\varphi}$ не приводит к ошибке на выходе всей схемы $S_{\bar{g}}$. Следовательно, чтобы произошла ошибка, нужно, чтобы ошибся выходной элемент подсхемы $S_{2\varphi}$ или чтобы ошиблись хотя бы две из других подсхем. Поэтому вероятность ошибки v_0 схемы $S_{\bar{g}}$ на наборе $(0, 0, 0)$ удовлетворяет неравенству

$$v_0 \leqslant \varepsilon + \mathrm{C}_4^2(5{,}2\varepsilon)^2 + 4 \cdot 5{,}2\varepsilon^2 \leqslant \varepsilon + 6(5{,}2\varepsilon)^2 + 20{,}8\varepsilon^2 \leqslant \varepsilon + 184\varepsilon^2.$$

Аналогично получается верхняя оценка для вероятности ошибки v_1 схемы $S_{\bar{g}}$ на наборе $(1,1,1)$, только в этом случае вероятность ошибки выходного элемента равна 0, поэтому $v_1 \leqslant 184\varepsilon^2$.

Пусть f — произвольная булева функция. По лемме 2 её можно реализовать схемой S с ненадёжностью $P(S) \leqslant 5{,}2\varepsilon$ при $\varepsilon \in (0, 1/960]$. Используя схему $S_{\bar{g}}$, с помощью леммы 1 по схеме S построим схему $\Phi(S)$ и оценим её ненадёжность при $\varepsilon \in (0, 1/960]$:

$$P(\Phi(S)) \leqslant \max\{v_0, v_1\} + 3pP(S_g) + 3p^2 \leqslant \varepsilon + 184\varepsilon^2 + 3 \cdot 5{,}2\varepsilon \cdot 23\varepsilon + 3(5{,}2\varepsilon)^2 \leqslant$$
$$\leqslant \varepsilon + 624\varepsilon^2 \leqslant 1{,}65\varepsilon.$$

Таким образом, любую булеву функцию f можно реализовать такой схемой S, что $P(S) \leqslant 1{,}65\varepsilon$ при $\varepsilon \leqslant 1/960$.

Э т а п 2. На первом этапе доказано, что схемы Σ и $\bar{G}$ можно построить с ненадёжностью $P(\Sigma) \leqslant 1{,}65\varepsilon$, $P(\bar{G}) \leqslant 1{,}65\varepsilon$. Тогда схему $S_{\bar{g}}$ на рис. 1 можно построить так, что $P(S_{\bar{g}}) \leqslant 2P(\bar{G}) + 2P(\bar{\Sigma}) + 2\varepsilon \leqslant 8{,}6\varepsilon$.

Оценим вероятность ошибки v_0 схемы $S_{\bar{g}}$ на наборе $(0,0,0)$ аналогично этапу 1:

$$v_0 \leqslant \varepsilon + \mathrm{C}_4^2(1{,}65\varepsilon)^2 + 4 \cdot 1{,}65\varepsilon^2 \leqslant \varepsilon + 23\varepsilon^2.$$

Так же получается верхняя оценка для вероятности ошибки v_1 схемы $S_{\bar{g}}$ на наборе $(1,1,1)$, только в этом случае вероятность ошибки выходного элемента равна 0, поэтому $v_1 \leqslant 23\varepsilon^2$.

Пусть f — произвольная булева функция. Реализуем её схемой S с ненадёжностью $P(S) \leqslant 1{,}65\varepsilon$ при $\varepsilon \in (0, 1/960]$. Используя схему $S_{\bar{g}}$, построенную на втором этапе, с помощью леммы 1 по схеме S построим схему $\Phi(S)$ и оценим её ненадёжность: $P(\Phi(S)) \leqslant \max\{v_0, v_1\} + 3pP(S_g) + 3p^2 \leqslant \varepsilon + 23\varepsilon^2 + 3 \cdot 1{,}65\varepsilon \cdot 8{,}6\varepsilon + 3(1{,}65\varepsilon)^2 \leqslant \varepsilon + 74\varepsilon^2$. Таким образом, любую булеву функцию f можно реализовать такой схемой S, что $P(S) \leqslant \varepsilon + 74\varepsilon^2$.

2) Пусть $\varphi^*(x_1, x_2, x_3) = x_1x_2x_3 \oplus x_1 \oplus x_2 \oplus x_3 \oplus 1 \in B$. Поскольку функции $x_1 \oplus x_2 \oplus x_3$ и $\bar{g}(x_1, x_2, x_3) = \overline{x_1x_2 \vee x_1x_3 \vee x_2x_3}$ являются самодвойственными, для доказательства леммы будем использовать схему на рис. 1, заменив функциональные элементы, реализующие φ, на элементы, реализующие φ^*. Все рассуждения такие же, как в предыдущем случае. Отметим только несущественное отличие для оценок v_0, v_1, а именно: на первом этапе $v_0, v_1 \leqslant \varepsilon + 184\varepsilon^2$, а на втором — $v_0, v_1 \leqslant \varepsilon + 23\varepsilon^2$. ■

Теорема 3. Если полный конечный базис B содержит функцию из множества $M_2 \cup M_2^*$, то любую булеву функцию в этом базисе можно реализовать схемой, ненадёжность которой не больше $\varepsilon + 100\varepsilon^2$ при всех $\varepsilon \in (0, 1/960]$.

Доказательство. Пусть базис содержит функцию множества $M_2 \cup M_2^*$, т. е. функцию, конгруэнтную одной из таких функций:

1) $\phi_1(x_1, x_2, x_3) = \bar{x}_1\bar{x}_2\bar{x}_3 \vee \bar{x}_1x_2x_3 \vee x_1\bar{x}_2x_3$;
2) $\phi_2(x_1, x_2, x_3) = \bar{x}_1\bar{x}_2x_3 \vee \bar{x}_1x_2\bar{x}_3 \vee x_1\bar{x}_2\bar{x}_3$;
3) $\phi_3(x_1, x_2, x_3) = x_1x_2\bar{x}_3 \vee x_1\bar{x}_2x_3 \vee \bar{x}_1x_2x_3$;
4) $\phi_4(x_1, x_2, x_3) = x_1x_2x_3 \vee x_1\bar{x}_2\bar{x}_3 \vee \bar{x}_1x_2\bar{x}_3$;
5) $\phi_5(x_1, x_2, x_3) = (\bar{x}_1 \vee \bar{x}_2 \vee \bar{x}_3)\&(\bar{x}_1 \vee x_2 \vee x_3)\&(x_1 \vee \bar{x}_2 \vee x_3)$;
6) $\phi_6(x_1, x_2, x_3) = (\bar{x}_1 \vee \bar{x}_2 \vee x_3)\&(\bar{x}_1 \vee x_2 \vee \bar{x}_3)\&(x_1 \vee \bar{x}_2 \vee \bar{x}_3)$;
7) $\phi_7(x_1, x_2, x_3) = (x_1 \vee x_2 \vee \bar{x}_3)\&(x_1 \vee \bar{x}_2 \vee x_3)\&(\bar{x}_1 \vee x_2 \vee x_3)$;
8) $\phi_8(x_1, x_2, x_3) = (x_1 \vee x_2 \vee x_3)\&(x_1 \vee \bar{x}_2 \vee \bar{x}_3)\&(\bar{x}_1 \vee x_2 \vee \bar{x}_3)$.

С л у ч а й 1. Поскольку $\phi_1(x_1, x_1, x_3) = \bar{x}_1 \& \bar{x}_3$, по результатам из работы [9] утверждение теоремы верно.

С л у ч а и 2, 5 и 6 рассмотрены в [12], для них утверждение теоремы верно.

С л у ч а й 4 рассмотрен в [11], для него утверждение теоремы верно.

С л у ч а и 3 и 7 рассмотрены в лемме 4, для них утверждение теоремы верно.

С л у ч а й 8. Нетрудно проверить, что $\phi_8(x_1, x_1, x_2) = x_1 \vee x_2$, а $\varphi(x_1, x_2, x_3) = \phi_8(x_1, x_3, \phi_8(x_1, x_1, x_2)) = \phi_8(x_1, x_3, x_1 \vee x_2) = \bar{x}_1 x_2 \oplus x_3$. Моделируя последнюю формулу, строим схему S_φ из двух элементов. Вычислим вероятности ошибок w_0 на наборе $(1, 0, 1)$ и w_1 на наборе $(1, 0, 0)$ (см. лемму 3) на выходе схемы S_φ и получим $w_0 = \varepsilon$ и $w_1 = \varepsilon(1 - \varepsilon)$. Следовательно, $\max\{w_0, w_1\} = \varepsilon$.

По условию B — полный базис, следовательно, функции $(x_1 \oplus x_2)^0, (x_1 \oplus x_3)^1 \in [B]$. По лемме 2 реализуем их такими схемами S_1 и S_2 соответственно, что $P(S_1) \leqslant 5{,}2\varepsilon$ и $P(S_2) \leqslant 5{,}2\varepsilon$ при всех $\varepsilon \in (0, 1/960]$. Нетрудно видеть, что $\varphi((x_1 \oplus x_2)^0, (x_1 \oplus x_3)^1, x_1) = x_1x_2 \vee x_1x_3 \vee x_2x_3 = g \in G$. Моделируя формулу $\varphi((x_1 \oplus x_2)^0, (x_1 \oplus x_3)^1, x_1)$, построим схему S_g, реализующую функцию $g \in G$. Легко проверить, что $P(S_g) \leqslant 2\varepsilon + 2 \cdot 5{,}2\varepsilon = 12{,}4\varepsilon$. С помощью леммы 3 оценим вероятности ошибок v_0 и v_1 схемы S_g на наборах $(0, 0, 0)$ и $(1, 1, 1)$ соответственно: $v_0, v_1 \leqslant \max\{w_0, w_1\} + 2p_\oplus^2 \leqslant \varepsilon + 2(5{,}2\varepsilon)^2 \leqslant \varepsilon + 54{,}1\varepsilon^2$.

Пусть f — произвольная булева функция. По лемме 2 её можно реализовать схемой S с ненадёжностью $P(S) \leqslant 5{,}2\varepsilon$ при всех $\varepsilon \in (0, 1/960]$.

Возьмём три экземпляра схемы S, реализующей функцию f. Используя их и схему S_g, построим схему $\Phi(S)$, которая реализует функцию f. Оценим ненадёжность схемы $\Phi(S)$, используя лемму 1. Получим неравенство $P(\Phi(S)) \leqslant \max\{v_0, v_1\} + 3pP(S_g) + 3p^2 \leqslant \varepsilon + 54{,}1\varepsilon^2 + 3(5{,}2\varepsilon)12{,}4\varepsilon + 3(5{,}2\varepsilon)^2 \leqslant \varepsilon + 328{,}7\varepsilon^2 \leqslant 1{,}35\varepsilon$ при $\varepsilon \in (0, 1/960]$. По схеме $\Phi(S)$ построим схему $\Phi^2(S)$. По лемме 1 оценим ненадёжность схемы $\Phi^2(S)$. Получим $P(\Phi^2(S)) \leqslant \varepsilon + 54{,}1\varepsilon^2 + 3(1{,}35\varepsilon)12{,}4\varepsilon + 3(1{,}35\varepsilon)^2 \leqslant \varepsilon + 110\varepsilon^2 \leqslant 1{,}12\varepsilon$ при всех $\varepsilon \in (0, 1/960]$. По схеме $\Phi^2(S)$ построим схему $\Phi^3(S)$. По лемме 1 оценим ненадёжность схемы $\Phi^3(S)$. Получим $P(\Phi^3(S)) \leqslant \varepsilon + 54{,}1\varepsilon^2 + 3(1{,}12\varepsilon)12{,}4\varepsilon + 3(1{,}12\varepsilon)^2 \leqslant \varepsilon + 100\varepsilon^2$. Схема $\Phi^3(S)$ — искомая. ■

Теорема 4. Если полный конечный базис B содержит функцию из множества $M_4 \cup M_4^*$, то любую булеву функцию можно реализовать схемой, ненадёжность которой не больше $\varepsilon + 100\varepsilon^2$ при всех $\varepsilon \in (0, 1/960]$.

Доказательство. Пусть базис B содержит функцию множества $M_4 \cup M_4^*$, т. е. функцию, конгруэнтную одной из следующих функций:

1) $\psi_1(x_1, x_2, x_3) = x_1x_2x_3 \vee \bar{x}_1\bar{x}_2\bar{x}_3$;
2) $\psi_2(x_1, x_2, x_3) = x_1x_2\bar{x}_3 \vee \bar{x}_1\bar{x}_2x_3$;
3) $\psi_3(x_1, x_2, x_3) = (x_1 \vee x_2 \vee x_3)\&(\bar{x}_1 \vee \bar{x}_2 \vee \bar{x}_3)$;
4) $\psi_4(x_1, x_2, x_3) = (x_1 \vee x_2 \vee \bar{x}_3)\&(\bar{x}_1 \vee \bar{x}_2 \vee x_3)$.

С л у ч а и 1 и 2 рассмотрены в [11], для них утверждение теоремы верно.

В с л у ч а е 3 имеем $\psi_3(x_1, x_1, x_3) = x_1 \oplus x_2$, $\psi_3(x_1, x_1, x_1) = 0$, $\psi_3(x_1, x_2, 0) = x_1 \vee x_2$ и по результатам из [17] утверждение теоремы верно.

В с л у ч а е 4 функция $\psi_4(x_1, x_2, x_3)$ представима в виде $x_1x_2 \oplus x_1x_3 \oplus x_2x_3 \oplus x_3 \oplus 1$, т. е. является особенной. По результатам из [18] утверждение теоремы верно. ■

Заключение

Из теорем 1–4 и замечания 1 следует, что при неисправностях типа 0 на выходах элементов в базисе B, $B \cap M \neq \varnothing$, почти любую функцию можно реализовать асимп-

тотически оптимальной по надёжности схемой, функционирующей с ненадёжностью, асимптотически равной ε при $\varepsilon \to 0$.

Таким образом, значительно расширено множество функций, при наличии каждой из которых в базисе возможна реализация почти любой булевой функции асимптотически оптимальной по надёжности схемой, функционирующей с ненадёжностью, асимптотически равной ε при $\varepsilon \to 0$.

Поскольку $M = M^*$, полученные результаты справедливы при неисправностях типа 1 на выходах базисных элементов [19].

ЛИТЕРАТУРА

1. *Von Neumann J.* Probabilistic logics and the synthesis of reliable organisms from unreliable components // Automata studies / eds. C. Shannon and J. McCarthy. Princeton University Press, 1956. P. 43–98.
2. *Добрушин Р. Л., Ортюков С. И.* Верхняя оценка для избыточности самокорректирующихся схем из ненадежных функциональных элементов // Пробл. передачи информ. 1977. Т. 13. № 3. С. 56–76.
3. *Ортюков С. И.* Об избыточности реализации булевых функций схемами из ненадежных элементов // Труды семинара по дискретной математике и её приложениям (Москва, 27–29 января 1987 г.). М.: Изд-во Моск. ун-та, 1989. С. 166–168.
4. *Uhlig D.* Reliable networks from unreliable gates with almost minimal complexity // LNCS. 1987. V. 278. P. 462–469.
5. *Pippenger N.* On networks of noisy gates // 26th Ann. Symp. Foundations of Computer Science. Portland, 21–23 Oct. 1985. P. 30–38.
6. *Яблонский С. В.* Асимптотически наилучший метод синтеза надежных схем из ненадежных элементов // Banach Center Publ. 1982. V. 7. No. 1. P. 11–19.
7. *Тарасов В. В.* К синтезу надежных схем из ненадежных элементов // Матем. заметки. 1976. Т. 20. № 3. С. 391–400.
8. *Алехина М. А.* О синтезе надежных схем из функциональных элементов x/y при однотипных константных неисправностях на выходах элементов // Вест. Моск. ун-та. Матем. Механ. 1991. № 5. С. 80–83.
9. *Алехина М. А.* Синтез асимптотически оптимальных по надежности схем. Пенза: ИИЦ ПГУ, 2006. 156 с.
10. *Васин А. В.* Асимптотически оптимальные по надежности схемы в полных базисах из трехвходовых элементов. Пенза: ИИЦ ПГУ, 2010. 100 с.
11. *Алехина М. А., Клянчина Д. М.* Достаточные условия реализации булевых функций асимптотически оптимальными схемами с тривиальной оценкой ненадежности // Междунар. симп. «Надежность и качество, 2010» (Пенза, 24–31 мая 2010). Пенза: ИИЦ ПГУ, 2010. Т. 1. С. 229–232.
12. *Алехина М. А., Клянчина Д. М.* Об асимптотически оптимальных по надежности схемах в некоторых специальных базисах // Известия высших учебных заведений. Поволжский регион. Физико-математические науки. 2010. № 4 (16). С. 3–13.
13. *Алехина М. А., Клянчина Д. М.* Об асимптотически оптимальных по надежности схемах в базисах, содержащих существенную линейную функцию и функцию вида $x_1^a \& x_2^b$ // Материалы XVI Междунар. конф. «Проблемы теоретической кибернетики» (Нижний Новгород, 20–25 июня 2011). Н. Новгород: Изд-во Нижегород. ун-та, 2011. С. 33–37.
14. *Гаврилов Г. П., Сапоженко А. А.* Задачи и упражнения по дискретной математике: учеб. пособие. 3-е изд., перераб. М.: Физматлит, 2006. 416 с.

15. *Алехина М. А., Васин А. В.* О базисах с коэффициентом ненадежности 2 // Матем. заметки. 2014. Т. 95. № 2. С. 170–201.
16. *Алехина М. А., Гусынина Ю. С., Шорникова Т. А.* Верхняя оценка ненадежности схем в полном конечном базисе (в P_2) при произвольных неисправностях элементов // Изв. вузов. Математика. 2017. № 12. С. 8–83.
17. *Алехина М. А.* О надежности схем в полном конечном базисе, содержащем линейную функцию двух переменных и обобщенную дизъюнкцию // Известия высших учебных заведений. Поволжский регион. Физико-математические науки. 2019. № 1. С. 64–73.
18. *Алехина М. А., Гусынина Ю. С., Шорникова Т. А.* О надежности схем при неисправностях типа 0 на выходах элементов в полном конечном базисе, содержащем особенную функцию // Изв. вузов. Математика. 2019. № 6. С. 85–88.
19. *Алехина М. А., Пичугина П. Г.* О надежности двойственных схем в полном конечном базисе // XVIII Междунар. школа-семинар «Синтез и сложность управляющих систем» (Пенза, 28 сентября–3 октября 2009). М.: Изд-во мех.-мат. ф-та МГУ, 2009. С. 10–13.

REFERENCES

1. *Von Neumann J.* Probabilistic logics and the synthesis of reliable organisms from unreliable components. Automata studies (eds. C. Shannon and J. McCarthy). Princeton University Press, 1956, pp. 43–98.
2. *Dobrushin R. L. and Ortyukov S. I.* Upper bound on the redundancy of self-correcting arrangements of unreliable functional elements. Problems Inform. Transmission, 1977, vol. 13, no. 3, pp. 203–218.
3. *Ortyukov S. I.* Ob izbytochnosti realizatsii bulevykh funktsiy skhemami iz nenadezhnykh elementov [On the redundancy of the Boolean functions implementation by circuits from unreliable elements]. Seminar on Discr. Math. and its Appl. (Moscow, 27–29 Jan. 1987). Moscow, MSU Publ., 1989, pp. 166–168. (in Russian)
4. *Uhlig D.* Reliable networks from unreliable gates with almost minimal complexity. LNCS, 1987, vol. 278, pp. 462–469.
5. *Pippenger N.* On networks of Noisy Gates. 26th Ann. Symp. Foundations of Computer Science, Portland, 21–23 Oct. 1985, pp. 30–38.
6. *Yablonskiy C. V.* Asimptoticheski nailuchshiy metod sinteza nadezhnykh skhem iz nenadezhnykh elementov [Asymptotically best method for synthesizing reliable circuits from unreliable elements]. Banach Center Publ., 1982, vol. 7, no 1, pp. 11–19. (in Russian)
7. *Tarasov V. V.* The synthesis of reliable circuits from unreliable elements. Math. Notes, 1976, vol. 20, iss. 3, pp. 775–780.
8. *Alekhina M. A.* O sinteze nadezhnykh skhem iz funktsional'nykh elementov x/y pri odnotipnykh konstantnykh neispravnostyakh na vykhodakh elementov[On the synthesis of reliable circuits of x/y functional elements at the same type constant faults at the element outputs]. Vestnic MSU, Matematika Mekhanika, 1991, no. 5, pp. 80–83. (in Russian)
9. *Alekhina M. A.* Sintez asimptoticheski optimal'nykh po nadezhnosti skhem [The synthesis of asymptotically optimal on reliability circuits]. Penza, Penz. State Univ., 2006. 156 p. (in Russian)
10. *Vasin A. V.* Asimptoticheski optimal'nyye po nadezhnosti skhemy v polnykh bazisakh trekhvkhodovykh elementov [Asymptotically optimal on reliability circuits in complete bases of three-input elements]: PhD Thesis. Penza, Penz. State Univ., 2010. 100 p. (in Russian)
11. *Alekhina M. A. and Klyanchina D. M.* Dostatochnyye usloviya realizatsii bulevykh funktsiy asimptoticheski optimal'nymi skhemami s trivial'noy otsenkoy nenadezhnosti [Sufficient conditions for the implementation of Boolean functions by asymptotically optimal circuits

with a trivial estimate of unreliability]. Int. Symp. “Reliability and quality, 2010” (Penza, Russia, 24–31 May 2010). Penza, Penz. State Univ., 2010, pp. 229–232. (in Russian)

12. *Alekhina M. A. and Klyanchina D. M.* Ob asimptoticheski optimal'nykh po nadezhnosti skhemakh v nekotorykh spetsial'nykh bazisakh [On asymptotically optimal on reliability circuits in some special bases]. Izvestiya Vysshikh Uchebnykh Zavedeniy. Povolzhskiy Region. Fiziko-Matematicheskie Nauki, 2010, no. 4 (16), pp. 3–13. (in Russian)
13. *Alekhina M. A. and Klyanchina D. M.* Ob asimptoticheski optimal'nykh po nadezhnosti skhemakh v bazisakh, soderzhashchikh sushchestvennuyu lineynuyu funktsiyu i funktsiyu vida $x_1^a \& x_2^b$ [On asymptotically optimal on reliability circuits in bases containing an essential linear function and a function of the form $x_1^a \& x_2^b$]. XVI Int. Conf. “Problems of Theoretical Cybernetics” (Nizhny Novgorod, 20–25 June 2011), Nizhny Novgorod, Nizhny Novgorod State Univ., 2011, pp. 33–37. (in Russian)
14. *Gavrilov G. P. and Sapozhenko A. A.* Zadachi i uprazhneniya po diskretnoy matematike [Tasks and Exercises in Discrete Mathematics]: tutorial, 3rd ed., revised. Moscow, Fizmatlit Publ., 2006. 416 p. (in Russian)
15. *Alekhina M. A. and Vasin A. V.* On bases with unreliability coefficient 2. Math. Notes, 2014, vol. 93, no. 2, pp. 147–173.
16. *Alekhina M. A., Gusynina Yu. S. and Shornikova T. A.* Upper estimate of unreliability of schemes in full finite basis (in P_2) for arbitrary faults of gates. Russian Mathematics, 2017, vol. 61, no. 12, pp. 70–72.
17. *Alekhina M. A.* O nadezhnosti skhem v polnom konechnom bazise, soderzhashchem lineynuyu funktsiyu dvukh peremennykh i obobshchennuyu diz”yunktsiyu [On the reliability of circuits in a complete finite basis containing a linear function of two variables and a generalized disjunction]. Izvestiya Vysshikh Uchebnykh Zavedeniy. Povolzhskiy Region. Fiziko-Matematicheskie Nauki, 2019, no. 1, pp. 64–73. (in Russian)
18. *Alekhina M. A. and Gusynina Yu. S. and Shornikova T. A.* O nadezhnosti skhem pri neispravnostyakh tipa 0 na vykhodakh elementov v polnom konechnom bazise, soderzhashchem osobennuyu funktsiyu [On the reliability of circuits in case of faults of type 0 at the outputs of elements in the complete finite basis, containing a special function]. Izv. Vyssh. Uchebn. Zaved. Mat., 2019, no. 6, pp. 85–88. (in Russian)
19. *Alekhina M. A. and Pichugina P. G.* O nadezhnosti dvoystvennykh skhem v polnom konechnom bazise [On the reliability of dual circuts in the complete finite basis]. XVIII Int. School-Seminar “Synthesis and Complexity of Control Systems” (Penza, 28 Sept.–3 Oct. 2009), Moscow, Faculty of Mechanics and Mathematics of Moscow State University, 2009, pp. 10–13. (in Russian)

ПРИКЛАДНАЯ ТЕОРИЯ ГРАФОВ

УДК 519.17

СРАВНЕНИЕ ДОСТАТОЧНЫХ УСЛОВИЙ ГАМИЛЬТОНОВОСТИ ГРАФА, ОСНОВАННЫХ НА СТЕПЕНЯХ ВЕРШИН

М. Б. Абросимов

Саратовский национальный исследовательский государственный университет имени Н. Г. Чернышевского, г. Саратов, Россия

Для всех графов с числом вершин до 12 сравниваются наиболее популярные достаточные условия гамильтоновости, основанные на степенях вершин графа: теоремы Дирака, Оре, Поша, Хватала и Бонди — Хватала. Для каждого условия подсчитано число графов, ему удовлетворяющих. Наилучшие результаты показывает достаточное условие гамильтоновости, предложенное Бонди и Хваталом в 1976 г., — этому условию удовлетворяют около 90 % гамильтоновых графов.

Ключевые слова: *гамильтонов граф, теорема Дирака, теорема Оре, теорема Поша, теорема Хватала, теорема Бонди — Хватала, FHCP.*

DOI 10.17223/20710410/45/6

COMPARISON OF SUFFICIENT DEGREE BASED CONDITIONS FOR HAMILTONIAN GRAPH

M. B. Abrosimov

Saratov State University, Saratov, Russia

E-mail: mic@rambler.ru

A graph G is said to be Hamiltonian if it contains a spanning cycle, i.e. a cycle that passes through all of its vertices. The Hamiltonian cycle problem is NP-complete, and many sufficient conditions have been found after the first sufficient condition proposed by Dirac in 1952. In this paper for all graphs with a number of vertices up to 12, the most popular sufficient degree based conditions for Hamiltonian graph are compared: theorems by Dirac, Ore, Posa, Chvatal and Bondy-Chvatal. The number of graphs which satisfy each condition is counted. With the number of vertices from 3 to 12, the number of graphs satisfying the Dirac condition is 1, 3, 3, 19, 29, 424, 1165, 108376, 868311, 495369040; the number of graphs satisfying the Ore condition is 1, 3, 5, 21, 68, 503, 4942, 128361, 5315783, 575886211; the number of graphs satisfying the Posha condition is 1, 3, 6, 31, 190, 2484, 53492, 2683649, 216082075, 40913881116; the number of graphs satisfying the Chvatal condition is 1, 3, 6, 34, 194, 2733, 54435, 2914167, 218674224, 43257613552 and the number of graphs satisfying the Bondy — Chvatal condition is 1, 3, 7, 45, 352, 5540, 157016, 8298805, 802944311, 141613919605. This result is the best one: about 90 % of the Hamiltonian graphs satisfy condition proposed by Bondy and Chvatal in 1976. The FHCP Challenge Set is a collection of 1001 instances of the Hamiltonian Cycle Problem, ranging in size from 66 vertices

up to 9528. All graphs from the FHCP Challenge Set were checked whether they satisfy considered conditions. It turned out that 11 graphs satisfy the Bondy — Chvatal condition: no. 59 (with 400 vertices), no. 72 (460), no. 79 (480), no. 84 (500), no. 90 (510), no. 96 (540), no. 128 (677), no. 134 (724), no. 150 (823), no. 162 (909), and no. 188 (with 1123 vertices). For these graphs we can check and find Hamiltonian cycle using Bondy — Chvatal's theorem with computational complexity $O(n^4)$ where n is the number of graph vertices.

Keywords: *Hamiltonian graph, Dirac's theorem, Ore's theorem, Posa's theorem, Chvatal's theorem, theorem by Bondy and Chvatal.*

Введение

В 1859 г. ирландский математик сэр Уильям Роуэн Гамильтон предложил игру, в которой требовалось найти обход додекаэдра по его ребрам с возвратом в исходную точку. В его честь позднее был назван соответствующий обход графа: гамильтоновым циклом называется остовный цикл в графе, то есть цикл, проходящий по всем вершинам графа. Граф, содержащий гамильтонов цикл, называется гамильтоновым. В 1952 г. Г. Дирак предложил достаточное условие гамильтоновости графа: если степень каждой вершины не меньше половины от общего числа вершин, то такой граф является гамильтоновым. Очевидно, что количество рёбер в графе, удовлетворяющем условию Дирака, должно быть достаточно большим. Впоследствии было получено много различных условий гамильтоновости, из которых большую группу образовывают так называемые условия типа Дирака или достаточные условия, сформулированные в терминах степеней вершин графа. Задача проверки гамильтоновости графа является NP-полной, поэтому эффективного алгоритма для её решения неизвестно. Это приводит к тому, что количество различных достаточных условий гамильтоновости продолжает увеличиваться. Зачастую по формулировке условий оценить эффективность того или иного утверждения представляется сложным. Под эффективностью достаточного условия будем понимать число гамильтоновых графов, удовлетворяющих этому условию.

В работе рассматривается эффективность достаточных условий гамильтоновости, которые чаще всего встречаются в литературе по теории графов: теоремы Дирака [1], Оре [2, 3], Поша [4], Хватала [5] и Бонди — Хватала [6]. Обзоры по другим достаточным условиям гамильтоновости можно найти в работах [7 – 10].

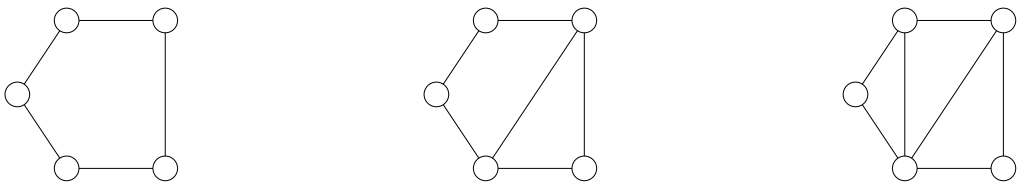
Напомним, что простым неориентированным графом (далее — просто графом) называется пара $G = (V, \alpha)$, где α — симметричное и антирефлексивное отношение на множестве вершин V, называемое отношением смежности. Если $(u, v) \in \alpha$, то говорят, что вершины u и v смежны и эти вершины соединены ребром (u, v). При этом (u, v) и (v, u) — это одно и то же ребро, которое обозначают $\{u, v\}$. Говорят, что ребро $\{u, v\}$ инцидентно каждой из вершин u и v, и эти вершины называются концевыми вершинами или концами ребра $\{u, v\}$. Степенью вершины v в графе G будем называть количество вершин в G, смежных с v, и обозначать через $\mathrm{d}(v)$. Число вершин и рёбер в графе будем обозначать n и m соответственно. Основные определения даются по работе [11].

1. Достаточные условия гамильтоновости

В 1952 г. Г. Дирак предложил первое достаточное условие гамильтоновости графа.

Теорема 1 (Г. Дирак, 1952 [1]). Если в графе G с числом вершин $n \geqslant 3$ степень любой вершины $\mathrm{d}(u) \geqslant n/2$, то граф G гамильтонов.

Граф, который удовлетворяет условию теоремы Дирака, будем называть графом Дирака. Минимальные по числу вершин гамильтоновы графы, которые не являются графами Дирака, — это пять 5-вершинных графов с векторами степеней (2, 2, 2, 2, 2), (3, 3, 2, 2, 2), (4, 3, 3, 2, 2), (4, 4, 3, 3, 2), (3, 3, 3, 3, 2). Первые три графа представлены на рис. 1. Вычислительная сложность проверки условия в теореме Дирака составляет $\mathrm{O}(n+m)$.



а *б* *в*

Рис. 1. Гамильтоновы графы, не являющиеся графами Оре

Теорема 2 (О. Оре, 1960 [2]). Если в графе G с числом вершин $n \geqslant 3$ для любых двух несмежных вершин u и v выполняется неравенство $\mathrm{d}(u) + \mathrm{d}(v) \geqslant n$, то граф G гамильтонов.

Граф, который удовлетворяет условию теоремы Оре, будем называть графом Оре. Легко видеть, что теорема 1 является следствием теоремы 2, то есть графы Дирака являются и графами Оре. Минимальные по числу вершин гамильтоновы графы, которые не являются графами Оре, — это три 5-вершинных графа (рис. 1). Вычислительная сложность проверки условия в теореме Оре составляет $\mathrm{O}(n^2)$.

Оре предложил ещё одно достаточное условие:

Теорема 3 (О. Оре, 1961 [3]). Если в графе G с числом вершин $n \geqslant 3$ число рёбер $m \geqslant (n^2 - 3n + 6)/2$, то граф G гамильтонов.

Это условие существенно слабее основной теоремы Оре и теоремы Дирака, а минимальный по числу вершин граф, который не удовлетворяет условию теоремы 3, — это 4-вершинный цикл C_4.

Теорема 4 (Л. Поша, 1962 [4]). Если граф G с числом вершин $n \geqslant 3$ и степенной последовательностью $d_1 \leqslant \ldots \leqslant d_n$ удовлетворяет следующим двум условиям, то он гамильтонов:

1) для всякого k, $1 \leqslant k < (n-1)/2$, число вершин со степенями, меньшими или равными k, меньше, чем k;
2) для нечётного n число вершин со степенями, меньшими или равными $(n-1)/2$, не превосходит $(n-1)/2$.

Граф, который удовлетворяет условию теоремы Поша, будем называть графом Поша. Теорема 2 является следствием теоремы 4, то есть графы Оре являются и графами Поша. Минимальные по числу вершин гамильтоновы графы, которые не являются графами Поша, — это два 5-вершинных графа с векторами (2, 2, 2, 2, 2) и (3, 3, 2, 2, 2) (рис. 1, *а* и *б*). Граф на рис. 1, *в* с вектором степеней (4, 3, 3, 2, 2) является минимальным по числу вершин графом Поша, который не является графом Оре. Вычислительная сложность проверки условия в теореме Поша составляет $\mathrm{O}(n+m)$. Достаточно часто теорема Поша встречается в следующей упрощённой формулировке:

Теорема 5 (Л. Поша, 1962 [4]). Если граф G с числом вершин $n \geqslant 3$ и степенной последовательностью $d_1 \leqslant \ldots \leqslant d_n$ удовлетворяет следующему условию, то он

гамильтонов: для всякого k, $1 \leqslant k \leqslant n/2$, число вершин со степенями, меньшими или равными k, меньше, чем k.

Как будет видно дальше, теорема 5 при нечётном числе вершин существенно ухудшает теорему 4, а минимальным по числу вершин графом Поша, который не удовлетворяет теореме 5, является граф на рис. 1, *в* с вектором степеней (4, 3, 3, 2, 2).

Теорема 6 (В. Хватал, 1972 [5]). Если граф G с числом вершин $n \geqslant 3$ и степенной последовательностью $d_1 \leqslant \ldots \leqslant d_n$ удовлетворяет следующему условию, то он гамильтонов: для любого k верна импликация $(d_k \leqslant k < n/2) \Rightarrow (d_{n-k} \geqslant n - k)$.

Граф, который удовлетворяет условию теоремы Хватала, будем называть графом Хватала. Минимальные по числу вершин гамильтоновы графы, которые не являются графами Хватала, — это два 5-вершинных графа с векторами (2, 2, 2, 2, 2) и (3, 3, 2, 2, 2) (рис. 1, *а* и *б*). Теорема 5 является следствием теоремы 6, то есть графы Поша являются и графами Хватала. Минимальные по числу вершин графы Хватала, которые не являются графами Поша, — это два 6-вершинных графа с векторами степеней (4, 4, 4, 4, 2, 2) и (5, 4, 4, 3, 2, 2) (рис. 2). Вычислительная сложность проверки условия в теореме Хватала составляет $\mathrm{O}(n + m)$.

а *б*

Рис. 2. Графы Хватала, не являющиеся графами Поша

Для формулировки следующего результата дадим определение замыкания графа.

Определение 1. Замыкание $[G]$ n-вершинного графа G получается из графа G добавлением рёбер $\{u, v\}$ для всех пар вершин u и v, для которых выполняется условие $\mathrm{d}(u) + \mathrm{d}(v) \geqslant n$.

Теорема 7 (Дж. Бонди, В. Хватал, 1976 [6]). Граф G является гамильтоновым тогда и только тогда, когда его замыкание $[G]$ является гамильтоновым.

Построение замыкания графа имеет вычислительную сложность $\mathrm{O}(n^4)$ [6]. Обычно теорема Бонди — Хватала используется в форме следующего достаточного условия гамильтоновости.

Теорема 8 (Дж. Бонди, В. Хватал, 1976 [6]). Если замыкание $[G]$ графа G является полным графом, то граф G гамильтонов.

Граф, который удовлетворяет достаточному условию теоремы Бонди — Хватала, будем называть графом Бонди — Хватала. Минимальным по числу вершин гамильтоновым графом, который не является графом Бонди — Хватала, является цикл C_5 (рис. 1, *а*). Теорема 6 является следствием теоремы 8, то есть графы Хватала являются и графами Бонди — Хватала. Минимальный по числу вершин граф Бонди — Хватала, который не является графом Хватала, — это 5-вершинный граф с вектором степеней (3, 3, 2, 2, 2) (рис. 1, *б*). Очевидно, что если в n-вершинном графе G все вершины имеют степень $\mathrm{d}(v) < n/2$, то замыкание $[G]$ изоморфно самому графу G. Заметим, что любой цикл с числом вершин $n > 4$ не удовлетворяет условию Бонди — Хватала. Если граф удовлетворяет условию Бонди — Хватала, то можно построить и соответствующий гамильтонов цикл с вычислительной сложностью $\mathrm{O}(n^3)$ [6].

2. Вычислительный эксперимент

В книгах по теории графов традиционно встречаются различные вариации теорем 1–8. Например, в [11] доказывается теорема Поша, а следствиями из неё выводятся теоремы Оре и Дирака. В [12] доказывается теорема Хватала, а следствием — теорема Дирака. В [13 – 15] доказывается теорема Хватала, а следствиями — теоремы Оре и Дирака. В [16] доказываются теоремы Оре, Дирака, Хватала и Бонди — Хватала. Интерес представляет то, насколько сильно каждый последующий результат улучшает предыдущий, а также насколько широко каждый результат охватывает всю совокупность гамильтоновых графов с заданным числом вершин. Некоторые условия на практике оказываются совсем малоэффективными [17].

Проведён вычислительный эксперимент по проверке условий теорем 1–6 и 8 для графов с числом вершин до 12. Для генерации графов использовалась программа geng из пакета nauty, разработанного Бренданом МакКеем [18]. Программа получает на вход целое число n, а результатом её работы являются все неизоморфные графы с заданным числом вершин n. Если дополнительно указать ключ -c, то будут построены только связные графы, а с ключом -C — только двусвязные. Напомним, что связный граф называется двусвязным, если в нём нет вершины, после удаления которой граф перестаёт быть связным. Так как все гамильтоновы графы являются двусвязными [11], то достаточно строить только их. Вычисления проводились с использованием кластера высокопроизводительных вычислений ПРЦ НИТ СГУ.

В табл. 1 приведены данные по количеству простых неориентированных графов с заданным числом вершин n, связных, двусвязных и гамильтоновых. Эти данные согласовываются с известными (последовательности A000088, A001349, A002218 и A003216 [19]). В табл. 2 приводятся эти же данные в процентах от общего числа графов с заданным числом вершин.

Таблица 1

Количество связных, двусвязных и гамильтоновых графов

n	Всего	Связных	Двусвязных	Гамильтоновых
3	4	3	1	1
4	11	6	3	3
5	34	21	10	8
6	156	112	56	48
7	1044	853	468	383
8	12346	11117	7123	6196
9	274668	261080	194066	177083
10	12005168	11716571	9743542	9305118
11	1018997864	1006700565	900969091	883156024
12	165091172592	164059830476	153620333545	152522187830

В табл. 3 приводятся данные по числу графов Дирака, Оре, Поша, Хватала и Бонди — Хватала, а в табл. 4 — эти же данные в процентах от общего числа гамильтоновых графов с заданным числом вершин. Последней строкой в табл. 3 указана вычислительная сложность соответствующего достаточного условия. Количество графов Оре с числом вершин до 10 представлено в OEIS (последовательность A264683 [19]).

Можно заметить, что теорема Оре значительно сильнее теоремы Дирака, особенно для нечётного числа вершин. Теоремы Поша и Хватала существенно усиливают теоремы Дирака и Оре, но между собой они отличаются не столь сильно. Наконец, теорема

Таблица 2

Количество связных, двусвязных и гамильтоновых графов в процентах

n	Всего	Связных, %	Двусвязных, %	Гамильтоновых, %
3	4	75	25	25
4	11	54,55	27,27	27,27
5	34	61,76	29,41	23,53
6	156	71,79	35,90	30,77
7	1044	81,70	44,83	36,69
8	12346	90,05	57,69	50,19
9	274668	95,05	70,65	64,47
10	12005168	97,60	81,16	77,51
11	1018997864	98,79	88,42	86,67
12	165091172592	99,38	93,05	92,39

Таблица 3

Количество графов Дирака, Оре, Поша, Хватала и Бонди — Хватала

n	Дирак	Оре	Поша	Хватал	Бонди — Хватал
3	1	1	1	1	1
4	3	3	3	3	3
5	3	5	6	6	7
6	19	21	31	34	45
7	29	68	190	194	352
8	424	503	2484	2733	5540
9	1165	4942	53492	54435	157016
10	108376	128361	2683649	2914167	8298805
11	868311	5315783	216082075	218674224	802944311
12	495369040	575886211	40913881116	43257613552	141613919605
Сложность	$O(n+m)$	$O(n^2)$	$O(n+m)$	$O(n+m)$	$O(n^4)$

Таблица 4

Количество графов Дирака, Оре, Поша, Хватала и Бонди — Хватала в процентах

n	Гамильтоновых	Дирак	Оре	Поша	Хватал	Бонди — Хватал
3	1	100	100	100	100	100
4	3	100	100	100	100	100
5	8	37,50	62,50	75,00	75,00	87,50
6	48	39,58	43,75	64,58	70,83	93,75
7	383	7,57	17,75	49,61	50,65	91,91
8	6196	6,84	8,12	40,09	44,11	89,41
9	177083	0,66	2,79	30,21	30,74	88,67
10	9305118	1,16	1,38	28,84	31,32	89,19
11	883156024	0,10	0,60	24,47	24,76	90,92
12	152522187830	0,32	0,38	26,82	28,36	92,85

Бонди — Хватала намного улучшает теоремы Поша и Хватала, причём покрывает около 90 % гамильтоновых графов.

В табл. 5 приводятся данные по числу графов, удовлетворяющих условиям Оре из теорем 2 и 3, а также условиям Поша из теорем 4 и 5. Можно заметить, что при чётном числе вершин теорема 5 существенно ухудшает теорему 4. Число графов, удовлетворяющих теореме 3, чрезвычайно мало даже по сравнению с числом графов Дирака.

Таблица 5

Количество графов, удовлетворяющих условиям из теорем 2–5

n	Теорема 2	Теорема 3	Теорема 4	Теорема 5
3	1	1	1	1
4	3	2	3	3
5	5	4	6	5
6	21	9	31	31
7	68	19	190	134
8	503	44	2484	2484
9	4942	108	53492	33869
10	128361	277	2683649	2683649
11	5315783	752	216082075	132370005
12	575886211	2179	40913881116	40913881116

Заключение

На основании исследования всех графов с числом вершин до 12 (более 166 миллиардов) можно сделать следующие выводы. Наилучшим из рассмотренных достаточных условий гамильтоновости графа является теорема Бонди — Хватала, которая, имея вычислительную сложность $O(n^4)$, позволяет установить гамильтоновость примерно 90 % графов. Из более простых условий можно использовать теоремы Поша или Хватала, которые имеют вычислительную сложность $O(n + m)$ и позволяют установить гамильтоновость более 25 % графов.

Рассмотренные достаточные условия позволяют установить гамильтоновость графа за полиномиальное время, если в нём достаточно большое число рёбер. В случае произвольного графа проверка графа на гамильтоновость является NP-полной задачей, которая сводится к перебору различных вариантов построения гамильтонового цикла. Разработка алгоритмов, которые могут определить гамильтоновость графа с большим числом вершин, представляет большой интерес. В таких алгоритмах применяются различные эвристики, которые позволяют уменьшить перебор [20]. Для оценки эффективности используются различные классы графов, например обобщённые графы Петерсена или специально составленные библиотеки сложных задач, например TSPLIB [21].

В 2015 г. был объявлен конкурс «FHCP Challenge Set», в рамках которого за один год требовалось найти гамильтоновы циклы в заданном наборе из 1001 специально подобранного графа с числом вершин от 66 до 9528 [20]. Победителям удалось справиться с 985 графами.

В рамках описываемого вычислительного эксперимента были проверены все 1001 граф из набора FHCP на предмет того, удовлетворяют ли они каким-либо из рассмотренных достаточных условий. Оказалось, что 11 графов удовлетворяют условию Бонди — Хватала: № 59 (400 вершин), № 72 (460), № 79 (480), № 84 (500), № 90 (510), № 96 (540), № 128 (677), № 134 (724), № 150 (823), № 162 (909) и № 188 (1123 вершин). Это позволяет достаточно быстро найти гамильтонов цикл в указанных графах (за $O(n^4)$ выполняется построение замыкания и проверка условия Бонди — Хватала, дополнительно за $O(n^3)$ выполняется построение гамильтонова цикла). Остальным достаточным условиям, которые рассмотрены в работе, графы из набора FHCP не удовлетворяют. В табл. 6 для каждого из перечисленных графов указывается его номер в наборе FHCP, число вершин n и рёбер m, максимальная степень вершины Δ и минимальная степень вершины δ.

Таблица 6

№ п/п	n	m	Δ	δ
59	400	40001	362	323
72	460	52901	268	134
79	480	57601	420	455
84	500	62501	200	387
90	510	65026	235	287
96	540	72901	474	458
128	677	114583	585	77
134	724	131045	312	206
150	823	169333	744	479
162	909	206571	274	195
188	1123	315283	908	2

ЛИТЕРАТУРА

1. *Dirac G. A.* Some theorems on abstract graphs // Proc. London Math. Soc. 1952. V. 2. P. 69–81.
2. *Ore O.* Note on Hamilton circuits // Amer. Math. Monthly. 1960. V. 67. P. 55.
3. *Ore O.* Arc coverings of graphs // Ann. Mat. Pura Appl. 1961. V. 55. P. 315–322.
4. *Posa L.* On the circuits of finite graphs // Magyar Tud. Akad. Mat. Kutatd Int. Kozl. 1963. V. 8. P. 355–361.
5. *Chvatal V.* On Hamilton's ideals // J. Combin. Theory. 1972. V. 12. P. 163–168.
6. *Bondy J. A. and Chvatal V.* A method in graph theory // Discr. Math. 1976. V. 15. Iss. 2. P. 111–135.
7. *Gould R. J.* Updating the Hamiltonian problem — A survey // J. Graph Theory. 1991. V. 15. No. 2. P. 121–157.
8. *Gould R. J.* Advances on the Hamiltonian problem — A survey // Graphs and Combinatorics. 2003. V. 19. P. 7–52.
9. *DeLeon M.* A study of sufficient conditions for Hamiltonian cycles // Rose — Hulman Undergraduate Mathematics J. 2000. V. 1. Iss. 1. Article 6. P. 129–145.
10. *Li H.* Generalizations of Diracs theorem in Hamiltonian graph theory — A survey // Discr. Math. 2013. V. 313. P. 2034–2053.
11. *Харари Ф.* Теория графов. М.: Мир, 1973. 300 с.
12. *Diestel R.* Graph Theory. Heidelberg: Springer Verlag, 2017. 447 p.
13. *Емеличев В. А., Мельников О. И., Сарванов В. И., Тышкевич Р. И.* Лекции по теории графов. М.: Наука, 1990. 384 с.
14. *Асанов М. О., Баранский В. А., Расин В. В.* Дискретная математика. Графы, матроиды, алгоритмы. Ижевск: НИЦ «РХД», 2001. 288 с.
15. *Касьянов В. Н., Евстигнеев В. А.* Графы в программировании: обработка, визуализация и применение. СПб.: БХВ-Петербург, 2003. 1104 с.
16. *Омельченко А. В.* Теория графов. М.: МЦНМО, 2018. 416 с.
17. *Абросимов М. Б.* О достаточном условии Гудмана — Хедетниеми гамильтоновости графа // Изв. Сарат. ун-та. Нов. сер. Сер. Математика. Механика. Информатика. 2018. Т. 18. Вып. 3. С. 347–353.
18. *McKay B. D. and Piperno A.* Practical graph isomorphism. II // J. Symbolic Computation. 2014. No. 60. P. 94–112.
19. `http://oeis.org` — The On-Line Encyclopedia of Integer Sequences, 2018.
20. *Haythorpe M.* FHCP Challenge Set: The first set of structurally difficult instances of the Hamiltonian cycle problem // Bulletin of the ICA. 2018. V. 83. P. 98–107.

21. *Reinelt G.* http://www.iwr.uni-heidelberg.de/groups/comopt/software/TSPLIB95/ — TSPLIB, 2018.

REFERENCES

1. *Dirac G. A.* Some theorems on abstract graphs. Proc. London Math. Soc., 1952, vol. 2, pp. 69–81.
2. *Ore O.* Note on Hamilton circuits. Amer. Math. Monthly, 1960, vol. 67, p. 55.
3. *Ore O.* Arc coverings of graphs. Ann. Mat. Pura Appl., 1961, vol. 55, pp. 315–322.
4. *Posa L.* On the circuits of finite graphs. Magyar Tud. Akad. Mat. Kutatd Int. Kozl., 1963, vol. 8, pp. 355–361.
5. *Chvatal V.* On Hamilton's ideals. J. Combin. Theory, 1972, vol. 12, pp. 163–168.
6. *Bondy J. A. and Chvatal V.* A method in graph theory. Discr. Math., 1976, vol. 15, iss. 2, pp. 111–135.
7. *Gould R. J.* Updating the Hamiltonian problem — A survey. J. Graph Theory, 1991, vol. 15, no. 2, pp. 121–157.
8. *Gould R. J.* Advances on the Hamiltonian problem — A survey. Graphs and Combinatorics, 2003, vol. 19, pp. 7–52.
9. *DeLeon M.* A study of sufficient conditions for Hamiltonian cycles. Rose — Hulman Undergraduate Mathematics J., 2000, vol. 1, iss. 1, article 6, pp. 129–145.
10. *Li H.* Generalizations of Diracs theorem in Hamiltonian graph theory — A survey. Discr. Math., 2013, vol. 313, pp. 2034–2053.
11. *Harary F.* Teoria grafov [Graph theory]. Moscow, Mir Publ., 1973. 300 p. (in Russian)
12. *Diestel R.* Graph Theory. Heidelberg, Springer Verlag, 2017. 447 p.
13. *Emelichev V. A., Melnikov O. I., Sarvanov V. I., and Tyshkevich R. I.* Lekcii po teorii grafov [Lections on Graph Theory]. Moscow, Nauka Publ., 1990. 384 p. (in Russian)
14. *Asanov M. O., Baranskij V. A., and Rasin V. V.* Diskretnaya matematika. Grafy, matroidy, algoritmy [Discrete Mathematics: Graphs, Matroids, Algorithms]. Izhevsk, NIC RHD Publ., 2001. 288 p. (in Russian)
15. *Kasyanov V. N. and Evstigneev V. A.* Graphy v programmirovanii: obrabotka, vizualizaciya i primenenie [Graphs in Programming: Processing, Visualization and Application]. SPb., BHV-Peterburg Publ., 2003. 1104 p. (in Russian)
16. *Omelchenko A. V.* Теория графов [Graph Theory]. Moscow, MCCME Publ., 2018. 416 p. (in Russian)
17. *Abrosimov M. B.* On a Goodman — Hedetniemi sufficient condition for the graph hamiltonicity. Izv. Saratov Univ. (N. S.), Ser. Math. Mech. Inform., 2018, vol. 18, iss. 3, pp. 347–353 (in Russian).
18. *McKay B. D. and Piperno A.* Practical graph isomorphism, II. J. Symbolic Computation, 2014, no. 60, pp. 94–112.
19. http://oeis.org — The On-Line Encyclopedia of Integer Sequences, 2018.
20. *Haythorpe M.* FHCP Challenge Set: The first set of structurally difficult instances of the Hamiltonian cycle problem. Bulletin of the ICA, 2018, vol. 83, pp. 98–107.
21. *Reinelt G.* http://www.iwr.uni-heidelberg.de/groups/comopt/software/TSPLIB95/ — TSPLIB, 2018.

УДК 519.8

АЛГОРИТМЫ ПРИБЛИЖЁННОГО РЕШЕНИЯ ОДНОЙ ЗАДАЧИ КЛАСТЕРИЗАЦИИ ГРАФА

В. П. Ильев*, С. Д. Ильева*, А. В. Моршинин**

*Омский государственный университет им. Ф. М. Достоевского, г. Омск, Россия,
**Институт математики им. С. Л. Соболева СО РАН, г. Омск, Россия

Изучается задача кластеризации графа. Для варианта задачи, в котором число кластеров не превосходит 3, разработаны три приближённых алгоритма. Первый алгоритм использует в качестве процедуры известный алгоритм Коулмана — Саундерсона — Вирта, который приближённо решает аналогичную задачу с числом кластеров, не превосходящим 2, и многократно применяет локальный поиск. Второй алгоритм основан на оригинальной идее и вообще не использует локальный поиск. В третьем алгоритме процедура локального поиска применяется к допустимому решению, возвращенному вторым алгоритмом. Получены априорные гарантированные оценки точности всех трёх алгоритмов, лучшая из которых равна $(6 - 12/n)$, где n — число вершин данного графа. Проведено также экспериментальное сравнение предложенных алгоритмов.

Ключевые слова: *граф, кластеризация, приближённый алгоритм, гарантированная оценка точности.*

DOI 10.17223/20710410/45/7

APPROXIMATE ALGORITHMS FOR GRAPH CLUSTERING PROBLEM

V. P. Il'ev*, S. D. Il'eva*, A. V. Morshinin**

*Dostoevsky Omsk State University, Omsk, Russia,
**Sobolev Institute of Mathematics SB RAS, Omsk, Russia

E-mail: iljev@mail.ru

In the paper, the graph clustering problem is studied. For the version of the problem when the number of clusters does not exceed 3, we develop three approximate algorithms. The first algorithm uses as a procedure the known Coleman — Saunderson — Wirth algorithm which approximately solves the similar problem when the number of clusters does not exceed 2, and repeatedly applies a local search. On the contrary, our second algorithm is based on an original idea and does not use a local search at all. The main difference between these algorithms is the following. The first algorithm looks over all vertices of a given graph, for each vertex forms the cluster involving this vertex and all its neighbors and on the rest of the vertices forms one or two clusters using the Coleman — Saunderson — Wirth algorithm. The second algorithm looks over all ordered pairs of vertices of a given graph and for every pair forms two clusters at once, each of which contains only one vertex of this pair with some of its neighbors, placing the rest of the vertices to the third cluster (the third cluster may be empty). Finally, the third algorithm applies the local search only once to the feasible solution returned by the second one. A priori performance guarantees of all approximate algorithms are obtained, the best is equal to $(6 - 12/n)$ for the second and the third

algorithms, where n is the number of vertices of a given graph. Also, experimental comparing of the developed algorithms was carried out. Experimental testing show that running time of our second and third algorithms is essentially less than running time of the first algorithm. At the same time the third algorithm demonstrated the best results in sense of accuracy of the solutions. Thus, the third algorithm has the best characterstics both from point of view of theoretical analysis and experimental study.

Keywords: *graph, clustering, approximation algorithm, performance guarantee.*

Введение

В *задаче кластеризации* требуется разбить заданное множество объектов на несколько подмножеств (кластеров) на основе сходства объектов друг с другом. Одной из наиболее наглядных формализаций этой задачи является *задача кластеризации графа* [1, 2]. В ней объектам взаимно однозначно сопоставлены вершины графа, а рёбра соединяют вершины, соответствующие похожим объектам.

Будем рассматривать только *обыкновенные графы*, т. е. графы без петель и кратных рёбер. Обыкновенный граф называется *кластерным графом* [3], если каждая его компонента связности является полным графом. Пусть $\mathcal{M}(V)$ — множество всех кластерных графов на множестве вершин V; $\mathcal{M}_k(V)$ — множество всех кластерных графов на V, имеющих ровно k непустых компонент связности; $\mathcal{M}_{\leqslant k}(V)$ — множество всех кластерных графов на V, имеющих не более k компонент связности, $2 \leqslant k \leqslant |V|$.

Если $G_1 = (V, E_1)$ и $G_2 = (V, E_2)$ — графы на одном и том же множестве вершин V, то *расстояние* $\rho(G_1, G_2)$ между ними определяется как

$$\rho(G_1, G_2) = |E_1 \Delta E_2| = |E_1 \setminus E_2| + |E_2 \setminus E_1|,$$

т. е. $\rho(G_1, G_2)$ — это число несовпадающих рёбер в графах G_1 и G_2.

В литературе в основном рассматривались следующие три варианта задачи кластеризации графа.

Задача GC. Для произвольного графа $G = (V, E)$ найти такой граф $M^* \in \mathcal{M}(V)$, что

$$\rho(G, M^*) = \min_{M \in \mathcal{M}(V)} \rho(G, M).$$

Задача $\mathbf{GC_k}$. Дан произвольный граф $G = (V, E)$ и целое число k, $2 \leqslant k \leqslant |V|$. Найти такой граф $M^* \in \mathcal{M}_k(V)$, что

$$\rho(G, M^*) = \min_{M \in \mathcal{M}_k(V)} \rho(G, M).$$

Задача $\mathbf{GC_{\leqslant k}}$. Дан произвольный граф $G = (V, E)$ и целое число k, $2 \leqslant k \leqslant |V|$. Найти такой граф $M^* \in \mathcal{M}_{\leqslant k}(V)$, что

$$\rho(G, M^*) = \min_{M \in \mathcal{M}_{\leqslant k}(V)} \rho(G, M).$$

В последние годы задачи кластеризации графов неоднократно переоткрывались и независимо изучались под разными названиями (Correlation Clustering [4], Cluster Editing [3, 5] и др.). В этих и других работах рассматривались также более общие постановки задач.

Вычислительная сложность задач кластеризации графов долгое время оставалась неизвестной. В 1986 г. М. Крживанек и Дж. Моравек [6] доказали, что задача $\mathbf{GC}$ является NP-трудной, однако их работа осталась незамеченной. В 2004 г. в [4] и независимо в [3] доказана NP-трудность задачи $\mathbf{GC}$. В [4] доказано также, что задача $\mathbf{GC_k}$ является NP-трудной при любом фиксированном $k \geqslant 2$; в 2006 г. в [7] опубликовано более простое доказательство этого результата. В том же году в [8] доказано, что задачи $\mathbf{GC_2}$ и $\mathbf{GC_{\leqslant 2}}$ NP-трудны уже на кубических графах, откуда сделан вывод, что все упомянутые ранее задачи кластеризации графа являются NP-трудными, включая и задачу $\mathbf{GC_{\leqslant k}}$.

В [4] приведён 3-приближённый алгоритм для задачи $\mathbf{GC_{\leqslant 2}}$, в [8] доказано существование рандомизированной полиномиальной приближённой схемы для задачи $\mathbf{GC_{\leqslant 2}}$, а в [7] предложена рандомизированная полиномиальная приближённая схема для задачи $\mathbf{GC_k}$ (для любого фиксированного $k \geqslant 2$), сложность которой лишает её перспективы практического применения. В [9] представлен 2-приближённый алгоритм решения задачи $\mathbf{GC_{\leqslant 2}}$, основанный на применении процедуры локального поиска к каждому допустимому решению, полученному с помощью 3-приближённого алгоритма из [4]. Для задачи $\mathbf{GC_2}$ в [10] предложен $(3 - 6/|V|)$-приближённый алгоритм.

Что касается задачи $\mathbf{GC}$, то в [11] показано, что задача $\mathbf{GC}$ является APX-трудной, и разработан 4-приближённый алгоритм её решения. В [12] предложен 2,5-приближённый алгоритм для задачи $\mathbf{GC}$, в [13] — $(2{,}06 - \varepsilon)$-приближённый алгоритм.

Более подробный обзор известных результатов по задачам кластеризации графов можно найти в [14].

В п. 1 настоящей работы приводится краткий обзор известных результатов для задачи кластеризации графа, в которой число кластеров не превосходит 2. В п. 2 рассматривается задача $\mathbf{GC_{\leqslant 3}}$, в которой число кластеров не превосходит 3. Для этой задачи предложены два полиномиальных приближённых алгоритма. Первый 6-приближённый алгоритм использует в качестве процедуры известный алгоритм Коулмана — Саундерсона — Вирта, который приближённо решает аналогичную задачу с числом кластеров, не превосходящим 2, и многократно применяет локальный поиск. Второй алгоритм основан на оригинальной идее и не использует локальный поиск. Доказаны априорные гарантированные оценки точности обоих алгоритмов, лучшая из которых, у второго алгоритма, равна $6 - 12/n$, где n — число вершин графа. В п. 3 предложен ещё один полиномиальный алгоритм приближённого решения задачи $\mathbf{GC_{\leqslant 3}}$, в котором процедура локального поиска применяется к допустимому решению, возвращённому вторым алгоритмом. Приведены также результаты экспериментального исследования всех трёх разработанных алгоритмов.

1. Приближённые алгоритмы для задачи $\mathbf{GC_{\leqslant 2}}$

Через $N_G(v)$ обозначим окрестность вершины v, т. е. множество вершин графа $G = (V, E)$, смежных с вершиной v. Через $\overline{N}_G(v)$ обозначим множество вершин графа G, не смежных с v: $\overline{N}_G(v) = V \setminus (N_G(v) \cup \{v\})$.

Пусть $G_1 = (V, E_1)$ и $G_2 = (V, E_2)$ — два графа на множестве вершин V, $n = |V|$. Через $D(G_1, G_2)$ обозначим граф на множестве вершин V с множеством рёбер $E_1 \Delta E_2$. Заметим, что $\rho(G_1, G_2)$ — это количество рёбер в графе $D(G_1, G_2)$.

Следующее утверждение легко доказывается с помощью леммы о рукопожатиях.

Лемма 1 [10]. Пусть $d_{\min}$ — минимум степеней вершин в графе $D(G_1, G_2)$. Тогда

$$\rho(G_1, G_2) \geqslant n d_{\min}/2.$$

Для множеств $V_1, \ldots, V_s \subseteq V$, таких, что $V_i \cap V_j = \varnothing$ для любых $i, j \in \{1, \ldots, s\}$, $i \neq j$, и $V_1 \cup \ldots \cup V_s = V$, обозначим через $M(V_1, \ldots, V_s)$ кластерный граф из $\mathcal{M}_{\leqslant s}(V)$ с компонентами связности, порождёнными множествами $V_1, \ldots, V_s$. Далее множества $V_1, \ldots, V_s$ будем называть *кластерами*. Некоторые из V_i могут быть пустыми.

Пусть $G = (V, E)$ — произвольный граф. Для вершины $v \in V$ и множества $A \subseteq V$ обозначим через A_v^+ количество таких вершин $u \in A$, что $(v, u) \in E$; A_v^- — число таких вершин $u \in A$, что $(v, u) \notin E$.

Будем говорить, что кластерный граф $M(\overline{V}_1, \ldots, \overline{V}_s)$ получен из графа $M(V_1, \ldots, V_s)$ путём переноса вершины $v \in V_i$ из кластера V_i в кластер V_j, если $\overline{V}_i = V_i \setminus \{v\}$, $\overline{V}_j = V_j \cup \{v\}$ и $\overline{V}_k = V_k$ для всех $k \notin \{i, j\}$.

В 2004 г. разработан 3-приближённый алгоритм 1 решения задачи $\mathbf{GC}_{\leqslant \mathbf{2}}$ [4].

Алгоритм 1. BBC

Вход: граф $G = (V, E)$.
Выход: кластерный граф $M \in \mathcal{M}_{\leqslant 2}(V)$.

1: Для каждой вершины $v \in V$ определить кластерный граф $M_v = M(V_1, V_2)$ следующим образом: $V_1 = \{v\} \cup N_G(v)$, $V_2 = \overline{N}_G(v)$ (возможно $V_2 = \varnothing$).
2: Среди всех графов M_v выбрать такой граф M, что $\rho(G, M) = \min\limits_{v \in V} \rho(G, M_v)$.

Справедлива следующая гарантированная оценка точности алгоритма **BBC**.

Теорема 1 [4]. Для любого графа $G = (V, E)$ имеет место неравенство

$$\rho(G, M) \leqslant 3\rho(G, M^*),$$

где $M^* \in \mathcal{M}_{\leqslant 2}(V)$ — оптимальное решение задачи $\mathbf{GC}_{\leqslant \mathbf{2}}$ на графе G; M — кластерный граф, построенный алгоритмом **BBC**.

В 2008 г. Т. Коулман, Дж. Саундерсон и А. Вирт [9] предложили 2-приближённый алгоритм для задачи $\mathbf{GC}_{\leqslant \mathbf{2}}$, применив процедуру локального поиска **LS** к каждому допустимому решению, полученному на шаге 1 алгоритма **BBC** (алгоритм 2).

Алгоритм 2. Процедура LS(M, X, Y)

Вход: кластерный граф $M = M(X, Y) \in \mathcal{M}_{\leqslant 2}(V)$.
Выход: кластерный граф $\overline{M} = M(\overline{X}, \overline{Y}) \in \mathcal{M}_{\leqslant 2}(V)$.

1: Для каждой вершины $v \in X$ вычислить $\delta(v) = X_v^- - X_v^+ + Y_v^+ - Y_v^-$. Выбрать такую вершину $v_0 \in X$, что $\delta(v_0) = \max\limits_{v \in X} \delta(v)$.
2: Для каждой вершины $u \in Y$ вычислить $\delta(u) = Y_u^- - Y_u^+ + X_u^+ - X_u^-$. Выбрать такую вершину $u_0 \in Y$, что $\delta(u_0) = \max\limits_{u \in Y} \delta(u)$.
3: **Если** $\delta(v_0) \leqslant 0$ и $\delta(u_0) \leqslant 0$, **то стоп**; положить $\overline{X} := X$ и $\overline{Y} := Y$, **иначе** переход на шаг 4.
4: **Если** $\delta(v_0) \geqslant \delta(u_0)$, **то** положить $X := X \setminus \{v_0\}$, $Y := Y \cup \{v_0\}$.
5: **Если** $\delta(v_0) < \delta(u_0)$, **то** положить $X := X \cup \{u_0\}$, $Y := Y \setminus \{u_0\}$. Переход на шаг 1.

С учётом **BBC** и процедуры локального поиска **LS** 2-приближённый алгоритм Коулмана, Саундерсона и Вирта примет следующий вид (алгоритм 3).

Алгоритм 3. CSW

Вход: граф $G = (V, E)$.
Выход: кластерный граф $M \in \mathcal{M}_{\leqslant 2}(V)$.

1: **Для** каждой вершины $v \in V$:
2: определить кластерный граф $M_v = M(V_1, V_2)$ следующим образом:

$$V_1 = \{v\} \cup N_G(v), \quad V_2 = \overline{N}_G(v);$$

3: выполнить процедуру $\mathbf{LS}(M_v, V_1, V_2)$. Построенный граф обозначить $\overline{M}_v$.
4: Среди всех графов $\overline{M}_v$ выбрать такой граф M, что $\rho(G, M) = \min\limits_{v \in V} \rho(G, \overline{M}_v)$.

Справедлива следующая гарантированная оценка точности алгоритма **CSW**.

Теорема 2 [9]. Для любого графа $G = (V, E)$ имеет место неравенство

$$\rho(G, M) \leqslant 2\rho(G, M^*),$$

где $M^* \in \mathcal{M}_{\leqslant 2}(V)$ — оптимальное решение задачи $\mathbf{GC_{\leqslant 2}}$ на графе G; M — кластерный граф, построенный алгоритмом **CSW**.

2. Приближённое решение задачи $\mathbf{GC_{\leqslant 3}}$

2.1. 6 - Приближённый алгоритм для задачи $\mathbf{GC_{\leqslant 3}}$

Используя идеи [4, 9], можно построить полиномиальный алгоритм 4 приближённого решения задачи $\mathbf{GC_{\leqslant 3}}$.

Алгоритм 4. $\mathbf{A_1}$

Вход: граф $G = (V, E)$, $|V| = n$.
Выход: кластерный граф $M_1 \in \mathcal{M}_{\leqslant 3}(V)$.

1: **Если** $n \leqslant 2$, **то** $M_1 := G$, **иначе** переход на шаг 2.
2: **Для** каждой вершины $w \in V$:
3: $V_1 := \{w\} \cup N_G(w)$. **Если** $V_1 = V$, **то** M_w — полный граф K_n, **иначе** переход на шаг 4.
4: Обозначить G_1 подграф графа G, порождённый множеством $V \setminus V_1$. Приближённо решить задачу $\mathbf{GC_{\leqslant 2}}$ на графе G_1 алгоритмом **CSW**, полученный кластерный граф обозначить $M = M(V_2, V_3)$ (возможно, $V_3 = \varnothing$). Положить $M_w := M(V_1, V_2, V_3)$.
5: Среди всех графов M_w выбрать ближайший к G кластерный граф M_1:

$$\rho(G, M_1) = \min_{w \in V} \rho(G, M_w).$$

Замечание 1. Трудоёмкость алгоритма $\mathbf{A_1}$ равна $O(n^6)$.

Справедлива следующая гарантированная оценка точности алгоритма $\mathbf{A_1}$.

Теорема 3. Для любого графа $G = (V, E)$ имеет место неравенство

$$\rho(G, M_1) \leqslant 6\rho(G, M^*),$$

где $M^* \in \mathcal{M}_{\leqslant 3}(V)$ — оптимальное решение задачи $\mathbf{GC_{\leqslant 3}}$ на графе G; M_1 — кластерный граф, построенный алгоритмом $\mathbf{A_1}$.

Доказательство. Пусть v — вершина минимальной степени $d_{\min}$ в графе $D = D(G, M^*)$. Рассмотрим кластерный граф $\overline{M} \in \mathcal{M}_{\leqslant 3}(V)$, полученный из M^* путём переноса $d_{\min}$ вершин в другие компоненты графа M^*: вершины множества $N_G(v) \cap N_D(v)$ переместим в компоненту, содержащую вершину v, а вершины, принадлежащие множеству $\overline{N}_G(v) \cap N_D(v)$, перенесём в любую из компонент, не содержащих v. Тогда кластер $V_1 = \{v\} \cup N_G(v)$ — один и тот же в $\overline{M}$ и в M_v.

Если $d_{\min} = 0$, то кластерный граф $\overline{M}$ совпадает с кластерным графом M^*, а значит,

$$\rho(G, \overline{M}) = \rho(G, M^*) \leqslant 3\rho(G, M^*).$$

Пусть $d_{\min} > 0$. Заметим, что при переносе $d_{\min}$ вершин значение целевой функции не может увеличиться более чем на $nd_{\min}$, так как перенос одной вершины увеличивает значение целевой функции не более чем на $n = |V|$. Отсюда с учётом леммы 1 получаем

$$\rho(G, \overline{M}) \leqslant \rho(G, M^*) + nd_{\min} \leqslant \rho(G, M^*) + 2\rho(G, M^*) = 3\rho(G, M^*).$$

Итак,

$$\rho(G, \overline{M}) \leqslant 3\rho(G, M^*) \tag{1}$$

Докажем, что $\rho(G, M_v) \leqslant 6\rho(G, M^*)$. Рассмотрим кластер $V_1 = \{v\} \cup N_G(v)$. Возможны два случая:

а) $V_1 = V$. В соответствии с шагом 3 алгоритма $\mathbf{A_1}$ полагаем $M_v := K_n = \overline{M}$. Следовательно,

$$\rho(G, M_v) = \rho(G, \overline{M}) \leqslant 3\rho(G, M^*) \leqslant 6\rho(G, M^*).$$

б) $V_1 \neq V$. Обозначим через M_1^* оптимальное решение задачи $\mathbf{GC}_{\leqslant \mathbf{2}}$ на графе G_1, порождённом множеством $V \setminus V_1$. Рассмотрим кластерный граф $\widetilde{M} = M(V_1) \cup M_1^*$, где $M(V_1)$ — полный граф на множестве V_1. Очевидно, что

$$\rho(G, \widetilde{M}) \leqslant \rho(G, \overline{M}).$$

Тогда с учётом (1)

$$\rho(G, \widetilde{M}) \leqslant 3\rho(G, M^*). \tag{2}$$

Пусть $M = M(V_2, V_3)$ — допустимое решение задачи $\mathbf{GC}_{\leqslant \mathbf{2}}$ на графе G_1, найденное алгоритмом **CSW** на шаге 4, если на шаге 3 выбрана вершина v. Тогда $M_v = M(V_1, V_2, V_3)$ (множество V_3 может быть пустым).

Обозначим через s сумму числа отсутствующих рёбер в подграфе графа G, порождённом множеством вершин $V_1 = \{v\} \cup N_G(v)$, и величины разреза $(V_1, V \setminus V_1)$ в графе G. Очевидно, что

$$\rho(G, M_v) = s + \rho(G_1, M),\ \rho(G, \widetilde{M}) = s + \rho(G_1, M_1^*). \tag{3}$$

Если $|V \setminus V_1| \leqslant 2$, то $M = G_1$ и $\rho(G_1, M) = \rho(G_1, M_1^*) = 0$. Тогда с учётом (2) и (3)

$$\rho(G, M_v) = \rho(G, \widetilde{M}) \leqslant 3\rho(G, M^*) \leqslant 6\rho(G, M^*).$$

Пусть $|V \setminus V_1| \geqslant 3$. Тогда по теореме 2

$$\rho(G_1, M) \leqslant 2\rho(G_1, M_1^*).$$

В силу (2) и (3) получим

$$\rho(G, M_v) = s + \rho(G_1, M) \leqslant s + 2\rho(G_1, M_1^*) \leqslant 2s + 2\rho(G_1, M_1^*) = \\ = 2(s + \rho(G_1, M_1^*)) = 2\rho(G, \widetilde{M}) \leqslant 6\rho(G, M^*).$$

На шаге 5 алгоритма $\mathbf{A_1}$ среди всех графов будет рассмотрен граф M_v, где v — вершина минимальной степени в графе $D = D(G, M^*)$. Отсюда получается требуемая оценка точности алгоритма $\mathbf{A_1}$. ■

В п. 2.2 представлен ещё один полиномиальный алгоритм приближённого решения задачи $\mathbf{GC}_{\leqslant \mathbf{3}}$ с лучшей гарантированной оценкой точности, основанный на другой идее и не использующий локальный поиск.

2.2. Приближённый алгоритм для задачи $\mathbf{GC}_{\leqslant \mathbf{3}}$, не использующий локальный поиск

Лемма 2. Пусть $G = (V, E)$ — n-вершинный граф, $n \geqslant 3$, $M = M(V_1, \ldots, V_s)$, вершина $v \in V_i$ такова, что её окрестность в графе $D = D(G, M)$ не пуста. Рассмотрим кластерный граф $M' \in \mathcal{M}_{\leqslant s}(V)$, полученный из графа M следующим образом: либо произвольная вершина $u \in N_G(v) \cap N_D(v)$ перенесена в кластер V_i, либо произвольная вершина $u \in \overline{N}_G(v) \cap N_D(v)$ перенесена в любой из кластеров V_j, не содержащий вершину v. Тогда

$$\rho(G, M') \leqslant \rho(G, M) + (n - 3).$$

Доказательство. Пусть $u \in N_D(v)$. Очевидно, что графы D и $D' = D(G, M')$ отличаются только рёбрами вида (u, w), $w \in V$, остальные рёбра у них совпадают. Следовательно,

$$\rho(G, M') - \rho(G, M) = |N_{D'}(u)| - |N_D(u)|.$$

Так как $v \in N_D(u)$, а $N_{D'}(u) \subseteq V \setminus \{u, v\}$, то $|N_D(u)| \geqslant 1$, $|N_{D'}(u))| \leqslant n - 2$, откуда

$$\rho(G, M') - \rho(G, M) \leqslant n - 2 - 1 = n - 3.$$

Лемма 2 доказана. ■

Пусть $M^* = M(V_1^*, V_2^*, V_3^*) \in \mathcal{M}_{\leqslant 3}(V)$ — оптимальное решение задачи $\mathbf{GC}_{\leqslant \mathbf{3}}$ на графе $G = (V, E)$, причём $M^* \neq K_n$, где $n = |V|$. Через $d_D(v)$ обозначим степень вершины v в графе $D = D(G, M^*)$.

Положим $d_1 = \min\{d_D(v) : v \in V\} = d_D(v_1)$; обозначим V_1^* тот кластер графа M^*, который содержит вершину v_1, $n_1 = |V_1^*|$. Пусть $d_2 = \min\{d_D(v) : v \in V \setminus V_1^*\} = d_D(v_2)$, V_2^* — тот кластер графа M^*, который содержит вершину v_2. Очевидно, что $d_1 \leqslant d_2$.

Так как $\rho(G, M^*)$ равно числу рёбер в графе $D(G, M^*)$, из леммы о рукопожатиях и леммы 1 вытекает

Лемма 3. Для произвольного n-вершинного графа $G = (V, E)$ имеет место неравенство

$$\rho(G, M^*) \geqslant \frac{d_1 n_1 + d_2(n - n_1)}{2} \geqslant \frac{n d_1}{2},$$

где $M*$ — оптимальное решение задачи $\mathbf{GC}_{\leqslant \mathbf{3}}$ на графе G.

Алгоритм 5 основан на идее, отличной от той, что использована в [4, 9] для приближённого решения задачи $\mathbf{GC}_{\leqslant \mathbf{2}}$.

Замечание 2. Трудоёмкость алгоритма $\mathbf{A_2}$ равна $O(n^3)$.

Алгоритм 5. $\mathbf{A_2}$

Вход: граф $G = (V, E)$, $n = |V|$, $n \geqslant 3$.
Выход: кластерный граф $M_2 \in \mathcal{M}_{\leqslant 3}(V)$.

1: **Для** каждой упорядоченной пары вершин $(u, v) \in V \times V$, такой, что $u \neq v$, выполнить:
2: положить $V_1 = \{u\} \cup (N_G(u) \setminus \{v\})$;
3: обозначить через G_1 подграф графа G, порождённый множеством $V \setminus V_1$. Положить $V_2 = \{v\} \cup N_{G_1}(v)$, $V_3 = V \setminus (V_1 \cup V_2)$ (V_3 может быть пустым). Положить $M_{uv} = M(V_1, V_2, V_3)$.
4: Среди построенных графов M_{uv} и графа K_n выбрать ближайший к G кластерный граф $M_2 \in \mathcal{M}_{\leqslant 3}(V)$:

$$\rho(G, M_2) = \min_{(u,v) \in V \times V} \{\rho(G, M_{uv}), \rho(G, K_n)\},$$

где минимум берется по всем парам $(u, v) \in V \times V$, таким, что $u \neq v$.

Справедлива следующая гарантированная оценка точности алгоритма $\mathbf{A_2}$.

Теорема 4. При $n \geqslant 3$ для любого n-вершинного графа $G = (V, E)$ имеет место неравенство

$$\rho(G, M_2) \leqslant \left(6 - \frac{12}{n}\right) \rho(G, M^*),$$

где $M^* \in \mathcal{M}_{\leqslant 3}(V)$ — оптимальное решение задачи $\mathbf{GC}_{\leqslant \mathbf{3}}$ на графе G; M_2 — кластерный граф, построенный алгоритмом $\mathbf{A_2}$.

Доказательство. Возможны два случая.

С л у ч а й 1: $M^* = K_n$. Тогда на шаге 2 алгоритма $\mathbf{A_2}$ в качестве M_2 будет выбран граф K_n, следовательно,

$$\rho(G, M_2) = \rho(G, M^*) \leqslant \left(6 - \frac{12}{n}\right) \rho(G, M^*).$$

С л у ч а й 2: $M^* = M(V_1^*, V_2^*, V_3^*) \neq K_n$. Пусть, как и ранее, $v_1, v_2 \in V$ — такие вершины, что $d_D(v_1) = d_1$, $d_D(v_2) = d_2$, причём $v_1 \in V_1^*$, $v_2 \in V_2^*$. Рассмотрим кластерный граф $M \in \mathcal{M}_{\leqslant 3}(V)$, полученный из графа M^* путём переноса не более чем d_1 вершин в другие кластеры графа M^*: вершины множества $(N_G(v_1) \setminus \{v_2\}) \cap N_D(v_1)$ перенесены в кластер, содержащий вершину v_1 (т. е. в кластер V_1^*), а вершины множества $\overline{N}_G(v_1) \cap N_D(v_1)$ — в кластер, не содержащий ни v_1, ни v_2 (т. е. в кластер V_3^*). Пусть $M = M(V_1, V_2, V_3)$, где $V_1 = \{v_1\} \cup (N_G(v_1) \setminus \{v_2\})$ и V_2 — кластер, содержащий вершину v_2.

Если $d_1 > 0$, то, применив не более чем d_1 раз лемму 2, получим

$$\rho(G, M) \leqslant \rho(G, M^*) + d_1(n - 3).$$

Если $d_1 = 0$, то граф M совпадает с графом M^*, а значит,

$$\rho(G, M) = \rho(G, M^*) = \rho(G, M^*) + d_1(n - 3).$$

Итак, показано, что в любом случае

$$\rho(G, M) \leqslant \rho(G, M^*) + d_1(n - 3). \tag{4}$$

Заметим, что $V_1 \neq V$, поскольку $v_2 \notin V_1$. Пусть G_1 — подграф графа G, порождённый множеством вершин $V \setminus V_1$. Рассмотрим кластерный граф $M_{v_2} \in \mathcal{M}_{\leqslant 2}(V \setminus V_1)$, построенный следующим образом: вершина v_2 и все смежные с ней вершины графа G_1 принадлежат одному кластеру графа M_{v_2}, а все несмежные с v_2 вершины — другому кластеру графа M_{v_2} (этот кластер может быть пустым).

Пусть $M_1 \in \mathcal{M}_{\leqslant 2}(V \setminus V_1)$ — подграф графа M, порождённый множеством вершин $V \setminus V_1$, т. е. $M_1 = M(V_2, V_3)$. Положим $D_1 = D(G_1, M_1)$. Обозначим через $\overline{d}_2$ степень вершины v_2 в графе D_1. Очевидно, что кластерный граф M_{v_2} получен из графа M_1 путём переноса $\overline{d}_2$ вершин: вершины множества $N_{G_1}(v_2) \cap N_{D_1}(v_2)$ перенесены в кластер, содержащий вершину v_2 (т. е. в компоненту, порождённую множеством V_2), а все вершины множества $\overline{N}_{G_1}(v_2) \cap N_{D_1}(v_2)$ — в кластер, не содержащий v_2 (т. е. в компоненту, порождённую множеством V_3).

Если $|V \setminus V_1| \leqslant 2$, то, очевидно, $M_{v_2} = G_1$ и $\rho(G_1, M_{v_2}) = 0$. Если при этом $\overline{d}_2 = 0$ то $M_{v_2} = M_1$. Таким образом,

$$\rho(G_1, M_{v_2}) = \rho(G_1, M_1) = \rho(G_1, M_1) + \overline{d}_2(|V \setminus V_1| - 3).$$

Пусть теперь $\overline{d}_2 > 0$. Это возможно только при $|V \setminus V_1| = 2$. В этом случае $\overline{d}_2 = 1$. Это означает, что $\rho(G_1, M_1) = 1$. Тогда

$$\rho(G_1, M_{v_2}) = 0 = \rho(G_1, M_1) - 1 = \rho(G_1, M_1) + \overline{d}_2(|V \setminus V_1| - 3).$$

Теперь рассмотрим случай $|V \setminus V_1| \geqslant 3$. Если $\overline{d}_2 > 0$ то, применив $\overline{d}_2$ раз лемму 2, получим

$$\rho(G_1, M_{v_2}) \leqslant \rho(G_1, M_1) + \overline{d}_2(|V \setminus V_1| - 3).$$

Если $\overline{d}_2 = 0$, то графы M_{v_2} и M_1 совпадают, следовательно,

$$\rho(G_1, M_{v_2}) = \rho(G_1, M_1) = \rho(G_1, M_1) + \overline{d}_2(|V \setminus V_1| - 3).$$

Итак, показано, что в любом случае

$$\rho(G_1, M_{v_2}) \leqslant \rho(G_1, M_1) + \overline{d}_2(|V \setminus V_1| - 3).$$

Так как граф M получен из графа M^* путём переноса не более чем d_1 вершин в другие кластеры, то $|V_1| \geqslant |V_1^*| - d_1 = n_1 - d_1$, откуда

$$\rho(G_1, M_{v_2}) \leqslant \rho(G_1, M_1) + \overline{d}_2(n - n_1 + d_1 - 3). \tag{5}$$

Покажем, что $\overline{d}_2 \leqslant d_2$. В самом деле, $d_2 = |N_D(v_2)|$, $\overline{d}_2 = |N_{D_1}(v_2)|$. Множество $N_D(v_2)$ состоит из всех вершин u, таких, что либо $u \in N_G(v_2) \cap (V_1^* \cup V_3^*)$, либо $u \in \overline{N}_G(v_2) \cap V_2^*$, а множество $N_{D_1}(v_2)$ — из всех вершин u, таких, что либо $u \in N_{G_1}(v_2) \cap V_3$, либо $u \in \overline{N}_{G_1}(v_2) \cap V_2$, т. е.

$$N_D(v_2) = (N_G(v_2) \cap (V_1^* \cup V_3^*)) \cup (\overline{N}_G(v_2) \cap V_2^*),$$
$$N_{D_1}(v_2) = (N_{G_1}(v_2) \cap V_3) \cup (\overline{N}_{G_1}(v_2) \cap V_2).$$

Очевидно, что $V_3 \subseteq V_1^* \cup V_3^*$ и $V_2 \subseteq V_2^*$, а так как G_1 — подграф графа G, то $N_{G_1}(v_2) \subseteq N_G(v_2)$ и $\overline{N}_{G_1}(v_2) \subseteq \overline{N}_G(v_2)$. Следовательно,

$$N_{G_1}(v_2) \cap V_3 \subseteq N_G(v_2) \cap (V_1^* \cup V_3^*),$$
$$\overline{N}_{G_1}(v_2) \cap V_2 \subseteq \overline{N}_G(v_2) \cap V_2^*.$$

Значит, $N_{D_1}(v_2) \subseteq N_D(v_2)$, откуда получаем $\overline{d}_2 = |N_{D_1}(v_2)| \leqslant |N_D(v_2)| = d_2$.

Итак, $\overline{d}_2 \leqslant d_2$. Тогда из (5) следует, что

$$\rho(G_1, M_{v_2}) \leqslant \rho(G_1, M_1) + d_2(n - n_1 + d_1 - 3). \tag{6}$$

В соответствии с шагом 4 алгоритма $\mathbf{A_2}$ положим $M_{v_1v_2} = M(V_1) \cup M_{v_2}$, где $M(V_1)$ — полный граф на множестве V_1. Обозначим через s сумму числа отсутствующих рёбер в подграфе графа G, порождённом множеством вершин V_1, и величины разреза $(V_1, V \setminus V_1)$ в графе G. Очевидно, что

$$\rho(G, M_{v_1v_2}) = s + \rho(G_1, M_{v_2}),\ \ \rho(G, M) = s + \rho(G_1, M_1).$$

Отсюда с учётом (4) и (6) имеем

$$\begin{gathered}\rho(G, M_{v_1v_2}) = s + \rho(G_1, M_{v_2}) \leqslant s + \rho(G_1, M_1) + d_2(n - n_1 + d_1 - 3) = \\ = \rho(G, M) + d_2(n - n_1 + d_1 - 3) \leqslant \rho(G, M^*) + d_1(n-3) + d_2(n - n_1 + d_1 - 3),\end{gathered}$$

т. е.

$$\rho(G, M_{v_1v_2}) \leqslant \rho(G, M^*) + d_1(n-3) + d_2(n - n_1 + d_1 - 3). \tag{7}$$

Теперь покажем, что $d_2 \leqslant n_1 + n/2$. Действительно,

$$d_2 = |N_D(v_2| = |N_D(v_2)) \cap V_1^*| + |N_D(v_2) \cap (V_2^* \cup V_3^*|.$$

Очевидно, что $|N_D(v_2) \cap V_1^*| \leqslant |V_1^*| = n_1$. Обозначим $d' = |N_D(v_2) \cap (V_2^* \cup V_3^*)|$ и докажем, что $d' \leqslant n/2$. Предположим противное: $d' > n/2$. Рассмотрим кластерный граф $\widetilde{M}$, полученный из M^* путём переноса вершины v_2 из кластера V_2^* в кластер V_3^*. Очевидно, что $V_2 \cup V_3 = V_2^* \cup V_3^*$ и графы D и $\widetilde{D} = D(G, \widetilde{M})$ отличаются только рёбрами вида (u, v_2), где $u \in V_2^* \cup V_3^*$, остальные рёбра у них совпадают. Следовательно,

$$\rho(G, \widetilde{M}) - \rho(G, M^*) = |N_{\widetilde{D}}(v_2) \cap (V_2^* \cup V_3^*)| - |N_D(v_2) \cap (V_2^* \cup V_3^*)|.$$

Так как

$$N_{\widetilde{D}}(v_2) \cap (V_2^* \cup V_3^*) = (V_2^* \cup V_3^*) \setminus (\{v_2\} \cup N_D(v_2)) = (V_2^* \cup V_3^*) \setminus ((\{v_2\} \cup N_D(v_2)) \cap (V_2^* \cup V_3^*)),$$

получим

$$|N_{\widetilde{D}}(v_2) \cap (V_2^* \cup V_3^*)| - |N_D(v_2) \cap (V_2^* \cup V_3^*)| = (n - n_1 - d' - 1) - d' = n - n_1 - 2d' - 1.$$

По предположению, $d' > n/2$, поэтому $\rho(G, \widetilde{M}) - \rho(G, M^*) < -n_1 - 1 < 0$, что противоречит оптимальности кластерного графа M^*. Следовательно, $d' \leqslant n/2$, откуда

$$d_2 \leqslant n_1 + n/2. \tag{8}$$

Оценим слагаемые из правой части неравенства (7) с учётом (8), леммы 3 и неравенства $d_1 \leqslant d_2$.

а) Дадим оценку суммы $\rho(G, M^*) + d_1(n-3)$:

$$\begin{gathered}\rho(G, M^*) + d_1(n-3) = \rho(G, M^*) + d_1 n(1 - 3/n) \leqslant \\ \leqslant \rho(G, M^*) + 2\rho(G, M^*)(1 - 3/n) = (3 - 6/n)\rho(G, M^*).\end{gathered}$$

б) Оценим слагаемое $d_2(n - n_1 + d_1 - 3)$:

$$d_2(n - n_1 + d_1 - 3) = d_1 n_1 - d_1 n_1 + d_2(n - n_1) + d_1 d_2 - 3d_2 \leqslant$$
$$\leqslant 2\rho(G, M^*) - d_1 n_1 + d_1 d_2 - 3d_2 \leqslant 2\rho(G, M^*) - d_1 n_1 + d_1(n_1 + n/2) - 3d_2 =$$
$$= 2\rho(G, M^*) + d_1 n/2 - 3d_2 \leqslant 2\rho(G, M^*) + d_1 n/2 - 3d_1 =$$
$$= 2\rho(G, M^*) + d_1 n/2(1 - 6/n) \leqslant 2\rho(G, M^*) + \rho(G, M^*)(1 - 6/n) = (3 - 6/n)\rho(G, M^*).$$

Итак,

$$\rho(G, M_{v_1 v_2}) \leqslant (6 - 12/n)\rho(G, M^*).$$

На шаге 4 алгоритмом $\mathbf{A_2}$ среди всех графов M_{uv} будет, в частности, рассмотрен и граф $M_{v_1 v_2}$. Отсюда получается требуемая оценка точности алгоритма $\mathbf{A_2}$. ■

3. Вычислительный эксперимент

Для сравнения точности приближённых алгоритмов $\mathbf{A_1}$ и $\mathbf{A_2}$ проведён вычислительный эксперимент. Исследован также алгоритм $\mathbf{A_3}$, являющийся модификацией алгоритма $\mathbf{A_2}$ (в нём к приближённому решению, полученному алгоритмом $\mathbf{A_2}$, применяется процедура локального поиска). Очевидно, что алгоритм $\mathbf{A_3}$ полиномиален и его гарантированная оценка точности не хуже, чем гарантированная оценка точности алгоритма $\mathbf{A_2}$. Все алгоритмы реализованы на языке программирования C++ в среде JetBrains Clion 2017.3.3. Вычисления проводились на персональном компьютере модели RT I5N с процессором Intel Core i5-2400, тактовой частотой 3,10 ГГц и объёмом оперативной памяти 8 Гбайт.

Опишем класс графов, на которых производились вычисления. Зафиксируем параметр $p \in (0, 1)$, случайный n-вершинный граф $G = (V, E)$ будем получать с использованием следующей процедуры. Для каждой пары вершин (u, v) проводится независимый случайный эксперимент, исходом которого является наличие ребра uv с вероятностью p и отсутствие ребра с вероятностью $1 - p$. В [15, 16] семейство n-вершинных графов с таким распределением обозначается $G(n, p)$ и используется как при теоретическом изучении графов, так и в экспериментальных исследованиях (модель Эрдеша — Реньи).

Параметр p представляет собой математическое ожидание плотности случайного графа $G = (V, E)$, которая определяется как $2|E|/(n(n - 1))$. В проведённых экспериментах использовались значения $p \in \{0{,}33, 0{,}5, 0{,}67\}$.

На графах малой размерности (при n от 10 до 20 вершин) алгоритмом полного перебора были найдены точные решения. Это позволило получить предварительные сведения об изменении точности алгоритмов в зависимости от параметров n и p. Для каждой пары значений n и p было решено по 50 задач.

Далее *точностью* (*отклонением от оптимума*) приближённого алгоритма будем называть отношение значения целевой функции, полученного данным алгоритмом, к её оптимальному значению. Обозначим через $\delta_1(n, p)$, $\delta_2(n, p)$ и $\delta_3(n, p)$ средние точности алгоритмов $\mathbf{A_1}$, $\mathbf{A_2}$ и $\mathbf{A_3}$ для n-вершинных графов при фиксированном значении параметра p.

По итогам первого эксперимента наилучшие показатели принадлежат алгоритму $\mathbf{A_3}$: его среднее отклонение от оптимума достигает максимума при значениях параметров $n = 18$, $p = 0{,}33$ и составляет 3,6 %, в то время как алгоритмы $\mathbf{A_1}$ и $\mathbf{A_2}$ при тех же параметрах отклоняются на 7,3 и 12,3 % соответственно. При всех остальных значениях параметров n и p алгоритм $\mathbf{A_3}$ также даёт решения с наименьшим средним отклонением от оптимума среди исследуемых алгоритмов (рис. 1).

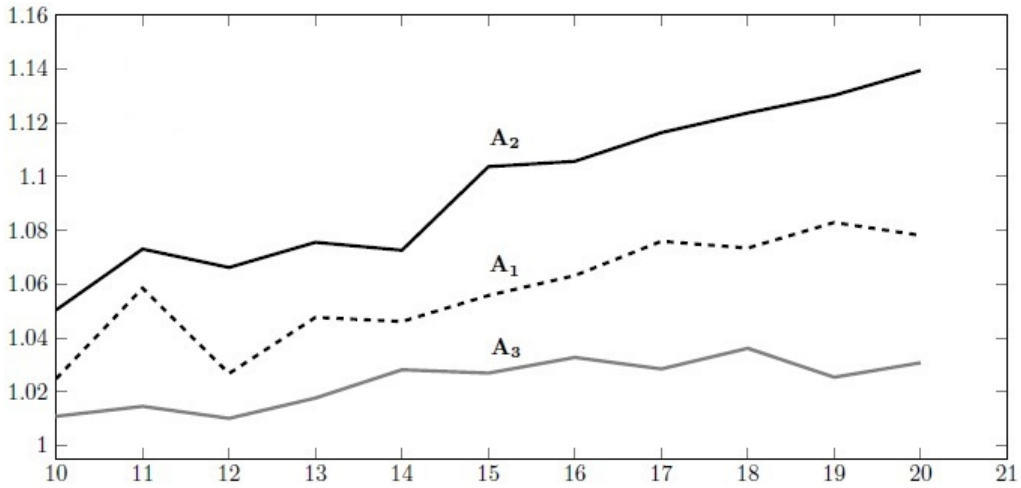


Рис. 1. Средняя точность приближённых алгоритмов при $p = 0{,}33$

Таким образом, наиболее перспективным приближённым алгоритмом можно считать алгоритм $\mathbf{A_3}$. Для подтверждения того, что он продолжает доминировать и при других значениях n, был проведён эксперимент на графах большей размерности (при n от 50 до 250 вершин). Однако в этом случае нет возможности найти оптимальное решение за приемлемое время, поэтому в качестве исследуемых были выбраны следующие величины:

$$d_1(n,p) = \frac{\delta_1(n,p)}{\delta_3(n,p)},\ d_2(n,p) = \frac{\delta_2(n,p)}{\delta_3(n,p)}.$$

Выбор таких величин объясняется тем, что, во-первых, их вычисление не требует знания оптимального решения, а во-вторых, они позволяют ответить на вопрос, действительно ли алгоритм $\mathbf{A_3}$ даёт в среднем лучшие решения, чем алгоритмы $\mathbf{A_1}$ и $\mathbf{A_2}$, на графах большей размерности. Для каждой пары значений n и p было решено по 100 задач.

При всех значениях n и p величины $d_1(n,p)$ и $d_2(n,p)$ оказались больше единицы. При этом в каждой решённой задаче алгоритм $\mathbf{A_3}$ даёт значение целевой функции не большее, чем значения целевых функций, полученных алгоритмами $\mathbf{A_1}$ и $\mathbf{A_2}$ (рис. 2). Это позволяет сделать вывод о том, что алгоритм $\mathbf{A_3}$ продолжает доминировать над алгоритмами $\mathbf{A_1}$ и $\mathbf{A_2}$ независимо от значений параметров n и p. При этом время работы алгоритмов $\mathbf{A_2}$ и $\mathbf{A_3}$ с ростом n увеличивается достаточно медленно (меньше 1 с при $n = 50$ и около 10 с при $n = 250$). Для алгоритма $\mathbf{A_1}$ рост количества вершин оказал существенное влияние на время работы (меньше 1 с при $n = 50$ и около 800 с при $n = 250$).

Из результатов вычислительного эксперимента можно сделать вывод, что наилучшим алгоритмом приближенного решения задачи $\mathbf{GC}_{\leqslant \mathbf{3}}$ является $\mathbf{A_3}$: решения, полученные этим алгоритмом, наиболее близки к оптимальным, при этом время его работы остаётся сравнительно небольшим при росте количества вершин.

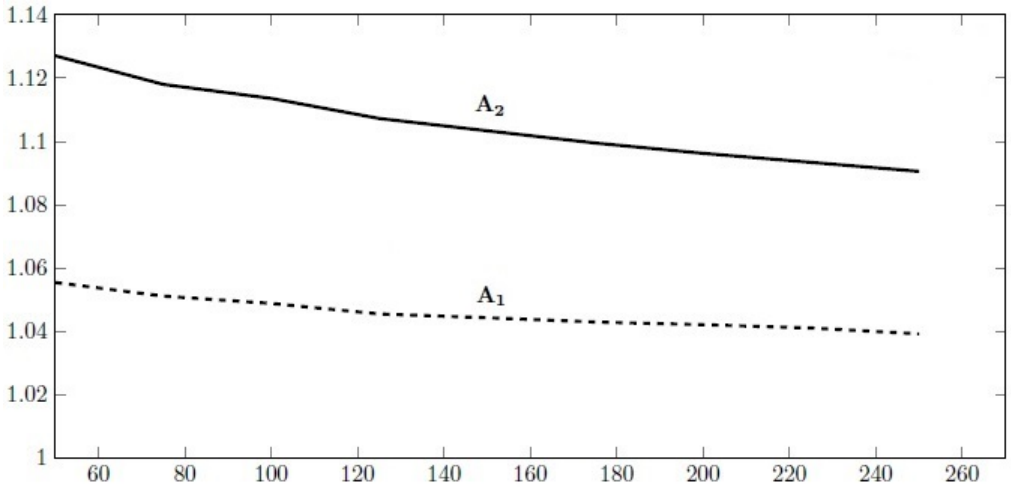


Рис. 2. Значения величин $d_1(n,p)$ и $d_2(n,p)$ при $p = 0{,}33$

Заключение

В работе рассматривается задача кластеризации графа, в которой число кластеров не превосходит трёх. Предложены три полиномиальных алгоритма приближённого решения этой задачи. Первый, 6-приближённый алгоритм $\mathbf{A_1}$, многократно использует локальный поиск и основан на идеях, применяемых для решения задачи, в которой число кластеров не превышает двух. Второй алгоритм $\mathbf{A_2}$ основан на оригинальной идее и не использует локальный поиск. Приближённый алгоритм $\mathbf{A_3}$ является модификацией второго и отличается тем, что к приближённому решению, найденному алгоритмом $\mathbf{A_2}$, применяется процедура локального поиска. Доказано, что алгоритмы $\mathbf{A_2}$ и $\mathbf{A_3}$ имеют гарантированную оценку точности $6 - 12/n$, где n — число вершин кластеризуемого графа.

ЛИТЕРАТУРА

1. *Kulis B., Basu S., Dhillon I., and Mooney R.* Semi-supervised graph clustering: a kernel approach // Machine Learning. 2009. V. 74. No. 1. P. 1–22.
2. *Schaeffer S. E.* Graph clustering // Computer Science Review. 2005. V. 1. No. 1. P. 27–64.
3. *Shamir R., Sharan R., and Tsur D.* Cluster graph modification problems // Discr. Appl. Math. 2004. V. 144. No. 1-2. P. 173–182.
4. *Bansal N., Blum A., and Chawla S.* Correlation clustering // Machine Learning. 2004. V. 56. P. 89–113.
5. *Ben-Dor A., Shamir R., and Yakhimi Z.* Clustering gene expression patterns // J. Comput. Biol. 1999. V. 6. No. 3-4. P. 281–297.
6. *Křivánek M. and Morávek J.* NP-hard problems in hierarchical-tree clustering // Acta Informatica. 1986. V. 23. P. 311–323.
7. *Giotis I. and Guruswami V.* Correlation clustering with a fixed number of clusters // Theory of Computing. 2006. V. 2. No. 1. P. 249–266.
8. *Агеев А. А., Ильев В. П., Кононов А. В., Талевнин А. С.* Вычислительная сложность задачи аппроксимации графов // Дискрет. анализ и исслед. операций. Сер. 1. 2006. Т. 13. № 1. С. 3–11.
9. *Coleman T., Saunderson J., and Wirth A.* A local-search 2-approximation for 2-correlation-clustering // LNCS. 2008. V. 5193. P. 308–319.
10. *Ильев В. П., Ильева С. Д., Навроцкая А. А.* Приближенные алгоритмы для задач аппроксимации графов // Дискрет. анализ и исслед. операций. 2011. Т. 18. № 1. С. 41–60.

11. *Charikar M., Guruswami V., and Wirth A.* Clustering with qualitative information // J. Comput. Syst. Sci. 2005. V. 71. No. 3. P. 360–383.
12. *Ailon N., Charikar M., and Newman A.* Aggregating inconsistent information: Ranking and clustering // J. ACM. 2008. V. 55. No. 5. P. 1–27.
13. *Chawla S., Makarychev K., Schramm T., and Yaroslavtsev G.* Near optimal LP algorithm for correlation clustering on complete and complete k-partite graphs // STOC'15 Symp. on Theory of Computing. N.Y.: ACM, 2015. P. 219–228.
14. *Il'ev V., Il'eva S., and Kononov A.* Short survey on graph correlation clustering with minimization criteria // LNCS. 2016. V. 9869. P. 25–36.
15. *Эрдёш П., Спенсер Дж.* Вероятностные методы в комбинаторике. М.: Мир, 1976.
16. *Alon N. and Spencer J.* The Probabilistic Method. N.Y.: Wiley and Sons, 1992.

REFERENCES

1. *Kulis B., Basu S., Dhillon I., and Mooney R.* Semi-supervised graph clustering: a kernel approach. Machine Learning, 2009, vol. 74, no. 1, pp. 1–22.
2. *Schaeffer S. E.* Graph clustering. Computer Science Review, 2005, vol. 1, no. 1, pp. 27–64.
3. *Shamir R., Sharan R., and Tsur D.* Cluster graph modification problems. Discr. Appl. Math., 2004, vol. 144, no. 1-2, pp. 173–182.
4. *Bansal N., Blum A., and Chawla S.* Correlation clustering. Machine Learning, 2004, vol. 56, pp. 89–113.
5. *Ben-Dor A., Shamir R., and Yakhimi Z.* Clustering gene expression patterns. J. Comput. Biol., 1999, vol. 6, no. 3-4, pp. 281–297.
6. *Křivánek M. and Morávek J.* NP-hard problems in hierarchical-tree clustering. Acta Informatica, 1986, vol. 23, pp. 311–323.
7. *Giotis I. and Guruswami V.* Correlation clustering with a fixed number of clusters. Theory of Computing, 2006, vol. 2, no. 1, pp. 249–266.
8. *Ageev A. A., Il'ev V. P., Kononov A. V., and Talevnin A. S.* Computational complexity of the graph approximation problem. J. Appl. Indust. Math., 2007, vol. 1, no. 1, pp. 1–8.
9. *Coleman T., Saunderson J., and Wirth A.* A local-search 2-approximation for 2-correlation-clustering. LNCS, 2008, vol. 5193, pp. 308–319.
10. *Il'ev V. P., Il'eva S. D., and Navrotskaya A. A.* Approximation algorithms for graph approximation problems. J. Appl. Indust. Math., 2011, vol. 5, no. 4, pp. 569–581.
11. *Charikar M., Guruswami V., and Wirth A.* Clustering with qualitative information. J. Comput. Syst. Sci., 2005, vol. 71, no. 3, pp. 360–383.
12. *Ailon N., Charikar M., and Newman A.* Aggregating inconsistent information: Ranking and clustering. J. ACM, 2008, vol. 55, no. 5, pp. 1–27.
13. *Chawla S., Makarychev K., Schramm T., and Yaroslavtsev G.* Near optimal LP algorithm for correlation clustering on complete and complete k-partite graphs. STOC'15 Symp. on Theory of Computing, New York, ACM, 2015, pp. 219–228.
14. *Il'ev V., Il'eva S., and Kononov A.* Short survey on graph correlation clustering with minimization criteria. LNCS, 2016, vol. 9869, pp. 25–36.
15. *Erdös P. and Spencer J.* Probabilistic Methods in Combinatorics. Budapest, Académiai Kiadó, 1974.
16. *Alon N. and Spencer J.* The Probabilistic Method. New York, Wiley and Sons, 1992.

UDC 519.1 DOI 10.17223/20710410/45/8

ON THE ONE EDGE ALGORITHM FOR THE ORTHOGONAL DOUBLE COVERS

R. El-Shanawany, A. El-Mesady

Menoufia University, Menouf, Egypt

E-mail: ramadan_elshanawany380@yahoo.com, ahmed_mesady88@yahoo.com

The existing problem of the orthogonal double covers of the graphs is well-known in the theory of combinatorial designs. In this paper, a new technique called the one edge algorithm for constructing the orthogonal double covers of the complete bipartite graphs by copies of a graph is introduced. The advantage of this algorithm is that it is accessible to discrete mathematicians not intimately familiar with the theory of the orthogonal double covers.

Keywords: *graph decomposition, symmetric starter, orthogonal double covers.*

Nomenclature:

K_m	complete graph on m vertices;
$K_{m,n}$	complete bipartite graph with independent sets of sizes m and n;
mG	m disjoint copies of G ;
$C_k(n_1, n_2, \ldots, n_k)$	caterpillar (tree) obtained from the path $P_k = x_1x_2 \ldots x_k$ by joining vertex x_i to n_i new vertices where $k \geqslant 1$, $n_1, n_2, \ldots, n_k$ are positive integers, $n_1, n_k \geqslant 1$ and $n_i \geqslant 0$ for $i \in \{2, 3, \ldots, k-1\}$.

Introduction

Graphs serve as a mathematical model to solve many real-world problems successfully. Some problems in chemistry, physics, computer technology, communication science, psychology, genetics, linguistics, and sociology can be formulated as problems in graph theory. Also, many branches of mathematics, such as topology, probability, group theory, and matrix theory, have close connections with the graph theory. Some puzzles of a practical nature have been instrumental in the development of various topics in graph theory. The cyclic graphs theory was developed for solving many problems of electrical networks, and the study of "trees" is considered a helping tool for enumerating isomers of organic compounds.

An *orthogonal double cover* (ODC) of H by G is a collection $\mathcal{G} = \{\pi(x) : x \in V(H)\}$ of isomorphic subgraphs (to G) of H (called pages) such that (i) every edge of H is contained in exactly two pages of $\mathcal{G}$ and (ii) $\pi(a)$ and $\pi(b)$share an edge if and only if a and b are adjacent in H.

The existence problem of ODC has attracted much attention during the last few years. The problem is known to be hard in general as it includes some long-standing open problems like the existing problems of biplanes. The ODC problem originally stems from problems in database optimization, statistical design of experiments, and design theory. In [1, 2], J. Demetrovics et al. have inspected the key of the Armstrong databases of minimum size. The ODC of K_n whose elements consist of distinct cliques is equivalent to Armstrong database of size n. Also, ODCs are related to several graph decomposition problems [3, 4]. The initial interest was concerned with the complete graphs, but the specialists have solved

the ODC problem for several graphs such as the Cayley graphs [5] and the complete bipartite graphs (e.g., [6, 7]).

The labeling of the vertices of the complete bipartite graph $K_{n,n}$ is defined by the bijective mapping $\phi : V(K_{n,n}) \to \mathbb{Z}_n \times \mathbb{Z}_2$. The product $(v, j) \in \mathbb{Z}_n \times \mathbb{Z}_2$ will be written as v_j referring to the corresponding vertex and the edge $(c_\gamma, d_\delta) \in E(K_{n,n})$ if $\gamma \neq \delta$ for all $c, d \in \mathbb{Z}_n$ and $\gamma, \delta \in \mathbb{Z}_2$. We shall denote by (c, d) the edge between the vertices c_0 and d_1. If G is a subgraph of $K_{n,n}$ and $i \in \mathbb{Z}_n$, then $G + i$ is called i-translate of G. The edges of $G + i$ are obtained from G by rotating the edges of G, i.e., by mapping the edge (c, d) in G to the edge $(c + i, d + i)$ in $G + i$ (with calculations modulo n). If $e = (p, q) \in E(G)$, then it has a length defined by $d(e) = q - p$ (with calculations modulo n). For the graph G, if $|E(G)| = n$ and the lengths of all edges in G are mutually distinct and equal to $\mathbb{Z}_n$, then G is said to be a half starter w.r.t. $\mathbb{Z}_n$. R. El-Shanawany et al. [8] have proved the following three results.

A. The union of all translates of G forms an edge decomposition of $K_{n,n}$, i.e., $\bigcup_{z \in \mathbb{Z}_n} E(G+z) = E(K_{n,n})$, if and only if G is a half starter.

In what follows, a half starter G will be represented by the vector $v(G) = (v_0, v_1, \ldots, v_{n-1}) \in \mathbb{Z}_n^n$ where $v_i, i \in \mathbb{Z}_n$ and $(v_i)_0$ is the unique vertex $((v_i, 0) \in \mathbb{Z}_n \times \{0\})$ that belongs to the unique edge of length i in G. The two half starter vectors $v(G_0)$ and $v(G_1)$ are said to be orthogonal if $\{v_i(G_0) - v_i(G_1) : i \in \mathbb{Z}_n\} = \mathbb{Z}_n$.

B. If $v(G_0)$ and $v(G_1)$ are orthogonal two half starter vectors, then $\mathcal{G} = \{G_{z,l} : (z, l) \in \in \mathbb{Z}_n \times \mathbb{Z}_2\}$ with $G_{z,l} = G_l + z$ is considered the ODC of $K_{n,n}$.

If the graph G_s is a subgraph of $K_{n,n}$ with $E(G_s) = \{(p_0, q_1) : (q_0, p_1) \in E(G)\}$, then G_s is the symmetric graph of G. It is easy to prove that if G is a half starter, then G_s is also a half starter. Also, if $v(G)$ and $v(G_s)$ are orthogonal, then the half starter G is called a symmetric starter w.r.t. $\mathbb{Z}_n$.

C. If n is any positive integer and G is a half starter with $v(G) = (v_0, v_1, \ldots, v_{n-1})$, then G is symmetric starter iff $\{v_i - v_{-i} + i : i \in \mathbb{Z}_n\} = \mathbb{Z}_n$.

For illustrative purposes, Fig. 1 exhibits the graph $K_{4,4}$ and its ODC by $K_{1,4}$.

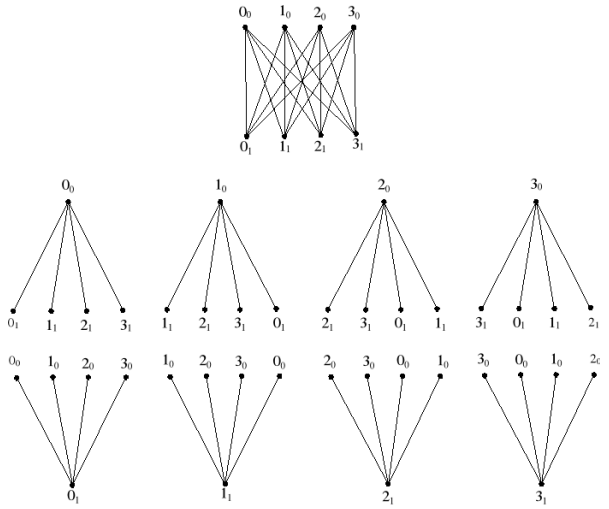


Fig. 1. $K_{4,4}$ and its ODC by $K_{1,4}$

The aim of this paper is to present a new algorithm which is called the one edge algorithm. This algorithm is a helping tool for constructing the ODCs of the complete

bipartite graphs. In addition, we present a number of new results as a direct application of this algorithm. The difference between our paper and [9] is as follows. In [9], El-Serafi et al. constructed the orthogonal double cover of complete bipartite graphs by a complete bipartite graph and the disjoint union of complete bipartite graphs with the cartesian product of symmetric starter vectors. For several results of ODCs by different graph classes, see [10 – 12].

1. Main results

In this section, we construct the orthogonal double covers of the complete bipartite graphs by copies of stars, copies of a caterpillar, and by copies of a complete bipartite graph.

Definition 1. For $k \geqslant 2$, a caterpillar graph $C_k(m_1, m_2, \ldots, m_k)$ is obtained from a path $P_k = v_1 v_2 \ldots v_k$ by attaching $n_i \geqslant 0$ pendant vertices $u_{i,j}$ $(1 \leqslant j \leqslant m_i)$ to each v_i.

Definition 2. A complete bipartite graph of the form $K_{1,n}$ is called a star.

In what follows, we prove the theorems on existence of ODCs for particular classes of complete bipartite graphs $K_{mn,mn}$ and $K_{2mn,2mn}$ with $n > 0$ and $m \equiv 1, 5 \pmod 6$. Really, our effort was concentrated on these classes and in the future we hope to solve the problem for the other classes to be used in the design theory. The proofs of these theorems base on the direct constructions of ODCs. The main strengths of represented algorithms is that they are clear for discrete mathematicians not intimately familiar with the theory of the orthogonal double covers.

Theorem 1. Let m and n be positive integers with $m \equiv 1, 5 \pmod 6$. Then, there is an ODC of $K_{mn,mn}$ by $mK_{1,n}$.

Proof. Algorithm 1 proves the Theorem 1. ■

Algorithm 1.	ODC of $K_{mn,mn}$ by $mK_{1,n}$
1: [Inauguration.]	Choose the values $m, n \in \mathbb{N}$, $m \equiv 1, 5 \pmod 6$.
2: [Initial edge.]	Construct the edge $(0, 0)$.
3: [Complementary edges.]	Add $(0, -\beta)$ to $(0, 0)$ where $\beta \in \mathbb{Z}_n \setminus \{0\}$.
4: [Initial graph $\cong G^0$.]	The union of the edge set in steps 2, 3 gives the edge set of $K_{1,n}$.
5: [Complementary graphs $\cong \bigcup_{\alpha=1}^{m-1} G^\alpha$.]	Add $(\alpha n, 2\alpha n)$ to the edge set of step 4 where $\alpha \in \mathbb{Z}_m \setminus \{0\}$.
6: [Symmetric starter $\cong \bigcup_{\alpha=0}^{m-1} G^\alpha$.]	The union of the edge set in steps 4, 5 gives the edge set of $G \cong mK_{1,n}$.
7: [Symmetric graph.]	If the edge (a, b) belongs to the edge set of step 6, then (b, a) is an edge of the graph $G_s \cong mK_{1,n}$.
8: [Translation.]	Add (γ, γ) to the edge set in steps 6, 7 where $\gamma \in \mathbb{Z}_{mn}$.
9: [Done.]	Output the edge set of step 8 which represents the ODC of $K_{mn,mn}$ by $mK_{1,n}$.

For more illustration to Theorem 1, let $m = 5$ and $n = 3$, then there is an ODC of $K_{15,15}$ by $5K_{1,3}$, see Fig. 2.

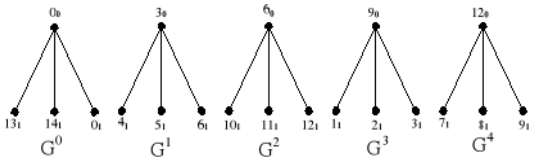


Fig. 2. Symmetric starter of an ODC of $K_{15,15}$ by $5K_{1,3}$

Theorem 2. Let m and $n \geqslant 5$ be positive integers with $m \equiv 1, 5 \pmod 6$. Then, there is an ODC of $K_{mn,mn}$ by $2mK_2 \cup mK_{1,n-2}$.

Proof. Algorithm 2 proves the Theorem 2. ■

Algorithm 2.	ODC of $K_{mn,mn}$ by $2mK_2 \cup mK_{1,n-2}$
1: [Inauguration.]	Choose the values of m and n where m and $n \geqslant 5$ are positive integers with $m \equiv 1, 5 \pmod 6$.
2: [Initial edge.]	Construct the edge $(0, 0)$.
3: [Complementary edges.]	Add $(1 + n, 2n - 1), (2n - 1, 3n - 2)$, and $(0, -2\beta)$ to $(0, 0)$ where $\beta \in \mathbb{Z}_n \setminus \mathbb{Z}_3$.
4: [Initial graph $\cong G^0$.]	The union of the edge set in steps 2, 3 gives the edge set of $2K_2 \cup K_{1,n-2}$.
5: [Complementary graphs $\cong \bigcup_{\alpha=1}^{m-1} G^\alpha$.]	Add $(\alpha n, 2\alpha n)$ to the edge set of step 4 where $\alpha \in \mathbb{Z}_m \setminus \{0\}$.
6: [Symmetric starter of $G \cong \bigcup_{\alpha=0}^{m-1} G^\alpha$.]	The union of the edge set in steps 4, 5 gives the edge set $\cong 2mK_2 \cup mK_{1,n-2}$.
7: [Symmetric graph.]	If the edge (a, b) belongs to the edge set of step 6, then (b, a) is an edge of the graph $G_s \cong 2mK_2 \cup mK_{1,n-2}$.
8: [Translation.]	Add (γ, γ) to the edge set in steps 6, 7 where $\gamma \in \mathbb{Z}_{mn}$.
9: [Done.]	Output the edge set of step 8 which represents the ODC of $K_{mn,mn}$ by $2mK_2 \cup mK_{1,n-2}$.

For more illustration to Theorem 2, let $m = 5$ and $n = 5$, then there is an ODC of $K_{25,25}$ by $5K_{1,3} \cup 10K_2$, see Fig. 3.

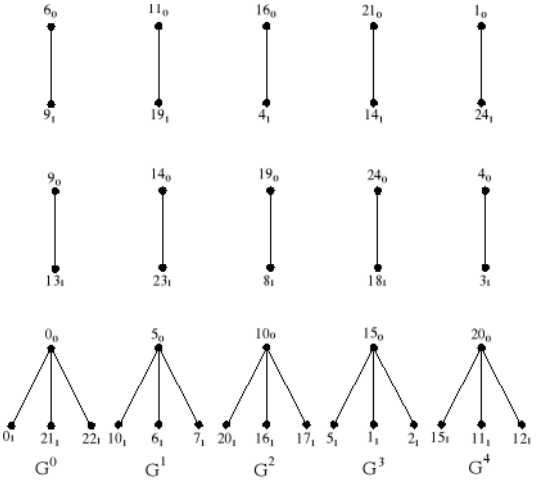


Fig. 3. Symmetric starter of an ODC of $K_{25,25}$ by $5K_{1,3} \cup 10K_2$

Theorem 3. Let m and $n \geqslant 4$ be positive integers with $m \equiv 1, 5 \pmod 6$. Then, there is an ODC of $K_{mn,mn}$ by $mC_2(1, n - 2)$.

Proof. Algorithm 3 proves the Theorem 3. ■

Algorithm 3.	ODC of $K_{mn,mn}$ by $mC_2(1, n-2)$
1: [Inauguration.]	Choose the values of m and n where m and $n \geqslant 4$ are positive integers with $m \equiv 1, 5 \pmod 6$.
2: [Initial edge.]	Construct the edge $(0, 0)$.
3: [Complementary edges.]	Add $(1, 1+\beta-n)$ to $(0, 0)$ where $\beta \in \mathbb{Z}_n \setminus \{0\}$.
4: [Initial graph $\cong G^0$.]	The union of the edge set in steps 2, 3 gives the edge set of $C_2(1, n-2)$.
5: [Complementary graphs $\cong \bigcup_{\alpha=1}^{m-1} G^\alpha$.]	Add $(\alpha n, 2\alpha n)$ to the edge set of step 4 where $\alpha \in \mathbb{Z}_m \setminus \{0\}$.
6: [Symmetric starter $\cong \bigcup_{\alpha=0}^{m-1} G^\alpha$.]	The union of the edge set in steps 4, 5 gives the edge set of $G \cong mC_2(1, n-2)$.
7: [Symmetric graph.]	If the edge (a, b) belongs to the edge set of step 6, then (b, a) is an edge of the graph $G_s \cong mC_2(1, n-2)$.
8: [Translation.]	Add (γ, γ) to the edge set in steps 6, 7 where $\gamma \in \mathbb{Z}_{mn}$.
9: [Done.]	Output the edge set of step 8 which represents the ODC of $K_{mn,mn}$ by $mC_2(1, n-2)$.

For more illustration to Theorem 3, let $m = 5$ and $n = 4$, then there is an ODC of $K_{20,20}$ by $5C_2(1, 2)$, see Fig. 4.

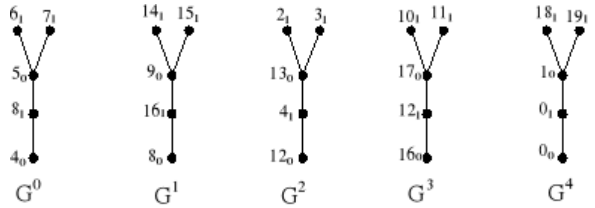


Fig. 4. Symmetric starter of an ODC of $K_{20,20}$ by $5C_2(1, 2)$

Theorem 4. Let m and $n \geqslant 2$ be positive integers with $m \equiv 1, 5 \pmod 6$. Then, there is an ODC of $K_{2mn,2mn}$ by $mK_{2,n}$.

Proof. Algorithm 4 proves the Theorem 4. ■

Algorithm 4.	ODC of $K_{2mn,2mn}$ by $mK_{2,n}$
1: [Inauguration.]	Choose the values of m and n where m and $n \geqslant 2$ are positive integers with $m \equiv 1, 5 \pmod 6$.
2: [Initial edge.]	Construct the edge $(0, 0)$.
3: [Complementary edges.]	Add $(0, -2\beta)$ and $(1, -2\beta)$ to $(0, 0)$ where $\beta \in \mathbb{Z}_n \setminus \{0\}$.
4: [Initial graph $\cong G^0$.]	The union of the edge set in steps 2, 3 gives the edge set of $K_{2,n}$.
5: [Complementary graphs $\cong \bigcup_{\alpha=1}^{m-1} G^\alpha$.]	Add $(2\alpha n, 4\alpha n)$ to the edge set of step 4 where $\alpha \in \mathbb{Z}_m \setminus \{0\}$.
6: [Symmetric starter $\cong \bigcup_{\alpha=0}^{m-1} G^\alpha$.]	The union of the edge set in steps 4, 5 gives the edge set of $G \cong mK_{2,n}$.
7: [Symmetric graph.]	If the edge (a, b) belongs to the edge set of step 6, then (b, a) is an edge of the graph $G_s \cong mK_{2,n}$.

8: [Translation.]	Add (γ, γ) to the edge set in steps 6, 7 where $\gamma \in \mathbb{Z}_{2mn}$.
9: [Done.]	Output the edge set of step 8 which represents the ODC of $K_{2mn,2mn}$ by $mK_{2,n}$.

For more illustration to Theorem 4, let $m = 5$ and $n = 3$, then there is an ODC of $K_{30,30}$ by $5K_{2,3}$, see Fig. 5.

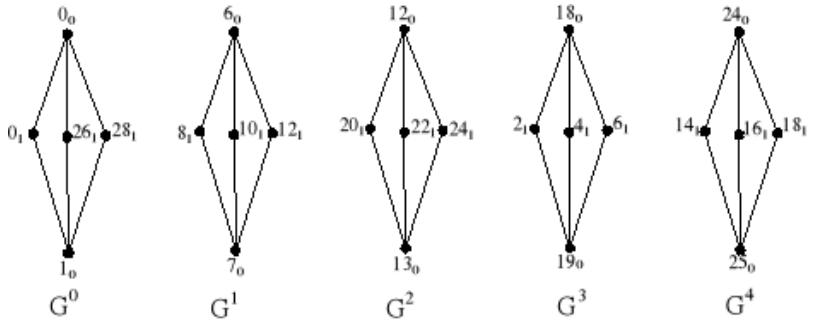


Fig. 5. Symmetric starter of an ODC of $K_{30,30}$ by $5K_{2,3}$

Theorem 5. Let m and $n \geqslant 5$ be positive integers with $m \equiv 1, 5 \pmod 6$. Then, there is an ODC of $K_{mn,mn}$ by $mC_5(1, 0, 0, 0, n-5)$.

Proof. Algorithm 5 proves the Theorem 5. ■

Algorithm 5.	ODC of $K_{mn,mn}$ by $mC_5(1, 0, 0, 0, n-5)$.
1: [Inauguration.]	Choose the values of m and n where m and $n \geqslant 5$ are positive integers with $m \equiv 1, 5 \pmod 6$.
2: [Initial edge.]	Construct the edge $(0, 0)$.
3: [Complementary edges.]	Add $(0, 2-n), (2, 2-n+\beta), ((m+7)n/2, \gamma - n)$ to $(0, 0)$ where $\beta \in \mathbb{Z}_n \setminus \{0, 2, n-2, n-1\}$ and $\gamma \in \{2, 3\}$.
4: [Initial graph $\cong G^0$.]	The union of the edge set in steps 2, 3 gives the edge set of $C_5(1, 0, 0, 0, n-5)$.
5: [Complementary graphs $\cong \bigcup_{\alpha=1}^{m-1} G^\alpha$.]	Add $(\alpha n, 2\alpha n)$ to the edge set of step 4 where $\alpha \in \mathbb{Z}_m \setminus \{0\}$.
6: [Symmetric starter $\cong \bigcup_{\alpha=0}^{m-1} G^\alpha$.]	The union of the edge set in steps 4, 5 gives the edge set of $G \cong mC_5(1, 0, 0, 0, n-5)$.
7: [Symmetric graph.]	If the edge (a, b) belongs to the edge set of step 6, then (b, a) is an edge of the graph $G_s \cong mC_5(1, 0, 0, 0, n-5)$.
8: [Translation.]	Add (γ, γ) to the edge set in steps 6, 7 where $\gamma \in \mathbb{Z}_{mn}$.
9: [Done.]	Output the edge set of step 8 which represents the ODC of $K_{mn,mn}$ by $mC_5(1, 0, 0, 0, n-5)$.

For more illustration to Theorem 5, let $m = 5$ and $n = 7$, then there is an ODC of $K_{35,35}$ by $5C_5(1, 0, 0, 0, 2)$, see Fig. 6.

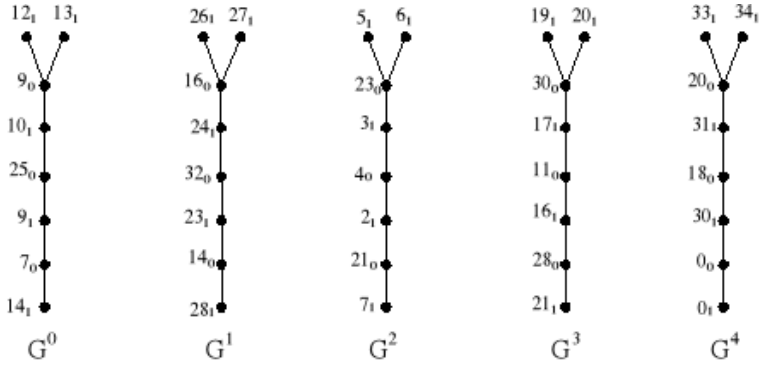


Fig. 6. Symmetric starter of an ODC of $K_{35,35}$ by $5C_5(1,0,0,0,2)$

Conclusion

In this paper, we study the orthogonal double covers of the complete bipartite graphs by algorithms for generating the orthogonal double covers by one edge.

REFERENCES

1. *Demetrovics J., Füredi Z., and Katona G. O. H.* Minimum matrix representations of closure operations. Discrete Appl. Math., 1985, no. 11, pp. 115–128.
2. *Demetrovics J. and Katona G. O. H.* External combinatorial problems in relational database. Fundamentals of Combinatorics of Computation Theory, Berlin, Springer, 1981, pp. 110–119.
3. *Alspach B., Heinrich K., and Liu G.* Orthogonal factorizations of graphs.Contemporary Design Theory, J. H. Dinitz and D. R. Stinson, eds., ch. 2, N.Y., Wiley, 1992.
4. *Heinrich K.* Graph decompositions and designs. CRC Handbook of Combinatorial Designs, C. J. Colbourn and J. H. Dinitz, eds., ch. IV.22, Boca Raton, CRC Press, 1996.
5. *Scapellato R., El-Shanawany R., and Higazy M.* Orthogonal double covers of Cayley graphs. Discrete Appl. Math., 2009, vol. 157, pp. 3111–3118.
6. *El-Shanawany R., Higazy M., Shabana H., and El-Mesady A.* Cartesian product of two symmetric starter vectors of orthogonal double covers. AKCE Intern. J. Graphs and Combinatorics, 2015, no. 12, pp. 59–63.
7. *Gronau H.-D. O. F., Hartman S., Grüttmüller M., et al.* On orthogonal double covers of graphs. Design Codes Cryptography, 2002, vol. 27, pp. 49–91.
8. *El-Shanawany R., Gronau H.-D. O. F., and Grüttmüller M.* Orthogonal double covers of $K_{n,n}$ by small graphs. Discrete Appl. Math., 2004, vol. 138, pp. 47–63.
9. *El-Serafi S., El-Shanawany R., and Shabana H.* Orthogonal double cover of complete bipartite graph by disjoint union of complete bipartite graphs. Ain Shams Engineering J., 2015, no. 6, pp. 657–660.
10. *Sampathkumar R. and Srinivasan S.* Cyclic orthogonal double covers of 4-regular circulant graphs. Discr. Math., 2011, vol. 311, pp. 2417–2422.
11. *El-Shanawany R., Higazy M., and El-Mesady A.* On cartesian products of orthogonal double covers. Intern. J. Math. and Math. Sci., 2013, vol. 2013, Article ID 265136, 4 p.
12. *Froncek D.* Orthogonal double covers of complete graphs by lobsters of diameter 4. Congr. Numer., 2005, vol. 177, pp. 25–32.

МАТЕМАТИЧЕСКИЕ ОСНОВЫ ИНФОРМАТИКИ И ПРОГРАММИРОВАНИЯ

УДК 510.52

О СЛОЖНОСТИ ЭКЗИСТЕНЦИАЛЬНОЙ И УНИВЕРСАЛЬНОЙ ТЕОРИЙ КОНЕЧНЫХ ПОЛЕЙ[1]

А. Н. Рыбалов

Институт математики им. С. Л. Соболева СО РАН, г. Омск, Россия

Конечные поля являются важнейшими математическими объектами, которые используются при решении многих практически важных задач оптимизации, информатики, передачи информации и криптографии. Многие такие задачи можно формулировать как задачи, связанные с решением систем уравнений над полями, что приводит к необходимости развития алгебраической геометрии. Алгебраическая геометрия над этими объектами тесным образом связана со свойствами экзистенциальных и универсальных теорий. С практической точки зрения важнейшими являются вопросы разрешимости и вычислительной сложности этих теорий. В работе изучается вычислительная сложность экзистенциальной и универсальной теорий конечных полей. Доказывается, что экзистенциальная теория класса всех конечных полей является NP-трудной, а универсальная теория этого класса является co-NP-трудной. Это означает, что, при условии неравенства классов сложности P, NP и co-NP, не существует полиномиальных алгоритмов, распознающих эти теории.

Ключевые слова: *конечные поля, универсальная теория, экзистенциальная теория, NP-трудность.*

DOI 10.17223/20710410/45/9

ON COMPLEXITY OF THE EXISTENTIAL AND UNIVERSAL THEORIES OF FINITE FIELDS

A. N. Rybalov

Sobolev Institute of Mathematics, Omsk, Russia

E-mail: alexander.rybalov@gmail.com

Finite fields are the most important mathematical objects that are used for solving many practical problems of optimization, computer science, information transfer and cryptography. Many such problems can be formulated as problems connected with the solving systems of equations over fields. This leads to the need for the development of algebraic geometry. Algebraic geometry over such objects is closely related to properties of existential and universal theories. From a practical point of view, the most important are questions about decidability and computational complexity of these theories. In this paper, we study the computational complexity of existential and

[1]Работа поддержана грантом РНФ № 19-11-00209.

universal theories of finite fields. We prove that the existential theory of finite fields is NP-hard, and the universal theory of finite fields is co-NP-hard. This means that there are no polynomial algorithms that recognize these theories, provided the inequality of classes P, NP and co-NP.

Keywords: *finite fields, universal theory, existential theory, NP-hardness.*

Введение

Большой пласт алгоритмических проблем связан с проблемами разрешения элементарных теорий различных алгебраических систем. Как правило, эти проблемы либо неразрешимы (формальная арифметика, теория поля рациональных чисел, теория графов и т. д.), либо разрешимы, но имеют очень высокую вычислительную сложность (арифметика Пресбургера, теория упорядоченного поля действительных чисел, теория конечных полей и т. д.). Поэтому большой интерес представляет изучение экзистенциальных и универсальных теорий. Экзистенциальные и универсальные теории тесно связаны с решением уравнений и систем уравнений над различными алгебраическими системами. Для классических алгебраических систем, таких, как поля вещественных, комплексных, рациональных чисел, кольца целых и алгебраических чисел изучение их экзистенциальных и универсальных теорий является классической темой исследований в различных областях математики в течение уже многих сотен лет. Классическая алгебраическая геометрия изучает множества решений алгебраических уравнений над полями вещественных и комплексных чисел. В рамках диофантовой геометрии и диофантова анализа изучаются решения алгебраических уравнений над целыми и рациональными числами.

В XX веке, в связи с бурным развитием компьютерной техники и прикладной математики, на первый план вышли исследования различных конечных комбинаторных и алгебраических объектов. Одними из важнейших таких объектов являются конечные поля, без которых немыслима теория передачи информации и современная криптография. Классическим подходом к изучению конечных полей является использование алгебраических методов. Новый подход к изучению этих объектов — логический и теоретико-модельный — родился в рамках так называемой универсальной алгебраической геометрии [1]. Многие практически важные задачи о конечных полях можно формулировать как задачи, связанные с решением систем уравнений над полями, что приводит к необходимости развития алгебраической геометрии. Алгебраическая геометрия над этими объектами тесным образом связана со свойствами экзистенциальных и универсальных теорий. С практической точки зрения важнейшими являются вопросы разрешимости и вычислительной сложности этих теорий.

В данной работе изучается вычислительная сложность экзистенциальной и универсальной теорий конечных полей. Доказывается, что экзистенциальная теория класса всех конечных полей является NP-трудной, а универсальная теория этого класса является co-NP-трудной. Это означает, что, при условии неравенства классов сложности P, NP и co-NP, не существует полиномиальных алгоритмов, распознающих эти теории.

1. Предварительные сведения

Напомним некоторые определения из математической логики и теории моделей [1]. *Сигнатурой* называется множество σ, состоящее из предикатных, функциональных и константных символов. *Алгебраическая система* сигнатуры σ есть набор $\mathfrak{A} = \langle A, \sigma \rangle$, где A — непустое множество, причём каждому предикатному символу сигнатуры σ со-

поставлен некоторый предикат на множестве A, каждому функциональному символу из σ сопоставлена функция на множестве A со значениями в множестве A, а каждому константному символу из σ — элемент из A. *Классом* алгебраических систем $\mathcal{K}$ сигнатуры σ называется множество алгебраических систем фиксированной сигнатуры σ.

Формула логики первого порядка сигнатуры σ, в которой на каждую переменную навешан квантор, называется *предложением*. Тот факт, что предложение Φ сигнатуры σ истинно в алгебраической системе $\mathfrak{A} = \langle A, \sigma \rangle$, обозначается $\mathfrak{A} \models \Phi$. Предложение Φ сигнатуры σ называется *экзистенциальным* (или $\exists$-предложением), если оно имеет вид

$$\Phi = \exists x_1 \ldots \exists x_n \varphi(x_1, \ldots, x_n),$$

где φ — бесквантор ная формула сигнатуры σ. Предложение Φ сигнатуры σ называется *универсальным* (или $\forall$-предложением), если оно имеет вид

$$\Phi = \forall x_1 \ldots \forall x_n \varphi(x_1, \ldots, x_n),$$

где φ — бесквантор ная формула сигнатуры σ. *Элементарной теорией* класса $\mathcal{K}$ называется множество $Th(\mathcal{K})$ всех предложений сигнатуры σ, истинных во всех алгебраических системах класса $\mathcal{K}$. Множество всех $\exists$-предложений теории $Th(\mathcal{K})$ называется *экзистенциальной* теорией $Th_\exists(\mathcal{K})$ класса $\mathcal{K}$. Множество всех $\forall$-предложений теории $Th(\mathcal{K})$ называется *универсальной* теорией $Th_\forall(\mathcal{K})$ класса $\mathcal{K}$. Теория (элементарная, экзистенциальная, универсальная) класса $\mathcal{K}$ разрешима, если существует алгоритм, который по произвольному предложению сигнатуры σ позволяет определить, принадлежит ли оно этой теории.

Приведём некоторые сведения из теории сложности вычислений [2]. *Время работы* $t_M(x)$ машины Тьюринга M над алфавитом A на входе $x \in A^*$ — это число шагов машины от начала работы до остановки. Если M на x не останавливается, полагаем $t_M(x) = \infty$. В дальнейших определениях для строки x под $|x|$ понимается длина строки x. Машина Тьюринга M *полиномиальна*, если существует полином $p(n)$, такой, что для любого $x \in A^*$ имеет место $t_M(x) < p(|x|)$. Множество $S \subseteq I$ принадлежит *классу* P, если существует полиномиальная машина Тьюринга, вычисляющая характеристическую функцию множества S. Множество $S \subseteq I$ принадлежит *классу* NP, если существует такое множество $S' \in$ P и такой полином $p(n)$, что

$$x \in S \Leftrightarrow \exists y \in \{0,1\}^* \ (|y| < p(|x|) \text{ и } (x, y) \in S').$$

Здесь строка y называется подсказкой (или сертификатом, или решением), а множество S' — проверятелем. Множество $S \subseteq I$ принадлежит *классу* co-NP, если $I \setminus S \in$ NP.

Множество $A \subseteq I$ *полиномиально сводится* к множеству $B \subseteq J$, если существует функция $f : I \to J$, вычислимая на полиномиальной машине Тьюринга, такая, что

$$\forall x \in I \ (x \in A \Leftrightarrow f(x) \in B).$$

Множество A называется *NP-трудным*, если любое множество $B \in$ NP полиномиально сводится к A. Если при этом $A \in$ NP, то A называется *NP-полным*. Соответствующая проблема распознавания NP-трудного множества также называется *NP-трудной*. NP-трудные проблемы являются самыми сложными проблемами в классе NP в том смысле, что если хотя бы одна из них лежит в классе P, то P = NP, а также, если P $\neq$ NP, то любая NP-трудная проблема не лежит в классе P. Множество A называется *co-NP-трудным*, если любое множество $B \in$ co-NP полиномиально сводится к A. Легко видеть, что $S \subseteq I$ co-NP-трудно тогда и только тогда, когда $I \setminus S$ NP-трудно.

2. Конечные поля

Рассмотрим класс $\mathcal{F}$ всех конечных полей $\mathrm{GF}(q)$ сигнатуры $\sigma_{\mathcal{F}} = \{=, +, -, /, 0, 1\}$. В [3] доказано, что элементарная теория класса всех конечных полей разрешима. Поэтому разрешимы экзистенциальная и универсальная теория класса конечных полей. Изучим вычислительную сложность этих теорий.

Булева формула $\varphi(x_1, \ldots, x_n)$ называется *выполнимой*, если существуют значения булевых переменных $x_1, \ldots, x_n$, при которых $\varphi(x_1, \ldots, x_n)$ является истинной. *Проблема выполнимости булевых формул* — это проблема распознавания множества выполнимых булевых формул на множестве всех булевых формул. Булева формула записана в виде 3-КНФ, если она имеет вид

$$\varphi(x_1, \ldots, x_n) = \bigwedge_{k=1}^{m} \left(x_{k1}^* \vee x_{k2}^* \vee x_{k3}^*\right),$$

где x_i^* есть литерал, то есть либо сама логическая переменная x_i, либо её отрицание $\neg x_i$. Известно [2], что проблема выполнимости 3-КНФ является NP-полной.

Определим отображение t из множества всех литералов от переменных $x_1, \ldots, x_n$ в множество многочленов от переменных $x_1, \ldots, x_n$ следующим образом:

$$t(x_i^*) = \begin{cases} x_i, & \text{если } x_i^* = x_i, \\ (1 - x_i), & \text{если } x_i^* = \neg x_i. \end{cases}$$

Построим по 3-КНФ $\varphi(x_1, \ldots, x_n)$ следующее экзистенциальное предложение сигнатуры $\sigma_{\mathcal{F}}$:

$$\Phi_\varphi = \exists x_1 \ldots \exists x_n \bigwedge_{i=1}^{n} \big((x_i = 0) \vee (x_i = 1)\big) \wedge \bigwedge_{k=1}^{m} \big(t(x_{k1}^*)t(x_{k2}^*)t(x_{k3}^*) = 0\big).$$

Лемма 1. Пусть φ — 3-КНФ. Тогда φ выполнима тогда и только тогда, когда $\Phi_\varphi \in Th_\exists(\mathcal{F})$.

Доказательство. Пусть 3-КНФ $\varphi(x_1, \ldots, x_n)$ выполнима и $a_1, \ldots, a_n$ — набор значений логических переменных, на которых φ истинна. Пусть $\mathfrak{A}_F$ — любое конечное поле. Тогда $\mathfrak{A}_F \models \Phi_\varphi$, при этом значения переменных $x_1, \ldots, x_n$, которые делают бескванторную часть предложения Φ_φ истинной, определяются следующим образом:

$$x_i = \begin{cases} 0, \text{ если } a_i = \text{истина}, \\ 1, \text{ если } a_i = \text{ложь}. \end{cases}$$

Таким образом, $\Phi_\varphi \in Th_\exists(\mathcal{F})$.

Обратно, пусть $\Phi_\varphi \in Th_\exists(\mathcal{F})$, то есть $\mathfrak{A}_F \models \Phi_\varphi$ для любого конечного поля $\mathfrak{A}_F$, в частности для двоичного поля $\mathrm{GF}(2)$. Но тогда 3-КНФ $\varphi(x_1, \ldots, x_n)$ выполнима, а значения логических переменных $a_1, \ldots, a_n$, на которых φ истинна, определяются по значениям переменных $x_1, \ldots, x_n$, которые делают бескванторную часть предложения Φ_φ истинной в $\mathrm{GF}(2)$:

$$a_i = \begin{cases} \text{истина}, & \text{если } x_i = 0, \\ \text{ложь}, & \text{если } x_i = 1. \end{cases}$$

Лемма доказана. ■

Теорема 1. Экзистенциальная теория конечных полей является NP-трудной.

Доказательство. Доказательство заключается в сведении проблемы выполнимости 3-КНФ к проблеме разрешимости экзистенциальной теории конечных полей. Определим сводимость f между двумя этими проблемами следующим образом. Если φ — 3-КНФ, то положим $f(\varphi) = \Phi_\varphi$. По лемме 1 φ выполнима тогда и только тогда, когда $\Phi_\varphi \in Th_\exists(\mathcal{F})$. Полиномиальность сводимости следует из того факта, что размер формулы Φ_φ линейно ограничен от размера 3-КНФ φ. Таким образом, проблема разрешимости экзистенциальной теории конечных полей является NP-трудной. ■

Теорема 2. Универсальная теория конечных полей является co-NP-трудной.

Доказательство. Докажем, что дополнение $Th_\exists(\mathcal{F})$ в множестве всех экзистенциальных предложений сигнатуры $\sigma_\mathcal{F}$ полиномиально сводится к $Th_\forall(\mathcal{F})$. Отсюда по теореме 1 будет следовать, что последняя теория co-NP-трудна. Эта сводимость f действует на предложение

$$\Phi = \exists x_1 \ldots \exists x_n \varphi(x_1, \ldots, x_n)$$

следующим образом:

$$f(\Phi) = \neg\Phi = \neg\big(\exists x_1 \ldots \exists x_n \varphi(x_1, \ldots, x_n)\big) = \forall x_1 \ldots \forall x_n \neg\varphi(x_1, \ldots, x_n).$$

Очевидно, что для любого $\exists$-предложения Φ имеет место $\Phi \notin Th_\exists(\mathcal{F}) \Leftrightarrow \neg\Phi \in Th_\forall(\mathcal{F})$. Полиномиальность сводимости f следует из того, что размер $\neg\varphi$ линейно ограничен от размера φ. ■

Автор выражает благодарность рецензенту за полезные замечания и предложения по улучшению текста статьи.

ЛИТЕРАТУРА

1. *Даниярова Э. Ю., Мясников А. Г., Ремесленников В. Н.* Алгебраическая геометрия над алгебраическими системами. Новосибирск: СО РАН, 2016. 288 с.
2. *Гэри М., Джонсон Д.* Вычислительные машины и труднорешаемые задачи. М.: Мир, 1982. 419 с.
3. *Ax J.* The elementary theory of finite fields // Ann. Math. 1968. V. 88. No. 2. P. 239–271.

REFERENCES

1. *Daniyarova E. Y., Myasnikov A. G., and Remeslennikov V. N.* Algebraicheskaya geometriya nad algebraicheskimi sistemami [Algebraic Geometry over Algebraic Structures]. Novosibirsk, SB RAS Publ., 2016. 288 p. (in Russian)
2. *Garey M. and Johnson D.* Computers and Intractability. N. Y., Freeman & Co, 1979. 340 p.
3. *Ax J.* The elementary theory of finite fields. Ann. Math., 1968, vol. 88, no. 2, pp. 239–271.

УДК 519.766.2

МИНИМИЗАЦИЯ КОНТЕКСТНО-СВОБОДНЫХ ГРАММАТИК[1]

Ю. Д. Рязанов, С. В. Назина

Белгородский государственный технологический университет им. В. Г. Шухова, г. Белгород, Россия

Решается задача преобразования исходной контекстно-свободной грамматики (КС-грамматики) без лишних символов в эквивалентную ей грамматику меньшей сложности. Предлагается способ минимизации КС-грамматики, основанный на введённом отношении на множестве нетерминалов, обладающим свойством эквивалентности. Это отношение разбивает множество нетерминалов на классы эквивалентности, и новая КС-грамматика строится на нетерминалах, являющихся представителями классов эквивалентности. В результате получается КС-грамматика с меньшим количеством нетерминалов и правил.

Ключевые слова: *формальный язык, формальная грамматика, отношение эквивалентности, минимизация.*

DOI 10.17223/20710410/45/10

MINIMIZATION OF CONTEXT-FREE GRAMMARS

Yu. D. Ryazanov, S. V. Nazina

Belgorod State Technological University named after V. G. Shukhov, Belgorod, Russia

E-mail: razanov.yd@bstu.ru

This paper solves the problem of transforming the initial context-free grammar (CF-grammar) without excess characters into equivalent CF-grammar with less complexity. To solve this problem, the following relation on the set of a CF-grammar non-terminals is introduced: $E = \{(X, Y) : (X = Y) \vee (X \to \alpha \Leftrightarrow Y \to \beta \wedge |\alpha| = |\beta| \wedge \forall i\,(\alpha(i) = = \beta(i) \vee (\alpha(i), \beta(i)) \in E))\}$ where X, Y are non-terminals, α, β are chains of terminal and non-terminals, possibly blank, $\alpha(i)$ is the i-th character in chain α, $\beta(i)$ is the i-th character in chain β. It is proved that the relation E has the equivalence property and splits the set of non-terminals into equivalence classes. An algorithm is proposed for splitting a set of non-terminals into equivalence classes based on the method of sequential decomposition of the set of non-terminals into subsets so that non-equivalent non-terminals fall into different subsets. New CF-grammar is built on a set of non-terminals N, which elements are representatives of equivalence classes. From the set of rules of the initial CF-grammar, the rules with the left parts belonging to the set N are chosen. If there is a non-terminal in the left side of any selected rule that does not belong to the set N, then it is replaced by its equivalent non-terminal from the set N. After such transformations in the CF-grammar, sets of identical rules may appear. From each set of identical rules, we leave only one rule. The result is a CF-grammar containing less rules and non-terminals than the initial CF-grammar. The paper provides an example of the implementation of the described transformations.

Keywords: *formal language, formal grammar, equivalence relation, minimization.*

[1]Работа поддержана грантом РФФИ № 16-07-00487.

Введение

Для построения трансляторов, интерпретаторов и других программ обработки формальных языков широко используются методы, основанные на применении контекстно-свободных грамматик [1, 2]. Формальный язык можно задать КС-грамматиками, которые могут существенно различаться по сложности. Оценкой сложности КС-грамматики может служить суммарное количество символов (терминалов и нетерминалов) в правых частях правил грамматики. Размер программ обработки языков зависит от сложности используемой КС-грамматики, поэтому для сокращения размера программ целесообразно применять КС-грамматики меньшей сложности. Одним из способов уменьшения сложности КС-грамматики является устранение лишних символов [1, 3]. В данной работе решается задача преобразования исходной КС-грамматики без лишних символов в эквивалентную ей КС-грамматику меньшей сложности за счёт минимизации мощности множества нетерминалов. Для решения этой задачи предлагается способ, основанный на введённом отношении на множестве нетерминалов, обладающем свойством эквивалентности. Это отношение разбивает множество нетерминалов на классы эквивалентности, и новая КС-грамматика строится на нетерминалах, являющихся представителями этих классов. В результате получается КС-грамматика меньшей сложности.

Подход к минимизации дискретных объектов на основе некоторого отношения эквивалентности используется в минимизации состояний конечных автоматов [4 – 6], состояний и магазинных символов распознавателей с магазинной памятью [7, 8], минимизации челночных сплайновых процессоров [9], минимизации КСР-грамматик [10], минимизации автоматной модели визуального описания синтаксического разбора [11] и других объектов. В данном случае объектом минимизации является КС-грамматика и отношение эквивалентности вводится на множестве нетерминалов грамматики.

1. Отношение на множестве нетерминалов контекстно-свободной грамматики

Контекстно-свободной грамматикой [1] называется четвёрка $G = \langle N, \Sigma, P, S\rangle$, где

N — алфавит нетерминальных символов (нетерминалов);

Σ — непересекающийся с N алфавит терминальных символов (терминалов);

P — конечное множество правил вывода вида $A \to \alpha$, где $A \in N$, α — цепочка символов из $N \cup \Sigma$;

S — начальный нетерминал, $S \in N$.

Приведём пример грамматики G_1:

$S \to AaC \,|D\,|\, SbF$
$A \to DaS \,|\, E \,|\, AbB$
$B \to EaC \,|\, D \,|\, BbA$
$C \to AaS \,|\, E \,|\, SbC$
$D \to bSEa \,|\, bBSa \,|\, bCDa \,|\, c$
$E \to bFDa \,|\, bASa \,|\, c$
$F \to BaF \,|\, D \,|\, FbF$

В этой грамматике $\{S, A, B, C, D, E, F\}$ — множество нетерминалов; S — начальный нетерминал; $\{a, b, c\}$ — множество терминалов и двадцать два правила. Сложность грамматики равна 57.

Определим отношение на множестве нетерминалов:

$$E = \{(X, Y) : (X = Y) \vee (X \to \alpha \Leftrightarrow Y \to \beta \wedge |\alpha| = |\beta| \wedge \forall i\, (\alpha(i) = \beta(i) \vee (\alpha(i), \beta(i)) \in E))\},$$

где X, Y — нетерминалы; α, β — цепочки терминалов и нетерминалов, возможно пустые; $\alpha(i)$ — i-й символ в цепочке α; $\beta(i)$ — i-й символ в цепочке β.

Правила $X \to \alpha$ и $Y \to \beta$, в которых $|\alpha| = |\beta|$ и $\forall i\,(\alpha(i) = \beta(i) \vee (\alpha(i), \beta(i)) \in E)$, назовём эквивалентными.

Цепочки одинаковой длины α^i и α^j назовём эквивалентными, если $\forall(k)\,(\alpha^i(k) = \alpha^j(k) \vee (\alpha^i(k), \alpha^j(k)) \in E)$, где $\alpha^i(k)$ — k-й символ в цепочке α^i; $\alpha^j(k)$ — k-й символ в цепочке α^j. Множество всех терминальных цепочек, выводимых из нетерминала X, обозначим $L(X)$.

Теорема 1. Если $(X_i, X_j) \in E$, то $L(X_i) = L(X_j)$.

Доказательство. Начнём один вывод из нетерминала X_i, а второй — из X_j. В исходном состоянии цепочка α^i состоит только из нетерминала X_i, а цепочка α^j — из X_j. Эти цепочки эквивалентны, так как $(X_i, X_j) \in E$. Пусть на очередном шаге вывода к k-му символу цепочки α^i, который является нетерминалом, применяется правило $\alpha^i(k) \to \beta$, в результате чего получается цепочка γ^i. Тогда на этом же шаге применим к k-му символу цепочки α^j правило $\alpha^j(k) \to \delta$, которое эквивалентно правилу $\alpha^i(k) \to \beta$, в результате чего получим цепочку γ^j, эквивалентную цепочке γ^i. Продолжая таким образом выводы, будем получать эквивалентные цепочки вплоть до терминальных. Терминальные эквивалентные цепочки равны. Таким образом, любую терминальную цепочку, выводимую из нетерминала X_i, можно вывести из нетерминала X_j, и наоборот, следовательно, $L(X_i) = L(X_j)$. ■

Из $L(X_i) = L(X_j)$ для $(X_i, X_j) \in E$ следует очевидное эквивалентное преобразование: все вхождения нетерминала X_j в правые части правил заменить на X_i, а правила с левой частью X_j исключить. В результате такого преобразования количество нетерминалов и правил грамматики уменьшится.

Теорема 2. Отношение E обладает свойством эквивалентности.

Доказательство. Отношение обладает свойством эквивалентности, если оно рефлексивно, симметрично и транзитивно. Рефлексивность и симметричность отношения следует из его определения. Отношение транзитивно, если $((X, Y) \in E \wedge \wedge (Y, Z) \in E) \Rightarrow (X, Z) \in E$ для всех X, Y, Z.

Пусть $(X, Y) \in E$ и $(Y, Z) \in E$. Если в грамматике есть правило $X \to \alpha$ и $(X, Y) \in \in E$, то правило $Y \to \beta$, такое, что $|\alpha| = |\beta|$ и $\alpha(i) = \beta(i)$, если i-й символ — терминал в цепочках α и β, и $(\alpha(i), \beta(i)) \in E$, если i-й символ — нетерминал, тоже есть в этой грамматике. Аналогично если правило $Y \to \beta$ есть в грамматике и $(Y, Z) \in E$, то в грамматике есть правило $Z \to \gamma$, такое, что $|\beta| = |\gamma|$ и для любого символа $\beta(i) = \gamma(i)$ или $(\beta(i), \gamma(i)) \in E$. Так как для любого нетерминала $\alpha(i)$ выполняется $(\alpha(i), \beta(i) \in E)$, а для любого терминала $\alpha(i) = \beta(i)$, то правило $X \to \alpha$ можно заменить правилом $X \to \beta$. Поскольку $|\beta| = |\gamma|$, а $\beta(i) = \gamma(i)$ или $(\beta(i), \gamma(i)) \in E$, условие принадлежности пары (X, Z) отношению E является истинным. ■

2. Разбиение множества нетерминалов на классы эквивалентности

Отношение E определяет разбиение множества нетерминалов на классы эквивалентности. Применим метод последовательного разбиения множества нетерминалов на подмножества так, чтобы неэквивалентные нетерминалы попали в разные подмножества.

Сначала разобьём множество нетерминалов на подмножества $Q_1, Q_2, \ldots, Q_n$ так, чтобы нетерминалы, для которых условие $X \to \alpha \Leftrightarrow Y \to \beta \wedge |\alpha| = |\beta| \wedge \forall i\,(\alpha(i) \in \in \Sigma \Rightarrow \alpha(i) = \beta(i))$ ложно, оказались в разных подмножествах. Для этого построим

таблицу, столбцы которой соответствуют нетерминалам, а строки — цепочкам, составленным из правых частей правил грамматики заменой нетерминалов символом «$*$», не принадлежащим множеству терминалов или нетерминалов грамматики. Цепочку, полученную из правой части правила $X \to \alpha$ заменой нетерминалов символом «$*$», обозначим α^*. Таблицу заполняем следующим образом: если в грамматике существует правило $X \to \alpha$, то в клетку, соответствующую нетерминалу X и цепочке α^*, записываем «1».

Если столбцы, соответствующие нетерминалам X и Y, различаются, то эти нетерминалы нужно включить в разные подмножества. Для грамматики G_1 эта таблица представлена в табл. 1.

Таблица 1

α^*	S	A	B	C	D	E	F
$*a*$	1	1	1	1			1
$*$	1	1	1	1			1
$*b*$	1	1	1	1			1
$b**a$					1	1	
c					1	1	

Анализируя табл. 1, получаем подмножества $Q_1 = \{S, A, B, C, F\}$ и $Q_2 = \{D, E\}$, которые образуют разбиение $R = \{Q_1, Q_2\}$ множества нетерминалов.

Нетерминалы, принадлежащие разным подмножествам, явно неэквивалентны. Так как мы не учитывали условие $(\alpha(i), \beta(i)) \in E$ для нетерминальных символов цепочек, нетерминалы, принадлежащие одному подмножеству, тоже могут быть неэквивалентными.

Для дальнейшего разбиения построим таблицу, структура которой аналогична таблице, построенной на предыдущем шаге, за исключением строк, которые не содержат символ «$*$». В клетку таблицы будем записывать множество упорядоченных последовательностей чисел $(i_1, i_2, \ldots, i_n)$, где i_k — номер подмножества в последнем разбиении, $k = 1, \ldots, n$, n — количество нетерминалов в правой части каждого правила, соответствующего строке таблицы. Если в грамматике есть правило $X \to \alpha$ и α содержит n нетерминалов $X_1, X_2, \ldots X_n$, то в клетку, соответствующую нетерминалу X и цепочке α^*, добавляем последовательность $(i_1, i_2, \ldots, i_n)$, где i_k — номер подмножества в последнем разбиении, которому принадлежит нетерминал X_k.

Если нетерминалы X и Y принадлежат одному подмножеству и соответствующие им столбцы различаются, то пара нетерминалов X и Y не принадлежит отношению E и их нужно включить в разные подмножества нового разбиения R'.

Анализируя таблицу, формируем новое разбиение R'. Подмножество Q'_k разбиения R' формируется из элементов некоторого подмножества Q_k разбиения R: Q'_k является максимальным по мощности подмножеством множества Q_k, таким, что для каждой пары нетерминалов X и Y, принадлежащих подмножеству Q'_k, столбцы таблицы, соответствующие X и Y, равны.

Если $R' \neq R$, то полагаем $R := R'$, заново строим таблицу по описанным правилам и формируем новое разбиение. Если $R' = R$, то R' представляет собой множество классов эквивалентности.

Построим новую таблицу для G_1 (табл. 2).

В этой таблице сначала расположены столбцы, соответствующие подмножеству Q_1, а затем — столбцы, соответствующие подмножеству Q_2. Рассмотрим правило $S \to AaC$

Таблица 2

α^*	Q_1					Q_2	
	S	A	B	C	F	D	E
$*a*$	(1, 1)	(2, 1)	(2, 1)	(1, 1)	(1, 1)		
$*$	(2)	(2)	(2)	(2)	(2)		
$*b*$	(1, 1)	(1, 1)	(1, 1)	(1, 1)	(1, 1)		
$b**a$						(1, 2) (1, 1)	(1,2) (1,1)

грамматики G_1. Правая часть соответствует цепочке $*a*$. Оба нетерминала правой части принадлежат подмножеству Q_1, поэтому в клетку, находящуюся в строке $*a*$ и столбце S, записываем последовательность (1,1). Клетка в строке $b**a$ и столбце S остаётся пустой, так как в грамматике нет правила с нетерминалом S в левой части, правая часть которого соответствует цепочке $b**a$. В клетку в строке $b**a$ и столбце D записываем последовательность (1,2), так как в грамматике есть правило $D \to bSEa$ и $S \in Q_1$, $E \in Q_2$. В эту же клетку записываем пару (1,1), так как в грамматике есть правило $D \to bBSa$ и $B \in Q_1$, $S \in Q_1$. Аналогично заполняем все клетки таблицы.

Анализируя табл. 2, видим, что Q_1 разбивается на два подмножества, а Q_2 не разбивается. В результате получаем новое разбиение $R' = \{Q'_1, Q'_2, Q'_3\}$, где $Q'_1 = \{S, C, F\}$, $Q'_2 = \{A, B\}$, $Q'_3 = \{D, E\}$. Поскольку $R' \neq R$, полагаем $R := R'$ и строим новую таблицу (табл. 3).

Таблица 3

α^*	Q_1			Q_2		Q_3	
	S	C	F	A	B	D	E
$*a*$	(2, 1)	(2, 1)	(2, 1)	(3, 1)	(3, 1)		
$*$	(3)	(3)	(3)	(3)	(3)		
$*b*$	(1, 1)	(1, 1)	(1, 1)	(2, 2)	(2, 2)		
$b**a$						(1, 3) (2, 1)	(1,3) (2,1)

Анализируя табл. 3, приходим к выводу, что подмножества Q_1, Q_2 и Q_3 не разбиваются и поэтому представляют собой классы эквивалентности.

3. Построение упрощенной грамматики

Новую грамматику G_2 формируем на множестве нетерминалов, являющихся представителями классов эквивалентности. Таким множеством может быть, например, $\{S, A, D\}$. Из множества правил грамматики G_1 берём правила, левые части которых принадлежат множеству $\{S, A, D\}$:

$S \to AaC \,|D\,|\, SbF$
$A \to DaS \,|\, E \,|\, AbB$
$D \to bSEa \,|\, bBSa \,|\, bCDa \,|\, c$

Последовательно просматриваем все правила и если в правой части есть нетерминал, который не принадлежит множеству $\{S, A, D\}$, то заменяем его на эквивалентный ему нетерминал из множества $\{S, A, D\}$:

$S \to AaS \,|D\,|\, SbS$
$A \to DaS \,|\, E \,|\, AbA$
$D \to bSDa \,|\, bASa \,|\, bSDa \,|\, c$

Как видно, в грамматике появились одинаковые правила. Оставляя из каждого множества одинаковых правил по одному, получаем грамматику G_2:

$S \to AaS \,|\, D \,|\, SbS$
$A \to DaS \,|\, E \,|\, AbA$
$D \to bSDa \,|\, bASa \,|\, c$

Полученная грамматика имеет сложность 23, что значительно меньше сложности исходной грамматики.

Заключение

В процессе выполнения описанных преобразований количество правил грамматики никогда не увеличивается. Уменьшение сложности грамматики достигается за счёт исключения из исходной грамматики правил, определяющих нетерминалы, не принадлежащие новому множеству нетерминалов, и замены возникающих в процессе преобразований множеств одинаковых правил одним. Следовательно, уменьшить сложность грамматики можно за счёт минимизации множества нетерминалов.

ЛИТЕРАТУРА

1. *Aho A. V. and Ullman J. D.* The Theory of Parsing, Translation and Compiling. NJ, USA: Prentice-Hall Inc., 1972. V. 1. 560 p.
2. *Aho A. V., Lam M. S., Sethi R., and Ullman J. D.* Compilers: Principles, Techniques, and Tools. Addison-Wesley, 2007, 1009 p.
3. *Конюхова О. В., Кравцова Э. А.* Программная реализация алгоритмов упрощения контекстно-свободных грамматик на языках программирования Haskell и Prolog // Информационные системы и технологии. 2017. № 4. С. 77–86.
4. *Hopcroft J. E.* An $n \log n$ algorithm for minimizing states in a finite automaton // Theory of Machines and Computations. N.Y.: Academic Press, 1971. P. 189–196.
5. *Hopcroft J. E., Motwani R., and Ullman J. D.* Introduction to Automata Theory, Languages, and Computation. Pearson, 2013, 496 p.
6. *Мартыненко Б. К.* Ещё один метод минимизации конечных автоматов // Компьютерные инструменты в образовании. 2017. № 1. С. 5–14.
7. *Polyakov V. M. and Ryazanov Yu. D.* Reducing the number of states in pushdown recognizers by means of equivalence relation // Intern. J. Pharmacy & Technology. 2016. V. 8. No. 4. P. 22578–22587.
8. *Рязанов Ю. Д.* Сокращение количества магазинных символов в распознавателях с магазинной памятью и одним состоянием // Вестник БГТУ им. В. Г. Шухова. 2017. № 6. С. 152–157.
9. *Мартыненко Б. К.* Синтаксически управляемая обработка данных. СПб.: Изд-во С.-Петербург. ун-та, 2004. 316 с.
10. *Федорченко Л. Н.* Минимизация трансляционной КСР-грамматики и состояний синтаксического анализатора КСР-языка // Вестник Бурятского государственного университета. Математика, информатика. 2013. № 2. С. 39–49.
11. *Стасенко А. П.* Автоматная модель визуального описания синтаксического разбора // Вычислительные технологии. 2008. Т. 13. № 5. С. 70–87.

REFERENCES

1. *Aho A. V. and Ullman J. D.* The Theory of Parsing, Translation and Compiling. New Jersey, Prentice-Hall, 1972, vol. 1, 560 p.

2. *Aho A. V., Lam M. S., Sethi R.,and Ullman J. D.* Compilers: Principles, Techniques, and Tools. Addison-Wesley, 2007, 1009 p.
3. *Konyuxova O. V. and Kravcova E. A.* Programmnaya realizaciya algoritmov uproshcheniya kontekstno-svobodnyh grammatik na yazykah programmirovaniya Haskell i Prolog [The implementation of the simplifying context-free grammars algorithms in Haskell and in Prolog]. Information Systems and Technologies, 2017, no. 4, pp. 77–86. (in Russian)
4. *Hopcroft J. E.* An $n \log n$ algorithm for minimizing states in a finite automaton Theory of Machines and Computations. N.Y., Academic Press, 1971, pp. 189–196.
5. *Hopcroft J. E., Motwani R., and Ullman J. D.* Introduction to Automata Theory, Languages, and Computation. Pearson, 2013. 496 p.
6. *Martynenko B. K.* Eshche odin metod minimizatsii konechnykh avtomatov [One more method for minimization of finite automata]. Computer Tools in Education, 2017, no. 1, pp. 5–14. (in Russian)
7. *Polyakov V. M. and Ryazanov Yu. D.* Reducing the number of states in pushdown recognizers by means of equivalence relation. Intern. J. Pharmacy & Technology, 2016, vol. 8, no. 4, pp. 22578–22587.
8. *Ryazanov Yu. D.* Sokrashchenie kolichestva magazinnykh simvolov v raspoznavatelyakh s magazinnoy pamyatyu i odnim sostoyaniem [Reducing the Number of Pushdown Symbols in One-state Pushdown Recognizers]. Bulletin of BSTU named after V. G. Shukhov, 2017, no. 6, pp. 152–157. (in Russian)
9. *Martynenko B. K.* Sintaksicheski upravlyaemaya obrabotka dannykh [Syntax-directed data processing]. Saint-Petersburg, St Petersburg University, 2004. 316 p. (in Russian)
10. *Fedorchenko L. N* Minimizatsiya translyatsionnoy KSR-grammatiki i sostoyaniy sintaksicheskogo analizatora KSR-yazyka [Minimization of Translational CFR-grammar and Conditions of Generated Parser CFR-language]. Bulletin of the Buryat State University. Mathematics, Informatics, 2013, no. 2, pp. 39–49. (in Russian)
11. *Stasenko A. P.* Avtomatnaya model vizualnogo opisaniya sintaksicheskogo razbora [Automaton model of visual description of syntax parsing]. Computational Technologies, 2008, vol. 13, no. 5. pp. 70–87. (in Russian)

УДК 621.396:621.372

ИНФОРМАЦИОННАЯ ЁМКОСТЬ СЕТИ ХОПФИЛДА С КВАНТОВАННЫМИ ВЕСАМИ

М. С. Тарков

Институт физики полупроводников им. А. В. Ржанова СО РАН, г. Новосибирск, Россия

Использование бинарных и многоуровневых мемристоров при аппаратной реализации нейронных сетей вызывает необходимость квантования их весовых коэффициентов. Исследуется влияние числа уровней квантования весов сети Хопфилда на её информационную ёмкость и устойчивость к искажениям входных данных. Показано, что при числе градаций весов порядка 20 ёмкость сети Хопфилда — Хебба с дискретными весами приближается к ёмкости её варианта с непрерывными весами. Для проекционной сети Хопфилда подобного результата удаётся достичь лишь при числе градаций порядка 100. Эксперименты показали, что: 1) бинарные мемристоры следует использовать в сетях Хопфилда — Хебба, редуцированных путём обнуления всех весов, модули которых строго меньше максимального для данной строки матрицы весов; 2) в проекционных сетях Хопфилда с дискретными весами следует использовать многоуровневые мемристоры с числом градаций (уровней) значительно больше двух, причём конкретное число уровней зависит от размерности хранимых эталонных векторов, их конкретного набора и допустимого уровня шума во входных данных.

Ключевые слова: *сети Хопфилда — Хебба, проекционные сети Хопфилда, информационная ёмкость, квантование весов, бинарные и многоуровневые мемристоры.*

DOI 10.17223/20710410/45/11

INFORMATIONAL CAPACITY OF THE HOPFIELD NETWORK WITH QUANTIZED WEIGHTS

M. S. Tarkov

Rzhanov Institute of Semiconductor Physics SB RAS, Novosibirsk, Russia

E-mail: tarkov@isp.nsc.ru

The use of binary and multilevel memristors in the hardware neural networks implementation necessitates their weight coefficients quantization. In this paper, we investigate the Hopfield network weights quantization influence on its information capacity and resistance to input data distortions. It is shown that, for a weight level number of the order of tens, the capacitance of Hopfield — Hebb network with the quantized weights approximates the capacitance of its version with continuous weights. For a Hopfield projection network, similar result can be achieved only for a weight levels number of the order of hundreds. Experiments show that: 1) binary memristors should be used in Hopfield — Hebb networks, reduced by zeroing all weights in a given row, in which absolute values are strictly less than the maximum weight in the row; 2) in the Hopfield projection networks with quantized weights, multilevel memristors with a weight levels number significantly more than two should be used, with a specific

levels number depending on the stored reference vectors dimension, their particular set and the permissible input data noise level.

Keywords: *Hopfield — Hebb networks, projection Hopfield networks, information capacity, weight quantization, binary and multilevel memristors.*

Введение

В процессе функционирования сети Хопфилда [1, 2] можно выделить два режима — обучение и классификацию. Веса сети вычисляются в режиме её обучения на основе известных векторов. В режиме классификации при фиксированных значениях весов и заданных значениях входов нейрона инициируется переходный процесс, завершающийся в одном из локальных энергетических минимумов. При вводе обучающих векторов x^k, $k = 1, \ldots, p$, вычисляются веса w_{ij} в соответствии с обобщённым правилом Хебба

$$w_{ij} = \frac{1}{N} \sum_{k=1}^{p} x_i^k x_j^k. \tag{1}$$

Важным параметром ассоциативной памяти является её информационная ёмкость, которая определяется как максимальное число хранимых в памяти образцов. Показано [1], что при использовании правила Хебба ёмкость памяти оценивается как $p_{\max} = N/(2 \ln N)$. Такую сеть будем называть далее сетью Хопфилда — Хебба.

Проекционный метод обучения сети Хопфилда [3, 4] порождает матрицу весов, итерационно зависящую от последовательности обучающих векторов x^k, $k = 1, \ldots, p$:

$$\begin{aligned} y^k &= (W^{k-1} - E)x^k, \\ W^k &= W^{k-1} + \frac{y^k \cdot y^{kT}}{y^{kT} \cdot y^k} \end{aligned} \tag{2}$$

при начальном условии $W^0 = 0$ (E — единичная матрица). В результате матрица весов сети получает значение $W = W^p$. Применение метода проекций увеличивает ёмкость сети Хопфилда до $N - 1$. Далее такую сеть будем называть проекционной сетью Хопфилда.

1. Бинарные и многоуровневые мемристоры

Аппаратная реализация нейронной сети требует много памяти для хранения матрицы весов слоя нейронов и является дорогостоящей. Решение этой проблемы упрощается при использовании в качестве ячейки памяти устройства, называемого мемристором (резистором с памятью). Мемристор был предсказан теоретически в 1971 г. Леоном Чуа [5]. Первую физическую реализацию мемристора продемонстрировала в 2008 г. лаборатория фирмы «Hewlett Packard» в виде тонкоплёночной структуры TiO_2 [6]. Мемристор ведёт себя подобно синапсу: он «запоминает» полный электрический заряд, прошедший через него. Память, основанная на мемристорах, может достигать степени интеграции 100 Гбит/см^2, в несколько раз более высокой, чем на основе технологии флэш-памяти. Эти уникальные свойства делают мемристор многообещающим устройством для создания массово-параллельных нейроморфных систем.

Бинарными называют мемристоры, реализующие два значения проводимости, многоуровневыми — мемристоры, реализующие множество дискретных уровней проводимости (количество уровней может достигать десятков и сотен). Бинарные и многоуровневые мемристоры [7 – 9] основаны на механизме переключения филамента и распространены более широко, чем аналоговые мемристоры, в которых проводимость можно

изменять непрерывно, но материалы для них встречаются значительно реже, к тому же они требуют более сложного процесса обработки. Многоуровневые мемристоры устойчивы к статистическим флуктуациям по сравнению с аналоговыми мемристорами. Использование бинарных и многоуровневых мемристоров для задания весовых коэффициентов сетей Хопфилда делает актуальной задачу исследования влияния числа уровней квантования весов на информационную ёмкость сети и её устойчивость к уровню шумов входных данных. При этом задачу квантования весов можно рассматривать как обобщение задачи их редукции [10].

2. Оценка ёмкости сети Хопфилда

Для оценивания ёмкости сети Хопфилда [11] порождается множество M случайных векторов с компонентами из множества $\{-1, +1\}$. На множестве M строится сеть Хопфилда согласно правилам (1) или (2). Каждый из векторов множества M подаётся на вход построенной сети Хопфилда. Если после одного срабатывания сети её выход совпадает с её входом для всех векторов множества M, то в M добавляется новый случайный вектор и процесс повторяется (изначально $|M| = 1$). В противном случае ёмкость сети полагаем равной $|M| - 1$.

Результаты апробации данного алгоритма приведены в табл. 1 и 2. Здесь C — оценка ёмкости, усреднённая по 100 испытаниям; $N/(2\ln N)$ — асимптотическая оценка ёмкости сети Хопфилда — Хебба; $N - 1$ — оценка максимальной ёмкости проекционной сети Хопфилда.

Таблица 1

Метод Хебба

N	32	64	128	256	512
C	5,14	7,51	11,22	18,07	29,88
$N/(2\ln N)$	4,62	7,69	13,19	23,08	41,04

Таблица 2

Проекционный метод

N	32	64	128	256	512
C	30,83	62,39	125,97	253,52	508,52
$N-1$	31	63	127	255	511

Из табл. 2 следует, что результаты алгоритма [11] очень близки к оценкам [4] для проекционной сети Хопфилда (разница составляет менее 1 %).

3. Ёмкость редуцированной сети Хопфилда

Редукция весов сети Хопфилда осуществляется следующим образом [10]:

1. Веса сети нормируются по формуле

$$w_{ij} \leftarrow \frac{w_{ij}}{\max\limits_{j=1,\ldots,N} w_{ij}}.$$

2. Нормированные веса редуцируются по правилу

$$w_{ij}^{'} = \begin{cases} +1, & w_{ij} = 1, \\ -1, & w_{ij} = -1, \\ 0, & -1 < w_{ij} < 1. \end{cases} \tag{3}$$

С целью сохранения симметрии матрицы весов правило редукции (3) для сетей Хопфилда приобретает следующий вид: вес w_{ij} обнуляется одновременно с весом w_{ji} при одновременном выполнении условий $w_{ij} < w_{i,\max}$ и $w_{ji} < w_{j,\max}$, где $w_{i,\max}$ и $w_{j,\max}$ — максимальные веса в строках i и j соответственно.

В результате редукции получаем сеть с весами из множества $\{-1, 0, +1\}$, которые могут быть реализованы на основе использования бинарных мемристоров. Значения ± 1 задаются максимальной проводимостью, а значение 0 — минимальной проводимостью мемристора. В табл. 3 приведены значения ёмкости редуцированной сети Хопфилда — Хебба, усреднённые по 100 экспериментам. Их значения близки к $\log_2 N$.

Таблица 3

Ёмкость редуцированной сети Хопфилда — Хебба

N	32	64	128	256	512
C	4,62	5,43	6,52	7,85	8,9
$\log_2 N$	5	6	7	8	9

4. Ёмкость сети Хопфилда с квантованными весами

Алгоритм квантования весов сети Хопфилда состоит в следующем:

1. Найти максимум $W_{\max}$ модуля весов сети Хопфилда.
2. Найти величину кванта (скачка) веса $\Delta = W_{\max}/(L-1)$, где $L \geqslant 2$ — число уровней квантования.
3. Преобразовать все веса w_{ij}, $i, j = 1, \ldots, N$, по правилу: если $(k-1)\Delta < w_{ij} < k\Delta$, то w_{ij} получает значение $(k-1)\Delta \cdot \operatorname{sign}(w_{ij})$, $k = 1, \ldots, L$, где

$$\operatorname{sign}(a) = \begin{cases} 1, & a > 0, \\ -1, & a < 0. \end{cases}$$

На рис. 1 представлены графики зависимости ёмкости сети Хопфилда от числа уровней (градаций) $L \in \{2, 4, 8, 16, 32, 64, 128, 256\}$ квантования весовых коэффициентов, $N \in \{32, 64, 128\}$. Эти графики показывают, что при малых значениях L ёмкость сети с квантованными весами мала, но для сети Хопфилда — Хебба, начиная с $L = 16$, становится близкой к ёмкости сети с непрерывными весами (табл. 1). Для проекционной сети Хопфилда зависимость ёмкости от L выражена гораздо сильнее. Проекционная сеть с квантованными весами по ёмкости приближается к проекционной сети с непрерывными весами только при приближении L к $128 \div 256$.

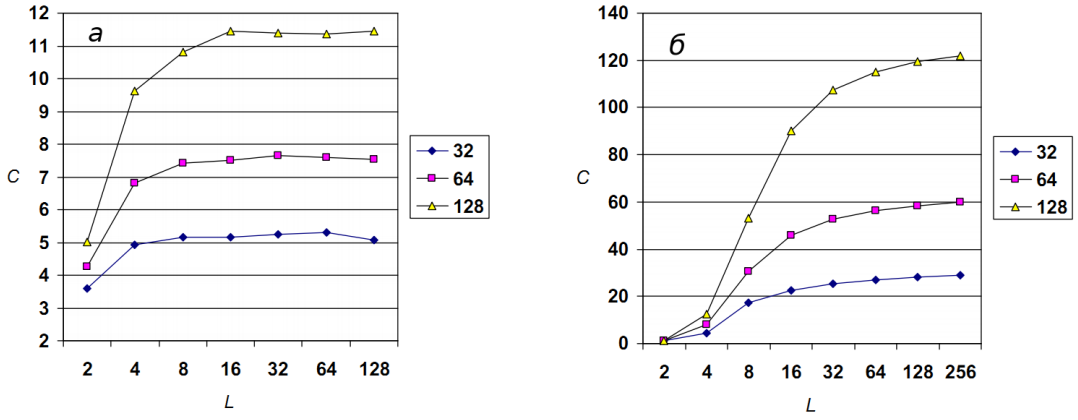


Рис. 1. Зависимость ёмкости сети Хопфилда — Хебба ($а$) и проекционной сети Хопфилда ($б$) от числа градаций весов

5. Выбор количества уровней квантования весов сети Хопфилда для фильтрации шумов

Выбор минимального количества уровней квантования весов для фильтрации шумов зависит от варианта сети Хопфилда (сеть Хопфилда — Хебба или проекционная сеть Хопфилда) и конкретного множества эталонных векторов. Для редуцированной сети Хопфилда — Хебба [10] количество уровней квантования равно двум независимо от размерности эталонных векторов. Веса таких сетей могут быть реализованы на бинарных мемристорах. При достаточно большой размерности векторов эти сети устойчивы к шумам при существенном сокращении числа связей (более 90 %) [10].

На рис. 2 и в табл. 4 приведены полученные экспериментально минимальные количества L уровней квантования весов проекционной сети Хопфилда в зависимости от доли искажённых пикселей ξ для набора из десяти эталонных векторов размерности $n \times n$, $n \in \{8, 16, 32, 64\}$, соответствующих изображениям цифр на рис. 3.

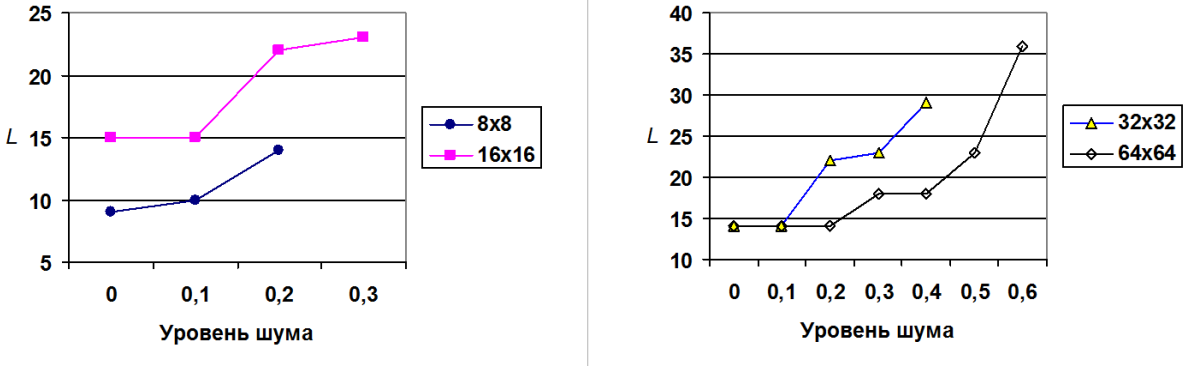


Рис. 2. Зависимость числа L уровней квантования от уровня шумов ξ

Таблица 4

Количество L уровней квантования весов проекционной сети Хопфилда для предельно допустимых долей ξ искажённых пикселей

$n \times n$	8×8	16×16	32×32	64×64
ξ	0,14	0,33	0,48	0,64
L	14	26	29	36

Рис. 3. Эталонные изображения

Результаты экспериментов показывают, что для обеспечения допустимых уровней шумов, соответствующих проекционной сети Хопфилда [10] с непрерывными весами, требуются десятки уровней весов.

Заключение

В настоящее время актуальной является задача построения нейронных сетей на основе использования многоуровневых (в том числе двухуровневых) мемристоров для аппаратной реализации синапсов. Поскольку число различимых уровней (градаций) проводимости мемристора ограничено, в работе исследуется влияние числа уровней квантования весов сети Хопфилда на её информационную ёмкость. Показано, что когда число градаций весов достигает 16, ёмкость квантованной сети Хопфилда — Хебба приближается к ёмкости её варианта с непрерывными весами. Для проекционной сети Хопфилда подобного результата удаётся достичь при числе градаций порядка сотен.

Эксперименты показали, что:

1. Бинарные мемристоры следует использовать в сетях Хопфилда — Хебба, редуцированных путём обнуления в строке матрицы всех весов, модули которых строго меньше максимального для данной строки.

2. В проекционных сетях Хопфилда с дискретными весами следует использовать многоуровневые мемристоры с числом уровней значительно больше двух, причём конкретное число уровней зависит от размерности хранимых эталонных векторов, их конкретного набора и допустимого уровня шума.

ЛИТЕРАТУРА

1. *Haykin S.* Neural Networks: A Comprehensive Foundation. Prentice Hall, Inc., 1999.
2. *Hopfield J.* Neural networks and physical systems with emergent collective computational abilities // Proc. NAS USA. 1982. V. 79. P. 2554–2558.
3. *Personnaz L., Guyon I., and Dreyfus G.* Collective computational properties of neural networks: New learning mechanisms // Phys. Rev. Ser. A. 1986. V. 34. No. 5. P. 4217–4228.
4. *Michel A. N. and Liu D.* Qualitative Analysis and Synthesis of Recurrent Neural Networks. N.Y.: Marcel Dekker Inc., 2002.
5. *Chua L.* Memristor — the missing circuit element // IEEE Trans. Circuit Theory. 1971. V. 18. P. 507–519.
6. *Strukov D. B., Snider G. S., Stewart D. R., and Williams R. S.* The missing memristor found // Nature. 2008. V. 453. P. 80–83.
7. *He W., Sun H., Zhou Y., et al.* Customized binary and multi-level HfO_{2-x}-based memristors tuned by oxidation conditions // Scientific Reports. 2017. V. 7. Article No. 10070.
8. *Yu S., Gao B., Fang Z., et al.* A low energy oxide-based electronic synaptic device for neuromorphic visual systems with tolerance to device variation // Adv. Mater. 2013. V. 25. P. 1774–1779.
9. *Тарков М. С.* Построение сети Хемминга на основе кроссбара с бинарными мемристорами // Прикладная дискретная математика. 2018. № 40. С. 105–113.
10. *Тарков М. С.* Редукция связей автоассоциативной памяти Хопфилда // Прикладная дискретная математика. 2017. № 37. С. 107–113.
11. *Folli V., Leonetti M., and Ruocco G.* On the maximum storage capacity of the Hopfield model // Frontiers in Comput. Neuroscience. 2017. V. 10. Article 144. `www.frontiersin.org`.

REFERENCES

1. *Haykin S.* Neural Networks: A Comprehensive Foundation. Prentice Hall, Inc., 1999.
2. *Hopfield J.* Neural networks and physical systems with emergent collective computational abilities. Proc. NAS, 1982, vol. 79, pp. 2554–2558.
3. *Personnaz L., Guyon I., and Dreyfus G.* Collective computational properties of neural networks: New learning mechanisms. Phys. Rev., ser. A, 1986, vol. 34, no. 5, pp. 4217–4228.

4. *Michel A. N. and Liu D.* Qualitative Analysis and Synthesis of Recurrent Neural Networks. N.Y., Marcel Dekker Inc., 2002.
5. *Chua L.* Memristor — the missing circuit element. IEEE Trans. Circuit Theory, 1971, vol. 18, pp. 507–519.
6. *Strukov D. B., Snider G. S., Stewart D. R., and Williams R. S.* The missing memristor found. Nature, 2008, vol. 453, pp. 80–83.
7. *He W., Sun H., Zhou Y., et al.* Customized binary and multi-level HfO_{2-x}-based memristors tuned by oxidation conditions. Scientific Reports, 2017, vol. 7, article number: 10070.
8. *Yu S., Gao B., Fang Z., et al.* A low energy oxide-based electronic synaptic device for neuromorphic visual systems with tolerance to device variation. Adv. Mater., 2013, vol. 25, pp. 1774–1779.
9. *Tarkov M. S.* Postroenie seti Khemminga na osnove krossbara s binarnymi memristorami [Construction of a Hamming network based on a crossbar with binary memristors]. Prikladnaya Diskretnaya Matematika, 2018, no. 40, pp. 105–113. (in Russian)
10. *Tarkov M. S.* Reduktsiya svyazey avtoassotsiativnoy pamyati Khopfilda [Reduction of synapses in the Hopfield autoassociative memory]. Prikladnaya Diskretnaya Matematika, 2017, no. 37, pp. 107–113. (in Russian)
11. *Folli V., Leonetti M., and Ruocco G.* On the maximum storage capacity of the Hopfield model. Frontiers in Comput. Neuroscience, 2017, vol. 10, article 144. `www.frontiersin.org`.

ВЫЧИСЛИТЕЛЬНЫЕ МЕТОДЫ В ДИСКРЕТНОЙ МАТЕМАТИКЕ

УДК 519.7

ЗАДАЧА КОММИВОЯЖЁРА: ПРИБЛИЖЁННЫЙ АЛГОРИТМ ПО МЕТОДУ ВЕТВЕЙ И ГРАНИЦ С ГАРАНТИРОВАННОЙ ТОЧНОСТЬЮ[1]

Ю. Л. Костюк

Национальный исследовательский Томский государственный университет, г. Томск, Россия

Для решения задачи коммивояжёра с матрицей расстояний порядка n предлагается приближённый алгоритм на основе метода ветвей и границ. Для отсечения используется увеличенная оценка снизу текущего частичного решения, гарантирующая заранее заданную величину ε погрешности всего решения. Проведён вычислительный эксперимент для матриц расстояний четырёх видов распределений, среди них для равномерного случайного (несимметричного) распределения, а также для матриц евклидовых расстояний между случайными точками (симметричного распределения). В последнем случае дополнительно применён локальный поиск. Получены оценки степени p для функции полиномиальной трудоёмкости $O(n^p)$ для разных видов распределений и различных величин погрешности ε. Для равномерного случайного распределения полученная оценка степени p оказалась близка к 2,8 в диапазоне n до 1000 и средней погрешности около 1 %.

Ключевые слова: *задача коммивояжёра, метод ветвей и границ, приближённый алгоритм, локальный поиск, вычислительный эксперимент.*

DOI 10.17223/20710410/45/12

THE TRAVELING SALESMAN PROBLEM: APPROXIMATE ALGORITHM BY BRANCH-AND-BOUND METHOD WITH GUARANTEED PRECISION

Yu. L. Kostyuk

Tomsk State University, Tomsk, Russia

E-mail: kostyuk_y_l@sibmail.com

To solve the traveling salesman problem with distance matrix of order n, we propose an approximate algorithm based on the branch-and-border method. For clipping, an increased least estimate of the current partial solution is used. This guarantees a predetermined value ε of the whole solution error. A computational experiment for distance matrices of four kinds of distributions was carried out. A uniform random (asymmetric) distribution as well as matrices of Euclidean distances between random

[1]Результаты получены в рамках выполнения госзадания Минобрнауки России, проект № 2.4218.2017/4.6.

points (a symmetric distribution) were used. In the latter case, a local search was additionally applied. Estimates for the power p in the polynomial computational complexity $O(n^p)$ of the algorithm for various kinds of distributions and various values of error ε are obtained. For a uniform random distribution and $n \leqslant 1000$, the obtained estimate of the power p turned out to be close to 2.8 and of an average value of error to be 1 %.

Keywords: *traveling salesman problem, branch-and-border method, approximate algorithm, local search, computational experiment.*

Введение

В известном алгоритме Литтла [1], реализующим метод ветвей и границ для задачи коммивояжёра (ЗК), на каждом шаге обработки матрицы расстояний вычисляется нижняя оценка стоимости оптимального решения. В [2] предложена модификация алгоритма Литтла путём вычисления более точных нижних оценок. В результате увеличивается эффективность отсечения бесперспективных маршрутов и тем самым значительно сокращается время вычислений. В [3] предложен способ дополнительного уточнения нижних оценок. Как показал вычислительный эксперимент, алгоритм [3] имеет преимущество перед алгоритмом Литтла как на симметричных, так и на несимметричных матрицах. Тем не менее его трудоёмкость экспоненциальная, поэтому единственным способом решения ЗК большой размерности остаётся поиск приближённого решения. При этом в общем случае приближённое решение с гарантированной точностью можно находить лишь алгоритмами с экспоненциальной трудоёмкостью [4, 5]. Таким образом, поиск более эффективных приближённых алгоритмов с заданной степенью приближения к точному решению представляется актуальным.

В настоящей работе рассмотрен приближённый алгоритм с гарантированной точностью решения, имеющий полиномиальную трудоёмкость в ограниченном диапазоне n, который построен на основе модифицированного алгоритма с уточнением нижней границы [3].

1. Приближённый алгоритм

Алгоритм реализует поиск вглубь на дереве решений, первоначально корень дерева содержит исходную матрицу расстояний M порядка n, а длина текущего маршрута полагается бесконечно большой. В результате обработки матрицы вычисляется нижняя оценка стоимости оптимального маршрута, а в каждой строке и каждом столбце матрицы появляется не менее чем по одному нулю. Затем выбирается один из нулевых элементов матрицы M_{ij}, который определяет переход в маршруте из вершины i в вершину j. Создаётся левый узел в дереве решений, содержащий копию матрицы расстояний без строки i и столбца j, т. е. порядка $n-1$. Альтернативный вариант маршрута, в котором переход из вершины i в вершину j заведомо отсутствует, сохраняется в исходной матрице M, для этого в ней элемент M_{ij} заменяется на бесконечность.

На следующем шаге алгоритм переходит к обработке левого узла дерева решений, если по матрице расстояний в узле можно построить новый маршрут коммивояжёра меньшей стоимости, чем текущий маршрут. В противном случае выполняется отсечение этих бесперспективных вариантов, для чего алгоритм переходит к обработке правого узла, т. е. осуществляет обратный ход. После обработки очередного узла с матрицей расстояний порядка n создаётся новый левый узел с матрицей порядка $n-1$ и т. д., пока не будет построен новый текущий маршрут коммивояжёра.

Таким образом, на всех шагах работы алгоритма дерево решений представляется списком узлов от корня влево до висячего узла.

Отсечение бесперспективных вариантов выполняется следующим образом. Пусть $S_{\min}$ — стоимость наилучшего на текущий момент построенного маршрута (в начальный момент $S_{\min}$ равно бесконечности), S_m — нижняя оценка для нового частично построенного маршрута в текущий момент, состоящего из отрезков. В методе ветвей и границ проверка $S_m \geqslant S_{\min}$ отсекает все бесперспективные варианты, которые можно построить на основе частичного маршрута. В предлагаемом приближённом алгоритме проверка $S_m(1+\varepsilon) \geqslant S_{\min}$ гарантирует, что наименьший найденный маршрут не будет превышать оптимальный маршрут более чем в $(1+\varepsilon)$ раз. Чтобы среднюю погрешность сделать на порядок меньше по сравнению с заданной величиной ε, предлагается более гибкая проверка в виде

$$S_m \geqslant S_{\min}/(1+\varepsilon(m)),$$

где функция $\varepsilon(m)$ зависит от количества m отрезков частично построенного маршрута, т. е. размера текущей матрицы, и от параметра n_0 следующим образом:

$$\varepsilon(m) = \begin{cases} 0, & \text{если } m \leqslant n_0, \\ \varepsilon(m-n_0)/(n-n_0), & \text{если } m > n_0. \end{cases}$$

Параметр n_0 имеет следующий смысл: если порядок m текущей матрицы не превышает n_0, то проверка на отсечение выполняется так же, как в точном алгоритме: $S_m \geqslant S_{\min}$, а если m больше n_0, то проверка выполняется с учётом погрешности, увеличивающейся линейно от нуля до ε при возрастании m от n_0 до n.

Общая структура алгоритма аналогична структуре алгоритма с улучшенной нижней оценкой [3].

Как показал вычислительный эксперимент, в некоторых случаях полученный приближённый маршрут можно значительно улучшить, применив к нему локальный поиск [6] с окном фиксированного размера k. Пусть нумерация вершин в маршруте $P_1, P_2, \ldots, P_n$. Для всех i от 0 до $n-1$ последовательно рассматривается группа вершин в окне $P_{i \bmod n+1}, \ldots, P_{(i+k-1) \bmod n+1}$, с учётом замыкания маршрута (после вершины P_n следует вершина P_1). Для этой группы вершин выбирается перестановка, минимизирующая маршрут. Так как k является константой, общее время работы алгоритма при этом возрастает на $\mathrm{O}(n)$.

Исходный текст модифицированного алгоритма с улучшенной нижней границей и приближённого алгоритма на языке Паскаль (Delphi) доступен в сети интернет [7].

2. Вычислительный эксперимент

В вычислительном эксперименте на модельных данных проверялась сравнительная трудоёмкость следующих четырёх алгоритмов:

1) алгоритма Литтла (А-1);
2) модифицированного алгоритма с улучшенной нижней границей (А-2);
3) приближённого алгоритма с заданной гарантированной точностью решения ε (А-3).
4) приближённого алгоритма с заданной величиной ε и с дополнительным локальным поиском (А-4).

Алгоритмы написаны на языке Паскаль в системе Delphi, вычисления производились на компьютере с процессором Athlon (частота 2,9 Ггц). Исходными данными для

всех алгоритмов являлись одни и те же сгенерированные экземпляры матриц расстояний (каждая матрица задаёт свою ЗК).

В табл. 1–8 представлено среднее время вычисления маршрута (в секундах) для следующих четырёх видов распределений элементов матриц:

1) случайные матрицы, элементы в них равномерно распределённые независимые случайные целые числа из диапазона от 0 до 1000, матрицы несимметричные;
2) случайные матрицы, такие же, как в первом распределении, в которых затем вычислены кратчайшие расстояния алгоритмом Флойда [8] и этими кратчайшими расстояниями заменены исходные элементы матрицы; в результате для расстояний выполняется неравенство треугольника;
3) матрицы евклидовых расстояний, в которых для n равномерно распределённых независимых случайных точек на плоскости с координатами из диапазона от 0 до 1000 вычислены расстояния между точками, матрицы симметричные;
4) матрицы евклидовых расстояний, такие же, как в первом распределении, в которых для каждой пары расстояний M_{ij} и M_{ji} случайно выбрано одно из них и его длина увеличена в 2 раза, после чего вычислены все кратчайшие расстояния алгоритмом Флойда и этими кратчайшими расстояниями заменены исходные элементы матрицы; полученные матрицы несимметричные, но с выполнением неравенства треугольника.

Распределения выбраны из следующих соображений. Равномерное случайное (первый вид распределения) и евклидовы расстояния (третий вид) — стандартные для ЗК. Для практических применений важен случай, когда расстояния несимметричные, но выполняется неравенство треугольника, как в распределениях 2 и 4. При этом четвёртый вид распределения задаёт почти евклидовы расстояния, когда отношение расстояний M_{ij} и M_{ji} не превышает 2.

Для каждого размера матриц n генерировалось от 400 до 4000 вариантов, чем меньше n, тем больше вариантов. В таблицах отсутствуют те результаты вычислений, для получения которых потребовалось бы слишком много времени.

Таблица 1

Результаты эксперимента для матриц первого вида, $n \leqslant 100$

n	40	50	60	70	80	90	100
A-1	0,077	0,69	5,2	–	–	–	–
A-2	0,010	0,036	0,12	0,40	1,24	3,65	8,1
A-3, $\varepsilon = 0{,}05$	0,0055	0,0113	0,021	0,043	0,078	0,12	0,18
$\bar{\varepsilon}$	0,05 %	0,08 %	0,11 %	0,14 %	0,16 %	0,21 %	0,21 %
A-3, $\varepsilon = 0{,}1$	0,0039	0,0058	0,0085	0,013	0,016	0,021	0,025
$\bar{\varepsilon}$	0,18 %	0,31 %	0,46 %	0,55 %	0,61 %	0,75 %	0,75 %
A-3, $\varepsilon = 0{,}15$	0,0032	0,0046	0,0062	0,0091	0,011	0,015	0,018
$\bar{\varepsilon}$	0,3 %	0,5 %	0,7 %	0,8 %	0,9 %	1,1 %	1,1 %

Таблица 2

Результаты эксперимента для матриц первого вида, $n \leqslant 1000$

n	150	200	300	400	500	700	1000
A-3, $\varepsilon = 0{,}1$	0,074	0,15	0,47	10	–	–	–
A-3, $\varepsilon = 0{,}15$	0,053	0,11	0,36	0,84	1,6	4,6	18
$\Delta\bar{\varepsilon}$	0,32 %	0,27 %	0,26 %	0,23 %	–	–	–

В табл. 1 и 2 представлены результаты вычислений для матриц первого вида. Здесь параметр n_0 задавался равным 25, это означает, что если некоторый частичный маршрут состоит не более чем из 25 отрезков, то для всех вариантов, строящихся на основе данного частичного маршрута, ищется в точности оптимальный маршрут. Оказалось, что здесь локальный поиск не даёт никакого улучшения маршрута, поэтому в таблицах он не отражён. Для алгоритма А-3 в том случае, когда получено точное решение, приведено среднее значение погрешности $\bar{\varepsilon}$ в процентах. В случае, когда точное решение отсутствует, приведено среднее превышение погрешности $\Delta\bar{\varepsilon}$ по сравнению с наиболее близким к точному решением, полученным для соответствующей размерности n.

Таблица 3

Результаты эксперимента для матриц второго вида, $n \leqslant 50$

n	25	30	35	40	45	50
А-1	0,017	0,13	1,2	4,3	21	–
А-2	0,003	0,015	0,076	0,6	1,7	14
А-3, $\varepsilon = 0{,}05$	0,001	0,005	0,015	0,09	0,2	0,7
$\bar{\varepsilon}$	0,17 %	0,19 %	0,28 %	0,31 %	0,38 %	0,4 %
А-3, $\varepsilon = 0{,}1$	0,0006	0,0018	0,003	0,0035	0,013	0,035
$\bar{\varepsilon}$	0,4 %	0,43 %	0,6 %	0,7 %	0,8 %	0,85 %
А-3, $\varepsilon = 0{,}15$	0,0005	0,0011	0,002	0,0028	0,004	0,006
$\bar{\varepsilon}$	0,56 %	0,6 %	0,8 %	0,85 %	1,1 %	1,1 %

Таблица 4

Результаты эксперимента для матриц второго вида, $n \leqslant 110$

n	60	70	80	90	100	110
А-3, $\varepsilon = 0{,}05$	1,7	–	–	–	–	–
А-3, $\varepsilon = 0{,}1$	0,1	3	–	–	–	–
$\Delta\bar{\varepsilon}$	0,5 %	–	–	–	–	–
А-3, $\varepsilon = 0{,}15$	0,02	0,17	0,5	0,7	1,2	3
$\Delta\bar{\varepsilon}$	0,7 %	0,3 %	–	–	–	–

В табл. 3 и 4 представлены результаты вычислений для матриц второго вида. Здесь параметр n_0 задавался равным 10. Оказалось, что здесь, как и для матриц первого распределения, локальный поиск не даёт никакого улучшения маршрута, поэтому в таблицах он не отражён. Для алгоритма А-3 приведена также погрешность $\bar{\varepsilon}$ или превышение погрешности $\Delta\bar{\varepsilon}$ в процентах.

В табл. 5 и 6 представлены результаты вычислений для матриц третьего вида. Здесь параметр n_0 задавался равным 10. Для алгоритмов А-3 и А-4 приведена также погрешность $\bar{\varepsilon}$ или превышение погрешности $\Delta\bar{\varepsilon}$ в процентах. Размер окна k для локального поиска в алгоритме А-4 задавался равным 11.

В табл. 7 и 8 представлены результаты вычислений для матриц четвёртого вида. Условия проведения эксперимента полностью идентичны предыдущему случаю.

Из таблиц видно, что среднее время работы как точных, так и приближённых алгоритмов зависит, в первую очередь, от вида распределения элементов матриц. При одинаковой размерности для первого вида матриц время работы точных алгоритмов наименьшее, а для третьего — наибольшее, причём разница достигает сотен раз. Ещё больше разница во времени для приближённых алгоритмов.

Таблица 5

Результаты эксперимента для матриц третьего вида, $n \leqslant 40$

n	25	30	35	40
A-1	0,58	13	–	–
A-2	0,11	0,57	4,3	17
A-3, $\varepsilon = 0{,}05$	0,036	0,14	0,46	2,1
$\bar{\varepsilon}$	0,5 %	0,52 %	0,58 %	0,62 %
A-4, $\varepsilon = 0{,}05$	0,086	0,19	0,52	2,2
$\bar{\varepsilon}$	0,06 %	0,08 %	0,09 %	0,18 %
A-3, $\varepsilon = 0{,}1$	0,014	0,04	0,11	0,3
$\bar{\varepsilon}$	1,8 %	1,9 %	2,3 %	2,4 %
A-4, $\varepsilon = 0{,}1$	0,07	0,09	0,17	0,37
$\bar{\varepsilon}$	0,16 %	0,3 %	0,7 %	0,8 %
A-3, $\varepsilon = 0{,}15$	0,007	0,017	0,035	0,08
$\bar{\varepsilon}$	3,2 %	3,6 %	4,3 %	4,5 %
A-4, $\varepsilon = 0{,}15$	0,06	0,07	0,095	0,15
$\bar{\varepsilon}$	0,3 %	0,6 %	1,2 %	1,4 %

Таблица 6

Результаты эксперимента для матриц третьего вида, $n \leqslant 60$

n	45	50	55	60
A-3, $\varepsilon = 0{,}05$	5,7	22	–	–
A-4, $\varepsilon = 0{,}05$	5,8	22	–	–
A-3, $\varepsilon = 0{,}1$	0,8	2,3	11	–
$\Delta\bar{\varepsilon}$	1,9 %	2,1 %	–	–
A-4, $\varepsilon = 0{,}1$	0,9	2,4	11	–
$\Delta\bar{\varepsilon}$	0,7 %	0,8 %	–	–
A-3, $\varepsilon = 0{,}15$	0,2	0,43	1,2	4
$\Delta\bar{\varepsilon}$	4 %	4 %	2,8 %	–
A-4, $\varepsilon = 0{,}15$	0,3	0,51	1,3	4,1
$\Delta\bar{\varepsilon}$	1,6 %	1,8 %	1,2 %	–

Таблица 7

Результаты эксперимента для матриц четвёртого вида, $n \leqslant 45$

n	30	35	40	45
A-1	4	26	–	–
A-2	0,32	1,5	3,6	15
A-3, $\varepsilon = 0{,}05$	0,05	0,2	0,3	1
$\bar{\varepsilon}$	0,5 %	0,5 %	0,55 %	0,55 %
A-4, $\varepsilon = 0{,}05$	0,11	0,26	0,36	1,1
$\bar{\varepsilon}$	0,15 %	0,2 %	0,2 %	0,2 %
A-3, $\varepsilon = 0{,}1$	0,018	0,06	0,08	0,2
$\bar{\varepsilon}$	1,7 %	2 %	2,2 %	2,2 %
A-4, $\varepsilon = 0{,}1$	0,08	0,13	0,15	0,3
$\bar{\varepsilon}$	0,6 %	0,7 %	0,9 %	0,9 %
A-3, $\varepsilon = 0{,}15$	0,007	0,02	0,03	0,04
$\bar{\varepsilon}$	3,2 %	3,4 %	3,9 %	4,1 %
A-4, $\varepsilon = 0{,}15$	0,07	0,09	0,11	0,13
$\bar{\varepsilon}$	1,1 %	1,3 %	1,5 %	1,8 %

Таблица 8

Результаты эксперимента для матриц четвёртого вида, $n \leqslant 80$

n	50	60	70	80
А-3, $\varepsilon = 0{,}05$	3,6	–	–	–
А-4, $\varepsilon = 0{,}05$	3,7	–	–	–
А-3, $\varepsilon = 0{,}1$	0,6	3,1	10	–
$\Delta\bar{\varepsilon}$	1,5 %	–	–	–
А-4, $\varepsilon = 0{,}1$	0,7	3,2	10,1	–
$\Delta\bar{\varepsilon}$	0,9 %	–	–	–
А-3, $\varepsilon = 0{,}15$	0,12	0,3	1,6	4
$\Delta\bar{\varepsilon}$	3,5 %	2 %	2,6 %	–
А-4, $\varepsilon = 0{,}15$	0,22	0,4	1,7	4,1
$\Delta\bar{\varepsilon}$	1,7 %	1,2 %	1,3 %	–

Для третьего и четвёртого вида матриц дополнительный локальный поиск уменьшает среднюю погрешность в 2–3 раза при незначительном увеличении времени. В целом, средняя погрешность $\bar{\varepsilon}$ оказывается меньше гарантированной погрешности ε на порядок.

На основе проведённого эксперимента установлено, что средняя трудоёмкость $T(n)$ приближённых алгоритмов в ограниченном диапазоне n имеет полиномиальную зависимость вида $T(n) = a \cdot n^p$. Для получения оценки константы p будем рассматривать отношение трудоёмкости при некотором n_i к трудоёмкости при n_{i-1} для возрастающей последовательности $n_1, n_2, \ldots$:

$$\frac{T(n_i)}{T(n_{i-1})} = \left(\frac{n_i}{n_{i-1}}\right)^p. \tag{1}$$

Будем минимизировать сумму квадратов отклонений логарифмов отношений экспериментальных данных времени работы t_i и трудоёмкости (1):

$$\sum_i \left(\ln\frac{t_i}{t_{i-1}} - p\ln\frac{n_i}{n_{i-1}}\right)^2 \to \min_p. \tag{2}$$

После дифференцирования левой части выражения (2) по параметру p и приравнивания производной к нулю получим формулу для вычисления оценки $\hat{p}$:

$$\hat{p} = \sum_i \ln\frac{t_i}{t_{i-1}} \ln\frac{n_i}{n_{i-1}} \Big/ \sum_i \left(\ln\frac{n_i}{n_{i-1}}\right)^2. \tag{3}$$

В табл. 9 приведены полученные по формуле (3) оценки $\hat{p}$ при гарантированной погрешности $\varepsilon = 0{,}15$ для различных видов матриц. Для третьего и четвёртого видов оценка $\hat{p}$ приведена для приближённого алгоритма с последующей локальной оптимизацией.

Таблица 9

Оценка параметра p для различных видов матриц

Вид матриц	1	2	3	4
n	От 40 до 1000	От 25 до 110	От 25 до 60	От 30 до 80
$\hat{p}$	2,8	5,9	4,0	4,2

Заключение

Как показал вычислительный эксперимент, для матриц с равномерно случайными расстояниями предложенный приближённый алгоритм способен находить решение ЗК со средней погрешностью, близкой к 1 %, в диапазоне до $n = 1000$, с трудоёмкостью, растущей как $n^{2,8}$. Для матриц расстояний с другими распределениями, где время работы приближённого алгоритма приемлемо (не более 20 с на обычном компьютере), диапазон n существенно меньше. Однако и в этом случае время работы (по сравнению с временем работы точного алгоритма) уменьшается в десятки и сотни раз при заданной гарантированной погрешности 15 % и средней погрешности, близкой к 2 %.

ЛИТЕРАТУРА

1. *Little J. D. C., Murty K. G., Sweeney D. W., and Karel C.* An algorithm for the Traveling Salesman Problem // Operat. Res. 1963. No. 11. P. 972–989.
2. *Костюк Ю. Л.* Эффективная реализация алгоритма решения задачи коммивояжёра методом ветвей и границ // Прикладная дискретная математика. 2013. № 2(20). С. 78–90.
3. *Костюк Ю. Л.* Задача коммивояжёра: улучшенная нижняя граница в методе ветвей и границ // Прикладная дискретная математика. 2013. № 4(22). С. 73–81.
4. *Меламед И. И., Сергеев С. И., Сигал И. Х.* Задача коммивояжёра. Точные алгоритмы // Автоматика и телемеханика. 1989. № 10. С. 3–29.
5. *Меламед И. И., Сергеев С. И., Сигал И. Х.* Задача коммивояжёра. Приближенные алгоритмы // Автоматика и телемеханика. 1989. № 11. С. 3–26.
6. *Пападимитриу Х., Стайглиц К.* Комбинаторная оптимизация: Алгоритмы и сложность: пер. с англ. М.: Мир, 1985. 510 с.
7. *Костюк Ю. Л.* Модифицированный алгоритм решения задачи коммивояжёра методом ветвей и границ с улучшенной нижней границей, вычисляющий точное или приближённое решение при заданной гарантированной точности. Программа на языке Паскаль в системе Delphi. 2017 г., апрель. `http://www.inf.tsu.ru/Decanat/Staff.nsf/people/KostjukJuL`
8. *Ахо А., Хопкрофт Дж., Ульман Дж.* Построение и анализ вычислительных алгоритмов: пер. с англ. М.: Мир, 1979. 536 с.

REFERENCES

1. *Little J. D. C., Murty K. G., Sweeney D. W., and Karel C.* An algorithm for the Traveling Salesman Problem. Operat. Res., 1963, no. 11, pp. 972–989.
2. *Kostyuk Yu. L.* Effektivnaya realizaciya algoritma resheniya zadachi kommivoyazhera metodom vetvei i granic [Effective implementation of algorithm for solving the travelling salesman problem by the branch-and-border method]. Prikladnaya Diskretnaya Matematika, 2013, no. 2(20), pp. 78–90. (in Russian)
3. *Kostyuk Yu. L.* Zadacha kommivoyazhera: uluchshennaya nizhnyaya granica v metode vetvei i granic [The travelling salesman problem: improved lover bounds in the branch-and-border method]. Prikladnaya Diskretnaya Matematika, 2013, no. 4(22), pp. 73–81. (in Russian)
4. *Melamed I. I., Sergeev S. I., and Sigal I. K.* The traveling salesman problem. Exact methods. Autom. Remote Contro, 1989, vol. 50, no. 10, pp. 1303–1324.
5. *Melamed I. I., Sergeev S. I., and Sigal I. K.* The traveling salesman problem. Approximate algorithModifitsirovannyy algoritm resheniya zadachi kommivoyazhera metodom vetvey i granits s uluchshennoy nizhney granitsey, vychislyayushchiy tochnoe ili priblizhennoe reshenie pri zadannoy garantirovannoy tochnosti. Programma na yazyke Paskal' v sisteme Delphi. 2017 g., aprel'.ms. Autom. Remote Control, 1989, vol. 50, no. 11, pp. 1459–1479.
6. *Papadimitriou C. H. and Steiglitz K.* Combinatorial Optimization: Algorithms and Complexity. N.J., Prentice Hall, 1982.

7. *Kostyuk Yu. L.* A modified algorithm for solving the traveling salesman problem using the branch and bound method with an improved lower bound, which calculates an exact or approximate solution for a given guaranteed accuracy. Pascal program in Delphi system. April, 2017. http://www.inf.tsu.ru/Decanat/Staff.nsf/people/KostjukJuL
8. *Aho A. V., Hopcroft J. E., and Ullman J. D.* The Design and Analysis of Computer Algorithms. Reading, Massachusetts, Addison-Wesley Publ., 1976.

МАТЕМАТИЧЕСКИЕ ОСНОВЫ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ

УДК 519.81

НЕПРОТИВОРЕЧИВОЕ АГРЕГИРОВАНИЕ ОТНОШЕНИЙ КВАЗИПОРЯДКА

В. Н. Нефёдов, С. О. Смерчинская, Н. П. Яшина

Московский авиационный институт, г. Москва, Россия

Работа относится к разделу «Групповой выбор» математической теории принятия решений. Предлагается методика непротиворечивого агрегирования отношений квазипорядка, основанная на построении нагруженного мажоритарного графа. Веса на дугах графа характеризуют степень превосходства одной альтернативы над другой и используются для разрушения противоречивых контуров, при этом сохраняются контуры, содержащие равноценные альтернативы. Удаление из контуров дуг минимального веса позволяет построить непротиворечивое отношение, учитывающее предпочтения большинства экспертов. Разработан алгоритм упорядочения альтернатив на основе отношения квазипорядка. Все процедуры могут быть использованы и при многокритериальном выборе в случае задания вербальной информации о попарном сравнении альтернатив по критериям качества.

Ключевые слова: *групповой выбор, мажоритарный граф, агрегированное отношение, квазипорядок, минимальное расстояние, противоречивый контур, уровни предпочтения.*

DOI 10.17223/20710410/45/13

NON-CONTRADICTORY AGGREGATION OF QUASI-ORDER RELATIONS

V. N. Nefedov, S. O. Smerchinskaya, N. P. Yashina

Moscow Aviation Institute, Moscow, Russia

E-mail: svetlana_os@mail.ru

The paper belongs to the “Collective choice” section of the mathematical theory of decision-making. A method for non-contradictory aggregation of expert preferences given by quasi-order relations is proposed. The aggregated relation is built according to the rule of “the majority of experts”, which satisfies the condition of the minimum distance from expert preferences. Let the expert preferences profile be given by the quasi-order relations $\rho_1, \rho_2, \ldots, \rho_m$ on the set of alternatives $A = \{a_1, a_2, \ldots, a_n\}$. Relations will be given by the preference matrices $P^1, P^2, \ldots, P^m$ and the vertex adjacency matrices $R^1, R^2, \ldots, R^m$ of the corresponding digraphs. The preference matrix, in contrast to the adjacency matrix, contains elements 1/2 for equivalent alternatives. The algorithm for constructing a non-contradictory aggregate relation contains the following steps.

1. Construction of a weighted majority digraph $G = \langle A, \rho_\Sigma \rangle$. The adjacency matrix $R_\Sigma = \|r_{ij}^\Sigma\|$ of the majority digraph is constructed on the basis of the matrix of total preferences $P_\Sigma = \sum\limits_{k=1}^{m} P^k$ according to the majority rule

$$r_{ij} = \begin{cases} 1, & \text{if } p_{ij} \geqslant p_{ji}, \\ 0, & \text{if } p_{ij} < p_{ji}, \end{cases} \text{ where } p_{ij} \text{ are the elements of the matrix } P_\Sigma.$$

The weights on the digraph arcs $l_{ij} = \sum\limits_{k=1}^{m} (p_{ij}^k - p_{ji}^k)$ characterize the degree of superiority of alternative a_i over a_j and are used to destroy contradictory cycles ($P^k = \|p_{ij}^k\|$; $i, j = 1, \ldots, n$).
2. The destruction of contradictory cycles. Arcs are removed from the cycles that have a minimum weight (with minimal advantage in expert preferences) and belong to the asymmetric part of the relation. In this case, arcs connecting equivalent alternatives are saved. We get a digraph $G' = \langle A, \rho \rangle$, R is the adjacency matrix.
3. Construction of aggregated quasi-order $\widehat{\rho}$. The adjacency matrix of an aggregate quasi-order is found by the formula $\widehat{R} = E \vee \operatorname{Tr} R$, where $\operatorname{Tr} R$ is the adjacency matrix of the transitive closure of the relation ρ without contradictory cycles.
The propositions about the uniqueness and non-contradictory of the constructed aggregated relation $\widehat{\rho}$ are proved. The computational complexity of the algorithm is $\mathrm{O}(n^3)$. Based on the constructed aggregated quasi-order relation, the ranking of alternatives is carried out. For this purpose, an algorithm for constructing digraph preference levels has been developed. The algorithm is based on the Demukron procedure of partitioning a digraph without contours into levels $N_0, N_1, \ldots, N_t$, where

$$N_0 = \{a_i : a_i \in A,\ \Gamma a_i = \varnothing\};\ N_k = \{a_i : a_i \in A \setminus \bigcup_{j=0}^{k-1} N_j,\ \Gamma a_i \subseteq \bigcup_{j=0}^{k-1} N_j\},\ k = 1, \ldots, t.$$

The propositions that allow to modify the Demukron procedure for partitioning into preference levels of an arbitrary digraph are proved. In this case, the condition that equivalent alternatives belong to the same level of preference is satisfied. Using this algorithm, it is possible in particular to build a nonstrict ranking of alternatives. The developed technique can be used to solve multi-criteria problems in case of verbal information about pairwise comparison of alternatives according to the quality criteria.

Keywords: *collective choice, weighted majority graph, aggregated relation, quasi-order, contradictory cycle, preference levels.*

Введение

При принятии сложных решений часто требуется агрегировать информацию о попарном сравнении альтернатив. Традиционно эти задачи относились к разделу «Групповой выбор»: исходная информация задаётся в результате опроса группы экспертов — квалифицированных специалистов в данной предметной области. В последнее время широкое распространение получили алгоритмы многокритериального выбора, использующие вербальную информацию о сравнении альтернатив по критериям качества [1]. И в том и в другом случае информация о предпочтениях на множестве альтернатив задаётся их попарным сравнением, а в алгоритмах принятия решений используется в виде бинарных отношений.

Методы группового выбора, такие, как процедуры Борда, Коупленда, Кемени и многие другие [2], в качестве входной информации используют только частный случай бинарных отношений — ранжирование альтернатив. Но при решении практических задач, особенно с большим числом альтернатив, не всегда удаётся найти экспертов, подробно знакомых с каждым вариантом. Целесообразнее потребовать от специалистов не ранжировать, а только сравнить между собой известные им альтернативы.

В работе [3] предлагается процедура построения мажоритарного графа, позволяющая в качестве исходных использовать произвольные бинарные отношения, что значительно расширяет круг рассматриваемых задач. Мажоритарный граф, построенный по правилу большинства, удовлетворяет требованиям к групповым решениям: сохранение отношения Парето, монотонность, независимость, ненавязанность — и при этом наиболее полно учитывает мнения экспертов. Основным недостатком является то, что мажоритарный граф с большой вероятностью не транзитивен и может содержать контуры, а это затрудняет выбор наилучших альтернатив.

Ещё Ж.-А. Кондорсе (см. [4]) разработал алгоритм построения агрегированного ранжирования, соответствующего пути максимальной длины в нагруженном орграфе. Если в графе существует контур, то Кондорсе разрушает его по «слабому звену», удаляя дугу с минимальным весом, т. е. дугу с минимальной разностью в предпочтениях экспертов. Но П. Янг привёл пример, в котором уже для четырёх альтернатив путь с наибольшей длиной содержит удалённую дугу.

Принцип построения непротиворечивого группового решения путём разрушения контуров используется в [5]. При этом в качестве критерия приближённости агрегированного отношения к исходным берётся суммарное расстояние до экспертных предпочтений [6]. В [5] предложена процедура непротиворечивого агрегирования бинарных отношений строгого порядка. Метод основан на предварительном построении мажоритарного графа, но позволяет получить транзитивное и не содержащее контуров агрегированное отношение. Данная работа является продолжением и в некоторой степени обобщением этого исследования. При задании исходных предпочтений отношениями строгого порядка не учитывается возможность наличия равноценных альтернатив. Равноценные варианты решений наряду со строгим предпочтением одной альтернативы над другой допускает отношение квазипорядка. Квазипорядок — рефлексивное и транзитивное отношение — является обобщением частичного порядка и эквивалентности [7]. В этой работе отношения квазипорядка выбраны в качестве экспертных предпочтений и группового решения. Как и в [5], метод агрегирования экспертных предпочтений основывается на предварительном построении нагруженного мажоритарного графа. Разработана процедура нахождения транзитивного агрегированного предпочтения с учётом особенностей отношения квазипорядка. Она основана на разрушении противоречивых контуров в мажоритарном графе, при этом контуры, содержащие только равноценные альтернативы, сохраняются. Предлагается также алгоритм упорядочения альтернатив по уровням предпочтения, который, в частности, позволяет построить их нестрогое ранжирование.

1. Постановка задачи

Рассматривается множество альтернатив $A = \{a_1, a_2, \ldots, a_n\}$ и множество экспертов $E = \{E_1, E_2, \ldots, E_m\}$. Профиль индивидуальных предпочтений экспертов на множестве A задан отношениями квазипорядка $\rho_1, \rho_2, \ldots, \rho_m$, причём $\langle a_i, a_j\rangle \in \rho_t$, $t = 1, \ldots, m$, если элемент a_i не более предпочтителен, чем элемент a_j (отношения ρ_t можно выбрать и по-другому: $\langle a_i, a_j\rangle \in \rho_t \Leftrightarrow$ элемент a_i не менее предпочтителен, чем

элемент a_j, но стандартные процедуры на графах для выбора наилучших альтернатив и их ранжирования (нахождение устойчивых множеств и разбиение вершин на уровни) удобнее реализовывать для отношения «не более предпочтителен»).

Требуется построить агрегированное отношение $\widehat{\rho}$ на множестве A, согласованное с предпочтениями $\rho_1, \rho_2, \ldots, \rho_m$ и также являющееся квазипорядком. Напомним, что квазипорядком называется рефлексивное и транзитивное отношение [7].

Бинарное отношение ρ будем задавать матрицей предпочтений и матрицей смежности соответствующего графа $G = \langle A, \rho \rangle$.

Определение 1. Под матрицей предпочтений будем понимать квадратную матрицу $P = \|p_{ij}\|$ порядка n (n — число альтернатив) с элементами

$$p_{ij} = \begin{cases} 1, & \text{если } a_i \text{ менее предпочтительна, чем } a_j, \\ 1/2, & \text{если } a_i \text{ и } a_j \text{ равноценны}, \\ 0, & \text{если } a_j \text{ менее предпочтительна, чем } a_i, \text{ или } a_i \text{ и } a_j \text{ не сравнимы} \end{cases}$$

при $i \neq j$. Элемент $p_{ii} = 1$ ($i = 1, \ldots, n$), если отношение ρ рефлексивно; в противном случае $p_{ii} = 0$.

Равноценность двух альтернатив a_i и a_j означает, что отношению ρ одновременно принадлежат пары $\langle a_i, a_j \rangle$ и $\langle a_j, a_i \rangle$ ($i, j = 1, \ldots, n$, $i \neq j$).

Обозначим $P^1, P^2, \ldots, P^m$ матрицы экспертных предпочтений, где m — число экспертов. Профиль экспертных предпочтений $\rho_1, \rho_2, \ldots, \rho_m$ можно также задавать матрицами смежности графов соответствующих отношений. Для отношения квазипорядка матрицы предпочтений и смежности не совпадают. Матрицу смежности можно получить из матрицы предпочтений заменой всех элементов, равных $1/2$, на 1.

2. Построение агрегированного отношения квазипорядка

Потребуем от агрегированного отношения квазипорядка $\widehat{\rho}$, чтобы оно удовлетворяло следующим двум условиям:

1) было непротиворечивым;
2) наиболее полно отражало предпочтения каждого эксперта.

Формализуем условие непротиворечивости отношения с учётом особенностей отношений квазипорядка. Введём понятия симметричной и асимметричной частей отношения. Симметричная часть отношения ρ, заданного на множестве A (обозначается $\operatorname{Sym}\rho$) содержит все такие пары, для которых $\langle a_i, a_j \rangle \in \rho$ и $\langle a_j, a_i \rangle \in \rho$ одновременно ($a_i, a_j \in A$). Асимметричная часть отношения $\operatorname{As}\rho = \rho \setminus \operatorname{Sym}\rho$. Отношение квазипорядка фактически объединяет отношения частичного порядка и эквивалентности, допуская пары из симметричной и асимметричной части [7].

Противоречивость отношения обычно связывают с наличием контуров в его графе [5]. Будем говорить «контур отношения ρ», подразумевая контур соответствующего графа $G = \langle A, \rho \rangle$. Заметим, что множество дуг графа совпадает с множеством упорядоченных пар отношения ρ. Отношение квазипорядка может содержать равноценные альтернативы, следовательно, в нём допускается наличие контуров, полностью принадлежащих симметричной части отношения. Кроме того, отношение квазипорядка содержит петли, так как является рефлексивным. Будем различать противоречивые и непротиворечивые контуры графа.

Определение 2. Контур отношения ρ называется противоречивым, если в нём хотя бы одна пара $\langle a_i, a_j \rangle \in \operatorname{As}\rho$ ($a_i, a_j \in A$).

На рис. 1 изображён непротиворечивый контур: все дуги принадлежат симметричной части отношения и, следовательно, альтернативы a_1, a_2 и a_3 равноценны. Контур на рис. 2 противоречивый: дуга $\langle a_2, a_3\rangle$ принадлежит асимметричной части отношения.

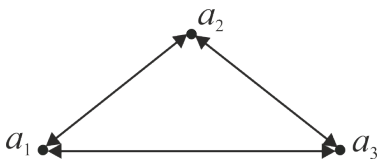


Рис. 1

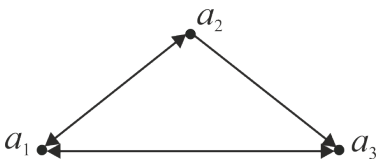


Рис. 2

Определение 3. Отношение ρ называется противоречивым, если оно содержит противоречивый контур.

Таким образом, выполнение условия о непротиворечивости агрегированного отношения означает, что $\widehat{\rho}$ не должно содержать контуров, в которых не все элементы равноценны. Отсутствие таких контуров — необходимое условие для однозначного непустого выбора наилучших альтернатив. Этому условию удовлетворяет отношение квазипорядка.

Лемма 1. Любой контур транзитивного отношения принадлежит симметричной части этого отношения.

Доказательство. Пусть K — контур транзитивного отношения ρ, заданного на множестве A. Покажем, что $K \subseteq \operatorname{Sym}\rho$. Докажем от противного. Пусть существует пара $\langle a_i, a_j\rangle \in (K \cap \operatorname{As}\rho)$, т. е. $\langle a_i, a_j\rangle \in K \subseteq \rho$, но $\langle a_j, a_i\rangle \notin \rho$ $(a_i, a_j \in A)$. Из того, что K — контур и $a_i, a_j \in K$, следует, что существует последовательность пар отношения ρ: $\langle a_j, a_{j_1}\rangle$, $\langle a_{j_1}, a_{j_2}\rangle, \ldots, \langle a_{j_t}, a_i\rangle$. Из транзитивности ρ получим $\langle a_j, a_i\rangle \in \rho$ и, следовательно, $\langle a_i, a_j\rangle \in \operatorname{Sym}\rho$. ■

Утверждение 1. Отношение квазипорядка непротиворечиво.

Доказательство. Следует из того, что отношение квазипорядка по определению транзитивно и, следовательно, по лемме 1 все его контуры принадлежат симметричной части отношения. Таким образом, отношение квазипорядка не может содержать противоречивых контуров, т. е. контуров, содержащих дуги из асимметричной части отношения. ■

Для формализации условия 2 о согласованности агрегированного отношения с экспертными предпочтениями введем понятие расстояния между отношениями.

Определение 4. Расстоянием между двумя отношениями ρ_k и ρ_t назовём величину $\mathrm{d}(\rho_k, \rho_t)$, вычисляемую по формуле

$$\mathrm{d}(\rho_k, \rho_t) = \sum_{i=1}^{n}\sum_{j=1}^{n} |p_{ij}^k - p_{ij}^t|,$$

где p_{ij}^k и p_{ij}^t — элементы матриц предпочтения P^k и P^t.

Под агрегированным отношением, наиболее полно отражающим предпочтения каждого эксперта, будем понимать отношение $\widehat{\rho}$, такое, что сумма расстояний между $\widehat{\rho}$ и отношениями $\rho_1, \rho_2, \ldots, \rho_m$ минимальна:

$$D(\widehat{\rho}) = \sum_{t=1}^{m} \mathrm{d}(\widehat{\rho},\, \rho_t) \to \min\,.$$

Алгоритм нахождения агрегированного отношения квазипорядка основывается на построении нестрогого нагруженного мажоритарного графа.

Определение 5. Нестрогим нагруженным мажоритарным графом назовём ориентированный граф $G = \langle A, \rho_\Sigma \rangle$ с множеством вершин-альтернатив $A = \{a_1, a_2, \ldots, a_n\}$ и дугами $\rho_\Sigma = \{\langle a_i, a_j \rangle : a_i, a_j \in A \text{ и } l_{ij} \geqslant 0\}$, где $l_{ij} = \sum\limits_{k=1}^{m} (p_{ij}^k - p_{ji}^k)$. Каждой дуге $\langle a_i, a_j \rangle \in \rho_\Sigma$ поставим в соответствие вес l_{ij}.

Из определения 5 следует:

$$\begin{aligned} \operatorname{As} \rho_\Sigma =& \{\langle a_i, a_j \rangle : a_i, a_j \in A \text{ и } l_{ij} > 0\}, \\ \operatorname{Sym} \rho_\Sigma =& \{\langle a_i, a_j \rangle : a_i, a_j \in A \text{ и } l_{ij} = l_{ji} = 0\}. \end{aligned}$$

Нестрогий нагруженный мажоритарный граф может содержать дуги с нулевым весом: они соединяют равноценные альтернативы. Это позволит включить равноценные альтернативы и в искомое отношение квазипорядка. Матрицу смежности нагруженного мажоритарного графа $G = \langle A, \rho_\Sigma \rangle$ обозначим R_Σ.

Зададим матрицу весов $C = \|c_{ij}\|$ порядка n, где n — число альтернатив и

$$c_{ij} = \begin{cases} l_{ij}, & \text{если } \langle a_i, a_j \rangle \in \rho_\Sigma, \\ \infty & \text{иначе.} \end{cases}$$

Введём матрицу суммарных предпочтений $P_\Sigma = \|p_{ij}^\Sigma\|$ порядка n, где $p_{ij}^\Sigma = \sum\limits_{k=1}^{m} p_{ij}^k$. С её помощью удобно находить матрицу весов C:

$$c_{ij} = \begin{cases} p_{ij}^\Sigma - p_{ji}^\Sigma, & \text{если } \langle a_i, a_j \rangle \in \rho_\Sigma, \\ \infty & \text{иначе.} \end{cases}$$

Для построения агрегированного квазипорядка $\widehat{\rho}$ найдём предварительно суммарное отношение, не содержащее противоречивых контуров (алгоритм 1).

Алгоритм 1. Построение отношения ρ, не содержащего противоречивых контуров, по ρ_Σ

1: Проверяем граф $G = \langle A, \rho_\Sigma \rangle$ на наличие противоречивых контуров. Если таких контуров нет, то граф $G = \langle A, \rho_\Sigma \rangle$ без весов на дугах и есть граф $G' = \langle A, \rho \rangle$ искомого отношения ρ. Если противоречивые контуры есть, переходим к п. 2.

2: Из противоречивых контуров графа $G = \langle A, \rho_\Sigma \rangle$ удаляем все дуги $\langle a_i, a_j \rangle \in \operatorname{As} \rho$ $(a_i, a_j \in A)$, имеющие наименьший вес среди дуг всех противоречивых контуров. Переходим к п. 1.

Заметим, что по построению нагруженного мажоритарного графа вес дуги из асимметричной части отношения всегда больше нуля, вес дуги из симметричной части равен нулю.

Утверждение 2. Не содержащее противоречивых контуров отношение ρ на множестве A, построенное из отношения ρ_Σ по алгоритму 1, является единственным.

Доказательство. Следует из однозначности действий на каждом шаге алгоритма. ■

При практической реализации алгоритма 1 удобно воспользоваться следующим утверждением и его следствием. Обозначим $\operatorname{Tr}\rho$ транзитивное замыкание отношения ρ, т. е. наименьшее транзитивное отношение, содержащее ρ.

Утверждение 3. Отношение ρ, заданное на множестве A, непротиворечиво тогда и только тогда, когда $\operatorname{As}\rho \subseteq \operatorname{As}(\operatorname{Tr}\rho)$.

Доказательство. Пусть отношение ρ не содержит противоречивых контуров, но не выполняется включение $\operatorname{As}\rho \subseteq \operatorname{As}(\operatorname{Tr}\rho)$. Следовательно, существует упорядоченная пара $\langle a_i, a_j\rangle$ элементов из множества A, такая, что $\langle a_i, a_j\rangle \in \operatorname{As}\rho$ и $\langle a_i, a_j\rangle \in \operatorname{Sym}(\operatorname{Tr}\rho)$. Но в этом случае $\langle a_j, a_i\rangle \in \operatorname{Sym}(\operatorname{Tr}\rho)$ и по определению транзитивного замыкания отношения существует цепочка пар, принадлежащих отношению ρ: $\langle a_j, a_{k_1}\rangle, \ldots, \langle a_{k_t}, a_i\rangle$. Учитывая, что $\langle a_i, a_j\rangle \in \operatorname{As}\rho$, получим противоречивый контур отношения ρ, что не соответствует предположению о непротиворечивости ρ.

Пусть теперь для отношения ρ выполняется $\operatorname{As}\rho \subseteq \operatorname{As}(\operatorname{Tr}\rho)$. Если отношение ρ противоречиво, то в нём существует контур, содержащий дугу из $\operatorname{As}\rho$ и $\operatorname{Sym}(\operatorname{Tr}\rho)$ одновременно, что противоречит предположению. ■

Заметим, что $\operatorname{Sym}\rho \subseteq \operatorname{Sym}(\operatorname{Tr}\rho)$ для произвольного отношения ρ.

Следствие 1. Отношение ρ является непротиворечивым тогда и только тогда, когда $\operatorname{As}\rho \cap \operatorname{Sym}(\operatorname{Tr}\rho) = \varnothing$.

Искомое агрегированное отношение предпочтения $\widehat{\rho}$ должно быть квазипорядком. Найдём $\widehat{\rho}$ по формуле

$$\widehat{\rho} = e \cup \operatorname{Tr}\rho, \tag{1}$$

где e — тождественное отношение (диагональ). Напомним, что отношение ρ получается из отношения ρ_Σ путём разрушения противоречивых контуров. Полученное по формуле (1) отношение рефлексивно и транзитивно, следовательно, является квазипорядком. Его матрицы смежности и предпочтений обозначим соответственно $\widehat{R}$ и $\widehat{P}$.

Будем рассматривать матрицы смежности как элементы булевой алгебры. Введём на множестве матриц операции конъюнкции & и дизъюнкции $\vee$:

$$R^k \& R^t = \|r^k_{ij} \,\&\, r^t_{ij}\|, \quad R^k \vee R^t = \|r^k_{ij} \vee r^t_{ij}\|.$$

Будем обозначать матрицы смежности графов отношений ρ, $\operatorname{Tr}\rho$, $\operatorname{As}\rho$, $\operatorname{Sym}\rho$ соответственно R, $\operatorname{Tr}R$, $\operatorname{As}R$, $\operatorname{Sym}R$. Найдём матрицу смежности $\widehat{R}$ отношения $\widehat{\rho}$ согласно формуле (1):

$$\widehat{R} = E \vee \operatorname{Tr}R.$$

Матрицу предпочтений $\widehat{P}$ получим заменой в матрице смежности $\widehat{R}$ элементов, симметричных относительно главной диагонали и равных 1, на элементы, равные 1/2.

Работа алгоритма 1 основывается на нахождении отношения ρ_K, содержащего контуры отношения ρ_Σ. Матрицу смежности R_K графа отношения ρ_K, содержащего все контуры отношения ρ_Σ, вычислим по следующей формуле [5]:

$$R_K = \widehat{R}_\Sigma \,\&\, (\widehat{R}_\Sigma)^{\mathrm{T}} \,\&\, R_\Sigma,$$

где $\widehat{R}_\Sigma$ — матрица смежности отношения $\operatorname{Tr}\rho_\Sigma$; T — транспонирование матрицы.

Если в отношении ρ_K существуют одновременно дуги $\langle a_i, a_j\rangle$ и $\langle a_j, a_i\rangle$, что соответствует равноценности альтернатив a_i и a_j (следовательно, эти дуги принадлежат $\operatorname{Sym}\rho_\Sigma$), то они из контура не удаляются. Заметим, что эти дуги легко распознать:

их вес равен нулю. Таким образом, из контура удаляются только дуги, вес которых минимальный, но больше нуля, т. е. принадлежащие $\operatorname{As}\rho_\Sigma$. Наличие таких дуг в отношении ρ_K свидетельствует о существовании противоречивых контуров в отношении ρ_Σ.

При реализации алгоритма 1 для построения отношения ρ, не содержащего противоречивых контуров, можно выделить только дуги противоречивых контуров из $\operatorname{As}\rho_\Sigma$.

Утверждение 4. Справедлива формула

$$\operatorname{As}\rho_K = \operatorname{As}\rho_\Sigma \cap \operatorname{Sym}(\operatorname{Tr}\rho_\Sigma). \tag{2}$$

Доказательство. Пусть $\langle a_i, a_j\rangle \in \operatorname{As}\rho_K$, тогда из построения ρ_K следует $\langle a_i, a_j\rangle \in \operatorname{As}\rho_\Sigma$ (если $\langle a_i, a_j\rangle \in \operatorname{Sym}\rho_\Sigma$, то $\langle a_i, a_j\rangle \in \operatorname{Sym}\rho_K$). Условие $\langle a_i, a_j\rangle \in \rho_K$ означает, что $\langle a_i, a_j\rangle \in K \subseteq \rho_K$, где K — некоторый контур из ρ_Σ. По лемме 1 $K \subseteq \operatorname{Sym}(\operatorname{Tr}\rho_\Sigma) \Rightarrow \langle a_i, a_j\rangle \in \operatorname{Sym}(\operatorname{Tr}\rho_\Sigma)$. С учётом $\langle a_i, a_j\rangle \in \operatorname{As}\rho_\Sigma$ получим $\langle a_i, a_j\rangle \in$ $\in \operatorname{As}\rho_\Sigma \cap \operatorname{Sym}(\operatorname{Tr}\rho_\Sigma)$.

Пусть $\langle a_i, a_j\rangle \in \operatorname{As}\rho_\Sigma \cap \operatorname{Sym}(\operatorname{Tr}\rho_\Sigma)$. Из $\langle a_i, a_j\rangle \in \operatorname{As}\rho_\Sigma$ следует $\langle a_i, a_j\rangle \in \rho_\Sigma$, но $\langle a_j, a_i\rangle \notin \rho_\Sigma$. Из $\langle a_i, a_j\rangle \in \operatorname{Sym}(\operatorname{Tr}\rho_\Sigma)$ следует, что $\langle a_j, a_i\rangle \in \operatorname{Tr}\rho_\Sigma$ и существует цепочка пар, принадлежащих отношению ρ_Σ: $\langle a_j, a_{k_1}\rangle, \ldots, \langle a_{k_t}, a_i\rangle$. С учётом $\langle a_i, a_j\rangle \in \rho_\Sigma$ получаем контур в ρ_Σ, т. е. $\langle a_i, a_j\rangle \in \rho_K$. Но $\langle a_j, a_i\rangle \notin \rho_\Sigma$, значит, $\langle a_j, a_i\rangle \notin \rho_K$ ($\rho_K \subseteq \rho_\Sigma$) и, следовательно, $\langle a_i, a_j\rangle \in \operatorname{As}\rho_K$. ■

В матричном виде формула (2) примет вид

$$\operatorname{As} R_K = \operatorname{As} R_\Sigma \,\&\, \operatorname{Sym}(\operatorname{Tr} R_\Sigma).$$

Согласно алгоритму 1, из отношения $\operatorname{As}\rho_K$ поочерёдно удаляются все дуги, имеющие минимальный вес, до тех пор, пока противоречивые контуры отношения ρ_Σ не будут разрушены.

Заметим, что для произвольного отношения q с матрицей смежности Q матрицы смежности отношений $\operatorname{As} q$ и $\operatorname{Sym} q$ находятся по формулам [5]

$$\operatorname{Sym} Q = Q \,\&\, Q^T, \quad \operatorname{As} Q = Q - \operatorname{Sym} Q.$$

Предлагаемый метод согласования экспертных предпочтений состоит из двух частей: построения нагруженного мажоритарного графа и разрушения контуров с целью построения непротиворечивого агрегированного отношения. Можно осуществить выбор наилучших альтернатив на основе мажоритарного графа, не прибегая к разрушению противоречивых контуров, например, для упорядочения альтернатив воспользоваться процедурой Коупленда. Для мажоритарного графа данная процедура сводится к вычислению индексов вершин, равных разности количеств входящих в вершину и исходящих из неё дуг (для отношения «не более предпочтительна»). Для нагруженного мажоритарного графа процедуру можно усилить, учитывая веса на дугах: складывать и вычитать веса соответствующих дуг. Затем упорядочить вершины графа в соответствии с вычисленными индексами: лучшая вершина-альтернатива имеет наибольший индекс.

Наибольшую вычислительную сложность имеет алгоритм нахождения транзитивного замыкания отношения — $\mathrm{O}(n^3)$. В случае построения ранжирования альтернатив через индексы вершин мажоритарного графа сложность алгоритма уменьшится до $\mathrm{O}(n^2)$.

Приведём пример, демонстрирующий возможности метода непротиворечивого агрегирования отношений квазипорядка. Во всех примерах в качестве экспертной информации будем брать отношения, в которых все альтернативы сравнимы между собой.

Это даёт возможность сравнить предложенную процедуру с известными, хотя метод работает и в случае, когда часть альтернатив попарно несравнимы.

Пример 1. Пусть множество A содержит пять альтернатив $A = \{a_1, a_2, a_3, a_4, a_5\}$ и профиль предпочтений трех экспертов — нестрогие ранжирования (отношение квазипорядка, при котором любые два элемента сравнимы между собой). Требуется построить агрегированное отношение, также являющееся квазипорядком. Нестрогие ранжирования альтернатив представлены в таблице. В верхней строке расположены наиболее предпочтительные альтернативы. Равноценные альтернативы расположены на одной строке.

ρ_1	ρ_2	ρ_3
a_4	a_5	a_1
a_1	a_3	a_2
$a_2 - a_3$	$a_1 - a_2$	$a_3 - a_5$
a_5	a_4	a_4

Представим отношения ρ_1, ρ_2, ρ_3 матрицами предпочтений:

$$P^1 = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1/2 & 1 & 0 \\ 1 & 1/2 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix};\ P^2 = \begin{pmatrix} 1 & 1/2 & 1 & 0 & 1 \\ 1/2 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix};\ P^3 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1/2 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1/2 & 0 & 1 \end{pmatrix}.$$

Матрица суммарных предпочтений $P_\Sigma = P^1 + P^2 + P^3$ имеет вид

$$P_\Sigma = \begin{pmatrix} 3 & 1/2 & 1 & 1 & 1 \\ 5/2 & 3 & 3/2 & 1 & 1 \\ 2 & 3/2 & 3 & 1 & 3/2 \\ 2 & 2 & 2 & 3 & 2 \\ 2 & 2 & 3/2 & 1 & 3 \end{pmatrix}.$$

Следовательно, матрицы смежности и весов мажоритарного графа можно записать так:

$$R_\Sigma = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 \end{pmatrix}, \quad C = \begin{pmatrix} 0 & \infty & \infty & \infty & \infty \\ 2 & 0 & 0 & \infty & \infty \\ 1 & 0 & 0 & \infty & 0 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & \infty & 0 \end{pmatrix}.$$

Мажоритарный граф (рис. 3) содержит противоречивые контуры. Разрушим их, удалив, согласно алгоритму, дуги с весом 1 (одна дуга, которая изображена на рис. 3 пунктирной линией). Получим матрицу смежности R и матрицу предпочтений R^p отношения, не содержащего противоречивых контуров. Вычислив транзитивное замыкание отношения ρ, получим искомый квазипорядок $\widehat{\rho}$. При этом к отношению ρ добавятся дуги $\langle a_2, a_5\rangle$ и $\langle a_5, a_2\rangle$. Матрицы смежности и предпочтений искомого агрегированного отношения квазипорядка имеют следующий вид:

$$\widehat{R} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 \end{pmatrix}, \quad \widehat{P} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1/2 & 0 & 1/2 \\ 1 & 1/2 & 1 & 0 & 1/2 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1/2 & 1/2 & 0 & 1 \end{pmatrix}.$$

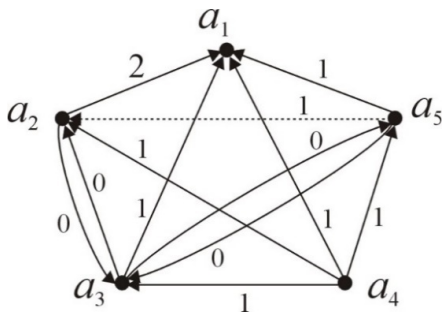


Рис. 3

Граф на рис. 3 не полностью соответствует своей матрице смежности: для большей наглядности производимых вычислений в нём отсутствуют петли.

Минимальное суммарное расстояние до экспертных предпочтений $D(\rho_\Sigma) = 19$ [3]; для агрегированного отношения $\widehat{\rho}$ суммарное расстояние до экспертных предпочтений $D(\widehat{\rho}) = 20$.

Сравним этот результат с суммарными расстояниями, полученными по простой процедуре Коупленда и модифицированной — с разностью весов дуг. Обе процедуры основаны на предварительном построении мажоритарного графа. Простая процедура Коупленда приписывает каждой вершине-альтернативе индекс, равный разности количеств входящих и исходящих из неё дуг. Модифицированная процедура приписывает каждой вершине-альтернативе индекс, равный разности сумм весов входящих и исходящих дуг. Нестрогое ранжирование, полученное по модифицированной процедуре Коупленда, — $\begin{pmatrix} a_1 \\ a_2 - a_3 \\ a_5 \\ a_4 \end{pmatrix}$ — даёт расстояние 20: $a_1(5)$, $a_2(0)$, $a_3(0)$, $a_4(-4)$, $a_5(-1)$.

Используя обычную процедуру Коупленда, получим $\begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_5 \\ a_4 \end{pmatrix}$, суммарное расстояние — 21: $a_1(4)$, $a_2(1)$, $a_3(0)$, $a_4(-4)$, $a_5(-1)$.

Нестрогое ранжирование с минимальным суммарным расстоянием 20 можно также получить, используя процедуру построения медианы Кемени, но она трудна алгоритмически и имеет экспоненциальную вычислительную сложность.

Процедура непротиворечивого агрегирования предпочтений может применяться и в случае задания профиля экспертных предпочтений произвольными бинарными отношениями. Заметим, что противоречивые исходные отношения могут привести к нарушению второй аксиомы Эрроу о сохранении в агрегированном отношении предпочтений, совпадающих у всех экспертов [2]. Рекомендуется в качестве исходных брать транзитивные отношения, что исключает их противоречивость.

Для применения разработанного алгоритма в случае задания численных оценок альтернатив необходимо воспользоваться предложенным в [8] способом формирования матриц предпочтения.

3. Процедура упорядочения альтернатив на основе отношения квазипорядка

Построение агрегированного отношения предпочтения предшествует процедуре выбора наилучших альтернатив или их ранжированию. Для реализации этих процедур обычно используются алгоритмы на графах. Подмножества независимых и доминирующих альтернатив можно найти, используя метод Магу нахождения внутренне и внешне устойчивых множеств вершин графа, а также ядра графа, являющегося одновременно внутренне и внешне устойчивым. Для графа без контуров ядро определяется однозначно. Упорядочить альтернативы по предпочтению, в частности ранжировать их, можно с помощью алгоритма Демукрона разбиения графа без контуров на уровни [9]. Применение этих процедур для графа, содержащего контуры, либо невозможно, либо может привести к неоднозначному или пустому выбору альтернатив.

Отношение квазипорядка допускает наличие непротиворечивых контуров, т. е. контуров, в которые входят только равноценные альтернативы. Модифицируем алгоритм Демукрона таким образом, чтобы его можно было применить при наличии в графе агрегированного отношения непротиворечивых контуров. Равноценные альтернативы при этом должны принадлежать одному уровню предпочтения. Это позволит на основе отношения квазипорядка упорядочить альтернативы, в частности построить их нестрогое ранжирование. Напомним, что под нестрогим ранжированием понимается такое отношение квазипорядка, что любые два элемента множества, на котором оно задано, сравнимы.

Дадим определение уровней графа. Пусть $G = \langle A, \rho \rangle$ — орграф с множеством вершин $A = \{a_1, a_2, \ldots, a_n\}$ и множеством дуг ρ.

Определение 6 [9]. Уровнями $N_0, N_1, \ldots, N_t$ графа $G = \langle A, \rho \rangle$ называются следующие непустые множества вершин графа:

$$\begin{aligned} &N_0 = \{a_i : a_i \in A,\ \Gamma a_i = \varnothing\},\\ &N_1 = \{a_i : a_i \in A \setminus N_0,\ \Gamma a_i \subseteq N_0\},\\ &\ldots\\ &N_t = \{a_i : a_i \in A \setminus \bigcup_{j=0}^{t-1} N_j,\ \Gamma a_i \subseteq \bigcup_{j=0}^{t-1} N_j\},\\ &\Gamma^{-1} N_t = \varnothing. \end{aligned} \tag{3}$$

Здесь для любой вершины a_i определены вершины из её образа Γa_i и прообраза $\Gamma^{-1} a_i$:

$$\Gamma a_i = \{a_j : \langle a_i, a_j \rangle \in \rho,\ a_j \in A\},$$
$$\Gamma^{-1} a_i = \{a_j : \langle a_j, a_i \rangle \in \rho,\ a_j \in A\}.$$

Утверждение 5 [9]. Орграф можно разбить на уровни тогда и только тогда, когда он не содержит контуров.

Рассмотрим транзитивный граф $G = \langle A, \rho \rangle$ с множеством вершин $A = \{a_1, a_2, \ldots, a_n\}$, который может содержать контуры, состоящие из равноценных альтернатив.

Утверждение 6. Пусть альтернативы a_i и a_j равноценны по транзитивному отношению ρ, заданному на множестве A. Тогда $\Gamma a_i = \Gamma a_j$ $(a_i, a_j \in A)$.

Доказательство. Пусть $a_k \in \Gamma a_i$. Тогда $\langle a_i, a_k \rangle \in \rho$. В силу транзитивности ρ из $\langle a_j, a_i \rangle \in \rho$ и $\langle a_i, a_k \rangle \in \rho$ получим $\langle a_j, a_k \rangle \in \rho$. Отсюда $a_k \in \Gamma a_j$ и, следовательно, $\Gamma a_i \subseteq \Gamma a_j$.

Аналогично доказывается, что $\Gamma a_j \subseteq \Gamma a_i$. ■

Утверждение 7. Равноценные по транзитивному отношению ρ альтернативы принадлежат одному уровню графа отношения $\operatorname{As}\rho$.

Доказательство. Граф $G = \langle A, \operatorname{As}\rho\rangle$ не содержит контуров, так как по лемме 1 в транзитивном отношении ρ все контуры принадлежат $\operatorname{Sym}\rho$. Следовательно, по утверждению 5 граф $G = \langle A, \operatorname{As}\rho\rangle$ можно разбить на уровни $N_0, N_1, \ldots, N_t$ (3).

Заметим, что уровни $N_0, N_1, \ldots, N_t$ полностью определяются множествами образов вершин Γa_i. Поскольку у равноценных альтернатив a_i и a_j в силу утверждения 6 $\Gamma a_i = \Gamma a_j$, то a_i и a_j принадлежат одному уровню графа отношения $\operatorname{As}\rho$ ($a_i, a_j \in A$). ■

Пусть $\widehat{\rho}$ — отношение квазипорядка, построенное на множестве альтернатив $A = \{a_1, a_2, \ldots, a_n\}$. Отношение $\widehat{\rho}$ транзитивно и, согласно утверждению 7, равноценные альтернативы находятся на одном уровне графа отношения $\operatorname{As}\widehat{\rho}$. Этот факт позволяет предложить для упорядочения альтернатив алгоритм 2.

Алгоритм 2. Упорядочение альтернатив на основе отношения квазипорядка

1: Найдем $\operatorname{As}\widehat{\rho}$ — асимметричную часть отношения квазипорядка $\widehat{\rho}$.
2: Разбить граф отношения $\operatorname{As}\widehat{\rho}$ на уровни $N_0, N_1, \ldots, N_t$ (3), используя алгоритм Демукрона.
3: Восстановить на каждом из уровней дуги из $\operatorname{Sym}\widehat{\rho}$.
4: Упорядочить альтернативы согласно уровням $N_0, N_1, \ldots, N_t$; наилучшие альтернативы будут принадлежать уровню N_0.

Отметим, что на каждом из уровней могут оказаться как равноценные, так и несравнимые вершины-альтернативы. Нестрогое ранжирование альтернатив можно получить только в случае, когда все альтернативы сравнимы по отношению квазипорядка.

Пример 2. Воспользовавшись полученным в примере 1 отношением квазипорядка $\widehat{\rho}$, проведём ранжирование альтернатив множества $A = \{a_1, a_2, a_3, a_4, a_5\}$. Матрица смежности графа отношения $\widehat{\rho}$ имеет вид

$$\widehat{R} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 \end{pmatrix}.$$

Найдём матрицу смежности асимметричной части отношения $\widehat{\rho}$:

$$\operatorname{As}\widehat{R} = \widehat{R} - \operatorname{Sym}\widehat{R} = \widehat{R} - (\widehat{R}\,\&\,\widehat{R}^{\mathrm{T}}) =$$

$$= \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 \end{pmatrix} - \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Используя алгоритм Демукрона, разобьем граф отношения $\operatorname{As}\widehat{\rho}$ на уровни (3). Этапы работы алгоритма легко проследить на примере.

Согласно алгоритму, просуммируем элементы каждой строки матрицы смежности $\operatorname{As}\widehat{R}$. Получим вектор $L^{(0)} = (0\ 1\ 1\ 4\ 1)^{\mathrm{T}}$. Элемент $l_1^{(0)}$ равен 0, следовательно, $a_1 \in N_0$.

Обнуляем первый столбец матрицы смежности (соответствующий вершине a_1):

$$\operatorname{As}\widehat{R}1 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Найдём вектор, компоненты которого равны сумме единиц каждой строки полученной матрицы (кроме первой строки): $L^{(1)} = (*\ 0\ 0\ 3\ 0)^{\mathrm{T}}$. Элемент $l_1^{(1)} = *$, так как $a_1 \in N_0$; $l_2^{(1)} = l_3^{(1)} = l_5^{(1)} = 0$, следовательно, $a_2, a_3, a_5 \in N_1$.

Обнуляем второй, третий и пятый столбцы матрицы $\operatorname{As}\widehat{R}1$, получаем

$$\operatorname{As}\widehat{R}2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

следовательно, $L^{(2)} = (*\ *\ *\ 0\ *)^{\mathrm{T}}$ и $a_4 \in N_2$.

Разбиение графа отношения $\operatorname{As}\widehat{\rho}$ на уровни с добавлением дуг графа отношения $\widehat{\rho}$ (кроме петель) представлено на рис. 4. Соответствующее нестрогое ранжирование альтернатив: $\begin{pmatrix} a_1 \\ a_2 - a_3 - a_5 \\ a_4 \end{pmatrix}$.

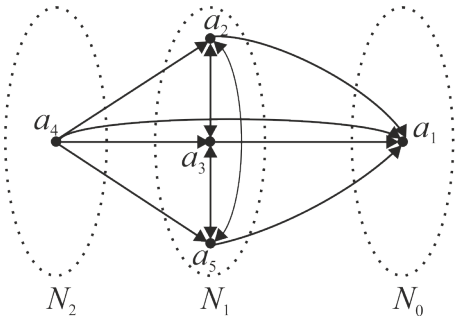


Рис. 4

Из рис. 4 видно, что построенное агрегированное отношение позволяет осуществить однозначный выбор наилучшей альтернативы. Существование дуг $\langle a_2, a_1\rangle$, $\langle a_3, a_1\rangle$, $\langle a_4, a_1\rangle$ и $\langle a_5, a_1\rangle$ указывает на то, что альтернатива a_1 предпочтительнее всех остальных — победитель по Кондорсе [2].

Заметим, что если не все альтернативы попарно сравнимы с помощью построенного квазипорядка, то, используя предложенный алгоритм, мы получим упорядочение, но не ранжирование альтернатив, при этом наилучшие альтернативы будут принадлежать уровню N_0.

Заключение

В работе предложена методика непротиворечивого агрегирования отношений квазипорядка, основанная на построении нагруженного мажоритарного графа с весами на

дугах, характеризующими различия в предпочтениях экспертов. Для построения транзитивного группового отношения используется алгоритм разрушения противоречивых контуров путём удаления дуг из асимметричной части отношения, имеющих наименьший вес. Предложен алгоритм упорядочения альтернатив по предпочтительности на основе отношения квазипорядка. Разработанные алгоритмы могут использоваться и при решении многокритериальных задач, в частности при задании численных оценок альтернатив по критериям качества [8]. Все алгоритмы обладают вычислительной сложностью не более $O(n^3)$ от числа альтернатив, что даёт возможность эффективно применять их для решения задач с большими массивами исходных данных.

ЛИТЕРАТУРА

1. *Ларичев О. И.* Вербальный анализ решений / под ред. А. Б. Петровского. М.: Наука, 2006. 400 с.
2. *Петровский А. Б.* Теория принятия решений. М.: Академия, 2009. 400 с.
3. *Миркин Б. Г.* Проблема группового выбора. М.: Наука, 1974. 256 с.
4. *Мулен Э.* Кооперативное принятие решений: Аксиомы и модели. М.: Мир, 1991. 464 с.
5. *Нефедов В. Н., Осипова В. А., Смерчинская С. О., Яшина Н. П.* Непротиворечивое агрегирование отношений строгого порядка // Изв. вузов. Математика. 2018. № 5. С. 71–85.
6. *Нефедов В. Н., Смерчинская С. О., Яшина Н. П.* Построение агрегированного отношения, минимально удаленного от экспертных предпочтений // Прикладная дискретная математика. 2018. № 42. С. 120–132.
7. *Шрейдер Ю. А.* Равенство, сходство, порядок. М.: Наука, 1971. 256 с.
8. *Smerchinskaya S. O. and Yashina N. P.* An algorithm for pairwise comparison of alternatives in multi-criteria problems // Intern. J. Modeling, Simulation, and Scientific Computing. 2018. V. 9. No. 1. https://www.worldscientific.com/doi/abs/10.1142/S179396231850006X.
9. *Нефедов В. Н., Осипова В. А.* Курс дискретной математики. М.: Изд-во МАИ, 1992. 262 с.

REFERENCES

1. *Larichev O. and Moshkovich H.* Verbal Decision Analysis for Unstructured Problems. Boston, Kluwer Acedemic Publishers, 1997. 272 p.
2. *Petrovskiy A. B.* Teoriya prinyatiya resheniy [Decision Making Theory]. Moscow, Akademiya Publ., 2009. 400 p. (in Russian)
3. *Mirkin B. G.* Group Choice. V. H. Winston & Sons Publ., 1979. 252 p.
4. *Moulin H.* Axioms of Cooperative Decision Making. Cambridge, Cambridge University Press, 1988. 348 p.
5. *Nefyodov V. N., Osipova V. A., Smerchinskaya S. O., and Yashina N. P.* Non-contradictory aggregation of strict order relations. Russian Mathematics, 2018, vol. 62, pp. 61–73.
6. *Nefyodov V. N., Smerchinskaya S. O., and Yashina N. P.* Postroenie agregirovannogo otnosheniya, minimal'no udalyonnogo ot ehkspertnyh predpochtenij [Constructing an aggregated relation with a minimum distance from the expert preferences]. Prikladnaya Diskretnaya Matematika, 2018, no. 42, pp. 120–132. (in Russian)
7. *Schreider Ju. A.* Equality, Resemblance and Order. Moscow, Mir Publ., 1979. 279 p.
8. *Smerchinskaya S. O. and Yashina N. P.* An algorithm for pairwise comparison of alternatives in multi-criteria problems. Intern. J. Modeling, Simulation, and Scientific Computing, 2018, vol. 9, no. 1. www.worldscientific.com/doi/abs/10.1142/S179396231850006X.
9. *Nefedov V. N. and Osipova V. A.* Kurs diskretnoy matematiki [Discrete Mathematics Course]. Moscow, MAI Publ., 1992. 262 p. (in Russian)

СВЕДЕНИЯ ОБ АВТОРАХ

АБРОСИМОВ Михаил Борисович — доктор физико-математических наук, доцент, заведующий кафедрой Саратовского национального исследовательского государственного университета им. Н. Г. Чернышевского, г. Саратов.
E-mail: **mic@rambler.ru**

АЛЕХИНА Марина Анатольевна — доктор физико-математических наук, профессор, заведующая кафедрой математики и физики Пензенского государственного технологического университета, г. Пенза. E-mail: **alekhina@penzgtu.ru, alekhina.marina19@yandex.ru**

ГРАБОВСКАЯ Светлана Михайловна — кандидат физико-математических наук, доцент кафедры компьютерных технологий Пензенского государственного университета, г. Пенза. E-mail: **swetazin@mail.ru**

ГУСЫНИНА Юлия Сергеевна — кандидат технических наук, доцент кафедры математики и физики Пензенского государственного технологического университета, г. Пенза. E-mail: **gusynina@mail.ru**

ИЛЬЕВ Виктор Петрович — доктор физико-математических наук, профессор, профессор Омского государственного университета им. Ф. М. Достоевского, г. Омск.
E-mail: **iljev@mail.ru**

ИЛЬЕВА Светлана Диадоровна — кандидат физико-математических наук, инженер-программист Омского государственного университета им. Ф. М. Достоевского, г. Омск. E-mail: **iljeva@mail.ru**

КОСОЛАПОВ Юрий Владимирович — кандидат технических наук, доцент Южного федерального университета, г. Ростов-на-Дону. E-mail: **itaim@mail.ru**

КОСТЮК Юрий Леонидович — доктор технических наук, профессор, профессор Национального исследовательского Томского государственного университета, г. Томск.
E-mail: **kostyuk_y_l@sibmail.com**

МОРШИНИН Александр Владимирович — аспирант Института математики им. С. Л. Соболева СО РАН, г. Омск. E-mail: **morshinin.alexander@gmail.com**

НАЗИНА Светлана Витальевна — студентка кафедры программного обеспечения вычислительной техники и автоматизированных систем Белгородского государственного технологического университета им. В. Г. Шухова, г. Белгород.
E-mail: **lanalana9808@gmail.com**

НЕФЕДОВ Виктор Николаевич — кандидат физико-математических наук, доцент, доцент Московского авиационного института, г. Москва.
E-mail: **nefedovvn54@yandex.ru**

НИКИТИН Алексей Юрьевич — инженер-исследователь лаборатории комбинаторных и вычислительных методов алгебры и логики Института математики им. С. Л. Соболева СО РАН, г. Омск. E-mail: **nikitinlexey@gmail.com**

РЫБАЛОВ Александр Николаевич — кандидат физико-математических наук, старший научный сотрудник лаборатории комбинаторных и вычислительных методов алгебры и логики Института математики им. С. Л. Соболева СО РАН, г. Омск.
E-mail: **alexander.rybalov@gmail.com**

РЯБОВ Владимир Геннадьевич — кандидат физико-математических наук, генеральный директор НП «ГСТ», Москва. E-mail: **4vryabov@gmail.com**

РЯЗАНОВ Юрий Дмитриевич — доцент, доцент Белгородского государственного технологического университета им. В. Г. Шухова, г. Белгород.
E-mail: **razanov.yd@bstu.ru**

СЕЛИВЕРСТОВ Александр Владиславович — кандидат физико-математических наук, ведущий научный сотрудник Института проблем передачи информации им. А. А. Харкевича Российской академии наук, г. Москва. E-mail: **slvstv@iitp.ru**

СМЕРЧИНСКАЯ Светлана Олеговна — кандидат физико-математических наук, старший преподаватель Московского авиационного института, г. Москва.
E-mail: **svetlana_os@mail.ru**

ТАРКОВ Михаил Сергеевич — кандидат технических наук, доцент, старший научный сотрудник Института физики полупроводников им. А. В. Ржанова СО РАН, г. Новосибирск. E-mail: **tarkov@isp.nsc.ru**

ТУРЧЕНКО Олег Юрьевич — аспирант Южного федерального университета, г. Ростов-на-Дону. E-mail: **olegmmcs@gmail.comu**

ЯШИНА Нина Павловна — кандидат физико-математических наук, доцент, доцент Московского авиационного института, г. Москва.
E-mail: **nina_p_ yashina@mail.ru**

El-MESADY Ahmed — Menoufia University, Menouf, Egypt.
E-mail: **ahmed_mesady88@yahoo.com**

El-SHANAWANY Ramadan — Menoufia University, Menouf, Egypt.
E-mail: **ramadan_elshanawany380@yahoo.com**