

ISSN 2222-8772

**МАТЕМАТИЧЕСКИЕ
СТРУКТУРЫ
И
МОДЕЛИРОВАНИЕ**

**№ 4(36)
2015**



**МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ**

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«ОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМ. Ф.М. ДОСТОЕВСКОГО»**

**МАТЕМАТИЧЕСКИЕ
СТРУКТУРЫ
И
МОДЕЛИРОВАНИЕ**

№ 4(36)

Омск
2015

Математические структуры и моделирование. — Омск : Омский государственный университет, 2015. — № 4(36). — 145 с.

ISSN 2222-8772 (print)

ISSN 2222-8799 (online)

Редакционная коллегия

- Н. Ф. Богаченко** технический редактор, канд. физ.-мат. наук, доцент, Омский государственный университет им. Ф.М. Достоевского.
- А. Ю. Веснин** д.ф.-м.н., профессор Новосибирского государственного университета, член-корреспондент РАН, заведующий лабораторией прикладного анализа, Институт математики им. С.Л. Соболева Сибирского отделения Российской академии наук, г. Новосибирск.
- В. Я. Волков** доктор техн. наук, профессор, зав. кафедрой начертательной геометрии, инженерной и машинной графики, Сибирская государственная автомобильно-дорожная академия (СибАДИ).
- В. П. Голубятников** доктор физ.-мат. наук, профессор Новосибирского государственного университета, главный научный сотрудник Института математики СО РАН, г. Новосибирск.
- С. И. Горлов** доктор физ.-мат. наук, профессор, ректор Нижневартковского государственного университета.
- А. Г. Гринь** доктор физ.-мат. наук, профессор, кафедра кибернетики, Омский государственный университет им. Ф.М. Достоевского.
- А. К. Гуц** главный редактор, председатель редакционной коллегии, доктор физ.-мат. наук, профессор, зав. кафедрой кибернетики, Омский государственный университет им. Ф.М. Достоевского.
- В. А. Еровенко** доктор физ.-мат. наук, профессор, зав. кафедрой общей математики и информатики Белорусского государственного университета, г. Минск, Республика Беларусь.
- V. Zilber** Dr.Sc.(Phys.-Math.), Professor of Mathematical Logic, Mathematical Institute, University of Oxford.
- А. Н. Кабанов** канд. физ.-мат. наук, кафедра кибернетики, Омский государственный университет им. Ф.М. Достоевского.
- П. А. Корчагин** доктор техн. наук, профессор, Сибирская государственная автомобильно-дорожная академия (СибАДИ).
- V. Kreinovich** Ph.D.(Phys.-Math.), Professor, Computer Science Department, University of Texas at El Paso, Texas, USA.
- Д. Н. Лавров** выпускающий редактор, канд. техн. наук, доцент, зав. каф. компьютерных технологий и сетей, Омский государственный университет им. Ф.М. Достоевского.
- А. А. Fedorenko** Ph.D.(Phys.-Math.), Researcher (CR1) at the French National Centre of Scientific Research (CNRS) Laboratoire de Physique de l'ENS-Lyon, France.

Учредитель

Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Омский государственный университет им. Ф. М. Достоевского».

Свидетельство о регистрации средства массовой информации ПИ № ФС77-57908 от 28 апреля 2014 г.

Адрес научной редакции

Россия, 644077, Омск,
пр. Мира 55А, Омский
государственный университет
им. Ф. М. Достоевского,
факультет компьютерных наук.

E-mail: guts@omsu.ru,
lavrov@omsu.ru

**МАТЕМАТИЧЕСКИЕ
СТРУКТУРЫ
И
МОДЕЛИРОВАНИЕ**

Журнал основан в 1998 году.

В журнале публикуются статьи, в которых излагаются результаты исследований по фундаментальной и прикладной математике, теоретической физике, компьютерным наукам, философии и истории математики и информатики, а также размышления, касающиеся окружающей нас природы и общества. Объекты исследования должны быть представлены в форме некоторых математических структур и моделей.

Все статьи журнала проходят обязательное рецензирование.

Журнал является реферлируемым. Рефераты статей публикуются в «Реферативном журнале», в журналах «Zentralblatt für Mathematik» (Германия), «Mathematical Reviews» (США), индексируется в РИНЦ (elibrary.ru).

Входит в Перечень научных изданий, в которых должны быть опубликованы основные результаты диссертаций на соискание учёных степеней в соответствии с приказом Минобрнауки России от 25 июля 2014 г. № 793.

Электронная версия журнала представлена в сети Интернет по адресам:

<http://msm.univer.omsk.su>
<http://msm.omsu.ru>

Подписной индекс по каталогу «Пресса России»: 94082

Электронная почта главного редактора:

guts@omsu.ru

Электронная почта выпускающего редактора:

lavrov@omsu.ru

СОДЕРЖАНИЕ

Фундаментальная математика и физика

- В.В. Варламов. *Комплексный момент и спин-зарядовое гильбертово пространство* 5
- А.Г. Гринь. *О моментах обобщённых сумм* 23
- В.А. Еровенко. *Концепция фрактала Мандельброта с математической и философской точек зрения* 29
- Н.В. Михайлова. *Теория бесконечных множеств Кантора в контексте генезиса философии математики* 40
- V. Kreinovich, O. Kosheleva. *Paradox of Choice: A Possible Explanation* 49
- G.A. Noskov. *Search Problems with a Promise and Graph Isomorphism* 53
- J. Urenda, D. Finston, V. Kreinovich. *Once We Know that a Polynomial Mapping Is Rectifiable, We Can Algorithmically Find a Rectification* 67

Прикладная математика и моделирование

- В.В. Гольпяпин, В.А. Шовин. *Алгоритмы КО-РА и ФОРДИАСИМПТ как методы распознавания двух образов в пространстве двоичных признаков* 74
- А.К. Гуц, Л.А. Паутова. *Социальная топология и топосы Гротендика* 93
- В.В. Коробицын, Ю.В. Фролова. *Метод локальных координат для вычисления траектории гибридной динамической системы при скольжении* 110
- В.А. Шовин. *Многомерное шкалирование на базе метода Верле* 117

Компьютерные науки

- Н.В. Манюкова. *Компьютерное зрение как средство извлечения информации из видеоряда* 123

Продолжение на следующей странице

Наши публикации



Информационная безопасность

Д.М. Бречка, В.В. Зубова. *Компьютерное моделирование системы разграничения доступа Харрисона-Рузсо-Ульмана*.....129

В.С. Виноградов, В.В. Коробицын, М.Н. Московцев. *Использование турбо-кодека для безопасной передачи данных*.....134

КОМПЛЕКСНЫЙ МОМЕНТ И СПИН-ЗАРЯДОВОЕ ГИЛЬБЕРТОВО ПРОСТРАНСТВО

В.В. Варламов

д.ф.-м.н., e-mail: varlamov@subsiu.ru

Сибирский государственный индустриальный университет

Аннотация. Массовая формула и собственные подпространства оператора энергии определяются в рамках комплексной оболочки (комплексного момента) групповой алгебры собственной группы Лоренца. Собственные подпространства оператора энергии отождествляются с энергетическими (массовыми) уровнями спектра элементарных частиц. Элементарная частица задаётся суперпозицией векторов состояния спин-зарядового гильбертова пространства.

Ключевые слова: комплексный момент, группа Лоренца, спинорная структура, гильбертово пространство.

1. Введение

Согласно господствующей ныне парадигме близкодействия, элементарные частицы взаимодействуют друг с другом посредством переносчиков полей взаимодействий. Ареной таких взаимодействий является пространство-время, что придаёт последнему ярко выраженный субстанциональный характер. Предполагается, что все процессы микромира разворачиваются в пространстве-времени¹. Неудивительно, что как следствие такого представления все развитие релятивистской квантовой теории происходило в рамках картины Шрёдингера, в которой волновая функция (вектор гильбертова пространства) зависит от времени. Долгое время считалось, что картина Гейзенберга, в которой векторы состояния не зависят от времени, эквивалентна картине Шрёдингера. Первым, кто обратил внимание на неэквивалентность этих двух картин, был Дирак. Дирак [2] утверждал, что картина Шрёдингера слишком узка, поскольку векторы состояния этой картины определены в *сепарабельном* гильбертовом пространстве, а для описания квантовых процессов необходимы векторы состояния, определённые в более общих *несепарабельных* гильбертовых пространствах. Картина Гейзенберга, в силу независимости векторов состояния от пространственно-временных координат, включает в себя несепарабельные пространства. Дирак

¹Против такой точки зрения в своё время возражал Дж. Чью [1], утверждая, что пространство-время имеет макроскопическую природу и по этой причине не является пригодным для описания микромира.

пишет: «Все это заставляет думать, что картина Гейзенберга хорошая картина, а картина Шрёдингера плохая и что обе картины неэквивалентны» [2, с. 14]. И далее: «Как только вы видите ссылку на картину Шрёдингера, сейчас же выбросьте её вон. При этом вы обнаружите, что без этих ссылок зачастую можно довольно хорошо обойтись и в результате такого выбрасывания вся теория становится более логичной и более вразумительной. Я сказал бы, что таким образом из обычной трактовки квантовой теории поля удаляется значительная часть хлама. То, что при этом остаётся, относится исключительно к гейзенберговской картине и составляет существо теории, и на нем нам следует сосредоточить наше внимание» [2, с. 15]. По сути, указанная Дираком неэквивалентность картин Гейзенберга и Шрёдингера, явилась первым строгим математическим доказательством существования двухуровневой структуры реальности, на что ранее указывали Гейзенберг и Фока (концепция Гейзенберга-Фока).

В данной статье, следуя концепции Гейзенберга-Фока, элементарная частица определяется как *несепарабельное состояние* в спин-зарядовом гильбертовом пространстве $\mathbf{H}^S \otimes \mathbf{H}^Q \otimes \mathbf{H}_\infty$. Пространство $\mathbf{H}^S \otimes \mathbf{H}^Q \otimes \mathbf{H}_\infty$ позволяет учитывать *спин*, *заряд* и *массу* элементарной частицы, при этом все эти три характеристики являются атрибутами *спинтензорного субстрата* (спинорной структуры), ассоциированного с каждым вектором гильбертова пространства $\mathbf{H}^S \otimes \mathbf{H}^Q \otimes \mathbf{H}_\infty$. Зарядовое сопряжение C интерпретируется как псевдоавтоморфизм спинорной структуры. Энергетические (массовые) уровни, задающие актуализированные состояния частиц, описываются в рамках собственных подпространств оператора энергии (комплексного момента).

2. Комплексный момент

Как известно, универсальное накрытие собственной группы Лоренца $SO_0(1, 3)$ (группа вращений четырёхмерного псевдоевклидова пространства $\mathbb{R}^{1,3}$) задаётся спинорной группой

$$\mathbf{Spin}_+(1, 3) \simeq \left\{ \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \mathbb{C}_2 : \det \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = 1 \right\} = \mathrm{SL}(2, \mathbb{C}).$$

Пусть $\mathfrak{g} \rightarrow T_{\mathfrak{g}}$ — произвольное линейное представление собственной группы Лоренца $SO_0(1, 3)$ и пусть $A_i(t) = T_{a_i(t)}$ — инфинитезимальный оператор, соответствующий вращению $a_i(t) \in SO_0(1, 3)$. Аналогично, пусть $B_i(t) = T_{b_i(t)}$, где $b_i(t) \in SO_0(1, 3)$ — гиперболическое вращение. Элементы A_i и B_i образуют

базис групповой алгебры $\mathfrak{sl}(2, \mathbb{C})$ и удовлетворяют соотношениям

$$\left. \begin{aligned} [A_1, A_2] &= A_3, & [A_2, A_3] &= A_1, & [A_3, A_1] &= A_2, \\ [B_1, B_2] &= -A_3, & [B_2, B_3] &= -A_1, & [B_3, B_1] &= -A_2, \\ [A_1, B_1] &= 0, & [A_2, B_2] &= 0, & [A_3, B_3] &= 0, \\ [A_1, B_2] &= B_3, & [A_1, B_3] &= -B_2, \\ [A_2, B_3] &= B_1, & [A_2, B_1] &= -B_3, \\ [A_3, B_1] &= B_2, & [A_3, B_2] &= -B_1. \end{aligned} \right\} \quad (1)$$

Перейдём к комплексной оболочке групповой алгебры $\mathfrak{sl}(2, \mathbb{C})$, определяя операторы

$$X_l = \frac{1}{2}i(A_l + iB_l), \quad Y_l = \frac{1}{2}i(A_l - iB_l), \quad (2)$$

$$(l = 1, 2, 3).$$

Используя соотношения (1), находим

$$[X_k, X_l] = i\varepsilon_{klm}X_m, \quad [Y_l, Y_m] = i\varepsilon_{lmn}Y_n, \quad [X_l, Y_m] = 0. \quad (3)$$

Из соотношений (3) следует, что каждое из множеств операторов X и Y генерирует группу $SU(2)$ и эти две группы коммутируют между собой. Таким образом, из соотношений (3) следует, что группа $SL(2, \mathbb{C})$, по существу, эквивалентна группе $SU(2) \otimes SU(2)$ (данный изоморфизм, безусловно, имеет локальный характер, т. е. вблизи единицы группы, поскольку $SL(2, \mathbb{C})$ некомпактна).

Как известно, наиболее важной *наблюдаемой* в квантовой механике является *энергия*. Пусть имеется эрмитов оператор энергии H^2 , определённый на гильбертовом пространстве H_∞^3 . Тогда все возможные значения энергии являются собственными значениями оператора H . При этом, если $E' \neq E''$ — собственные значения H , а $|\psi'\rangle, |\psi''\rangle$ — принадлежащие им собственные векторы в пространстве H_∞ , то $\langle \psi' | \psi'' \rangle = 0$. Все собственные векторы, принадлежащие данному собственному значению E , образуют вместе с нулевым вектором *собственное подпространство* H_E гильбертова пространства H_∞ . Все собственные подпространства $H_E \in H_\infty$ конечномерны. Размерность H_E называется *кратностью* собственного значения E , если эта размерность r больше единицы, собственное значение E является *r-кратно вырожденным*. Как известно [3], оператор энергии H перестановочен со всеми операторами в H_∞ , представляющими алгебру Ли группы $SU(2)$. Следовательно, H перестановочен также со всеми операторами в H_∞ , представляющими групповую алгебру группы $SU(2) \otimes SU(2)$.

²Спектр этого оператора образует некоторое множество действительных чисел.

³Будем предполагать, что это пространство является сепарабельным, т.е. обладает счётной базой.

Рассмотрим произвольное собственное подпространство H_E оператора энергии H . Поскольку операторы X_l, Y_l и H коммутируют между собой, то, как известно [2], для этих операторов можно построить общую систему собственных функций. Это значит, что подпространство H_E инвариантно относительно операторов X_l, Y_l ⁴. Каждый из операторов X_l, Y_l , например, X_3 (или Y_3) есть эрмитов оператор на H_E и, следовательно, имеет на H_E r_E независимых собственных векторов.

Далее, предположим, что дано некоторое *локальное представление* группы $SU(2) \otimes SU(2)$ операторами, действующими в H_∞ , т. е. представление, определённое в некоторой окрестности единицы группы. Потребуем, чтобы все представляющие операторы были перестановочны с H . Тогда каждое собственное подпространство H_E оператора энергии инвариантно относительно операторов комплексного момента X_l, Y_l .

Далее, вводя операторы вида («повышающие» и «понижающие» операторы группы $SL(2, \mathbb{C})$)

$$\left. \begin{aligned} X_+ &= X_1 + iX_2, & X_- &= X_1 - iX_2, \\ Y_+ &= Y_1 + iY_2, & Y_- &= Y_1 - iY_2, \end{aligned} \right\} \quad (4)$$

видим, что

$$\begin{aligned} [X_3, X_+] &= X_+, & [X_3, X_-] &= -X_-, & [X_+, X_-] &= 2X_3, \\ [Y_3, Y_+] &= Y_+, & [Y_3, Y_-] &= -Y_-, & [Y_+, Y_-] &= 2Y_3. \end{aligned}$$

В силу коммутативности соотношений (3) пространство неприводимого конечномерного представления группы $SL(2, \mathbb{C})$ может быть натянуто на совокупность $(2l+1)(2\dot{l}+1)$ базисных кет-векторов $|l, m; \dot{l}, \dot{m}\rangle$ и базисных бра-векторов $\langle l, m; \dot{l}, \dot{m}|$, где l, m, \dot{l}, \dot{m} — целые или полуцелые числа, $-l \leq m \leq l$, $-\dot{l} \leq \dot{m} \leq \dot{l}$. Следовательно,

$$\begin{aligned} X_- |l, m; \dot{l}, \dot{m}\rangle &= \sqrt{(l+m)(l-m+1)} |l, m-1; \dot{l}, \dot{m}\rangle \quad (m > -l), \\ X_+ |l, m; \dot{l}, \dot{m}\rangle &= \sqrt{(l-m)(l+m+1)} |l, m+1; \dot{l}, \dot{m}\rangle \quad (m < l), \\ X_3 |l, m; \dot{l}, \dot{m}\rangle &= m |l, m; \dot{l}, \dot{m}\rangle, \\ \langle l, m; \dot{l}, \dot{m}| Y_- &= \langle l, m; \dot{l}, \dot{m}-1| \sqrt{(\dot{l}+\dot{m})(\dot{l}-\dot{m}+1)} \quad (\dot{m} > -\dot{l}), \\ \langle l, m; \dot{l}, \dot{m}| Y_+ &= \langle l, m; \dot{l}, \dot{m}+1| \sqrt{(\dot{l}-\dot{m})(\dot{l}+\dot{m}+1)} \quad (\dot{m} < \dot{l}), \\ \langle l, m; \dot{l}, \dot{m}| Y_3 &= \langle l, m; \dot{l}, \dot{m}| \dot{m}. \end{aligned} \quad (5)$$

Таким образом, комплексная оболочка групповой алгебры $\mathfrak{sl}(2, \mathbb{C})$, порождая комплексный угловой момент, приводит к *двойственности*, отражающейся в появлении двух пространств: пространства кет-векторов $|l, m; \dot{l}, \dot{m}\rangle$ и дуального к нему пространства бра-векторов $\langle l, m; \dot{l}, \dot{m}|$.

⁴Более того, операторы X_l, Y_l можно рассматривать *только* на H_E .

2.1. Спинорная структура

Как известно [4, 5], спинтензорные представления группы $SL(2; \mathbb{C}) \simeq \mathbf{Spin}_+(1, 3)$ образуют основу всех конечномерных представлений группы Лоренца. Рассмотрим связь спинтензорных представлений с комплексными алгебрами Клиффорда. С каждой комплексной алгеброй Клиффорда $\mathbb{C}_n = \mathbb{C} \otimes \mathcal{A}_{p,q}$ ($n = p + q$) ассоциировано спинпространство $\mathbb{S}_{2^{n/2}}$, которое является комплексификацией минимального левого идеала вещественной подалгебры $\mathcal{A}_{p,q}$: $\mathbb{S}_{2^{n/2}} = \mathbb{C} \otimes I_{p,q} = \mathbb{C} \otimes \mathcal{A}_{p,q} f_{pq}$, где f_{pq} — примитивный идемпотент подалгебры $\mathcal{A}_{p,q}$. Далее, спинпространство, соответствующее бикватернионной алгебре \mathbb{C}_2 , имеет вид $\mathbb{S}_2 = \mathbb{C} \otimes I_{2,0} = \mathbb{C} \otimes \mathcal{A}_{2,0} f_{20}$ или $\mathbb{S}_2 = \mathbb{C} \otimes I_{1,1} = \mathbb{C} \otimes \mathcal{A}_{1,1} f_{11}$ ($\mathbb{C} \otimes I_{0,2} = \mathbb{C} \otimes \mathcal{A}_{0,2} f_{02}$). Следовательно, тензорное произведение k алгебр \mathbb{C}_2 индуцирует тензорное произведение k спинпространств \mathbb{S}_2 : $\mathbb{S}_2 \otimes \mathbb{S}_2 \otimes \dots \otimes \mathbb{S}_2 = \mathbb{S}_{2^k}$. Векторы спинпространства \mathbb{S}_{2^k} (или элементы минимального левого идеала алгебры \mathbb{C}_{2^k}) являются спинтензорами следующего вида:

$$\mathbf{s}^{\alpha_1 \alpha_2 \dots \alpha_k} = \sum \mathbf{s}^{\alpha_1} \otimes \mathbf{s}^{\alpha_2} \otimes \dots \otimes \mathbf{s}^{\alpha_k}, \tag{6}$$

где суммирование производится по всем наборам индексов $(\alpha_1 \dots \alpha_k)$, $\alpha_i = 1, 2$.

Далее, пусть $\overset{*}{\mathbb{C}}_2$ — алгебра бикватернионов, коэффициенты которой комплексно сопряжены коэффициентам алгебры \mathbb{C}_2 . Тензорное произведение $\overset{*}{\mathbb{C}}_2 \otimes \overset{*}{\mathbb{C}}_2 \otimes \dots \otimes \overset{*}{\mathbb{C}}_2 \simeq \overset{*}{\mathbb{C}}_{2^r}$ r алгебр $\overset{*}{\mathbb{C}}_2$ индуцирует тензорное произведение r спинпространств $\overset{*}{\mathbb{S}}_2$: $\overset{*}{\mathbb{S}}_2 \otimes \overset{*}{\mathbb{S}}_2 \otimes \dots \otimes \overset{*}{\mathbb{S}}_2 = \overset{*}{\mathbb{S}}_{2^r}$. Векторы спинпространства $\overset{*}{\mathbb{S}}_{2^r}$ имеют вид

$$\mathbf{s}^{\dot{\alpha}_1 \dot{\alpha}_2 \dots \dot{\alpha}_r} = \sum \mathbf{s}^{\dot{\alpha}_1} \otimes \mathbf{s}^{\dot{\alpha}_2} \otimes \dots \otimes \mathbf{s}^{\dot{\alpha}_r}. \tag{7}$$

В общем случае имеем тензорное произведение k алгебр \mathbb{C}_2 и r алгебр $\overset{*}{\mathbb{C}}_2$:

$$\underbrace{\mathbb{C}_2 \otimes \mathbb{C}_2 \otimes \dots \otimes \mathbb{C}_2}_{k \text{ раз}} \otimes \underbrace{\overset{*}{\mathbb{C}}_2 \otimes \overset{*}{\mathbb{C}}_2 \otimes \dots \otimes \overset{*}{\mathbb{C}}_2}_{r \text{ раз}} \simeq \mathbb{C}_{2^k} \otimes \overset{*}{\mathbb{C}}_{2^r}, \tag{8}$$

которое индуцирует спинпространство

$$\underbrace{\mathbb{S}_2 \otimes \mathbb{S}_2 \otimes \dots \otimes \mathbb{S}_2}_{k \text{ раз}} \otimes \underbrace{\overset{*}{\mathbb{S}}_2 \otimes \overset{*}{\mathbb{S}}_2 \otimes \dots \otimes \overset{*}{\mathbb{S}}_2}_{r \text{ раз}} = \mathbb{S}_{2^{k+r}} \tag{9}$$

с векторами

$$\mathbf{S} = \mathbf{s}^{\alpha_1 \alpha_2 \dots \alpha_k \dot{\alpha}_1 \dot{\alpha}_2 \dots \dot{\alpha}_r} = \sum \mathbf{s}^{\alpha_1} \otimes \mathbf{s}^{\alpha_2} \otimes \dots \otimes \mathbf{s}^{\alpha_k} \otimes \mathbf{s}^{\dot{\alpha}_1} \otimes \mathbf{s}^{\dot{\alpha}_2} \otimes \dots \otimes \mathbf{s}^{\dot{\alpha}_r}. \tag{10}$$

Для каждого $A \in SL(2, \mathbb{C})$ определим линейное преобразование спинтензора \mathbf{s} посредством формулы

$$\mathbf{s}^{\alpha_1 \alpha_2 \dots \alpha_k \dot{\alpha}_1 \dot{\alpha}_2 \dots \dot{\alpha}_r} \longrightarrow \sum_{(\beta)(\dot{\beta})} A^{\alpha_1 \beta_1} A^{\alpha_2 \beta_2} \dots A^{\alpha_k \beta_k} \overline{A}^{\dot{\alpha}_1 \dot{\beta}_1} \overline{A}^{\dot{\alpha}_2 \dot{\beta}_2} \dots \overline{A}^{\dot{\alpha}_r \dot{\beta}_r} \mathbf{s}^{\beta_1 \beta_2 \dots \beta_k \dot{\beta}_1 \dot{\beta}_2 \dots \dot{\beta}_r}, \tag{11}$$

где символы (β) и $(\dot{\beta})$ означают $\beta_1, \beta_2, \dots, \beta_k$ и $\dot{\beta}_1, \dot{\beta}_2, \dots, \dot{\beta}_r$. Спинтензорные представления группы $SL(2, \mathbb{C})$, определяемые формулой (11), действуют в спинпространстве $\mathbb{S}_{2^{k+r}}$ размерности 2^{k+r} . Как правило, каждое из этих представлений редуцируется на симметричную и антисимметричную части. Выделим подпространства $\text{Sym}_{(k,r)} \subset \mathbb{S}_{2^{k+r}}$ симметрических спинтензоров. Представления группы $SL(2, \mathbb{C})$ в пространствах $\text{Sym}_{(k,r)}$ образуют *полную* систему неприводимых представлений этой группы. Пространство $\text{Sym}_{(k,r)}$ симметрических спинтензоров имеет размерность

$$\dim \text{Sym}_{(k,r)} = (k+1)(r+1). \quad (12)$$

Размерность пространства $\text{Sym}_{(k,r)}$ называется *степенью* представления τ_{li} группы $SL(2, \mathbb{C})$. Легко видеть, что группа $SL(2, \mathbb{C})$ имеет представления *любой степени*. Это представление группы $SL(2, \mathbb{C})$ обозначим как $\tau_{\frac{k}{2}, \frac{r}{2}} = \tau_{li}$. Произведения (8) и (9) определяют *комплексную спинорную структуру*. Представления группы $SL(2, \mathbb{C})$, генерируемые в рамках этой структуры, также являются комплексными. Каждое *неприводимое* конечномерное представление группы $SL(2, \mathbb{C})$ эквивалентно одному из представлений $\tau_{k/2, r/2}$.

2.2. Уравнения движения

В 1945 г. Баба [6] ввёл релятивистские волновые уравнения

$$i\Gamma_{\mu} \frac{\partial \psi}{\partial x_{\mu}} + m\psi = 0, \quad \mu = 0, 1, 2, 3, \quad (13)$$

описывающие системы с многими массами и спинами⁵. В свою очередь, уравнения движения для произвольных спиновых цепочек (спиновых мультиплетов)

$$\tau_{li}, \tau_{l+\frac{1}{2}, i-\frac{1}{2}}, \tau_{l+1, i-1}, \tau_{l+\frac{3}{2}, i-\frac{3}{2}}, \dots, \tau_{li}$$

в бивекторном пространстве \mathbb{R}^6 имеют вид [7, 8]

$$\begin{aligned} \sum_{j=1}^3 \Lambda_j^{ii} \frac{\partial \psi}{\partial a_j} - i \sum_{j=1}^3 \Lambda_j^{ii} \frac{\partial \psi}{\partial a_j^*} + m^{(s)} \psi &= 0, \\ \sum_{j=1}^3 \Lambda_j^{l+\frac{1}{2}, i-\frac{1}{2}} \frac{\partial \psi}{\partial a_j} - i \sum_{j=1}^3 \Lambda_j^{l+\frac{1}{2}, i-\frac{1}{2}} \frac{\partial \psi}{\partial a_j^*} + m^{(s)} \psi &= 0, \\ \sum_{j=1}^3 \Lambda_j^{l+1, i-1} \frac{\partial \psi}{\partial a_j} - i \sum_{j=1}^3 \Lambda_j^{l+1, i-1} \frac{\partial \psi}{\partial a_j^*} + m^{(s)} \psi &= 0, \end{aligned}$$

⁵Уравнения Бабы определены в пространстве-времени Минковского $\mathbb{R}^{1,3}$. Эти уравнения представляют собой (как и вся теория релятивистских волновых уравнений) попытку описания элементарных частиц в терминах пространства-времени, т. е. в рамках картины Шрёдингера.

$$\sum_{j=1}^3 \Lambda_j^{l+\frac{3}{2}, i-\frac{3}{2}} \frac{\partial \psi}{\partial a_j} - i \sum_{j=1}^3 \Lambda_j^{l+\frac{3}{2}, i-\frac{3}{2}} \frac{\partial \psi}{\partial a_j^*} + m^{(s)} \dot{\psi} = 0,$$

.....

$$\sum_{j=1}^3 \Lambda_j^{li} \frac{\partial \dot{\psi}}{\partial a_j} + i \sum_{j=1}^3 \Lambda_j^{li} \frac{\partial \dot{\psi}}{\partial a_j^*} + m^{(s)} \psi = 0, \tag{14}$$

где спин $s = l - \dot{l}$ изменяется следующим образом:

$$l - \dot{l}, l - \dot{l} + 1, l - \dot{l} + 2, l - \dot{l} + 3, \dots, \dot{l} - l.$$

Система (14) описывает состояния частиц с различными массами и спинами. Масса состояния $m^{(s)}$, соответствующая уровню энергии $H_E \simeq \text{Sym}_{(k,r)}$, определяется формулой

$$m^{(s)} = \mu^0 \left(l + \frac{1}{2} \right) \left(\dot{l} + \frac{1}{2} \right), \tag{15}$$

где $s = |l - \dot{l}|$. Явный вид элементов матриц Λ_i^{li} находится посредством вычисления коммутаторов

$$\begin{aligned} \left[\left[\Lambda_3^{li}, X_-^{li} \right], X_+^{li} \right] &= 2\Lambda_3^{li}, & \left[\left[\Lambda_3^{li}, Y_-^{li} \right], Y_+^{li} \right] &= 2\Lambda_3^{li}, \\ \left[A_2^{li}, \Lambda_3^{li} \right] &= \Lambda_1^{li}, & \left[A_1^{li}, \Lambda_3^{li} \right] &= -\Lambda_2^{li} \end{aligned}$$

относительно кет- и бра-векторов спирального базиса (более подробно см. [8]), здесь $A_i^{li} = A_i^l \otimes \mathbf{1}_{2i+1} - \mathbf{1}_{2i+1} \otimes A_i^{\dot{l}}$.

3. Концепция Гейзенберга-Фока

Как известно, копенгагенская интерпретация квантовой механики утверждает, что не стоит искать более глубокого описания и понимания реальности, данной нам в эксперименте. Только феномены являются реально существующими, и помимо них нет никакой более глубокой реальности. Согласно Гейзенбергу [9], за квантовым феноменом стоит более глубокий уровень реальности — бытие в возможности, что приводит к двухуровневой, двухмодусной онтологической картине: имеется модус бытия в возможности и модус бытия действительного, мир фактически существующего⁶. Гейзенберг в своих

⁶Гейзенберг отмечал, что это возвращает нас к философии Аристотеля, согласно которой реальность имеет двухуровневую структуру: *δυναμις* (динамис) — бытие в возможности (потенция) и *εντελεχεια* (энтелехия) — бытие действительного (проявленный мир). Согласно А.Ю. Севальникову [10], это приводит к изначальной «расщеплённости», многомодусности или *полионтичности бытия*. Ту же двухуровневую структуру реальности мы видим у Зурека [11] в теории декогеренции, согласно которой в основании реальности находится нелокальный квантовый субстрат (*квантовый домен*), а весь видимый мир (*классический домен*) возникает из квантового домена в результате процесса декогеренции. Двухмодусность бытия является ключевым моментом в теории физических структур Ю.И. Кулакова [12] и реляционной теории пространства-времени Ю.С. Владимирова [13].

философских работах неоднократно подчёркивал необходимость использования в квантовотеоретической онтологии аристотелевского понятия «бытие в возможности», однако, сам он не развил последовательным образом эти мысли. По сути к тем же идеям независимо пришёл В.А. Фок, но значительно более последовательным образом. Фок подчёркивает объективный характер существования потенциальных возможностей: «Описываемое волновой функцией состояние объекта является объективным в том смысле, что оно представляет объективную (**независимую от наблюдателя**) характеристику потенциальных возможностей того или иного результата взаимодействия атомного объекта с прибором. В том же смысле оно относится именно к данному, единичному объекту. Но это объективное состояние не является ещё действительным, в том смысле, что для объекта в данном состоянии указанные потенциальные возможности ещё не осуществились. Переход от потенциальной возможности к осуществившемуся происходит в заключительной стадии эксперимента» [14, с. 13]. Фок настаивает на объективном существовании потенциальной возможности, причём чётко указывает (что особенно важно!) на её соответствие с **единичным квантовым объектом** (например, с элементарной частицей).

4. Гильбертово пространство $\mathbf{H}_{2s+1}^S \otimes \mathbf{H}_\infty$

Согласно концепции Гейзенберга-Фока, описывающей двухмодусную структуру реальности, модус потенциального соотнесён с единичным квантовым объектом (в нашем случае этим объектом будет элементарная частица). Прежде чем приступить к построению абстрактного гильбертова пространства элементарной частицы, рассмотрим соотношение картин Шрёдингера и Гейзенберга согласно двухмодусной структуре. *Картина Шрёдингера принадлежит модусу актуального*, здесь есть пространство-время и сепарабельные гильбертовы пространства. *Картина Гейзенберга принадлежит модусу потенциального*, здесь уже нет пространства-времени в его обычном понимании, векторы состояния в этой картине не зависят от времени и определены в несепарабельных гильбертовых пространствах⁷.

Согласно фон Нейману [15], «природа» векторов *абстрактного гильбертова пространства* \mathbf{H}_∞ несущественна, существенно лишь выполнение аксиом сходимости, сепарабельности, сложения, умножения и скалярного (внутренне-

⁷Здесь необходимо сделать следующее уточнение. Согласно стандартному пониманию картины Гейзенберга, если векторы состояния этой картины не зависят от времени, то операторы, представляющие наблюдаемые величины, от времени зависят. Поэтому картина Гейзенберга в её стандартном понимании не может быть целиком соотнесена с модусом потенциального. С другой стороны, наблюдаемые величины есть феномены фактически существующего, проявленного мира, т. е. принадлежащими к модусу актуального. По этой причине математика линейных самосопряжённых операторов, представляющих наблюдаемые величины, не имеет смысла для модуса потенци. Эта ситуация, в некотором смысле, близка к аналитической теории S -матрицы. Вот что пишет Чью: «Теория S -матрицы не использует всего аппарата квантовой механики: в ней сохраняется только принцип суперпозиции. В теории нет ни гамильтониана, ни каких-либо других операторов, нет также изменяющихся во времени векторов состояния» [1, с. 17].

го) произведения.

Пусть $|A\rangle, |B\rangle, |C\rangle, \dots$ — векторы пространства \mathbf{H}_∞ , удовлетворяющие следующим аксиомам.

I. Сложение.

$$|A\rangle + |B\rangle = |B\rangle + |A\rangle, \quad (|A\rangle + |B\rangle) + |C\rangle = |A\rangle + (|B\rangle + |C\rangle).$$

II. Умножение.

$$c_1 (|A\rangle + |B\rangle) = c_1 |A\rangle + c_1 |B\rangle, \quad (c_1 + c_2) |A\rangle = c_1 |A\rangle + c_2 |A\rangle,$$

$$(c_1 c_2) |A\rangle = c_1 (c_2 |A\rangle), \quad 0 \cdot |A\rangle = 0, \quad 1 \cdot |A\rangle = |A\rangle.$$

III. Внутреннее произведение.

$$\langle (c_1 A + c_2 B) | C \rangle = \bar{c}_1 \langle A | C \rangle + \bar{c}_2 \langle B | C \rangle, \quad \langle C | (c_1 |A\rangle + c_2 |B\rangle) \rangle = c_1 \langle C | A \rangle + c_2 \langle C | B \rangle.$$

Всегда имеет место $\langle A | A \rangle \geq 0$. При этом, если $|A\rangle \neq 0$, то $\langle A | A \rangle > 0$. Длина вектора $|A\rangle$ определяется формулой $\| |A\rangle \| = \sqrt{\langle A | A \rangle}$.

IV. Условие сепарабельности⁸. В \mathbf{H}_∞ существует такая бесконечная последовательность базисных векторов $\xi^1, \xi^2, \dots, \xi^n, \dots$, что для любого вектора $|A\rangle$ существует однозначное разложение

$$|A\rangle = \sum_{i=1}^{\infty} |\xi^i\rangle \langle \xi^i | A \rangle,$$

где суммирование ряда понимается в смысле

$$\lim_{n \rightarrow \infty} \left| |A\rangle - \sum_{i=1}^n |\xi^i\rangle \langle \xi^i | A \rangle \right| = 0.$$

V. Сходимость. Пусть

$$|A_m\rangle = \sum_{i=1}^m |\xi^i\rangle \langle \xi^i | A \rangle, \quad |A_n\rangle = \sum_{i=1}^n |\xi^i\rangle \langle \xi^i | A \rangle.$$

Если

$$\lim_{m, n \rightarrow \infty} ||A_m\rangle - |A_n\rangle| = 0,$$

то существует вектор $|A_0\rangle$, к которому $|A_n\rangle$ сходятся:

$$\lim_{n \rightarrow \infty} ||A_n\rangle - |A_0\rangle| = 0.$$

Как известно, функции вероятности⁹ могут быть интерпретированы как векторы пространства \mathbf{H}_∞ (эта интерпретация возможна при выполнении аксиом

⁸При выполнении аксиомы IV имеем сепарабельное гильбертово пространство. В случае отсутствия аксиомы IV приходим к несепарабельному гильбертову пространству, которое и будем использовать в дальнейшем.

⁹Здесь мы следуем терминологии Гейзенберга, который избегал использовать термин волновая функция применительно к описанию квантового объекта, считая, что этот термин является рудиментом картины Шрёдингера, т. е. затемняющим существо дела.

I–III). Следуя обозначениям Дирака [2], запишем функции вероятности как $|\psi\rangle = |A\rangle$, где $|A\rangle \in \mathbf{H}_\infty$. Векторы $|A\rangle$ могут представлять тензоры, функции на группе, представления и т. д. Мы возьмём $|A\rangle$ как функции на группе Лоренца, где каждый вектор $|A\rangle \in \mathbf{H}_\infty$ соответствует некоторому представлению группы $SL(2, \mathbb{C})$. В данном контексте условие фон Неймана принимает вид

$$\int_{SL(2, \mathbb{C})} \langle \psi | U(g) | \psi \rangle dg < \infty, \quad (16)$$

где dg — мера Хаара на группе $SL(2, \mathbb{C})$, $U(g)$ — представление группы $SL(2, \mathbb{C})$ ¹⁰.

4.1. Спиновые мультиплеты

Перейдём к дальнейшему обобщению абстрактного гильбертова пространства \mathbf{H}_∞ . В 1927 г. Паули [17] ввёл первую теорию спина электрона. Главная идея этой теории состоит в *удвоении* пространства функций вероятности. Пусть $|\psi_1\rangle$ и $|\psi_2\rangle$ — векторы пространства \mathbf{H}_∞ . Тогда удвоенное пространство определяется формальными линейными комбинациями вида

$$c_1 |\psi_1\rangle + c_2 |\psi_2\rangle, \quad (17)$$

где c_1 и c_2 — комплексные коэффициенты. Отсюда следует, что пространство \mathbf{H}_∞ необходимо заменить тензорным произведением

$$\mathbf{H}_2^S \otimes \mathbf{H}_\infty. \quad (18)$$

Далее, первым простейшим спиновым мультиплетом является *спиновый дублет*, конструируемый в рамках пространства $\mathbf{H}_2^S \otimes \mathbf{H}_\infty$ (пространства Паули). Мы имеем здесь два спиновых состояния: одно состояние принадлежит линии спина $1/2$, а другое состояние принадлежит дуальной линии спина $1/2$. Первый спиновый дублет (см. рис. 1) определяется следующей схемой:

$$\begin{array}{ccc} \left(\frac{1}{2}, 0\right) & \text{-----} & \left(0, \frac{1}{2}\right) \\ \bullet & & \bullet \\ \frac{1}{2} & \text{-----} & -\frac{1}{2} \end{array} \quad (19)$$

Здесь вторая строка означает, что представление $(1/2, 0)$ описывает состояние частицы (например, электрона) со значением спина $1/2$. В свою очередь, представление $(0, 1/2)$ описывает состояние частицы (электрона) со значением

¹⁰В рамках метода усовершенствованного алгебраического квантования внутреннее произведение состояний определяется посредством техники группового усреднения. В групповом усреднении используется интеграл $\int_G \langle \phi_1 | U(g) | \phi_2 \rangle dg$ над калибровочной группой G , где dg — так называемая «симметричная» мера Хаара на G , $U(g)$ — представление группы G , ϕ_1 и ϕ_2 — векторы состояния, принадлежащие вспомогательному гильбертову пространству \mathcal{H}_{aux} . Сходящееся групповое усреднение даёт алгоритм для построения полного множества наблюдаемых квантовой системы [16].

спина $-1/2$. Дублет (19) соответствует уравнению Дирака и по этой причине должен быть назван *фундаментальным дублетом*. С другой стороны, мы можем построить спиновый дублет, используя $(1, 1/2)$ - и $(1/2, 1)$ -представления:

$$\begin{array}{ccc} (1, \frac{1}{2}) & \text{-----} & (\frac{1}{2}, 1) \\ \bullet & & \bullet \\ \frac{1}{2} & \text{-----} & -\frac{1}{2} \end{array}$$

Легко видеть, что существует бесконечно много спиновых дублетов, в которых различные спиновые состояния $1/2$ и $-1/2$ принадлежат соответственно линии спина $1/2$ и дуальной линии спина $1/2$:

$$\begin{array}{ccc} (\frac{3}{2}, 1) & \text{-----} & (1, \frac{3}{2}) \\ \bullet & & \bullet \\ \frac{1}{2} & \text{-----} & -\frac{1}{2} \\ \\ (2, \frac{3}{2}) & \text{-----} & (\frac{3}{2}, 2) \\ \bullet & & \bullet \\ \frac{1}{2} & \text{-----} & -\frac{1}{2} \\ \dots & & \dots \\ \\ (\frac{59}{2}, 29) & \text{-----} & (29, \frac{59}{2}) \\ \bullet & & \bullet \\ \frac{1}{2} & \text{-----} & -\frac{1}{2} \\ \dots & & \dots \end{array}$$

Следующим спиновым мультиплетом является *спиновый триплет*, конструируемый в рамках пространства $\mathbf{H}_3^S \otimes \mathbf{H}_\infty$. Здесь имеем три спиновых состояния: два состояния 1 и -1 принадлежат линии спина 1 и дуальной линии спина 1 , а третье спиновое состояние 0 принадлежит линии спина 0 . Согласно рис. 1, первый спиновый триплет (*фундаментальный триплет*) задаётся следующей схемой:

$$\begin{array}{ccccc} (1,0) & \text{-----} & (\frac{1}{2}, \frac{1}{2}) & \text{-----} & (0,1) \\ \bullet & & \bullet & & \bullet \\ 1 & \text{-----} & 0 & \text{-----} & -1 \end{array} \tag{20}$$

Очевидно, что имеется бесконечно много спиновых триплетов. Спиновый триплет, следующий после (20), определяется схемой

$$\begin{array}{ccccc} (\frac{3}{2}, \frac{1}{2}) & \text{-----} & (1,1) & \text{-----} & (\frac{1}{2}, \frac{3}{2}) \\ \bullet & & \bullet & & \bullet \\ 1 & \text{-----} & 0 & \text{-----} & -1 \end{array}$$

и т. д.

Подобным образом строятся спиновые квадруплеты и другие спиновые мультиплеты. Например, первый шестимерный мультиплет (6-плет), определяемый в $\mathbf{H}_6^S \otimes \mathbf{H}_\infty$, задаётся схемой

$$\begin{array}{ccccccccc} (\frac{5}{2}, 0) & \text{-----} & (2, \frac{1}{2}) & \text{-----} & (\frac{3}{2}, 1) & \text{-----} & (1, \frac{3}{2}) & \text{-----} & (\frac{1}{2}, 2) & \text{-----} & (0, \frac{5}{2}) \\ \bullet & & \bullet \\ \frac{5}{2} & \text{-----} & \frac{3}{2} & \text{-----} & \frac{1}{2} & \text{-----} & -\frac{1}{2} & \text{-----} & -\frac{3}{2} & \text{-----} & -\frac{5}{2} \end{array}$$

Продолжая это построение, приходим к следующему обобщению абстрактного гильбертова пространства:

$$\mathbf{H}_{2s+1}^S \otimes \mathbf{H}_\infty, \quad (21)$$

где $s = 0, \frac{1}{2}, 1, \frac{3}{2}, 2, \dots$ и $s = |l - \dot{l}|$. Следует отметить, что также имеется бесконечное множество *спиновых синглетов* в $\mathbf{H}_{2s+1}^S \otimes \mathbf{H}_\infty$. Все эти синглеты принадлежат линии спина 0 и определяются представлениями $(0, 0), (1/2, 1/2), (1, 1), \dots, (s, s), \dots$. Синглет, определяемый представлением $(0, 0)$, называется *фундаментальным синглетом*.

Далее, при нечётном s имеем *фермионные мультиплеты* в $\mathbf{H}_{2s+1}^S \otimes \mathbf{H}_\infty$ и, соответственно, *бозонные мультиплеты* при s чётном. Все пространства $\mathbf{H}_{2s+1}^S \otimes \mathbf{H}_\infty$ являются несепарабельными гильбертовыми пространствами.

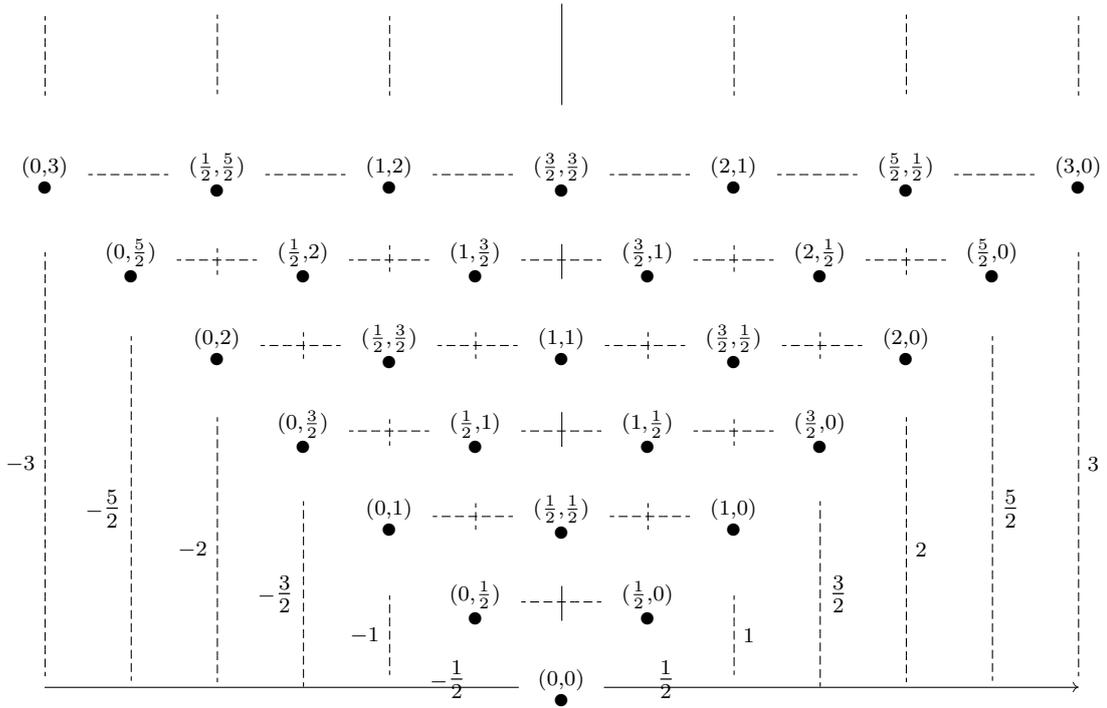


Рис. 1: Собственные подпространства $\mathbf{H}_E \simeq \text{Sym}_{(k,r)}$ оператора энергии H . Значение массы для каждого уровня \mathbf{H}_E даётся формулой (15).

4.2. Первое определение элементарной частицы

Дадим первое определение элементарной частицы¹¹. **Частицей является суперпозиция векторов состояния в несепарабельном гильбертовом про-**

¹¹ Следует отметить, что словосочетание *элементарная частица* является рудиментом редукционистского мировоззрения. Гейзенберг в статье «Что такое элементарная частица?» пишет: «Нам неминуемо приходится пользоваться языком, корнящимся в традиционной философии. Мы спрашиваем: из чего состоит протон? Можно ли разделить электрон, или он неделим?»

пространстве $\mathbf{H}_{2s+1}^S \otimes \mathbf{H}_\infty$. Суперпозиция векторов состояния задаёт квантовую систему U элементарной частицы. Это определение учитывает спин и массу частицы. Сама частица есть несепарабельное состояние в пространстве $\mathbf{H}_{2s+1}^S \otimes \mathbf{H}_\infty$. Из этого определения следует, что частица, согласно концепции Гейзенберга-Фока, есть объект модуса потенциального. Переход с этого уровня на модус актуального осуществляется посредством редукции суперпозиции¹². При редукции суперпозиции актуализируется одно из возможных состояний квантовой системы U , которое описывается в рамках собственного подпространства $\mathbf{H}_E \simeq \text{Sym}_{(k,r)}$ оператора энергии H . В свою очередь на подпространстве $\mathbf{H}_E \simeq \text{Sym}_{(k,r)}$ действует представление τ_{ii} группы $\text{SL}(2, \mathbb{C}) \simeq \mathbf{Spin}_+(1, 3)$ с ассоциированной спинорной структурой, задаваемой произведениями (8) и (9). Таким образом, при редукции суперпозиции имеем $U \rightarrow \tau_{ii}$ и $\mathbf{H}_{2s+1}^S \otimes \mathbf{H}_\infty \rightarrow \mathbf{H}_E \simeq \text{Sym}_{(k,r)}$.

В качестве примера рассмотрим простейший случай, соответствующий фундаментальному дублету (19), т. е. электрону¹³. Итак, электрон является су-

Прост или составен квант света? И так далее. Но все эти вопросы поставлены неправильно, потому что слова «делить» и «состоит из» в значительной мере утратили свой смысл» [9, с. 172].

¹²Как известно, постулат Дирака-фон Неймана о редукции квантовой суперпозиции (так называемый *коллапс волновой функции*) явился объектом многочисленных споров. Утверждалось, что этот постулат противоречит логике квантовой механики и нарушает линейность квантово-механической эволюции. Предпринимались попытки исключить постулат редукции из квантовой механики, например, посредством введения многомировой интерпретации Эверетта, что есть по сути не более чем расщепление картины Шрёдингера (модуса актуального) на множество картин Шрёдингера (эвереттовских миров). В свою очередь против «квантово-механической эволюции», якобы описываемой уравнением Шрёдингера, Дирак [2] выдвинул веские доказательства, заключающиеся в том, что изменяющийся во времени вектор состояния будет выбит из сепарабельного гильбертова пространства за кратчайшее время, т. е. по сути ни о какой эволюции речи быть не может, и решения уравнения Шрёдингера не существует. С другой стороны, предпринимались попытки пространственно-временного описания редукции квантовой суперпозиции, т. е. рассматривалась возможность представить редукцию как процесс, протекающий в пространстве-времени, что, как показало время, не увенчалось успехом. Отсюда с необходимостью следует вывод: природа *коллапса функции вероятности* находится за пределами пространства и времени, т. е. принадлежит модусу потенциального.

¹³Следует отметить интересную аналогию между электроном и кубитом. Как известно, *кубит* есть несепарабельное состояние, заданное суперпозицией $|\psi\rangle = a|0\rangle + b|1\rangle$, где $a, b \in \mathbb{C}$. Кубит есть минимально возможный (элементарный) вектор состояния. Любой вектор состояния может быть представлен множеством таких элементарных векторов, по этой причине кубит является исходным «строительным кирпичом» для всех остальных векторов состояния любой размерности. Квантовое состояние N кубитов может быть описано вектором гильбертова пространства размерности 2^N . Очевидно, что это пространство совпадает со спинпространством \mathbb{S}_{2^N} . Выберем в качестве ортонормированного базиса для этого пространства состояния, в которых каждый кубит имеет определённое значение $|0\rangle$ или $|1\rangle$. Тогда векторы базиса могут быть представлены двоичными строками вида $|01110010 \dots 1001\rangle$. Общий нормированный вектор может быть выражен в этом базисе как $\sum_{x=0}^{2^N-1} a_x |x\rangle$, где a_x — комплексные числа, удовлетворяющие условию $\sum_x |a_x|^2 = 1$. Уже здесь видна глубокая аналогия между кубитами и 2-компонентными спинорами. Подобно кубитам 2-компонентные спиноры являются «строительными кирпичиками» фундаментальной спинорной структуры (посредством тензорных произведений алгебр \mathbb{C}_2 и \mathbb{C}_2^* , см. (8) и (9)). Более того, векторы гильбертова пространства $\mathbf{H}_{2s+1}^S \otimes \mathbf{H}_\infty$, представляющие актуализированные состояния частиц («двоичные строки»), строятся таким же образом.

перпозицией векторов состояния в несепарабельном гильбертовом пространстве $\mathbf{H}_2^S \otimes \mathbf{H}_\infty$. В случае электрона имеем два спиновых состояния: состояние $1/2$, описываемое представлением $\tau_{1/2,0}$ на линии спина $1/2$, и состояние $-1/2$, описываемое представлением $\tau_{0,1/2}$ на дуальной линии спина $1/2$. Представления $\tau_{1/2,0}$ и $\tau_{0,1/2}$ действуют в пространствах $\text{Sym}_{(1,0)}$ и $\text{Sym}_{(0,1)}$, соответственно. Следовательно, имеются два вектора состояния: кет-вектор $|\Psi\rangle = |\tau_{1/2,0}, \text{Sym}_{(1,0)}, \mathbb{C}_2, \mathbb{S}_2, \rangle$ и бра-вектор $\langle \dot{\mathbb{S}}_2, \dot{\mathbb{C}}_2, \text{Sym}_{(0,1)}, \tau_{0,1/2} | = \langle \dot{\Psi} |$. Спинорная структура задаётся прямой суммой $\mathbb{C}_2 \oplus \mathbb{C}_2$. Уравнения движения имеют вид (частный случай системы (14) для спиновой цепочки $\tau_{1/2,0}, \tau_{0,1/2}$)

$$\sum_{j=1}^3 \Lambda_j^{*0, \frac{1}{2}} \frac{\partial \psi}{\partial \tilde{a}_j} + i \sum_{j=1}^3 \Lambda_j^{*0, \frac{1}{2}} \frac{\partial \psi}{\partial \tilde{a}_j^*} + m_e \psi = 0,$$

$$\sum_{j=1}^3 \Lambda_j^{\frac{1}{2}, 0} \frac{\partial \psi}{\partial a_j} - i \sum_{j=1}^3 \Lambda_j^{\frac{1}{2}, 0} \frac{\partial \psi}{\partial a_j^*} + m_e \psi = 0,$$

где

$$\Lambda_1^{\frac{1}{2}, 0} = \frac{1}{2} c_{\frac{1}{2} \frac{1}{2}} \left\| \begin{array}{cc} 0 & 1 \\ 1 & 0 \end{array} \right\|, \quad \Lambda_2^{\frac{1}{2}, 0} = \frac{1}{2} c_{\frac{1}{2} \frac{1}{2}} \left\| \begin{array}{cc} 0 & -i \\ i & 0 \end{array} \right\|, \quad \Lambda_3^{\frac{1}{2}, 0} = \frac{1}{2} c_{\frac{1}{2} \frac{1}{2}} \left\| \begin{array}{cc} 1 & 0 \\ 0 & -1 \end{array} \right\|,$$

$$\Lambda_1^{*0, \frac{1}{2}} = \frac{1}{2} \dot{c}_{\frac{1}{2} \frac{1}{2}} \left\| \begin{array}{cc} 0 & 1 \\ 1 & 0 \end{array} \right\|, \quad \Lambda_2^{*0, \frac{1}{2}} = \frac{1}{2} \dot{c}_{\frac{1}{2} \frac{1}{2}} \left\| \begin{array}{cc} 0 & -i \\ i & 0 \end{array} \right\|, \quad \Lambda_3^{*0, \frac{1}{2}} = \frac{1}{2} \dot{c}_{\frac{1}{2} \frac{1}{2}} \left\| \begin{array}{cc} 1 & 0 \\ 0 & -1 \end{array} \right\|.$$

Легко видеть, что эти матрицы совпадают с матрицами Паули σ_i при $c_{\frac{1}{2} \frac{1}{2}} = 2$.

При редукции имеем $U \rightarrow \tau_{1/2,0}$ (или $U \rightarrow \tau_{0,1/2}$) и $\mathbf{H}_2^S \otimes \mathbf{H}_\infty \rightarrow \mathbf{H}_E \simeq \text{Sym}_{(1,0)}$ (или $\mathbf{H}_2^S \otimes \mathbf{H}_\infty \rightarrow \mathbf{H}_E \simeq \text{Sym}_{(0,1)}$). Собственное подпространство \mathbf{H}_E оператора энергии H представляет собой массовый уровень со значением массы, определяемой формулой (15). Пусть $s = l = 1/2$ и пусть $m_e = \mu^0 (l + \frac{1}{2}) = \mu^0$ — масса электрона. Из массовой формулы (15) прямо следует, что масса электрона является минимальной массой покоя μ^0 .

5. Гильбертово пространство $\mathbf{H}^S \otimes \mathbf{H}^Q \otimes \mathbf{H}_\infty$

Пространство $\mathbf{H}_{2s+1}^S \otimes \mathbf{H}_\infty$ и первое определение элементарной частицы позволяют учитывать только две важнейшие характеристики частицы: спин и массу. Для введения **заряда** (третьей важнейшей характеристики) требуется дальнейшее обобщение абстрактного гильбертова пространства.

5.1. Зарядовые мультиплеты

В 1932 г. Гейзенберг [18] предложил рассматривать протон и нейтрон как два различных состояния одной и той же частицы (нуклона). Теория Гейзенберга протонно-нейтронных состояний (*зарядовый дублет*) формально совпадает

с теорией электронного спина, предложенной Паули. Главным объектом теории Гейзенберга является абстрактное гильбертово пространство следующего типа:

$$\mathbf{H}_2^Q \otimes \mathbf{H}_\infty,$$

где \mathbf{H}_2^Q — зарядовое пространство, ассоциированное с фундаментальным представлением группы $SU(2)$.

В 1938 г. Кеммер [19] расширил теорию Гейзенберга для случая частицы с тремя различными зарядовыми состояниями 1, 0, -1 . Согласно Кеммеру, *зарядовый триплет* строится в рамках абстрактного гильбертова пространства

$$\mathbf{H}_3^Q \otimes \mathbf{H}_\infty,$$

где \mathbf{H}_3^Q — зарядовое пространство, ассоциированное с представлением τ_1 группы $SU(2)$.

5.2. Введение заряда. Псевдоавтоморфизм спинорной структуры

В существующей ныне форме квантовой теории заряженным частицам отводят комплексные поля [20]. Также как и спин, заряд является внутренним понятием, принадлежащим уровню спинорной структуры (спинтензорного субстрата). Прообразом зарядового сопряжения C на этом уровне является псевдоавтоморфизм $\mathcal{A} \rightarrow \overline{\mathcal{A}}$ спинорной структуры [21, 22, 25, 26]. Тензорный субстрат (8) соответствует *комплексной спинорной структуре*. В согласии с [21, теорема 1] алгебры Клиффорда над полем $\mathbb{F} = \mathbb{C}$ соответствуют **заряженным частицам**, таким как электрон, протон и т. д. Переходя к *вещественной спинорной структуре*, которая, очевидно, является подструктурой субстрата (8), мы видим, что преобразование C (псевдоавтоморфизм $\mathcal{A} \rightarrow \overline{\mathcal{A}}$) для алгебр $\mathcal{C}_{p,q}$ над полем $\mathbb{F} = \mathbb{R}$ и кольцом $\mathbb{K} \simeq \mathbb{H}$ (типы $p - q \equiv 4, 6 \pmod{8}$) редуцируется к **обмену частица-античастица** C' (см. [21, 22]). Как известно, нейтральные частицы описываются в рамках вещественных представлений группы Лоренца. Существуют два класса нейтральных частиц: 1) частицы, имеющие античастицы, такие как нейтроны, нейтрино и т. д.; 2) частицы, совпадающие со своими античастицами (например, фотоны, K^0 -мезоны и т. д.). Первый класс описывается алгебрами $\mathcal{C}_{p,q}$ над полем $\mathbb{F} = \mathbb{R}$ с кольцами $\mathbb{K} \simeq \mathbb{H}$ и $\mathbb{K} \simeq \mathbb{H} \oplus \mathbb{H}$ (типы $p - q \equiv 4, 6 \pmod{8}$ и $p - q \equiv 5 \pmod{8}$), а второй класс (**истинно нейтральные частицы**) описывается алгебрами $\mathcal{C}_{p,q}$ над полем $\mathbb{F} = \mathbb{R}$ с кольцами $\mathbb{K} \simeq \mathbb{R}$ и $\mathbb{K} \simeq \mathbb{R} \oplus \mathbb{R}$ (типы $p - q \equiv 0, 2 \pmod{8}$ и $p - q \equiv 1 \pmod{8}$) (см. [21–26]).

5.3. Второе определение элементарной частицы

Пространства $\mathbf{H}_{2s+1}^S \otimes \mathbf{H}_\infty$ (спиновые мультиплеты) и $\mathbf{H}^Q \otimes \mathbf{H}_\infty$ (зарядовые мультиплеты), рассмотренные в предыдущих параграфах, естественным образом приводят к следующему обобщению абстрактного гильбертова пространства. Пусть

$$\mathbf{H}^S \otimes \mathbf{H}^Q \otimes \mathbf{H}_\infty \tag{22}$$

— тензорное произведение гильбертова пространства \mathbf{H}_∞ и спин-зарядового пространства $\mathbf{H}^S \otimes \mathbf{H}^Q$. Векторы состояния пространства (22) описывают актуализированные состояния частиц спина $s = |l - \bar{l}|$ и заряда Q с массой m , определяемой формулой (15). Более того, векторы состояния группируются в спиновые линии: линия спина 0, линия спина 1/2, линия спина 1 и т. д. (бозонные и фермионные линии). Каждый вектор состояния представляет собой неприводимое представление τ_{ii} группы $\mathbf{Spin}_+(1, 3)$ в пространстве $\text{Sym}_{(k,r)} \simeq \mathbf{H}_E$ (собственное подпространство оператора энергии H). Заряд Q принимает три значения $-1, 0, +1$ ¹⁴, где значения $-1, +1$ соответствуют актуализированным состояниям заряженных частиц, а значение 0 соответствует нейтральным (или истинно нейтральным) состояниям частиц. На уровне спинорной структуры заряженные частицы описываются комплексными представлениями группы $\mathbf{Spin}_+(1, 3)$, для которых псевдоавтоморфизм $\mathcal{A} \rightarrow \overline{\mathcal{A}}$ нетривиален ($\mathbb{F} = \mathbb{C}$). Действие псевдоавтоморфизма $\mathcal{A} \rightarrow \overline{\mathcal{A}}$ заменяет комплексные представления (зарядовое состояние -1) на комплексно сопряжённые представления (зарядовое состояние $+1$). Нейтральные частицы (зарядовое состояние 0) описываются вещественными представлениями группы $\mathbf{Spin}_+(1, 3)$, для которых преобразование $\mathcal{A} \rightarrow \overline{\mathcal{A}}$ также нетривиально ($\mathbb{F} = \mathbb{R}, \mathbb{K} \simeq \mathbb{H}$), т. е. мы имеем здесь обмен частица-античастица. В свою очередь истинно нейтральные частицы описываются вещественными представлениями группы $\mathbf{Spin}_+(1, 3)$, для которых действие псевдоавтоморфизма $\mathcal{A} \rightarrow \overline{\mathcal{A}}$ является тривиальным ($\mathbb{F} = \mathbb{R}, \mathbb{K} \simeq \mathbb{R}$). Чтобы отличить этот случай от нейтральных частиц (состояние 0), обозначим это зарядовое состояние через $\bar{0}$. Следовательно, спинорная структура с помощью псевдоавтоморфизма $\mathcal{A} \rightarrow \overline{\mathcal{A}}$ позволяет разделить вещественные представления для нейтральных (зарядовое состояние 0) и истинно нейтральных (зарядовое состояние $\bar{0}$) частиц.

Учёт заряда позволяет дать второе определение элементарной частицы. **Частицей является суперпозиция векторов состояния в несепарабельном гильбертовом пространстве $\mathbf{H}^S \otimes \mathbf{H}^Q \otimes \mathbf{H}_\infty$** ¹⁵. Векторы пространства $\mathbf{H}^S \otimes \mathbf{H}^Q \otimes \mathbf{H}_\infty$ имеют вид

$$|A\rangle = |\tau_{ii}, \text{Sym}_{(k,r)}, \mathcal{O}_{p,q}, \mathbb{S}_{(p+q)/2}, C^{a,b,c,d,e,f,g}, \dots\rangle, \quad (23)$$

где τ_{ii} — представление собственной группы Лоренца, $\text{Sym}_{(k,r)}$ — пространство

¹⁴В случае зарядового квадруплета (например, Δ -квадруплет спина 3/2) имеем четыре значения $-1, 0, 1, 2$.

¹⁵Суперпозицию векторов состояния в пространстве $\mathbf{H}^S \otimes \mathbf{H}^Q \otimes \mathbf{H}_\infty$ можно рассматривать как унитарное (бесконечномерное) представление некоторой группы. Например, согласно Вигнеру [27], квантовая система, описываемая неприводимым унитарным представлением группы Пуанкаре, является элементарной частицей (интерпретация Вигнера). Очевидно, что для нашего случая группа Пуанкаре (группа движений пространства-времени), являющаяся фундаментальной симметрией модуса актуального, недостаточна. С другой стороны, группы $SU(N)$, описывающие приближенные динамические симметрии (так называемые «внутренние» симметрии), также недостаточны в силу их компактности и, как следствие этого, отсутствия бесконечномерных представлений. Наиболее вероятным кандидатом на роль фундаментальной симметрии модуса потенциального является группа Румера-Фета $SU(2, 2) \otimes SU(2) \otimes SU(2)$ [28], имеющая унитарные представления и содержащая в качестве подгруппы группу Лоренца ($SU(2, 2)$ — универсальное накрытие конформной группы $SO_0(2, 4)$).

представления τ_{ij} степени (12), $\mathcal{C}_{p,q}$ — алгебра Клиффорда, ассоциированная с τ_{ij} , $\mathbb{S}_{(p+q)/2}$ — спинпространство, ассоциированное с алгеброй $\mathcal{C}_{p,q}$, $C^{a,b,c,d,e,f,g}$ — *SPT* группа, определяемая в рамках алгебры $\mathcal{C}_{p,q}$. Очевидно, что главным объектом, определяющим векторы вида (23), является представление τ_{ij} . Объекты $\mathcal{C}_{p,q}$, $\mathbb{S}_{(p+q)/2}$ и $C^{a,b,c,d,e,f,g}$ принадлежат спинорной структуре. Другими словами, представление τ_{ij} соответствует каждому вектору пространства $\mathbf{H}^S \otimes \mathbf{H}^Q \otimes \mathbf{H}_\infty$, а множество всех возможных представлений τ_{ij} генерирует $\mathbf{H}^S \otimes \mathbf{H}^Q \otimes \mathbf{H}_\infty$ с векторами (23).

ЛИТЕРАТУРА

1. Чью Дж. Аналитическая теория S-матрицы. М. : Мир, 1968.
2. Дирак П.А.М. Лекции по квантовой теории поля. М. : Мир, 1971.
3. Румер Ю.Б., Фет А.И. Теория унитарной симметрии. М. : Наука, 1970.
4. Гельфанд И.М., Минлос Р.А., Шапиро З.Я. Представления группы вращений и группы Лоренца, их применения. М. : Физматлит, 1958.
5. Румер Ю.Б., Фет А.И. Теория групп и квантованные поля. М. : Наука, 1977.
6. Bhabha H.J. Relativistic Wave Equations for the Elementary Particles // Rev. Mod. Phys. 1945. V. 17. P. 200–216.
7. Varlamov V.V. General Solutions of Relativistic Wave Equations // Int. J. Theor. Phys. 2003. V. 42. P. 583–633; arXiv:math-ph/0209036 (2002).
8. Varlamov V.V. General Solutions of Relativistic Wave Equations II: Arbitrary Spin Chains // Int. J. Theor. Phys. 2007. V. 46. P. 741–805; arXiv:math-ph/0503058 (2005).
9. Гейзенберг В. Шаги за горизонт. М. : Прогресс, 1987.
10. Севальников А.Ю. Интерпретации квантовой механики: В поисках новой онтологии. М. : Книжный дом «ЛИБРОКОМ», 2009.
11. Zurek W.H. Decoherence, Einselection, and the Quantum Origins of the Classical // Rev. Mod. Phys. 2003. V. 75. P. 715; arXiv:quant-ph/0105127 (2001).
12. Кулаков Ю.И. Теория физических структур. Новосибирск : Изд-во «Альфа Виста», 2004.
13. Владимиров Ю.С. Реляционная теория пространства-времени и взаимодействий. Ч. 2. Теория физических взаимодействий. М. : МГУ, 1998.
14. Фок В.А. Об интерпретации квантовой механики. М., 1957.
15. Фон Нейман И. Математические основы квантовой механики. М. : Наука, 1964.
16. Варламов В.В. Групповое усреднение для свободных полей в терминах гиперсферических функций над группой де Ситтера // ТМФ. 2010. Т. 164:3. С. 473–480.
17. Pauli W. Zur Quantenmechanik des magnetischen Elektrons // Z. Phys. 1927. V. 43. P. 601–623.
18. Heisenberg W. Über den Bau der Atomkerne // Z. Phys. 1932. V. 77. P. 1–11.
19. Kemmer N. The charge-dependence of nuclear forces // Proc. Camb. Phil. Soc. 1938. V. 34. P. 354–364.
20. Wick G.G., Wigner E.P., Wightman A.S. Intrinsic Parity of Elementary Particles // Phys. Rev. 1952. V. 88. P. 101.
21. Варламов В.В. Дискретные симметрии на пространствах фактор-представлений группы Лоренца // Математические структуры и моделирование. 2001. Вып. 7. С. 114–127.

22. Varlamov V.V. Universal Coverings of Orthogonal Groups // Adv. Appl. Clifford Algebras. 2004. V. 14. P. 81–168; arXiv:math-ph/0405040 (2004).
23. Varlamov V.V. CPT groups for spinor field in de Sitter space // Phys. Lett. B. 2005. V. 631. P. 187–191; arXiv:math-ph/0508050 (2005).
24. Varlamov V.V. CPT Groups of Higher Spin Fields // Int. J. Theor. Phys. 2012. V. 51. P. 1453–1481; arXiv: 1107.4156 [math-ph] (2011).
25. Varlamov V.V. CPT groups of spinor fields in de Sitter and anti-de Sitter spaces // Adv. Appl. Clifford Algebras. 2015. V. 25. P. 487–516; arXiv: 1401.7723 [math-ph](2014).
26. Varlamov V.V. Cyclic structures of Cliffordian supergroups and particle representations of $\mathbf{Spin}_+(1, 3)$ // Adv. Appl. Clifford Algebras. 2014. V. 24. P. 849–874; arXiv: 1207.6162 [math-ph] (2012).
27. Wigner E.P. On unitary representations of the inhomogeneous Lorentz group // Ann. Math. 1939. V. 40. P. 149–204.
28. Фет А.И. Группа симметрии химических элементов. Новосибирск : Наука, 2010.

COMPLEX MOMENTUM AND SPIN-CHARGE HILBERT SPACE

V.V. Varlamov

Dr.Sc.(Phys.-Math.), e-mail: varlamov@sibsiu.ru

Siberian State Industrial University

Abstract. Mass formula and eigenspaces of energy operator are defined within a complex envelope (complex momentum) of the group algebra of the proper orthochronous Lorentz group. Eigenspaces of the energy operator are identified with the mass levels of the elementary particle spectrum. An elementary particle is understood as a superposition of the state vectors in a spin-charge Hilbert space.

Keywords: complex momentum, Lorentz group, spinor structure, mass spectrum, Hilbert space.

О МОМЕНТАХ ОБОБЩЁННЫХ СУММ

А.Г. Гринь

д.ф.-м.н., профессор, e-mail: griniran@gmail.com

Омский государственный университет им. Ф.М. Достоевского

Аннотация. В работе получены оценки для моментов так называемых обобщённых сумм слабо зависимых величин в терминах моментов меньших порядков. Оценки получены с помощью аналога известного неравенства М. Пелиград.

Ключевые слова: обобщённые суммы, равномерно сильное перемешивание, оценки для моментов.

Пусть $\{\xi_n\} = \{\xi_n, n = 1, 2, \dots\}$ — стационарная в узком смысле последовательность и пусть $\mathcal{F}_{\leq n}$ и $\mathcal{F}_{\geq n}$ — σ -алгебры, порождённые семействами $\{\xi_i : i \leq n\}$ и $\{\xi_i : i \geq n\}$. Говорят, что последовательность $\{\xi_n\}$ удовлетворяет *условию равномерно сильного перемешивания* (φ -перемешивания) с коэффициентом перемешивания $\varphi(n)$, если

$$\varphi(n) = \sup \left\{ \frac{|\mathbf{P}(AB) - \mathbf{P}(A)\mathbf{P}(B)|}{\mathbf{P}(A)} : A \in \mathcal{F}_{\leq 0}, B \in \mathcal{F}_{\geq n} \right\} \rightarrow 0, \quad n \rightarrow \infty.$$

Если $\{\xi_n\}$ — стационарная последовательность с φ -перемешиванием, $S_n = \sum_{i=1}^n \xi_i$, $\mathbf{E}\xi_1 = 0$, $\sigma_n^2 = \mathbf{E}S_n^2 \rightarrow \infty$, $n \rightarrow \infty$, то $\mathbf{E}|S_n|^p \leq C\sigma_n^p$, $p > 2$, где $C > 0$ не зависит от n . Такие оценки впервые получены И. А. Ибрагимовым (см., например, [1, лемма 18.5.1]); на этих оценках базировалось доказательство центральной предельной теоремы для последовательностей с φ -перемешиванием. В дальнейшем после появления известного неравенства М. Пелиград [2], с его помощью оценки такого типа доказывались различными авторами в различных модификациях (см., например, [3]). В настоящей работе на основе некоторого аналога неравенства М. Пелиград оценки подобного типа получены для так называемых обобщённых сумм (см. [4]).

Обобщённой суммой $x \oplus y$ будем называть бинарную операцию на $\mathbb{D} \subseteq \mathbb{R}$, удовлетворяющую условиям $A_1 - A_4$ (условия **(A)**):

A_1 . Ассоциативность: $x \oplus (y \oplus z) = (x \oplus y) \oplus z$, $x, y, z \in \mathbb{D}$;

A_2 . Коммутативность: $x \oplus y = y \oplus x$, $x, y \in \mathbb{D}$;

A_3 . $x \oplus 0 = x$, $x \in \mathbb{D}$;

A_4 . Равномерная непрерывность в следующем смысле: для любого $\varepsilon > 0$ найдётся $\delta > 0$ такое, что из $|y| < \delta$ следует $|x \oplus y - x| < \varepsilon$, $\forall x \in \mathbb{D}$;

Этим условиям удовлетворяют, например, $x \oplus y = x + y$, $\mathbb{D} = \mathbb{R}$, $x \vee y = \max\{x, y\}$, $\mathbb{D} = \mathbb{R}_+ = [0, +\infty)$, $x \wedge y = \min\{x, y\}$, $\mathbb{D} = \mathbb{R}_- = (-\infty, 0]$,

а не удовлетворяют, скажем, $x \oplus y = xy$, (не выполняются A_3 и A_4) и $x \oplus y = x + y \pmod{d}$, $d > 0$, $\mathbb{D} = \mathbb{R}$ (не выполняется A_4).

Если бинарная операция $x \otimes y$ удовлетворяет условиям (А), а $f(x)$ возрастающая выпуклая (вниз) функция такая, что $f(0) = 0$, $f(\mathbb{D}) \subseteq \mathbb{D}$, то бинарная операция $x \oplus y = f^{-1}(f(x) \otimes f(y))$ также удовлетворяет условиям (А).

Будем обозначать

$$X_{k,m}(b) = \left(\frac{\xi_k}{b}\right) \oplus \dots \oplus \left(\frac{\xi_m}{b}\right), \quad X_n(b) = X_{1,n}(b),$$

$$X_n = X_n(1), \quad \bar{X}_n(b) = \max_{1 \leq k \leq n} |X_k(b)|, \quad k, m, n \in \mathbb{N}, \quad b > 0.$$

В дальнейшем будем предполагать, что $\mathbb{D} = \mathbb{R}$.

Лемма 1. Для любого $\varepsilon > 0$ найдётся $\delta > 0$ такое, что если

$$\max_{1 \leq k \leq n} \mathbf{P}\{|X_k(xc_n)| \geq \delta\} + \varphi(m) \leq \gamma < 1, \quad x > 0,$$

то при любых $a > 0$

$$\mathbf{P}\{\bar{X}_n(xc_n) \geq a + \varepsilon\} \leq \frac{1}{1 - \gamma} \left(\mathbf{P}\{|X_n(xc_n)| \geq a\} + \mathbf{P}\left\{\max_{1 \leq k \leq n} |\xi_k| \geq \delta xc_n\right\} \right).$$

Доказательство. Из свойств $A_1 - A_4$ выводится, что при любом натуральном m для любого $\varepsilon > 0$ найдётся $\delta > 0$ такое, что при любом $x > 0$

$$\{|\xi| \geq x + \varepsilon, |\eta| < \delta\} \subseteq \{|\xi \oplus \eta| \geq x\}. \quad (1)$$

$$\{|\xi| \geq x + \varepsilon, |\eta_1| < \delta, \dots, |\eta_m| < \delta\} \subseteq \{|\xi \otimes \eta_1 \oplus \dots \oplus \eta_m| \geq x\}. \quad (2)$$

Пусть $E_k = \{\bar{X}_{k-1}(xc_n) < a + \varepsilon \leq |X_k(xc_n)|\}$, $k = 1, \dots, n$. Тогда $E_i E_j = \emptyset$, $i \neq j$, $\bigcup_{k=1}^n E_k = \{\bar{X}_n(xc_n) \geq a + \varepsilon\}$, а в силу (2) найдётся $\delta > 0$ такое, что

$$\{|X_k(xc_n)| \geq a + \varepsilon, \max_{1 \leq k \leq n} |\xi_k| < \delta xc_n, |X_{k+m,n}(xc_n)| < \delta\} \subseteq \{|X_m(xc_n)| \geq a\},$$

то есть

$$\{|X_n(xc_n)| < a\} \subseteq \{|X_k(xc_n)| < a + \varepsilon\} \cup \{|X_{k+m,n}(xc_n)| \geq \delta\} \cup \left\{ \max_{1 \leq k \leq n} |\xi_k| \geq \delta xc_n \right\},$$

$k = 1, \dots, n - 1$, откуда

$$\{|X_n(xc_n)| < a, E_k\} \subseteq \{|X_{k+m,n}(xc_n)| \geq \delta, E_k\} \cup \left\{ \max_{1 \leq k \leq n} |\xi_k| \geq \delta xc_n, E_k \right\}. \quad (3)$$

С помощью (3) получаем

$$\mathbf{P}\{\bar{X}_n(xc_n) \geq a + \varepsilon\} \leq \mathbf{P}\{|X_n(xc_n)| \geq a\} + \sum_{k=1}^n \mathbf{P}\{|X_n(xc_n)| < a, E_k\} +$$

$$\begin{aligned}
 & +\mathbf{P}\left\{\max_{1 \leq k \leq n} |\xi_k| \geq \delta x c_n\right\} \leq \mathbf{P}\{|X_m(x c_n)| \geq a\} + \sum_{k=1}^n \mathbf{P}\{|X_{k+m,n}(x c_n)| \geq \delta, E_k\} \leq \\
 & \leq \mathbf{P}\{|X_n(x c_n)| \geq a\} + \left(\max_{1 \leq k \leq n} \mathbf{P}\{|X_k(x c_n)| \geq \delta\} + \varphi(m)\right) \sum_{k=1}^n \mathbf{P}\{E_k\} + \\
 & \quad +\mathbf{P}\left\{\max_{1 \leq k \leq n} |\xi_k| \geq \delta x c_n\right\} \leq \mathbf{P}\{|X_m(x c_n)| \geq a\} + \\
 & \quad +\gamma \mathbf{P}\{\bar{X}_n(x c_n) \geq a + \varepsilon\} + \mathbf{P}\left\{\max_{1 \leq k \leq n} |\xi_k| \geq \delta x c_n\right\},
 \end{aligned}$$

откуда следует утверждение леммы. ■

Следующее предложение — это модификация для обобщённых сумм неравенства М. Пелиград (леммы 3.1 из [2]).

Лемма 2. Для любого $\varepsilon > 0$ найдётся $\delta > 0$ такое, что если

$$\max_{1 \leq k \leq n} \mathbf{P}\{|X_k(x c_n)| \geq \delta\} + \varphi(m) \leq \gamma < 1, \quad x > 0,$$

то при любом $a > 0$

$$\mathbf{P}\{|X_n(x c_n)| \geq a + 2\varepsilon\} \leq \frac{\gamma}{1 - \gamma} \mathbf{P}\{|X_n(x c_n)| \geq a\} + \frac{1}{1 - \gamma} \mathbf{P}\left\{\max_{1 \leq k \leq n} |\xi_k| \geq \delta x c_n\right\}.$$

Доказательство. Пусть $E_k = \{\bar{X}_{k-1}(x c_n) < a + \varepsilon \leq |X_k(x c_n)|\}, k = 1, \dots, n$. Тогда $E_i E_j = \emptyset, i \neq j, \bigcup_{k=1}^n E_k = \{\bar{X}_n(x c_n) \geq a + \delta\}$. В силу (2) для любого $\varepsilon > 0$ найдётся $\delta > 0$ такое, что при $1 \leq k \leq n - m$

$$\begin{aligned}
 & \{|X_{k+m,n}(x c_n)| < \delta, E_k, \max_{1 \leq k \leq n} |\xi_k| < \delta x c_n\} \subseteq \\
 & \subseteq \{|X_n(x c_n)| < a + 2\varepsilon, E_k, \max_{1 \leq k \leq n} |\xi_k| < \delta x c_n\},
 \end{aligned}$$

откуда

$$\{|X_n(x c_n)| \geq a + 2\varepsilon, E_k, \max_{1 \leq k \leq n} |\xi_k| < \delta x c_n\} \subseteq \{E_k, |X_{k+m,n}(x c_n)| \geq \delta\}. \quad (4)$$

Аналогично выводится

$$\{|X_n(x c_n)| \geq a + 2\varepsilon, \max_{1 \leq k \leq n} |\xi_k| < \delta x c_n\} \subseteq \{\bar{X}_{n-m}(x c_n) \geq a + \varepsilon, \max_{1 \leq k \leq n} |\xi_k| < \delta x c_n\}.$$

Отсюда

$$\begin{aligned}
 & \{|X_n(x c_n)| \geq a + 2\varepsilon, \max_{1 \leq k \leq n} |\xi_k| < \delta x c_n\} = \\
 & = \{|X_n(x c_n)| \geq a + 2\varepsilon, \bar{X}_{n-m}(x c_n) \geq a + \varepsilon, \max_{1 \leq k \leq n} |\xi_k| < \delta x c_n\}. \quad (5)
 \end{aligned}$$

$$\begin{aligned}
& \text{С помощью (7) и (8) получаем } \mathbf{P}\{|X_n(xc_n)| \geq a + 2\varepsilon\} \leq \\
& \leq \mathbf{P}\{|X_n(xc_n)| \geq a + 2\varepsilon, \max_{1 \leq k \leq n} |\xi_k| < \delta xc_n\} + \mathbf{P}\{\max_{1 \leq k \leq n} |\xi_k| \geq \delta xc_n\} = \\
& = \mathbf{P}\{|X_n(c_n)| \geq a + 2\varepsilon, \bar{X}_{n-m}(c_n) \geq a + \varepsilon, \max_{1 \leq k \leq n} |\xi_k| < \delta xc_n\} + \\
& + \mathbf{P}\{\max_{1 \leq k \leq n} |\xi_k| \geq \delta xc_n\} = \sum_{k=1}^{n-m} \mathbf{P}\{|X_n(xc_n)| \geq a + 2\varepsilon, E_k, \max_{1 \leq k \leq n} |\xi_k| < \delta xc_n\} + \\
& + \mathbf{P}\{\max_{1 \leq k \leq n} |\xi_k| \geq \delta xc_n\} \leq \mathbf{P}\left\{\max_{1 \leq k \leq n} |\xi_k| \geq \delta xc_n\right\} + \sum_{k=1}^{n-m} \mathbf{P}\{|X_{k+m,n}(xc_n)| \geq \delta, E_k\} \leq \\
& \leq \mathbf{P}\left\{\max_{1 \leq k \leq n} |\xi_k| \geq \delta xc_n\right\} + \left(\max_{1 \leq k \leq n} \mathbf{P}\{|X_k(xc_n)| \geq \delta\} + \varphi(m)\right) \sum_{k=1}^n \mathbf{P}\{E_k\} \leq \\
& \leq \gamma \mathbf{P}\{\bar{X}_n(c_n) \geq a + \delta\} + \mathbf{P}\left\{\max_{1 \leq k \leq n} |\xi_k| \geq \delta xc_n\right\}.
\end{aligned}$$

Отсюда с помощью леммы 1 получаем утверждение леммы 2. \blacksquare

Покажем, как с помощью леммы 2 можно получать оценки для моментов обобщённых сумм. Пусть $\delta > 0$, $N > 0$ и натуральное m таковы, что

$$\max_{1 \leq k \leq n} \mathbf{P}\{|X_k(Nc_n)| \geq \delta\} + \varphi(m) \leq \gamma < 1,$$

где $\gamma > 0$ такое, что $\frac{\gamma}{1-\gamma} < 1$. Предположим сначала, что $\max_{1 \leq k \leq n} |\xi_k| < \delta Nc_n$ почти наверное. Из леммы 2 следует тогда

$$\begin{aligned}
\mathbf{P}\{|X_n(Nc_n)| \geq a + 2k\varepsilon\} & \leq \frac{\gamma}{1-\gamma} \mathbf{P}\{|X_n(Nc_n)| \geq a + 2(k-1)\varepsilon\} \leq \dots \leq \\
& \leq \frac{\gamma^k}{(1-\gamma)^k} \mathbf{P}\{|X_n(Nc_n)| \geq a\},
\end{aligned}$$

то есть

$$\mathbf{P}\{|X_n(Nc_n)| \geq y\} \leq \exp\{-\alpha y\}, \quad \alpha > 0, \quad y > 0.$$

Отсюда следует, что если $\int_0^\infty f(x) \exp\{-\alpha y\} dx < \infty$, $f(x) \geq 0$, то $\sup_n \mathbf{E}f(|X_n(Nc_n)|) < \infty$, в частности $\sup_n \mathbf{E}|X_n(Nc_n)|^p < \infty$ при любом $p > 0$.

Будем говорить, что выполнено условие A_5 , если при любых $x > 0$, $y_i \in \mathbb{D}$, $i = 1, \dots, n$, $n \geq 2$

$$(xy_1) \oplus \dots \oplus (xy_n) = x(y_1 \oplus \dots \oplus y_n).$$

Например, $x \oplus y = x + y$, $x \oplus y = (|x|^p + |y|^p)^{1/p}$, $p \geq 1$, $\mathbb{D} = \mathbb{R}$, $x \oplus y = x \vee y$, $\mathbb{D} = \mathbb{R}_+$, удовлетворяют условию A_5 .

Теорема 1. Пусть бинарная операция \oplus удовлетворяет условиям $A_1 - A_5$ и пусть $0 < q < p$, $\mathbf{E}|X_n|^p < \infty$. Тогда

$$\mathbf{E}|X_n|^p \leq A (\mathbf{E}|X_n|^q)^{p/q} + B \mathbf{E} \max_{1 \leq k \leq n} |\xi_k|^p,$$

где A и B не зависят от n .

Доказательство. Для бинарной операции, удовлетворяющей условиям $A_1 - A_5$, утверждение леммы 2 можно переписать так: для любого $\varepsilon > 0$ найдётся $\delta > 0$ такое, что если

$$\max_{1 \leq k \leq n} \mathbf{P}\{|X_k| \geq \delta N c_n\} + \varphi(m) \leq \gamma < 1,$$

то при $x \geq N$

$$\mathbf{P}\{|X_n| \geq x c_n\} \leq \frac{\gamma}{1 - \gamma} \mathbf{P}\{|X_n| \geq \frac{x c_n}{1 + 2\varepsilon}\} + \frac{1}{1 - \gamma} \mathbf{P}\left\{\max_{1 \leq k \leq n} |\xi_k| \geq \frac{\delta x c_n}{1 + 2\varepsilon}\right\}. \quad (6)$$

Если $\gamma > 0$ таково, что $\frac{\gamma(1 + 2\varepsilon)^p}{1 - \gamma} < 1$, то с помощью (6) получаем

$$\begin{aligned} \mathbf{E}|X_n|^p &= -p \int_0^\infty x^{p-1} \mathbf{P}\{|X_n| \geq x\} dx \leq (N c_n)^p + p c_n^p \int_N^\infty x^{p-1} \mathbf{P}\{|X_n| \geq x c_n\} dx \leq \\ &\leq (N c_n)^p + \frac{\gamma(1 + 2\varepsilon)^p}{1 - \gamma} \mathbf{E}|X_n|^p + \frac{\gamma(1 + 2\varepsilon)^p}{\delta^p(1 - \gamma)} \mathbf{E} \max_{1 \leq k \leq n} |\xi_k|^p. \end{aligned}$$

Отсюда

$$\mathbf{E}|X_n|^p \leq A' c_n^p + B' \mathbf{E} \max_{1 \leq k \leq n} |\xi_k|^p, \quad (7)$$

где A' и B' не зависят от n .

Пусть $0 < q < p$, $c_n^q = \max_{1 \leq k \leq n} \mathbf{E}|X_k|^q$. Тогда

$$\max_{1 \leq k \leq n} \mathbf{P}\{|X_k| \geq \delta N c_n\} \leq \frac{c_n^q}{(\delta N c_n)^q} = (\delta N)^{-q},$$

так что m и N можно выбрать такими, что $\frac{\gamma(1 + 2\varepsilon)^p}{1 - \gamma} < 1$ и из (7) следует теперь

$$\mathbf{E}|X_n|^p \leq A' \left(\max_{1 \leq k \leq n} \mathbf{E}|X_n|^q \right)^{p/q} + B' \mathbf{E} \max_{1 \leq k \leq n} |\xi_k|^p. \quad (8)$$

В силу леммы 1

$$\mathbf{P}\{\bar{X}_n \geq x c_n\} \leq \frac{1}{1 - \gamma} \left(\mathbf{P}\left\{|X_n| \geq \frac{x c_n}{1 + \varepsilon}\right\} + \mathbf{P}\left\{\max_{1 \leq k \leq n} |\xi_k| \geq \frac{\delta x c_n}{1 + \varepsilon}\right\} \right), \quad x > 0,$$

откуда

$$\max_{1 \leq k \leq n} \mathbf{E}|X_n|^q \leq \mathbf{E}\bar{X}_n^q \leq \frac{(1 + \varepsilon)^q}{1 - \gamma} \left(\mathbf{E}|X_n|^q + \delta^{-q} \mathbf{E} \max_{1 \leq k \leq n} |\xi_k|^q \right),$$

что вместе с (8) даёт утверждение теоремы. ■

Пусть $x \oplus y = x + y$, $X_n = S_n$, $\mathbf{E}\xi_1 = 0$, $\sigma_n^2 = \mathbf{E}S_n^2 \rightarrow \infty$, $n \rightarrow \infty$. Тогда σ_n является правильно меняющейся последовательностью порядка $1/2$ [1, теорема 18.2.3], так что

$$\mathbf{E} \max_{1 \leq k \leq n} |\xi_k|^p \leq n \mathbf{E}|\xi_1|^p = o(\sigma_n^p), \quad p > 2,$$

и из теоремы 1 следуют оценки И.А. Ибрагимова.

ЛИТЕРАТУРА

1. Ибрагимов И.А., Линник Ю.В. Независимые и стационарно связанные величины. М. : Наука, 1965. 524 с.
2. Peligrad M. An invariance principle for φ -mixing sequences // Ann. Probab. 1985. V. 13, № 4. P. 1304–1313.
3. Гринь А.Г. Нормирующие последовательности в предельных теоремах для слабо зависимых величин // Теория вероят. и её примен. 1991. Т. 36, № 2. С. 285–300.
4. Гринь А.Г. Условия слабой зависимости в предельных теоремах для обобщённых сумм // Математические структуры и моделирование. 2014. Вып. 29. С. 4-12.
5. Сенета Е. Правильно меняющиеся функции. М. : Наука, 1985, 141 с.

ON THE MOMENTS OF GENERALIZED SUMS

A.G. Grin

Dr.Sc.(Phys.-Math.), Professor, e-mail: griniran@gmail.com

Omsk State University n.a. F.M. Dostoevskiy

Abstract. We obtain estimates for the moments of so-called generalized sums of weakly dependent variables in terms of moments of smaller order. Estimates are obtained with the aid of an analogue of the well-known M. Peligrad inequality.

Keywords: generalized sums, estimates for the moments, uniformly strong mixing condition.

КОНЦЕПЦИЯ ФРАКТАЛА МАНДЕЛЬБРОТА С МАТЕМАТИЧЕСКОЙ И ФИЛОСОФСКОЙ ТОЧЕК ЗРЕНИЯ

В.А. Еровенко

профессор, д.ф.-м.н., e-mail: erovenko@bsu.by

Белорусский государственный университет, г. Минск, Беларусь

Аннотация. Статья посвящена философско-математическим аспектам концепции фрактала Мандельброта, которая стала реализацией синтеза геометрического и виртуально-социокультурного восприятия мира, возродившего экспериментальную математику «неидеализированных» объектов.

Ключевые слова: фрактал Мандельброта, фрактальная геометрия, фрактальная размерность, фрактальная реальность, современный компьютер.

Выдающийся французский и американский математик польского происхождения Бенуа Мандельброт (1924–2010) основатель и создатель теории фракталов изменил мировоззренческую парадигму традиционного восприятия хаоса, показав, что хаос и сложность могут возникать в результате действия простых математических законов. Современная наука переживает эволюцию, связанную с заменой «непрерывной» парадигмы на «фрактальную». Введя в научный оборот понятие фрактала, он не только продемонстрировал универсальность фрактальных структур в природе, дав им в 1975 году название «фрактальная геометрия», но и создал математику для её описания, с помощью которой появилась возможность выделения порядка в многообразии хаоса, который до этого выглядел непрístupным. Методологической целью фрактальной геометрии был математический анализ изломанных, «морщинистых» и нечётких геометрических форм, для которых Мандельброт использовал слово «фрактал», наткнувшись однажды в словаре на латинское прилагательное *fractus*, являющееся производным от глагола *fracture*, что означает «ломать». Философско-математическая суть фракталов в самоподобии, когда их геометрическая структура остаётся приблизительно той же в любом масштабе рассмотрения, поэтому, несмотря на то, что понятие фрактала появилось сравнительно недавно, как принцип мироустройства, они существовали всегда, так как самоподобие — это универсальное свойство природы.

Высшим достижением фрактальной геометрии стало открытие Мандельбротом математической структуры, которая, с одной стороны, обладает поразительной сложностью, а с другой стороны, может быть воспроизведена с помощью очень простой итеративной процедуры. Речь идёт о множестве Мандельброта — примере чисто математической абстракции, которая, однако, представима графически и может рассматриваться как геометрический объект. Неровность

этого множества является центральной темой фрактальной геометрии, изучая которую, Мандельброт обнаружил «фрактальный порядок» там, где до него видели только беспорядок. Основным методом для построения математических фракталов служит итерация, то есть многократное повторение определённой геометрической операции, а так как геометрические фигуры зависимы от пространства, то в философском контексте они помещаются между чувственным миром и миром идеальным. Это, вообще говоря, неудивительно, так как в абстрактной математике многие из умопостигаемых математических сущностей оказываются в то же время представимы пространственно, то есть имеют аналоги в чувственно воспринимаемом мире. Следует также отметить, что поскольку Мандельброт не дал окончательного определения фрактала, то его широкое философское толкование позволило включить в число фракталов хорошо известные в современной математике множества Кантора, Коха, Серпинского, Пеано, Жюлиа и много других.

Интерпретация фрактальных структур в конкретных познавательных установках заставила философов математики по-новому оценить хорошо известные математические понятия и сущности, в частности, различные типы размерности, парадоксы измерений, а также такой простейший фрактал, как знаменитое канторово множество, возникшее при попытке решения проблемы гипотезы континуума. С философской точки зрения концепция фрактала Мандельброта дистанцируется от таких традиционных понятий задания и описания геометрической формы, как граница, длина, ширина. Этим смысловых понятий там просто нет, так как они перестают работать внутри концепции фрактала, и потому совершенно не понятно, как их применять. Неожиданное продолжение эти наблюдения получили в новой области математического знания — теории фракталов, которая смещает познавательные установки от строгой рациональности до интуитивно-образного мышления. С математической точки зрения интересно то, что для введения понятия «фрактал» не потребовалось изобретать каких-то абсолютно новых формализаций или математических понятий. Заметим, что фракталами принято иногда называть объекты дробной размерности, которые обладают свойством масштабной инвариантности, или «самоподобия», когда изменение масштаба не меняет их структуры. Современный геометрический аспект теории масштабной инвариантности был разработан Бенуа Мандельбротом до математической теории новых геометрических форм, имеющих дробную размерность, которая с полным на то основанием была названа им «фрактальной геометрией». Это научное направление, которое изначально опиралось на математические парадоксы, совершенно неожиданно подарило неповторимые интеллектуальные и эстетические впечатления от красоты многих фракталов. Они связаны со свойством фрактальных множеств «выглядеть» в любом масштабе примерно одинаково, которое сейчас называется «масштабной инвариантностью». По этому поводу Мандельброт писал: «Моя увлечённость масштабной инвариантностью постоянно подпитывалась новым энтузиазмом и, обогащаясь, благодаря сменам области исследований, новыми инструментами и идеями, постепенно подводила меня к созданию полноценной общей теории» [1, с. 587]. Другими словами, масштабная инвариантность означает, что

малый фрагмент структуры фрактального объекта подобен другому более крупному фрагменту этого объекта или даже структуре в целом, как например, в «снежинке Коха». Граница снежинки, придуманной шведским математиком Гельгомом фон Кохом в 1904 году, описывается кривой, составленной из трёх одинаковых фракталов. Каждая треть снежинки строится итеративно, начиная с одной из сторон равностороннего треугольника, которая в итоге даёт стандартный пример известного фрактала — «кривую Коха».

Математическое построение этой кривой начинается с единичного отрезка, который заменяется ломаной из четырёх отрезков, каждый из которых имеет длину равную $1/3$, а именно, убирается средняя треть отрезка, и, рассматривая её как основание равностороннего треугольника, добавляются два новых отрезка, строящихся как боковые стороны треугольника. Затем многократно повторяем данную процедуру, на каждом шаге заменяя среднюю треть двумя новыми отрезками. Длина полученной ломаной линии после n -го шага будет равна $(4/3)^n$, поэтому результирующая длина кривой Коха бесконечна: $\lim(4/3)^n = \infty$ при $n \rightarrow \infty$. Если построение начинается с треугольника, то получается снежинка Коха, которую ещё называют «островом Коха». Мы можем фотографировать этот остров в океане из космоса с любым увеличением, хотя мелкие детали в крупном масштабе, естественно, будут теряться, но поскольку фракталы обладают масштабной инвариантностью, то при увеличении мы вновь и вновь будем видеть масштабно некорректируемую конструкцию. С точки зрения философии и методологии современной математики, важна не сама длина, а то, как она зависит от размеров линейки, то есть важно некое число, называемое «фрактальной размерностью».

В классической математике понятие размерности употребляется, в силу его определения, сугубо дискретно, например, $n = 1, 2, 3, \dots$ и так далее. В современной математике существуют математические объекты, реально используемые в математической практике, для вычисления размерности которых классического определения явно недостаточно. Для решения такого рода современных проблем неклассической математики используется понятие хаусдорфовой размерности. Строго математическое определение этой размерности даётся с помощью технически сложного понятия меры Хаусдорфа, но его суть можно прояснить следующим образом. Пусть отрезок длины A покрывается отрезками длины ε , или квадрат площадью A покрывается квадратами со стороной длины ε , или куб площадью A покрывается кубами со стороной длины ε . Тогда, соответственно, для отрезка $A = N(\varepsilon)\varepsilon$, для квадрата $A = N(\varepsilon)\varepsilon^2$ и для куба $A = N(\varepsilon)\varepsilon^3$, где $N(\varepsilon)$ — обозначает минимальное число отрезков, соответственно, квадратов или кубов, необходимых для покрытия множества A . В приведённых примерах показатель степени в каждом случае совпадает с классической размерностью рассмотренных математических объектов, поэтому можно записать, что $A = N(\varepsilon)\varepsilon^d$, где d — размерность, и, следовательно, при $\varepsilon \rightarrow \infty$ имеем, что $N(\varepsilon)$ растёт пропорционально ε^{-d} . Прологарифмировав последнее равенство для A , получим, что $d = (\ln A - \ln N(\varepsilon)) / \ln \varepsilon$. Поэтому хаусдорфова размерность

произвольного объекта в n -мерном пространстве определяется по формуле:

$$d = -\lim(\ln N(\varepsilon))/\ln \varepsilon \text{ при } \varepsilon \rightarrow 0.$$

Заметим, что так называемая хаусдорфова размерность кривой Коха равна: $d = \ln 4/\ln 3 \approx 1,262$, то есть для береговой линии острова Коха она лежит между числами 1 и 2. В частности, для этой линии, как и для кривой, описываемой функцией Вейерштрасса, нельзя провести касательную ни в одной точке. Экзотический пример острова Коха показывает, что конечная площадь фигуры может не влиять на бесконечную длину её периметра.

Отметим, что первоначальное определение фрактала, опиралось на классическое представление о хаусдорфовой размерности, у которой есть другое название — размерность Хаусдорфа-Безиковича, поскольку математик Феликс Хаусдорф ввёл это понятие, а Абрам Безикович придал ему окончательный вид. Использование Мандельбротом дробной размерности в 1975 году шокировало математический мир. Хотя тем, кто возражает против «нелепой», с их точки зрения, дробной размерности, аргументируя это тем, что ей нельзя придать наглядный смысл, можно также возразить, что никто «не присягал» на незыблемость целочисленной размерности, лишь потому, что топологическая размерность принимает исключительно целые значения. Фрактальная размерность уже доказала свою методологическую полезность. Эффективность новых теоретических инструментов заслуживает особого внимания с точки зрения переноса «фрактальной геометрии природы» на абстрактную математическую основу. Но возникает вопрос: так что такое фрактал? Общепринятого определения этого понятия не существует, так как даже определение фрактала по Мандельброту — «множество, хаусдорфова размерность которого превышает его топологическую размерность», вряд ли можно считать удовлетворительным. Действительно, со временем, как заметил популяризатор математики Ю.А. Данилов, выяснилось «обстоятельство почти криминального характера: размерность Хаусдорфа-Безиковича некоторых фракталов оказалась целой» [2, с. 189]. Хотя точного определения фрактала до сих пор не предложено, такое определение, при наличии разрушающих его контрпримеров, особо и не нужно в силу сложившейся «интерсубъективной практики» научного применения этой категории математики. Это тот случай, когда нужно продолжать знакомство с новым понятием не с уточнения его определения, а с рассмотрения конкретизирующих его контрпримеров.

Во фрактальной геометрии понятие самоподобия, говорящее о том, что часть в каком-то смысле подобна целому, в новом методологическом контексте меняет познавательный статус философского понятия «часть», а также его смысл последнего, установившийся в евклидовой геометрии. Отметим, что «канторово множество» задаётся с помощью двузначного отображения вида $f(x) = f_1(x) \cup f_2(x)$, где $f_1(x) = (1/3)x$ и $f_2(x) = (1/3)x + 2/3$, а это и есть преобразования подобия отрезка $[0, 1]$, которое влечёт свойство самоподобия канторова множества. Геометрическое свойство, присущее фрактальным математическим объектам, — это самоподобие, состоящее в том, что структура, которую объект имеет на «микроуровне», повторяется в нем и на «макроуровне»,

поэтому оно обладает определённым эвристическим потенциалом, поскольку такое свойство присуще некоторым природным формам более общей концепции. Методологически понятие «самоподобие», вообще говоря, неприменимо для описания некоторых фрактальных множеств, например, для множества Мандельброта, фрагменты которых, строго говоря, не отображаются во все множество с помощью преобразования подобия.

Множество Мандельброта, граница которого представляет собой фрактальное множество, описывает конкретные процессы и явления реальности, связанные со свойствами устойчивости и хаоса в динамических системах, стимулируя тем самым исследования фрактальных множеств на комплексной плоскости. Специалист в области концепции обучения фрактальной геометрии В.С. Секованов считает: «Полезны комплексные фракталы для реализации дидактических целей, поскольку математические исследования здесь органически переплетаются с разработкой алгоритмов, реализуемых с помощью современных информационных и коммуникативных технологий...» [3, с. 37]. Бесконечный итерационный процесс как повторное применение операций, порождающий множество Мандельброта, задаётся с помощью квадратичного отображения $z \rightarrow z^2 + c$, где z и c — комплексные числа. Итерационный процесс начинается с какого-нибудь фиксированного числа z , например, в канонической форме это $z = 0$. Затем, в нём каждый раз в алгебраическое выражение $z^2 + c$ подставляется значение z , полученное на предыдущем шаге, а если выбирать различные значения c , то полученные последовательности комплексных чисел стремятся к бесконечности или ограничены в своём движении определённой областью, принимающей форму «множества Мандельброта». Математики, которые с помощью компьютера получили графическое изображение множества Мандельброта, считают, что это множество представляет собой самый сложный геометрический объект в математике. Если следовать противоположному правилу, когда значение c фиксировано, а z принимает различные значения, то получающееся в результате такого итерационного процесса множество отличается от множества Мандельброта, а его граница тоже является фрактальным множеством и называется «множеством Жюлиа» по имени французского математика Гастона Жюлиа.

Несмотря на определённую аналогию в несущей конструкции, эти множества различаются тем, что структура множества Жюлиа подобна самой себе, естественно, в различном масштабе, в то время как для множества Мандельброта, даже для его границы, это не так. Для преодоления этого затруднения, отставив релевантность принятого методологического подхода, можно расширить описания фракталов не только через преобразования подобия, но и через другие виды геометрических преобразований, например, аффинных. Аффинное преобразование, как теоретический конструкт, представляет собой линейное преобразование вместе с последующим преобразованием сдвига, то есть это не что иное, как формула для изменения масштаба, поворотов и перемещений геометрических фигур. Если последовательность таких преобразований применять к фигуре бесконечно, то результат будет подобен самому себе, а именно, ее увеличенная часть выглядит аналогично целому, что характерно

для фракталов. Математическая проблематика, которая проявляется в том, что надо знать, какие преобразования следует выбирать, тесно связана с философскими императивами. В качестве иллюстративного примера можно показать, как, используя небольшой набор аффинных преобразований, можно создавать абстракции, подобные простому самоподобному фракталу под названием «салфетка Серпинского», придуманному известным польским математиком Вацлавом Серпинским. Пусть исходное множество — равносторонний треугольник вместе с областью, которую он замыкает. Разобьём его на четыре меньшие треугольные области, соединив отрезками середины сторон исходного треугольника, и удалим внутренность маленькой центральной треугольной области. Затем повторим процесс такого построения для каждой из трёх оставшихся треугольных областей. Бесконечно продолжая итерации подобным образом, получим последовательность вложенных множеств, чьё пересечение и образует нужное множество.

Конструктивное построение этого фрактала показывает, что полученное множество представляет собой объединение трёх непересекающихся и уменьшенных в два раза копий, следовательно, это пример самоподобного фрактала с размерностью $d = \ln 3 / \ln 2 \approx 1,585$. Заметим, что «площадь» салфетки Серпинского равна нулю, но она слишком «дырява», чтобы иметь площадь в смысле стандартного определения. Действительно, на первом шаге выбросили $1/4$ часть площади, на втором шаге выбросили 3 треугольника, причём площадь каждого из них равна $1/4^2$ площади исходного треугольника. Рассуждая таким образом, получим, что полная доля выкинутой площади составила: $1/4 + 3(1/4^2) + 3^2(1/4^3) + \dots + 3^{n-1}(1/4^n) + \dots = 1$. Поэтому можно утверждать, что оставшееся множество имеет площадь меры нуль. В математике хорошо известна ещё одна конструкция под названием «ковёр Серпинского», который определяется совершенно аналогично, а разница состоит только в том, что салфетка строится на основе треугольника, а ковёр — на основе квадрата. В частности, можно описать совершенно случайный процесс, в результате которого появляется симметричный узор фрактала Серпинского, то есть можно говорить о репрезентации философской импликации — «случайность порождает порядок». Даже если феномен самоподобия может возникнуть там, где его совершенно не ждут, почему математики и философы в своих исследованиях обратились к концепции фракталов? «В этой концепции философско-математического синтеза изумлённому взгляду открывается совершенно новый мир чистой пластической красоты, заложенный давным-давно в слово «геометрия»; открывается особый, парадоксальный тип сложности, удивительный мир синергии хаоса и порядка» [4, с. 37]. Фрактальную геометрию природы, как философско-мировоззренческую концепцию, заставившую исследователей по-новому взглянуть на мир, можно отнести к философии постнеклассической математики, возникшей в конце XX века.

Рассмотрим теперь другой популярный пример простейшего фрактала из неклассической математики, изучаемый студентами-математиками в классическом университете. Для этого посчитаем хаусдорфову размерность канторова множества, которое было построено Кантором, при попытке найти промежу-

точное по мощности множество между рядом натуральных чисел N и континуумом вида $[0, 1]$. Оно получается последовательным выбрасыванием третей из сегмента $[0, 1]$. То есть единичный отрезок прямой делится на три равные части, и средняя часть выбрасывается. С каждой из двух оставшихся частей совершается такая же операция (см. рис. 1).

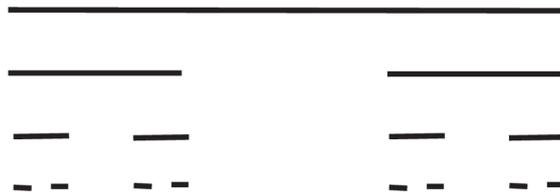


Рис. 1. Пример простейшего фрактала

Продолжая этот процесс, получим, что на n -м шаге имеем $N(\varepsilon) = 2^n$ кусков отрезка длины $\varepsilon = 3^{-n}$. Таким образом, их суммарная длина будет равна $A_n = (2/3)^n$, а при $n \rightarrow \infty$ получаем, что $A_n \rightarrow 0$. Нетрудно понять, что в пределе от отрезка $[0, 1]$ почти ничего не остаётся, а то, что остаётся, называется «канторовым множеством». Действительно, так как длина выброшенных третей равна $1/3 + 2/9 + 4/27 + \dots + (2n - 1)/(3n) + \dots = 1$, то, по существу, «вся длина» отрезка $[0, 1]$ выбрасывается. Однако феноменальность этой процедуры состоит в том, что оставшееся после выбрасывания третей «тощее множество», которое было названо канторовым, оказалось равномоощным континууму $[0, 1]$. Для Георга Кантора это был настоящий шок! Но пример канторова множества мы рассматривали ради того, чтобы посчитать его хаусдорфову размерность, которая равна $d = \ln 2 / \ln 3 = \log_3 2 = 0,631\dots < 1$, так что размерность канторова множества больше, чем у точки, но меньше, чем у линии. Напомним, что аналитические характеристики математических объектов выражаются через алгебраические, топологические и порядковые структуры математики, хотя «дробномерный» мир фракталов является нестандартным обобщением классического понятия размерности, совершенно не похожим на линейность математических объектов с единственным измерением.

Во фрактальных объектах переплелись современные проблемы математического познания в целом — логика и вычисление, независимость от человека и зависимость от него, сложность и простота. Визуализация такого сложного объекта, как множество Мандельброта, стала возможной лишь благодаря современному компьютеру, но это не означает, что его можно полностью изобразить. Поскольку теоретические структуры современной математики для «тонких» моделей слишком сложны, то переход к компьютерному моделированию стимулирует принятие сравнительно «грубых» моделей, которые начали подменять собой реальность. Акцент в этом предположении надо поставить на слове «подменять», так как непосредственно вычислить фрактальные множества Мандельброта и Жюлиа непросто. По мнению специалиста по вычислению фрактальных множеств А.К. Дьюдни, «точность выполнения арифметических операций компьютером может не позволить точно указать точки, которые с

самого начала должны находиться на границе, а когда начинается итерационный процесс, точность ещё более снижается и итерируемые переменные уходят из поля зрения» [5, с. 92]. Тем не менее, считает он, даже с помощью сравнительно несложной программы современный компьютер можно превратить в «своеобразный микроскоп» и наблюдать с его помощью за поведением границы фрактального множества.

Философия фрактальной геометрии рационализирует и конкретизирует восточный принцип «одно во всем и все в одном», указывая на область его применимости и одновременно методологически открывая простоту сложного. Поэтому благодаря этому новому разделу математики, зная ход процессов на малых масштабах, можно открывать возможность их определения на больших масштабах и, соответственно, наоборот. Заметим, что число различных масштабов длины природных объектов для практических целей бесконечно, так как реальность обладает не просто большей сложностью, а, вообще говоря, сложностью другого уровня. Фрактальная геометрия, занимающаяся изучением инвариантов группы «самоаффинных» преобразований, описывает весьма широкий класс природных явлений. С точки зрения принципа дополнительности, вопросы нельзя ставить в плане логических исключений, поскольку «линейная» или «фрактальная» установки наблюдателя оказывают влияние на результаты измерения. Длину и фрактальную размерность измерить одновременно, при одном и том же масштабном преобразовании, вообще говоря, нельзя, поскольку соответствующие утверждения об измерениях дополняют друг друга. Кроме того, разные способы задания размерностей могут конституировать разные математические понятия. Но если предположить «линейность» объекта измерения, выражаемую целой размерностью, то тем самым конструируется определение длины, лишаящей смысла понятие фрактала, так как тогда фрактал теоретически не наблюдаем.

В определённом смысле множество Мандельброта и подобные ему фрактальные множества нельзя считать созданием компьютера, так как вычислительный процесс должен продолжаться бесконечно долго, и поэтому эти множества невозможно окончательно и точно вычислить. Существование, приписанное множеству Мандельброта, можно трактовать как свойство его абсолютной природы. Когда Мандельброт увидел самые первые компьютерные изображения, он счёл полученные размытые структуры результатом сбоя и только потом убедился в том, что они действительно являются частью строящегося множества, хотя сложную структуру множества Мандельброта во всех её деталях не под силу охватить никому, так как её невозможно полностью отобразить на компьютере. Результатом исследования любого математика, работающего на компьютере, будет приближение к единой фундаментальной математической структуре, которая уже существует где-то вне нас. В этом случае даже самая изощёренная математическая интуиция, возможно, представляет собой всего лишь некоторое знание поведения математических объектов в «виртуальной реальности», создающей впечатление человека о пребывании в искусственно созданном мире. На мировоззренческом уровне фрактал уже становится философской категорией, и «по степени сложности традиционно выделяют следующие

типы фрактальных образований: геометрический, алгебраический, стохастический» [6, с. 211]. Философская концепция фрактала Мандельброта изменила и компьютеризованную социокультурную действительность, в которой виртуальная реальность отражает целостное единство субъективного и объективного в человеке.

Наибольшей степенью наглядности обладают фракталы геометрического типа, примерами которых являются снежинка Коха и салфетка Серпинского. Наиболее известными фракталами алгебраического типа, которые называют динамическими, являются множество Мандельброта и множество Жюлиа. В математическом отношении фракталы стохастического типа самые сложные, так как образуются в том случае, когда в процессе случайным образом изменяется какой-либо параметр, например, представление о фрактальных признаках стохастического поведения физической системы возникает при рассмотрении броуновского движения. Современная вычислительная математика, сочетающая в своём методологическом обосновании формализм теоретической математики и конструктивизм прикладной математики, с учётом все более совершенствующихся компьютеров, вообще говоря, не ограничена в своём движении к надёжности и согласованности с имеющимися доказательствами математического существования. Несущую конструкцию современной математики можно представить в виде постепенного накопления «некорректируемых структур», к которым, несомненно, относятся теория чисел, аналитическая и алгебраическая геометрия, дифференциальные и интегральные уравнения, вещественный, комплексный и функциональный анализ и многие другие теории. Отличительным признаком полной структурной стабилизации развитых математических теорий является их аксиоматизация. Но если понимать современную математику только как науку об абстрактных аксиоматических структурах, то большая часть вычислительной математики в нее явно не укладывается. Поэтому в контексте системного подхода к обоснованию математических теорий следует стремиться к их объединению, синтезу и замыканию их, в конце концов, в системную целостность.

С точки зрения философского и математического подходов к обоснованию математических теорий, спекуляции о компьютерных неожиданностях в сфере фракталов постепенно утихают, хотя обнаруженные эффекты не только нетривиальны, но и в определённой степени фундаментальны в современной теории познания. Хотя фрактальные объекты были известны довольно давно, познавательный интерес к ним появился после активной популяризаторской деятельности Бенуа Мандельброта, работавшего в корпорации IBM, благодаря которому математики и информатики сумели открыть удивительный мир, по-новому взглянув на хорошо знакомые математические объекты и явления. Возможно поэтому в конце прошлого века на страницах журнала "Mathematical Intelligencer" прошла дискуссия под названием «Кто же открыл фрактал Мандельброта?» [7]. Хотя Мандельброт опубликовал свою пионерскую работу в 1980 году, некоторые математики, упомянутые в указанной дискуссии, заявляли, что наблюдали множество Мандельброта на дисплее своего компьютера еще раньше, и даже опубликовали соответствующую работу. Но именно Ман-

дельброт развил интуитивное понимание хаусдорфовой размерности до практической реализации концепции фрактальной размерности. Вполне аргументированные возражения Мандельброта сводились к тому, что определения и построения сами по себе «ничего не значат», если при этом не было сказано, почему они важны, а ещё лучше было бы убедить других в этой важности. Но, все равно, не следует утверждать, что новую науку создал один человек, имея в виду его предшественников.

С помощью фрактальных объектов природа на языке математики демонстрирует не просто значительно более высокую степень сложности, соответствующей современному уровню развития науки, а принципиально другой уровень. Когда появляется дополнительная мотивация в контексте новых философских подходов к репрезентации фрактальной геометрии природы, созданной, прежде всего, для нужд естествознания и играющей ведущую роль в возрождении теории итераций, иногда удаётся отойти от шаблона, которому обычно следуют при изучении различных тем классической математики, и вернуться к анализу новых математических объектов с меньшей предубежденностью. Особый интерес в философии математики представляет один из основных принципов методологии современной науки — «принцип нелинейности», согласно которому, любая сложная развивающаяся система не может быть эксплицирована аддитивным способом, поскольку её целостность нельзя описать через суммативность составляющих её частей. Поэтому философская составляющая в практической реализации математики выступает как конструктивный методологический ориентир, что очень важно для формирования познавательного и преобразовательного мировоззрения студентов, поскольку в каждый конкретный период исторической эволюции математического знания можно выявить философско-методологическое сосуществование математических методов, находящихся на разных этапах эволюции.

ЛИТЕРАТУРА

1. Мандельброт Б. Фрактальная геометрия природы. М. : Институт компьютерных исследований, 2002. 656 с.
2. Данилов Ю.А. Красота фракталов // Синергетическая парадигма. Многообразие поисков и подходов. М. : Прогресс-Традиция, 2000. С. 186–190.
3. Секованов В.С. О множествах Мандельброта и Жюлиа для многочленов комплексной переменной // Вестник Костромского государственного университета им. Н.А. Некрасова. 2011. Т. 17, № 3. С. 37–43.
4. Дунаев В.Ю., Курганская В.Д. Фрактальная геометрия природы и архитектура смыслового пространства культуры // Вестник Новосибирского государственного университета. Серия: Философия. 2012. Т. 10, Вып. 4. С. 34–41.
5. Дьюдни А.К. Множество Мандельброта и родственные ему множества Жюлиа // В мире науки. 1988. № 1. С. 88–93.
6. Елхова О.И. Фрактальность виртуальной реальности // Вестник Башкирского университета. 2014. Т. 19, № 1. С. 210–214.
7. Андрианов И. Кто же открыл фрактал Мандельброта? // Знание — сила. 1997. № 11. С. 70–73.

**CONCEPT OF MANDELBROT FRACTAL FROM THE MATHEMATICAL
AND PHILOSOPHICAL POINTS OF VIEW**

V.A. Erovenko

Dr.Sc.(Phys.-Math.), Professor, e-mail: erovenko@bsu.by

Belarusian State University, Minsk, Belarus'

Abstract. The article is devoted to the philosophical and mathematical aspects of the concept of Mandelbrot fractal which was the implementation of the synthesis of geometric and virtual-sociocultural perception of the world revived the experimental mathematics of "nonidealized" objects.

Keywords: Mandelbrot fractal, fractal geometry, fractal dimension, fractal reality modern computer.

ТЕОРИЯ БЕСКОНЕЧНЫХ МНОЖЕСТВ КАНТОРА В КОНТЕКСТЕ ГЕНЕЗИСА ФИЛОСОФИИ МАТЕМАТИКИ

Н.В. Михайлова

к.филос.н., доцент, e-mail: michailova_mshrc@mail.ru

Белорусский государственный университет информатики и радиоэлектроники,
г. Минск, Беларусь

Аннотация. В работе предпринята попытка философско-математического анализа теории бесконечных множеств Кантора, оказавшей существенное влияние на развитие математической науки двадцатого столетия с методологической точки зрения генезиса философии современной математики.

Ключевые слова: теория множеств Кантора, бесконечные множества, философия математики.

В начале XX века математика столкнулась с определёнными трудностями при решении некоторых внутренних задач, проливающих свет на степень точности формализованных теорий. Речь идёт, прежде всего, о проблемах обоснования современной математики, связанных с теорией бесконечных множеств Кантора. Следует подчеркнуть, что математическая теория выдающегося немецкого учёного Георга Кантора (1845 – 1918) оказала существенное влияние не только на самоопределение науки двадцатого века, но и стала общефилософским событием. Он был чрезвычайно настойчив, трудолюбив и одарён в искусствах, играл на скрипке и великолепно рисовал. Очень начитанный и хорошо образованный, всегда наполненный светлой радостью осенявших его мыслей, он не знал усталости. По свидетельству современников, участие импозантного Кантора в любом собрании математиков всегда было захватывающим духовным наслаждением для собравшихся благодаря гениальной интуиции его мысли как в области математических, так и нематематических интересов. Когда математики осознали, что натуральные, вещественные числа и многое другое можно трактовать как множества, то это побудило их заняться теорией множеств. В частности, по поводу фундаментального в математике понятия «множество» можно сказать, что весь опыт становления современной математики и анализ её направлений обоснования показывают, что множества служат тем основным элементарным материалом, из которого строятся все основные математические объекты.

Отсюда вытекает универсальность идеи множества и языка теории бесконечных множеств для всей современной математики. По существу, Кантор выдвинул плодотворную методологическую идею построения всей современной математики на базе теории множеств. Кроме того, понятия теории множеств по степени общности сравнялись с понятиями логики, хотя при этом пришлось

расстаться с привычными нормами мышления. Например, в теории бесконечных множеств высказывание «целое больше своей части» потеряло свой прежний смысл. Философские определения понятия множества содержатся в переписке Георга Кантора с Давидом Гильбертом, где вводится важное понятие «завершённое множество», которое, в философской интерпретации Кантора, является «актуально существующей целостностью». Философско-методологические трудности специальной математической терминологии для профессионалов не так уж сложны, и их всегда можно устранить. Взаимодействие математики и философии в учении Кантора проявлялось в том, что, полемизируя с философами, он использовал новые математические конструкции, пытаясь обосновать ограниченность прежних представлений, а говоря с математиками, был вынужден использовать философскую терминологию в оправдание своих нетрадиционных подходов.

Известно, что древнегреческая философия осознала идею потенциальной бесконечности, но критически отнеслась к актуальной бесконечности, и лишь в рамках христианской культуры идея актуальной бесконечности постепенно становится привычной. В работе «О различных точках зрения на актуально бесконечное» Георг Кантор писал: «Несмотря на существенное различие понятий потенциального и актуального бесконечного — причём первое означает «переменную» конечную величину, «растущую» сверх всяких конечных границ, а последнее — некоторое «замкнутое в себе, постоянное», лежащее по ту сторону всех конечных величин, количество, — к сожалению, слишком часто встречаются случаи смешения этих понятий» [1, с. 210]. Кроме того, как считал сам Кантор, часто происходит смешение двух форм актуально бесконечного другого рода, когда смешивается «трансфинитное» с «абсолютным». Их отличие в том, что первое следует, конечно, мыслить «бесконечным», хотя еще доступным увеличению, а второе следует мыслить «недоступным увеличению» и поэтому «математически неопределимым». Канторовскую теорию множеств можно охарактеризовать как теорию конечных и актуально бесконечных множеств, в которой актуализируются основные принципы, принимаемые математиками в качестве верных утверждений относительно множеств. Основной предмет канторовской теории множеств — это бесконечные множества, существование которых предполагается. С тех пор вся философия математики колеблется в своих формах выражения между математикой и философией. Гильберт считал теорию Кантора «высочайшим проявлением математического гения». Хотя канторовская теория множеств стала «миром», вместившим почти всю математику, постепенно угасает тот восторг, который в своё время охватил многих математиков после знаменитого канторовского «прорыва в бесконечность».

Методологическая сложность ситуации состоит в том, что теория бесконечных множеств привнесла в математику целый набор частных случаев актуальной бесконечности, большинство из которых нельзя разумно интерпретировать в реальном мире. У Кантора множество рассматривается как «единство», в котором нет взаимодействия элементов, а имеет место лишь внешняя унификация, а любое бесконечное множество может иметь столько же элементов, сколько и некоторые его собственные подмножества, то есть его части, отличные

от множества. Даже в книгах по математике, посвящённых «наивной теории множеств», избегают точного определения понятия множества. Роль понятия «множество» в математике была осознана лишь во второй половине XIX века. В процессе размышления над логическими основами математики и её структурой осознавалась важность понятия множества, особенностью которого, в частности, является то, что оно не требует вычислений. При этом Кантор заложил основы теории абстрактных множеств и внёс существенный вклад в основание математического анализа и в изучение множества вещественных чисел. Следует отметить, что со времён Аристотеля философы отвергали концепцию «завершённых бесконечностей» в основном из-за логических парадоксов, к которым, как казалось, они приводят. Поэтому, чтобы избежать подобных философских возражений, математики стремились различать бесконечность, рассматриваемую в качестве завершённой величины, и бесконечность, рассматриваемую как потенциально осуществимый процесс, например, сходимости членов ряда, стремящихся к некоторому пределу, то есть бесконечность рассматривалась как способ выражения математического понятия предела последовательности.

В частности, Кантор предположил, что всякое иррациональное число может быть представлено бесконечной последовательностью рациональных чисел. Но «самое значительное достижение Кантора состояло в доказательстве того, что не все бесконечные множества количественно эквивалентны, т. е. имеют одинаковую мощность, а потому их можно сравнивать друг с другом» [2, с. 76]. Другими словами, было впервые показано, что не все бесконечные множества имеют одинаковый «размер», или, как говорят об этом математики, «мощность». Идеи Георга Кантора оказались настолько радикальными и неожиданными для математики и так противоречащими интуиции, что Анри Пуанкаре в своё время даже назвал теорию трансфинитных чисел «болезнью», от которой, как он надеялся, математика когда-нибудь должна будет излечить. Понятие трансфинитных чисел в математике представляет собой распространение понятия порядкового числа на бесконечные множества, также как обобщение понятия количественного числа приводит к соответствующему понятию мощности множества. Свои индексы для обозначения множеств фактически образовывали у Кантора новый вид чисел, и он назвал их трансфинитными числами, которые являются самостоятельным и систематическим обобщением натуральных чисел. Для понимания такого философско-критического отношения к работам Кантора достаточно привести следующий пример. Согласно его теории множеств, несмотря на «плотность» расположения рациональных чисел на прямой и «разреженность» множества целых чисел, может показаться крайне противоречащим интуиции то, что эти два множества оказываются количественно эквивалентными.

Даже сам Георг Кантор сначала воздерживался от введения трансфинитных чисел, обоснованно считая, что идею актуальной бесконечности нельзя сформулировать непротиворечиво, поэтому ей нет места в строгой математической науке. Но именно эти первоначальные сомнения позволили Кантору предвосхитить критическую оппозицию разных сторон, тем самым дав ему возможность «вооружиться» философскими и математическими аргументами,

преодолев своё «предубеждение» в отношении трансфинитных чисел. Математическая сущность трудов Кантора состоит в том, что, разработав арифметику трансфинитных чисел, он придал математическое содержание идее актуальной бесконечности. В существовании актуально бесконечных множеств был убеждён Готфрид Лейбниц, считавший, что природа не только не терпит бесконечного, а напротив, выражает свою любовь к ней, чтобы таким способом наглядно продемонстрировать совершенство Творца. Кроме того, считая эквивалентными любые два множества, для которых существует взаимно однозначное соответствие, теория множеств не принимает во внимание природу элементов этих множеств. С одной стороны, это позволило применить результаты математики к разнообразным абстрактным математическим объектам, равнозначным с точки зрения теории множеств, а с другой стороны, приходилось постоянно преодолевать шлейф сомнения насчёт того, насколько содержательны полученные математические утверждения. Всякое множество чисел, элементы которого можно расположить один за другим, то есть фактически сосчитать, используя для этого множество целых положительных чисел, Кантор назвал «счётным множеством». Счётные множества, например, натуральные или рациональные числа, допускают трактовку с соответствующими модификациями, как в смысле актуальной бесконечности, так и в смысле потенциальной бесконечности, а несчётные множества можно интерпретировать только в смысле актуальной бесконечности.

Даже метод математической индукции применим только к счётным множествам, но не применим к несчётным множествам. Ярким примером последнего является континуум, но в отличие от счётных множеств, где точки или элементы этих множеств первичны по отношению к целому, для континуума первичным является целое множество. Невзирая на профессиональные возражения против актуализации бесконечности, Георг Кантор сформулировал тезис: «существует актуальная бесконечность», то есть все бесконечные множества современной математики, включая любую аксиоматическую теорию множеств, являются актуально-бесконечными множествами. Полагая, что потенциальная бесконечность в действительности зависит от логически предшествующей ей актуальной бесконечности, Кантор не только стал изучать бесконечные множества как «готовые», но и занялся задачей классификации бесконечных множеств. Актуальная бесконечность, по определению Кантора, есть некоторая вещь-для-себя. «И эта «вещь-для-себя» никогда не становится «вещью-для-нас». Мы познаем её только на основании её, так сказать, «проявлений» и «внешних» свойств» [3, с. 63]. Концепция Кантора понятия бесконечности показывает, что «платоновские симпатии» создателя теории множеств, в специфической форме «христианизированного платонизма», имели свои границы. Они навязывались актуальной бесконечностью, как предметом рассмотрения, и потому были обусловлены соответствующей традицией «приручения» этого объекта познания, понимание которого основывалась на двух дополнительных потоках идей, один из которых был чисто математического содержания, а другой состоял из философских интерпретаций. Основная идея математического проекта Кантора сводилась к установлению взаимнооднозначного соответствия

между множествами.

В соответствии с этим он определил бесконечное множество как такое множество, которое можно поставить во взаимнооднозначное соответствие со своим собственным подмножеством, отличным от всего множества. Кроме того, Кантор построил одну из конструкций вещественных чисел, доказал несчётность континуума и равномощность пространств разного числа измерений и привёл доказательство существования трансцендентных чисел. С одной стороны, для математики понятие бесконечности важно и является необходимым, так как большинство математических утверждений, не имеющих отношения к бесконечности, с точки зрения теоретического оснащения математики конца двадцатого века, можно считать «тривиальными». С другой стороны, появление большого комплекса вычислительных наук, в силу своей специфики, должно учитывать финитность ресурсов вычислительной компьютерной техники, а это формирует естественное стремление к «финитизации» математики, что явно проглядывает, например, у интуиционистов и конструктивистов. Начиная с Аристотеля, математики проводили различие между актуальной бесконечностью и потенциальной бесконечностью, которые Кантор называл также «несобственной бесконечностью» и «собственной бесконечностью». Обе они, в свою очередь, делятся на бесконечно большие и бесконечно малые, появившиеся с возникновением математического анализа. Кроме того, бесконечность в математике может проявляться в форме величины, числа, множества и последовательности, на которые распространяются арифметические, геометрические, алгебраические и логические процедуры или правила действий. Но мы имеем дело с актуальной бесконечностью, например, когда, по мнению Георга Кантора, рассматриваем натуральный ряд так, как если бы все натуральные числа были даны нам одновременно, или в частности, когда рассматриваем произвольное функциональное пространство как бесконечное множество элементов-точек.

Понятие о потенциальной бесконечности состоит в рассмотрении бесконечной совокупности математических объектов, исходя из процесса последовательного построения этих объектов. Каноническим примером может служить бесконечность натурального ряда, рассматриваемого как процесс образования натуральных чисел, учитывая, что построение слишком больших натуральных чисел в реальных условиях невозможно. Понятие об актуальной бесконечности состоит в рассмотрении бесконечной совокупности математических объектов как завершённой совокупности, независимо от процесса построения этих объектов. В связи с этим академик П.С. Александров утверждал, что во второй половине XIX века не существовало математика, оказавшего наибольшее влияние на развитие математических теорий, чем создатель абстрактной теории множеств Георг Кантор: «Огромное большинство математиков той эпохи считали, что работы Кантора вообще относятся не к математике, а в лучшем случае к философии. Но история науки, как и во многих более ранних аналогичных случаях, вынесла свой приговор, в силу которого сделанное Кантором является неотъемлемой и фундаментальной частью математики» [4, с. 291]. В частности, для профессиональных математиков понятие «бесконечное» — это специфический элемент математического метода или «метафора конечного», поскольку

при изучении «бесконечное» заменяется множественным числом «конечного». Заметим, что представители разных направлений философии математики стараются избежать явных определений конечности и бесконечности, в отличие от попыток математиков конца XIX — начала XX веков. Хотя и на этом, и на предшествующих этапах развития современной математики так и не были выработаны дефиниции конечности и бесконечности, которые удовлетворили бы большинство учёного сообщества философов и математиков, в конце девятых годов XIX века Кантор осознает наличие парадоксов теории множеств. К этому времени относятся его работы по созданию теории трансфинитных чисел, и тогда же Георг Кантор обнаруживает первые парадоксы теории множеств.

После обнаружения парадоксов теории бесконечных множеств, которые Кантор обсуждает с Гильбертом, математики и логики предприняли отчаянные попытки для «профилактического ремонта» всей теории в целом. Даже обнаруженные изъяны теории, успешно применявшейся в математике XX века, сами математики предпочитали называть «парадоксами» или «антиномиями», хотя они и производили тягостное впечатление противоречий. Эти противоречия заключались в том, что в рамках теории множеств можно было предъявить два высказывания о множествах, одновременно истинных и таких, что одно из них являлось отрицанием другого. Наибольшую известность среди этих парадоксов получил «парадокс Рассела». В надежде на скорое устранение этих «нелепостей в математике» сами математики пытались смягчить неприятную для них ситуацию «чисто лингвистическими средствами». Но предпринятый «ремонт» теории, к сожалению, оказался лишь «косметическим» и не дал никаких надёжных гарантий на будущее, поэтому теорию множеств пришлось «деформировать» ценою её естественности и простоты. Поэтому Кантор, предпринимая попытки осмыслить и защитить свои исследования бесконечного, пишет цикл философских работ, в частности, работу «Принципы теории порядковых чисел», опубликованную уже после его смерти. Все структуры, изучаемые в математических теориях, основанных на канторовской теории множеств, априорно жёстко заданы. Например, после того, как было обнаружено, что «множество всех множеств» является нечётко заданной совокупностью, оно было «запрещено», хотя аксиома выбора, постулирующая существование функции выбора, имеющей явно нечёткий характер, до сих пор активно используется в современной математике.

Бесконечность процесса познания, вообще говоря, не обязана означать чрезмерного обилия его событий. Математики уже сталкивались с такими феноменами в обосновании своей науки, для понимания которых они нуждаются в какой-то другой более глубокой уверенности. Историки науки указывают на определённый религиозный пафос научной деятельности Кантора, хотя в математических работах Кантора практически нет обращений к теологии или теологам. Когда Кантор отказался читать лекции по математике, то стал читать философию, хотя его лекции по философии не пользовались успехом, и он отказался от этой затеи. С этого времени его научные интересы смещаются в область философского обоснования теории множеств. Анализируя теорию множеств, Ю.И. Манин высказался о Канторе так: «Встроенная в теорию ауто-

референтность и мощное расширение области, доступной математической интуиции (принципы построения новых множеств), вносят дополнительные штрихи в эту картину художественной смелости, соединённой с самоограничением» [5, с. 110]. Заметим, что несмотря на некоторые философско-математические проблемы теоретико-множественная концепция позволила разобраться в разнообразии возможных математических теорий, а также систематизировать и структурировать их. Отметим также, что одним из требований, предъявляемым философами математики к системной классификации, является структурированность, которая связана с генезисом методологической концепции абстрактной математической структуры. Французский математик Рене Том считал, что одним из важнейших философских утверждений, на которые должна опираться современная математика, является утверждение о существовании математических структур независимо от человеческого разума. Это положение он объясняет тем, что философские надежды показать, как математические структуры естественно вытекают из иерархии множеств, их подмножеств и их комбинаций не осуществились.

Поэтому нельзя ни по каким разумным причинам отказаться от мысли, что важные математические структуры существуют во внешнем мире, и их огромное многообразие находит единственное оправдание в реальности. Обратим также внимание на мнение влиятельной группы математиков, объединившейся под именем Бурбаки, которая считала, что математика говорит не о специфических математических объектах, а только о структурах. Точка зрения группы Бурбаки состояла в том, что существует одна математика с разнообразными математическими дисциплинами, а объединяют эти разные математические разделы в единую науку понятие «математическая структура» и аксиоматический метод. Структуру, например, можно интерпретировать как список операций, отношений и их свойств, которые обычно выражены аксиомами и сформулированы так, что представляются как свойства, которым удовлетворяет некоторый класс специфических математических объектов, возможно, даже различных между собой. Наиболее важной чертой математических структур является то, что, пользуясь аксиоматическим методом, они репрезентируют, образно говоря, философскую «экологию мысли». Теоретико-множественная аксиоматика позволяет с единой точки зрения рассмотреть строение математических теорий, предметное содержание которых раскрывается через систему аксиом. Поэтому можно заключить, что в настоящее время внутренняя эволюция современной математики, которая является открытой, связанной и целостной системой, вопреки видимости более чем когда-либо упрочила единство своих направлений и создала своего рода фундаментальное ядро, что не всегда заметно в её внешних проявлениях.

Математиков иногда упрекают в том, что они присвоили себе право решать, какие утверждения о бесконечных множествах справедливы, поставив себя на место творца, который эти множества может созерцать. Анализируя работы по теории множеств, историки науки сделали следующие три главных вывода. Во-первых, основные понятия теории множеств «имеют математические истоки», и интерес к канторовской теории множеств проявили, прежде всего, сами ма-

тематики. Во-вторых, в силу абстрактности и теоретической общности своих понятий, в теории множеств «звучат философские мотивы», не случайно Кантор в своих философских изысканиях рассматривал такие категории, как бесконечность и континуум, интересующие философов со времён античности. В связи с этим уместно заметить, что в ранний период становления теории множеств, отсутствие в работах Георга Кантора позитивных результатов и обилие философских рассуждений служило препятствием в её признании со стороны крупных и влиятельных математиков. В-третьих, следует отметить, что, несмотря на решающий вклад самого Кантора, теория множеств — это совместное творение многих выдающихся математиков, то есть она оказалась закономерным этапом в развитии математики XIX века. Теория множеств, созданная Георгом Кантором, была самым глубоким проникновением в природу бесконечного и поэтому имела более философский, чем математический характер, в духе платоновской онтологии. В качестве дополнительной аргументации напомним о влиянии существования самостоятельного «мира идеальных объектов», содержащего абстрактные понятия математики, о котором говорили Кантор и его последователи.

Георг Кантор провозгласил принцип, согласно которому «сущность математики заключается в её свободе», выражая тем самым свою философскую установку на то, что всякое свободное творение математического разума имеет идеальное платоническое существование. Но если бы ему был предоставлен такой выбор, то «чистую математику» Кантор, по его собственному признанию, назвал бы «свободной» математикой, поскольку математика в своём развитии совершенно свободна. Она связана лишь тем условием, что её понятия должны быть «свободны» от внутренних противоречий и должны находиться в неизменных, установленных определениями отношениях к образованным и испытанным раньше понятиям. Принцип свободы Кантора удобен в философском отношении, так как он не стесняет математического творчества, заранее оправдывая любые абстрактные построения. Георг Кантор не видел в этом принципе также какой-либо опасности для науки и считал, что большая опасность заключается во всяком «излишнем ограничении» математического стремления к творчеству, поскольку в пользу таких ограничений нельзя привести убедительных доводов, исходя из сущности науки.

Хотя теория бесконечных множеств лежит в основе всех математических наук, и практически все математики верят в то, что она непротиворечива, Георг Кантор также говорил, что свобода математики — это вовсе не произвол. Генезис философии математики показывает, что такого никогда и не было. Поэтому этот философский принцип Кантора вовсе не стесняет математического творчества, тем самым как бы оправдывая возможность построения новых абстрактных теорий, а многие математики по-прежнему продолжают придерживаться философской и математической сути канторовского принципа «свободы математики». Поэтому и знаменитое утверждение Кантора о «свободе математики» отчасти утратило свой пафосный и гордый смысл «свободного полёта математической мысли», постепенно выходя из употребления. Но необходимо признать, что построение мира классической математической науки было бы

невозможно без допущения гипотетически мощного интеллекта, не ограниченного ни пространством, ни временем.

ЛИТЕРАТУРА

1. Кантор Г. О различных точках зрения на актуально бесконечное // Парадоксы бесконечного. Минск : Издатель В.П. Ильин, 2000. С. 204–213.
2. Даубен Дж.У. Георг Кантор и рождение теории трансфинитных множеств // В мире науки. 1983. № 8. С. 76–86.
3. Катасонов В.Н. Боровшийся с бесконечным. Философско-религиозные аспекты генезиса теории множеств Г. Кантора. М. : Мартис, 1999. 207 с.
4. Александров П.А. О вкладе Георга Кантора в математику // Историко-математические исследования. 1983. Вып. 27. С. 290–292.
5. Манин Ю.И. Георг Кантор и его наследие // Математика как метафора. М. : Изд-во МЦНМО, 2008. С. 110–124.

THE THEORY OF INFINITE SETS OF CANTOR IN THE CONTEXT OF GENESIS OF PHILOSOPHY OF MATHEMATICS

N.V. Michailova

Ph.D.(Philosophy), Associate Professor, e-mail: michailova_mshrc@mail.ru

Belarusian State University of Informatics and Radioelectronics, Minsk

Abstract. The article shows an attempt of philosophical and mathematical analysis of the theory of infinite sets of Cantor which had a significant impact on the development of the mathematics of the twentieth century from a methodological point of view of the genesis of philosophy of modern mathematics.

Keywords: Cantor's set theory, infinite sets, philosophy of mathematics.

PARADOX OF CHOICE: A POSSIBLE EXPLANATION

V. Kreinovich

Ph.D. (Phys.-Math.), Professor, e-mail: vladik@utep.edu

O. Kosheleva

Ph.D. (Phys.-Math.), Associate Professor, e-mail: olgak@utep.edu

University of Texas at El Paso, El Paso, TX 79968, USA

Abstract. At first glance, we would expect that the more choices we have, the happier we will be. Experiments show, however, that when the number of choices increases, customers become *less* happy. In this paper, we provide a possible explanation for this paradox.

Keywords: paradox of choice, customer choice, customer satisfaction, uniform distribution, independent random variables.

1. Formulation of the Problem

Intuitively, the more choices we have, the better we should feel about it. In practice, however, if we add many additional choices to the current ones, the customers become less happy; see, e.g., [2] and references therein. The decrease of happiness is relatively small — so small that while in many cases, it is barely above the level of statistical significance, and often below this level [3], but in many cases, it is there. How can we explain this counterintuitive phenomenon?

2. Possible Explanation

How to describe customer choice? There are many possible settings of the user choice. For example, a customer has a fixed amount of money that he or she is willing to spend on a certain product, and the customer is looking for the best value for this amount. In this case, the customer is looking for the largest value per unit price.

Alternatively, a customer may be interested in buying a certain product, and he/she is looking for the cheapest option (among those options that satisfy his/her requirements). In this case, the customer is looking for the smallest value of the price per unit, i.e., equivalently, for the largest number of units per dollar.

There may be other possible settings. In all these settings, a customer wants to maximize his or her gain (or, equivalently, minimize his or her loss). In the general case, let us denote the quantity that we want to maximize by q , and the values of this quantity corresponding to different choices by q_1, \dots, q_n .

Few choices vs. multiple choices: the main difference. When the number n of available choices is small, the customer can simply consider all the options and select the one with the largest value of the desired quantity $\max(q_1, \dots, q_n)$.

The decrease of happiness starts when the number of choices becomes so large that it is not realistically possible to seriously consider all these choices in detail — a situation often happening in supermarkets. In such situations, since the customer cannot consider all the options, he/she considers only *some* of these options; among the considered options, the customer selects the one with the largest value q . Let us denote the number of considered options by k . Without losing generality, let us assume that the considered options are q_1, \dots, q_k . In this case, the resulting quality is equal to $\max(q_1, \dots, q_k)$.

Paradox of choice reformulated in precise terms. In these terms, the paradox of choice can be reformulated as follows. We consider two possible situations:

- in the first situation, the number n of choices is relatively small, so, among the options with values q_1, \dots, q_n , the customer selects the option with quality $Q = \max(q_1, \dots, q_n)$;
- in the second situation, the number n of choices is large, so, among the options with values q'_1, \dots, q'_n , the customer selects an option with quality $Q' = \max(q'_1, \dots, q'_k)$ for some $k < n$.

The empirical fact is that even when $Q = Q'$ — i.e., when the selected product is of the same quality in both cases — a customer is usually less happy in the second situation.

Towards a possible explanation. In general, the values of the quality q corresponding to different products are bounded. Let us denote the lower bound by \underline{q} and the upper bound by \bar{q} . The actual quality of different choices is randomly distributed in the interval $[\underline{q}, \bar{q}]$.

In general, we have no reason to believe that some values from this interval are more frequent than others. So, it is reasonable to assume that all the values from this interval are equally probable, i.e., that we have a uniform distribution on this interval; see, e.g., [1].

Similarly, we have no reason to believe that there is a correlation between different options, so all these quantities can be considered independent [1].

If we choose between n choices, then the resulting quality is equal to $Q = \max(q_1, \dots, q_n)$. What is the expected value of Q ? To find the expected value, let us find the corresponding cumulative distribution $F(q) \stackrel{\text{def}}{=} \text{Prob}(Q \leq q)$.

The maximum Q of k values q_i is smaller than or equal to a given number q if and only if each of these values is $\leq q$. Thus, due to independence assumption (see, e.g., [4]):

$$\begin{aligned} F(q) &\stackrel{\text{def}}{=} \text{Prob}(Q \leq q) = \text{Prob}((q_1 \leq q) \& \dots \& (q_n \leq q)) = \\ &= \text{Prob}(q_1 \leq q) \cdot \dots \cdot \text{Prob}(q_n \leq q). \end{aligned}$$

For the uniform distribution, $\text{Prob}(q_i \leq q) = \frac{q - \underline{q}}{\bar{q} - \underline{q}}$, so $F(q) = \left(\frac{q - \underline{q}}{\bar{q} - \underline{q}}\right)^n$. Thus, the corresponding probability density function $f(q)$ has the form

$$f(q) = \frac{dF(q)}{dq} = n \cdot \frac{(q - \underline{q})^{n-1}}{(\bar{q} - \underline{q})^n}.$$

Therefore, the mean grade $E[Q]$ is equal to

$$E[Q] = \int_{\underline{q}}^{\bar{q}} q \cdot f(q) dq = \int_{\underline{q}}^{\bar{q}} q \cdot n \cdot \frac{(q - \underline{q})^{n-1}}{(\bar{q} - \underline{q})^n} dq.$$

By introducing a new variable $x \stackrel{\text{def}}{=} q - \underline{q}$, for which $q = \underline{q} + x$, we can explicitly compute the corresponding integral, and get

$$E[Q] = \underline{q} + \frac{n}{n+1} \cdot (\bar{q} - \underline{q}).$$

Let us show that this formula enables us to explain the paradox of choice.

Resulting explanation. When we have a small number of choices, we select the option with the largest value Q . Since we did consider all available options, we know that this is the best choice we could have made.

When we have a large number of options n , then we select the value $Q = \max(q_1, \dots, q_k)$ for some $k < n$. The expected value of this choice is equal to

$$Q = \underline{q} + \frac{k}{k+1} \cdot (\bar{q} - \underline{q}).$$

In this case, we did not consider all available options, and thus, we are not sure that the choice we made is the best possible one — maybe we could get a better result if we considered more options.

Theoretically, if we were able to consider all n options in detail, we would be able to get an option with the average quality of

$$Q_0 = \underline{q} + \frac{n}{n+1} \cdot (\bar{q} - \underline{q}).$$

Since $n > k$, we have $\frac{1}{n+1} < \frac{1}{k+1}$, thus,

$$\frac{n}{n+1} = 1 - \frac{1}{n+1} > 1 - \frac{1}{k+1} = \frac{k}{k+1},$$

and $Q_0 > Q$.

So, while we got exactly the same quality Q as in the first case, we also know that:

- in the first case, we did select the best of available options, while

- in the second case, we could have attained better quality if we tested more options.

Since our goal is to maximize quality, and in the second case, we know that we did not reach the maximum – so that we could, e.g., have gotten more value per dollar – we thus naturally feel less happy.

This also explains why the difference is small: the potential relative increase

$$\frac{Q_0 - Q}{Q} = \frac{\frac{1}{k+1} - \frac{1}{n+1}}{1 - \frac{1}{k+1}}$$

is small: e.g., for $k = 10$ and $n = 100$, it is about 10%, not much.

Acknowledgments

This work was supported in part by the National Science Foundation grants HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and DUE-0926721.

REFERENCES

1. Jaynes E.T., Bretthorst G.L. Probability Theory: The Logic of Science. Cambridge University Press, Cambridge, UK, 2003.
2. Schwarz B. The Paradox of Choice: Why More Is Less. Harper Perennial, New York, 2005.
3. Scheibehenne B., Gerfeneder R., Todd P.M. Can there ever be too many options? A meta-analytic review of choice overload // Journal of Consumer Research. 2010. Vol. 37. P. 409–425.
4. Sheskin D.J. Handbook of Parametric and Nonparametric Statistical Procedures. Chapman and Hall, CRC Press, Boca Raton, Florida, 2011.

ПАРАДОКС ВЫБОРА: ВОЗМОЖНОЕ ОБЪЯСНЕНИЕ

В. Крейнович

к.ф.-м.н., профессор, e-mail: vladik@utep.edu

О. Кошелева

к.ф.-м.н., доцент, e-mail: olgak@utep.edu

Техасский университет в Эль Пасо, США

Аннотация. С первого взгляда, мы можем ожидать, что чем больше выбора у нас есть, тем счастливее мы будем. Однако, эксперименты показывают, что с ростом числа выборов, потребители становятся менее счастливыми. В данной статье представлено возможное объяснение этого парадокса.

Ключевые слова: парадокс выбора, выбор потребителя, удовлетворённость потребителя, равномерное распределение, независимые случайные переменные.

SEARCH PROBLEMS WITH A PROMISE AND GRAPH ISOMORPHISM

G.A. Noskov

Dr.Sc.(Phys.-Math.), S.R., e-mail: g.noskov@googlemail.com

Institute of Mathematics, SORAN

Abstract. The Graph isomorphism problem is considered from the point of view of the theory of "problems with a promise", developed by Even, Selman and Yacobi [4]. The "tau-invariant" of graphs is studied and with its help the search graph isomorphism problem is solved for asymptotically almost all graphs.

Keywords: decision problem, search problem with a promise, graph isomorphism.

Introduction

In this paper we consider the search graph isomorphism problem SGI in the context of "search problems with a promise" as they defined in [4,6,7].

In existing literature (see for instance [9]) the decision and search variants of the graph isomorphism problem are formulated as follows:

P1: Given two graphs G and H with n vertices each, decide whether they are isomorphic.

P2: Given two graphs G and H , decide whether they are isomorphic, and if so, construct an isomorphism from G to H .

Note that P2 contains P1 as a subproblem. Recently, A.N. Rybalov suggested to isolate the search version from the decision version as follows [12]. By definition, the input set for search graph isomorphism problem SGI is the set of all pairs of isomorphic graphs. Given two isomorphic graphs G and H , one needs to construct the isomorphism between G and H . Thus in Rybalov's setup of the SGI the input graphs are already assumed to be isomorphic, whereas in version P2 above the input graphs are arbitrary. The seemingly unusual input set in SGI can be elegantly explained in the framework of "promise problems" in the sense of [4,6]. Promise problems are a natural generalization of search and decision problems, where one explicitly considers a set of legitimate instances (rather than considering any string as a legitimate instance). Informally, a promise decision problem has the following structure: input x , promise $P(x)$, property $R(x)$, where P and R are unary predicates. An algorithm solves the promise problem if, given an input x , it answers the question whether $R(x)$ given that $P(x)$. The behavior of such an algorithm may be arbitrary on instances x for which the promise P is false [10].

In this paper we formulate the search graph isomorphism problems "with a promise" and study their reducibility and generic solvability in polynomial time. We make use the "type-invariant" of a graph introduced in [2, 13, 14]. Our main result is Theorem 5, which presents the polynomial-time algorithm A , solving the problem $SGI(P_{isom}, R)$ for asymptotically almost all inputs $P_{isom} \cap P_{oblique}$.

In Section 1 we give the necessary definitions about computation problems. Section 2 is devoted to the formulation of the search graph isomorphism problem (with a promise). Section 3 contains definitions of computational problems in the case of graphs. Section 4 contains the proof of polynomial reducibility of search graph isomorphism problem SGI to the graph isomorphism problem. In Section 5 we discuss the main tool - the graph invariant τ and oblique graphs. In Section 6 the efficient solvability of our problems in case of oblique graphs. In Section 7 we solve the SGI problem in generic case. Finally, in the last section we apply our results to probabilistic algorithms.

1. Decision and search with a promise

We start with the standard definitions of decision and search computational problems as they presented in [11] and [7]. Let $I = \{0, 1\}^*$ be the set of all **words** (=binary strings) in the alphabet $\{0, 1\}$. We consider algorithms as means of computing functions. Specifically, an algorithm A computes the function $f_A : \{0, 1\}^* \rightarrow \{0, 1\}^*$ defined by $f_A(x) = y$ if, when invoked on input x , algorithm A halts with output y . We associate the algorithm A with the function f_A computed by it; that is, we write $A(x)$ instead of $f_A(x)$.

Decision problems. Let's denote by I^k the k -th direct power $I \times \dots \times I$ of I for $k \geq 1$. A **decision problem** for a subset (=language) $L \subseteq I^k$ is to determine for a given tuple $w \in I^k$ whether w belongs to L or not. An algorithm A solving this problem is the **decision algorithm** for L , and in this case the decision problem for L , as well as the language L , is called **decidable**.

If L is decidable and additionally there are positive constants c, q such that for every instance $x \in I^k$ the algorithm A determines the membership of x to L in at most $c|x|^q$ steps, then the decision problem for L , as well as L is called **polynomial-time decidable** (or **decidable in polynomial time**). Here, for a tuple $x = (x_1, \dots, x_k)$ we let denote by $|x|$ the maximum $\max |x_i|$ of lengths of words x_i .

Decision problems with a promise. More general class of partial decision problems was introduced in [4] under the name of "promise problems". Formally, a **partial decision problem** is a pair of decidable subsets (L, P) of I^k , where P is the set of **allowed** or **promised** tuples and $\bar{P} = I^k - P$ is the set of **disallowed** tuples. The promise problem (L, P) is **solved by algorithm** A if for every $x \in P \cap L$ it holds that $A(x) = 1$ and for every $x \in P - L$ it holds that $A(x) = 0$. Shortly,

$$A(x) = \begin{pmatrix} 1 & \text{if } x \in P \cap L, \\ 0 & \text{if } x \in P - L \end{pmatrix}. \quad (1)$$

Thus, the algorithm A is required to distinguish **yes-instances** $P \cap L$ from **no-**

instances $P - L$ and A is allowed to have arbitrary behavior on inputs \bar{P} that are neither yes-instances nor no-instances. A set L_1 is called a **solution** to the partial decision problem (L, P) if $L_1 \cap P = L$. Clearly, when $P = I^k$ we have the notion of a standard decision problem. A partial decision problem (L, P) is **polynomial-time decidable** if there is a polynomial-time decidable solution L_1 of it.

Remark 1. In [7, 8] a more restrictive definition is used, namely the promise problems (L, P) are considered with the condition $L \subseteq P$. It is not enough for our purposes, see Section 5.

Search problems. A **search computational** problem can be described by a binary relation $R \subseteq I^k \times I^l$ for some fixed $k, l \geq 1$. The problem is "given an input $x \in I^k$, find y such that $R(x, y)$ holds, if such y exists". More precisely, one requires, for a given $x \in I^k$ to decide first whether there exists $y \in I^l$ such that $R(x, y)$ holds, and only after that to find such y if it exists.

Let us consider R as a multi-valued function $R(x) = \{y : (x, y) \in R\}$. The associated **solution set** is $S_R = \{x : R(x) \neq \emptyset\}$. A function $f : I^k \rightarrow I^l \cup \{\perp\}$ is called a **branch** of $R(x)$ if $f(x) \in R(x)$ for all $x \in S_R$ and $f(x) = \perp$ for all $x \notin S_R$ (thus $f(x) = \perp$ indicates that x has no solution). We say that the branch function f of R **solves** the search problem of R . As before, we write $A(x)$ for $f(x)$ for an algorithm A , computing the function f .

Note, that the search problem for R contains the decision problem for a **solution set** $S_R = \{x : R(x) \neq \emptyset\}$ as a subproblem. Indeed, if A computes the branch of R , then $x \in S_R$ iff $A(x) \neq \perp$.

A relation $R \subseteq I^k \times I^l$ is **polynomially bounded** if there exists a polynomial p such that for every $(x, y) \in R$ it holds that $|y| \leq p(|x|)$. The search problem of a polynomially bounded relation $R \subseteq I^k \times I^l$ is **efficiently solvable** if there exists a branch function f of R which is polynomial-time computable. In this case an algorithm A , computing this function, is called a **polynomial-time algorithm** for the search problem R . We denote by \mathcal{PF} the class of polynomially bounded search problems that are efficiently solvable [8].

Search problems with a promise [8, p.143]. A search problem with a **promise** consists of the input set I^k , a binary relation $R \subseteq I^k \times I^l$ and a **promise set** $P \subseteq I^k$. Such a problem is also referred to as the **search problem R with promise P** and is denoted by (P, R) . The search problem (P, R) with promise P is **solved** by algorithm A if for every $x \in P$ it holds that $(x, A(x)) \in R$ if $x \in S_R$ and $A(x) = \perp$ otherwise. Thus, the restriction function $A|P$ is the branch of the relation $R \cap (P, I^l)$. "We stress that nothing is required of the solver in the case that the input violates the promise (i.e., $x \notin P$); in particular, in such a case the algorithm may halt with a wrong output" [8, p.143]. Note that the **full search problem** (I^k, R) with a promise I^k is the standard search problem. And at the other extreme there is the so called **candid search problem** (S_R, R) with promise S_R .

The **time complexity of A on inputs in P** is defined by

$$T_{A|P}(n) = \max\{t_A(x) : x \in P \cap \{0, 1\}^n\}, \tag{2}$$

where $t_A(x)$ is the running time of $A(x)$. A problem (P, R) is **polynomially-time decidable** if there is an algorithm A such that function $T_{A|P}$ is bounded by a polynomial in n . "In this case, it does not matter whether the time complexity of A is defined on inputs in P or on all possible strings. Suppose that A has (polynomial) time complexity T on inputs in P ; then we can modify A to halt on any input x after at most $T(|x|)$ steps. This modification may only affect the output of A on inputs not in P (which are inputs that do not matter anyhow). The modification can be implemented in polynomial time by computing $t = T(|x|)$ and emulating the execution of $A(x)$ for t steps." [7, p.88].

Algorithms which are polynomial-time on asymptotically almost all inputs. By the very definition of the search problem (P, R) , it becomes easier when decreasing the promise P . If the given problem (P, R) is hard, then it makes sense to consider a **restriction** problem (P', R) with a promise set $P' \subset P$ and ask whether (P', R) is polynomial-time decidable. This may happen when P' is small enough, for instance when P' is finite. Thus, it is highly desirable to find out a polynomial-time decidable restriction (P', R) with P' being maximally close to P . For a rigorous definition of closedness we need a **size function** on the promise set P , i.e. any computable function $x \mapsto \|x\| \in \mathbb{N}$ such that for every $n \in \mathbb{N}$ the set $P_n = \{x \in P : \|x\| = n\}$ is finite. Furthermore, we need an **ensemble of distributions** $\mathcal{D} = (D_n)$ with D_n a distribution on the set P_n [1]. The closedness of P' to P can be measured by the **asymptotic density (=asymptotic probability)** (if exists)

$$D(P') = \lim_{n \rightarrow \infty} D_n(P' \cap P_n). \quad (3)$$

A subset $P' \subseteq P$ is said to be **asymptotically almost certain (=generic)** if $D(P') = 1$. In the theory of random graphs another terminology is accepted: the event $P' \subseteq P$ as above **happens asymptotically almost surely (=a.a.s)** or **with high probability (=w.h.p.)**. The complement of a generic set is said to be **negligible**.

A search problem (P, R) is **polynomial-time decidable with high probability (=for asymptotically almost all inputs)** if it admits a polynomial-time decidable restriction (P', R) such that P' is asymptotically almost certain in P (relative to a fixed size function and fixed distribution ensemble. Similar definitions can be given in case of promise decision problems.

2. Graph problems

Let \mathcal{G}_n denote the set of all graphs on the set of vertices $[1; n] = \{1, \dots, n\}$. The set \mathcal{G}_n is naturally acted upon by the symmetric group S_n on the set $[1; n]$. The orbits of this action are precisely the isomorphism classes of graphs from \mathcal{G}_n .

We encode graphs as binary strings as usual. Namely, let (g_{ij}) be the adjacency matrix of G then we encode G by the string $(g_{12}g_{13} \cdots g_{1n}g_{23} \cdots g_{2n} \cdots g_{n-1,n})$. This encodes \mathcal{G}_n bijectively onto $I_{\frac{n(n-1)}{2}}$. The permutations $\phi \in S_n$ are in one-one correspondence with $n \times n$ monomial matrices and the last ones can be encoded row by row by binary strings in I_{n^2} . We consider the **graph isomorphism problem**

$GI(L, P)$ as a partial decision problem (L, P) whose input set is $I \times I$, whose promise set is a decidable subset $P \subseteq P_{all}$, where

$$P_{all} = \cup_n (\mathcal{G}_n \times \mathcal{G}_n) = \cup_n \left(I_{\frac{n(n-1)}{2}} \times I_{\frac{n(n-1)}{2}} \right) \quad (4)$$

and whose language L of yes-instances is defined by

$$L = P_{isom} \stackrel{def}{=} \{(G, H) \in P_{all} : G \simeq H\}. \quad (5)$$

The **standard** GI-problem is (P_{isom}, P_{all}) . It is to determine for a given tuple $(G, H) \in I^2$ whether $G \simeq H$ or not.

In view of exceptional hardness of (P_{isom}, P_{all}) the problems (P_{isom}, P) have been considered with P being the pairs of graphs from some interesting classes of graphs, such as trees, planar graphs, graphs of bounded valence, etc. For instance, it is well known, that promise decision problem (P_{isom}, P_{trees}) , where P_{trees} is the class of all finite trees, is polynomial-time decidable.

Accompanying to GI is SGI - the **search graph isomorphism problem with a promise**. This problem is denoted by $SGI(P, R)$ and consists of the input set I^2 , relation $R \subseteq I^2 \times I$ and a promise set $P \subseteq I^2$. The relation R is defined by

$$R = \{(G, H, \phi) : \exists n \text{ such that } G, H \in \mathcal{G}_n, \phi \in S_n \subseteq I_{n^2} \text{ and } \phi : G \simeq H\}. \quad (6)$$

The promise set P is an arbitrary decidable subset of P_{all} . The problem requires, for a given pair of graphs $(G, H) \in P$ to decide first whether they are isomorphic and then to find an isomorphism $\phi : G \simeq H$.

The problem $SGI(P, R)$ with promise P is **solved** by algorithm A if for every $(G, H) \in P$ it holds that $((G, H), A(G, H)) \in R$ if $G \simeq H$ and $A(G, H) = \perp$ otherwise. We see, that the **full search problem** (P_{all}, R) with a promise P_{all} is the standard SGI-problem as it is defined for instance in [9]. This problem contains the standard decision problem $GI(P_{isom}, P_{all})$ as a subproblem. Indeed, if an algorithm A solves $SGI(P_{all}, R)$ then $(G, H) \in P_{isom}$ if and only if $A(G, H) = \perp$, thus A recognizes whether an isomorphism between G and H exists or not.

At the other extreme there is the candid search problem (S_R, R) with promise S_R , where $S_R = \{x : R(x) \neq \emptyset\}$ is the solution set of R . In our case it is clear that

$$S_R = P_{isom} = \{(G, H) \in I^2 : G \simeq H\}. \quad (7)$$

And again, the full search problem $SGI(P_{all}, R)$ contains the candid search problem $SGI(S_R, R)$ as a subproblem. We consider $SGI(S_R, R)$ as an adequate formulation (in terms of promise problems) for the problem SGI from [12]. We find the following problem rather intriguing:

Problem 1. *Is the standard isomorphism problem $GI(P_{isom}, P_{all})$ polynomial-time reducible to $SGI(S_R, R)$?*

3. The problem $SGI(P_{all}, R)$ is polynomial-time reducible to $GI(P_{isom}, P_{all})$

Theorem 1. *The full search isomorphism problem $SGI(P_{all}, R)$ is polynomial-time reducible to the standard isomorphism problem $GI(P_{isom}, P_{all})$.*

Proof. (Cf. [9, Thm.6, p.29]) It is easy to see that SGI can be reduced to the case of graphs without isolated vertices.

Ascent. Suppose we have a polynomial-time algorithm A recognizing the graph isomorphism. Given two isomorphic graphs G, H on n vertices and the list of vertices $V_G = \{g_1, \dots, g_n\}$, we will construct inductively and efficiently (=in polynomial time) the sets of graphs $G = G_0 < G_1 < \dots < G_n$, $H = H_0 < H_1 < \dots < H_n$ and the vertices $h_1, \dots, h_n \in V_H$ such that:

- 1) $G_i \simeq H_i$ for all i ,
- 2) Every isomorphism between G_i and H_i takes G_{i-1} to H_{i-1} for all $i = 1, \dots, n$,
- 3) Every isomorphism between G_i and H_i takes g_i to h_i for all $i = 1, \dots, n$.

The construction is defined as follows. Let K_{n+1} denote the complete graph on $n+1$ vertices. Fix $k_1 \in K_{n+1}$ and consider

$$G_1 = G \cup_{g_1=k_1} K_{n+1}, \quad H_1(h) = H \cup_{h=k_1} K_{n+1},$$

where $\cup_{x=y}$ denote the gluing operation on graphs with identified vertices $x = y$. Since $G \simeq H$, there exists h such that $G_1 \simeq H_1 = H_1(h)$, namely h is the image of g_1 under isomorphism $G \simeq H$. We find out h by applying \mathcal{A} to all n pairs $(G_1, H_1(h))$. Set $h_1 = h$. By assumption, G, H have no isolated vertices, so g_1, h_1 are unique vertices of highest degree in G_1, H_1 . Hence every isomorphism between G_1 and H_1 takes g_1 to h_1 . Moreover, all vertices of G distinct from g_1 have degree (in G_1) no greater than $n-1$, whereas all the vertices of K_{n+1} have degree (in G_1) at least n . Therefore, every isomorphism between G_1 and H_1 takes G_0 to H_0 . Thus 1)-3) hold for $i = 1$.

Fix $k_2 \in K_{n+2}$. For all $h \in V_H$ construct the graphs

$$G_2 = G_1 \cup_{g_2=k_2} K_{n+2}, \quad H_2(h) = H_1 \cup_{h=k_2} K_{n+2}. \quad (8)$$

Since G_1, H_1 are isomorphic, there exists h such that $G_2 \simeq H_2$ and we can find h applying \mathcal{A} to all n pairs of graphs $G_2, H_2(h)$. Fix any such h and denote it by h_2 . Continuing, we find out the desired $h_1, \dots, h_n \in H$ and $G = G_0 < G_1 < \dots < G_n$, $H = H_0 < H_1 < \dots < H_n$.

Descent. The map $g_i \mapsto h_i$ is an isomorphism between G and H ! Indeed, by 1) there exists an isomorphism $\phi : G_n \rightarrow H_n$. By 2) ϕ takes g_n to h_n . The restriction $\phi|_{G_{n-1}}$ takes G_{n-1} to H_{n-1} by 2) and takes g_{n-1} to h_{n-1} by 3). Then $\phi|_{G_{n-1}}$ takes G_{n-2} to H_{n-2} and we can continue this descent process until obtaining that ϕ takes every g_i to h_i . Hence the map $g_i \mapsto h_i$ is an isomorphism in question. ■

4. Graph invariant τ and oblique graphs

Graph invariant τ [2, 13, 14].

Let \mathbb{N} denote the set of all natural numbers (taken as $1, 2, 3, 4, \dots$). A **string** over \mathbb{N} is a finite sequence of natural numbers, that is, an integer $n > 0$ and a mapping $\{1, \dots, n\} \rightarrow \mathbb{N}$. If $n = 0$, the domain is the nullset and there is a unique such mapping, called the **nullstring** and generally denoted ε [3]. Let \mathbb{N}^* denote the set of all finite strings of natural numbers (including the nullstring ε). The **lexicographic order** ranks strings of the same length in \mathbb{N}^* , by comparing the letters in the first position where the strings differ. We define the **ShortLex order** on \mathbb{N}^* : $v < w$ if and only if v is shorter than w , or they have the same length and v comes before w in lexicographical order. ShortLex order is a well-ordering.

Let $G = (V, E)$ be a simple graph. By $d(x)$ we denote the degree of the vertex x . The set of all vertices adjacent to v is denoted by $N(v)$.

The **type-string** $\tau_G(v) \in \mathbb{N}^*$ of a vertex v is the string of degrees of vertices of $N(v)$ in non-decreasing order. We set $\tau_G(v) = \varepsilon$ for isolated vertex v . In detail, $\tau_G(v) = (d_1, \dots, d_{d(v)})$ is the degree sequence of the vertices adjacent to v , arranged in non-decreasing order: $d_1 \leq d_2 \leq \dots \leq d_{d(v)}$. Clearly a type-string is preserved under graph isomorphisms: if $\phi : G \rightarrow H$ is a graph isomorphism then $\tau_H(\phi(v)) = \tau_G(v)$ for all $v \in V(G)$, i.e. $\tau_H \circ \phi = \tau_G$.

The **type-vector** τ_G of G is the ShortLex ordered sequence $(\tau_G(v_1), \dots, \tau_G(v_n))$, $n = |V|$ of all type-strings of all vertices (thus τ_G is used to denote a vector as well as a map). Clearly the type-vector function $G \mapsto \tau_G$ is a **graph invariant**, i.e. $\tau_G = \tau_H$ for isomorphic graphs G, H . However, as the next example will show, the type-vector is not a complete graph invariant, i.e. there exist non-isomorphic graphs with the same type-vector. The examples will be found among regular graphs. A graph G is said to be **k -regular** if all the vertices of G have the same degree k . In this case $\tau_G = (k^{*k}, \dots, k^{*k})$ has n coordinates, $n = |V_G|$, where $k^{*k} = (k, \dots, k) \in \mathbb{N}^*$ is a string with k coordinates. Thus all k -regular graphs on n vertices have the same τ -invariant.

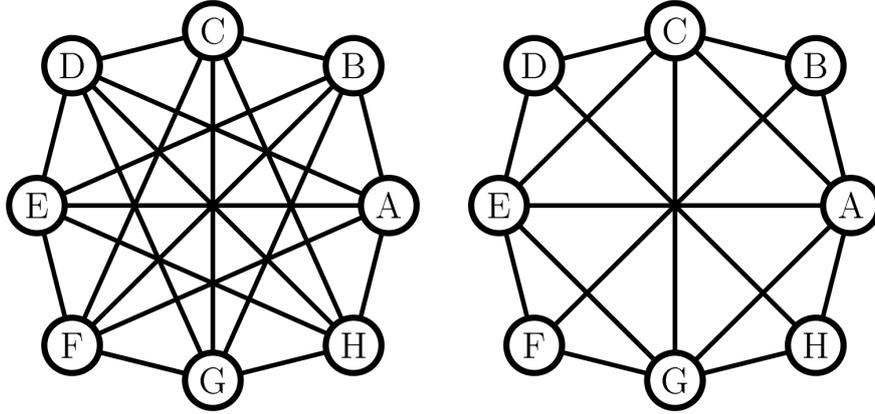
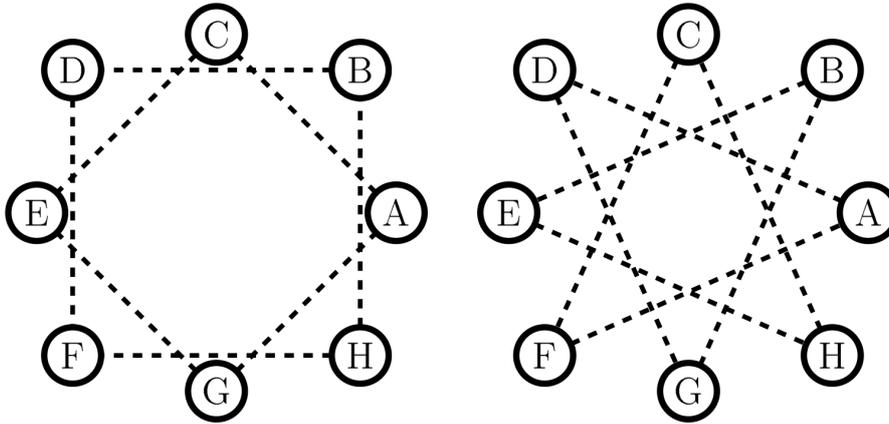
Proposition 1. *There exist two non-isomorphic graphs G_1, G_2 on 8 vertices all of whose degrees are equal 5. In particular G_1, G_2 have the same type-vector $(5^{*5}, 5^{*5}, 5^{*5}, 5^{*5}, 5^{*5}, 5^{*5}, 5^{*5}, 5^{*5})$, where $5^{*5} = (5, 5, 5, 5, 5) \in \mathbb{N}^*$.*

Proof. Define G_1, G_2 as on Figure 1.

Degree conditions are clearly fulfilled. It is left to ensure that the graphs are non-isomorphic. For this it is enough to ensure that the complement graphs are non-isomorphic. The complement to the first graph is two disjoint 4-cycles ACHF and BDGE (so it is disconnected), whereas the complement to the second graph is an 8-cycle AFEBHCDG (hence it is connected). The non-isomorphism is clear now. ■

Despite the fact that τ is not complete, it turns out to be **asymptotically complete**, i.e. for asymptotically almost all graphs G, H the equality $\tau_G = \tau_H$ implies that G, H are isomorphic (see [2, 13] and Section 7).

A graph G is said to be **oblique** if the map $\tau_G : V \rightarrow \mathbb{N}^*$ is injective or, in other words, there are no distinct vertices $u, v \in V(G)$ such that $\tau_G(u) = \tau_G(v)$ [14]. Figure 3 shows an example of oblique graph on 8 vertices [5].

Figure 1. Graphs G_1 and G_2 .Figure 2. Complement graphs $\overline{G_1}$ and $\overline{G_2}$.

Lemma 1 (uniqueness of the type-vector). *If a graph G is oblique then its type vector $\tau_G = (\tau_G(v_1), \dots, \tau_G(v_n))$ is unique in the sense that $(\tau_G(v_1), \dots, \tau_G(v_n)) = (\tau_G(w_1), \dots, \tau_G(w_n))$ implies that $v_i = w_i$ for all i .*

Proof. Indeed, by assumption $\tau_G(v_i) = \tau_G(w_i)$ for all i , hence by obliqueness $v_i = w_i$ for all i . ■

Lemma 2 (retrieving an isomorphism). *Let $\tau_G = (\tau_G(g_1), \dots, \tau_G(g_n)), \tau_H = (\tau_H(h_1), \dots, \tau_H(h_n))$ be the type-vectors of graphs G, H respectively. If G is oblique and $G \simeq H$ then the map $g_i \mapsto h_i$ is the isomorphism of G onto H and there is no other isomorphism between G and H can be drawn. In particular, every oblique graph has a trivial automorphism group.*

Proof. Let $\tau_G = (\tau_G(g_1), \dots, \tau_G(g_n)), \tau_H = (\tau_H(h_1), \dots, \tau_H(h_n))$ be the type-vectors of graphs G, H respectively, let G be oblique and let $\phi : G \rightarrow H$ be an

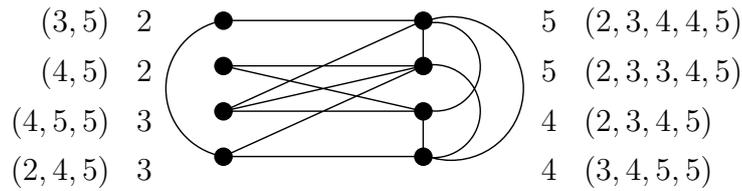


Figure 3. A vertex-oblique graph on 8 vertices. The degrees of vertices and their neighbors are written in columns

arbitrary isomorphism. By assumption $\tau_G : V(G) \rightarrow \mathbb{N}^*$ is injective, so $\tau_H = \tau_G \circ \phi^{-1} : V(H) \rightarrow \mathbb{N}^*$ is injective also and thus H is oblique. The vector of strings

$$(\tau_H(\phi g_1), \dots, \tau_H(\phi g_n)) = (\tau_G(g_1), \dots, \tau_G(g_n)) \tag{9}$$

is the type-vector of H because its coordinates are all type-strings of all vertices of H (looking at the left-hand side) and they are ShortLex ordered (looking at the right-hand side). Thus we have two type-vectors for H :

$$(\tau_H(\phi g_1), \dots, \tau_H(\phi g_n)) = \tau_H = (\tau_H(h_1), \dots, \tau_H(h_n)). \tag{10}$$

By the uniqueness Lemma $\phi g_i = h_i$ for all i . Hence the map $g_i \mapsto h_i = \phi g_i$ is an isomorphism of G onto H ; moreover, there is no other isomorphism between G and H . ■

5. Efficient solvability of $GI(P_{isom}, P_{oblique})$ and $SGI(P_{isom} \cap P_{oblique}, R)$

Consider the following sets of pairs of graphs:

$$P_{all} = \cup_n (\mathcal{G}_n \times \mathcal{G}_n), \tag{11}$$

$$P_{isom} = \{(G, H) \in P_{all} : G \simeq H\}, \tag{12}$$

$$P_{oblique} = \{(G, H) \in P_{all} : G \text{ and } H \text{ are oblique}\}. \tag{13}$$

It is easy to see that all these sets are decidable.

Theorem 2. *The promise decision problem $GI(P_{isom}, P_{oblique})$ is polynomial-time solvable.*

Proof. (Cf. [2,13]) This time the promise set $P_{oblique}$ does not contain the language P_{isom} , so we need the formalism of promise problems in full generality. Precisely, an algorithm A solves instance (G, H) of $P_{oblique}$ iff

$$A(G, H) = \begin{pmatrix} 1 & \text{if } (G, H) \in P_{oblique} \cap P_{isom} \\ 0 & \text{if } (G, H) \in P_{oblique} - P_{isom} \end{pmatrix}. \tag{14}$$

On instances out of $P_{oblique}$ the behavior of A may be rather arbitrary. Since $P_{oblique}$ is decidable, we can construct the required algorithm A by splitting into cases depending on whether a given instance (G, H) belongs to $P_{oblique}$ or not. Precisely, we put $A(G, H) = 0$ on instances $(G, H) \in I^2 - P_{oblique}$. The idea of the restriction algorithm $A|_{P_{oblique}}$ is to distinguish all vertices of a graph using the invariant τ . Given a pair $(G, H) \in P_{oblique}$ the required algorithm A first computes the type-vectors

$$\tau_G = (\tau_G(g_1), \dots, \tau_G(g_n)), \tau_H = (\tau_H(h_1), \dots, \tau_H(h_n)). \quad (15)$$

If $\tau_G = \tau_H$ then A checks whether the map $\phi : g_i \mapsto h_i$ is an isomorphism from G to H or not. If it is an isomorphism then we put $A(G, H) = 1$. Otherwise Lemma 2 asserts that G, H are not isomorphic, so $A(G, H) = 0$. Thus A is well defined on all inputs I^2 and it solves the given problem. It is easy to see that the running time of A is linear in $|V| + |E|$. ■

Theorem 3. *The problem SGI $(P_{isom} \cap P_{oblique}, R)$ with relation*

$$R = \{((G, H), \phi) \in I^2 \times I : \exists n \text{ such that } G, H \in \mathcal{G}_n, \phi \in S_n \subseteq I_{n^2} \text{ and } \phi : G \simeq H\} \quad (16)$$

is polynomial-time solvable.

Proof. By definition the given problem is solved by algorithm A if for every $(G, H) \in P_{isom} \cap P_{oblique}$ the inclusion $((G, H), A(G, H)) \in R$ holds in case $G \simeq H$ and $A(G, H) = \perp$ otherwise. Given the n -vertex graphs G, H the required algorithm A first computes the type-vectors

$$\tau_G = (\tau_G(g_1), \dots, \tau_G(g_n)), \tau_H = (\tau_H(h_1), \dots, \tau_H(h_n)). \quad (17)$$

If the coordinates either of τ_G or τ_H are not distinct then at least one of graphs is not oblique i.e. $(G, H) \notin P_{isom} \cap P_{oblique}$. In this case we let $A(G, H) = id \in S_n$. Otherwise both of graphs are oblique and then algorithm compares the type-vectors. If $\tau_G \neq \tau_H$, then G, H are not isomorphic and thus $A(G, H)$ is already defined. If $\tau_G = \tau_H$ then the algorithm checks whether the map $\phi : g_i \mapsto h_i$ is an isomorphism or not. If this map turns out to be an isomorphism then the algorithm gives ϕ as the answer to the problem. Otherwise, if the map is not an isomorphism, Lemma 2 asserts that there is no isomorphism between the given graphs at all. That is this case is impossible by the very setup of the problem. Thus the algorithm A gives the right answer for the set of inputs $P_{isom} \cap P_{oblique}$ and $A(G, H) = id$ on the complement set $\overline{P_{isom} \cap P_{oblique}}$, which may be the wrong answer (but also may be right answer occasionally). ■

6. Generic algorithms for GI and SGI and their failure probability

Theorem 4. *Endow the set P_{all} with the ensemble of distributions (D_n) , where D_n is the uniform distribution on the set P_{all}^n of pairs of n -vertex graphs. Then $P_{isom} \cap P_{oblique}$ is asymptotically almost certain in P_{isom} .*

Proof. Let \mathcal{G}_n^o denote the set of oblique graphs on vertices $1, 2, \dots, n$. We rely on the result from [2, 13] which says that the property of a graph $G \in \mathcal{G}_n$ to be oblique holds asymptotically almost surely, i.e.

$$\lim_{n \rightarrow \infty} \frac{|\mathcal{G}_n^o|}{|\mathcal{G}_n|} = 1, \tag{18}$$

or, equivalently,

$$\lim_{n \rightarrow \infty} \frac{|\mathcal{S}_n|}{|\mathcal{G}_n|} = 0, \tag{19}$$

where $\mathcal{S}_n = \mathcal{G}_n - \mathcal{G}_n^o$.
We have to prove that

$$\lim_{n \rightarrow \infty} \frac{|P_{isom}^n \cap P_{oblique}^n|}{|P_{isom}^n|} = 1, \tag{20}$$

or, equivalently,

$$\lim_{n \rightarrow \infty} \frac{|P_{isom}^n \cap (\mathcal{S}_n \times \mathcal{S}_n)|}{|P_{isom}^n|} = 0. \tag{21}$$

Consider the natural projection map onto the first coordinate $pr_1 : P_{isom}^n \rightarrow \mathcal{G}_n$. What is the cardinality of the inverse image (=fiber over G) of a graph G under pr_1 ? Each isomorphism class of $G \in \mathcal{G}_n$ consists of at most $n!$ elements, so $|pr_1^{-1}G| \leq n!$ for every $G \in \mathcal{G}_n$, hence

$$|P_{isom}^n| \geq n! |\mathcal{G}_n|. \tag{22}$$

In case G is oblique, we can say more: $|pr_1^{-1}G| = n!$. Indeed, by Lemma 2, $Aut(G) = 1$ and thus the orbit $S_n G$ consists of $n!$ graphs and the set $(G, S_n G)$ is the inverse image of G under pr_1 . Thus each fiber of the map $pr_1 : P_{isom}^n \cap (\mathcal{G}_n^o \times \mathcal{G}_n^o) \rightarrow \mathcal{G}_n^o$ has cardinality $n!$. Hence $|P_{isom}^n \cap (\mathcal{G}_n^o \times \mathcal{G}_n^o)| = n! |\mathcal{G}_n^o|$ and so

$$|P_{isom}^n| \geq n! |\mathcal{G}_n^o|. \tag{23}$$

Considering the fibers of pr_1 over the set \mathcal{S}_n , we conclude that

$$|P_{isom}^n \cap (\mathcal{S}_n \times \mathcal{S}_n)| \leq n! |\mathcal{S}_n|. \tag{24}$$

Finally, combining inequalities 23 and 24 and the result 19, we obtain

$$\frac{|P_{isom}^n \cap (\mathcal{S}_n \times \mathcal{S}_n)|}{|P_{isom}^n|} \leq \frac{n! |\mathcal{S}_n|}{n! |\mathcal{G}_n^o|} = \frac{|\mathcal{S}_n|}{|\mathcal{G}_n^o|} = \frac{|\mathcal{S}_n|}{|\mathcal{G}_n| - |\mathcal{S}_n|} = \tag{25}$$

$$= \frac{|\mathcal{S}_n|}{|\mathcal{G}_n|} \left(\frac{1}{1 - \frac{|\mathcal{S}_n|}{|\mathcal{G}_n|}} \right) \rightarrow 0, \quad n \rightarrow \infty. \tag{26}$$

■

Remark 2. It is shown in [2] a linear time, high probability canonical labeling algorithm for $G(n, p)$ graphs for $p = \omega(\ln^4 n/n \ln \ln n)$ and $p \leq 1/2$. Here, high probability means probability at least $1 - O(n^{-c})$ for every $c > 0$. It follows that $P_{isom} \cap P_{oblique}$ is asymptotically almost certain with the rate of convergence at least $1 - O(n^{-c})$ for every $c > 0$.

Here we prove the main theorem in the following more precise formulation.

Theorem 5. *Any algorithm A , constructed in Section 5. to solve the problem $SGI(P_{isom} \cap P_{oblique}, R)$, solves also the problem $SGI(P_{isom}, R)$ for asymptotically almost all inputs $P_{isom} \cap P_{oblique}$. In other words, the failure probability of the algorithm A tends to zero as n tends to infinity.*

7. Applications to probabilistic algorithms

Recall that the input set of the problem $SGI(P_{isom}, R)$ is $I^2 \times I$, the promise set is

$$P_{isom} = \{(G, H) : \exists n \text{ such that } G, H \in \mathcal{G}_n, G \simeq H\}, \quad (27)$$

and the relation is

$$R = \{(G, H, \phi) : \exists n \text{ such that } G, H \in \mathcal{G}_n, \phi \in S_n \subseteq I_{n^2} \text{ and } \phi : G \simeq H\}. \quad (28)$$

A.N. Rybalov considered a search graph isomorphism problem with particularly small promise set $P_\gamma \subseteq P_{isom}$ [12]. Namely, fix an infinite sequence of graphs $\gamma = (G_n)_{n \in \mathbb{N}}$, such that $G_n \in \mathcal{G}_n$ for all n and define

$$P_\gamma = \{(G, G_n) : n \in \mathbb{N}, \text{ and } G, G_n \in \mathcal{G}_n, G \simeq G_n\}. \quad (29)$$

A.N. Rybalov studied the problems of the type $SGI(P_\gamma, R)$ which clearly are all the restrictions of the candid search problem $SGI(P_{isom}, R)$.

The main result of A.N.Rybalov is the following.

Theorem 6. [12] *If there exists a polynomial generic algorithm for $SGI(P_\gamma, R)$, then there exists a polynomial probabilistic algorithm computing $SGI(P_\gamma, R)$ for all inputs.*

From this and from our Theorem 5 we derive the following

Theorem 7. *If the sequence of graphs $\gamma = (G_n)_{n \in \mathbb{N}}$ consists of oblique graphs, then the problem $SGI(P_\gamma, R)$ is polynomial-time solvable and there exists a polynomial probabilistic algorithm computing $SGI(P_\gamma, R)$ for all inputs.*

Acknowledgements

The author is thankful to A.N.Rybalov for stating the SGI.

REFERENCES

1. Bogdanov A., Trevisan L. Average-case complexity // *Found. Trends Theor. Comput. Sci.* October 2006. V. 2(1). P. 1–106.
2. Czajka T., Pandurangan G. Improved random graph isomorphism // *J. of Discrete Algorithms.* March 2008. V. 6(1). P. 85–92.
3. Epstein D.B.A., Cannon J.W., Holt D.F., Levy S.V.F., Paterson M.S., Thurston W.P. *Word processing and group theory.* Jones and Bartlet Publishers, 1992.
4. Even S., Selman A.L., Yacobi Y. The complexity of promise problems with applications to public-key cryptography // *Inform. and Control.* 1984. V. 61(2). P. 159–173.
5. Farrugia A. Dually vertex-oblique graphs // *Discrete Mathematics.* 2007. V. 307(11–12). P. 1323–1331. The Fourth Caracow Conference on Graph Theory Czorsztyn 2002.
6. Goldreich O. On promise problems: a survey. In *Theoretical computer science // Lecture Notes in Comput. Sci.* Springer, Berlin, 2006. V. 3895. P. 254–290.
7. Goldreich O. *Computational complexity – a conceptual perspective.* Cambridge University Press, 2008.
8. Goldreich O. *P, NP, and NP-Completeness: The Basics of Complexity Theory.* Cambridge University Press, 2010.
9. Hoffmann C.M. *Group-theoretic algorithms and graph isomorphism // Lecture Notes in Computer Science.* Springer-Verlag, Berlin, 1982. V. 136.
10. Longpré L. Selman A.L. Hard promise problems and nonuniform complexity. In *STACS 90 (Rouen, 1990) // Lecture Notes in Comput. Sci.* Springer, Berlin, 1990. V. 415. P. 216–226.
11. Myasnikov A., Shpilrain V., Ushakov A. *Group-based Cryptography. Advanced Courses in Mathematics CRM Barcelona.* Birkhauser Verlag, 2008.
12. Rybalov A. *On the generic complexity of the searching graph isomorphism problem.* Omsk State Technical University, 2015.
13. Schreyer J. Almost every graph is vertex-oblique // *Discrete Math.* 2007. V. 307(7-8). P. 983–989.
14. Schreyer J., Walther H., Mel'nikov L.S. Vertex-oblique graphs // *Discrete Mathematics.* 2007. V. 307(11–12). P. 1538 – 1544. The Fourth Caracow Conference on Graph Theory Czorsztyn 2002.

ПРОБЛЕМЫ ПОИСКА С ПОСУЛОМ И ИЗОМОРФИЗМ ГРАФОВ**Г.А. Носков**

д.ф.-м.н., с.н.с., e-mail: g.noskov@googlemail.com

Институт Математики им. С.Л. Соболева СОРАН

Аннотация. Проблема изоморфизма графов рассматривается с точки зрения теории «проблем с посулом», развитой Ивеном, Зелманом и Якоби [4]. Изучается «тау-инвариант» графов и с его помощью решается проблема поиска изоморфизма для асимптотически почти всех графов.

Ключевые слова: проблема разрешимости, проблема поиска с посулом, изоморфизм графов.

ONCE WE KNOW THAT A POLYNOMIAL MAPPING IS RECTIFIABLE, WE CAN ALGORITHMICALLY FIND A RECTIFICATION

J. Urenda¹

Ph.D.(Math.), Instructor, e-mail: jcurenda@utep.edu

D. Finston²

Ph.D.(Math.), Full Professor, e-mail: dfinston@nmsu.edu

V. Kreinovich¹

Ph.D. (Phys.-Math.), Professor, e-mail: vladik@utep.edu

¹University of Texas at El Paso, El Paso, Texas 79968, USA

²New Mexico State University, Las Cruces, New Mexico 88003, USA

Abstract. It is known that some polynomial mappings $\varphi : \mathbb{C}^k \rightarrow \mathbb{C}^n$ are *rectifiable* in the sense that there exists a polynomial mapping $\alpha : \mathbb{C}^n \rightarrow \mathbb{C}^n$ whose inverse is also polynomial and for which $\alpha(\varphi(z_1, \dots, z_k)) = (z_1, \dots, z_k, 0, \dots, 0)$ for all z_1, \dots, z_k . In many cases, the existence of such a rectification is proven indirectly, without an explicit construction of the mapping α .

In this paper, we use Tarski-Seidenberg algorithm (for deciding the first order theory of real numbers) to design an algorithm that, given a polynomial mapping $\varphi : \mathbb{C}^k \rightarrow \mathbb{C}^n$ which is known to be rectifiable, returns a polynomial mapping $\alpha : \mathbb{C}^n \rightarrow \mathbb{C}^n$ that rectifies φ .

The above general algorithm is not practical for large n , since its computation time grows faster than 2^{2^n} . To make computations more practically useful, for several important case, we have also designed a much faster alternative algorithm.

Keywords: polynomial mapping, rectification, algorithmically computable, Tarski-Seidenberg algorithm.

1. Formulation of the Problem

It is known that several classes of polynomial mappings are rectifiable in the following sense.

Definition 1. Let \mathbb{C} denote the field of all complex numbers. A polynomial mapping $\alpha : \mathbb{C}^n \rightarrow \mathbb{C}^n$ is called a polynomial automorphism if this mapping a bijection, and the inverse mapping $\beta = \alpha^{-1}$ is also polynomial.

Definition 2. A polynomial mapping $\varphi : \mathbb{C}^k \rightarrow \mathbb{C}^n$ is called rectifiable if there exists a polynomial automorphism $\alpha : \mathbb{C}^n \rightarrow \mathbb{C}^n$ for which $\alpha(\varphi(t_1, \dots, t_k)) = (t_1, \dots, t_k, 0, \dots)$ for all (t_1, \dots, t_k) .

Most existing proofs of rectifiability just prove the existence of a rectifying automorphism α , without explaining how to actually compute it. In this paper, we show how to compute α .

2. Main Result

We will formulate two versions of the main result: for the case when the coefficients of the original polynomial mapping are algebraic numbers, and for the general case, when these coefficients are not necessarily algebraic and may not even be computable.

Definition 3. *A real number is called algebraic if this number is a root of a non-zero polynomial with integer coefficients. A complex number $a + b \cdot i$ is called algebraic if both a and b are algebraic.*

Comment. In the computer, an algebraic real number can be represented by the integer coefficients of the corresponding polynomial and — if this polynomial has several roots — by a rational-valued interval that contains this particular root and does not contain any other roots of this polynomial.

Once this information is given, we can compute the corresponding root with any given accuracy.

Lemma 1. *If a polynomial mapping φ with algebraic coefficients is rectifiable, then there exists a rectifying polynomial automorphism α with algebraic coefficients.*

Proposition 1. *There exists an algorithm that, given a rectifiable polynomial mapping φ with algebraic coefficients, computes the coefficients of a polynomial automorphism α that rectifies φ .*

Discussion. It is desirable to extend this algorithm to the general case, when the coefficients of the original mapping φ are not necessarily algebraic and may not even be computable. When the coefficients are not necessarily computable, we cannot represent them in a computer, so we need to extend the usual notion of an algorithm to cover this case.

Definition 4. *By a generalized algorithm, we mean a sequence of the following elementary operations with real numbers:*

- *adding, subtracting, multiplying, and dividing numbers;*
- *checking whether a number is equal to 0, whether it is positive, and whether it is negative;*
- *given the coefficients of a polynomial that has a root, returning one of the roots.*

Comment. Of course, when the real numbers are algebraic, these operations are algorithmically computable.

Proposition 2. *There exists a generalized algorithm that, given the coefficients of a rectifiable polynomial mapping φ , computes the coefficients of a polynomial automorphism α that rectifies φ .*

Discussion. Propositions 1 and 2 show that if a polynomial mapping is rectifiable, then the corresponding rectification can be algorithmically computed.

Comments. Our proof uses the Tarski algorithm. While this algorithm produces the desired results, it is known to be hyper-exponential: as the length ℓ of the formula increases, its running time grows faster than 2^{2^ℓ} . Thus, from the application viewpoint, it is desirable to come up with a faster algorithm. For some important cases, such faster algorithm was proposed in [3]; it should be mentioned that, in contrast to our algorithms which are limited to the field of all complex numbers, algorithms from [3] can be applied to other fields (and rings) as well.

Comment. The main results were first announced in [3].

3. Proofs

Tarski-Seidenberg algorithm: reminder. In this paper, we will use Tarski-Seidenberg algorithm; see, e.g., [1, 2]. This algorithm deals with the *first-order theory of real numbers*. Formulas of this theory are defined as follows:

- we start with real-valued variables x_1, \dots, x_n ;
- *elementary formulas* are formulas of the type $P = 0$, $P > 0$, or $P \geq 0$, where P is a polynomial with integer coefficients;
- finally, a general formula can be obtained from elementary formulas by using logical connectives (“and” $\&$, “or” \vee , “implies” \rightarrow , and “not” \neg) and quantifiers over real numbers ($\forall x_i$ and $\exists x_i$).

For example, a formula describing that the given polynomial $P(x_1, \dots, x_n)$ with integer coefficients has a solution with $x_i > 0$ for all i is a first-order formula:

$$\exists x_1 \dots \exists x_n ((P(x_1, \dots, x_n) = 0) \& (x_1 > 0) \& \dots \& (x_n > 0)).$$

Another example is a formula that shows that every quadratic polynomial with non-negative determinant has a solution:

$$\forall a \forall b \forall c ((b^2 - 4a \cdot c \geq 0) \rightarrow \exists x (a \cdot x^2 + b \cdot x + c = 0)).$$

Tarski designed an algorithm that, given a formula from this theory, returns 0 or 1 depending on whether this formula is true or not.

Seidenberg noticed that Tarski’s algorithm works by “eliminating” quantifiers one by one, i.e., by sequentially reducing a given formula to a one with one fewer quantifier. Because of this fact, he showed that we can use a similar construction to reduce each first-order formula with free variables to a quantifier-free form.

Tarski-Seidenberg algorithm: corollary. From the above reduction, it follows that if a formula with free variables has a solution, then it also has an algebraic solution. Namely, we can reduce the original formula to a quantifier-free formula $F(x_1, \dots, x_n)$.

The formula $\exists x_2 \dots \exists x_n F(x_1, x_2, \dots, x_n)$ can be similarly reduced to a quantifier-free expression, i.e., to a combination of equalities and inequalities of the type $P(x_1) = 0$, $P(x_1) > 0$, and $P(x_1) \geq 0$. If one of them is an equality, then we get an algebraic number x_1 ; if all of them are strict inequalities, then the whole range of values satisfies these inequalities and thus, we can select a rational (hence, algebraic) value from this interval.

Once we plug in the algebraic value x_1 into the original formula, we can then similarly find an algebraic value x_2 , etc. – and after n stages, we will get a tuple of algebraic numbers x_1, \dots, x_n that satisfies the original formula $F(x_1, \dots, x_n)$.

Proof of Lemma 1 and Proposition 1. Let us show that by using the Tarski-Seidenberg algorithm, we can come up with the desired algorithm for proving Proposition 1.

Let d be the largest degree of polynomials α_i and β_i forming the mappings α and $\beta = \alpha^{-1}$. Each of these polynomial can be described by listing all the coefficients – to be precise, by listing real and imaginary values of all these coefficients. The condition that α and β are inverse to each other means that

$$\forall z_1 \dots, \forall z_n ((\alpha_1(\beta(z_1, \dots, z_n)) = z_1) \& \dots \& (\alpha_n(\beta(z_1, \dots, z_n)) = z_n))$$

and

$$\forall z_1 \dots, \forall z_n ((\beta_1(\alpha(z_1, \dots, z_n)) = z_1) \& \dots \& (\beta_n(\alpha(z_1, \dots, z_n)) = z_n)).$$

Substituting the expressions for α and β in terms of their coefficients, we get a first order formula.

Similarly, the condition that α rectifies φ , i.e., that

$$\forall t_1 \dots \forall t_k ((\alpha_1(\varphi(t_1, \dots, t_k)) = t_1) \& \dots \& (\alpha_k(\varphi(t_1, \dots, t_k)) = t_k)),$$

is clearly a first-order formula. Thus, due to the above result, if there exists a solution, then there exists a solution in which all the coefficients of all the polynomials α_i and β_i are algebraic numbers.

For each tuple of algebraic numbers, checking whether the corresponding polynomials constitute a rectifying automorphism means checking whether a given first order formula is true, and this checking can be done by using the original Tarski's algorithm.

To find the desired polynomial mappings α and β with algebraic coefficients, it is sufficient to enumerate all possible tuples of such coefficients, and try them one by one, until we find a tuple which corresponds to the rectifying automorphism. Since we assumed that a rectification is possible, we will eventually find the desired coefficient.

The only thing that needs to be clarified is how to enumerate all possible tuples of algebraic numbers. This can be easily done if we take into account that each

algebraic number is represented in a computer as a sequence of integers. Thus, an arbitrary finite sequence of algebraic numbers can also be represented as a sequence of integers.

It is easy to come with an algorithm that enumerates all possible sequences of integers. For example, for $M = 0, 1, \dots$, we can enumerate all the sequences (n_1, \dots, n_k) for which $|n_1| + \dots + |n_k| + k = M$. For each M , there are finitely many such sequences, and it is easy to enumerate them all.

The proposition is thus proven.

Proof of Proposition 2. For each degree d , the Tarski-Seidenberg algorithm reduces the formula describing the existence of a rectifying polynomial automorphism of degree d to a finite list of equalities and inequalities between expressions which polynomially depend on the given coefficients and 0 . In our definition of a generalized algorithm, we allowed:

- additions and multiplications (all we need to compute the value of a polynomial) and
- checking whether a given value is equal to 0 or greater than 0 .

Thus, for each d , we have a generalized algorithm that checks whether a rectifying polynomial automorphism of degree d is possible.

Since we assume that a rectification is possible, by trying all possible degrees $d = 0, 1, 2, \dots$, we will eventually find d for which there exists a rectifying polynomial automorphism of degree d .

To complete the proof, we need to show how we can compute the coefficients of the corresponding polynomial mapping α . We want to find the coefficients c_1, \dots, c_N that satisfy a quantifier-free formula $F(c_1, \dots, c_N) = 0$. Let us start with computing c_1 . We want to find c_1 for which

$$\exists c_2 \dots \exists c_N (F(c_1, c_2, \dots, c_N) = 0).$$

We can use Tarski-Seidenberg theorem to reduce this formula to a quantifier-free one, i.e., to a sequence of polynomial equalities and inequalities $P_i(c_1) = 0$ and $P_j(c_1) > 0$. All equalities $P_i(c_1)$ can be combined into a single equality $P(c_1) = 0$, where $P(c_1) \stackrel{\text{def}}{=} \sum_i (P_i(c_1))^2$. We know that this polynomial equation has a solution.

We can therefore use one of the elementary steps of a generalized algorithm to compute a solution to this polynomial equation. If the solution s produced by this elementary step does not satisfy the inequalities, then we get a new polynomial of a smaller degree by dividing $P(c_1)$ by $c_1 - s$; it is clear that c_1 is a root of this polynomial. Division is algorithmic since it can also be reduced to (allowed) arithmetic operations with coefficients. We can then repeat this procedure with the new polynomial of smaller degree, etc. At each step, either we find the desired c_1 or the degree decreases. Since the degree cannot decrease below 0 , this means that we will eventually find c_1 .

Substituting this value c_1 into the above formula, we will then similarly compute a value c_2 that satisfies the formula

$$\exists c_3 \dots \exists c_N (F(c_1, c_2, c_3, \dots, c_N) = 0),$$

etc. After N steps, we will compute all the coefficients of the rectifying polynomial α . The proposition is proven.

Acknowledgments

This work was supported in part by the National Science Foundation grants HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and DUE-0926721.

The authors are thankful to all the participants of the NMSU/UTEP Workshop on Mathematics, Computer Science, and Computational Science (Las Cruces, New Mexico, April 11, 2015) for valuable suggestions.

REFERENCES

1. Basu S., Pollack R., Roy M.-F. Algorithms in Real Algebraic Geometry. Springer-Verlag, Berlin, 2006.
2. Tarski A., A Decision Method for Elementary Algebra and Geometry. 2nd ed., Berkeley and Los Angeles, 1951. 63 p.
3. Urenda J. Algorithmic Aspects of the Embedding Problem. PhD Dissertation, Department of Mathematical Sciences, New Mexico State University, Las Cruces, New Mexico, May 2015.
4. Urenda J., Finston D., Kreinovich V. Once We Know that a Polynomial Mapping Is Rectifiable, We Can Algorithmically Find a Rectification. Abstracts of the 16th Joint NMSU/UTEP Workshop on Mathematics, Computer Science, and Computational Science, Las Cruces, New Mexico, April 11, 2015.

**ЕСЛИ ПОЛИНОМИАЛЬНОЕ ПРЕОБРАЗОВАНИЕ СПРЯМЛЯЕМО, МЫ
МОЖЕМ АЛГОРИТМИЧЕСКИ НАЙТИ СПРЯМЛЕНИЕ****Х. Уренда**¹

к.ф.-м.н., преподаватель, e-mail: jcurenda@utep.edu

Д. Финстон²

к.ф.-м.н., профессор, e-mail: dfinston@nmsu.edu

В. Крейнович¹

к.ф.-м.н., профессор, e-mail: vladik@utep.edu

¹Техасский университет в Эль Пасо, США²Университет штата Нью-Мексико, США

Аннотация. Известно, что некоторые полиномиальные преобразования $\varphi : \mathbb{C}^k \rightarrow \mathbb{C}^n$ спрямляемы в том смысле, что существует полиномиальное преобразование $\alpha : \mathbb{C}^n \rightarrow \mathbb{C}^n$ такое, что обратное к нему тоже полиномиально выполняется $\alpha(\varphi(z_1, \dots, z_k)) = (z_1, \dots, z_k, 0, \dots, 0)$ для всех z_1, \dots, z_k . Во многих случаях существование такого спрямления доказано косвенно, без явного построения преобразования α .

В данной статье использован алгоритм Тарского-Зайденберга (для разрешимости теории действительных чисел первого порядка) для построения алгоритма, который по данному спрямляемому полиномиальному преобразованию $\varphi : \mathbb{C}^k \rightarrow \mathbb{C}^n$ возвращает полиномиальное преобразование $\alpha : \mathbb{C}^n \rightarrow \mathbb{C}^n$, спрямляющее φ .

Указанный выше общий алгоритм не практичен для больших n , так как его временная сложность растёт быстрее, чем 2^{2^n} . Чтобы сделать вычисления более приемлемыми для некоторых важных случаев был также разработан более быстрый альтернативный алгоритм.

Ключевые слова: полиномиальное преобразование, спрямление, алгоритмическая вычислимость, алгоритм Тарского-Зайденберга..

АЛГОРИТМЫ КОРА И ФОРДИАСИМПТ КАК МЕТОДЫ РАСПОЗНАВАНИЯ ДВУХ ОБРАЗОВ В ПРОСТРАНСТВЕ ДВОИЧНЫХ ПРИЗНАКОВ

В.В. Гольяпин

к.ф.-м.н, доцент, e-mail: golyapin@mail.ru

В.А. Шовин

научный сотрудник, e-mail: v.shovin@mail.ru

Омский филиал Федерального государственного бюджетного учреждения науки
Института математики им. С.Л. Соболева Сибирского отделения РАН

Аннотация. В данной статье в рамках теории латентного анализа сформулированы и доказаны утверждение, лемма и теорема, позволяющие находить апостериорные вероятности на базе альтернативных показателей с использованием ортогональной факторной структуры. На основе полученных теоретических выкладок построен вычислительный алгоритм ФОРДИАСИМПТ, позволяющий строить диагностические симптомокомплексы на базе вероятностного метода распознавания образов. Проведен сравнительный анализ алгоритмов КОРА и ФОРДИАСИМПТ как методов распознавания двух образов в пространстве двоичных признаков в случае независимых симптомокомплексов при адекватной статистической информации.

Ключевые слова: симптомокомплекс, факторная модель, латентная модель, корреляционный анализ, маргинальное распределение, маргинал, алгоритм КОРА, алгоритм ФОРДИАСИМПТ.

Введение

Основная цель работы заключается в обосновании практической целесообразности применения алгоритма формирования диагностических симптомокомплексов для случаев независимых симптомокомплексов. Ввиду длинного названия алгоритма предлагается использовать сокращение ФОРДИАСИМПТ. Для демонстрации достоинств и недостатков алгоритма ФОРДИАСИМПТ было проведено сравнительное исследование с алгоритмом КОРА.

Предлагаемый алгоритм решает следующие задачи. Во-первых, формирует набор симптомокомплексов, опираясь на ортогональную факторную структуру и на уровень значимости φ коэффициента по χ^2 критерию. Во-вторых, для каждого симптомокомплекса находит диагностическую шкалу на базе простейшей латентно-структурной модели.

В силу обоснованности использования факторного анализа для альтернативных показателей, считаем известным матрицу ортогонального факторного отображения. С полным изложением теоретических основ алгоритма ФОРДИАСИМПТ, относящихся к поиску факторной структуры, можно ознакомиться в работах [1-4].

Математический аппарат простейшей латентной модели и алгоритма ФОРДИАСИМПТ

Особое внимание в данной статье уделим непосредственно математическому аппарату, который используется в построении латентной модели на базе альтернативных данных. Обозначим количество объектов исследования n — объем выборки, а количество измеряемых параметров m — размерность выборки. Тогда исходные альтернативные данные представляются в виде матрицы $Y = \{y_{ij}\}$ размерности $m \times n$, столбцы которой — объекты исследования, а строки — значения измеряемых параметров у конкретного объекта.

Далее введём следующие обозначения:

p_i — отношение количества объектов к n , у которых i -ый показатель равен 1;

p_{ij} — отношение количества объектов к n , у которых i -ый и j -ый показатели равны 1;

$p_{i\bar{j}}$ — отношение количества объектов к n , у которых i -ый показатель равен 1, j -ый показатель равен 0;

$p_{\bar{i}\bar{j}}$ — отношение количества объектов к n , у которых i -ый и j -ый показатели равны 0;

p_{ijk} — отношение количества объектов к n , у которых i -ый, j -ый и k -ый показатели равны 1;

$p_{i\bar{j}k}$ — отношение количества объектов к n , у которых i -ый и k -ый показатели равны 1, а j -ый показатель равен 0;

$p_{\bar{i}\bar{j}k}$ — отношение количества объектов к n , у которых i -ый и j -ый показатели равны 0, а k -ый показатель равен 1;

$\tilde{\phi}(x_l)$ — частота, соответствующая относительному объёму l -го класса;

$\tilde{f}_i(x_l)$ — вероятность значения 1 по i -му показателю у объекта, находящегося в l -ом классе;

$\tilde{f}_{ik}(x_l)$ — вероятность значения 1 по i -му и k -му показателям у объекта, находящегося в l -ом классе;

$\tilde{f}_{ijk}(x_l)$ — вероятность значения 1 по i -му, j -му и k -му показателям у объекта, находящегося в l -ом классе.

Основываясь на теории латентного анализа, можно говорить об однозначном разделении объектов по трём альтернативным показателям на два латентных класса и сформировать разрешимую систему уравнений с дискретными переменными [5,6]:

$$\begin{cases} \tilde{\phi}(x_1) + \tilde{\phi}(x_2) = 1, \\ p_1 = \tilde{f}_1(x_1)\tilde{\phi}(x_1) + \tilde{f}_1(x_2)\tilde{\phi}(x_2), \\ p_2 = \tilde{f}_2(x_1)\tilde{\phi}(x_1) + \tilde{f}_2(x_2)\tilde{\phi}(x_2), \\ p_3 = \tilde{f}_3(x_1)\tilde{\phi}(x_1) + \tilde{f}_3(x_2)\tilde{\phi}(x_2), \\ p_{12} = \tilde{f}_{12}(x_1)\tilde{\phi}(x_1) + \tilde{f}_{12}(x_2)\tilde{\phi}(x_2), \\ p_{13} = \tilde{f}_{13}(x_1)\tilde{\phi}(x_1) + \tilde{f}_{13}(x_2)\tilde{\phi}(x_2), \\ p_{23} = \tilde{f}_{23}(x_1)\tilde{\phi}(x_1) + \tilde{f}_{23}(x_2)\tilde{\phi}(x_2), \\ p_{123} = \tilde{f}_{123}(x_1)\tilde{\phi}(x_1) + \tilde{f}_{123}(x_2)\tilde{\phi}(x_2). \end{cases} \quad (1)$$

Определение. Параметры латентно-структурной модели отношения p_i, p_{ij}, p_{ijk} называются маргиналами.

Определение. Латентно-структурная модель называется простейшей, если для её построения используются три альтернативных показателя.

Утверждение. Нахождение частоты $\tilde{\phi}(x_i)$ в простейшей модели латентно-структурного анализа сводится к каноническому уравнению прямой с точкой (p_1, p_2, p_3) и направляющим вектором

$$\vec{n} = \left(\tilde{f}_1(x_i) - \tilde{f}_1(x_j), \tilde{f}_2(x_i) - \tilde{f}_2(x_j), \tilde{f}_3(x_i) - \tilde{f}_3(x_j) \right), \text{ где } i \neq j.$$

Доказательство. Выразим из уравнения (1) одну из частот (например, $\tilde{\phi}(x_2)$) и подставим в остальные три уравнения:

$$p_1 = \tilde{f}_1(x_1)\tilde{\phi}(x_1) + \tilde{f}_1(x_2) \left(1 - \tilde{\phi}(x_1) \right) = \tilde{\phi}(x_1) \left(\tilde{f}_1(x_1) - \tilde{f}_1(x_2) \right) + \tilde{f}_1(x_2),$$

$$p_2 = \tilde{f}_2(x_1)\tilde{\phi}(x_1) + \tilde{f}_2(x_2) \left(1 - \tilde{\phi}(x_1) \right) = \tilde{\phi}(x_1) \left(\tilde{f}_2(x_1) - \tilde{f}_2(x_2) \right) + \tilde{f}_2(x_2),$$

$$p_3 = \tilde{f}_3(x_1)\tilde{\phi}(x_1) + \tilde{f}_3(x_2) \left(1 - \tilde{\phi}(x_1) \right) = \tilde{\phi}(x_1) \left(\tilde{f}_3(x_1) - \tilde{f}_3(x_2) \right) + \tilde{f}_3(x_2).$$

Осуществив последовательно элементарные преобразования, получим требуемое:

$$\tilde{\phi}(x_1) = \frac{p_1 - \tilde{f}_1(x_2)}{\tilde{f}_1(x_1) - \tilde{f}_1(x_2)} = \frac{p_2 - \tilde{f}_2(x_2)}{\tilde{f}_2(x_1) - \tilde{f}_2(x_2)} = \frac{p_3 - \tilde{f}_3(x_2)}{\tilde{f}_3(x_1) - \tilde{f}_3(x_2)}.$$

В целях дальнейшего изложения теоретического аппарата введём следующие обозначения:

$$A_{ij} = \begin{pmatrix} p_{ij} & p_{i\bar{j}} \\ p_{\bar{i}j} & p_{\bar{i}\bar{j}} \end{pmatrix},$$

$$A_{ij|k} = \begin{pmatrix} p_{ijk} & p_{i\bar{j}k} \\ p_{\bar{i}jk} & p_{\bar{i}\bar{j}k} \end{pmatrix},$$

$$A_{ij|\bar{k}} = \begin{pmatrix} p_{ij\bar{k}} & p_{i\bar{j}\bar{k}} \\ p_{\bar{i}j\bar{k}} & p_{\bar{i}\bar{j}\bar{k}} \end{pmatrix}.$$

Тогда определители вышеуказанных матриц равны $\|A_{ij}\| = p_{ij}p_{\bar{i}\bar{j}} - p_{\bar{i}j}p_{i\bar{j}}$, $\|A_{ij|k}\| = p_{ijk}p_{\bar{i}\bar{j}k} - p_{\bar{i}jk}p_{i\bar{j}k}$, $\|A_{ij|\bar{k}}\| = p_{ij\bar{k}}p_{\bar{i}\bar{j}\bar{k}} - p_{\bar{i}j\bar{k}}p_{i\bar{j}\bar{k}}$ и называются произведением i -го и j -го показателей при условии (или без такового), что k -ый показатель равен 0 или 1.

В основе поиска неизвестных вероятностей простейшей модели можно воспользоваться нижеследующими таблицами совместных распределений:

Таблица 1. Совместное распределение двух альтернативных показателей i и j .

i -ый показатель / j -ый показатель	1	0	
1	p_{ij}	$p_{i\bar{j}}$	p_i
0	$p_{\bar{i}j}$	$p_{\bar{i}\bar{j}}$	$1 - p_i$
	p_j	$1 - p_j$	

Таблица 2. Совместное распределение двух альтернативных показателей i -го и j -го при условии, что k -ый показатель равен 1.

i -ый показатель / j -ый показатель	1	0	
1	p_{ijk}	$p_{i\bar{j}k}$	p_{ik}
0	$p_{\bar{i}jk}$	$p_{\bar{i}\bar{j}k}$	$p_k - p_{ik}$
	p_{jk}	$p_k - p_{jk}$	

Таблица 3. Совместное распределение двух альтернативных показателей i -го и j -го при условии, что k -ый показатель равен 0.

i -ый показатель / j -ый показатель	1	0	
1	$p_{ij\bar{k}}$	$p_{i\bar{j}\bar{k}}$	$p_{i\bar{k}}$
0	$p_{\bar{i}j\bar{k}}$	$p_{\bar{i}\bar{j}\bar{k}}$	$1 - p_k - p_{i\bar{k}}$
	$p_{j\bar{k}}$	$1 - p_k - p_{j\bar{k}}$	

Лемма. Отношение определителей матриц $\|A_{ij}\|$ и $\|A_{ij|k}\|$ равно произведению вероятностей $\tilde{f}_i(x_1)$ и $\tilde{f}_j(x_1)$.

Доказательство. Анализ таблицы 1 и таблицы 2 позволяет сформировать следующий список вспомогательных равенств:

$$\begin{aligned}
 p_{\bar{i}j} &= p_j - p_{ij}; \\
 p_{i\bar{j}} &= p_i - p_{ij}; \\
 p_{\bar{i}\bar{j}} &= 1 - p_i - p_j + p_{ij}; \\
 p_{\bar{i}jk} &= p_{jk} - p_{ijk};
 \end{aligned}$$

$$p_{i\bar{j}k} = p_{ik} - p_{ijk};$$

$$p_{\bar{i}jk} = p_k - p_{ik} - p_{jk} + p_{ijk}.$$

Используя эти равенства, можно преобразовать определители $\|A_{ij}\|$ и $\|A_{ij|k}\|$ к следующему виду:

$$\begin{aligned} \|A_{ij}\| &= p_{ij}p_{\bar{i}j} - p_{\bar{i}j}p_{ij} \\ &= p_{ij}(1 - p_i - p_j + p_{ij}) - (p_j - p_{ij})(p_i - p_{ij}) \\ &= p_{ij} - p_{ij}p_i - p_{ij}p_j + p_{ij}^2 - (p_jp_i - p_jp_{ij} - p_{ij}p_i + p_{ij}^2) \\ &= p_{ij} - p_{ij}p_i - p_{ij}p_j + p_{ij}^2 - p_jp_i + p_jp_{ij} + p_{ij}p_i - p_{ij}^2 = p_{ij} - p_i p_j; \end{aligned}$$

$$\begin{aligned} \|A_{ij|k}\| &= p_{ijk}p_{\bar{i}j\bar{k}} - p_{\bar{i}jk}p_{i\bar{j}k} \\ &= p_{ijk}(p_k - p_{jk} - p_{ik} + p_{ijk}) - (p_{ik} - p_{ijk})(p_{jk} - p_{ijk}) \\ &= p_{ijk}(p_k - p_{jk} - p_{ik} + p_{ijk}) - (p_{ik}p_{jk} - p_{ik}p_{ijk} - p_{ijk}p_{jk} + p_{ijk}^2) \\ &= p_{ijk}p_k - p_{ijk}p_{jk} - p_{ijk}p_{ik} + p_{ijk}^2 - p_{ik}p_{jk} + p_{ik}p_{ijk} + p_{ijk}p_{jk} - p_{ijk}^2 \\ &= p_{ijk}p_k - p_{ik}p_{jk}. \end{aligned}$$

Подставив вместо маргиналов в данных определителях соответствующие выражения вероятностей $\tilde{f}_j(x_i)$ и частот $\tilde{\phi}(x_i)$ и осуществив элементарные преобразования, получим:

$$\begin{aligned} \|A_{ij}\| &= p_{ij} - p_i p_j = \tilde{f}_{ij}(x_1)\tilde{\phi}(x_1) + \tilde{f}_{ij}(x_2)\tilde{\phi}(x_2) - \left(\tilde{f}_i(x_1)\tilde{\phi}(x_1) + \tilde{f}_i(x_2)\tilde{\phi}(x_2) \right) \times \\ &\quad \times \left(\tilde{f}_j(x_1)\tilde{\phi}(x_1) + \tilde{f}_j(x_2)\tilde{\phi}(x_2) \right) \\ &= \tilde{f}_i(x_1)\tilde{f}_j(x_1)\tilde{\phi}(x_1) + \tilde{f}_i(x_2)\tilde{f}_j(x_2)\tilde{\phi}(x_2) - \tilde{f}_i(x_1)\tilde{\phi}(x_1)\tilde{f}_j(x_1)\tilde{\phi}(x_1) - \\ &\quad - \tilde{f}_i(x_1)\tilde{\phi}(x_1)\tilde{f}_j(x_2)\tilde{\phi}(x_2) - \tilde{f}_i(x_2)\tilde{\phi}(x_2)\tilde{f}_j(x_1)\tilde{\phi}(x_1) - \tilde{f}_i(x_2)\tilde{\phi}(x_2)\tilde{f}_j(x_2)\tilde{\phi}(x_2) \\ &= \left(\tilde{\phi}(x_1) + \tilde{\phi}(x_2) \right) \left(\tilde{f}_i(x_1)\tilde{f}_j(x_1)\tilde{\phi}(x_1) + \tilde{f}_i(x_2)\tilde{f}_j(x_2)\tilde{\phi}(x_2) \right) - \\ &\quad - \tilde{f}_i(x_1)\tilde{f}_j(x_1)\tilde{\phi}^2(x_1) - \tilde{f}_i(x_1)\tilde{\phi}(x_1)\tilde{f}_j(x_2)\tilde{\phi}(x_2) - \tilde{f}_i(x_2)\tilde{\phi}(x_2)\tilde{f}_i(x_1)\tilde{\phi}(x_1) \\ &\quad - \tilde{f}_i(x_2)\tilde{f}_j(x_2)\tilde{\phi}^2(x_2) \\ &= \tilde{\phi}(x_1)\tilde{f}_i(x_1)\tilde{f}_j(x_1)\tilde{\phi}(x_1) + \tilde{\phi}(x_1)\tilde{f}_i(x_2)\tilde{f}_j(x_2)\tilde{\phi}(x_2) + \tilde{\phi}(x_2)\tilde{f}_i(x_1)\tilde{f}_j(x_1)\tilde{\phi}(x_1) \\ &\quad + \tilde{\phi}(x_2)\tilde{f}_i(x_2)\tilde{f}_j(x_2)\tilde{\phi}(x_2) - \tilde{f}_i(x_1)\tilde{f}_j(x_1)\tilde{\phi}^2(x_1) - \tilde{f}_i(x_1)\tilde{\phi}(x_1)\tilde{f}_j(x_2)\tilde{\phi}(x_2) \\ &\quad - \tilde{f}_i(x_2)\tilde{\phi}(x_2)\tilde{f}_i(x_1)\tilde{\phi}(x_1) - \tilde{f}_i(x_2)\tilde{f}_j(x_2)\tilde{\phi}^2(x_2) \\ &= \tilde{\phi}^2(x_1)\tilde{f}_i(x_1)\tilde{f}_j(x_1) + \tilde{\phi}(x_1)\tilde{f}_i(x_2)\tilde{f}_j(x_2)\tilde{\phi}(x_2) + \tilde{\phi}(x_2)\tilde{f}_i(x_1)\tilde{f}_j(x_1)\tilde{\phi}(x_1) \\ &\quad + \tilde{\phi}^2(x_2)\tilde{f}_i(x_2)\tilde{f}_j(x_2) - \tilde{f}_i(x_1)\tilde{f}_j(x_1)\tilde{\phi}^2(x_1) - \tilde{f}_i(x_1)\tilde{\phi}(x_1)\tilde{f}_j(x_2)\tilde{\phi}(x_2) \\ &\quad - \tilde{f}_i(x_2)\tilde{\phi}(x_2)\tilde{f}_i(x_1)\tilde{\phi}(x_1) - \tilde{f}_i(x_2)\tilde{f}_j(x_2)\tilde{\phi}^2(x_2) \end{aligned}$$

$$\begin{aligned}
 &= \tilde{\phi}(x_1)\tilde{f}_i(x_2)\tilde{f}_j(x_2)\tilde{\phi}(x_2) + \tilde{\phi}(x_2)\tilde{f}_i(x_1)\tilde{f}_j(x_1)\tilde{\phi}(x_1) - \tilde{f}_i(x_1)\tilde{\phi}(x_1)\tilde{f}_j(x_2)\tilde{\phi}(x_2) \\
 &\quad - \tilde{f}_i(x_2)\tilde{\phi}(x_2)\tilde{f}_i(x_1)\tilde{\phi}(x_1) \\
 &= \tilde{\phi}(x_1)\tilde{\phi}(x_2) \left(\tilde{f}_i(x_2)\tilde{f}_j(x_2) + \tilde{f}_i(x_1)\tilde{f}_j(x_1) - \tilde{f}_i(x_1)\tilde{f}_j(x_2) - \tilde{f}_i(x_2)\tilde{f}_i(x_1) \right) \\
 &= \tilde{\phi}(x_1)\tilde{\phi}(x_2) \left(\tilde{f}_i(x_2)\tilde{f}_j(x_2) + \tilde{f}_i(x_1)\tilde{f}_j(x_1) - \tilde{f}_i(x_1)\tilde{f}_j(x_2) - \tilde{f}_i(x_2)\tilde{f}_i(x_1) \right) \\
 &= \tilde{\phi}(x_1)\tilde{\phi}(x_2) \left(\tilde{f}_i(x_2) \left(\tilde{f}_j(x_2) - \tilde{f}_i(x_1) \right) - \tilde{f}_i(x_1) \left(\tilde{f}_j(x_2) - \tilde{f}_j(x_1) \right) \right) \\
 &= \tilde{\phi}(x_1)\tilde{\phi}(x_2) \left(\tilde{f}_i(x_2) - \tilde{f}_i(x_1) \right) \left(\tilde{f}_j(x_2) - \tilde{f}_j(x_1) \right).
 \end{aligned}$$

$$\|A_{ijk}\| = p_{ijk}p_k - p_{ik}p_{jk}$$

$$\begin{aligned}
 &= \left(\tilde{f}_{ijk}(x_1)\tilde{\phi}(x_1) + \tilde{f}_{ijk}(x_2)\tilde{\phi}(x_2) \right) \left(\tilde{f}_k(x_1)\tilde{\phi}(x_1) + \tilde{f}_k(x_2)\tilde{\phi}(x_2) \right) - \\
 &\quad - \left(\tilde{f}_{ik}(x_1)\tilde{\phi}(x_1) + \tilde{f}_{ik}(x_2)\tilde{\phi}(x_2) \right) \left(\tilde{f}_{jk}(x_1)\tilde{\phi}(x_1) + \tilde{f}_{jk}(x_2)\tilde{\phi}(x_2) \right) \\
 &= \left(\tilde{f}_{ijk}(x_1)\tilde{\phi}(x_1) + \tilde{f}_{ijk}(x_2)\tilde{\phi}(x_2) \right) \left(\tilde{f}_k(x_1)\tilde{\phi}(x_1) + \tilde{f}_k(x_2)\tilde{\phi}(x_2) \right) - \\
 &\quad - \tilde{f}_{ik}(x_1)\tilde{\phi}(x_1)\tilde{f}_{jk}(x_1)\tilde{\phi}(x_1) - \tilde{f}_{ik}(x_1)\tilde{\phi}(x_1)\tilde{f}_{jk}(x_2)\tilde{\phi}(x_2) - \\
 &\quad - \tilde{f}_{ik}(x_2)\tilde{\phi}(x_2)\tilde{f}_{jk}(x_1)\tilde{\phi}(x_1) - \tilde{f}_{ik}(x_2)\tilde{\phi}(x_2)\tilde{f}_{jk}(x_2)\tilde{\phi}(x_2) \\
 &= \tilde{f}_k(x_1)\tilde{\phi}(x_1)\tilde{f}_{ijk}(x_1)\tilde{\phi}(x_1) + \tilde{f}_k(x_1)\tilde{\phi}(x_1)\tilde{f}_{ijk}(x_2)\tilde{\phi}(x_2) + \\
 &\quad + \tilde{f}_k(x_2)\tilde{\phi}(x_2)\tilde{f}_{ijk}(x_1)\tilde{\phi}(x_1) + \tilde{f}_k(x_2)\tilde{\phi}(x_2)\tilde{f}_{ijk}(x_2)\tilde{\phi}(x_2) \\
 &\quad - \tilde{\phi}^2(x_1)\tilde{f}_{ik}(x_1)\tilde{f}_{jk}(x_1) - \tilde{f}_{ik}(x_1)\tilde{\phi}(x_1)\tilde{f}_{jk}(x_2)\tilde{\phi}(x_2) - \\
 &\quad - \tilde{f}_{ik}(x_2)\tilde{\phi}(x_2)\tilde{f}_{jk}(x_1)\tilde{\phi}(x_1) - \tilde{\phi}^2(x_2)\tilde{f}_{ik}(x_2)\tilde{f}_{jk}(x_2) \\
 &= \tilde{f}_k(x_1)\tilde{\phi}^2(x_1)\tilde{f}_{ijk}(x_1) + \tilde{f}_k(x_1)\tilde{\phi}(x_1)\tilde{f}_{ijk}(x_2)\tilde{\phi}(x_2) + \\
 &\quad + \tilde{f}_k(x_2)\tilde{\phi}(x_2)\tilde{f}_{ijk}(x_1)\tilde{\phi}(x_1) + \tilde{f}_k(x_2)\tilde{\phi}^2(x_2)\tilde{f}_{ijk}(x_2) \\
 &\quad - \tilde{\phi}^2(x_1)\tilde{f}_{ik}(x_1)\tilde{f}_{jk}(x_1) - \tilde{f}_{ik}(x_1)\tilde{\phi}(x_1)\tilde{f}_{jk}(x_2)\tilde{\phi}(x_2) - \\
 &\quad - \tilde{f}_{ik}(x_2)\tilde{\phi}(x_2)\tilde{f}_{jk}(x_1)\tilde{\phi}(x_1) - \tilde{\phi}^2(x_2)\tilde{f}_{ik}(x_2)\tilde{f}_{jk}(x_2) \\
 &= \tilde{\phi}^2(x_1) \left(\tilde{f}_k(x_1)\tilde{f}_{ijk}(x_1) \right) - \tilde{f}_{ik}(x_1)\tilde{f}_{jk}(x_1) \Big) + \\
 &\quad + \tilde{\phi}^2(x_2) \left(\tilde{f}_k(x_2)\tilde{f}_{ijk}(x_2) \right) - \tilde{f}_{ik}(x_2)\tilde{f}_{jk}(x_2) \Big) + \\
 &\quad + \tilde{f}_k(x_1)\tilde{\phi}(x_1)\tilde{f}_{ijk}(x_2)\tilde{\phi}(x_2) + \tilde{f}_k(x_2)\tilde{\phi}(x_2)\tilde{f}_{ijk}(x_1)\tilde{\phi}(x_1) - \\
 &\quad - \tilde{f}_{ik}(x_1)\tilde{\phi}(x_1)\tilde{f}_{jk}(x_2)\tilde{\phi}(x_2) - \tilde{f}_{ik}(x_2)\tilde{\phi}(x_2)\tilde{f}_{jk}(x_1)\tilde{\phi}(x_1) \\
 &= \tilde{f}_k(x_1)\tilde{f}_k(x_2)\tilde{\phi}(x_1)\tilde{\phi}(x_2) \left(\tilde{f}_{ij}(x_1) + \tilde{f}_{ij}(x_2) - \tilde{f}_i(x_1)\tilde{f}_j(x_2) - \tilde{f}_i(x_2)\tilde{f}_j(x_1) \right) \\
 &= \tilde{f}_k(x_1)\tilde{f}_k(x_2)\|A_{ij}\|.
 \end{aligned}$$

■

Теорема. Наличие всех маргиналов в простейшей латентно-структурной модели позволяет свести поиск всех неизвестных вероятностей к решению трёх квадратных уравнений.

Доказательство. В первую очередь обратим внимание на, то что из вышедоказанной леммы следует выполнение следующих двух равенств:

$$\|A_{ij|k}\| = \tilde{f}_k(x_1)\tilde{f}_k(x_2)\|A_{ij}\|, \quad \|A_{ij|\bar{k}}\| = \tilde{f}_{\bar{k}}(x_1)\tilde{f}_{\bar{k}}(x_2)\|A_{ij}\|.$$

Или в более наглядной форме

$$\begin{aligned} \|A_{ij|k}\| &= \tilde{f}_k(x_1)\tilde{f}_k(x_2)\|A_{ij}\|, \\ \|A_{ij|\bar{k}}\| &= \left(1 - \tilde{f}_k(x_1)\right) \left(1 - \tilde{f}_k(x_2)\right) \|A_{ij}\|. \end{aligned}$$

Раскрыв скобки и осуществив элементарные преобразования, получим следующий вид системы уравнений с двумя неизвестными:

$$\begin{aligned} \|A_{ij|k}\| &= \tilde{f}_k(x_1)\tilde{f}_k(x_2)\|A_{ij}\|, \\ \|A_{ij|\bar{k}}\| &= \left(1 - \tilde{f}_k(x_1) - \tilde{f}_k(x_2) + \tilde{f}_k(x_1)\tilde{f}_k(x_2)\right) \|A_{ij}\|. \end{aligned}$$

Поделив обе стороны уравнений данной системы с двумя неизвестными на определитель $\|A_{ij}\|$, получим

$$\begin{aligned} \frac{\|A_{ij|k}\|}{\|A_{ij}\|} &= \tilde{f}_k(x_1)\tilde{f}_k(x_2), \\ \frac{\|A_{ij|\bar{k}}\|}{\|A_{ij}\|} &= \left(1 - \tilde{f}_k(x_1) - \tilde{f}_k(x_2) + \tilde{f}_k(x_1)\tilde{f}_k(x_2)\right). \end{aligned}$$

Если осуществить подстановку первого уравнения во второе и разместить неизвестные в левой части уравнений, а известные – в правой, то получим классическую теорему Виета:

$$\begin{aligned} \tilde{f}_k(x_1)\tilde{f}_k(x_2) &= \frac{\|A_{ij|k}\|}{\|A_{ij}\|}, \\ \tilde{f}_k(x_1) + \tilde{f}_k(x_2) &= \left(1 - \frac{\|A_{ij|\bar{k}}\|}{\|A_{ij}\|} + \frac{\|A_{ij|k}\|}{\|A_{ij}\|}\right), \end{aligned}$$

где $\tilde{f}_k(x_1)$ и $\tilde{f}_k(x_2)$ – корни квадратного уравнения

$$x^2 - \left(1 - \frac{\|A_{ij|\bar{k}}\|}{\|A_{ij}\|} + \frac{\|A_{ij|k}\|}{\|A_{ij}\|}\right)x + \frac{\|A_{ij|k}\|}{\|A_{ij}\|} = 0.$$

Если же рассматривать каждый показатель как условный, то соответственно получим ещё два квадратных уравнения:

$$y^2 - \left(1 - \frac{\|A_{jk|\bar{i}}\|}{\|A_{jk}\|} + \frac{\|A_{jk|i}\|}{\|A_{jk}\|}\right)y + \frac{\|A_{jk|i}\|}{\|A_{jk}\|} = 0,$$

$$z^2 - \left(1 - \frac{\|A_{ik|\bar{j}}\|}{\|A_{ik}\|} + \frac{\|A_{ik|j}\|}{\|A_{ik}\|}\right) z + \frac{\|A_{ik|j}\|}{\|A_{ik}\|} = 0,$$

где $\tilde{f}_i(x_1)$ и $\tilde{f}_i(x_2)$ – корни первого уравнения, а $\tilde{f}_j(x_1)$ и $\tilde{f}_j(x_2)$ – второго.

Далее предполагается совместное использование латентной модели и ортогональной факторной структуры для построения алгоритма метода ФОРДИАСИМПТ вероятностного метода распознавания на базе альтернативных показателей.

Во-первых, необходимо сформулировать определение диагностической шкалы, симптомокомплекса и независимости симптомокомплексов в методе ФОРДИАСИМПТ.

Определение. *Диагностической шкалой называется набор апостериорных вероятностей, полученных с помощью простейшей латентно-структурной модели и формулы Байеса, позволяющей отнести объект исследования к одному из двух сформированных классов.*

Определение. *Симптомокомплекс – тройка альтернативных показателей, используемых для построения диагностической шкалы в методе ФОРДИАСИМПТ.*

Определение. *Два симптомокомплекса считаются зависимыми, если они содержат один и более общих параметров, в противном случае они независимы.*

Во-вторых, в целях упрощения дальнейшего изложения, введем функцию

$$\gamma_l(y_{i_kj}) = \begin{cases} \tilde{f}_i^{(k)}(x_l) & \text{если } y_{i_kj} = 1, \\ 1 - \tilde{f}_i^{(k)}(x_l) & \text{если } y_{i_kj} = 0, \end{cases}$$

где l – номер класса и может принимать значение 1 или 2, k – номер номер симптомокомплекса, $\tilde{f}_i^{(k)}(x_l)$ – вероятность значения 1 по i_k -му показателю у объекта из l -ого класса. Выбранный исследователем i_k -ый показатель входит в состав k -го симптомокомплекса, полученного на основании анализа факторной структуры.

Тогда вероятность принадлежности к первому классу можно определить посредством формулы Байеса с использованием введённой функции

$$P(1|y_{a_kj}, y_{b_kj}, y_{c_kj}) = \frac{\gamma_1(y_{a_kj})\gamma_1(y_{b_kj})\gamma_1(y_{c_kj})\tilde{\phi}_k(x_1)}{\sum_{i=1}^2 \gamma_i(y_{a_kj})\gamma_i(y_{b_kj})\gamma_i(y_{c_kj})\tilde{\phi}_k(x_i)}, \quad (2)$$

где a_k, b_k, c_k – номера трёх параметров k -го симптомокомплекса.

Алгоритм ФОРДИАСИМПТ:

1. Из матрицы Y путём элементарного преобразования получаем стандартизованную матрицу Z размерности $m \times n$ [1-3].
2. Вычисляем корреляционную матрицу R размерности $m \times m$ [1-3].

3. С целью исключения незначимых показателей вычисляем вероятностные значения уровней зависимости по формуле $\chi^2 = n \cdot \varphi$ при единичной степени свободы.
4. Определяем наименьшее количество выделяемых факторов (критерий Гуттмана, критерий «каменной осыпи» или другой адекватный критерий) [1-3].
5. Находим общности любым из известных методов (лучше взять метод минимальных остатков) [1-3].
6. Вычисляем первичную ортогональную матрицу весовых нагрузок факторов A размерности $m \times r$ (метод главных факторов, метод минимальных остатков или любой другой адекватный метод) [1-3].
7. Полученную на предыдущем шаге матрицу весовых нагрузок подвергаем ортогональному вращению в соответствии с варимакс критерием [2-3,7,8].
8. Осуществляем анализ ортогональной факторной структуры, полученной после вращения и формируем зависимые и независимые симптомокомплексы [3].
9. Для каждого симптомокомплекса формируем диагностическую шкалу, вычисляя маргиналы и решая систему уравнений (1), используя результаты теоремы.
10. По формуле (2) вычисляем частные апостериорные вероятности для всех объектов исследования.

Утверждение. Алгоритм ФОРДИАСИМПТ используется для распознавания двух образов в пространстве двоичных признаков при совпадении количества выделенных измеряемых факторов и полученных независимых симптомокомплексов.

Доказательство. В силу ортогональности выделяемых факторов получаем, что группы параметров, наполняющие тот или иной фактор, очень слабо коррелируют между собой. Тогда можно использовать условия независимости частных апостериорных вероятностей для получения формулы общей апостериорной вероятности:

$$P(1|y_{a1j}, y_{b1j}, y_{c1j}, \dots, y_{arj}, y_{brj}, y_{crj}) = \prod_{i=1}^r P(1|y_{a_{ij}}, y_{b_{ij}}, y_{c_{ij}}). \quad (3)$$

■
Замечание. Объект распознавания в случае независимых симптомокомплексов относится к первому классу, если общая апостериорная вероятность из формулы (3) меньше или равна 0,5, и ко второму классу – в противоположном случае.

Вычислительный алгоритм КОРА

Алгоритм Кора (комбинаторного распознавания) — алгоритм классификации (взвешенного голосования правил), предложенный М. Вайнцвайгом и М. Бонгардом в 1973 г. [9-10], который применяется как метод распознавания двух образов в пространстве двоичных признаков путём построения и анализа набора конъюнктивных закономерностей.

Полагаем, что даны две таблицы «объект-свойство». Обозначим количество объектов исследования в каждой таблице за n и k соответственно, а количество измеряемых параметров за m — размерность выборки. Тогда исходные альтернативные данные представляются в виде таблиц, столбцы которой — объекты исследования, а строки — значения измеряемых параметров у конкретного объекта. На первом шаге алгоритма осуществляется объединение исходных таблиц в одну матрицу Y размерности $m \times (n + k)$ с сохранением информации о классовой принадлежности объектов. Далее просматриваются всевозможные матрицы Y_{ruv} размерности $3 \times (n + k)$, получаемые извлечениями из матрицы Y , что составляет C_m^3 извлечения.

$$Y_{ruv} = \left(\begin{array}{cccc|cccc} y_{r1} & y_{r2} & \dots & y_{rn} & y_{r(n+1)} & \dots & y_{r(n+k)} & \\ y_{u1} & y_{u2} & \dots & y_{un} & y_{u(n+1)} & \dots & y_{u(n+k)} & \\ y_{v1} & y_{v2} & \dots & y_{vn} & y_{v(n+1)} & \dots & y_{v(n+k)} & \end{array} \right),$$

На следующем шаге среди первых n столбцов матрицы Y_{ruv} выделяются и фиксируются все тройки, не совпадающие ни с одной из троек в столбцах с $(n + 1)$ по $(n + k)$. Аналогичная операция осуществляется со столбцами с $(n + 1)$ по $(n + k)$. Из полученных двух совокупностей троек формируются два множества, которые обозначим за $A_{ruv} = \{(y_{ri}, y_{ui}, y_{vi})\}$ и $B_{ruv} = \{(y_{rj}, y_{uj}, y_{vj})\}$. Множества A_{ruv} и B_{ruv} называются характеристиками классов A и B . Эти характеристики формируются по всевозможным матрицам Y_{ruv} .

Основной этап алгоритма КОРА заключается в распознавании некоторого объекта $x = (x_1 \dots x_r \dots x_u \dots x_v \dots x_m)$ относительно классов A и B . Число совпадений $(x_r, x_u, x_v) = (y_{ri}, y_{ui}, y_{vi})$ обозначим $V(x, A)$ — число голосов, поданных для x за класс A , аналогично $V(x, B)$ — число голосов $(x_r, x_u, x_v) = (y_{rj}, y_{uj}, y_{vj})$. Если $V(x, A) > V(x, B)$, то объект относится к первому классу, если $V(x, A) < V(x, B)$, то к второму, при равенстве алгоритм отказывается от классификации.

Некоторые авторы предлагают вводить пороговое значение L . В этом случае если $V(x, A) - L > V(x, B)$, то объект относится к первому классу, если $V(x, A) < V(x, B) - L$, то к второму, при равенстве или нулевом значении голосов алгоритм отказывается от классификации.

Сравнительный анализ алгоритмов КОРА и ФОРДИАСИМПТ на независимых симптомокомплексах

Отметим, что алгоритм ФОРДИАСИМПТ статистического типа, а алгоритм КОРА относится к категории эмпирических алгоритмов. В этой связи не имеет особого смысла сравнивать их на объёмах малых и средних выборок. На очень малых выборках алгоритм ФОРДИАСИМПТ нецелесообразно использовать из-за низкого уровня значимости коэффициента φ по χ^2 критерию. Вычислительные эксперименты на независимых симптомокомплексах с средними выборками показали, что результаты обоих алгоритмов приблизительно одинаковы. А в том случае, когда обрабатывается выборка большого объёма, выявляются существенные расхождения. В качестве примера таких исходных данных для вычислительного эксперимента предлагаются альтернативные показатели с объёмом $n = 150$. Основные результаты работы алгоритма ФОРДИАСИМПТ представлены в нижеследующих таблицах (для краткости будем обозначать симптомокомплекс как С-комплекс).

Таблица 4. Матрица коэффициентов корреляции.

1	2	3	4	5	6
1	0,293	0,362	-0,0611	-0,0215	-0,0159
0,293	1	0,418	-0,173	0,0189	0,0586
0,362	0,418	1	-0,0161	0,0509	0
-0,0611	-0,173	-0,0161	1	0,23	0,207
-0,0215	0,0189	0,0509	0,23	1	0,39
-0,0159	0,0586	0	0,207	0,39	1

Таблица 5. Матрица значимости φ коэффициентов по χ^2 распределению.

1	2	3	4	5	6
1,000	0,999	0,999	0,545	0,207	0,154
0,999	1,000	0,999	0,167	0,183	0,527
0,999	0,999	1,000	0,156	0,466	0,000
0,545	0,167	0,156	1,000	0,995	0,988
0,207	0,183	0,466	0,995	1,000	0,999
0,154	0,527	0,000	0,988	0,999	1,000

При использовании алгоритма КОРА получены следующие характеристики классов:

Таблица 6. Матрица ортогонального факторного отображения после варимакс вращения.

	Фактор №1	Фактор №2
1	0,699	0,0441
2	0,769	0,012
3	0,785	-0,061
4	-0,198	-0,597
5	0,0517	-0,784
6	0,0505	-0,765

Таблица 7. Основные показатели симптомокомплекса №1.

Маргиналы	Значения частот и априорных вероятностей	Варианты			Апостериорная вероятность
		ответов			
$p_1 = 0,266$	$\tilde{\phi}_1 = 0,770$	1	1	1	0,007
$p_2 = 0,166$	$\tilde{\phi}_2 = 0,230$	0	1	1	0,086
$p_3 = 0,3$	$\tilde{f}_1(x_1) = 0,048$	0	0	1	0,705
$p_{12} = 0,113$	$\tilde{f}_2(x_1) = 0,174$	0	0	0	0,992
$p_{13} = 0,153$	$\tilde{f}_3(x_1) = 0,096$	1	0	0	0,914
$p_{23} = 0,1$	$\tilde{f}_1(x_2) = 0,849$	1	1	0	0,294
$p_{123} = 0,08$	$\tilde{f}_2(x_2) = 0,564$	1	0	1	0,162
	$\tilde{f}_3(x_2) = 0,723$	0	1	0	0,837

Таблица 8. Основные показатели симптомокомплекса №2.

Маргиналы	Значения частот и априорных вероятностей	Варианты			Апостериорная вероятность
		ответов			
$p_1 = 0,66$	$\tilde{\phi}_1 = 0,746$	1	1	1	0,995
$p_2 = 0,7$	$\tilde{\phi}_2 = 0,254$	0	1	1	0,972
$p_3 = 0,51$	$\tilde{f}_1(x_1) = 0,842$	0	0	1	0,486
$p_{12} = 0,546$	$\tilde{f}_2(x_1) = 0,858$	0	0	0	0,046
$p_{13} = 0,393$	$\tilde{f}_3(x_1) = 0,615$	1	0	0	0,220
$p_{23} = 0,406$	$\tilde{f}_1(x_2) = 0,125$	1	1	0	0,913
$p_{123} = 0,333$	$\tilde{f}_2(x_2) = 0,235$	1	0	1	0,857
	$\tilde{f}_3(x_2) = 0,214$	0	1	0	0,643

Таблица 9. Таблица исходных объектов исследования с частотой встерчаемости в каждом классе.

Класс А						Частота	Класс В						Частота
0	1	1	1	1	1	0.015625	0	0	0	1	1	1	0.372093
0	0	0	1	0	0	0.078125	1	0	0	0	1	1	0.034884
1	1	1	1	1	1	0.046875	1	0	0	1	1	1	0.058140
1	0	1	1	0	1	0.046875	0	1	0	0	1	0	0.011628
0	0	0	0	0	0	0.171875	0	0	0	0	1	0	0.046512
1	0	0	0	0	1	0.015625	0	0	0	0	1	1	0.197674
0	0	0	0	0	1	0.093750	0	0	1	1	0	1	0.011628
1	0	0	1	0	0	0.015625	1	0	0	1	1	0	0.023256
0	1	0	0	0	1	0.031250	0	0	1	1	1	0	0.034884
1	1	1	0	1	1	0.093750	0	0	0	1	0	1	0.058140
1	0	1	0	0	1	0.015625	0	0	0	1	1	0	0.034884
1	0	1	0	0	0	0.031250	1	0	0	1	0	1	0.011628
1	1	0	0	1	1	0.015625	0	0	1	1	1	1	0.034884
1	0	0	0	0	0	0.078125	0	1	0	1	1	1	0.011628
1	0	1	1	1	1	0.062500	0	0	1	0	1	1	0.034884
1	1	1	0	0	0	0.015625	0	1	0	1	0	1	0.011628
1	1	0	0	0	1	0.015625	1	0	0	0	1	0	0.011628
1	1	0	1	1	1	0.015625							
1	1	1	0	1	0	0.015625							
0	1	1	0	1	1	0.015625							
0	0	1	0	1	0	0.015625							
0	1	1	0	0	1	0.015625							
0	1	1	0	0	0	0.031250							
0	0	1	1	0	0	0.015625							
1	1	1	1	1	0	0.015625							
1	0	1	0	1	1	0.015625							

$$\begin{aligned}
A_{123} &= \{(011), (111), (101), (110)\}, A_{124} = \{(111), (110)\}, A_{125} = \{(111), (110)\}, \\
A_{126} &= \{(111), (110)\}, A_{134} = \{(111), (110)\}, A_{135} = \{(111), (110)\}, \\
A_{136} &= \{(111), (110)\}, A_{145} = \{(000), (100)\}, A_{156} = \{(000), (100)\}, \\
A_{234} &= \{(111), (110)\}, A_{235} = \{(111), (110)\}, A_{236} = \{(111), (110)\}, \\
A_{245} &= \{(000), (100)\}, A_{246} = \{(101), (110)\}, A_{256} = \{(000), (100)\}, \\
A_{345} &= \{(000), (100)\}, A_{346} = \{(100)\}, A_{356} = \{(000), (100)\}, \\
A_{456} &= \{(100), (000), (001)\}, \\
B_{135} &= \{(001)\}, B_{235} = \{(001)\}, B_{236} = \{(100)\}, B_{245} = \{(110)\}, B_{356} = \{(010)\}.
\end{aligned}$$

В таблицах №11 и №12 вместо количества голосов в алгоритме КОРА для нераспознанных объектов стоят вопросы.

Т. о. при работе алгоритма КОРА имеются объекты, которые остались не распознанными, а алгоритм ФОРДИАСИМПТ позволяет распознать все объекты исследования. За исключением нераспознанных объектов, алгоритмы сработали одинаково. Для того чтобы понять, почему алгоритм КОРА не смог распознать некоторые объекты, достаточно сравнить нижеследующие распознанные объекты из класса A с первым из нераспознанных объектов из класса B .

Анализ данной таблицы показывает, что объект класса B не распознан в силу полного перекрытия тройками объектов класса A . Аналогичная ситуация наблюдается и у остальных нераспознанных объектов. Возможно, подобная ситуация редко встречается в геологии или сейсмологии [11-12], однако, в медицине и социологии этот случай встречается довольно часто [3,5].

Резюмируя изложенное, отметим, что преимуществом алгоритма ФОРДИАСИМПТ является возможность подтвердить или опровергнуть саму возможность разбиения объектов на два класса, основываясь на полученной статистической информации. Требование минимального пересечения классов является, например, ключевым для медицинских исследований [13,14].

Выводы

В рамках теории латентно-структурной модели сформулированы и доказаны:

1. Утверждение, позволяющее находить относительный объем соответствующего класса через каноническое уравнение прямой и утверждение, позволяющее использовать алгоритм ФОРДИАСИМПТ для распознавания двух образов в пространстве двоичных признаков при независимых симптомокомплексах.

2. Лемма о связи отношения определителей матриц $\|A_{ij}\|$ и $\|A_{ij|k}\|$ с произведением вероятностей $\tilde{f}_i(x_l)$ и $\tilde{f}_j(x_l)$

3. Теорема о сведении решения системы уравнений латентно-структурной модели к решению трёх квадратных уравнений.

На базе полученных теоретических выкладок построен и апробирован вычислительный алгоритм ФОРДИАСИМПТ, позволяющий строить диагностические симптомокомплексы на базе альтернативных данных, ортогональной

Таблица 10. Таблица исходных объектов класса А с результатом распознавания алгоритмами ФОРДИАСИМПТ и КОРА.

Класс А						Вероятность в первом С-комплексе	Вероятность во втором С-комплексе	Байесовская вероятность	Число голосов в КОРА
0	1	1	1	1	1	0.0861	0.9952	0.0856	4
0	0	0	1	0	0	0.9924	0.2199	0.2182	4
1	1	1	1	1	1	0.0076	0.9952	0.0075	10
1	0	1	1	0	1	0.1622	0.8471	0.1374	4
0	0	0	0	0	0	0.9924	0.0458	0.0455	7
1	0	0	0	0	1	0.9138	0.4856	0.4437	4
0	0	0	0	0	1	0.9924	0.4856	0.4819	4
1	0	0	1	0	0	0.9138	0.2199	0.2009	4
0	1	0	0	0	1	0.8375	0.4856	0.4067	5
1	1	1	0	1	1	0.0076	0.9725	0.0073	11
1	0	1	0	0	1	0.1622	0.4856	0.0788	8
1	0	1	0	0	0	0.1622	0.0458	0.0074	12
1	1	0	0	1	1	0.2942	0.9725	0.2861	5
1	0	0	0	0	0	0.9138	0.0458	0.0419	7
1	0	1	1	1	1	0.1622	0.9952	0.1614	4
1	1	1	0	0	0	0.0076	0.0458	0.0003	18
1	1	0	0	0	1	0.2942	0.4856	0.1429	9
1	1	0	1	1	1	0.2942	0.9952	0.2928	4
1	1	1	0	1	0	0.0076	0.6427	0.0049	11
0	1	1	0	1	1	0.0861	0.9725	0.0837	5
0	0	1	0	1	0	0.7053	0.6427	0.4533	1
0	1	1	0	0	1	0.0861	0.4856	0.0418	9
0	1	1	0	0	0	0.0861	0.0458	0.0039	12
0	0	1	1	0	0	0.7053	0.2199	0.1551	4
1	1	1	1	1	0	0.0076	0.9134	0.0069	11
1	0	1	0	1	1	0.1622	0.9725	0.1577	4

Таблица 11. Таблица исходных объектов класса В с результатом распознавания алгоритмами ФОРДИАСИМПТ и КОРА.

Класс В						Вероятность в первом С-комплексе	Вероятность во втором С-комплексе	Байесовская вероятность	Число голосов в КОРА
0	1	1	1	1	1	0.0861	0.9952	0.0856	4
0	0	0	1	0	0	0.9924	0.2199	0.2182	4
1	1	1	1	1	1	0.0076	0.9952	0.0075	10
1	0	1	1	0	1	0.1622	0.8471	0.1374	4
0	0	0	0	0	0	0.9924	0.0458	0.0455	7
1	0	0	0	0	1	0.9138	0.4856	0.4437	4
0	0	0	0	0	1	0.9924	0.4856	0.4819	4
1	0	0	1	0	0	0.9138	0.2199	0.2009	4
0	1	0	0	0	1	0.8375	0.4856	0.4067	5
1	1	1	0	1	1	0.0076	0.9725	0.0073	11
1	0	1	0	0	1	0.1622	0.4856	0.0788	8
1	0	1	0	0	0	0.1622	0.0458	0.0074	12
1	1	0	0	1	1	0.2942	0.9725	0.2861	5
1	0	0	0	0	0	0.9138	0.0458	0.0419	7
1	0	1	1	1	1	0.1622	0.9952	0.1614	4
1	1	1	0	0	0	0.0076	0.0458	0.0003	18
1	1	0	0	0	1	0.2942	0.4856	0.1429	9
1	1	0	1	1	1	0.2942	0.9952	0.2928	4
1	1	1	0	1	0	0.0076	0.6427	0.0049	11
0	1	1	0	1	1	0.0861	0.9725	0.0837	5
0	0	1	0	1	0	0.7053	0.6427	0.4533	1
0	1	1	0	0	1	0.0861	0.4856	0.0418	9
0	1	1	0	0	0	0.0861	0.0458	0.0039	12
0	0	1	1	0	0	0.7053	0.2199	0.1551	4
1	1	1	1	1	0	0.0076	0.9134	0.0069	11
1	0	1	0	1	1	0.1622	0.9725	0.1577	4

Таблица 12. Таблица объектов независимого контроля с результатом распознавания алгоритмами ФОРДИАСИМПТ и КОРА.

Класс В							Вероятность в первом С-комплексе	Вероятность во втором С-комплексе	Байесовская вероятность	Число голосов в КОРА
1	1	0	0	0	0	0	0.2942	0.0458	0.0135	11(A) > 1(B)
0	0	1	0	1	1		0.7053	0.9725	0.6859	?
0	0	1	0	0	1		0.7053	0.4856	0.3425	4(A) > 0(B)
1	1	1	0	1	1		0.0076	0.9725	0.0073	11(A) > 0(B)
0	0	0	1	1	1		0.9924	0.9952	0.9877	0(A) < 2(B)
0	0	0	0	1	1		0.9924	0.9725	0.9651	0(A) < 2(B)
0	0	0	0	0	0		0.9924	0.0458	0.0455	7(A) > 0(B)
1	0	0	1	1	0		0.9138	0.9134	0.8347	0(A) < 2(B)
0	0	0	0	1	0		0.9924	0.6427	0.6378	0(A) < 3(B)
1	1	1	1	1	1		0.0076	0.9952	0.0075	10(A) > 0(B)
1	0	1	1	1	0		0.1622	0.9134	0.1481	4(A) > 0(B)
1	0	0	0	1	0		0.9138	0.6427	0.5872	0(A) < 2(B)
0	1	1	1	0	0		0.0861	0.2199	0.0189	9(A) > 1(B)
0	0	1	1	0	0		0.7053	0.2199	0.1551	4(A) > 0(B)
0	1	0	1	1	1		0.8375	0.9952	0.8335	0(A) < 1(B)
0	1	1	1	0	1		0.0861	0.8471	0.0729	4(A) > 1(B)
1	0	1	0	0	0		0.1622	0.0458	0.0074	12(A) > 0(B)
1	0	0	0	0	0		0.9138	0.0458	0.0419	7(A) > 0(B)

Таблица 13. Первый нераспознанный объект класса *B* и объекты класса *A*, имеющие пересечение по тройкам.

Часть объектов класса <i>A</i> и нераспознанные объекты класса <i>B</i>						Вероятность в первом С-комплексе	Вероятность во втором С-комплексе	Байесовская вероятность	Число голосов в КОРА
1	0	1	1	0	1	0.1622	0.8471	0.1374	4
0	0	1	0	1	0	0.7053	0.6427	0.4533	1
0	0	1	1	0	0	0.7053	0.2199	0.1551	4
0	1	1	1	1	1	0.0861	0.9952	0.0856	4
0	0	0	0	0	1	0.9924	0.4856	0.4819	4
0	1	0	0	0	1	0.8375	0.4856	0.4067	5
0	1	1	0	1	1	0.0861	0.9725	0.0837	5
0	0	1	1	0	1	0.7053	0.8471	0.5974	?

факторной структуры, простейшей латентно-структурной модели и формулы Байеса.

Проведён сравнительный анализ алгоритмов КОРА и ФОРДИАСИМПТ на выборке объёмом 150 объектов. Показана целесообразность применения алгоритма ФОРДИАСИМПТ для независимых симптомокомплексов при адекватной статистической информации.

ЛИТЕРАТУРА

1. Иберла К. Факторный анализ. М. : Статистика, 1989.
2. Харман Г. Современный факторный анализ. М. : Статистика, 1972.
3. Гольцяпин В.В. Реализация вычислительного алгоритма метода ФОРДИАСИМПТ на примере альтернативных показателей артериальной гипертензии // Современные наукоёмкие технологии 2014. № 11. С. 50–55.
4. Гольцяпин В.В. Вероятностный метод формирования симптомокомплексов. // Математические структуры и моделирование. 2014. № 4(32). С. 53–59.
5. Осипов Г.В. Методы измерения в социологии. М. : Наука, 2003.
6. Lazarsfeld P.F. The logical and mathematical foundation of latent structure analysis. 1950 In: Measurement and Prediction. N.Y.
7. Kaiser H.F. The varimax criterion for analytic rotation in factor analysis // Psychometrika. 1958. № 23. С. 187–200.
8. Saunders D. The rationale for an "oblimax" method of transformation in factor analysis // Psychometrika. 1961. № 26. С. 317–324.
9. Вапник В.Н. Алгоритм обучения распознаванию образов. М. : Советское радио, 1973.
10. Журавлев Ю.И. Математические основы теории прогнозирования (курс лекций). 2008, МГУ.

11. Платоненко И.М. Исследование и реализация алгоритмов распознавания по представительным наборам на базе решения специальных систем булевых уравнений: диссертация на соискание учёной степени кандидата физико-математических наук: 01.01.09. 1973. Москва.
12. Завьялов А.Д. Среднесрочный прогноз землетрясений по комплексу признаков: диссертация на соискание ученой степени доктора физико-математических наук: 25.00.10. 2003. Москва.
13. Гольдяпин В.В., Друк И.В., Нечаева Г.И. Возможности практической реализации принципа динамического наблюдения пациентов с недифференцированной дисплазией соединительной ткани и риском развития неблагоприятных сердечно-сосудистых проявлений // Справочник врача общей практики. 2014. № 19. С. 27–37.
14. Гольдяпин В.В., Шовин В.А. Косоугольная факторная модель артериальной гипертензии первой стадии // Вестник Омского университета. 2010. № 4. С. 120–128.

KORA AND FORDIASIMPT ALGORITHMS AS METHODS OF RECOGNITION OF THE TWO IMAGES IN THE SPACE OF BINARY FEATURES

V.V. Golyapin

Ph.D.(Phys.-Math.), Associate Professor, Senior Researcher, e-mail: golyapin@mail.ru

V.A. Shovin

Researcher, e-mail: v.shovin@mail.ru

Omsk Branch of the Federal State budget institution Science Institute of Mathematics
S.L. Soboleva of Siberian Branch of RAS

Abstract. In this article we proved the assertions, lemma and theorem allowing us to find posterior probabilities based on binary data of orthogonal factor structure. On the basis of theoretical calculations there was build FORDIASIMPT computational algorithm allowing us to build diagnostic symptom complexes based on probabilistic method of pattern recognition. A comparative analysis of FORDIASIMPT and KORA algorithms as a methods of recognition of the two images in the space of binary characters in the case of independent symptom complexes with adequate statistical information is conducted.

Keywords: symptom, factor model, the latent model, correlation analysis, KORA algorithm, the algorithm FORDIASIMPT.

СОЦИАЛЬНАЯ ТОПОЛОГИЯ И ТОПОСЫ ГРОТЕНДИКА

А.К. Гуц¹

д.ф.-м.н., профессор, e-mail: guts@omsu.ru

Л.А. Паутова²

д.с.н., e-mail: pautoval@yandex.ru

¹Омский государственный университет им. Ф.М. Достоевского

²Москва, Фонд общественного мнения

Аннотация. Обсуждаются методология социальной топологии и наиболее общий подход к топологическому видению пространства, предложенному Гротендиком. Описываются топология Гротендика и топосы Гротендика. Показано, как они способствуют «топологическому повороту» в современной социологии.

Ключевые слова: социальная топология, социология, топосы, топос Гротендика, топос пучков, формализация социального пространства.

Введение

Социология представляет собой социальную топологию

П. Бурдьё

Топология Гротендика представляется наиболее естественно как модальный оператор типа «локально имеет место...»

Ф. У. Ловер

В теоретической социологии ещё в конце XX века оформилось направление так называемой социальной топологии. Социология, как заявил француз Пьер Бурдьё, представляет собой социальную топологию [1]. Идею топологии Бурдьё взял из математики, в рамках которой изучаются топологические свойства фигур, то есть свойства, не изменяющиеся, остающиеся инвариантными при любых непрерывных деформациях или при так называемых гомеоморфизмах.

Понятие «социальная топология» позволяет социологам по-новому рассматривать социальные отношения и их различия. Топология выводит на передний план не классы или социальные структуры, а производство, изменчивость и множественность социальных практик. Понятия «социальное пространство», «социальное поле», «топология» прочно прижились в социологии.

В последние годы вновь возник теоретический интерес к топологии, например, в работах Брюно Латура [4] или Джона Ло [5]. Заимствования из топологии дают возможность социологам-теоретикам описывать социальную реальность как сеть отношений, как многополярный, гетерогенный, находящийся в постоянном движении мир. Эта новая для социологии объяснительная схема позволяет анализировать не только субъекты отношений, но и материальное (предметы, вещи) как полноправного участника отношений, «носителя способности действия» (В. Вахштайн) [6].

В математике наука «топология» – это один из самых абстрактных разделов фундаментальной математики. Она изучает *топологические пространства*, под которыми понимаются либо множества или иные абстрактные объекты, которые снабжены *топологией*, или, как ещё говорят, в которых задана топология. Топология в данном случае – это степень близости точек пространства. При рассмотрении перехода (отображения) от одного топологического пространства к другому происходит смена представления о степени близости, и в силу этого важно изучать те переходы, которые близкое переводят в близкое. Такие переходы (отображения) называют *непрерывными*.

Из-за возможности описания непрерывных изменений на науку «топологию», и обратили свои взоры социологи.

Социальный мир видится в социологии как многомерное пространство – социальное пространство, измерения которого соответствуют свойствам, действующим в социальном мире. Действующие свойства, взятые за принцип построения социального пространства, у Бурдьё являются различными видами власти или капиталов¹. Социальный агент, оказавшийся в некоторой позиции (точке) социального пространства, наделён определённой силой и властью (Бурдьё [1]).

Социальное пространство определяется по взаимоисключению (или различению) позиций, которые его образуют, так сказать, как структура рядоположенности социальных позиций (Бурдьё [1]).

Структура социального пространства формируется объективно посредством существующих социальных отношений и субъективно – представлениями людей об окружающем мире.

Для современного социолога «описать социальный мир – значит выявить множество явлений и задать в нем топологию. Социологический смысл топологических структур исчерпывается социальными отношениями» [7, с. 16].

Социальная топология – это такая система подмножеств множества всех явлений социального мира, когда любые объединения и конечные пересечения подмножеств будут принадлежать данной системе² (Н.А. Шматко [7, с. 16]).

¹Капитал – у Бурдьё это продукт объективации социальных отношений, являющихся условиями и предпосылками практик агентов.

²Мы исправили данное Н.А. Шматко определение топологии, уточняя, что пересечения должны быть конечными.

Социальная топология концептуализирует, т. е. осмысливает, определяет социальные явления в терминах математической науки «топологии» через уточнение, разъяснение социологических различий между ними, через установление близости (по основаниям активных свойств) феноменов, которые до этого воспринимались как весьма далёкие с точки зрения «здравого смысла». Разрывая с обыденным смыслом и стихийными социологическими предпонятиями, социальная топология проводит различия между феноменами, «приближает» или «противопоставляет» явления в социальном пространстве. Она позволяет выстроить эффективное социологическое видение социального мира, выступающее, на наш взгляд, действенной основой научного объяснения. Поскольку можно задавать в социальном пространстве различные топологии, то, выстраивая все более «тонкие» и дробные топологии на полной совокупности известных нам социальных явлений, разрывая с ложными отождествлениями и обнаруживая новые различия, социологи получают инструмент для более полного социологического объяснения современной России [7, с. 15-16].

В математической топологии имеются конструкции, моделирующие ситуацию непрерывно меняющегося сложного объекта, под которым можно иметь в виду изменяющуюся или «изменчивую социальную практику». Их называют *пучками*. Совокупность всех пучков над топологическим пространством образует *категорию*.

Французский математик Александр Гротендик нашёл крайне абстрактное обобщения пучков, заменяя широко используемые в XX веке множества на объекты *категорий*, частным случаем которых является теория множеств **Sets**. Иначе говоря, он заменил теоретико-множественные пучки на теоретико-категорные пучки, придумав для этого топологию Гротендика. Построенную совокупность, или категорию пучков он назвал *топосом*.

Каким образом был выбран данный термин? Гротендик пишет: «Название «топос» было выбрано (в связи с понятием «топология» или «топологический»), чтобы наводить на мысль о том, что речь идёт об объекте, в полном смысле слова относящемся к области топологической интуиции. По обилию мысленных образов, которые слово «топос» вызывает, его можно рассматривать как более или менее эквивалент термину «пространство» (топологическое), просто сильнее подчёркивая «топологическую» специфику понятия» [2].

Для Гротендика совокупность (категория) пучков над пространством, т. е. топос в его понимании, – это «арсенал измерительных приборов», содержащий, «воплощающей» то, что наиболее существенно для пространства», «поскольку оказывается возможным «воссоздать» полностью исходное топологическое пространство в терминах «категории пучков» (или арсенала измерительных приборов), ему соответствующей» [2].

В результате появилась возможность изучать непрерывно меняющиеся «социальные практики» на самом общем абстрактном уровне. Категории пучков были названы Гротендиком, как уже говорилось выше, топосами.

Топос, точнее *элементарный топос*, в понимании математиков – это специальная категория; сами же категории изучает наука «теория категорий». Категории состоят из *объектов* и *морфизмов* между ними, которые характеризуют

отношения между объектами. Топосы очень близки к теории множеств **Sets**, но их внутренняя логика не является классической двузначной, точнее, в ней не действует закон исключённого третьего. Математики такие логики называют интуиционистскими.

О возможности использовать категорию пучков в социологии конфликта говорится в работах Саллаха [14].

Социологи используют понятия топология и топос, но у них эти понятия допускают множественное словесное описание, отсутствуют их строгие математические определения, и в силу этого, особенно понятие «топос», трактуется разными авторами так, как они его понимают.

В данной статье показано, как математический топос и топосы в понимании социологов можно свести к единой согласованной картине.

1. Математическая теория категорий

Теория категорий изучает объекты A, B, C, \dots и отношения между ними, называемые морфизмами.

Определение 1. Категория \mathcal{K} включает в себя:

- 1) объекты A, B, C, \dots ;
- 2) морфизмы f, g, h, \dots ;
- 3) каждый морфизм f связан с двумя объектами A, B ; первый называют областью определения морфизма, а второй – областью значений. Используются обозначения $f : A \rightarrow B$ или $A \xrightarrow{f} B$;
- 4) для каждого объекта A имеется тождественный морфизм $1_A : A \rightarrow A$;
- 5) для каждой пары морфизмов $f : A \rightarrow B$ и $g : B \rightarrow C$ определена композиция морфизмов $g \circ f : A \rightarrow C$. Композиция должна удовлетворять двум условиям:
 - (i) Закон идентичности. Если дан морфизм $f : A \rightarrow B$, то $1_B \circ f = f$ и $f \circ 1_A = f$.
 - (ii) Закон ассоциативности. Если даны морфизмы $f : A \rightarrow B$, $g : B \rightarrow C$, $h : C \rightarrow D$, то $(h \circ g) \circ f = h \circ (g \circ f)$.

Совокупность морфизмов из A в B обозначают как $\text{Hom}_{\mathcal{K}}(A, B)$ или просто как $\text{Hom}(A, B)$.

Двойственная категория. Если дана категория \mathcal{K} , то легко строится *двойственная категория* \mathcal{K}^{op} . Она имеет те же самые объекты, что категория \mathcal{K} , а морфизмы получаются из морфизмов категории \mathcal{K} «обращением стрелки», т. е. если в \mathcal{K} есть морфизм $f : A \rightarrow B$, то в \mathcal{K}^{op} рассматривается морфизм $f : B \rightarrow A$.

Подкатегория. Категория \mathcal{K} называется *подкатегорией* категории \mathcal{E} (обозначается это через $\mathcal{K} \subseteq \mathcal{E}$), если

- (i) каждый \mathcal{K} -объект является \mathcal{E} -объектом и
- (j) для любых двух \mathcal{K} -объектов A и B морфизм $A \rightarrow B$ категории \mathcal{K} есть морфизм категории \mathcal{E} .

Подкатегория \mathcal{K} называется *полной подкатегорией* категории \mathcal{E} , если для любых \mathcal{K} -объектов A и B выполняется равенство $\text{Hom}_{\mathcal{K}}(A, B) = \text{Hom}_{\mathcal{E}}(A, B)$, т. е. \mathcal{E} не имеет морфизмов $A \rightarrow B$, не лежащих в \mathcal{K} .

1.1. Примеры категорий

Пример 1. Простым, но очень важным примером категории является теория множеств Кантора **Sets**. Объекты этой категории – это множества A, B, C, D, \dots , а морфизмы – отображения $f : A \rightarrow B, g : C \rightarrow D, \dots$ из множества в множество.

Эта категория является основным инструментом, используемым математиками для описания и представления окружающей нас реальности, а также нашего внутреннего мира. Она основа парадигмы мышления современных математиков, в рамках которой они строят свои абстрактные конструкции, и в рамках которой воспринимают внешний и внутренний мир.

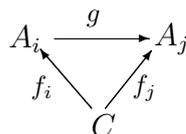
Пример 2. Теория топологических пространств **Тор** также является примером категории. Её объекты – это топологические пространства X, Y, Z, T, \dots , а морфизмы – непрерывные отображения $f : X \rightarrow Y, g : Z \rightarrow T, \dots$

Пример 3. Экзотическим примером категории является категория **N**. Она обладает только одним объектом, обозначаемым как N , а морфизмами являются натуральные числа $0, 1, \dots, n, \dots$, и имеют они запись в следующем виде: $0 : N \rightarrow N, 1 : N \rightarrow N, \dots, n : N \rightarrow N, \dots$ Композицией морфизмов n и m является натуральное число $n \circ m = n + m$. Единичная стрелка 1_N объекта N задаётся числом 0 .

1.2. Диаграммы

Диаграммой в категории \mathcal{K} называется совокупность объектов A_i, A_j, \dots вместе с некоторыми морфизмами $f : A_i \rightarrow A_j$ между отдельными объектами из этой диаграммы (между данной парой объектов может быть несколько морфизмов, а может и не быть их вовсе).

Конусом для диаграммы \mathcal{D} с объектами A_i, A_j, \dots называется такой объект C вместе с морфизмами $f_i : C \rightarrow A_i$ для каждого объекта A_i из \mathcal{D} , что диаграмма



коммукативна для любого морфизма $g : A_i \rightarrow A_j$ из \mathcal{D} , т.е. $g \circ f_i = f_j$.

Конус для диаграммы \mathcal{D} обозначаем через $(f_i : C \rightarrow A_i)$.

Предел диаграммы \mathcal{D} есть конус $(f_i : C \rightarrow A_i)$, такой, что для любого другого конуса $(f'_i : C' \rightarrow A_i)$ для \mathcal{D} существует единственный морфизм $f :$

$C' \rightarrow C$, для которого диаграмма

$$\begin{array}{ccc} & A_i & \\ f'_i \nearrow & & \nwarrow f_i \\ C & \xrightarrow{f} & C' \end{array}$$

коммутативна для каждого объекта A_i из \mathcal{D} .

Диаграмма

$$\begin{array}{ccc} A & \xrightarrow{f} & X \\ g \downarrow & & \downarrow f \\ Y & \xrightarrow{h} & B \end{array}$$

называется *декартовым квадратом*, если

- 1) она коммутативна, т.е. $k \circ f = h \circ g$;
- 2) для любых $\phi : C \rightarrow X$, $\psi : C \rightarrow Y$, для которых коммутативна диаграмма

$$\begin{array}{ccc} C & \xrightarrow{\phi} & X \\ \psi \downarrow & & \downarrow f \\ Y & \xrightarrow{h} & B \end{array}$$

существует единственный морфизм $j : C \rightarrow A$ такой, что диаграмма

$$\begin{array}{ccccc} C & & \phi & & X \\ & \searrow j & & \searrow & \\ & A & \xrightarrow{f} & & X \\ & \psi \searrow & & \searrow g & \\ & & Y & & \end{array}$$

коммутативна.

Обратный образ (pullback) пары морфизмов $h : X \rightarrow B$, $k : Y \rightarrow B$ есть предел диаграммы $(h : X \rightarrow B, k : Y \rightarrow B)$

$$\begin{array}{ccc} & X & \\ & \downarrow k & \\ Y & \xrightarrow{h} & B \end{array}$$

Другими словами – это такие объект A и морфизмы $f : A \rightarrow X$, $g : A \rightarrow Y$ такие, что квадрат

$$\begin{array}{ccc}
 A & \xrightarrow{f} & X \\
 g \downarrow & & \downarrow f \\
 Y & \xrightarrow{h} & B
 \end{array}$$

является декартовым.

Говорим, что в данном случае f обратный образ h относительно k , а g обратный образ k относительно h . Обозначение: $A = X \times_B Y$.

1.3. Произведение объектов

Определение 2. Произведением объектов A и B называется предел диаграммы, которая состоит только из двух объектов A и B и не имеет ни одного морфизма. Для произведения используется обозначение $A \times B$.

Поскольку $A \times B$ – конус, то имеем диаграмму

$$\begin{array}{ccc}
 A & & B \\
 pr_A \swarrow & & \nearrow pr_B \\
 & A \times B &
 \end{array}$$

где морфизмы pr_A, pr_B называются проекциями.

1.4. Функторы

Пусть даны две категории \mathcal{K}, \mathcal{E} . Если смотреть на категории как на множества, то хотелось бы иметь аналог отображения одной категории в другую. Таким аналогом является понятие функтора.

Функтором F из категории \mathcal{K} в категорию \mathcal{E} называется функция, которая ставит в соответствие:

- 1) каждому объекту A категории \mathcal{K} объект $F(A)$ категории \mathcal{E} ;
- 2) каждому морфизму $f : A \rightarrow B$ категории \mathcal{K} морфизм³ $F(f) : F(A) \rightarrow F(B)$ категории \mathcal{E} такой, что
 - а) $F(1_A) = 1_{F(A)}$;
 - б) $F(g \circ f) = F(g) \circ F(f)$ для любых морфизмов f, g , для которых определена композиция $g \circ f$.

Тождественный функтор $1_{\mathcal{K}} : \mathcal{K} \rightarrow \mathcal{K}$ определяется правилами $1_{\mathcal{K}}(A) = A, 1_{\mathcal{K}}(f) = f$. Эти же правила задают функтор включения $\mathcal{K} \hookrightarrow \mathcal{E}$ в случае, когда \mathcal{K} – подкатегория в \mathcal{E} .

Категории \mathcal{K}, \mathcal{E} называются эквивалентными, если существуют функторы $F : \mathcal{K} \rightarrow \mathcal{E}, G : \mathcal{E} \rightarrow \mathcal{K}$, для которых $G \circ F = 1_{\mathcal{K}}, F \circ G = 1_{\mathcal{E}}$.

³Мы определяем ковариантный функтор; если направление морфизма меняется, т. е. $F(f) : F(B) \rightarrow F(A)$, то имеем контрвариантный функтор.

1.5. Категория функторов $\mathcal{E}^{\mathcal{K}}$

Пусть даны две категории \mathcal{K} , \mathcal{E} . Построим категорию функторов $\mathcal{E}^{\mathcal{K}}$, объектами которой являются функторы из \mathcal{K} в \mathcal{E} .

Определим морфизмы категории $\mathcal{E}^{\mathcal{K}}$. Возьмём два функтора $F : \mathcal{K} \rightarrow \mathcal{E}$ и $G : \mathcal{K} \rightarrow \mathcal{E}$.

Морфизм $\tau : F \rightarrow G$ объекта F в объект G называется *естественным преобразованием* функтора F в функтор G и состоит из семейства морфизмов $\{\tau_A : F(A) \rightarrow G(A), \text{ где } A \text{ — любой объект категории } \mathcal{K}\}$. Причём морфизм $\tau_A : F(A) \rightarrow G(A)$ таков, что для любого морфизма $f : A \rightarrow B$ диаграмма

$$\begin{array}{ccccc} A & & F(A) & \xrightarrow{\tau_A} & G(A) \\ f \downarrow & & F(f) \downarrow & & \downarrow G(f) \\ B & & F(B) & \xrightarrow{\tau_B} & G(B) \end{array}$$

коммутативна.

Морфизмы τ_A называются *компонентами* преобразования τ .

2. Теория топосов

2.1. Экспоненциал

Определение 3. Категория \mathcal{K} допускает *экспоненцирование*, если

- 1) в ней существует произведение любых двух объектов;
- 2) если для любых двух объектов A и B существует объект B^A («отображение» объекта A в объект B), называемый *экспоненциалом*, и морфизм $ev : B^A \times A \rightarrow B$, называемый *морфизмом значения*, такие, что для любых объекта C и морфизма $g : C \times A \rightarrow B$ существует единственный морфизм $\hat{g} : C \rightarrow B^A$, для которого коммутативна диаграмма

$$\begin{array}{ccc} B^A \times A & \xrightarrow{ev} & B \\ \hat{g} \times 1_A \uparrow & & \nearrow g \\ C \times A & & \end{array}$$

2.2. Классификатор

Мы считаем, что в реальном мире действует двузначная классическая логика с двумя истинностными значениями «истина» = 1 или «ложь» = 0. Однако ещё Аристотель знал о модальной логике, в которой отражён более гибкий многозначный взгляд на мир, при котором между истиной и ложью всегда находится место множеству промежуточных вариантов, выдаваемые под предлогом целесообразности либо за ложь, либо за истину. Впрочем, в связи с тем, что в прикладных исследованиях все больше используется нечёткая логика

(нечёткая математика), то вполне разумно вместо теоретико-множественного объекта истинности $\{0, 1\}$ ввести многозначный объект истинности Ω .

Определим объект $\mathbf{1}$ как такой объект топоса, что для любого объекта A существует морфизм $! : A \rightarrow \mathbf{1}$. Назовём $\mathbf{1}$ *конечным объектом*.

Объект $\mathbf{1}$ – это аналог *1-точечного* множества.

Определение 4. *Классификатор подобъектов* для категории \mathcal{K} называется объект Ω вместе с морфизмом $\top : \mathbf{1} \rightarrow \Omega$ такой, что выполняется следующая

Ω -**аксиома**. Для каждого мономорфизма $f : A \rightarrow D$ существует единственный морфизм $\chi_f : D \rightarrow \Omega$, для которого диаграмма

$$\begin{array}{ccc} A & \xrightarrow{f} & D \\ ! \downarrow & & \downarrow \chi_f \\ \mathbf{1} & \xrightarrow{\top} & \Omega \end{array}$$

является декартовым квадратом.

Здесь через $! : A \rightarrow \mathbf{1}$ обозначен единственный морфизм из A в $\mathbf{1}$.

Морфизм χ_f называется *характеристическим*, или *характером* морфизма f , а объект Ω – *классифицирующим объектом*. Морфизм \top носит название «истина».

Термин «классификатор подобъектов» для Ω связан с тем, что любой мономорфизм $f : A \rightarrow D$ носит название *подобъекта* объекта D . Поэтому часто говорят, что морфизм $\chi_f : D \rightarrow \Omega$ *классифицирует* подобъект $f : A \rightarrow D$.

В теории категорий подобъекты – это аналоги подмножеств $A \subset D$ в теории множеств.

Допущение неклассического классификатора Ω означает переход к интуиционистской логике, в которой не действует закон исключённого третьего.

3. Определение элементарного топоса

Определение топоса, приводимое ниже, принадлежит Ловеру. Оно является более общим, чем понятие топоса Гротендика. Топос Гротендика – это частный случай топоса Ловера, и он тесным образом связан с геометрией и топологией, чего нельзя сказать о всех топосах Ловера. С работами социологов в области социальной топологии следует связывать топос Гротендика, но и другие топосы, негротендиковские, также имеет смысл использовать при описании обществ.

Определение 5. Категория с произведениями, в которых каждая пара объектов имеет экспоненциал, а сама категория обладает классификатором, называется *элементарным топосом*.

Элементарные топосы – это категории, которые очень близки по своим свойствам к категории теории множеств **Sets**. Действительно, если для топоса, выполняется аксиома выбора, то он является булевым [13, с. 161].

4. Топос Гротендика

Изучим теперь топосы Гротендика, которые были построены им под влиянием идеи пучков Лерэ раньше [3], чем появились топосы Ловера.

4.1. Топология Гротендика

Топология Гротендика – это структура на категории, которая делает её объекты похожими на открытые множества топологического пространства

С каждым объектом A связываем некоторую совокупность $\{A_x : x \in X\}$, которая в какой-то мере «покрывает» объект A , и это покрытие отражает внутреннюю связь и степень близости каждого A_x к A . Математики в таком случае говорят о топологии в «обобщённом» пространстве, в категории \mathcal{C} , которая задаёт указанные связи и близость.

Аксиоматически топологию в категории ввёл Александр Гротендик, которая носит теперь его имя. Формально речь идёт о следующем определении.

Определение 6. Топологией Гротендика J на категории \mathcal{C} называется функция, сопоставляющая каждому объекту A совокупность $J(A)$ множеств морфизмов, оканчивающихся в A , такая, что

- (i) одноэлементное множество $\{1_A : A \rightarrow A\}$ принадлежит $J(A)$;
- (ii) если $\{f_x : A_x \rightarrow A \mid x \in X\} \in J(A)$ и для каждого $x \in X$

$$\{f_y^x : A_y^x \rightarrow A_x \mid y \in X_x\} \in J(A_x),$$

то

$$\{f_x \circ f_y^x : A_y^x \rightarrow A \mid x \in X \text{ и } y \in X_x\} \in J(A);$$

(ij) если $\{f_x : A_x \rightarrow A \mid x \in X\} \in J(A)$ и $g : B \rightarrow A$ – произвольный морфизм, то для каждого $x \in X$ существует обратный образ морфизма f_x относительно g

$$\begin{array}{ccc} B \times_A A_x & \longrightarrow & A_x \\ g_x \downarrow & & \downarrow f_x \\ B & \xrightarrow{g} & A \end{array}$$

и

$$\{g_x : B \times_A A_x \rightarrow B \mid x \in X\} \in J(B).$$

Определение 7. Категория \mathcal{C} с топологией J , т. е. пара (\mathcal{C}, J) называется *ситусом* (site), а элементы из J – *покрытиями*.

4.2. Топология Гротендика для пространства и классическая топология пространства

Топология Гротендика говорит, каким образом объекты ситуса \mathcal{C} могут быть «покрыты» посредством морфизмов другими объектами. В случае, когда I – топологическое пространство и рассматриваемые объекты – это открытые множества в I , то топология Гротендика говорит, каким образом открытые множества A покрываются в смысле операции включения $A_x \hookrightarrow A$ семейством открытых множеств $J(A) = \{A_x\}_{x \in X}$ таким, что

$$\bigcup_{x \in X} A_x = A.$$

Следовательно, «топология Гротендика» – это структура на категории \mathcal{C} , которая делает её объекты похожими на открытые множества топологического пространства I , образующие категорию $\mathcal{O}(I)$. Топологии Гротендика аксиоматизируют определение открытого покрытия.

Существует естественный способ сопоставить топологическому пространству I топологию Гротендика $(\mathcal{O}(I), J)$. Поэтому топологии Гротендика часто рассматривают как обобщение обычных топологий. Объекты категории $\mathcal{O}(I)$ – это открытые множества топологии пространства I , морфизмы между двумя объектами – это включения $V \hookrightarrow U$, а покрытиями $J(U)$ являются так называемые *решёта*.

Для большого класса топологических пространств действительно можно восстановить топологию пространства I по её топологии Гротендика, однако, уже для антидискретного пространства это не так»⁴.

4.3. Топос пучков

Гротендик ввёл в математику совершенно новое представление о пространстве. Для этого он воспользовался идеей Лерэ о пучках, с помощью которых возможно установление отношений между локальными и глобальными данными, относящимися к пространству. Пространство по Гротендику снабжается «наиболее очевидной структурой, которую ему можно приписать, так сказать, «на глазок» – именно, структурой, называемой «категорией» [2].

Рассмотрим некоторую категорию \mathcal{C} . С каждым объектом $A \in \mathcal{C}$ мы связываем *множество различных данных* $F(A)$, которые охватывают возможности данного объекта в характеристике категории \mathcal{C} в целом.

В случае, когда $\mathcal{C} = \mathcal{O}(I)$ и $A \in \mathcal{O}(I)$ – открытое множество, то топология, определяющая степень близости точек пространства (в пределах A), менее предпочтительна, чем наличие в пространстве способа измерения расстояния между точками (в пределах A), т. е. геометрия имеет, конечно, преимущество перед топологией. При этом множество $F(A)$ может состоять из разных геометрических характеристик, но, упрощая, Гротендик пишет о измерительном приборе, или об эталоне метра, посредством которого делается измерение.

⁴См. статью «Топология Гротендика» в Википедии.

В рамках стремления к наиболее общему подходу с объектом A связывают просто какое-то множество данных, т. е. берём $F(A) \in \mathbf{Sets}$, а природа этих данных подлежит уточнению в зависимости от решаемой с помощью теории пучков задачи.

Таким образом, определена категория,

$$\mathbf{Sets}^{C^{op}}, \quad (1)$$

объектами которой являются *функторы* $F : C^{op} \rightarrow \mathbf{Sets}$, называемые *предпучками* и сопоставляющие объектам A множество «эталонных метр, или измерительных приборов», т. е. $F(A)$. (Значок «ор» (opposite) говорит, что направление морфизма надо заменить на противоположное).

Обратим внимание, что если мы возьмём другой предпучок $G : C^{op} \rightarrow \mathbf{Sets}$, то для каждого этноса A имеем иное множество $G(A)$ «эталонных метр или измерительных приборов», чем $F(A)$.

Категория $\mathbf{Sets}^{C^{op}}$ является топосом и называется категорией предпучков. Её часто обозначают $St(C)$.

Предпучок $F : C^{op} \rightarrow \mathbf{Sets}$ разрозненным, беспорядочным образом сопоставляет объектам B их «эталонные метр, или измерительные приборы» $F(B)$.

Для наведения порядка воспользуемся топологией категории C . Рассмотрим покрытие $J(A)$. Ясно, что объекты из $\{A_x : x \in X\}$ также имеют свои «эталонные метр, или измерительные приборы» той же структуры, что и A , т. е. имеем множества $F(A_x)$. Очевидно, должна быть увязка «эталонных метр, или измерительных приборов» $F(A_x)$ в нечто единое целое, определяемое «центральными» «эталонами метр, или измерительными приборами» $F(A)$ покрываемого «центрального» объекта A .

Увязка «эталонных метр, или измерительных приборов» в рамках одного предпучка F и в пределах одного объекта и близкого к нему семейства объектов $J(A)$ в *единый пучок*, именуемая в теории категории *склеиванием* элементов $s_x \in F(A_x)$ в единый элемент $s \in F(A)$, осуществляется посредством следующего определения.

Определение 8. Предпучок F называется *пучком над ситусом* (C, J) , если он удовлетворяет следующему условию *склейки*:

СОМ. Пусть $\{f_x : A_x \rightarrow A \mid x \in X\} \in J(A)$ – произвольное покрытие объекта A и $\{s_x : x \in X\}$ – произвольное семейство элементов $s_x \in F(A_x)$, являющихся попарно совместимыми, т. е. для любых $x, x' \in X$ выполняется равенство $F_{x'x}(s_x) = F_{x'x}(s_{x'})$. Тогда существует единственный $s \in F(A)$, удовлетворяющий равенству $F_x(s) = s_x$ при любом $x \in X$.

Здесь использованы функции ограничения $F_{x'x} : F(A_x) \rightarrow F(A_x \times_A A_{x'})$ и $F_{x'x} : F(A_{x'}) \rightarrow F(A_x \times_A A_{x'})$, являющиеся F -образами двух новых морфизмов $f_x, f_{x'}$, получаемых при построении декартова квадрата.

$$\begin{array}{ccc}
 A_x \times_A A_{x'} & \longrightarrow & A_{x'} \\
 \downarrow & & \downarrow f_{x'} \\
 A_x & \xrightarrow{f_x} & A
 \end{array}$$

Обозначим также через F_x стрелку $F(f_x) : F(A) \rightarrow F(A_x)$.

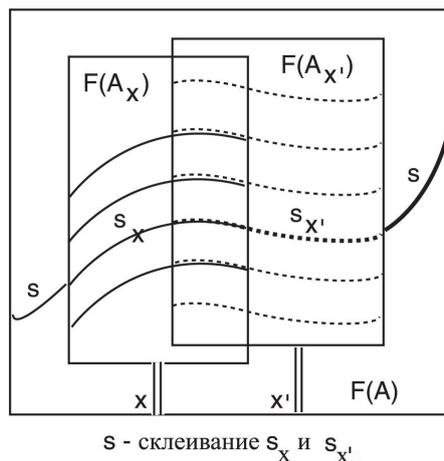


Рис. 1. Пучок – это предпучок со склеенным семейством «сечений» $\{s_x : x \in X\}$ в единое «сечение» s

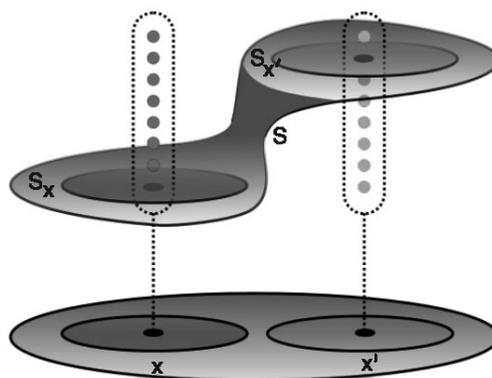


Рис. 2. Склеивание «сечений» s_x и $s_{x'}$ в пучке F позволяет объекту A_x непрерывно меняться вдоль «сечения» s при изменении $x \rightarrow x'$, переходя от A_x к $A_{x'}$ (рис. из [14])

Пучки – это предпучки специального вида, подобно тому, как абелевы группы являются специальным случаем групп. Пучки образуют полную подкатеорию в категории предпучков.

Определение 9. Подкатеорию категории предпучков $St(\mathcal{C})$, образованную пучками над ситусом (\mathcal{C}, J) , обозначаем через $Sh_J(\mathcal{C})$ (или $Sh(\mathcal{C}, J)$) и называем *категорией пучков*, или *топосом Гротендика*.

$$Sh_J(\mathcal{C}) \subset \mathbf{Sets}^{cop}.$$

Топос Гротендика, выражаясь его языком, – это арсенал всевозможных «эталонных метра, или измерительных приборов», соответствующим образом склеенных, т. е. согласованных.

5. Топосы и пространства

Топос Гротендика – это «пространство в новом стиле», обобщающее традиционное топологическое пространство. «То есть «пространство в новом стиле» (или топос), обобщающее традиционные топологические пространства, будет описываться попросту как «категория», которая, не вытекая с необходимостью из обыкновенного пространства, тем не менее обладает всеми хорошими свойствами (единожды чётко для всех определёнными, разумеется) этой «категории пучков» [2].

Благодаря топосам пучков, придуманных Гротендиком, «непрерывное», воплощённое в образе пространства, предстаёт «переданным» или «выраженным» структурой категории, по природе «алгебраической», воспринимавшейся до сих пор как существенно «разрывная» или «дискретная».

На самом деле самыми важными в топологическом пространстве являются отнюдь не его «точки» и не его «подмножества» с отношениями близости между ними, а пучки над этим пространством и категория, которую они образуют [2].

Понятие топоса Гротендика – расширение или, лучше сказать, метаморфоза понятия пространства.

Оно, во-первых, позволяет более или менее очевидным образом переносить на топосы самые важные интуитивные представления и геометрические конструкции, знакомые по старым добрым пространствам прежних времён, а во-вторых, новое понятие достаточно широко, чтобы охватить все множество ситуаций, в которых, как раньше считалось, не место интуитивным представлениям «тополого-геометрической» природы – именно, тем, какие тогда связывались только с обыкновенными топологическими пространствами [2]. Последнее и даёт шанс обеспечить «топологический поворот» в социологии «дискретным» образом развивая социальную топологию.

6. Топос в социологии

В российской социологической и философской литературе в последнее десятилетие распространился термин «топос». Как и в случае с понятием топологии, строгих определений топологии, топологического пространства, топоса в работах социологов⁵ и философов мы не находим. Поэтому можно соотнести

⁵Отметим достаточно строгое определение социальной топологии в работе Н.А. Шматко [7, с. 16].

имеющиеся в литературе расплывчатые определения с точными математическими определениями.

В социологии термин «топос» появился в работе Бурдьё, и изначально он вводился как греческий перевод слова «место».

Так Бурдьё пишет: «Место, *topos* может быть определено абсолютно, как то, где находится агент или предмет, где он «имеет место», существует, короче, как «локализация», или же относительно, релятивно, как позиция, как ранг в порядке. Занимаемое место может быть определено как площадь, поверхность и объем, который занимает агент или предмет, его размеры или, еще лучше, его габариты (как иногда говорят о машине или о мебели) [1].

В российской социологической литературе находим следующие определения топоса:

Топосы – локальные социальные порядки, поддерживаемые и воспроизводимые при участии публичной политики (Н.А. Шматко [7]).

Топосы – структурные единицы публичной политики, обуславливающие практики входящих в них агентов (Н.А. Шматко [7]).

Топосы – структурные единицы социальной коммуникации, обуславливающие практики, входящие в состав местных сообществ (Т.И. Макогон [9]).

Поскольку «локальные социальные порядки», местные сообщества можно локализовать посредством открытого множества A в топологическом социальном пространстве I , то скорее всего под «единицами социальной коммуникации, обуславливающими практики, входящими в состав местных сообществ» следует понимать пучок $F : \mathcal{O}(I) \rightarrow \mathbf{Sets}$, поскольку в данном случае для местного сообщества, которое *локально имеет место* в $A \subset I$, обусловленные им практики – это совокупность элементов, образующих множество $F(A)$. В пользу этого говорит то, что сами математики пучки используют для установления отношений между локальными и глобальными данными. Пучки содержат данные, отнесённые к каждому открытому подмножеству A социального пространства I .

Как пучки связывают локальное с глобальным, т. е. с тем, что характеризует все пространство I в целом? Ответ прост: например, пучкам (коммутативных групп) отвечают коммутативные группы, называемые *группами когомологий со значениями в пучке*. А группы когомологий определяют важнейшие фундаментальные топологические и геометрические свойства пространства (многообразия) I .

Иначе говоря, определение Т.И. Макогон следует переформулировать, заменив слово «топосы» на слово «пучки»:

Пучки – структурные единицы социальной коммуникации, обуславливающие практики, входящие в состав местных сообществ.

Топос – совокупность пучков, т. е. всевозможные варианты единиц социальной коммуникации.

7. Заключение

Социальная топология для многих социологических и философских исследований становится основной методологической установкой. Этот существенный «топологический поворот» в методологии требует соотнесения созвучных терминов, используемых математиками и социологами. И если с топологическими терминами ситуация более или менее удовлетворительная, то с термином «топос» наблюдается значительное расхождение. Конечно, каждая наука, используя одни и те же слова, часто придаёт им различный смысл. Тем не менее желательно иметь «переводчики», которые устанавливают требуемые соответствия между используемыми терминами. В данной статье показано, что термину «топос» в социологии в математике соответствует термин «пучок». Как результат, мы получаем возможность правильно использовать одну из самых блестящих математических теорий – теорию пучков и топосы Гротендика, в частности, в целях решения задач, стоящих перед теоретической социологией и философией, сформулированных в рамках методологии социальной топологии.

ЛИТЕРАТУРА

1. Бурдьё П. Социология политики: Пер. с фр. / Сост., общ. ред. и предисл. Н.А. Шматко. М. : Socio-Logos, 1993.
2. Гротендик А. Урожаи и посевы. Ижевск : Издательский дом: НИЦ «Регулярная и хаотическая динамика», 2002.
3. Artin M., Grothendieck A., Verdier J.L. Theorie des topos et cohomologie etale des schemas. Vols. 1, 2, and 3 // Lecture Notes in Mathematics. Springer-Verlag. V. 269. 1972; V. 270. 1972 and V. 305. 1973.
4. Латур Б. Пересборка социального. Введение в акторно-сетевую теорию. М. : ГУ-ВШЭ, 2014.
5. Ло Дж. Объекты и пространства / Пер. с англ. В. Вахштайна // Социологическое обозрение. 2006. № 1.
6. Вахштайн В. Социология вещей и «поворот к материальному» в социальной теории // Социология вещей. Сборник статей. М. : Издательский дом «Территория будущего», 2006.
7. Шматко Н.А. Плюрализация социального порядка и социальная топология // Социологические исследования. 2001. № 9. С. 14–18.
8. Шматко Н.А. Феномен публичной политики // Социол. исслед. 2001. № 7.
9. Макогон Т.И. Топологические перспективы описания местных сообществ // Теория и практика общественного развития. 2013. Вып. 1. С. 24–30.
10. Макогон Т.И. Процессы глобализации и собственная логика местных сообществ // Вестник КемГУ. 2013. Т. 1, № 3 (55). С. 226–232.
11. Бетев А.Н. Социальная топология свободы: диссертация ... канд. филос. наук. Екатеринбург, 2009. 167 с.
12. Гольдблатт Р. Топосы. М. : Мир, 1983.
13. Джонсон П.Т. Теория топосов. М. : Наука, 1986.

14. Sallach D.L. Recognition-Based Logic and Social Conflict: Toward a Topos Model // The 8th International Workshop on Agent-based Approach in Economic and Social Complex Systems. Tokyo: Shibaura Institute of Technology, 2013.

SOCIAL TOPOLOGY AND THE GROTHENDIECK TOPOI

A.K. Guts¹

Dr.Sc.(Phys.-Math.), Professor, e-mail: guts@omsu.ru

L.A. Pautova²

D.Sc.(Sociology), e-mail: pautoval@yandex.ru

¹Omsk State University n.a. F.M. Dostoevskiy

²Moscow, Fund of Public Opinion

Abstract. The methodology of social topology and most the common approach to the topological vision of space that is offered by Grothendieck are discussed. The Grothendieck topology and the Grothendieck topoi are considered. It is shown how they promote topological turn in the modern sociology.

Keywords: social topology, sociology, topoi, the Grothendieck topos, sheaves topoi, formalization of social space.

МЕТОД ЛОКАЛЬНЫХ КООРДИНАТ ДЛЯ ВЫЧИСЛЕНИЯ ТРАЕКТОРИИ ГИБРИДНОЙ ДИНАМИЧЕСКОЙ СИСТЕМЫ ПРИ СКОЛЬЖЕНИИ

В.В. Коробицын

к.ф.-м.н., доцент, e-mail: korobits@rambler.ru

Ю.В. Фролова

к.ф.-м.н., доцент, e-mail: juliavfrolova@rambler.ru

Омский государственный университет им. Ф.М. Достоевского

Аннотация. Предлагается метод вычисления траектории решения гибридной динамической системы, основанный на переходе к системе меньшей размерности, и вычисление траектории в локальных координатах поверхности скольжения. Метод позволяет избежать проблемы дрожания траектории решения при движении вдоль поверхности скольжения.

Ключевые слова: гибридная динамическая система, режим скольжения, система обыкновенных дифференциальных уравнений с разрывной правой частью, численный метод.

Введение

Современные подходы в области моделирования процессов ориентированы на использование составных систем, где область определения системы разбивается на подобласти. В каждой отдельной подобласти система имеет непрерывное поле направлений, но на границе этих областей непрерывность нарушается. Такие системы получили различные названия: сшитые, склеенные, гибридные, кусочно-гладкие, системы с разрывной правой частью [1, 12, 20, 21]. Существует большое количество гибридных систем, описывающих модели процессов в различных прикладных областях: механике, электротехнике, управлении, биологии [13, 14, 16–18].

Исследование гибридных систем, как и непрерывных динамических систем, ведётся двумя способами: аналитическое исследование поведения решений и численное исследование решений. При численном решении гибридных систем возникает много нерешённых вопросов. Некоторые системы могут иметь множество решений, некоторые могут не иметь решений, многие имеют не гладкие решения. Отметим, что гибридные системы имеют так называемый «скользящий режим», при котором траектория решения системы движется не вдоль векторного поля направлений, а вдоль поверхности раздела областей непрерывности системы. На этих поверхностях системы обычно доопределяют в смысле эквивалентного управления [11].

До сих пор не разработан алгоритм численного решения гибридных динамических систем в общем виде. Имеется множество работ, посвящённых отдельным задачам, связанным с построением такого алгоритма. Это задачи поиска пересечения траектории решения поверхности разрыва векторного поля направлений [2–4, 8], точному вычислению скользящего режима [5, 19], разработки концепции общего алгоритма [7, 9, 18]. Данная статья посвящена одной из этих задач: вычисление точного решения при скольжении по поверхности разрыва.

1. Определение склеенной динамической системы

Будем рассматривать динамическую систему, которая получается путём склеивания двух гладких систем по некоторой гладкой поверхности. Размерность системы равна 3, размерность поверхности — 2. Полученные ниже результаты можно обобщить на случай бóльших размерностей.

Пусть S — гладкая регулярная поверхность, заданная в параметрическом виде $x = r(u, v)$, $x \in \mathbb{R}^3$, $(u, v) \in \mathbb{R}^2$. Поверхность S делит пространство \mathbb{R}^3 на две непересекающиеся области G_1 и G_2 , расположенные по разные стороны от поверхности S . Поведение динамической системы в областях G_1 и G_2 будет описываться системой обыкновенных дифференциальных уравнений

$$\frac{dx}{dt} = \begin{cases} f_1(x) & \text{при } x \in G_1, \\ f_2(x) & \text{при } x \in G_2. \end{cases} \quad (1)$$

Векторные поля $f_1(x)$ и $f_2(x)$ непрерывны в G_1 и G_2 соответственно, где $\Gamma_1 = G_1 \cup S$, $\Gamma_2 = G_2 \cup S$. Траектория решения системы (1) внутри каждой из областей G_1 и G_2 будет описываться отдельным уравнением из (1) с непрерывным полем направления. И решение будет определяться традиционным способом, как дифференцируемая кривая, касательный вектор к которой в каждой точке совпадает с векторным полем правой части уравнения. Однако если траектория выходит на поверхность S , то поведение системы не может быть описано системой (1), поскольку в точке пересечения (касания) кривой решения поверхности S не определена правая часть системы. Её необходимо доопределить на поверхности S .

Доопределив систему в виде

$$\frac{dx}{dt} = \begin{cases} f_1(x) & \text{при } x \in G_1, \\ \frac{1}{2}(f_1(x) + f_2(x)) & \text{при } x \in S, \\ f_2(x) & \text{при } x \in G_2, \end{cases} \quad (2)$$

можно описать поведение системы при пересечении траектории решения поверхности S . То есть если траектория переходит из G_1 в G_2 или обратно. При этом в точке пересечения x_0 поверхности S траектория будет непрерывной, но может быть недифференцируемой, поскольку в этой точке векторное поле

скачком меняет своё значение с $f_1(x_0)$ на $f_2(x_0)$. При этом заданное на S векторное поле $\frac{1}{2}(f_1(x_0) + f_2(x_0))$ не имеет особого смысла, поскольку это значение задаёт направления лишь в одной точке, правая часть при этом все равно остаётся разрывной. Тем не менее, решение системы (2) будет представлено непрерывной кривой, составленной из гладких кривых в G_1 и G_2 , склеенных на поверхности S . Условие пересечения траектории решения поверхности S определяется направлениями векторов $f_1(x_0)$ и $f_2(x_0)$. Если оба вектора направлены по одну сторону от поверхности S , то происходит пересечение поверхности в направлении той стороны, куда указывают вектора.

В случае, когда векторы $f_1(x_0)$ и $f_2(x_0)$ направлены по разные стороны от S , траектория не может сойти с поверхности и возникает так называемый скользящий режим. Траектория решения движется по поверхности S , не имея возможности с неё уйти. Обозначим область $B \subseteq S$, на которой реализуется скользящий режим. В этом случае доопределение системы (2) не отвечает физическому смыслу, поскольку вектор $\frac{1}{2}(f_1(x) + f_2(x))$ не будет совпадать с касательной к траектории решения, если он не будет лежать в касательной плоскости K к поверхности S . Поэтому необходимо доопределить систему следующим образом

$$\frac{dx}{dt} = \begin{cases} f_1(x) & \text{при } x \in G_1, \\ \frac{1}{2}(f_1(x) + f_2(x)) & \text{при } x \in S \setminus B, \\ \frac{1}{2}\text{Pr}_K(f_1(x) + f_2(x)) & \text{при } x \in B, \\ f_2(x) & \text{при } x \in G_2, \end{cases} \quad (3)$$

здесь $\text{Pr}_K f(x)$ обозначает результат операции проецирования вектора $f(x)$ на касательную плоскость K . Решение системы (3) представляет собой непрерывную кривую, склеенную из гладких кривых, лежащих в G_1 , G_2 и на поверхности S в области B в случае скольжения.

2. Разложение векторов в локальном базисе

Поскольку поверхность S — гладкая регулярная, то в каждой точке (u, v) существуют векторы $r'_u := \frac{\partial r}{\partial u}(u, v)$ и $r'_v := \frac{\partial r}{\partial v}(u, v)$, они не равны нулю и не параллельны. Поэтому единичный вектор нормали n к касательной плоскости K можно вычислить как нормированное векторное произведение векторов r'_u, r'_v

$$n = \frac{[r'_u, r'_v]}{|[r'_u, r'_v]|}.$$

Векторы r'_u, r'_v, n образуют базис в \mathbb{R}^3 . Поэтому в любой точке поверхности $r(u, v)$ вектор f векторного поля $f_i(r(u, v)), i = 1, 2$ можно разложить в базисе $\{r'_u, r'_v, n\}$: $f = \alpha r'_u + \beta r'_v + \gamma n$. Вектор $h = (\alpha, \beta, \gamma)^T$ можно найти из системы

линейных алгебраических уравнений

$$Ah = f, \text{ где } A = \begin{pmatrix} r'_u{}^1 & r'_v{}^1 & n^1 \\ r'_u{}^2 & r'_v{}^2 & n^2 \\ r'_u{}^3 & r'_v{}^3 & n^3 \end{pmatrix}.$$

Поскольку векторы r'_u, r'_v, n не компланарны, то матрица A обратима и $h = A^{-1}f$. Значение координаты γ вектора h можно вычислить как скалярное произведение вектора Ah и n : $\gamma = \langle Ah, n \rangle = \langle f, n \rangle$. Знак величины γ для векторов f_1 и f_2 будет показывать направление этих векторов относительно касательной плоскости K . Условия пересечения и скольжения траектории решения на поверхности S можно записать, определяя знак произведения значений γ .

В области B векторные поля f_1 и f_2 направлены в противоположные стороны от касательной плоскости K . Условие скольжения можно сформулировать так

$$\begin{aligned} & (\langle f_1, n \rangle \cdot \langle f_2, n \rangle)(u, v) < 0 \text{ или} \\ & \text{одновременно } \langle f_1, n \rangle(u, v) = 0 \text{ и } \langle f_2, n \rangle(u, v) = 0 \end{aligned} \quad (4)$$

для любых $(u, v) \in B$.

Здесь значения векторов f_1 и f_2 определяются в точке (u, v) как значения векторных полей на поверхности S , соответствующих локальным координатам u, v : $f_i := f_i(r(u, v))$. Условие пересечения траектории решения системы (2) поверхности S вне области B выглядит так

$$\begin{aligned} & (\langle f_1, n \rangle \cdot \langle f_2, n \rangle)(u, v) > 0 \text{ или} \\ & \text{выполнено одно из условий } \langle f_1, n \rangle(u, v) = 0 \text{ или } \langle f_2, n \rangle(u, v) = 0 \end{aligned}$$

для любых $(u, v) \in S \setminus B$.

Проекцию вектора $f_i(x)$ на касательную плоскость K в любой точке $x = r(u, v)$ можно определить так

$$\text{Pr}_K(f_i(x)) = f_i(r(u, v)) - (\langle n, f_i \rangle \cdot n)(u, v).$$

3. Система в локальных координатах

В каждой точке $x = r(u, v)$ поверхности S можно ввести локальную систему координат (u, v, w) в базисе $\{r'_u, r'_v, n\}$.

Пусть $g(t) = (u(t), v(t))$ вектор-функция в \mathbb{R}^2 , описывающая движение точки вдоль кривой, лежащей на поверхности S . Тогда координаты точек этой кривой в объемлющем пространстве \mathbb{R}^3 будут $x(t) = r(u(t), v(t))$. Вычислим полную производную функции $x(t)$ в некоторой точке t

$$\frac{dx}{dt}(t) = r'_u \frac{du}{dt}(t) + r'_v \frac{dv}{dt}(t).$$

А это в свою очередь является разложением вектора $\frac{dx}{dt}$ по базису $\{r'_u, r'_v, n\}$. При этом третья координата равна нулю, поскольку кривая лежит на поверхности S . $\frac{dx}{dt}(t)$ — это вектор $(\frac{du}{dt}(t), \frac{dv}{dt}(t), 0)^T$ в базисе $\{r'_u, r'_v, n\}$. Можно записать

$$\begin{pmatrix} \frac{du}{dt} \\ \frac{dv}{dt} \\ 0 \end{pmatrix} = A^{-1} \frac{dx}{dt}(t).$$

Поскольку на поверхности S в области B кривая решения системы (3) должна удовлетворять третьему уравнению, то получаем

$$\begin{pmatrix} \frac{du}{dt} \\ \frac{dv}{dt} \\ 0 \end{pmatrix} = A^{-1} \left(\frac{f_1 + f_2}{2} - \left\langle n, \frac{f_1 + f_2}{2} \right\rangle \cdot n \right) (u(t), v(t)). \quad (5)$$

Для любого вектора f вектор $f - \langle n, f \rangle \cdot n$ будет ортогонален вектору n , поэтому третья координата вектора правой части системы (5) также будет равна нулю.

Заключение

Представленный метод нахождения траектории решения гибридной системы в режиме скольжения позволяет избежать проблемы, связанной с дрожанием траектории решения вдоль поверхности скольжения. Переход к решению задачи меньшей размерности, совпадающей с размерностью поверхности скольжения, даёт гарантию отсутствия схода траектории с поверхности скольжения раньше, чем это потребуется при выходе со скольжения.

Возникающая проблема с обращением матрицы A в системе (5) при вычислениях правой части на каждом шаге численного метода может быть решена вычислением обратной матрицы в аналитическом виде. Поскольку поверхность обычно описывается простой функцией от точки r , то нахождение частных производных от этой функции не составит особого труда. Однако если придётся вычислять эту матрицу численно, то это может привести к нежелательным эффектам при вычислении траектории.

ЛИТЕРАТУРА

1. Баутин Н.Н., Леонтович Е.А. Методы и приемы качественного исследования динамических систем на плоскости. М. : Наука, 1990. 488 с.
2. Коробицын В.В., Идрисов Р.Г. Линейное приближение точки пересечения кривой решения системы ОДУ с поверхностью разрыва правой части // Математические структуры и моделирование. 2010. Вып. 21. С. 17–23.
3. Коробицын В.В., Маренич В.Б., Фролова Ю.В. Численный алгоритм склеивания решения на поверхности разрыва для систем с клеточной структурой // Вестник Омского университета. 2010. № 4. С. 156–163.

4. Коробицын В.В. Об определении точки пересечения кривой решения задачи Коши с поверхностью разрыва // Вычисл. технологии. 2011. Т. 16, № 4. С. 50–63.
5. Коробицын В.В., Фролова Ю.В. Алгоритм вычисления скользящего режима для системы с гладкой границей разрыва // Вычисл. технологии. 2010. Т. 15, № 2. С. 56–72.
6. Коробицын В.В., Фролова Ю.В. Выбор режима продолжения траектории динамической системы при пересечении двух поверхностей разрыва // Математические структуры и моделирование. 2011. Вып. 24. С. 33–39.
7. Коробицын В.В., Фролова Ю.В., Маренич В.Б. Алгоритм численного решения кусочно-сшитых систем // Вычисл. технологии. 2008. Т. 13, № 2. С. 70–81.
8. Коробицын В.В., Фролова Ю.В. Оценка погрешности вычисления точки пересечения продолжения кривой решения задачи Коши с поверхностью разрыва // Математические структуры и моделирование. 2011. Вып. 22. С. 5–14.
9. Коробицын В.В., Фролова Ю.В. Представление алгоритма численного решения гибридной динамической системы с конечным множеством дискретных состояний в виде конечного автомата // Вестник Омского университета. 2013. № 4(70). С. 221–227.
10. Коробицын В.В., Фролова Ю.В. Программа МЕР2 для исследования динамических систем с клеточной структурой // Вестник Омского университета. 2011. № 2. С. 167–172.
11. Уткин В.И. Скользящие режимы в задачах оптимизации и управления. М. : Наука, 1981. 368 с.
12. Филиппов А.Ф. Дифференциальные уравнения с разрывной правой частью. М. : Наука, 1985. 224 с.
13. Brogliato B. *Nonsmooth Mechanics* (2nd edn). London: Springer-Verlag, 1999.
14. Gouze J.-L., Sari. T. A class of piecewise linear differential equations arising in biological models // *Dynamical Systems*. 2002. Vol. 17. P. 299–319.
15. Korobitsin V., Frolova J. Numerical algorithm for solving discontinuous dynamical systems: Automata approach // *Proceedings of the BIT 50 Conference: Trends in Numerical Mathematics*. June 17-20, 2010, Lund, Sweden. P. 66.
16. Lygeros J., Tomlin C., Sastry S. Controllers for reachability specifications for hybrid systems // *Automatica*. 1999. Vol. 35, № 3. P. 349–370.
17. Melin J., Hultgren A. A limit cycle of a resonant converter // In Engell H.G.S., Zaytoon J. *Analysis and design of hybrid systems 2003*. IFAC, Elsevier, 2003. P. 169–174.
18. Navarro-López E.M., Carter R. Hybrid automata: an insight into the discrete abstraction of discontinuous systems // *International Journal of Systems Science*. 2011. V. 42, № 11. P. 1883–1898.
19. Piironen P.T., Kuznetsov Y.A. An event-driven method to simulate Filippov systems with accurate computing of sliding motions // *ACM Trans. Math. Softw.* 2008. V. 34, № 3. Article 13. 24 p.
20. Witsenhausen H.S. A class of hybrid-state continuous-time dynamic system // *IEEE Transactions on Automatic Control*. 1966. Vol. AC-11, № 2. P. 161–167.
21. Zhusubaliyev Z., Mosekilde E. *Bifurcations and chaos in piecewise-smooth dynamical systems*. Singapore: World Scientific, 2003.

**METHOD OF LOCAL COORDINATES FOR COMPUTING THE SLIDING
TRAJECTORY OF HYBRID DYNAMICAL SYSTEM**

V.V. Korobitsyn¹

Ph.D.(Phys.-Math.), Associate Professor, e-mail: korobits@rambler.ru

J.V. Frolova¹

Ph.D.(Phys.-Math.), Associate Professor, e-mail: juliavfrolova@rambler.ru

¹Omsk State University n.a. F.M. Dostoevskiy

Abstract. The method for computing the sliding trajectory of hybrid dynamical systems is suggested. The method is based on transition to the local coordinate system on sliding surface with smaller dimension. The method solves the problem of chattering the trajectory near the sliding surface.

Keywords: hybrid dynamical system, sliding mode, system of ordinary differential equations with discontinuous right-hand side, numerical method.

МНОГОМЕРНОЕ ШКАЛИРОВАНИЕ НА БАЗЕ МЕТОДА ВЕРЛЕ

В.А. Шовин

научный сотрудник, e-mail: v.shovin@mail.ru

Омский филиал Федерального государственного бюджетного учреждения науки
Института математики им. С.Л. Соболева Сибирского отделения РАН

Аннотация. Предложен новый метод факторизации посредством многомерного шкалирования на базе метода Верле. Результаты факторизации корреляционной матрицы с помощью данного метода находятся в соответствии с классическими методами (метод главных компонент, метод минимальных остатков).

Ключевые слова: метод Верле, многомерное шкалирование, факторный анализ.

Введение

Многомерное шкалирование позволяет в рамках гипотезы о размерности целевого пространства расположить объекты по их взаимным расстояниям таким образом, чтобы восстанавливаемые расстояния между объектами приближались к эмпирическим.

Метод Верле — это итерационный метод вычисления следующего местоположения точки по текущему и прошлому. Этот метод позволяет учитывать дополнительные ограничения, накладываемые на системы точек, например, расстояния между точками. На базе метода Верле предлагается осуществить многомерное шкалирование, тем самым взаимные расстояния между точками будут учтены с наибольшей точностью.

В качестве матрицы взаимных расстояний будет выступать матрица корреляций. С помощью многомерного шкалирования будет осуществлена факторизация корреляционной матрицы, тем самым будет восстановлена факторная структура данных в факторном пространстве. Чтобы получить интерпретабельное решение, предлагается использовать отдельные методы факторного вращения, применённые к восстановленной факторной структуре.

1. Метод Верле

Алгоритм Верле используется для вычисления следующего положения точки по текущему и прошлому:

$$\bar{x}_j^i = \bar{x}_j^{i-1} + \bar{v}_j,$$

$\bar{x}_j^i = (x_{j1}^i, x_{j2}^i, \dots, x_{jm}^i)$ — вычисляемые координаты j -ой точки на i -ой итерации,

m — размерность пространства,

$\bar{v}_j = \bar{x}_j^{i-1} - \bar{x}_j^{i-2}$ — вектор скорости j -ой точки.

На систему точек накладываются ограничения.

Некоторые из точек связаны упругими стержнями заданной длины.

Алгоритм работает следующим образом:

1. Вычисляются новые положения точек.
2. Для каждой связи удовлетворяется соответствующее условие.
3. Шаг 2 повторяется s раз.

Например, $s = 16$.

Процедура релаксации связи описывается следующими формулами:

Если связь представлена точками \bar{a} и \bar{b} с равновесным расстоянием между ними t , то

$$\bar{a}^i = \bar{a}^{i-1} + \bar{r},$$

$$\bar{b}^i = \bar{b}^{i-1} - \bar{r},$$

$$\bar{r} = f \cdot q \cdot \frac{t - |\bar{a}^{i-1} - \bar{b}^{i-1}|}{|\bar{a}^{i-1} - \bar{b}^{i-1}|} (\bar{a}^{i-1} - \bar{b}^{i-1}),$$

$f = 0.7$ — коэффициент упругости связи,

$q = \frac{1}{s}$ — коэффициент, зависящий от числа s повторений шага 2.

2. Многомерное шкалирование

Многомерное шкалирование (МНШ) — это способ наиболее эффективного размещения объектов, приближённо сохраняющий наблюдаемые между ними расстояния. МНШ размещает объекты в пространстве заданной размерности и проверяет, насколько точно полученная конфигурация сохраняет расстояния между объектами. МНШ использует алгоритм минимизации некоторой функции, оценивающей качество получаемых вариантов отображения.

Мерой, наиболее часто используемой для оценки качества подгонки модели (отображения), измеряемого по степени воспроизведения исходной матрицы сходств, является так называемый стресс. Величина стресса φ для текущей конфигурации определяется так:

$$\varphi = \sum_{i=1, j=i+1}^m (d_{ij} - f(\delta_{ij}))^2.$$

Здесь d_{ij} — воспроизведённые расстояния в пространстве заданной размерности, а δ_{ij} — исходное расстояние, m — количество объектов. Функция $f(\delta_{ij})$ обозначает неметрическое монотонное преобразование исходных данных (расстояний). МНШ воспроизводит не количественные меры сходств объектов, а лишь их относительный порядок. Чем меньше значение стресса, тем лучше матрица исходных расстояний согласуется с матрицей результирующих расстояний.

3. Главные компоненты и факторная модель

Модель главных компонент описывается следующими формулами

$$\vec{z}_i = a_{i1}\vec{p}_1 + a_{i2}\vec{p}_2 + \dots + a_{ig}\vec{p}_g + d_i\vec{u}_i$$

m — число переменных,

g — число факторов,

\vec{z}_i — исходные переменные,

\vec{p}_i — общие факторы,

\vec{u}_i — специфичные факторы.

Корреляции между исходными переменными могут быть определены как скалярное произведение нормализованных последних с нулевым средним и единичной дисперсией:

$$r_{ij} = \frac{1}{n+1} \vec{z}_i \cdot \vec{z}_j,$$

n — размерность исходного пространства переменных.

Коэффициенты корреляций между исходными переменными определяют матрицу корреляций, сходную с матрицей взаимных расстояний метода многомерного шкалирования. Поскольку ближним объектам в факторном пространстве соответствуют большие значения коэффициентов корреляций, то элементы матрицы взаимных расстояний d_{ij} получаются из соответствующих элементов матрицы корреляций r_{ij} по формуле:

$$d_{ij} = 1 - |r_{ij}|.$$

С помощью метода Верле будет восстановлена факторная структура в рамках гипотезы о размерности g факторного пространства.

Факторная структура

Элементы факторной структуры a_{ij} могут быть определены как коэффициент корреляции между j -ой факторной осью и i -ой переменной:

$$a_{ij} = \frac{\bar{s}_i \cdot \bar{f}_j}{|\bar{s}_i| \cdot |\bar{f}_j|},$$

$$\bar{f}_i = (\delta_{i1}, \dots, \delta_{ig}),$$

$$\delta_{ij} = \begin{cases} 1, i = j, \\ 0, i \neq j \end{cases} \quad \text{— символ Кронекера,}$$

$\bar{s}_j = \bar{h}_j - \bar{h}_c$ — вектор направления j -ой переменной в факторном пространстве.

\bar{h}_j — j -ая переменная в факторном пространстве.

\bar{h}_c — центр масс факторной структуры переменных в факторном пространстве.

4. Численный эксперимент

В качестве исходных параметров были взяты 15 биофизических показателей для 131 лица с артериальной гипертензией начальной стадии:

- 1) *вес*,
- 2) *индекс массы тела (ИМТ)*,
- 3) *частота дыхания (ЧД)*,
- 4) *сегментоядерные нейтрофилы (С)*,
- 5) *лимфоциты (Л)*,
- 6) *конечно-систолический размер левого желудочка (КСР)*,
- 7) *конечно-систолический объем левого желудочка (КСО)*,
- 8) *конечно-диастолический размер левого желудочка (КДР)*,
- 9) *конечно-диастолический объем левого желудочка (КДО)*,
- 10) *ударный объем (УО)*,
- 11) *минутный объем сердца (МОС)*,
- 12) *общее периферическое сосудистое сопротивление (ОПСС)*,
- 13) *индекс Хильдебрандта (ИХ)*,
- 14) *фракция выброса левого желудочка (ФВ)*,
- 15) *фракция укорочения левого желудочка (ФУ)*.

Программная реализация

Метод Верле был реализован программно с использованием общедоступной JavaScript библиотеки Verlet.js, которая была усовершенствована для многомерного случая. Web-приложение многомерного шкалирования на базе метода Верле доступно по адресу: <http://svlaboratory.org/application/multscal> — после регистрации нового пользователя. Приложение позволяет визуализировать процесс сходимости метода Верле в заданной плоскости координат (рис. 1).

Результирующая факторная структура для данных артериальной гипертензии представлена в таблице 1.

Факторное решение после факторного вращения по критерию интерпретативности, предложенное в работе [1], представлено в таблице 2. Данные факторные структуры подтверждаются предыдущими работами [2].

5. Заключение

Метод многомерного шкалирования на базе метода Верле, примененный к корреляционной матрице, является альтернативным методом факторизации.

Таблица 1. Исходное факторное решение (метод Верле)

	<i>F1</i>	<i>F2</i>	<i>F3</i>	<i>F4</i>	<i>F5</i>
Вес	-0,469	0,442	0,227	-0,667	0,296
ИМТ	-0,336	0,172	0,421	-0,748	0,348
ЧД	-0,258	0,239	0,479	0,68	0,429
С	-0,854	-0,444	0,034	0,032	-0,265
Л	-0,766	-0,584	0,093	0,005	-0,254
КСР	0,914	0,151	-0,27	-0,155	-0,214
КСО	0,931	0,11	-0,251	-0,176	-0,165
КДР	0,604	0,096	-0,783	-0,105	-0,044
КДО	0,667	0,109	-0,727	-0,121	-0,011
УО	0,334	0,024	-0,931	-0,099	0,104
МОС	0,376	0,003	-0,908	0,062	0,175
ОПСС	0,156	0,248	-0,855	0,4	-0,155
ИХ	-0,189	-0,067	0,492	0,592	0,605
ФВ	0,576	-0,123	0,687	-0,211	-0,37
ФУ	0,266	0,085	0,679	0,075	-0,674

Таблица 2. Факторная структура по критерию интерпретабельности (косоугольный случай)

	<i>F1</i>	<i>F2</i>	<i>F3</i>	<i>F4</i>	<i>F5</i>
Вес	-0,274	-0,251	-0,227	0	0,899
ИМТ	-0,352	0	0	-0,198	0,902
ЧД	-0,430	-0,203	0,361	0,765	-0,238
С	-0,818	-0,118	-0,209	-0,412	-0,197
Л	-0,822	0	-0,102	-0,515	-0,220
КСР	0,889	0,342	-0,061	-0,042	-0,023
КСО	0,885	0,353	0	-0,067	0
КДР	0,940	-0,210	-0,061	-0,076	-0,039
КДО	0,963	-0,163	-0,019	-0,057	-0,009
УО	0,809	-0,487	-0,009	-0,110	0
МОС	0,812	-0,507	0,105	0	-0,114
ОПСС	0,631	-0,531	-0,254	0,281	-0,417
ИХ	-0,453	-0,149	0,653	0,548	-0,205
ФВ	0	0,967	0,003	-0,219	0,002
ФУ	-0,218	0,867	-0,364	0	-0,257

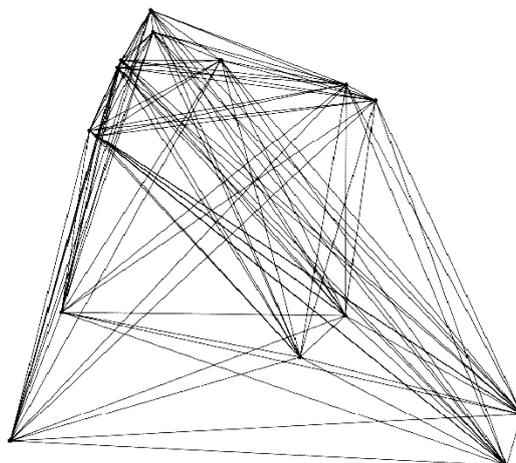


Рис. 1. Визуализация метода Верле в web-приложении

Для 15 биофизических показателей артериальной гипертензии начальной стадии была получена факторная структура на базе данного метода. Полученная факторная структура находится в соответствии с другими методами факторизации.

ЛИТЕРАТУРА

1. Шовин В.А., Гольдяпин В.В. Методы вращения факторных структур // Математические структуры и моделирование. 2015. № 2(34). С. 75–83.
2. Гольдяпин В.В., Шовин В.А. Косоугольная факторная модель артериальной гипертензии первой стадии // Вестник Омского университета. 2010. № 4. С. 120–128.

MULTIDIMENSIONAL SCALING BASED METHOD VERLET

V.A. Shovin

Researcher, e-mail: v.shovin@mail.ru

Omsk Branch of the Federal State budget institution Science Institute of Mathematics
S.L. Soboleva of Siberian Branch of RAS

Abstract. A new method of factorization through multidimensional scaling based on Verlet method is proposed. The results of the correlation matrix factorization using this method are in accordance with the classical methods (principal component, the minimal residual method).

Keywords: Verlet method, multidimensional scaling, factor analysis.

КОМПЬЮТЕРНОЕ ЗРЕНИЕ КАК СРЕДСТВО ИЗВЛЕЧЕНИЯ ИНФОРМАЦИИ ИЗ ВИДЕОРЯДА

Н.В. Манюкова

кандидат педагогических наук, доцент, e-mail: manukovanv@mail.ru

ФГБОУ ВПО «Нижевартовский государственный университет»

Аннотация. В статье рассмотрены области применения компьютерного зрения, представлены этапы распознавания информации из видеоряда, в частности номерных знаков транспортных средств.

Ключевые слова: компьютерное зрение, оператор Собеля, распознавание образов.

Компьютерное зрение призвано решать задачи, связанные со сбором и анализом зрительной информации в различных областях производства, при этом частично или полностью заменяя человека. Вопросы, касающиеся компьютерного зрения, были популярны ещё в 60-е годы XX века, но сейчас они находятся в особой стадии своего развития. Поскольку недавно появилась возможность доступного использования компьютеров и программного обеспечения большим количеством пользователей для обработки изображений и создания программных средств на основе идеи компьютерного зрения.

Под компьютерным зрением понимают теорию и технологию создания искусственных компьютерных систем, которые осуществляют обнаружение, классификацию и отслеживание объектов. Информацию они получают из изображений, которые могут быть представлены определённым видеорядом, изображением с различных камер или трехмерными данными, отсканированными изображениями и т. п.

Цель компьютерного зрения заключается в формировании полезных выводов относительно объектов и сцен реального мира на основе анализа изображений, полученных с помощью датчиков [8].

Имеется множество публикаций, посвящённых компьютерному зрению. Данными вопросами занимаются как отечественные учёные С.Ю. Желтов [3], А.Н. Писаревский [5] и др., так и зарубежные Л. Шапиро, Дж. Стокман [8], Д. Форсайт [7] и др., рассматривая современные подходы в этой области.

Б.В. Анисимов, В.Д. Курганов, В.К. Злобин в своих работах особое внимание уделяют идентификации точечных изображений при распознавании образов и распознаванию пространственных объектов по их плоским изображениям [1]. Они рассматривают особенности цифровой обработки точечных изображений, методы обработки плоских изображений пространственных объектов.

Процесс обработки и анализ изображений в задачах компьютерного зрения детально представлены в исследованиях Ю.В. Визильтера [2], С.Ю. Желто-

Таблица 1. Алгоритм работы приложения

Этап	Шаги
Инициализация	1. Инициализация захвата видео и чтение параметров камеры.
Главный цикл	2. Захват кадра видеопотока.
	3. Обнаружение автомобильного номера в видеопотоке.
	4. Распознавание автомобильного номера.
	5. Занесение информации в базу данных.
Выключение	6. Остановка видеопотока.

ва [3]. Авторы в анализе изображений используют математические модели, описывают задачи линейной и нелинейной фильтрации, применяя приёмы выделения контурных точек, гистограммную обработку изображений, бинаризацию полутоновых изображений, операторы их вычисления, проблемы постобработки контурного изображения.

С.Ю. Желтов [3] и С.И. Протасов [6] рассматривают алгоритмы и методы анализа, визуализации и передачи данных в системах компьютерного зрения.

Многие системы компьютерного зрения реализуют следующие функции:

- Получение изображений;
- Предварительная обработка;
- Выделение деталей;
- Детектирование/Сегментация;
- Высокоуровневая обработка [4].

Результатом компьютерного зрения может являться видоизмененное изображение или список значений определённых параметров изображения, например, размер объекта, цвет, ориентация по отношению к камере, скорость и т. п. Рассмотрим пример идентификации автомобиля через распознавание его номерного знака. Программа распознавания автомобильных номеров работает по следующему принципу (табл. 1).

Описание алгоритма распознавания номера. Программа распознавания автомобильных номеров состоит из нескольких этапов. По окончании их выполнения реализуется распознавание автомобильного номера.

На первом этапе происходит инициализация захвата видео и чтение параметров камеры с помощью диалогового окна, библиотеки AForge.NET, выбора источника видеопотока.

Второй этап состоит из 4 шагов.

Шаги со 2 по 5 повторяются непрерывно, пока приложение не завершит свою работу, в то время как шаги 1 и 6 выполняются только при инициализации и завершении работы приложения.

На первом шаге из видеопотока кадр сохраняется в переменную для его дальнейшей обработки (рис. 1).

На втором шаге к изображению применяется бинаризация. Это попиксель-

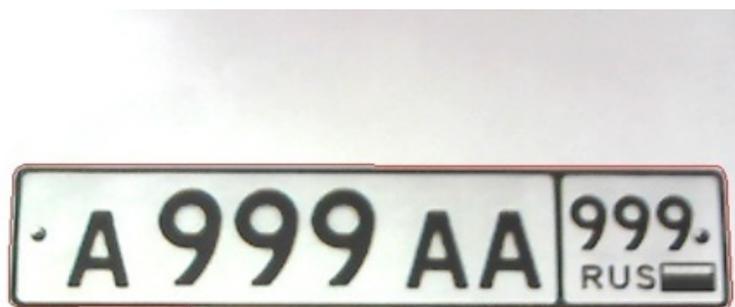


Рис. 1. Исходное изображение

ная операция, в которой значение пикселя сравнивается с фиксированным порогом. Бинаризация применяется ко всему изображению. Пиксели сохраняются в формате 8 бит RGB. Их составляющие значения (красный, зелёный и синий) суммируются и сравниваются с пороговым значением.

$$\text{Бинаризованное значение} = \begin{cases} 1, & \text{если } (Red + Green + Blue) \leq \text{порог}, \\ 0, & \text{если } (Red + Green + Blue) > \text{порог}. \end{cases} \quad (1)$$

Далее используется оператор Собеля (рис. 2), который определяет градиент яркости изображения в каждой точке. Так выявляется направление наибольшего увеличения яркости и величина её изменения. Результат показывает, насколько изменяется яркость изображения в каждой точке и вероятность нахождения точки на грани, а также ориентацию границы.

Завершением второго шага является выделение контуров и поиск контура подходящего по параметрам номерного знака. Для всех замкнутых областей выделяются контуры. Далее координаты всех подходящих контуров записываются в массив. Используя массив с полученными координатами из предыдущего алгоритма, обрабатываются участки, удовлетворяющие параметрам номерного знака. Из исходного изображения по координатам вырезаем предполагаемый номер (рис. 3).

На третьем шаге к данному изображению также применяется бинаризация, оператор Собеля и выделяются все контуры (рис. 4).

Затем происходит распознавание. Для этих целей используются шаблоны (рис. 5).

По координатам каждого контура берётся часть изображения, приводится к размеру шаблона и сравнивается. Для сравнения используется функция, переводящая изображение в массив. Далее массив сравнивается с шаблонными массивами. По результатам сравнения происходит распознавание символов. В итоге получается текстовое представление распознанного номерного знака.

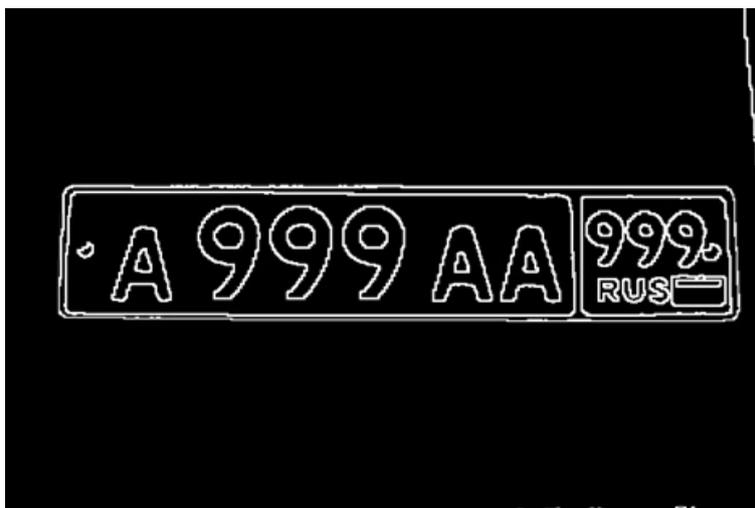


Рис. 2. Бинаризация изображения (порог) и применение оператора Собеля



Рис. 3. Вырезанный участок изображения



Рис. 4. Выделение контуров

1234567890
ABCDEFGHIJKLMORTXU

Рис. 5. Шаблоны букв

Имя столбца	Тип данных	Длина	Разрешить ...	Уникальн...	Первичный...
Номер	nchar	12	Нет	Нет	Нет
Дата	datetime	8	Нет	Нет	Нет
Время	datetime	8	Нет	Нет	Нет

Рис. 6. Структура базы данных

По завершении второго этапа последним шагом текстовое значение номерного знака заносится в базу данных (рис. 6). Таким образом, компьютерное зрение широко используется в различных системах управления, в системах видеонаблюдения и видеорегистрации, применяется при разработке роботов и извлечения информации из видеоряда.

ЛИТЕРАТУРА

1. Анисимов Б.В., Курганов В.Д., Злобин В.К. Распознавание и цифровая обработка изображений: учеб. пособие для студентов вузов. М. : Высш. шк., 1983. 295 с.
2. Визильтер Ю.В., Желтов С.Ю., Бондаренко А.В., Осоков М.В., Моржин А.В. Обработка и анализ изображений в задачах машинного зрения: Курс лекций и практических занятий. М. : Физматкнига, 2010. 672 с.
3. Желтов С.Ю. Разработка теории, методов и алгоритмов машинного зрения в задачах обнаружения объектов: диссертация ... доктора технических наук: 05.13.01 / Желтов Сергей Юрьевич. Москва, 2002. 338 с.
4. Компьютерное зрение [Электронный ресурс]. Режим доступа: http://machinopedia.org/index.php/Компьютерное_зрение (Дата обращения: 25.10.2015).
5. Писаревский А.Н., Чернявский А.Ф., Афанасьев Г.К. и др. Системы технического зрения (принципиальные основы, аппаратное и математическое обеспечение). Л. : Машиностроение. Ленингр. отд-ние, 1988.
6. Протасов С.И. Методы и алгоритмы анализа, передачи и визуализации данных в системах компьютерного стереозрения: автореферат диссертации ... кандидата физико-математических наук: 05.13.17 / Протасов Станислав Игоревич. Воронеж, 2013. 16 с.
7. Форсайт Д., Понс Ж. Компьютерное зрение. Современный подход. Пер. с англ. – М. : Издательский дом «Вильямс», 2004. 928с.
8. Шапиро Л., Стокман Дж. Компьютерное зрение. Пер. с англ. М. : Лаборатория знаний, 2006. 752с.

**COMPUTER VISION AS A MEANS OF EXTRACTING INFORMATION FROM
THE VIDEO****N.V. Manukova**

Ph.D.(Pedagogics), Associate Professor, e-mail: manukovanv@mail.ru

Federal State Budgetary Educational Institution of Higher Professional Education
"Nizhnevartovsk state University"

Abstract. The article discusses the applications of computer vision, presents the stages of the recognition of information from video sequences, in particular registration plates of vehicles.

Keywords: computer vision, Sobel operator, pattern recognition.

КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ СИСТЕМЫ РАЗГРАНИЧЕНИЯ ДОСТУПА ХАРРИСОНА-РУЗЗО-УЛЬМАНА

Д.М. Бречка

к.т.н., доцент кафедры кибернетики, e-mail: dbrechkawork@yandex.ru

В.В. Зубова

студентка, e-mail: zubovaleria60@gmail.com

Омский государственный университет им. Ф.М. Достоевского

Аннотация. Статья посвящена разработке компьютерной модели системы разграничения доступа Харрисона-Руззо-Ульмана. Такая модель необходима для изучения принципов работы систем с дискреционным разграничением доступа.

Ключевые слова: безопасность, HRU, компьютерные системы, программная реализация, объект, матрица доступа, разделение доступа.

Введение

Разделение доступа является одним из основных механизмов защиты информации в современных компьютерных системах. Для формального описания правил разделения доступа часто используются модели разделения доступа. Одной из таких моделей является модель Харрисона-Руззо-Ульмана (Harrison-Ruzzo-Ulman, HRU). Данная модель относится к дискреционным моделям разделения доступа [1].

В работах [2–7] была показана возможность применения модели HRU для анализа безопасности современных операционных систем. Работы [8, 9] показывают возможность составления матрицы доступов в операционных системах, матрица доступов является одним из составных элементов модели HRU.

Целью данной работы является программная реализация системы HRU. Программная реализация необходима для моделирования систем, работающих согласно принципам HRU, что позволит более глубоко изучать вопросы безопасности компьютерных систем.

1. Краткое описание модели HRU

Модель HRU использует субъект-объектный подход к моделированию безопасности компьютерной системы. Основными компонентами HRU являются:

- множество субъектов S — множество активных сущностей системы;
- множество объектов O — множество пассивных сущностей системы, при этом $S \subseteq O$;

- множество прав доступа R — множество действий, которые субъекты могут совершать над объектами;
- матрица доступа M — таблица, в строках которой расположены все субъекты системы, в столбцах — объекты, а в ячейках — соответствующие права доступа.

Функционирование системы рассматривается как последовательное изменение матрицы доступов. При этом над матрицей доступа возможно производить следующие 6 примитивных действий:

- добавить право r в матрицу доступа;
- удалить право r из матрицы доступа;
- добавить субъект;
- добавить объект;
- удалить субъект;
- удалить объект.

Примитивные действия объединяются в команды HRU. Таким образом, функционирование компьютерной системы рассматривается как последовательное применение команд HRU к матрице доступа.

2. Программная реализация модели HRU

Для реализации программной модели был выбран язык C#. Данный язык реализует объектно-ориентированную парадигму программирования, что позволяет составить удобную для разработки и анализа модель системы.

Для моделирования основных сущностей системы создадим три класса: *Subject*, *Objects*, *Monitor*. Класс *Subjects* представляет собой структуру, содержащую в себе набор полей и методов, необходимых для задания субъекта. А класс *Objects* — для объекта соответственно. Класс *Monitor* содержит в себе матрицу доступа M , реализованную как двумерный массив типа *char*, строки которого соответствуют субъектам, а столбцы — объектам доступа. При этом, учитывая, что множество S является подмножеством O , в столбцах матрицы M также содержатся и субъекты доступа.

В данной программе класс *Monitor* является основным, так как в нем отслеживаются запросы на изменение матрицы доступов и непосредственно производятся эти изменения. Диаграмма классов данной программы представлена на рисунке 1.

Для простоты и удобства обработки элементов системы в классе *Monitor* реализованы две структуры *Dictionary : subjects* и *objects*. Каждый экземпляр *Dictionary* имеет два параметра: ключ (*key*) и значение (*value*). В нашем случае параметр *value* будет нести в себе номер, название элемента *subjects* или *objects*, а параметр *key* — идентификатор данного элемента.

Для моделирования примитивных действий с матрицей доступа каждому действию сопоставлен метод класса *Monitor*. Рассмотрим подробнее их реализацию.

Метод *createS(int)*. Данный метод реализует добавление нового субъекта в матрицу доступа M . В первую очередь, происходит проверка того, что субъект

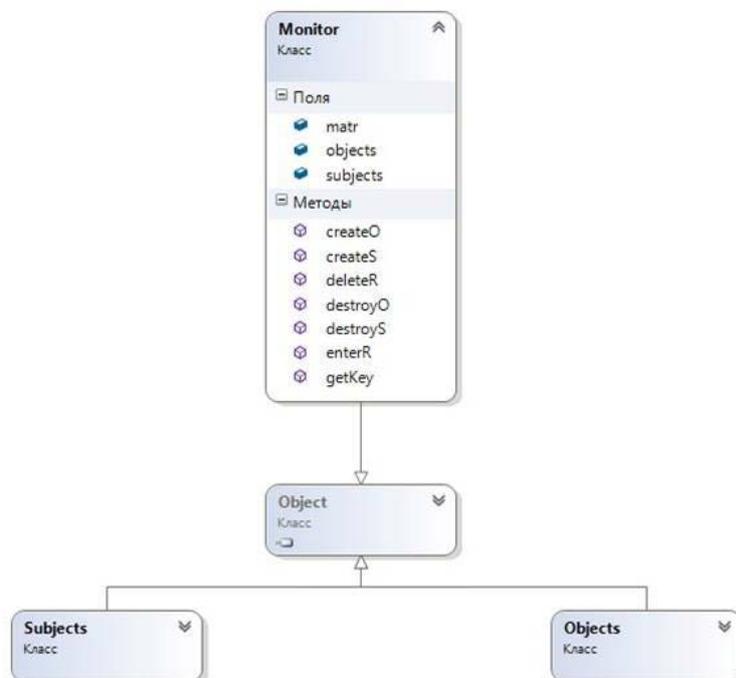


Рис. 1. Диаграмма классов

с данным номером ещё не был создан в системе. Если условие выполнено, то в словарь субъектов добавляется новый элемент, значение параметра *value* которого равно номеру субъекта. Номер субъекта посредством интерфейса программы задаёт пользователь (рис. 2). В качестве значения параметра *key* выбирается количество уже созданных субъектов в словаре. После вызова метода в матрицу доступов добавляются новая строка и столбец с номером *subjects.Key* и названием *subjects.Value*.

Метод *createO(int)*. Реализация данного метода аналогична предыдущему, однако, отличие в том, что для работы с объектами используется словарь *objects*. В результате выполнения в матрицу доступов добавляется только новый столбец с названием *objects.Value* и идентификатором *objects.Key*.

Метод *destroyS(int)* — метод удаления субъекта из системы и из матрицы доступа М. В начале проверяется, существует ли субъект с номером, заданным пользователем, в словаре *subjects*. Проверка выполняется по ключу. С помощью метода *While()* класса *Dictionary*, по заданному пользователем значению *value*, возвращается параметр *key*. Если условие выполнено, то, используя метод *Dictionary.Remove()*, субъект удаляется из словаря *subjects*. Затем из матрицы доступа удаляются строка и столбец с параметром *key*, который был получен на предыдущем шаге.

Метод *destroyO(int)* — метод, аналогичный предыдущему. В отличие от *destroyS()* здесь рассматривается словарь объектов *objects*, а из матрицы до-

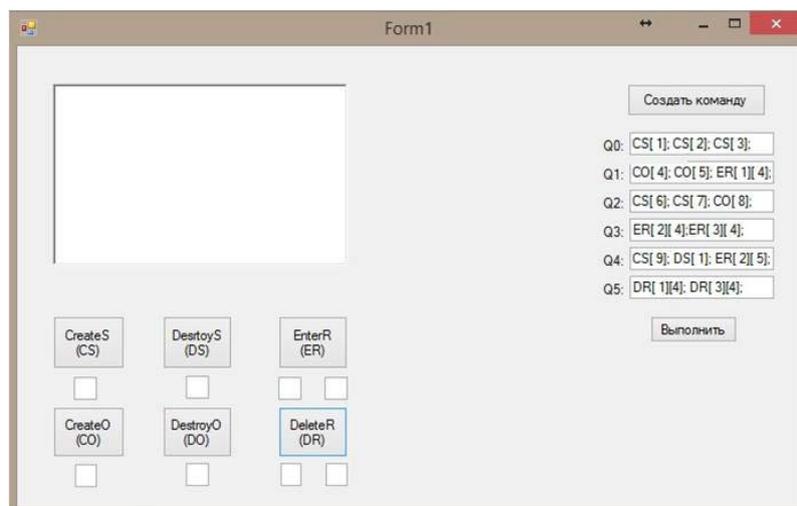


Рис. 2. Интерфейс программы

ступа удаляется только столбец с идентификатором объекта.

Метод $enterR(int, int)$. Условием для выполнения метода $enterR()$ является существование в матрице доступа строки и столбца с номерами, заданными пользователем. Пусть субъект имеет номер i , а объект — номер j . Если условие выполнено, то в матрице на пересечении данных строки и столбца ставится некоторый символ ' r '. Это символизирует, что субъект i теперь имеет право r на объект j .

Метод $deleteR(int, int)$ — метод удаления права r из матрицы доступа M . Пользователь вводит номер i -ой строки и j -го столбца. В методе осуществлена проверка — есть ли в системе субъект и объект с такими номерами и существует ли у субъекта i право r над объектом j . Если условие выполнено, то право r удаляется из матрицы доступа M .

Набор примитивных операторов образует команду HRU. Последовательное выполнение таких команд переводит систему из состояния Q_0 в новое состояние Q_1 . Таким образом, реализовав в классе *Monitor* методы, которые имитируют работу примитивных операторов модели Харрисона-Руззо-Ульмана, можно перевести систему в новое состояние, изменив при этом матрицу доступа M .

Заключение

Результатом проделанной работы является реализация программной модели компьютерной системы, работающей согласно принципам HRU. Тестирование модели показало корректность работы. В дальнейшем планируется использовать данную модель для детального изучения безопасных систем HRU, таких как монотонные и монооперационные системы, а также монооперационных систем в базе [3–5].

ЛИТЕРАТУРА

1. Девянин П.Н. Модели безопасности компьютерных систем: учебное пособие для студентов высших учебных заведений. М. : Издательский центр «Академия», 2005. 143 с.
2. Бречка Д.М., Белим С.В. Исследование безопасности компьютерных систем в модели дискреционного разделения доступа HRU // Математические структуры и моделирование. 2009. Вып. 19. С. 97–103.
3. Бречка Д.М., Белим С.В. Базисный подход в модели безопасности HRU // Проблемы информационной безопасности. Компьютерные системы. 2010. Вып. 2. С. 18–23.
4. Бречка Д.М., Белим С.В. Расширение класса безопасных систем в модели HRU // В мире научных открытий. 2010. № 4(10). Часть 4. С. 9–11.
5. Бречка Д.М., Белим С.В. Классы безопасности в модели HRU // Безопасность информационных технологий. 2010. Вып. 3. С. 26–31.
6. Бречка Д.М., Белим С.В. Исследование безопасности дискреционного разделения доступа в ОС Windows // Математические структуры и моделирование. 2011. Вып. 22. С. 121–130.
7. Проблемы обработки и защиты информации. Книга 1. Модели политик безопасности компьютерных систем / С.В. Белим [и др.]. Омск : ООО «Полиграфический центр КАН», 2010. 164 с.
8. Бречка Д.М., Сыргий Е.В. Формирование матрицы доступов на основе внутренних структур операционной системы LINUX // Материалы I междунар. науч.-практ. конф. «Информационная безопасность в свете Стратегии Казахстан-2050», Астана, 12 сентября 2013. С. 506–511.
9. Бречка Д.М., Сыргий Е.В. Система составления матрицы доступов запущенных процессов в операционной системе Windows // Вопросы защиты информации. 2014. Вып. 3(106) С. 17–24.

COMPUTER MODEL OF HARRISON-RUZZO-ULMAN ACCESS CONTROL SYSTEM**D.M. Brechka**

Ph.D.(Eng.), Associate Professor, e-mail: dbrechkawork@yandex.ru

V.V. Zubova

Student, e-mail: zubovaleria60@gmail.com

Omsk State University n.a. F.M. Dostoevskiy

Abstract. The article describes development of a computer model for Harrison-Ruzzo-Ulman access control system. This model is necessary to study the principles of systems with discretionary access control.

Keywords: security, HRU, computer system, implementation, object, access matrix, access control.

ИСПОЛЬЗОВАНИЕ ТУРБО-КОДЕКА ДЛЯ БЕЗОПАСНОЙ ПЕРЕДАЧИ ДАННЫХ

В.С. Виноградов

студент, e-mail: it.vinogradov@yandex.ru

В.В. Коробицын

доцент, к.ф.-м.н., e-mail: korobits@rambler.ru

М.Н. Московцев

преподаватель, e-mail: mnorthwind@gmail.com

Омский государственный университет им. Ф.М. Достоевского

Аннотация. В современных информационных системах, связанных с передачей информации по различным каналам связи, очень остро стоит вопрос обеспечения информационной безопасности передаваемых данных. В работе предложен алгоритм защиты данных от несанкционированного доступа при передаче в каналах связи с помехами, основанный на алгоритмах класса турбо-кодов. Рассмотрены варианты использования подобного алгоритма, эффективность прикладной реализации, а также вспомогательная информационная инфраструктура для разработанного метода, например, алгоритм генерации ключей.

Ключевые слова: компьютерная безопасность, турбо-коды, турбо-декодер, генерация ключей, алгоритмы защиты данных.

Введение

Современные информационные технологии (например, сотовые сети стандартов 3G [1], 4G [2] или спутниковая радиосвязь [3]) прошли огромный эволюционный путь за последнее десятилетие, а их распространение подтолкнуло развитие систем и средств передачи данных. При этом коммуникационные технологии в определённых условиях должны работать в средах с низким соотношением сигнал-шум, что обеспечивает необходимость использования помехоустойчивого кодирования. Одним из наиболее востребованных на настоящий момент классов помехоустойчивых кодов является класс турбо-кодов. Также нельзя не отметить, что в современном мире все большую актуальность приобретает вопрос защиты информации от несанкционированного доступа или модификации в процессе передачи по сетям связи.

Цель статьи — изучение особенностей передачи данных в средах с низким соотношением сигнал-шум при использовании помехоустойчивого кодирования с помощью алгоритмов класса турбо-кодов, а также изучение аспектов защиты передаваемых таким образом данных.

1. Модель канала передачи данных

Источник данных на выходе обычно выдаёт поток бит информации. Далее следует этап подготовки сообщения для отправки по зашумлённому каналу связи, который включает в себя выполнение двух основных (интересующих нас в контексте данной статьи) операций: сжатие информации и применение помехоустойчивого кодирования.

Сжатие информации необходимо для уменьшения избыточности сообщения, что позволяет снизить затраты на передачу информации. Далее производится отправка данных по каналу связи. После осуществления передачи данных по каналу нужно провести обратную операцию, которая называется декодирование.

Общая схема модели канала передачи данных в средах с наличием шумов представлена на рис. 1.

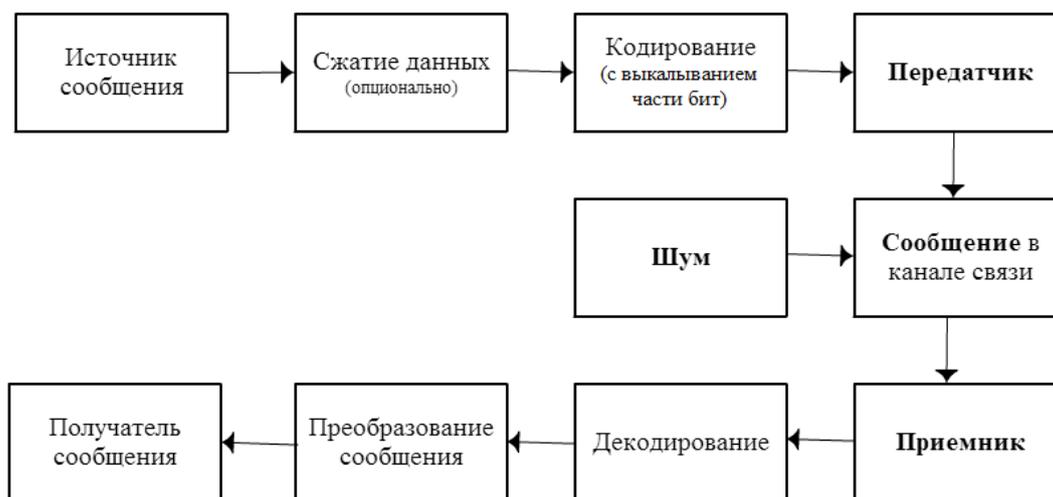


Рис. 1. Общая схема модели канала передачи данных

2. Введение в турбо-коды

Приведём базовые теоретические сведения об особенностях работы и реализации турбо-кодов на практике, необходимые для понимания исследовательской части данной работы.

Турбо-код – это параллельный каскадный блочный систематический код со способностью исправления ошибок. Он является классом высокоэффективных помехоустойчивых кодов. Турбо-коды успешно применяются в спутниковой, цифровой связи и электротехнике, эффективно решая задачи достижения максимальной скорости передачи данных в зашумлённом канале в ограниченной полосе частот.

Основным преимуществом турбо-кодов является то, что они позволяют при разумных вычислительных затратах приблизиться к границе Шеннона [4], ко-

торая представляет собой теоретический предел помехоустойчивости. Также характерной особенностью является то, что сложность декодирования сообщения при использовании турбо-кодов не зависит от длины самого блока передаваемых данных, что позволяет снизить вероятность ошибки декодирования с помощью увеличения длины блока.

Главным же их недостатком является высокая сложность декодирования.

Программная реализация турбо-кодера включает в себя непосредственно турбо-кодер и турбо-декодер.

Параллельный турбо-кодер (рис. 2) имеет в своей структуре два компонентных кодера, перемежитель и мультиплексор.



Рис. 2. Схема турбо-кодера

Компонентные коды чаще всего являются рекурсивными систематическими свёрточными кодами (RSC). Важно заметить, что все входные биты передаются напрямую в выходной поток (далее информационные или систематические биты). Биты выходных потоков кодера называются паритетными.

Мультиплексор используется для слияния битов из трёх выходов в один выходной поток. Он размещает биты по определённому алгоритму, например, первый систематический бит, первый бит из первого потока паритетных битов, первый бит из второго потока паритетных битов и так далее.

В состав турбо-декодера входят демультиплексор, два компонентных декодера, соединённых через перемежитель и деперемежитель (рис. 3).

Декодер получает на вход отчёты, соответствующие принятому из канала связи демодулированному сигналу. Демультиплексор разделяет входные биты на «систематические биты», «паритетные биты №1» и «паритетные биты №2». На вход каждого компонентного декодера подаётся 3 потока бит: систематические, паритетные и поток от другого компонентного декодера.

Турбо-декодер работает итеративно. На вход первого декодера подаётся поток бит, принятый из канала связи, при этом значения априорной информации неизвестны. Выход первого декодера используется для вычисления априорной информации для второго декодера. Второй декодер принимает на вход информацию из канала связи, рассчитанную для него, а также априорную информацию от первого декодера. Далее начинается вторая итерация.

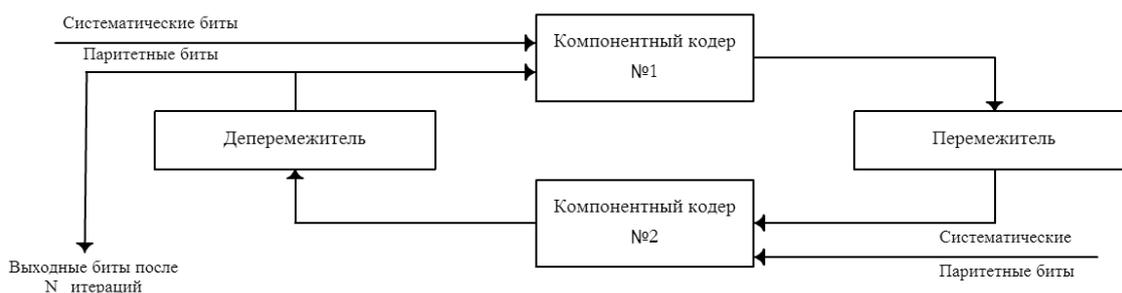


Рис. 3. Схема турбо-декодера

3. Использование турбо-декодера для защиты данных при передаче

Турбо-коды активно используются для передачи данных в средах с низким соотношением сигнал-шум и большим влиянием помех в каналах связи. Инкапсуляция несущего сигнала в шум сама по себе может служить некоторым средством защиты передаваемых данных, потому что она затрудняет отделение информационных битов от помех. Но это не даёт достаточного уровня защиты передаваемых данных (в случае перехвата сигнала при передаче).

Очевидным выглядит использование шифрования критически важных данных с помощью известных стойких криптографических алгоритмов перед процессом кодирования. Но нужно заметить, что шифрование больших объёмов данных (а тем более с последующим кодированием) является вычислительно ресурсоёмкой операцией. Поскольку мы не можем отказаться от кодирования, первоначальной идеей стало объединение этих двух процессов в один для сокращения вычислительных затрат.

Было предложено 2 пути использования турбо-декодера для защиты информации. Их использование целесообразно для разных типов сообщений и разных каналов передачи данных. Отправной точкой для определения этих вариантов стал график зависимости скорости кодирования при достоверном декодировании сообщения от соотношения сигнал-шум (рис. 4), из которого видно, что мы можем строить либо шумоподобную систему с низкой скоростью кодирования, либо систему с повышенной скоростью передачи данных при высоких соотношениях сигнал-шум.

4. Варианты построения защищённых систем

Первым вариантом использования турбо-декодера для защиты передаваемых данных стала организация шумоподобной системы. Особенность такой системы состоит в чрезвычайно низких соотношениях сигнал-шум (левая часть кривой на рис. 4), что требует применения сильных помехоустойчивых кодов. Следствием использования высокой избыточности является низкая скорость передачи данных (при, как правило, небольших размерах сообщения). При этом

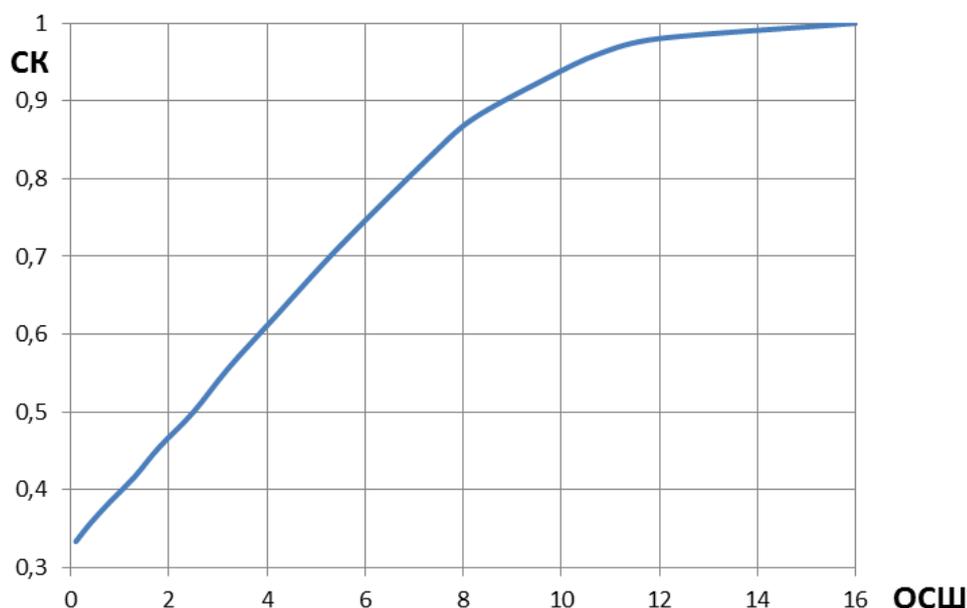


Рис. 4. Зависимость скорости кодирования от отношения сигнал-шум при достоверном декодировании сообщения

помехи выполняют определённую маскирующую роль. Сообщение инкапсулировано в шум, что затрудняет выделение несущих битов из перехваченного потока. Построение таких систем имеет весьма неплохие перспективы и востребовано для реализации определённых задач на практике. Однако больший интерес представляло исследование систем с повышенной скоростью передачи данных.

В таких системах мы по-прежнему вынуждены использовать помехоустойчивое кодирование, поскольку влияние шумов на передаваемое сообщение остаётся значительным. Однако инкапсуляция информационных битов в шум уже не является достаточно надёжным методом защиты от раскрытия сообщения в случае перехвата.

На практике нужно обеспечивать более высокую скорость передачи данных, по сравнению с первым вариантом (из-за больших размеров сообщений и большей стоимости операции шифрования перед передачей). Для увеличения скорости применяется алгоритм выкалывания части паритетных битов (например, выкалывается каждый второй проверочный бит) из сообщения перед отправкой, что позволяет уменьшить количество передаваемых бит.

5. Постановка задачи

Требовалось обеспечить выполнение следующих задач:

1. Исключить передачу информационных бит по каналу связи в открытом виде;

2. Ввести параметры алгоритма, позволяющие на должном уровне обеспечить уникальность логического канала для конкретного сеанса связи;
3. Исследовать восстанавливающую возможность алгоритма в данных условиях.

Для решения первой задачи было предложено не передавать исходное сообщение вместе с проверочными битами (как это происходит при обычном кодировании информации). Была выработана идея восстанавливать оригинальное сообщение, используя только полученные из канала паритетные биты. Практическую реализацию этой задумки можно осуществить с помощью операции выкалывания. В существующей реализации турбо-декодера использовалось выкалывание каждого второго паритетного бита. Было принято решение выколоть все информационные биты и исследовать работу алгоритма в случае восстановления сообщения только по паритетным битам.

6. Выбор ключа для разрабатываемого алгоритма

Под ключом понимаются параметры алгоритма, которые будут обеспечивать уникальность работы алгоритма для конкретного соединения при передаче данных. В качестве ключа можно использовать параметры используемого кодера или перестановку перемежителя.

Напомним, что параметры используемого кодера определяют позиции выводов между сдвигowymi регистрами для каждого выходного потока и их изменение вызывает изменение выхода алгоритма кодирования.

Изменение перестановки также вызывает изменение выходного потока, поскольку выходной поток из перемежителя подаётся на компонентный кодер №2, который участвует в формировании выходного потока турбо-кодера путём мультиплексирования результатов работы обоих компонентных кодеров.

При выполнении операции турбо-кодирования параметры кодера, обычно, остаются статическими для большей эффективности кодирования и декодирования (они подбираются специальным образом для конкретных условий). Поэтому было решено исследовать возможность использования в качестве ключа алгоритма перестановки перемежителя.

7. Результаты тестирования

Для проверки корректности работы описанного алгоритма было проведено моделирование. Необходимо было оценить восстанавливающую способность полученного алгоритма. Для этого были построены кривые помехоустойчивости (рис. 5) и проведено их сравнение с аналогичными кривыми (рис. 6) для обычного декодирования (с выкалыванием только половины паритетных бит).

В эксперименте использовался кодер со следующими параметрами: скорость кодера $R = \frac{1}{2}$, количество состояний 16 (4 блока задержки), генератор $G_1 = 23$, $G_2 = 35$, блочный перемежитель с размером 65536 бит. Цифры 1-18 определяют число итераций декодирования. Тесты были проведены 10 раз с округлением результатов.

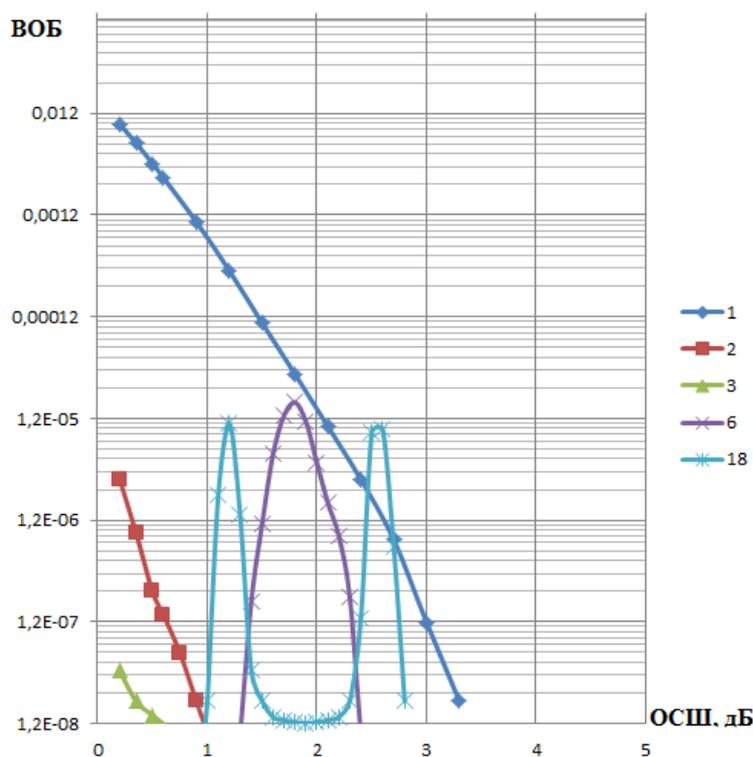


Рис. 5. График зависимости вероятности ошибки от соотношения сигнал-шум для системы с повышенной скоростью передачи данных

Как можно заметить из сравнения полученного графика с эталонным, при числе итераций от 1 до 3 декодер обеспечивает достаточное качество декодирования. Проблемы начинаются только при увеличении числа итераций. Предположительно, это происходит из-за того, что декодер не может только по паритетным битам восстановить информацию, и ошибки «размазываются» по другим битам в ходе следующих итераций декодирования. Решить эту проблему можно, отказавшись от использования более трёх итераций декодирования, либо отказавшись от использования декодера при определённых соотношениях сигнал-шум. Также потенциально более эффективным может оказаться использование усовершенствованного алгоритма декодирования с использованием хвостовых бит.

Второй проблемой оказался размер перестановки, которую предлагалось использовать в качестве ключа. Размер перестановки составляет 65536 бит. Использование ключа такого размера неприемлемо, поэтому было принято решение разработать алгоритм генерации перестановки из ключа меньшего размера.

8. Алгоритм генерации ключа

Размер перестановки определяется размером блока кодируемых данных. Его размер может достигать достаточно больших значений (порядка 65 кбайт в реальных условиях), что делает невозможным использование непосредствен-

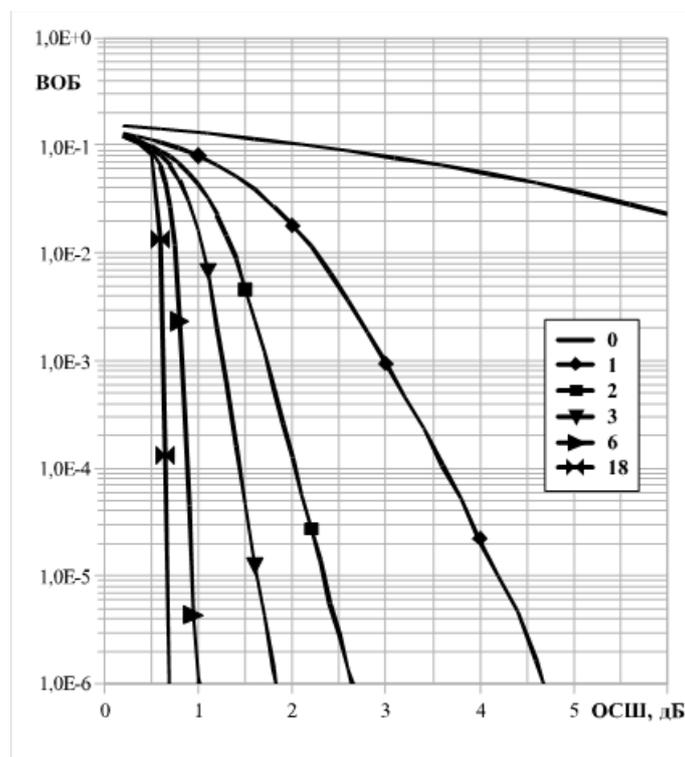


Рис. 6. Кривые помехоустойчивости для стандартного режима декодирования

но самой перестановки, например, из-за сложности передачи ключа, либо его хранения на устройстве. При этом размер ключа должен обеспечивать эффективную защиту от взлома методом прямого перебора.

Перестановка в памяти компьютера представляется как линейный уникальный массив без повторения элементов. Поэтому поставленная задача свелась к задаче генерации линейного уникального массива заданного размера.

Изначально в качестве тестового генератора перестановок был использован генератор псевдослучайных последовательностей, построенный по модели, предложенной канадским исследователем в области стохастического моделирования и оптимизации Пэрри Леквером (Pierre L'Esuyer). Этот генератор использует метод перемешивания Байеса-Дарема (Bays-Durham Shuffling [5]) и имеет достаточную для наших задач длину периода. В качестве зерна генератора используется число типа long размерностью 4 байта (32 бита).

Было предложено в качестве ключа использовать зерно, а также внутренние параметры генератора. Очевидны следующие проблемы: параметры генератора подобраны специальным образом для максимальной эффективности алгоритма генерации псевдослучайных последовательностей. Поэтому использование случайных значений параметров не является эффективным. Длина ключа в 32 бита (ключ — зерно одного генератора) для современных алгоритмов не считается достаточной, поскольку позволяет осуществить взлом методом полного перебора за ограниченное время. Учитывая изложенные обстоятельства, от подобного подхода построения ключа было решено отказаться.

Был разработан алгоритм построения ключа, основанный на тех же принципах генерации перестановки перемешителя, которые использовались ранее. Суть алгоритма заключается в использовании нескольких генераторов псевдослучайных чисел. При этом ключ для алгоритма формируется с помощью конкатенации зёрен каждого генератора в бинарном виде.

Приведем описание алгоритма генерации перестановки с использованием сформированного ключа:

1. Берём полученный ключ (обозначим его длину за $keyL$) и разбиваем его на числа типа long, по 32 бита каждое. Количество таких чисел обозначим как $blockN$
2. Инициализируем массив perm, представляющий необходимую перестановку длины $permL$.
3. Инициализируем $blockN$ генераторов случайных чисел, в качестве зерен используем числа, полученные в пункте 1 алгоритма.
4. Вычисляем длину вспомогательных массивов $arrL$, $arrL = \frac{permL}{blockN}$. Данные массивы будут строиться с помощью генераторов случайных чисел (инициализированных в пункте 3). Каждый массив должен содержать уникальные числа от 1 до $arrL$.
5. Генерируем $(blockN - 1)$ вспомогательных массивов $arr[j]$ ($j = 1..arrL$) длины $arrL$. Последний генератор, инициализированный в пункте 3, используется для расчёта $(permL - (blockN - 1) \cdot arrL)$ чисел, чтобы заполнить «конец» перестановки в случае, если длина перестановки не кратна $blockN$. При этом если число на выходе генератора не попадает в определённый в пункте 4 диапазон, оно просто отбрасывается.
6. Делаем пересчёт массивов, сгенерированных в пункте 5. Для j -го массива и его i -го элемента формула для пересчёта будет равна: $arr[j][i] = arr[j][i] + (j - 1) \cdot (arrL)$.
7. Определяем первые $arrL \cdot (blockN - 1)$ элементов финальной перестановки по формуле $perm[j] = arr[i, k]$, где $i = j \bmod (blockN - 1)$, $k = \frac{j}{arrL}$. При этом если $i = 0$, оно принимается равным $(blockN - 1)$.
8. Формируем последний фрагмент необходимой перестановки, используя последний генератор псевдослучайных чисел. Для этого последовательно генерируем $(permL - (blockN - 1) \cdot arrL)$ чисел, лежащих в диапазоне от 1 до $(permL - (blockN - 1) \cdot arrL)$. Каждое число увеличиваем на $(blockN - 1) \cdot (arrL)$ и дописываем в конец массива perm последовательно. Выходы генератора, не попадающие в нужный диапазон, отбрасываются.
9. Полученный массив perm будет содержать готовую перестановку для перемешителя, используемого в алгоритме турбо-кодирования.

Наглядно алгоритм генерации перестановки по ключу представлен на рис. 7.

Как можно заметить, полученный ключ имеет следующие свойства:

1. Минимальная длина ключа — 64 бита. Данная особенность возникает из необходимости иметь как минимум 2 генератора псевдослучайных последовательностей с разным инициализирующим зерном. Это свойство не накладывает существенных ограничений на использование алгоритма.

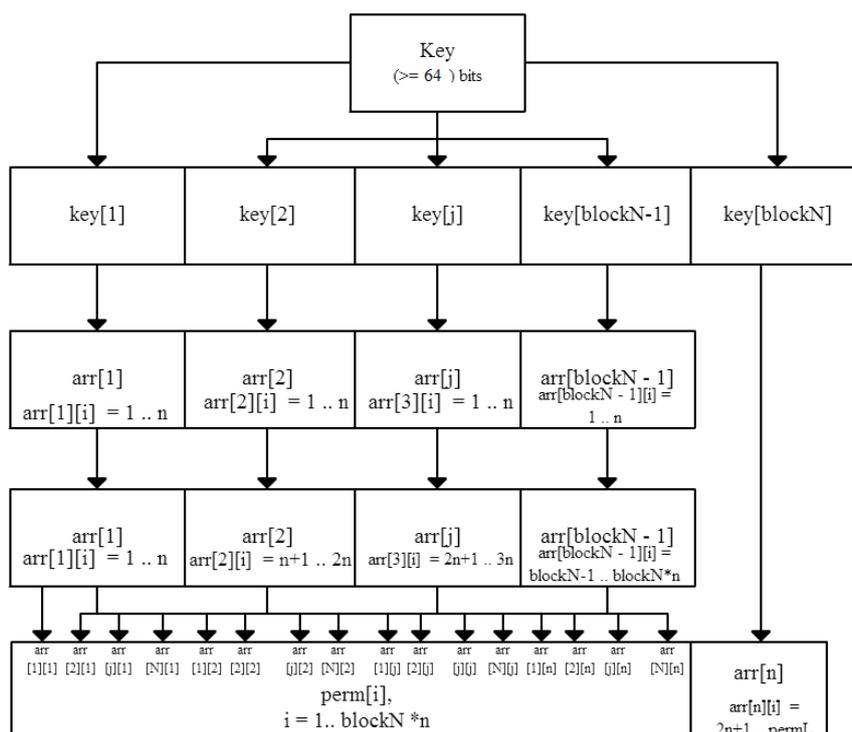


Рис. 7. Схема алгоритма генерации перестановки по ключу

2. Длина ключа должна быть кратна 32 битам.
3. Длина ключа сверху никак не ограничивается алгоритмом, но при этом необходимо отметить, что использование слишком больших ключей не является целесообразным.

При этом подобный алгоритм легко поддается распараллеливанию, что является очень удобным для организации системы управления ключами. Таким образом, с помощью алгоритмов параллельных вычислений можно формировать банки ключей для подобных систем связи.

ЛИТЕРАТУРА

1. An Investigation of Code Matched Interleaver for 3G Turbo Code Systems [Электронный ресурс]. URL: http://www.researchgate.net/profile/Balamuralithara_Balakrishnan/publication/26560302_An_Investigation_of_Code_Matched_Interleaver_for_3G_Turbo_Code_Systems/links/00b4953070a93a2ce3000000.pdf (дата обращения: 03.03.2015).
2. Особенности стандарта LTE [Электронный ресурс]. URL: <http://сmpo.vlsu.ru/edu/2013/A7.pdf> (дата обращения: 05.03.2015).
3. Спецификация DVB-RCS [Электронный ресурс]. URL: https://www.dvb.org/resources/public/factsheets/DVB-RCS2_Factsheet.pdf (дата обращения: 14.03.2015).

4. Морелос-Сарагоса Р. Искусство помехоустойчивого кодирования. Методы, алгоритмы, применение. М. : Техносфера, 2005. 320 с.
5. Метод перемешивания Байеса-Дарема [Электронный ресурс]. URL: http://en.wikibooks.org/wiki/Statistics/Numerical_Methods/Random_Number_Generation#Bays-Durham_Shuffling_of_Uniform_Deviates (дата обращения: 10.05.2015).

TURBO CODEC USING FOR THE SECURITY DATA TRANSMISSION

V.S. Vinogradov

Student, e-mail: it.vinogradov@yandex.ru

V.V. Korobitsin

Ph.D.(Phys.-Math.), Associate Professor, e-mail: korobits@rambler.ru

M.N. Moskovtsev

Instructor, e-mail: mnorthwind@gmail.com

Omsk State University n.a. F.M. Dostoevskiy

Abstract. The question of safe data transmission is very actual in the modern communication systems. In this article we offer the algorithm of safe data transmission which is based on turbocodes class algorithms. Options of such algorithm using, efficiency of current implementation and some information infrastructure for the developed method (for example, algorithm of key generation) are considered.

Keywords: computer security, turbo-codes, turbo-decoder, key generation, data protection algorithms.

Научный журнал

Математические структуры И моделирование

№4(36)

Главный редактор

А.К. Гуц

Выпускающий редактор

Д.Н. Лавров

Технический редактор

Н.Ф. Богаченко

Корректор:

И.Н. Баловнева

Проверка корректности перевода:

Е.А. Илюшечкин

А.Н. Кабанов

Адрес научной редакции

Россия, 644077, Омск, пр. Мира, 55А

Омский государственный университет

E-mail: guts@omsu.ru, lavrov@omsu.ru

Электронная версия журнала:

<http://msm.univer.omsk.su>

<http://msm.omsu.ru>



Подписано в печать 07.12.2015. Формат 60 × 84 1/8.

Усл. печ. л. 16,8. Тираж 100 экз. Заказ № 231.

Отпечатано на полиграфической базе издательства ОмГУ им. Ф.М. Достоевского
644077, г. Омск, пр. Мира, 55А

ISSN 2222-8772



9 772222 877005



15036 >