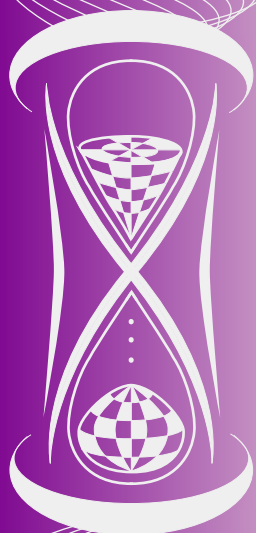


ISSN 2222-8772

МАТЕМАТИЧЕСКИЕ
СТРУКТУРЫ
И
МОДЕЛИРОВАНИЕ

№ 3(39)
2016



**МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ**

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

**МАТЕМАТИЧЕСКИЕ
СТРУКТУРЫ
И
МОДЕЛИРОВАНИЕ**

№ 3(39)

Омск
2016

Математические структуры и моделирование. — Омск : Омский государственный университет, 2016. — № 3(39). — 192 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. 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. |
| B. Zilber | Dr.Sc. (Phys.-Math.), Professor of Mathematical Logic, Mathematical Institute, University of Oxford, UK. |
| 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. |
| A. Jadczyk | Ph.D., Professor, Researcher, Laboratoire de Physique, Universite de Toulouse III et CNRS, France. |

Учредитель

Федеральное государственное бюджетное
образовательное учреждение
высшего 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 реферируемым. Рефераты статей публикуются в «Реферативном журнале», в журналах «Zentralblatt für Mathematik» (Германия), 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. *Спектр материи Гейзенберга в абстрактно-алгебраическом подходе* 5
- В.Н. Бородихин. *Исследование поведения систем при фазовых переходах на стадии коалесценции* 24
- В.Н. Степанов. *Оценка устойчивости для выпуклых поверхностей* 40
- А.К. Гуц. *Не-гёделевская машина времени* 48
- Ю.А. Лебедев. *Специфика моделирования эвретических пространств* 59
- М.О. Escobar, V. Kreinovich. *How to Compute Von Neumann-Morgenstern Solutions* 68

Прикладная математика и моделирование

- С.В. Белим, С.Б. Ларионов. *Сегментация изображений на основе алгоритма выделения сообществ на графе* 74
- К.Т. Кошеков, Н.В. Астапенко. *Метод и алгоритм вычисления объёма заданной дискретно 3D поверхности с помощью полиномов Лагранжа* 86
- Н.С. Новаковский. *Математическое моделирование сильного сжатия одномерных слоёв газа в конфигурации Р. Мизеса* 93
- А.К. Гуц, Л.А. Володченкова. *Дифференциальные игры в экологии человека и в социологии* . 110

Компьютерные науки

- С.В. Белим, Д.М. Бречка, Т.А. Горбунова, И.Б. Ларионов, И.В. Шмидт. *Интеллектуальная геоинформационная система археологических объектов* 119
- Д.С. Сиберт. *Пошаговая инструкция написания Shell Namespace Extension* 127
- A.V. Taran, D.N. Lavrov. *Future Internet Architecture: Clean-Slate vs Evolutionary Design* 142
- A. Taran, D.N. Lavrov. *Adaptive Video Multicast Streaming over Multirate Wireless LANs* 152

Продолжение на следующей странице

Наши публикации



Информационная безопасность

- К.И. Лейфер, Д.Н. Лавров. *Идентификация личности по биометрическим характеристикам жестов на сенсорном экране*.....163
- А.Е. Дизер, Е.С. Дизер, Т.М. Опарина. *Модификация метода Куттера-Джордана-Боссена скрытого хранения информации в изображениях формата JPEG*177
- Т.В. Вахний, А.К. Гуц. *DDOS-атаки как дифференциальная игра*.....184

СПЕКТР МАТЕРИИ ГЕЙЗЕНБЕРГА В АБСТРАКТНО-АЛГЕБРАИЧЕСКОМ ПОДХОДЕ

В.В. Варламов

д.ф.-м.н., e-mail: varlamov@subsui.ru

Сибирский государственный индустриальный университет

Аннотация. Исследуется алгебраическая структура спектра материи. Показывается, что базовой конструкцией, лежащей в основании понятия спектра материи, введённого Гейзенбергом, является двухуровневое гильбертово пространство. Двухуровневая структура гильбертова пространства задаётся следующей парой: 1) сепарабельное гильбертово пространство, в рамках которого определяются операторные алгебры и фундаментальные симметрии; 2) несепарабельное (физическое) гильбертово пространство, т.е. пространство состояний (энергетических уровней) спектра материи, в котором действуют динамические и калибровочные симметрии. Приведено разложение физического гильбертова пространства на когерентные подпространства, что позволяет единым образом охватить весь наблюдаемый спектр состояний, включая лептонный, мезонный и барионный секторы спектра материи.

Ключевые слова: спектр материи, гильбертово пространство, когерентные подпространства, принцип суперпозиции, принцип редукции, симметрии.

1. Введение

В настоящее время накоплен огромный наблюдательный материал по спектроскопии элементарных частиц. Открытие динамических симметрий ($SU(3)$ - и $SU(6)$ -симметрии и т.д.) позволило частично упорядочить и систематизировать эти наблюдательные данные, главным образом в области барионного спектра. Однако до сих пор остаётся неясной общая структура спектра элементарных частиц. По мнению Гейзенберга [1], главной причиной создавшейся ситуации является непонимание природы элементарной частицы. Кварковые модели, базирующиеся на приближенных динамических $SU(N)$ -симметриях, не дают ответа на этот вопрос, поскольку эти модели не включают в себя лептонный сектор. В стандартной модели фиксируется разделение адронного (кваркового) и лептонного секторов, плюс к этим двум добавляется калибровочный сектор (сектор полей-переносчиков взаимодействий). Такое тройственное разделение спектра частиц в стандартной модели на три класса «фундаментальных частиц» привело к ещё большей путанице в понимании общей структуры этого спектра, что и послужило главным мотивом в стремлении многих физиков-теоретиков

выйти за пределы стандартной модели, т.е. найти другую альтернативную схему описания спектра элементарных частиц.

Одной из наиболее интересных и многообещающих альтернативных схем является исследовательская программа, предложенная Гейзенбергом [2]. Главной идеей этой программы является представление о том, что всё огромное множество элементарных частиц есть **спектр материи**, а каждая элементарная частица представляет собой тот или иной энергетический уровень этого спектра. Принципиально важным моментом всей программы является отказ Гейзенберга от понятия фундаментальной частицы (все уровни спектра материи равноправны). Гейзенберг утверждает, что понятие «состоит из» уже не работает в физике элементарных частиц. Если продолжить применять это понятие, то получим ответ, что каждая данная частица состоит из всех известных частиц. Таким образом, физическое знание подошло к границам той области, где понятие «состоит из» оказывается уже не имеющим смысла. Гейзенберг пишет: «... при столкновении двух элементарных частиц высоких энергий в процессе их распада могут возникать разнообразные частицы, однако, эти частицы не обязательно окажутся меньше тех, что подверглись делению. Речь идёт фактически о возникновении новых элементарных частиц из кинетической энергии сталкивающихся объектов. Понятие деления тем самым утратило свой смысл, как и понятие наименьшей частицы. Когда энергия становится материей, возможность чего была давно уже признана теорией относительности, энергия принимает форму элементарной частицы» [2, с. 346]. И далее: «... все многообразие элементарных частиц объясняется некоторой универсальной первоматерией, которую можно назвать энергией или материей. В этом случае ни одна из элементарных частиц принципиально не выделяется среди других в качестве фундаментальной частицы» [3, с. 29]. Место фундаментальных частиц занимают фундаментальные симметрии: «Спрашивается, чем же тогда заменить понятие фундаментальной частицы. Полагаю, что нам следовало бы заменить его понятием фундаментальной симметрии. Фундаментальными симметриями определяется основополагающий закон, обуславливающий спектр элементарных частиц ... Тщательный анализ наблюдений даёт мне основание заключить, что, помимо Лоренцевой группы подлинными симметриями являются также SU_2 , принцип масштабной инвариантности и дискретные преобразования P , C , T , но я не стал бы причислять к фундаментальным симметриям SU_3 или более высокие симметрии этого рода, поскольку они могут возникать благодаря динамике системы в качестве приближенных симметрий» [2, с. 106].

В настоящей статье представлена абстрактно-алгебраическая формулировка основных положений спектра материи. Общая алгебраическая структура спектра материи, определяемая двухуровневым гильбертовым пространством, исследуется в п.2. Показывается, что основные энергетические уровни (состояния) спектра материи строятся на основе циклических представлений в рамках конструкции Гельфанда-Найма-Сигала. Конкретная реализация операторной алгебры осуществляется посредством спинорной структуры, ассоциированной с каждым циклическим представлением. Чистые состояния (циклические представления), задающие уровни спектра материи, подразделяются

относительно заряда (действия псевдоавтоморфизма спинорной структуры) на подмножества заряженных, нейтральных и истинно нейтральных состояний. Показывается, что структура спектра материи задаётся разбиением физического гильбертова пространства (пространства состояний) на когерентные подпространства, при этом принцип суперпозиции имеет место в ограниченной форме, т.е. в пределах когерентных подпространств. Ни один уровень (состояние) спектра материи не является выделенным или «фундаментальным», все уровни представляют собой актуализированные (локализованные) состояния квантовых микрообъектов (квантовой системы $\equiv$ спектра материи). Ключевую роль играет понятие симметрии. Симметрии спектра материи разделяются на три вида: *фундаментальные* симметрии, участвующие в формировании состояний, *динамические* и *калибровочные* симметрии, связывающие состояния между собой.

2. Общая алгебраическая структура спектра материи

В этом параграфе рассмотрим общую структуру спектра материи, которая задаётся строением гильбертова пространства элементарной частицы. В свою очередь, это гильбертово пространство имеет двухуровневую структуру. На первом уровне имеем *сепарабельное гильбертово пространство* H_∞ , в котором согласно стандартным правилам локальной квантовой феноменологии (картины Шредингера) определяются **наблюдаемые** (C^* -алгебры), состояния, спектры наблюдаемых и *фундаментальные симметрии*. Основной наблюдаемой является *энергия* (эрмитов оператор H), фундаментальная симметрия задаётся группой Лоренца $SO_0(1,3)$. На втором уровне имеем *несепарабельное гильбертово пространство* $H^S \otimes H^Q \otimes H_\infty^1$, в котором главными структурными составляющими являются **состояния**² (лучи). Векторы состояния в $H^S \otimes H^Q \otimes H_\infty$ конструируются из неприводимых конечномерных представлений группы $SL(2, \mathbb{C})$. Эти представления определены в собственных подпространствах $H_E \subset H_\infty$ оператора энергии H . Таким образом, векторы состояния в $H^S \otimes H^Q \otimes H_\infty$ задают спиновые и зарядовые степени свободы элементарной

¹Как известно, фундаментальная двойственность природы, впервые обнаруженная в эксперименте Штерна-Герлаха, привела к введению спина в квантовой механике. В 1927 г. Паули [4] дал строгий метод введения спина в квантовую механику посредством определения удвоенного гильбертова пространства $H_2 \otimes H_\infty$. В свою очередь, гильбертово пространство $H^S \otimes H^Q \otimes H_\infty$ является естественным обобщением пространства Паули $H_2 \otimes H_\infty$.

²Как отмечал Сигал [5], в качестве основного объекта, связанного с физической системой, можно взять или наблюдаемую или состояние: «*Относительно первичности состояний или наблюдаемых*. Что более фундаментально: наблюдаемые или состояния — этот вопрос во многом подобен аналогичному вопросу о курице или яйце. Если оставить в стороне метафизику, ни одно из этих понятий не имеет решающего преимущества как фундаментальная концепция. Однако в настоящее время не существует аналитической трактовки, исходящей из состояний и развитой в такой же мере, как трактовка, основанная на понятии наблюдаемой» [5, с. 32]. Сигал отмечает также, что в работах Биркгофа и фон Неймана [6] и Макки [7] сформулированы предварительные положения трактовки, основанной на понятии состояния. Однако трактовка Биркгофа-фон Неймана-Макки не получила дальнейшего развития в силу очевидных (главным образом исторических) обстоятельств.

частицы. Сама же элементарная частица представляет собой суперпозицию векторов состояния в $\mathbf{H}^S \otimes \mathbf{H}^Q \otimes \mathbf{H}_\infty$, т.е. в случае чистых (запутанных) состояний имеем несепарабельное (нелокальное) состояние. Следовательно, двухуровневая структура гильбертова пространства элементарной частицы задаётся парой $(\mathbf{H}_\infty, \mathbf{H}^S \otimes \mathbf{H}^Q \otimes \mathbf{H}_\infty)^3$.

2.1. Гильбертово пространство $\mathbf{H}_\infty$

Итак, на первом уровне имеем сепарабельное гильбертово пространство $\mathbf{H}_\infty$, т.е. $\mathbf{H}_\infty$ является банаховым пространством со счётной базой, всюду плотной в $\mathbf{H}_\infty$ (любой элемент из $\mathbf{H}_\infty$ представляется как предел последовательности элементов из счётного множества). На уровне $\mathbf{H}_\infty$ главную роль играют наблюдаемые. Исходным объектом рассмотрения является C^* -алгебра $\mathfrak{A}$ с единицей (*алгебра наблюдаемых* или алгебра ограниченных наблюдаемых). Эрмитовы элементы этой алгебры являются ограниченными *наблюдаемыми*⁴. Положительный функционал ω над $\mathfrak{A}$, нормированный условием $\|\omega\| \equiv \omega(1) = 1$, называется *состоянием* алгебры $\mathfrak{A}$. Множество всех состояний алгебры $\mathfrak{A}$ будем обозначать через $S(\mathfrak{A})$. Величина $\omega(A)$ при $A = A^*$ понимается как *среднее значение* наблюдаемой A в состоянии ω . Множество $S(\mathfrak{A})$ является *выпуклым*, т.е. для любых двух состояний ω_1, ω_2 и $\lambda_1, \lambda_2 \geq 0, \lambda_1 + \lambda_2 = 1$, имеем $\lambda_1\omega_1 + \lambda_2\omega_2 \in S(\mathfrak{A})$. Состояние ω называется *смешанным* (или статистической смесью), если оно представимо в *виде* $\omega = \lambda\omega_1 + (1 - \lambda)\omega_2$, где $0 < \lambda < 1$ и ω_1, ω_2 — два различных состояния алгебры $\mathfrak{A}$. Состояния, не являющиеся смешанными, называются *чистыми* (чистые состояния есть экстремальные точки множества $S(\mathfrak{A})$). Множество всех чистых состояний C^* -алгебры $\mathfrak{A}$ обозначим через $PS(\mathfrak{A})$. Пусть $\mathfrak{S}$ — множество таких состояний алгебры $\mathfrak{A}$, для которых выполняется условие $\omega(A) \geq 0$ для всех $\omega \in \mathfrak{S}$, т.е. A есть положительный элемент алгебры $\mathfrak{A}$ (A представимо в виде $A = B^*B$). Тогда множество $\mathfrak{S}$ будем называть множеством *физических состояний* алгебры наблюдаемых $\mathfrak{A}$, а пару $(\mathfrak{A}, \mathfrak{S})$ будем также называть *физической системой*.

Для произвольной C^* -алгебры $\mathfrak{A}$ вероятность перехода между двумя чистыми состояниями $\omega_1, \omega_2 \in PS(\mathfrak{A})$ определяется формулой $|\langle \Phi_1 | \Phi_2 \rangle|^2 = \omega_1 \cdot \omega_2 = 1 - 1/4 \|\omega_1 - \omega_2\|^2$ (см. [10]), где $|\Phi_1\rangle$ и $|\Phi_2\rangle$ — единичные векторы пространства $\mathbf{H}_\infty$. При этом $\omega_1 \cdot \omega_2 = \omega_2 \cdot \omega_1$ и $\omega_1 \cdot \omega_2$ всегда заключено на отрезке $[0, 1]$. Соответственно $\omega_1 \cdot \omega_2 = 1$ в точности тогда, когда $\omega_1 = \omega_2$. Будем называть два чистых состояния ω_1 и ω_2 *ортгональными*, если вероятность перехода $\omega_1 \cdot \omega_2$ равна нулю. Соответственно два подмножества S_1 и S_2 в $PS(\mathfrak{A})$ являются взаимно ортгональными, если $\omega_1 \cdot \omega_2 = 0$ для всех $\omega_1 \in S_1$ и $\omega_2 \in S_2$. Далее, непустое под-

³В согласии с концепцией Гейзенберга-Фока [2, 8], реальность имеет двухуровневую структуру: *потенциальная реальность* и *актуальная реальность*. Гейзенберг утверждал, что любой квантовый объект (например, элементарная частица) принадлежит обеим сторонам реальности: во-первых, потенциальной реальности как суперпозиция, и, во-вторых, актуальной реальности после редукции суперпозиции, т.е. измерения (более подробно см. [2, 8, 9]).

⁴Эрмитовы элементы C^* -алгебры $\mathfrak{A}$ образуют *алгебру Иордана* $\mathfrak{A}_h$. В $\mathfrak{A}_h$ определяются линейные комбинации с вещественными коэффициентами и квадрат каждого элемента с симметрическим произведением (псевдопроизведением) $A \circ B = 1/4[(A + B)^2 - (A - B)^2]$.

множество $S \in PS(\mathfrak{A})$ называется *нераспадающимся*, если его нельзя разбить на два непустых ортогональных подмножества. Следуя [10], будем считать, что всякое максимальное нераспадающееся множество (т.е. нераспадающееся множество, не являющееся собственным подмножеством другого нераспадающегося множества чистых состояний из $PS(\mathfrak{A})$) является *сектором*. Итак, $PS(\mathfrak{A})$ разбивается на секторы, следовательно, в $PS(\mathfrak{A})$ существует отношение эквивалентности $\omega_1 \sim \omega_2$ в точности тогда, когда существует нераспадающееся множество в $PS(\mathfrak{A})$, содержащее ω_1 и ω_2 . Следовательно, $PS(\mathfrak{A})$ единственным образом разбивается на попарно непересекающиеся и взаимно ортогональные секторы, которые в точности совпадают с *классами эквивалентности* в $PS(\mathfrak{A})$.

Одним из важнейших аспектов теории C^* -алгебр является двойственность между состояниями и представлениями. Связь между состояниями и неприводимыми представлениями операторных алгебр была впервые явно сформулирована Сигалом [11]. Пусть π есть некоторое представление алгебры $\mathfrak{A}$ в гильбертовом пространстве H_∞ , тогда для любого ненулевого вектора $|\Phi\rangle \in H_\infty$ выражение

$$\omega_\Phi(A) = \frac{\langle \Phi | \pi(A) \Phi \rangle}{\langle \Phi | \Phi \rangle} \quad (1)$$

определяет состояние $\omega_\Phi(A)$ алгебры $\mathfrak{A}$, называемое *векторным состоянием*, ассоциированным с представлением π и соответствующим вектору $|\Phi\rangle$. Если ρ — матрица плотности в H_∞ , то

$$\omega_\rho(A) = \text{Tr}(\rho\pi(A)).$$

Аналогично, $\omega_\rho(A)$ является состоянием, ассоциированным с представлением π и соответствующим матрице плотности ρ . Состояния $\omega_\rho(A)$ суть статистические смеси векторных состояний (1). Пусть S_π — множество всех состояний, ассоциированных с представлением π . Два представления π_1 и π_2 с одним и тем же множеством ассоциированных состояний (т.е. с $S_{\pi_1} = S_{\pi_2}$) называются феноменологически эквивалентными, что соответствует унитарно эквивалентным представлениям. Более того, множество $PS(\mathfrak{A})$ всех чистых состояний C^* -алгебры $\mathfrak{A}$ совпадает с множеством всех векторных состояний, ассоциированных со всеми *неприводимыми представлениями* алгебры $\mathfrak{A}$. Всякий сектор в $PS(\mathfrak{A})$ в точности совпадает с множеством всех векторных состояний, ассоциированных с некоторым неприводимым представлением алгебры $\mathfrak{A}$. При этом два унитарно эквивалентные (соответственно унитарно неэквивалентные) неприводимые представления определяют один и тот же сектор (соответственно взаимно ортогональные секторы).

Далее, пусть π — представление C^* -алгебры $\mathfrak{A}$ в H_∞ и пусть $|\Phi\rangle$ — циклический вектор⁵ представления π , определяющий состояние ω_Φ . Тогда согласно

⁵Вектор $|\Phi\rangle \in H_\infty$ называется циклическим для представления π , если все векторы вида $|\pi(A)\Phi\rangle$ (где $A \in \mathfrak{A}$) образуют тотальное множество в H_∞ , т.е. такое множество, замыкание линейной оболочки которого всюду плотно в H_∞ . Представление π с циклическим вектором называется циклическим.

конструкции Гельфанда-Наймарка-Сигала (см. [10]) каждое состояние определяет некоторое представление алгебры $\mathfrak{A}$, причём результирующее представление неприводимо в точности тогда, когда состояние является чистым. Тесная связь между состояниями и представлениями C^* -алгебры, следующая из конструкции ГНС, позволяет рассматривать представления алгебры как эффективный способ организации состояний. Это станет ещё более очевидным при конкретной реализации $\pi(\mathfrak{A})$.

2.1.1. Фундаментальные симметрии

В абстрактно-алгебраической формулировке под симметрией понимается преобразование физической системы, не изменяющее ее структурных свойств. В свою очередь, физическая система характеризуется алгеброй наблюдаемых $\mathfrak{A}$ и множеством состояний $S(\mathfrak{A})$. Следуя Гейзенбергу, условимся считать, что на уровне сепарабельного гильбертова пространства H_∞ мы имеем *фундаментальные (первичные) симметрии*. Определим фундаментальную симметрию как пару биекций $\alpha : \mathfrak{A} \rightarrow \mathfrak{A}$ и $\alpha' : S(\mathfrak{A}) \rightarrow S(\mathfrak{A})$, удовлетворяющих условию согласования: $(\alpha'\omega)(\alpha A) = \omega(A)$ для всех $A \in \mathfrak{A}$, $\omega \in S(\mathfrak{A})$. Множество всех симметрий физической системы образует группу с умножением, определяемым композицией биекций. Произведение двух симметрий (α, α') и (β, β') есть симметрия $(\alpha\beta, \alpha'\beta')$, где $(\alpha\beta)(A) \equiv \alpha[\beta(A)]$ и $(\alpha'\beta')(\omega) = \alpha'[\beta'(\omega)]$.

Будем называть группу G *группой фундаментальной симметрии C^* -алгебры наблюдаемых $\mathfrak{A}$* , если задан гомоморфизм $g \rightarrow (\alpha_g, \alpha'_g)$ группы G в группу всех симметрий системы $(\mathfrak{A}, S(\mathfrak{A}))$. Будем всегда предполагать, что G — некомпактная группа Ли (например, группа Лоренца, группа Пуанкаре или конформная группа). Тогда выполняется следующее условие непрерывности: при любом физическом состоянии $\omega \in \mathfrak{S}$ и любом фиксированном $A \in \mathfrak{A}$ функция $g \rightarrow \omega(\alpha_g(A))$ непрерывна по g . При этом группа G унитарно-антиунитарно реализована, если существует непрерывное представление $g \rightarrow U_g$ группы G унитарными или антиунитарными операторами (в соответствии с тем, являются ли α_g алгебраическими автоморфизмами или антиавтоморфизмами) в гильбертовом пространстве H_∞ такое, что для всех $A \in \mathfrak{A}$, $g \in G$ имеет место $\alpha_g(A) = U_g A^{(*)} U_g^{-1}$, где $A^{(*)}$ есть A для унитарного U_g и A^* для антиунитарного U_g .

2.1.2. Конкретная реализация $\pi(\mathfrak{A})$

В этом пункте рассмотрим конкретную реализацию операторной алгебры $\mathfrak{A}$. Переход $\mathfrak{A} \Rightarrow \pi(\mathfrak{A})$ от $\mathfrak{A}$ к конкретной алгебре $\pi(\mathfrak{A})$ иногда называют «одеванием». Итак, основной наблюдаемой спектра материи является *энергия*, которой соответствует эрмитов оператор H . Пусть $G = SO_0(1, 3) \simeq SL(2, \mathbb{C})/\mathbb{Z}_2$ — группа фундаментальной симметрии, где $SO_0(1, 3)$ — группа Лоренца. В силу локального изоморфизма $SL(2, \mathbb{C}) \simeq SU(2) \otimes SU(2)$ будем рассматривать *универсальную накрывающую* фундаментальной симметрии $\tilde{G} \simeq SU(2) \otimes SU(2)$, т.е. вблизи единицы группы. Пусть оператор энергии H определён на сепара-

бильном гильбертовом пространстве H_∞ . Тогда все возможные значения энергии (состояния) являются собственными значениями оператора H . При этом, если $E_1 \neq E_2$ — собственные значения H , а $|\Phi_1\rangle$ и $|\Phi_2\rangle$ — принадлежащие им собственные векторы в пространстве H_∞ , то $\langle \Phi_1 | \Phi_2 \rangle = 0$. 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*. Далее, пусть X_l, Y_l — инфинитезимальные операторы комплексной оболочки групповой алгебры $\mathfrak{sl}(2, \mathbb{C})$ универсальной накрывающей $\tilde{G}$, $l = 1, 2, 3$. Как известно [12], оператор энергии H перестановочен со всеми операторами в H_∞ , представляющими алгебру Ли группы $\tilde{G}$. Рассмотрим произвольное собственное подпространство H_E оператора энергии H . Поскольку операторы X_l, Y_l и H коммутируют между собой, то, как следствие, для этих операторов можно построить общую систему собственных функций. Это значит, что подпространство H_E инвариантно относительно операторов X_l, Y_l (более того, операторы X_l, Y_l можно рассматривать *только на* H_E). Далее, пусть дано некоторое *локальное представление* группы $\tilde{G}$ операторами, действующими в H_∞ . 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 . Это позволяет отождествить подпространства H_E с симметрическими пространствами $\text{Sym}_{(k,r)}$ зацепляющихся представлений $\tau_{k/2, r/2}$ группы Лоренца и тем самым получить конкретную реализацию («одевание») операторной алгебры $\pi(\mathfrak{A}) \rightarrow \pi(H)$, где $\pi \equiv \tau_{k/2, r/2}$. На рис. 1 представлена система зацепляющихся представлений группы Лоренца. Отсюда следует, что каждое возможное значение энергии (энергетический уровень) является векторным состоянием вида (1):

$$\omega_\Phi(H) = \frac{\langle \Phi | \pi(H) \Phi \rangle}{\langle \Phi | \Phi \rangle} = \frac{\langle \Phi | \tau_{k/2, r/2}(H) \Phi \rangle}{\langle \Phi | \Phi \rangle}, \quad (2)$$

ассоциированным с представлением $\pi \equiv \tau_{k/2, r/2}$ и соответствующим ненулевому (циклическому) вектору $|\Phi\rangle \in H_\infty$. Аналогично, если ρ — матрица плотности в H_∞ , то

$$\omega_\rho(H) = \text{Tr}(\rho \tau_{k/2, r/2}(H))$$

– статистическая смесь векторных состояний (2).

Далее, в силу изоморфизма $\text{SL}(2, \mathbb{C}) \simeq \mathbf{Spin}_+(1, 3)$ будем рассматривать универсальную накрывающую $\tilde{G}$ как *спинорную группу*, что позволит в дальнейшем дополнительно ассоциировать с каждым циклическим вектором $|\Phi\rangle \in H_\infty$ *спинорную структуру* (в некотором смысле это будет второй слой в процессе «одевания» операторной алгебры). Спинтензорные представления группы $\tilde{G} \simeq \mathbf{Spin}_+(1, 3)$ образуют *субстрат* зацепляющихся представлений $\tau_{k/2, r/2}$ группы Лоренца, реализуемых в пространствах $\text{Sym}_{(k,r)} \subset \mathbb{S}_{2k+r}$, где $\mathbb{S}_{2k+r}$ — спинпространство. В свою очередь, как известно [13], спинпространство является минимальным левым идеалом алгебры Клиффорда $\mathcal{O}_{p,q}$, т.е. существует

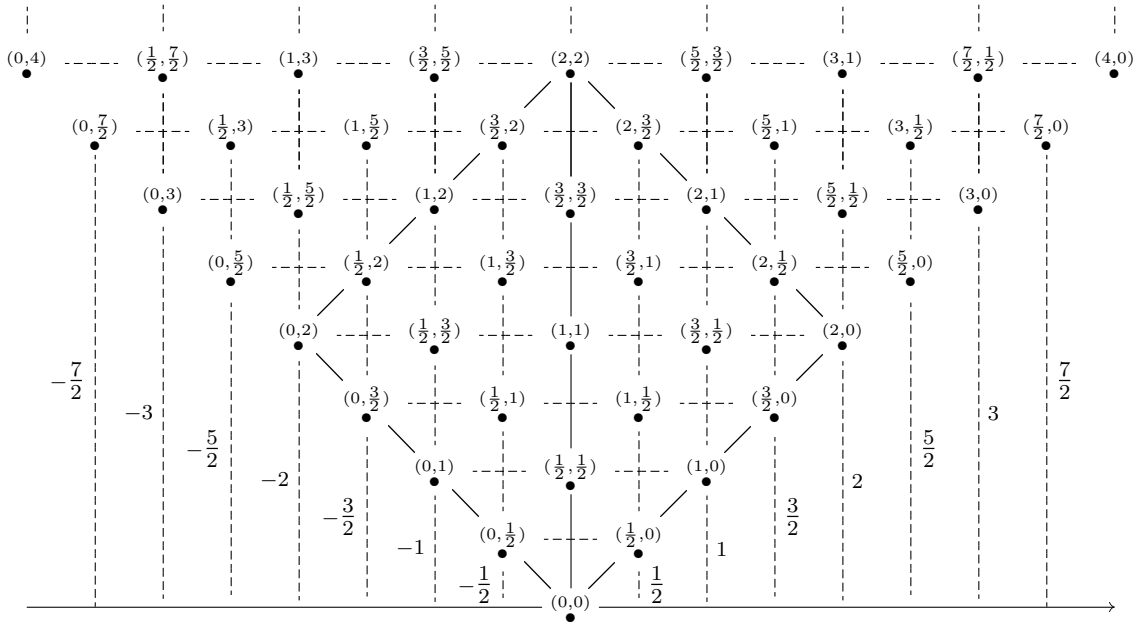


Рис. 1: Собственные подпространства $H_E \simeq \text{Sym}_{(k,r)}$ оператора энергии H . Каждое подпространство H_E является пространством неприводимого представления $\tau_{k/2,r/2}$ из системы зацепляющихся представлений группы Лоренца (группы фундаментальной симметрии). Сплошными линиями обозначена первая клетка спинорной шахматной доски второго порядка.

изоморфизм $S_{2^m}(\mathbb{K}) \simeq I_{p,q} = \mathcal{O}_{p,q}f$, где f — примитивный идемпотент алгебры $\mathcal{O}_{p,q}$, а $\mathbb{K} = f\mathcal{O}_{p,q}f$ — кольцо деления для $\mathcal{O}_{p,q}$, $m = (p+q)/2$. Комплексное спинпространство $S_{2^m}(\mathbb{C})$ является комплексификацией $\mathbb{C} \otimes I_{p,q}$ минимального левого идеала $I_{p,q}$ вещественной подалгебры $\mathcal{O}_{p,q}$. Так, $S_{2^{k+r}}$ является минимальным левым идеалом комплексной алгебры $\mathbb{C}_{2k} \otimes \mathbb{C}_{2r}^* \simeq \mathbb{C}_{2(k+r)}$ (более подробно см. [14, 15]). Определим систему *базисных циклических* векторов, наделённых комплексной спинорной структурой и соответствующих системе зацепляющихся представлений группы Лоренца:

$$|\mathbb{C}_0, \tau_{0,0}(H)\Phi\rangle;$$

$$|\mathbb{C}_2, \tau_{1/2,0}(H)\Phi\rangle, \quad |\mathbb{C}_2^*, \tau_{0,1/2}(H)\Phi\rangle;$$

$$|\mathbb{C}_2 \otimes \mathbb{C}_2, \tau_{1,0}(H)\Phi\rangle, \quad |\mathbb{C}_2 \otimes \mathbb{C}_2^*, \tau_{1/2,1/2}(H)\Phi\rangle, \quad |\mathbb{C}_2^* \otimes \mathbb{C}_2^*, \tau_{0,1}(H)\Phi\rangle;$$

$$|\mathbb{C}_2 \otimes \mathbb{C}_2 \otimes \mathbb{C}_2, \tau_{3/2,0}(H)\Phi\rangle, \quad |\mathbb{C}_2 \otimes \mathbb{C}_2 \otimes \mathbb{C}_2^*, \tau_{1,1/2}(H)\Phi\rangle, \quad |\mathbb{C}_2 \otimes \mathbb{C}_2^* \otimes \mathbb{C}_2^*, \tau_{1/2,1}(H)\Phi\rangle, \quad |\mathbb{C}_2^* \otimes \mathbb{C}_2^* \otimes \mathbb{C}_2^*, \tau_{0,3/2}(H)\Phi\rangle;$$

.....

Следовательно, согласно конструкции ГНС имеем комплексные векторные состояния вида

$$\omega_{\Phi}^c(H) = \frac{\langle \Phi | \mathbb{C}_{2(k+r)}, \tau_{k/2, r/2}(H) \Phi \rangle}{\langle \Phi | \Phi \rangle}, \quad (3)$$

ассоциированные с комплексными представлениями $\tau_{k/2, r/2}(H)$ (т.е. снабжёнными комплексной спинорной структурой) и соответствующими циклическим векторам $|\Phi\rangle \in \mathcal{H}_{\infty}$.

Как известно, в лагранжевом формализме стандартной квантовой теории поля *заряженным частицам* соответствуют *комплексные поля*. В нашем случае чистые состояния вида (3) соответствуют *заряженным состояниям*. При этом знак заряда меняется на обратный под действием псевдоавтоморфизма $\mathcal{A} \rightarrow \overline{\mathcal{A}}$ комплексной спинорной структуры (более подробно см. [16–18]). Продолжая аналогию с лагранжевым формализмом, в котором *нейтральным частицам* соответствуют *вещественные поля*, введём векторные состояния следующего вида:

$$\omega_{\Phi}^r(H) = \frac{\langle \Phi | \mathcal{O}_{p,q}, \tau_{k/2, r/2}(H) \Phi \rangle}{\langle \Phi | \Phi \rangle}. \quad (4)$$

Состояния (4) ассоциированы с вещественными представлениями $\tau_{k/2, r/2}(H)$, т.е. с представлениями, снабжёнными *вещественной спинорной структурой*, где $\mathcal{O}_{p,q}$ — вещественная подалгебра алгебры $\mathbb{C}_{2(k+r)}$. Состояния вида (4) соответствуют *нейтральным состояниям*. Поскольку вещественная спинорная структура появляется в результате редукции $\mathbb{C}_{2(k+r)} \rightarrow \mathcal{O}_{p,q}$, то, как следствие, *зарядовое сопряжение* C (псевдоавтоморфизм $\mathcal{A} \rightarrow \overline{\mathcal{A}}$) для алгебр $\mathcal{O}_{p,q}$ над вещественным числовым полем $\mathbb{F} = \mathbb{R}$ и кватернионным кольцом деления $\mathbb{K} \simeq \mathbb{H}$ (типы $p - q \equiv 4, 6 \pmod{8}$) редуцируется к *обмену частица-античастица* C' (см. [16–18]). Как известно, существуют два класса нейтральных частиц: 1) частицы, имеющие античастицы, такие как нейтроны, нейтрино и т.д.; 2) частицы, совпадающие со своими античастицами (например, фотоны, π^0 -мезоны и т.д.), т.е. так называемые *истинно нейтральные частицы*. Первый класс описывается нейтральными состояниями $\omega_{\Phi}^r(H)$ с алгебрами $\mathcal{O}_{p,q}$ 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 $p - q \equiv 5 \pmod{8}$). Для описания второго класса нейтральных частиц введём *истинно нейтральные состояния* $\omega_{\Phi}^{r0}(H)$ с алгебрами $\mathcal{O}_{p,q}$ над числовым полем $\mathbb{F} = \mathbb{R}$ и вещественными кольцами деления $\mathbb{K} \simeq \mathbb{R}$ и $\mathbb{K} \simeq \mathbb{R} \oplus \mathbb{R}$ (типы $p - q \equiv 0, 2 \pmod{8}$ и $p - q \equiv 1 \pmod{8}$). В случае состояний $\omega_{\Phi}^{r0}(H)$ псевдоавтоморфизм $\mathcal{A} \rightarrow \overline{\mathcal{A}}$ редуцируется к тождественному преобразованию (частица совпадает со своей античастицей).

Далее, если ρ — матрица плотности в $\mathcal{H}_{\infty}$, то $\omega_{\rho}^c(H)$, $\omega_{\rho}^r(H)$ и $\omega_{\rho}^{r0}(H)$ — статистические смеси заряженных, нейтральных и истинно нейтральных состояний.

Прежде чем приступить к построению физического гильбертова пространства, рассмотрим более подробно структуру состояний вида (2)-(4). Состояния (2)-(4) представляют собой уровни спектра материи, т.е. актуализированные (локализованные) состояния квантовых микрообъектов («элементарных частиц»). Каждое из состояний вида (2)-(4) обладает следующими характери-

ками (свойствами): энергия (масса), спин и заряд (первые два слоя «одевания» операторной алгебры). На этом уровне описания *состояние* приобретает первичное значение (в духе трактовки Биркгофа-фон Неймана-Макки), а *наблюдаемые* характеристики (энергия, спин, заряд, ...) являются свойствами состояния. Дальнейшее «одевание» операторной алгебры приводит к введению новых свойств (характеристик) состояния. Например, дискретные симметрии (инверсия пространства P , обращение времени T , зарядовое сопряжение C и их комбинации) появляются как автоморфизмы спинорной структуры, ассоциированной с каждым состоянием вида (2)-(4) [19–22]. Далее, фрактальная структура спектра материи определяется периодичностью Картана-Ботта спинорной структуры⁶. Однако подробное рассмотрение этих характеристик состояний выходит за рамки данной статьи.

2.2. Физическое гильбертово пространство

Множество всех чистых состояний $\omega_\Phi(H)$, определённых согласно конструкции ГНС равенством (2), при выполнении условия $\omega_\Phi(H) \geq 0$ образует *физическое гильбертово пространство*

$$\mathbf{H}_{\text{phys}} = \mathbf{H}^S \otimes \mathbf{H}^Q \otimes \mathbf{H}_\infty.$$

Легко проверить, что при этом выполняются аксиомы сложения, умножения и скалярного (внутреннего) произведения для векторов $\omega_\Phi(H) \rightarrow |\Psi\rangle \in \mathbf{H}_{\text{phys}}$. Будем предполагать, что определённое таким образом гильбертово пространство является *несепарабельным*, т.е. в общем случае в $\mathbf{H}_{\text{phys}}$ не выполняется аксиома сепарабельности⁷.

Пространство $\mathbf{H}_{\text{phys}}$, будучи вторым членом пары $(\mathbf{H}_\infty, \mathbf{H}_{\text{phys}})$, задающей двухуровневую структуру гильбертова пространства элементарной частицы, описывает спиновые и зарядовые степени свободы этой частицы. В соответствии с зарядовыми степенями свободы выделим три *базовых подпространства* в $\mathbf{H}_{\text{phys}}$.

1) *Подпространство заряженных состояний* $\mathbf{H}_{\text{phys}}^\pm = \mathbf{H}^S \otimes \mathbf{H}^\pm \otimes \mathbf{H}_\infty$.

⁶Фрактальная структура энергетического спектра образуется действием группы Брауэра-Уолла $BW_{\mathbb{R}} \simeq \mathbb{Z}_8$. Действие группы $BW_{\mathbb{R}} \simeq \mathbb{Z}_8$ связывает различные типы вещественных алгебр Клиффорда, которые образуют вещественную подструктуру спинтензорного субстрата. Циклическое действие группы $BW_{\mathbb{R}} \simeq \mathbb{Z}_8$ генерирует фрактальную структуру на спинтензорном субстрате. Эта структура аналогична ковру Серпинского с фрактальной размерностью (размерность Безиковича-Хаусдорфа) $D = \ln 63 / \ln 8 \approx 1,9924$ (см. [23, 24]). Далее, в силу отображения $\mathcal{C}_{p,q} \xrightarrow{\gamma} \text{End}_{\mathbb{K}}(\mathbb{S}), \quad u \rightarrow \gamma(u), \quad \gamma(u)s = us$, где $\mathbb{S} = \mathbb{S}_{2r}(\mathbb{K}) \simeq I_{p,q} = \mathcal{C}_{p,q}f$ — *вещественное спинпространство*, $s = s^{\alpha_1 \alpha_2 \dots \alpha_r} \in \mathbb{S}_{2r}$, фрактальная структура спинтензорного субстрата переносится на представления группы $\mathbf{Spin}_+(1, 3) \simeq \text{SL}(2, \mathbb{C})$.

⁷Следует отметить, что первым к идее введения в квантовую теорию несепарабельных гильбертовых пространств пришёл Дирак. Дирак показал [25], что в квантовой теории поля в силу интенсивности взаимодействий шредингеровский вектор состояния будет выбит из сепарабельного гильбертова пространства за кратчайшее время. Как следствие этого не существует решения уравнения Шредингера и в целом шредингеровская картина, на которой базируется вся аксиоматика локальной квантовой теории поля, является неработоспособной.

- 2) Подпространство нейтральных состояний $\mathbf{H}_{\text{phys}}^0 = \mathbf{H}^S \otimes \mathbf{H}^0 \otimes \mathbf{H}_\infty$.
- 3) Подпространство истинно нейтральных состояний $\mathbf{H}_{\text{phys}}^{\bar{0}} = \mathbf{H}^S \otimes \mathbf{H}^{\bar{0}} \otimes \mathbf{H}_\infty$.
- Базисные векторы $|\Psi\rangle \in \mathbf{H}_{\text{phys}}^\pm$ образованы состояниями $\omega_\Phi^c(H)$ (см. (3)), соответственно $|\Psi\rangle \in \mathbf{H}_{\text{phys}}^0$ и $|\Psi\rangle \in \mathbf{H}_{\text{phys}}^{\bar{0}}$ образуют состояния $\omega_\rho^r(H)$ и $\omega_\rho^{r_0}(H)$.

Следуя трактовке Биркгофа-фон Неймана-Макки, будем считать, что на уровне физического гильбертова пространства $\mathbf{H}_{\text{phys}}$ исходным (первичным) понятием является *состояние (луч)*. Пусть $|\Psi\rangle$ — вектор состояния в пространстве $\mathbf{H}_{\text{phys}}$, тогда $\Psi = e^{i\alpha} |\Psi\rangle$, где α пробегает все вещественные числа и $\sqrt{\langle \Psi | \Psi \rangle} = 1$, называется *единичным лучом*. Следовательно, единичный луч Ψ — это совокупность базисных векторов состояния $\{\lambda |\Psi\rangle\}$, $\lambda = e^{i\alpha}$, $|\Psi\rangle \in \mathbf{H}_{\text{phys}}$. Величины, связанные с наблюдаемыми эффектами, являются абсолютными значениями полубилинейной формы $|\langle \Psi_1 | \Psi_2 \rangle|^2$, не зависящими от параметров λ , характеризующих луч. Таким образом, *пространство лучей* есть фактор-пространство $\hat{H} = \mathbf{H}_{\text{phys}}/S^1$, т.е. проективное пространство одномерных подпространств из $\mathbf{H}_{\text{phys}}$. Все состояния физической (квантовой) системы (в нашем случае элементарной частицы) описываются единичными лучами. Предположим, что основное соответствие между физическими состояниями и элементами пространства $\mathbf{H}_{\text{phys}}$ включает *принцип суперпозиции* квантовой теории, т.е. существует набор базисных состояний, таких, что произвольные состояния могут быть построены из них при помощи линейных суперпозиций. Дадим общее определение элементарной частицы. *Элементарная частица (единичная квантовая микросистема) является суперпозицией векторов состояния в физическом гильбертовом пространстве $\mathbf{H}_{\text{phys}}$.*

2.2.1. Групповое действие в $\mathbf{H}_{\text{phys}}$

Предположим, что одна и та же квантовая система может быть описана двумя различными способами в одном и том же подпространстве $\mathbf{H}_{\text{phys}}^\pm$ ($\mathbf{H}_{\text{phys}}^0$ или $\mathbf{H}_{\text{phys}}^{\bar{0}}$) пространства $\mathbf{H}_{\text{phys}}$ один раз лучами $\Psi_1, \Psi_2, \dots$ и другой раз лучами $\Psi'_1, \Psi'_2, \dots$. Это значит, что имеет место симметрия квантовой системы, когда одно и то же физическое состояние описывается при помощи луча Ψ_1 в первом случае, а во втором — при помощи Ψ'_1 , так что вероятности переходов одинаковы. Следовательно, имеем отображение $\hat{T}$ между лучами Ψ_1 и Ψ'_1 . Так как инвариантны только абсолютные значения, преобразование $\hat{T}$ в гильбертовом пространстве $\mathbf{H}_{\text{phys}}$ может быть унитарным или антиунитарным. Обе эти возможности реализуются в случае подпространства $\mathbf{H}_{\text{phys}}^\pm$, векторы состояния которого наделены комплексной спинорной структурой, поскольку комплексное поле имеет два (и только два) автоморфизма, сохраняющих абсолютные значения: тождественный автоморфизм и комплексное сопряжение. Также обе эти возможности реализуются в подпространстве $\mathbf{H}_{\text{phys}}^0$, поскольку в этом случае векторы состояния наделены вещественной спинорной структурой с кватернионным кольцом деления. В случае подпространства $\mathbf{H}_{\text{phys}}^{\bar{0}}$ имеем только унитарные преобразования $\hat{T}$, поскольку вещественная спинорная структура с вещественным кольцом деления допускает только тождественный автоморфизм.

Пусть $|\Psi_1\rangle, |\Psi_2\rangle, \dots$ — единичные векторы, выбранные из первой совокупности лучей $\Psi_1, \Psi_2, \dots$, и пусть $|\Psi'_1\rangle, |\Psi'_2\rangle, \dots$ — единичные векторы, выбранные из второй совокупности лучей $\Psi'_1, \Psi'_2, \dots$, так что соответствие $|\Psi_1\rangle \leftrightarrow |\Psi'_1\rangle, |\Psi_2\rangle \leftrightarrow |\Psi'_2\rangle, \dots$ является унитарным или антиунитарным. Первый набор соответствует состояниям $\{\omega\}$, а второй набор соответствует преобразованным состояниям $\{g\omega\}$. Выберем векторы $|\Psi_1\rangle \in \Psi_1, |\Psi_2\rangle \in \Psi_2, \dots$ и $|\Psi'_1\rangle \in \Psi'_1, |\Psi'_2\rangle \in \Psi'_2, \dots$ так что

$$|\Psi'_1\rangle = T_g |\Psi_1\rangle, \quad |\Psi'_2\rangle = T_g |\Psi_2\rangle, \quad \dots \quad (5)$$

Это значит, что если $|\Psi_1\rangle$ — вектор, ассоциированный с лучом Ψ_1 , то $T_g |\Psi_1\rangle$ — вектор, ассоциированный с лучом Ψ'_1 . Если существуют два оператора T_g и $T_{g'}$ со свойством (5), то они могут отличаться только постоянным множителем, по модулю равным единице. Следовательно,

$$T_{gg'} = \phi(g, g') T_g T_{g'}, \quad (6)$$

где $\phi(g, g')$ — фазовый множитель. Представления вида (6) называются *лучевыми (проективными) представлениями*. Это означает также, что имеет место соответствие между физическими состояниями и лучами гильбертова пространства $\mathbf{H}_{\text{phys}}$. Отсюда следует, что лучевое представление T топологической группы G является непрерывным гомоморфизмом $T : G \rightarrow L(\hat{H})$, где $L(\hat{H})$ — множество линейных операторов в проективном пространстве $\hat{H}$ с фактортопологией относительно отображения $\hat{H} \rightarrow \mathbf{H}_{\text{phys}}$, т.е. $|\Psi\rangle \rightarrow \Psi$. Однако при $\phi(g, g') \neq 1$ мы не можем непосредственно применить математическую теорию обычных представлений (в силу неоднозначности фазы). С целью обойти это препятствие построим более широкую группу $\mathcal{E}$ таким образом, что обычные представления группы $\mathcal{E}$ дадут все неэквивалентные лучевые представления (6) группы G . Эта задача может быть решена посредством *подъёма* проективных представлений группы G в обычные представления группы $\mathcal{E}$. Пусть $\mathcal{K}$ — абелева группа, порождаемая умножением неэквивалентных фаз $\phi(g, g')$, удовлетворяющих условию

$$\phi(g, g')\phi(gg', g'') = \phi(g', g'')\phi(g, g'g'').$$

Рассмотрим пары (ϕ, x) , где $\phi \in \mathcal{K}$, $x \in G$, в частности, $\mathcal{K} = \{(\phi, e)\}$, $G = \{(e, x)\}$. Пары (ϕ, x) образуют группу с законом умножения типа полупрямого произведения: $(\phi_1, x_1)(\phi_2, x_2) = (\phi_1\phi(x_1, x_2)\phi_2, x_1x_2)$. Группа $\mathcal{E} = \{(\phi, x)\}$ называется *центральной расширением* группы G посредством группы $\mathcal{K}$. Векторные представления группы $\mathcal{E}$ содержат все лучевые представления группы G . Следовательно, группа симметрии G физической системы индуцирует представление T обратимых отображений пространства $\mathbf{H}_{\text{phys}}$ на себя, которое является унитарным или антиунитарным и является представлением центрального расширения $\mathcal{E}$ группы G либо группы, накрывающей группу G .

На уровне физического гильбертова пространства $\mathbf{H}_{\text{phys}}$ под группой симметрии G будем понимать группу из последовательности унитарных унимодулярных групп: $SU_T(2)$ (группа изоспина), $SU(3)$, $\dots$, $SU(N)$, $\dots$ (группы так

называемых «внутренних» симметрий). Согласно Гейзенбергу, группы $SU(N)$ определяют *динамические (вторичные) симметрии*. Таким образом, в соответствии с двухуровневой структурой гильбертова пространства элементарной частицы (единичной квантовой микросистемы), задаваемой парой $(\mathbf{H}_\infty, \mathbf{H}_{\text{phys}})$, все множество групп симметрий G разбивается на два класса: 1) *группы фундаментальных (первичных) симметрий* G_f , участвующих в образовании векторов состояния квантовой микросистемы; 2) *группы динамических (вторичных) симметрий* G_d , описывающие приближенные симметрии между векторами состояния квантовой системы⁸.

2.2.2. Принцип редукции

Итак, динамические симметрии G_d связывают между собой различные состояния квантовой системы (векторы состояния $|\Psi\rangle \in \mathbf{H}_{\text{phys}}$). Симметрия G_d квантовой системы может быть представлена как квантовый *переход* между её состояниями (уровнями спектра материи). Например, если взять барионы и пренебречь лептонами и мезонами, то реакция $N \rightarrow Pe^- \nu$, в которой нейтрон распадается на протон, электрон и нейтрино, может рассматриваться как переход $N \rightarrow P$, реакция $\Sigma^- \rightarrow N\pi^-$ как переход $\Sigma \rightarrow N$ и т.д. При этом предполагается, что представляющие операторы комплексной оболочки алгебры Ли группы G_d осуществляют все возможные квантовые переходы системы. Например, если $G_d = SU(3)$, то операторы Окубо $\mathbf{A}_\tau^\sigma$ переводят друг в друга состояния (частицы) октета. Естественно считать, что операторы группы G_d или её подгруппы соединяют между собой родственные состояния. Цепочка вложенных подгрупп приводит к иерархической классификации состояний.

Динамическая симметрия задаётся цепочкой вложенных в друг друга групп Ли:

$$G = G_0 \supset G_1 \supset G_2 \supset \dots \supset G_k.$$

Система с данной динамической симметрией задаётся неприводимым унитарным представлением $\mathfrak{P}$ группы G в физическом гильбертовом пространстве $\mathbf{H}_{\text{phys}}$. Редукция G/G_1 представления $\mathfrak{P}$ группы G по подгруппе G_1 приводит к разложению $\mathfrak{P}$ в ортогональную сумму неприводимых представлений $\mathfrak{P}_i^{(1)}$

⁸Согласно Вигнеру [26], квантовая система, описываемая неприводимым унитарным *представлением* группы Пуанкаре $\mathcal{P}$, называется элементарной частицей. С другой стороны, в согласии с $SU(3)$ -теорией элементарная частица описывается *вектором* неприводимого представления группы $SU(3)$. Например, в так называемом «восьмеричном пути» Гелл-Манна–Неемана [27] адроны (барионы и мезоны) представлены векторами восьмимерного регулярного представления $\text{Sym}_{(1,1)}^0$ группы $SU(3)$. Таким образом, имеем две взаимно исключающие друг друга интерпретации элементарной частицы: как *представления* группы $\mathcal{P}$ и как *вектора* представления группы $SU(3)$. Это противопоставление исчезает, если принять, что все «элементарные частицы» суть локализованные состояния (уровни) спектра материи, реализованного посредством физического гильбертова пространства $\mathbf{H}_{\text{phys}}$, векторы состояния которого задаются циклическими представлениями операторной алгебры (оператора энергии H). На этом уровне описания группа $SU(3)$, задаваемая в $\mathbf{H}_{\text{phys}}$ посредством центрального расширения, описывает динамические (приближенные) симметрии между различными состояниями (как будет показано ниже, между состояниями из различных когерентных подпространств в $\mathbf{H}_{\text{phys}}$).

подгруппы G_1 :

$$\mathfrak{P} = \mathfrak{P}_1^{(1)} \oplus \mathfrak{P}_2^{(1)} \oplus \dots \oplus \mathfrak{P}_i^{(1)} \oplus \dots$$

В свою очередь редукция G_1/G_2 представления группы G_1 по подгруппе G_2 приводит к разложению представлений $\mathfrak{P}_i^{(1)}$ на неприводимые представления $\mathfrak{P}_{ij}^{(2)}$ группы G_2 :

$$\mathfrak{P}_i^{(1)} = \mathfrak{P}_{i1}^{(2)} \oplus \mathfrak{P}_{i2}^{(2)} \oplus \dots \oplus \mathfrak{P}_{ij}^{(2)} \oplus \dots$$

и т.д.⁹

3. Когерентные подпространства

В этом параграфе продолжим исследование структуры физического гильбертова пространства $\mathbf{H}_{\text{phys}}$. Как известно [30], не все единичные лучи являются физически реализуемыми. Существуют физические ограничения (*правила суперотбора*) на выполнение принципа суперпозиции. В 1952 г. Вигнер, Уайтман и Вик [30] показали, что существование правил суперотбора связано с измеримостью относительной фазы суперпозиции. Это означает, что чистое состояние не может быть реализовано в виде суперпозиции некоторых состояний, например, не существует чистого состояния (когерентной суперпозиции) из бозонного $|\Psi_b\rangle$ и фермионного $|\Psi_f\rangle$ состояний (правило суперотбора по спину). Однако, если в $\mathbf{H}_{\text{phys}}$ определена матрица плотности ρ , то суперпозиция $|\Psi_b\rangle + |\Psi_f\rangle$ определяет смешанное состояние.

Теорема 1. *Физическое гильбертово пространство $\mathbf{H}_{\text{phys}}$ разлагается в прямую сумму (ненулевых) когерентных подпространств*

$$\mathbf{H}_{\text{phys}} = \mathbf{H}_{\text{phys}}^{\pm} \bigoplus \mathbf{H}_{\text{phys}}^0 \bigoplus \mathbf{H}_{\text{phys}}^{\bar{0}}, \quad (7)$$

где

$$\mathbf{H}_{\text{phys}}^Q = \bigoplus_{s=-|l-i|}^{|l-i|} \mathbf{H}^{2|s|+1} \otimes \mathbf{H}^Q \otimes \mathbf{H}_{\infty}, \quad Q = \{\pm, 0, \bar{0}\}. \quad (8)$$

При этом принцип суперпозиции имеет место в ограниченной форме, т.е. в пределах когерентных подпространств. Ненулевая линейная комбинация векторов чистых состояний есть вектор чистого состояния при условии, что все исходные векторы лежат в одном и том же когерентном подпространстве. Суперпозиция векторов чистых состояний из различных когерентных подпространств определяет смешанное состояние.

⁹Например, один из базовых супермультиплетов SU(3)-теории (барионный октет $F_{1/2}$), основанный на восьмимерном регулярном представлении $\text{Sym}_{(1,1)}^0$, допускает следующую SU(3)/SU(2)-редукцию на изотопические мультиплеты подгруппы SU(2): $\text{Sym}_{(1,1)}^0 = \Phi_3 \oplus \Phi_2 \oplus \Phi_2^* \oplus \Phi_0$, где Φ_3 — триплет, Φ_2 и Φ_2^* — дублеты, Φ_0 — синглет. Аналогично, для гипермультиплетов SU(6)-теории (56-плет барионов и 35-плет мезонов) имеют место SU(6)/SU(3)- и SU(6)/SU(4)-редукции, где SU(4) — подгруппа Вигнера [28, 29].

Доказательство. Исходным пунктом доказательства является соответствие $\omega_{\Phi}(H) \leftrightarrow |\Psi\rangle$ между состояниями операторной алгебры и базисными векторами пространства $\mathbf{H}_{\text{phys}}$. Как было показано в п. 2.1, множество всех чистых состояний $PS(\mathfrak{A})$ операторной алгебры $\mathfrak{A}$ разбивается на попарно непересекающиеся и взаимно ортогональные секторы в силу отношения эквивалентности $\omega_1 \sim \omega_2$, которые совпадают с классами эквивалентности в $PS(\mathfrak{A})$ (по сути сектор является алгебраическим двойником когерентного подпространства). Далее, пусть некоторое множество векторов в физическом гильбертовом пространстве $\mathbf{H}_{\text{phys}}$, содержащее чистые состояния алгебры наблюдаемых $\mathfrak{A}$, образует тотальное множество в $\mathbf{H}_{\text{phys}}$, т.е. такое множество X , замыкание линейной оболочки которого всюду плотно в $\mathbf{H}_{\text{phys}}$. Тогда X нельзя представить в виде объединения двух (или более) непустых взаимно ортогональных подмножеств. Будем говорить, что векторы $|\Psi_1\rangle, |\Psi_2\rangle \in X$ связаны соотношением $|\Psi_1\rangle \sim |\Psi_2\rangle$, если $|\Psi_1\rangle$ и $|\Psi_2\rangle$ принадлежат линейной оболочке из X . Легко видеть, что соотношение $|\Psi_1\rangle \sim |\Psi_2\rangle$ индуцируется соотношением эквивалентности $\omega_1 \sim \omega_2$ из $PS(\mathfrak{A})$. Следовательно, отношение $|\Psi_1\rangle \sim |\Psi_2\rangle$ есть отношение эквивалентности и классы эквивалентности в X образуют разбиение X на взаимно ортогональные системы X_{ν} , где $\{\nu\} = N$ — некоторое семейство индексов. Взяв теперь в качестве $\mathbf{H}_{\text{phys}}^{\nu}$ замкнутую линейную оболочку множества X_{ν} , мы приходим к искомому разложению $\mathbf{H}_{\text{phys}}$ в прямую сумму взаимно ортогональных подпространств $\mathbf{H}_{\text{phys}}^{\nu}$:

$$\mathbf{H}_{\text{phys}} = \bigoplus_{\nu \in N} \mathbf{H}_{\text{phys}}^{\nu}.$$

Таким образом, чистые состояния находятся во взаимно однозначном соответствии с единичными лучами в $\cup_{\nu} \mathbf{H}_{\text{phys}}^{\nu}$. Отсюда следует ограниченная форма принципа суперпозиции (а именно, в пределах подпространств $\mathbf{H}_{\text{phys}}^{\nu}$). Очевидно, что три базовых подпространства $\mathbf{H}_{\text{phys}}^{\pm}$, $\mathbf{H}_{\text{phys}}^0$ и $\mathbf{H}_{\text{phys}}^{\bar{0}}$ являются когерентными подпространствами (относительно заряда) исходного физического пространства $\mathbf{H}_{\text{phys}}$, отсюда следует разложение (7). Дальнейшее разложение $\mathbf{H}_{\text{phys}}$ на когерентные подпространства осуществляется относительно спина, т.е. относительно составляющей $\mathbf{H}^S$ в $\mathbf{H}^S \otimes \mathbf{H}^Q \otimes \mathbf{H}_{\infty}$ (формула (8)). ■

Пример. Согласно современным данным, нейтрино является суперпозицией трёх массовых нейтринных состояний: электронного, мюонного и τ -лептонного нейтрино. Все три состояния лежат в когерентных подпространствах $\mathbf{H}^2 \otimes \mathbf{H}^0 \otimes \mathbf{H}_{\infty}$, которые принадлежат линии спина 1/2 базового подпространства $\mathbf{H}_{\text{phys}}^0$ (подпространство нейтральных состояний). Все три нейтринных уровня спектра материи, принадлежа одному и тому же когерентному подпространству, тем не менее, отличаются друг от друга значением энергии, что характеризуется различным расположением соответствующих представлений на линии спина 1/2.

3.1. Калибровочные симметрии

Пусть в $\mathbf{H}_{\text{phys}}$ определена (посредством центрального расширения) связанная компактная n -параметрическая абелева группа $G = U(1)^n \equiv U(1) \times \dots \times U(1)$ (*калибровочная группа*). Произвольный элемент этой группы представляется набором n фазовых множителей:

$$g(s_1, \dots, s_n) \equiv (e^{i\alpha_1}, \dots, e^{i\alpha_n}), \quad 0 \leq \alpha_j < 2\pi.$$

В пространстве $\mathbf{H}_{\text{phys}}$ зададим точное унитарное представление $\mathfrak{U}$ группы G . Соответствующие калибровочные преобразования в $\mathbf{H}_{\text{phys}}$ имеют вид

$$\mathfrak{U}(g) = s_1^{Q_1} \dots s_n^{Q_n} \equiv e^{i(\alpha_1 Q_1 + \dots + \alpha_n Q_n)}.$$

Генераторы $Q_1, \dots, Q_n$ калибровочных преобразований являются взаимно коммутирующими самосопряжёнными операторами с целочисленным спектром (они называются *зарядами*, соответствующими данной калибровочной группе). Тогда $\mathbf{H}_{\text{phys}}$ разлагается в прямую сумму

$$\mathbf{H}_{\text{phys}} = \bigoplus_{q_1, \dots, q_n \in \mathbb{Z}} \mathbf{H}_{\text{phys}}(q_1, \dots, q_n) \quad (9)$$

соответствующих спектральных подпространств, состоящих из всех векторов $|\Psi\rangle$ таких, что $(Q_j - q_j)|\Psi\rangle = 0$, $j = 1, \dots, n$. При этом произвольный ненулевой вектор $|\Psi\rangle \in \mathbf{H}_{\text{phys}}$ определяет чистое состояние алгебры наблюдаемых в точности тогда, когда он является собственным вектором для всех зарядов. Таким образом, в $\mathbf{H}_{\text{phys}}$ задаются *стандартные правила суперотбора*, причём (9) есть разложение $\mathbf{H}_{\text{phys}}$ в прямую сумму когерентных подпространств $\mathbf{H}_{\text{phys}}(q_1, \dots, q_n)$ ¹⁰. Согласно современному состоянию в физике элементарных частиц правила суперотбора могут быть довольно полно описаны электрическим Q ($= Q_1$), барионным B ($= Q_2$) и лептонным L ($= Q_3$) зарядами, так что разложение на когерентные подпространства имеет вид

$$\mathbf{H}_{\text{phys}} = \bigoplus_{q, b, \ell \in \mathbb{Z}} \mathbf{H}_{\text{phys}}(q, b, \ell).$$

¹⁰В настоящее время популярна гипотеза о существовании строгой (калибровочной) *неабелевой* SU(3)-симметрии, называемой *цветной* SU(3). Рассмотрение такой симметрии призвано объяснить, почему с точки зрения модели кварков все наблюдаемые адроны являются связанными состояниями либо трёх кварков, либо кварка и антикварка. Ненарушенность «цветной симметрии» ответственна по этим представлениям за то, что свободные кварки и (цветные) глюоны не наблюдаются (конфайнмент). Здесь по умолчанию предполагается, что кварки не являются исключительно математическими объектами (например, как спиноры), а обладают некоторой «реальностью», хотя и ненаблюдаемой в опыте. Однако, прежде всего, кварки математически — это векторы фундаментального представления группы SU(3), а кварковая модель является производной конструкцией от динамической (приближенной) SU(3)-симметрии (по так называемому *аромату*). Более того, все основные факты SU(3)- и SU(6)-теорий, касающиеся систематизации адронных спектров, могут быть получены без использования кварковой гипотезы (см. [28]). Кроме того, следует добавить, что открытие экзотических барионов (пентакварков) поставило под сомнение фундаментальность цветной SU(3)-симметрии.

Разложение физического гильбертова пространства $\mathbf{H}_{\text{phys}}$ на когерентные подпространства относительно заряда и спина даётся формулой (7) (теорема 1). С целью описать весь спектр наблюдаемых состояний (уровней спектра материи) введём 2-параметрическую калибровочную группу $G = U(1)^2 \equiv U(1) \times U(1)$ относительно барионного B и лептонного L зарядов. Тогда разложение $\mathbf{H}_{\text{phys}}$ на когерентные подпространства имеет вид:

$$\mathbf{H}_{\text{phys}} = \bigoplus_{b, \ell \in \mathbb{Z}} \left[\mathbf{H}_{\text{phys}}^{\pm}(b, \ell) \oplus \mathbf{H}_{\text{phys}}^0(b, \ell) \oplus \mathbf{H}_{\text{phys}}^{\bar{0}}(b, \ell) \right], \quad (10)$$

где

$$\mathbf{H}_{\text{phys}}^Q(b, \ell) = \bigoplus_{s=-|l-i|}^{|l-i|} \mathbf{H}^{2|s|+1} \otimes \mathbf{H}^Q(b, \ell) \otimes \mathbf{H}_{\infty}, \quad Q = \{\pm, 0, \bar{0}\}.$$

Разложение (10) позволяет охватить практически весь наблюдаемый спектр состояний (см. Particle Data Group). Прежде всего, спектр материи подразделяется на три сектора: лептонный, мезонный и барионный секторы. Лептонный сектор включает в себя заряженные лептоны: электрон e^- , мюон μ^- , τ^- -лептон (и их античастицы). Все заряженные лептоны принадлежат когерентным подпространствам вида $\mathbf{H}_{\text{phys}}^{\pm}(0, \ell)$. Нейтральные лептоны (три сорта нейтрино) принадлежат когерентному подпространству $\mathbf{H}_{\text{phys}}^0(0, \ell)$. Лептонный сектор также включает в себя одно истинно нейтральное состояние: фотон γ (подпространство $\mathbf{H}_{\text{phys}}^{\bar{0}}(0, \ell)$). В отличие от лептонного сектора, мезонный и барионный секторы (адронный сектор в совокупности) включают в себя огромное множество состояний (частиц). Мезонный сектор разбивается относительно заряда на три множества когерентных подпространств. Во-первых, заряженные мезоны ($\pi^{\pm}$ (пионы), $K^{\pm}$ (каоны), $\rho^{\pm}$, ...) принадлежат когерентным подпространствам вида $\mathbf{H}_{\text{phys}}^{\pm}(0, 0)$ с целым спином (все мезоны имеют целый спин). Далее, нейтральные мезоны (K^0 , $\bar{K}^0$, ...) принадлежат подпространствам $\mathbf{H}_{\text{phys}}^0(0, 0)$ целого спина. В свою очередь, истинно нейтральные мезоны (π^0 , η , φ , ρ^0 , ...) являются состояниями, принадлежащими когерентному подпространству $\mathbf{H}_{\text{phys}}^{\bar{0}}(0, 0)$. Барионный сектор разбивается относительно заряда на два множества когерентных подпространств: заряженные барионы (p (протон), $\Sigma^{\pm}$, $\Xi^{\pm}$, ...) образуют подпространства вида $\mathbf{H}_{\text{phys}}^{\pm}(b, 0)$ с полуцелым спином (все барионы имеют полуцелый спин); нейтральные барионы (n (нейтрон), Σ^0 , Ξ^0 , ...) являются состояниями из когерентных подпространств вида $\mathbf{H}_{\text{phys}}^0(b, 0)$ полуцелого спина. Истинно нейтральные барионы (майорановские фермионы) до сих пор не обнаружены в природе.

В заключении следует отметить, что добавление калибровочных симметрий приводит к тройственному разделению симметрий спектра материи (не частиц, как это имеет место в стандартной модели, а симметрий). А именно, фундаментальные симметрии G_f участвуют в образовании чистых состояний (лучей) квантовой системы и соответствующих когерентных подпространств в $\mathbf{H}_{\text{phys}}$, динамические симметрии G_d описывают переходы между состояниями из раз-

личных когерентных подпространств, калибровочные симметрии G_g связывают чистые состояния внутри когерентных подпространств.

ЛИТЕРАТУРА

1. Гейзенберг В. Природа элементарных частиц // УФН. 1977. Т. 121. С. 657–668.
2. Гейзенберг В. Шаги за горизонт. М. : Прогресс, 1987.
3. Гейзенберг В. Физика и философия. Часть и целое. М. : Наука, 1990.
4. Pauli W. Zur Quantenmechanik des magnetischen Elektrons // Z. Phys. 1927. V. 43. P. 601–623.
5. Сигал И. Математические проблемы релятивистской физики. М. : Мир, 1968.
6. Birkhoff G., von Neumann J. The logic of quantum mechanics // Ann. of Math. 1936. V. 37. P. 823–843.
7. Mackey G.W. Quantum mechanics and Hilbert space // Amer. Math. Monthly. 1957. V. 64. P. 45–57.
8. Фок В.А. Об интерпретации квантовой механики // УФН. 1957. Т. 62. С. 461–474.
9. Варламов В.В. Комплексный момент и спин-зарядовое гильбертово пространство // Математические структуры и моделирование. 2015. № 4(36). С. 5–22.
10. Боголюбов Н.Н., Логунов А.А., Оксак А.И., Тодоров И.Т. Общие принципы квантовой теории поля. М. : Наука, 1987.
11. Segal I.E. Irreducible representations of operator algebras // Bull. Amer. Math. Soc. 1947. V. 53. P. 73–88.
12. Born M., Heisenberg W., Jordan P., Zur Quantenmechanik. II // Zs. Phys. 1926. V. 35. P. 557–615.
13. Lounesto P. Clifford Algebras and Spinors. Cambridge : Cambridge Univ. Press, 2001.
14. Варламов В.В. Спинорная структура и SU(3)-симметрия // Математические структуры и моделирование. 2015. № 1(33). С. 18–33.
15. Varlamov V.V. Spinor Structure and Internal Symmetries // Int. J. Theor. Phys. 2015. V. 54. P. 3533–3576; arXiv:1409.1400 [math-ph] (2014).
16. Варламов В.В. Дискретные симметрии на пространствах фактор-представлений группы Лоренца // Математические структуры и моделирование. 2001. Вып. 7. С. 114–127.
17. Varlamov V.V. Universal Coverings of Orthogonal Groups // Adv. Appl. Clifford Algebras. 2004. V. 14. P. 81–168; arXiv:math-ph/0405040 (2004).
18. 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).
19. Varlamov V.V. Discrete Symmetries and Clifford Algebras // Int. J. Theor. Phys. 2001. V. 40. P. 769–805; arXiv:math-ph/0009026 (2000).
20. Varlamov V.V. The CPT Group in the de Sitter Space // Annales de la Fondation Louis de Broglie. 2004. V. 29. P. 969–987; arXiv:math-ph/0406060 (2004).
21. 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).
22. 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).

23. 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).
24. Варламов В.В. Спинорная структура и периодичность алгебр Клиффорда // Математические структуры и моделирование. 2015. № 3(35). С. 4–20.
25. Дирак П.А.М. Лекции по квантовой теории поля. М. : Мир, 1971.
26. Wigner E.P. On unitary representations of the inhomogeneous Lorentz group // Ann. Math. 1939. V. 40. P. 149–204.
27. Gell-Mann M., Ne'eman Y. The Eightfold Way. Benjamin, New York, 1964.
28. Румер Ю.Б., Фет А.И. Теория унитарной симметрии. М. : Наука, 1970.
29. Фет А.И. Группа симметрии химических элементов. Новосибирск : Наука, 2010.
30. Wick G.G., Wigner E.P., Wightman A.S. Intrinsic Parity of Elementary Particles // Phys. Rev. 1952. V. 88. P. 101.

HEISENBERG'S MATTER SPECTRUM IN ABSTRACT-ALGEBRAIC APPROACH

V.V. Varlamov

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

Siberian State Industrial University

Abstract. An algebraic structure of matter spectrum is studied. It is shown that a basic mathematical construction in the ground of matter spectrum is a two-level Hilbert space. Two-level structure of the Hilbert space is defined by the following pair: 1) separable Hilbert space with operator algebras and fundamental symmetries; 2) nonseparable (physical) Hilbert space with dynamical and gauge symmetries, that is, a space of states (energetic levels) of the matter spectrum. A decomposition of the physical Hilbert space onto coherent subspaces is given. This decomposition allows one to describe the all observed spectrum of the states on an equal footing, including lepton, meson and baryon sectors of the matter spectrum.

Keywords: matter spectrum, Hilbert space, coherent subspaces, superposition principle, reduction principle, symmetries.

ИССЛЕДОВАНИЕ ПОВЕДЕНИЯ СИСТЕМ ПРИ ФАЗОВЫХ ПЕРЕХОДАХ НА СТАДИИ КОАЛЕСЦЕНЦИИ

В.Н. Бородихин

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

Омский государственный университет им. Ф.М. Достоевского

Аннотация. Предложена феноменологическая модель, позволяющая описывать кинетику фазовых переходов на стадии коалесценции с помощью системы дробно-дифференциальных уравнений. Найдено общее решение, описывающее поведение зародышей при фазовых переходах на стадии коалесценции с учётом немарковости процессов. Получен скейлинговый закон, позволяющий описать практически весь диапазон возможных асимптотик поведения систем при фазовых переходах.

Ключевые слова: фазовые переходы, коалесценция, кинетика фазовых переходов, дробно-дифференциальные уравнения.

Введение

Исследование кинетики фазовых переходов представляет большой теоретический и экспериментальный интерес. В общем случае процесс фазового перехода является немарковским. В частности это может проявляться в скейлинговом поведении, отличном от закона Лифшица-Слёзова [1, 2]. Такое отличие может проявляться в том числе и в средах, обладающих фрактальными свойствами.

Для исследования фазовых переходов важную роль может играть нахождение аналитических решений, позволяющих глубже понять суть протекающих процессов.

В [1] было получено кинетическое уравнение и его аналитическое решение, описывающее кинетику фазовых переходов первого рода на стадии коалесценции.

В данной работе предложено феноменологическое обобщение этого уравнения с учётом дробно-дифференциальных производных, что формально позволяет описывать скейлинговое поведение систем, имеющих асимптотику, отличную от закона Лифшица-Слёзова, а также с учётом немарковских процессов и фрактальных свойств среды. Поскольку как известно, дробные показатели порядка дифференцирования можно связать с фрактальной размерностью [3]. Учёт дробных показателей порядка дифференцирования может позволить описывать некоторые дополнительные эффекты, которые могут происходить в данных системах. В частности, это позволит описать эффекты нелокальности и

дальнодействующие корреляционные эффекты, проявляющиеся в не гауссовом поведении кинетических процессов вблизи неравновесных квазистационарных состояний [4]. Кроме этого, в таких системах могут возникать эффекты неадекватности и самоорганизующей критичности [6].

Следует отметить, что реальные физические процессы являются немарковскими. К таким процессам можно отнести, к примеру, броуновское движение, где на движение частицы оказывает влияние её прошлое поведение в вязкой среде [7], фликкер-шум [8], также любые процессы, находящиеся под воздействием внешнего процесса, в том числе марковского. Поведение систем с учётом эффектов памяти можно описывать, вводя в уравнение дробную производную по времени [4].

Динамика реальных фазовых переходов также подразумевает существование эффектов памяти предыдущих состояний системы, и для адекватного описания кинетики фазовых переходов следует использовать дробную производную по времени. В частности, в большинстве реальных фазовых переходов наблюдаются эффекты размытия [9], фазовый переход становится нечётким. Для чётких фазовых переходов, соответствующих классификации Эренфеста, слагаемые, описывающие скачок термодинамического потенциала, включают в себя дельта функцию [9], что соответствует отсутствию памяти при фазовом переходе. Таким образом, введение дробно-дифференциальных величин будет соответствовать учёту эффектов памяти и, соответственно, размытию фазовых переходов.

1. Дробная коалесценция

В работе [1, 10] рассмотрена кинетика фазового перехода первого рода на поздней стадии развития — стадии коалесценции. Данная стадия характеризуется тем, что на ней основную роль играет процесс поглощения более мелких зародышей новой фазы вещества более крупными. Получено кинетическое уравнение, описывающее поведение зародышей на стадии коалесценции и найдено его аналитическое решение. Рассмотрим кратко данную теорию, а затем проведём обобщение полученного в [1] уравнения на случай фрактальных сред.

Рассмотрим твёрдый раствор, в котором выпадающие зёрна неподвижны и растут только за счёт диффузии из окружающего раствора. Зёрна приближённо будем считать сферическими и пренебрежём упругими напряжениями вокруг них.

Равновесная концентрация раствора c_0 у поверхности зерна с радиусом a определяется:

$$c_0 = c_\infty \left(1 + \frac{2\alpha'v}{aT} \right), \quad (1)$$

где v — молекулярный объём вещества зародыша, c_∞ — концентрация насыщенного раствора над плоской поверхностью растворяемого вещества, α' — коэффициент поверхностного натяжения на межфазной границе, v — молекулярный объём растворяемого вещества, T — температура. Кон-

центрация определяется по объёмному количеству вещества, растворённого в 1 см^3 раствора.

Диффузный поток i у поверхности зерна совпадает со скоростью изменения его радиуса:

$$i = \frac{da}{dt} = D \frac{\partial c}{\partial r} \Big|_{r=a}, \quad (2)$$

где D — коэффициент диффузии растворённого вещества. Поскольку концентрация предполагается малой, то данная скорость настолько мала, что распределение концентрации вокруг зародыша можно считать в каждый момент времени совпадающим со стационарным распределением $c(r)$, отвечающим данному значению a [10]:

$$c(r) = c - \frac{(c - c_0)a}{r}, \quad (3)$$

где c — средняя концентрация раствора. Следовательно, диффузный поток с учётом 1 и 2 примет вид:

$$\frac{da}{dt} = \frac{D(c_0 - c)}{a} = \frac{D}{a} \left(\Delta - \frac{\sigma}{a} \right), \quad (4)$$

где $\sigma = \frac{2\alpha v c_\infty}{T}$, $\Delta = c - c_\infty$.

Определим $a_k = \sigma/\Delta(t)$ — критический радиус. При $a > a_k$ — зерно растёт, при $a < a_k$ — растворяется.

Далее время будем измерять в единицах $\frac{a_k^3(0)}{D\sigma}$, где $a_k(0)$ — значение критического радиуса в начале стадии коалесценции. В итоге получаем уравнение:

$$\frac{da}{dt} = \frac{a_k^3(0)}{a} \left(\frac{1}{a_k} - \frac{1}{a} \right). \quad (5)$$

Введём функцию распределения зёрен по размерам, нормируемую на число зёрен в единице объёма:

$$N(t) = \int_0^\infty f(t, a) da. \quad (6)$$

Определим скорость перемещения зерна в пространстве размеров $v_a = da/dt$, запишем уравнение непрерывности:

$$\frac{\partial f}{\partial t} + \frac{\partial}{\partial a}(f v_a) = 0. \quad (7)$$

К этим уравнениям необходимо добавить закон сохранения полного количества вещества:

$$\Delta + q = Q = \text{const}, \quad (8)$$

где Q — полное начальное пресыщение, q — объём выпавших зёрен

$$q(t) = \int a^3 f(t, a) da. \quad (9)$$

Рассмотрим обобщение данного уравнения на случай фрактальных сред. Это подразумевает введение вместо производной по времени — дробную производную по времени порядка γ и, соответственно, вместо производной по a введение дробной производной по радиусу зародыша порядка α .

В обеих частях уравнения будем использовать производную Капуто [4]:

$$\frac{\partial^\gamma f(t)}{\partial t^\gamma} = \frac{1}{\Gamma(\gamma - n)} \int_0^t \frac{f^{(n)}(\tau) d\tau}{(t - \tau)^{\gamma+1-n}}. \quad (10)$$

Поступление вещества к надкритическому зародышу осуществляется с помощью диффузии из окружающего раствора. В стационарном режиме с учётом фрактальности среды сферически-симметричное распределение концентрации $c(r)$ вокруг зародыша может быть определено решением дробно-дифференциального диффузного уравнения:

$$D\Delta^\alpha c(r) = 0, \quad (11)$$

где Δ^α — дробно-дифференциальный оператор Лапласа [5], его радиальная компонента: $\Delta^\alpha = \frac{1}{r^{2\alpha}} \frac{\partial^\alpha}{\partial r^\alpha} r^{2\alpha} \frac{\partial^\alpha}{\partial r^\alpha}$. Граничные условия запишем: $c(\infty) = c$, $c(a) = c_0$. Тогда решение (11) имеет вид:

$$c(r) = c - \frac{\Gamma(1 - \alpha)(c - c_0) a^\alpha}{\Gamma(1 - 2\alpha) r^\alpha}. \quad (12)$$

Дробный диффузный поток у поверхности зародыша:

$$i^\gamma = \frac{d^\gamma a}{dt^\gamma} = D \frac{\partial^\alpha c(r)}{\partial r^\alpha} \Big|_{r=a}. \quad (13)$$

В итоге получим для дробного потока:

$$i^\gamma = D \frac{k_1(c - c_0)}{a^\alpha}, \quad (14)$$

где $k_1 = \frac{\Gamma^2(1 - \alpha)}{\Gamma^2(1 - 2\alpha)}$, $\Gamma(\alpha)$ — гамма-функция. Время будем измерять в единицах $\frac{a_k^{1+\alpha+\gamma}(0)}{D\sigma}$. В итоге с учётом (1) обобщённое на случай дробных производных (5) примет вид:

$$\frac{d^\gamma a}{dt^\gamma} = \frac{a_k^{1+\alpha+\gamma}(0)k_1}{a^\alpha} \left(\frac{1}{a_k} - \frac{1}{a} \right). \quad (15)$$

К данному уравнению необходимо добавить дробное-дифференциальное уравнение непрерывности:

$$\frac{\partial^\gamma f}{\partial t^\gamma} + \frac{\partial^\alpha}{\partial a^\alpha} (f v_a^\gamma) = 0, \quad (16)$$

где $v_u^\gamma = d^\gamma a / dt^\gamma$, и закон сохранения полного количества вещества (8). Дробный аналог уравнения непрерывности следует понимать как уравнение непрерывности в среде, проявляющей фрактальные свойства, а также допускающей протекание немарковских процессов [4].

Для удобства решения перейдём к новым переменным. Введём безразмерную величину $x(t) = a_k(t)/a_k(0)$. При $t \rightarrow \infty$ пресыщение $\Delta(t) \rightarrow 0$, а $a_k \rightarrow \infty$. Соответственно, при изменении переменной t от 0 до ∞ монотонно меняется от 0 до ∞ также величина $\tau = 3 \ln x(t)$, которую можно выбрать в качестве новой переменной по времени. В качестве неизвестной функции будем использовать $u = a/a_k(t)$.

В итоге с учётом формулы дробного дифференцирования произведения функций получим уравнение:

$$\frac{d^\gamma u^3}{d\tau} = -d^{\gamma-1} u^3 + \frac{2k_1 \bar{\gamma} u^{3-\alpha-\gamma}}{\Gamma(4-\gamma)} - \frac{2k_1 \bar{\gamma} u^{2-\alpha-\gamma}}{\Gamma(4-\gamma)}. \quad (17)$$

Проинтегрируем данное уравнение с помощью дробного интеграла $I_u^{\gamma-1}$

$$I_t^\gamma f(t) = \frac{1}{\Gamma(\gamma)} \int_0^t \frac{f(\tau) d\tau}{(t-\tau)^{1-\gamma}}; \quad (18)$$

получим:

$$\frac{du^3}{d\tau} = -u^3 + \frac{2k_1 \Gamma(4-\alpha-\gamma) \bar{\gamma} u^{2-\alpha}}{\Gamma(4-\gamma) \Gamma(3-\alpha)} - \frac{2k_1 \bar{\gamma} \Gamma(3-\alpha-\gamma) u^{1-\alpha}}{\Gamma(4-\gamma) \Gamma(3-\alpha)}. \quad (19)$$

Видно, что в левой части вместо дробной производной во времени появилась производная первого порядка. Также в правой части исчез дробный дифференциал и из степеней функции u также ушла зависимость от дробного временного показателя γ . Тем не менее зависимость от временного показателя γ сохранилась в правой части уравнения в параметрах гамма-функции.

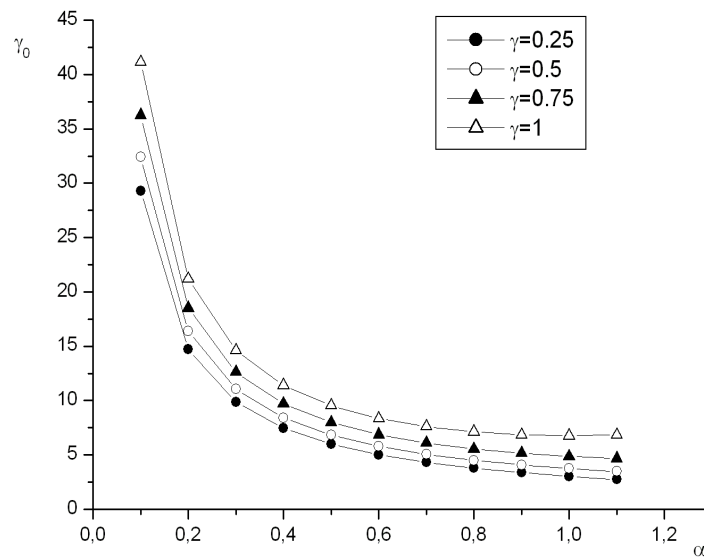
В правую часть (17), (19) входит функция $\bar{\gamma}$, определяемая как

$$\bar{\gamma}(\tau) = \frac{dt^\gamma}{x^{\gamma+\alpha} dx} > 0. \quad (20)$$

В отличие от соответствующей функции, полученной в рамках дифференцирования целого порядка [10], в (20) присутствует дифференциал по времени порядка γ , и в показатель степени x входит показатель γ и α . При $\alpha=\gamma=1$ (20) переходит в $\bar{\gamma}(\tau) = \frac{dt}{x^2 dx}$.

Определим зависимость $\bar{\gamma}_0$ — точки, в которой функция $\frac{d^\gamma u^3}{d\tau^\gamma}$ касается оси абсцисс от значения дробных параметров $\alpha + \gamma$. Для этого решим систему уравнений:

$$-u^3 + \frac{2k_1 \Gamma(4-\alpha-\gamma) \bar{\gamma} u^{2-\alpha}}{\Gamma(4-\gamma) \Gamma(3-\alpha)} - \frac{2k_1 \bar{\gamma} \Gamma(3-\alpha-\gamma) u^{1-\alpha}}{\Gamma(4-\gamma) \Gamma(2-\alpha)} = 0,$$

Рис. 1. Графики зависимости от α функции $\bar{\gamma}_0$ при различных γ .

$$-u^{2+\alpha} + \frac{2k_1(2-\alpha)\Gamma(4-\alpha-\gamma)\bar{\gamma}u}{\Gamma(4-\gamma)\Gamma(3-\alpha)} - \frac{2k_1(1-\alpha)\bar{\gamma}\Gamma(3-\alpha-\gamma)}{\Gamma(4-\gamma)\Gamma(2-\alpha)} = 0, \quad (21)$$

где 2-е уравнение получено путём дифференцирования 1-го по u для нахождения максимума функции $\frac{d^\gamma u^3}{d\tau^\gamma}$. Функция $\bar{\gamma}_0$ имеет вид:

$$\bar{\gamma}_0 = \frac{\Gamma(4-\gamma)}{2k_1} \left(\frac{\Gamma(3-\alpha-\gamma)}{\Gamma(2-\alpha)} \right)^{1+\alpha} \left(\frac{\Gamma(3-\alpha)}{\Gamma(4-\alpha-\gamma)} \right)^{2+\alpha} \frac{(2+\alpha)^{2+\alpha}}{(1+\alpha)^{1+\alpha}}. \quad (22)$$

При $\alpha = \gamma = 1$ $\bar{\gamma}_0 = \frac{27}{4}$, что совпадает со значением, найденным в [10].

Графики функции $\bar{\gamma}_0$ в зависимости от показателя α при различных γ показаны на рис. 1. Видно, что с ростом α функции плавно спадают, а также с уменьшением γ значения функции $\bar{\gamma}_0$ располагаются ниже на координатной оси.

При $\tau \rightarrow \infty$ функция $\bar{\gamma}(\tau)$ должна стремиться к определённому конечному пределу. Согласно [10], $\bar{\gamma}(\tau)$ можно представить в виде:

$$\bar{\gamma} = \bar{\gamma}_0(1 - \epsilon^2(\tau)), \quad (23)$$

где $\epsilon \rightarrow 0$ при $\tau \rightarrow \infty$.

При $\tau^2 \gg 1$, пренебрегая поправочным членом в (23), найдём предельный закон зависимости критического радиуса от времени:

$$x = \left(\frac{2k_1}{\Gamma(4-\gamma)} \left(\frac{\Gamma(2-\alpha)}{\Gamma(3-\alpha-\gamma)} \right)^{1+\alpha} \left(\frac{\Gamma(4-\alpha-\gamma)}{\Gamma(3-\alpha)} \right)^{2+\alpha} \frac{(1+\alpha)^{1+\alpha}}{(2+\alpha)^{2+\alpha}} \frac{t}{(\gamma+\alpha)} \right)^{1/(1+\alpha+\gamma)}. \quad (24)$$

Видно, что при $\alpha = \gamma = 1$ показатель степени основания t примет значение $1/3$ и (24) переходит в $x = \left(\frac{4t}{9}\right)^{1/3}$, что совпадает с целочисленным случаем [1]. Вблизи «запирающей точки» [1] решение определяется соответствующими поправками к данному закону.

В зависимости от показателей α и γ , определяемых свойствами фрактальной среды, при фазовых переходах уже в основном приближении меняется предельный закон зависимости критического радиуса от времени согласно (24).

С физической точки зрения здесь имеет место конкуренция между процессами, описываемыми дробными производными по времени и по размерам частиц, что может приводить в итоге к сложной динамике изменения размеров зародышей. Рост зародыша на стадии коалесценции может существенно задерживаться, не меняясь подолгу, затем он может резко измениться, что в итоге приводит к общему замедлению стадии коалесценции в системе. Данное поведение отвечает так называемому перемежаемому движению [3].

2. Функция распределения зёрен по размерам

Рассмотрим вычисление функции распределения зёрен по размерам. Функция распределения в переменных u и τ связана с функцией распределения в переменных a и t :

$$\varphi(\tau, u) = f(t, a)da, \quad (25)$$

$$f = \varphi/a_k.$$

Дробное уравнение непрерывности для функции φ имеет вид:

$$\frac{\partial^\gamma \varphi}{\partial \tau^\gamma} + \frac{\partial^\alpha}{\partial u^\alpha}(\varphi v_u) = 0, \quad (26)$$

$$\text{где } v_u = \frac{du}{d\tau}.$$

За исключением близкой ε -окрестности точки u_0 скорость v_u^γ определяется уравнением (с $\bar{\gamma} = \gamma_0$):

$$v_u = \frac{du}{d\tau} = -\frac{u}{3} + \frac{2k_1\Gamma(4-\alpha-\gamma)\bar{\gamma}}{\Gamma(4-\gamma)\Gamma(3-\alpha)u^\alpha} - \frac{2k_1\bar{\gamma}\Gamma(3-\alpha-\gamma)}{\Gamma(4-\gamma)\Gamma(2-\alpha)u^{1+\alpha}} = 0. \quad (27)$$

Используя обобщённую формулу Лейбница для дробных производных [11],

$$\frac{\partial^\alpha [f(x)g(x)]}{\partial x^\alpha} = \sum_{k=0}^{\infty} \frac{\Gamma(\alpha+1)}{k!\Gamma(\alpha-k+1)} f^{(\alpha-k)}(x)g^{(k)}(x), \quad (28)$$

и ограничиваясь первыми двумя слагаемыми, предполагая малость вклада слагаемых, содержащих производные второго и более высоких порядков (см. приложение), перепишем (26) в виде:

$$\frac{\partial^\gamma \varphi}{\partial \tau^\gamma} = \frac{\partial^\alpha v_u}{\partial u^\alpha} \varphi + \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} \frac{\partial^{\alpha-1} v_u}{\partial u^{\alpha-1}} \frac{\partial \varphi}{\partial u}. \quad (29)$$

Решение (29) будем искать в виде:

$$\varphi(u, \tau) = \varphi_1(u)\varphi_2(\tau). \quad (30)$$

Подставим (30) в (29) и поделим на (30), получим:

$$\frac{1}{\varphi_2} \frac{\partial^\gamma \varphi_2}{\partial \tau^\gamma} = -\frac{\partial^\alpha v_u}{\partial u^\alpha} - \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} \frac{\partial^{\alpha-1} v_u}{\partial u^{\alpha-1}} \frac{\partial \varphi_1}{\varphi_1 \partial u} = -\lambda. \quad (31)$$

Таким образом, приходим к системе двух уравнений для функций $\varphi_1(u)$ и $\varphi_2(\tau)$ с разделяющимися переменными:

$$\frac{d^\gamma \varphi_2}{\varphi_2} = -\lambda d\tau^\gamma, \quad (32)$$

$$\frac{d\varphi_1}{\varphi_1} = \frac{\Gamma(\alpha)}{\Gamma(\alpha+1)} \left(\frac{du}{d^{\alpha-1} v_u / du^{\alpha-1}} - \frac{\frac{d^\alpha v}{du^\alpha} du}{d^{\alpha-1} v_u / du^{\alpha-1}} \right). \quad (33)$$

Общее решение есть произведение решений уравнений (32) и (33). В итоге получим:

$$\phi(\tau, u) = A\tau^{\gamma-1} E_{\gamma, \gamma}(-\lambda\tau^\gamma) e^{\frac{\Gamma(\alpha)}{\Gamma(\alpha+1)} |I_1 - I_2|}, \quad (34)$$

где A — нормировочная константа, $\lambda = 1$ — константа, $E_{\gamma, \gamma}$ — обобщённая функция Миттаг-Леффлера [12]:

$$I_1 = \int \frac{du}{-A_2 u^{2-\alpha} + B_2 u^{1-2\alpha} - C_2 u^{-2\alpha}}, \quad (35)$$

$$I_2 = \int \frac{-A_1 u^{1-\alpha} + B_1 u^{-2\alpha} + C_1 u^{-1-2\alpha}}{-A_2 u^{2-\alpha} + B_2 u^{1-2\alpha} - C_2 u^{-2\alpha}} du, \quad (36)$$

где

$$\begin{aligned} A_1 &= \frac{1}{3\Gamma(2-\alpha)}, \\ B_1 &= \frac{2k_1 \bar{\gamma} \Gamma(4-\alpha-\gamma)}{3\Gamma(4-\gamma)\Gamma(3-\alpha)} \left(\frac{(\alpha+1)\Gamma(\alpha+1)}{\Gamma(\alpha)\Gamma(3-\alpha)} - \frac{1}{\Gamma(2-\alpha)} \right), \\ C_1 &= \frac{2k_1 \bar{\gamma} \Gamma(3-\alpha-\gamma)\Gamma(1+\alpha)(1+\alpha)}{3\Gamma(4-\gamma)\Gamma(2-\alpha)^2\Gamma(\alpha)}, \\ A_2 &= \frac{1}{3\Gamma(3-\alpha)}, \\ B_2 &= \frac{2k_1 \bar{\gamma} \Gamma(4-\alpha-\gamma)\Gamma(1-\alpha)}{3\Gamma(4-\gamma)\Gamma(3-\alpha)\Gamma(2-2\alpha)}, \\ C_2 &= \frac{2k_1 \bar{\gamma} \Gamma(3-\alpha-\gamma)B_z(1-\alpha, -\alpha)}{3\Gamma(4-\gamma)\Gamma(2-\alpha)\Gamma(1-\alpha)}, \end{aligned}$$

и где $B_z(1-\alpha, -\alpha)$ — неполная бета-функция, $z = \frac{u-c}{u}$, c — произвольная константа.

Данные интегралы можно посчитать явно при заданных значениях степеней α и γ . При $\alpha=\gamma=1$ $d^{\alpha-1}v_u/du^{\alpha-1} = v_u$, обобщённая функция Миттаг-Леффлера переходит в экспоненту, и решение совпадает с аналитическим решением, полученным в [1]:

$$\varphi(u, \tau) = Ae^{-\tau}P(u), \quad (37)$$

где $P(u)$ имеет вид:

$$P(u) = \frac{A_0 u^2 e^{\frac{1}{2u-3}}}{(u+3)^{7/3}(3-2u)^{11/3}}, \quad u < \frac{3}{2}, \quad (38)$$

$$P(u) = 0, \quad u > \frac{3}{2}.$$

Нормировочная константа A может быть найдена следующим образом. В уравнении сохранения вещества (8) $\Delta(\tau) \rightarrow 0$ при $\tau \rightarrow \infty$. Выразив q через переменные τ и u и подставляя $\varphi(u, \tau)$ из (37) получим:

$$\kappa A \int_0^{u_0} u^3 P(u) du = 1, \quad (39)$$

где $\kappa = \frac{4\pi a_k^{1+\alpha+\gamma}(0)}{3Q}$, u_0 — значение u , в котором кривая графика скорости $\frac{du^3}{d\tau}$ касается оси абсцисс ($\bar{\gamma} = \gamma_0$) [10].

Константа A_0 может быть найдена из условия нормировки:

$$\int_0^{u_0} P(u) du = 1. \quad (40)$$

График данной функции $P(u)$ изображён на рис. 2. Видно, что функция совпадает с кривой, полученной в [10] в рамках модели дифференцирования целочисленного порядка.

Рассмотрим решения для значений α и γ отличных от единицы. Для случая $\alpha = \frac{1}{2}$ получено действительное аналитическое решение:

$$P_{1/2}(u) = A_0 e^2 \frac{(u - \sqrt{u}\beta_1^{1/3} + \beta_1^{2/3})^{\frac{1}{6\beta_2}} (\beta_1^{2/3} - u)^{\frac{1}{3\beta_2}} (B_2 - A_2 u^{3/2})^{\beta_3} (u + \sqrt{u}\beta_1^{1/3} + (\sqrt{u} + \beta_1^{1/3})^{\frac{1}{3\beta_2}} (u^2 + u\beta_1^{2/3} + \beta_1^{4/3})^{\frac{1}{6\beta_2}} + \beta_1^{2/3})^{\beta_4} e^{\beta_5 \arctg(\frac{1}{\sqrt{3}}(\sqrt{u}\beta_1^{1/3}+1)) + \frac{1}{\sqrt{3}\beta_2} \arctg(\frac{1}{\sqrt{3}\beta_2}(2\sqrt{u}-1)) - \frac{1}{\sqrt{3}\beta_2} \arctg(\frac{1}{\sqrt{3}}(u\beta_1^{2/3}+1))}}{(\beta_1 - \sqrt{u})^{\beta_6}}, \quad (41)$$

$$u < u_0$$

$$P_{1/2}(u) = 0, \quad u > u_0,$$

$$\text{где } \beta_1 = \frac{B_2}{A_2}, \beta_2 = \frac{1}{3A_2^{2/3}B_2^{1/3}}, \beta_3 = \frac{1}{3} \left(\frac{2A_1}{A_2} - \frac{B_1}{B_2} \right), \beta_4 = \frac{1}{3B_2^{1/3}} \left(\frac{C_1 A_2^{2/3}}{B_2^{4/3}} - \frac{1}{2A_2^{2/3}} \right),$$

$$\beta_5 = \frac{1}{3B_2^{1/3}} \left(\frac{2C_1 A_2^{2/3}}{B_2^{4/3}} + \frac{1}{A_2^{2/3}} \right), \beta_6 = \frac{1}{3B_2^{1/3}} \left(\frac{2C_1 A_2^{2/3}}{B_2^{4/3}} - \frac{1}{A_2^{2/3}} \right).$$

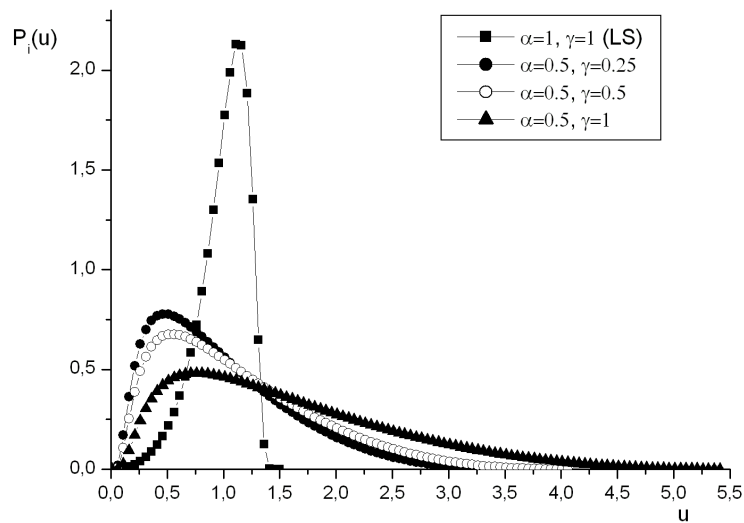


Рис. 2. Графики функций $P_{1/2}(u)$ при различных γ . Квадратиками приведено распределение Лифшица-Слэзова ($\alpha = \gamma = 1$)

Соответственно, нормировочная константа A_0 определяется из условия нормировки функции $P_{1/2}(u)$ на единицу. Для случая $\alpha = 0.5$, $\gamma = 1$ $A_0 = 880.72$.

Графики функций $P_{1/2}(u)$ при различных γ приведены на рис. 2.

Видно, что с ростом γ от 0.25 до 1 пик функции $P_{1/2}(u)$ уменьшается, при этом смещается точка u_0 , в которой функция обращается в ноль от $u_0 = 3.631$ ($\gamma = 0.25$) до $u_0 = 5.449$ ($\gamma = 1$). Относительно сравнения между собой графиков $\alpha = 0.5$ и $\alpha = 1$ можно отметить, что пики графиков $\alpha = 0.5$ расположены значительно ниже и смещены влево относительно функции $\alpha = 1$. Точка u_0 для $P(u)$ равна $3/2$, что значительно меньше, чем значения соответствующих точек функций $P_{1/2}(u)$. Графики $P_{1/2}(u)$ значительно шире, чем $P(u)$. Можно сказать, что функция $P_{1/2}(u)$ является как бы размытой относительно функции $P(u)$.

Число зёрен в единице объёма в случае $\alpha = \gamma = 1$ можно определить следующим образом:

$$N = \int_0^{u_0} \varphi(u, \tau) du = A e^{-\tau} = \frac{9A}{4t}. \quad (42)$$

Для случая $\alpha = 0.5$, $\gamma = 1$ обобщённая функция Миттаг-Леффлера $E_{\gamma\gamma}(-\tau^\gamma)$ переходит в $e^{-\tau}$, и для этого случая число зёрен может быть определено:

$$N = A \left(\frac{\alpha + 1}{\gamma_0 t} \right)^{\frac{3}{2+\alpha}} = A \left(\frac{3}{2\gamma_0 t} \right)^{\frac{6}{5}}. \quad (43)$$

Видно, что с ростом t число зёрен в случае $\alpha = 0.5$ спадает немного быстрее, чем в случае $\alpha = 1$.

Следует отметить, что в проведённом рассмотрении общий объём раствора рассматривался как неограниченный, а поэтому неограничен полный запас растворяемого вещества. В конечном объёме процесс должен заканчиваться за конечное время, когда всё растворённое вещество выпадет в виде одного тела.

В [13] было показано, что при обобщении различных уравнений с использованием дробных производных Римана-Лиувилля, могут возникать функции распределения, допускающие отрицательную плотность распределения. В данной работе применялась производная Капуто, использование которой также может приводить к нефизическим решениям. Тем не менее, в данном случае, полученное решение для *плотности* распределения (зёрен по размерам), как и решение [1] – как видно из приведённых графиков функций (рис. 2.) – является положительно определённым.

3. Скейлинговое поведение

Рассмотрим скейлинговое поведение. В [1] предсказано асимптотическое поведение радиуса зерна как $\sim t^{1/3}$, что подтверждается множеством экспериментальных исследований [14]. В данной работе получено асимптотическое поведение $\sim t^{1/1+\alpha+\gamma}$. Такой вид закона скейлинга позволяет в общем случае формально описывать практически весь диапазон найденных теоретически и подтверждённых экспериментально асимптотик. Асимптотический закон скейлинга можно представить как $\sim t^\varphi$, где показатель φ зависит от природы законов сохранения, наличия различных дефектов или примесей, свойств протекания полей [2].

Так закон Лифшица-Слёзова $\sim t^{1/3}$ характерен для систем с сохранением параметра порядка таких, которые в конечном состоянии имеют две сосуществующие фазы. Это соответствует $\alpha = 1$, $\gamma = 1$.

Закон Аллена-Кана $\sim t^{1/2}$ характерен для систем с несохраняющимся параметром порядка, который эволюционирует к конечному однофазному, равновесному упорядоченному состоянию с нарушенной симметрией [15]. Данный закон соответствует показателям $\alpha = 0.5$, $\gamma = 0.5$, т.е. проявлению полупространственных эффектов нелокальности (корреляции) размеров зёрен, а также эффектов памяти.

Для бинарных жидких смесей и бинарных сплавов были найдены асимптотики $\sim t^{1/3}$, $\sim t^{2/3}$ и $\sim t^1$ [2, 16], что соответствует диффузному режиму, вязкому гидродинамическому режиму и инерциальному гидродинамическому режиму. Последним двум случаям могут соответствовать индексы α или γ принимающие отрицательные значения (к примеру, $\alpha = 0.5$ или $\gamma = -0.5$), что соответствует переходу от дробных дифференциальных уравнений к интегродифференциальным. В промежуточных между этими асимптотиками значениях показателя $\varphi > 1/3$ индексы α и γ оба могут принимать дробные положительные значения.

Следует также отметить существование асимптотик с показателем

$\varphi < 1/3 \sim t^{1/4}$, а также с меньшим показателем φ , соответствующих, к примеру, поведению фрактальных кластеров, образующихся в процессе затвердевания переохлаждённой жидкости [17], фрактальным кластерам на подложке, выращенным методом осаждения [18], фрактальным вязким направленным узорам, формируемым неустойчивостью Саффмана-Тейлора [17]. Здесь индексы α и γ должны принимать значения ≥ 1 .

Таким образом, видно, что при описании асимптотического поведения различных систем с помощью дробно-дифференциальных уравнений важную роль играет как учёт дробного показателя по размерам зёрен, так и учёт дробного показателя по времени.

4. Заключение

Задача описания кинетики фазовых переходов на стадии коалесценции [1], как показано выше, может быть обобщена с помощью введения дробных производных как по времени, так и по радиусу зародышей. Получающиеся в итоге уравнения допускают аналитическое решение. Однако только решение для дробно-дифференциального индекса радиуса зародышей $\alpha = 1/2$ является физическим. Решения для других α являются комплексными и имеют нефизическую природу. Вероятно для получения действительных решений для случаев $\alpha \neq 0.5$ в данной модели необходимо учитывать какие-то дополнительные поправки. Однако это уже не гарантирует возможность получения решений в аналитическом виде.

Кроме этого, как показано в [13], использование дробных производных, в частности, дробной производной Римана-Лиувилля, существенно меняет свойства плотностей распределения, ухудшая свойства итогового уравнения, что также может являться причиной отсутствия действительных решений для показателей $\alpha \neq 0.5$.

Тем не менее удалось получить некоторые общие зависимости от дробно-дифференциальных показателей. Как следует из графиков зависимости функции $\bar{\gamma}_0$ (рис. 1) с ростом α от 0 до 1 функции плавно спадают, с уменьшением γ значения функции $\bar{\gamma}_0$ располагаются ниже на координатной оси. Было найдено, что в зависимости от показателей α и γ при фазовых переходах уже в основном приближении меняется предельный закон зависимости критического радиуса от времени. Показатель α в общем случае меняет закон зависимости количества зёрен в единице объёма от времени. Также в зависимости от свойств фрактальной среды меняется функция распределения зёрен по размерам. Вместо экспоненциальной зависимости от времени возникает обобщённая функция Миттаг-Леффлера, определяемая показателем дробной производной по времени.

С физической точки зрения имеет место конкуренция между процессами, описываемыми дробными производными по времени и по размерам частиц, что может приводить в итоге к сложной динамике изменения размеров зародышей. Рост зародыша может существенно задерживаться, не меняясь подолгу, затем он может резко измениться, что в итоге приводит к общему замедлению стадии

коалесценции в системе. Данное поведение отвечает так называемому перемещаемому движению [3].

Что касается интерпретации процессов, связанных с дробной производной по времени, в [19] выдвинуто предположение, что временная степень свободы является стохастической и представляет собой сумму случайных временных отрезков, каждый из которых является случайной величиной устойчивого вероятностного распределения. При переходе к непрерывному времени строятся процессы со случайным операционным временем. Для описания таких процессов в кинетические уравнения следует добавлять дробную производную по времени, где показатель дробной производной связан с параметром соответствующего распределения вероятности. Такой характер распределения вероятности порождает долговременные эффекты памяти.

Данная интерпретация может быть использована в том числе и для описания немарковской странной кинетики фазовых переходов, что соответствует поведению реальных систем при фазовых переходах.

Для показателя $\alpha = 1/2$ получено аналитическое решение при различных значениях γ (рис. 2.). Функция распределения $P(u)$ меняет свой вид по сравнению с $\alpha = 1$. Можно отметить, что пики графиков $\alpha = 0.5$ расположены значительно ниже и смещены влево относительно функции $\alpha = 1$. Точка u_0 для $P(u)$ равна $3/2$, что значительно меньше, чем значения соответствующих точек функций $P_{1/2}(u)$. Графики $P_{1/2}(u)$ значительно шире, чем $P(u)$. Можно сказать, что функция $P_{1/2}(u)$ является как бы размытой относительно функции $P(u)$. С ростом γ от 0.25 до 1 пик функции $P_{1/2}(u)$ уменьшается, при этом смещается точка u_0 , в которой функция обращается в ноль от $u_0 = 3.631$ ($\gamma = 0.25$) до $u_0 = 5.449$ ($\gamma = 1$).

Сопоставление с экспериментальными данными позволило выявить ещё одну универсальную стадию фазового перехода первого рода — так называемую начальную стадию коалесценции.

Полученный в работе скейлинговый закон $\sim t^{1/1+\alpha+\gamma}$ позволил описать практически весь диапазон возможных асимптотик [2]. Это означает, что при описании асимптотического поведения различных систем с помощью дробно-дифференциальных уравнений важную роль играет как учёт дробного показателя по размерам зёрен, так и учёт дробного показателя по времени.

5. ПРИЛОЖЕНИЕ

Бесконечную сумму ряда дробной производной произведения функций (правило Лейбница) можно записать в виде конечного ряда с остаточным членом [20]:

$$F(x) = \frac{\partial^\alpha [fg]}{\partial x^\alpha} = \sum_{k=0}^n \frac{\Gamma(\alpha+1)}{k! \Gamma(\alpha-k+1)} f^{(\alpha-k)} g^{(k)} + \frac{(-1)^{n+1}}{\Gamma(-\alpha)} \int_a^x f^{(n+1)}(t) [I_a^{n+1}(x-t)^{-\alpha-1} g(t)] dt, \quad (44)$$

$0 < \alpha < 1$, $f^{(n+1)}$ — производная $n+1$ -го порядка от функции f , I_a^β — оператор дробного интегрирования

$$I_a^\beta f = \frac{1}{\Gamma(\beta)} \int_a^x \frac{f(t) dt}{(x-t)^{-\beta+1}}. \quad (45)$$

(44) можно записать как $F(x) = P(x) + R_n(x)$, где $P(x)$ — конечная сумма, $R_n(x)$ — остаточный член.

Оценим остаточный член. Используя первую формулу среднего значения [21] для I_a^{n+1} , выберем на интервале $[a, t]$ точку z , в результате получим:

$$R_n(x) = \frac{(-1)^{n+1} g(z)}{\Gamma(-\alpha)(n+1)!(x-z)^{\alpha+1}} \int_a^x f^{(n+1)}(t)(t-a)^{n+1} dt, \quad (46)$$

Снова используем формулу среднего значения, выберем на интервале $[a, x]$ точку y , в итоге получим остаточный член в следующем виде:

$$R_n(x) = \frac{(-1)^{n+1} f^{(n+1)}(y) g(z)}{\Gamma(-\alpha)(n+2)!} \frac{(x-a)^{n+2}}{(x-z)^{\alpha+1}}, \quad (47)$$

Можно доказать, что при $x \rightarrow a$ остаточный член представляет собой бесконечно малую порядка выше n : $R_n(x) = o((x-a)^n)$. Пусть $H_n(x) = \frac{R_n(x)}{(x-a)^n}$ при $x \neq a$ и $H_n(x) = 0$ при $x = a$. Достаточно показать, что $\lim_{x \rightarrow a} H_n(x) = 0$. Заметим, что для всех $i = 0, 1, 2, \dots, n$ $R_n^{(i)}(a) = 0$. Применим $n-1$ раз правило Лопиталья, получим

$$\begin{aligned} \lim_{x \rightarrow a} \frac{R_n(x)}{(x-a)^n} &= \lim_{x \rightarrow a} \frac{\frac{d}{dx} R_n(x)}{\frac{d}{dx} (x-a)^n} = \dots = \lim_{x \rightarrow a} \frac{\frac{d^{n-1}}{dx^{n-1}} R_n(x)}{\frac{d^{n-1}}{dx^{n-1}} (x-a)^n} = \\ &= \frac{1}{n!} \lim_{x \rightarrow a} \frac{R_n^{(n-1)}(x)}{(x-a)} = \frac{1}{n!} \lim_{x \rightarrow a} \frac{R_n^{(n-1)}(x) - R_n^{(n-1)}(a)}{(x-a)} = \frac{1}{n!} R_n^{(n)}(a) = 0. \end{aligned} \quad (48)$$

Каждая следующая производная числителя стремится к 0 в точке $x = a$, и также каждая следующая производная знаменателя. Последнее равенство в (48) следует из определения производной в точке $x = a$.

ЛИТЕРАТУРА

1. Лифшиц И.М., Слёзов В.В. Кинетика диффузионного распада пересыщенных твердых растворов // Журнал эксперим. и теор. физики. 1958. Т. 35, С. 379.
2. Kinetic of phase transitions. / ed. Puri S., Wadhavan V., London: CRC Press, 2009.
3. Зелёный Л.М., Милованов А.В. Фрактальная топология и странная кинетика: от теории перколяции к проблемам космической электродинамики // Успехи физ. наук 2004. Т. 174, С. 809.
4. Tarasov V.E. Fractional Dynamics: Application of Fractional Calculus to Dynamics of Particles, Fields and Media. Beijing: Springer, Higher Education Press, 2010.
5. Li M-F., Ren Ji-R., Zhu T. Fractional Vector Calculus and Fractional Special Function. e-print arXiv:1001.2889, 2010.
6. Олемской А.И. Синергетика сложных систем. Москва: URSS, 2009.
7. Morozov A.N., Skripkin A.V. Spherical particle Brownian motion in viscous medium as non-Markovian random process // Phys. Lett. A. 2011. V. 375, P. 4113.
8. Bak P., Tang C., Wiesenfeld K. Self-organized criticality: An explanation of the $1/f$ noise // Phys. Rev. Lett. 1987. V. 59, P. 381.
9. Ролов Б.Н., Юркевич В.Э. Физика размытых фазовых переходов. Рига: РГУ, 1983.
10. Лифшиц Е.М., Питаевский Л.П. Теоретическая физика, т. 10, Физическая кинетика, Москва: Физматлит, 2002.
11. Oldham K.B., Spanier J. The fractional calculus. San Diego: Academic Press, 1974.
12. Kilbas A.A., Srivastava H.M., Trujillo J.J. Theory and Application of Fractional Differential Equations. Amsterdam: Elsevier, 2006.
13. Зенюк Д.А., Орлов Ю.Н. О применении дробного исчисления Римана–Лиувилля для описания распределений вероятностей // Препринт ИПМ РАН им. М.В. Келдыша. 2014. Т. 18.
14. Слёзов В.В., Сагалович В.В. Диффузионный распад твёрдых растворов // Успехи физ. наук 1987, Т. 151, Р. 67.
15. Gouyet, J.-F. Generalized Allen–Cahn equations to describe far-from-equilibrium order–disorder dynamics // Phys. Rev. E. 1995. V. 51, P. 1695.
16. Hashimoto T. Jinnai H., Hasegawa H. and Han C.C. Late stage spinodal decomposition in binary critical fluids: scaling function obtained over a wide q -space of 4 orders of magnitude // Physica A. 1994. V. 204, P. 261.
17. Langer J.S. Instabilities and pattern formation in crystal growth // Rev. Mod. Phys. 1980. V. 52, P. 1.
18. Brechignac C. et al. Instability Driven Fragmentation of Nanoscale Fractal Islands // Phys. Rev. Lett. 2002. V. 88, P. 196103.
19. Станиславский А.А. Вероятностная интерпретация интеграла дробного порядка // Теоретическая и математическая физика 2004. Т. 138, С. 491.
20. Bassam M.A. Some properties of Holmgren-Riesz transform // Annali della Scuola Normale Super di Pisa. 1961. V. 15, P. 1.
21. Ильин В.А., Садовничий В.А., Сендов Б.Х. Математический анализ, т.1. Москва: МГУ, 1985.

STUDY OF THE SYSTEMS BEHAVIOR IN PHASE TRANSITIONS ON THE COALESCENCE STAGE

V.N. Borodikhin

Ph.D. (Phys.-Math.), Associate Professor, e-mail: borodikhin@inbox.ru

Dostoevsky Omsk State University

Abstract. A phenomenological model that allows to describe the kinetics of phase transitions in the coalescence stage using fractional differential equations was proposed. A general solution, describing nucleation behavior in phase transitions on the coalescence stage, taking into account non-markov processes, was found. A scaling law, which allows to describe the entire range of possible asymptotic behavior of systems in phase transitions, was received.

Keywords: phase transitions, coalescence, kinetics of phase transitions, fractional differential equations.

ОЦЕНКА УСТОЙЧИВОСТИ ДЛЯ ВЫПУКЛЫХ ПОВЕРХНОСТЕЙ

В.Н. Степанов

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

Омский государственный технический университет

Аннотация. В статье даны необходимые и достаточные условия аналитичности решения уравнения первого рода типа свёртки на сфере с ядром $K(t) \in L_2[-1, 1]$. Доказана аналитичность некоторых функционалов замкнутой выпуклой аналитической поверхности. С помощью леммы Витушкина получена оценка для разности опорных функций двух замкнутых выпуклых аналитических поверхностей, если для них известны отклонения функционалов поверхностей и их производных на конечном множестве точек. Приведены некоторые известные результаты по проблеме устойчивости в определении выпуклой поверхности по томографическим данным.

Ключевые слова: уравнение первого рода, аналитичность, опорная функция, выпуклая поверхность, устойчивость, лемма Витушкина.

В последние годы получены различные результаты по устойчивости в определении выпуклых тел по некоторым функционалам выпуклой поверхности (томографическим данным: меры сечений и проекций, опорная функция, поверхностная функция выпуклой поверхности и т.д.). Эти теоремы утверждают, что при определённых условиях отклонение двух выпуклых тел мало (с точностью до сдвига), если томографические данные близки друг к другу. Подход, используемый для этого, может быть основан на исследовании устойчивости интегральных преобразований, тесно связанных с восстановлением выпуклого тела по томографическим данным [17, 18]. Частными случаями таких преобразований являются: сферическое преобразование Радона, полусферическое преобразование и косинус-преобразование функций, определённых на сфере.

Приведём некоторые известные и важные результаты по этой проблеме. Большой объём информации о проблеме содержится в работах [11–22].

В работах S. Campi [11, 12] даны оценки расстояния в L_2 — метрике и в метрике Хаусдорфа между двумя выпуклыми центрально-симметричными телами, телами вращения через L_2 — норму разности периметров и L_2 — норму разности площадей ортогональных проекций на плоскости всевозможных направлений.

Глубокий результат, обобщающий оценки S. Campi по размерности, был доказан J. Bourgain, J. Lindenstrauss J. [10]. Пусть K — выпуклое тело в $\mathbb{R}^n$, $\Pi(K)$ — проекционное тело (это тело с опорной функцией $H_{\Pi(K)}(u) = \{(n-1)$ -мерному объёму ортогональной проекции K на гиперплоскость ортогональную

$u\}$, $\delta(K, H) = \min\{\log \lambda : \lambda^{-1}H \subset K \subset \lambda H\}$ — метрика. Тогда для любого $n \geq 3$ и для любого $\gamma < 2/n(n+4)$ существует такая постоянная $c(n, \gamma)$, что $\delta(K, H) \leq c(n, \gamma)\delta(Z(K), Z(H))^\gamma$.

В случае, когда томографические данные даются формулой $(T_\Phi \sigma_{n-1}(K))(u) = \int_{S^{n-1}} \Phi(\langle u, v \rangle) \sigma_{n-1}(K, dv)$, где $\sigma_{n-1}(K, \cdot)$ — поверхностная функция [6, с. 18], $\Phi(t)$ ограниченная измеримая функция на $[-1, 1]$, оценка устойчивости была получена D. Hug и R. Schneider [19]: для некоторого $\gamma \in (0, 1)$ существует постоянная c , зависящая только от n, r, γ, R, Φ такая, что, если $K, L \in K^d(r, R)$, то $\delta(K, L + x) \leq c \|T_\Phi(\sigma_{n-1}(K) - \sigma_{n-1}(L))(u)\|^\gamma$. Здесь $\delta(K, L)$ — метрика Хаусдорфа, $\|\cdot\|$ — L_2 -норма, $K^d(r, R)$ — множество выпуклых тел, содержащихся внутри шара радиуса R и содержащих шар радиуса r .

Оценка в метрике L_2 индикатора отклонения опорных функций H_K и H_L выпуклых тел K и L в $\mathbb{R}^n$ в предположении, что мала разность $\|M(K_u) - M(L_u)\|_{L_2(S^{n-1})}$, где $M(K_u)$, $M(L_u)$ — ширины проекций тел K и L на подпространство ортогональное u , получена P. Goodey, H. Groemer в [16].

Алгоритмы реконструкции выпуклого тела по конечному набору значений томографических данных рассматривались в работах [14, 15]. Основной идеей этих алгоритмов является построение выпуклого многогранника, чья функция томографических данных наилучшим образом приближается данными измерений.

В случае когда томографическими данными являются значения функции яркости, основанием для построения таких алгоритмов может служить следующее утверждение [13, 20]. Пусть K — выпуклое тело в $\mathbb{R}^n$ и $U = \{u_1, \dots, u_k\} \subset S^{n-1}$ — линейная оболочка $\mathbb{R}^n$. Среди всех выпуклых тел с одинаковыми значениями функции яркости, как для тела K в направлениях из U , существует единственный центрально-симметричный выпуклый многогранник P максимального объёма, каждая грань которого ортогональна к одному из векторов из U .

В представленной работе получена оценка для $\sup$ — нормы разности опорных функций двух выпуклых тел в $\mathbb{R}^3$ с аналитической границей по значениям площадей ортогональных проекций и освещённых частей на конечном множестве направлений.

Пусть S^{n-1} — единичная сфера в $\mathbb{R}^n$; $u, v \in S^{n-1}$; $\langle u, v \rangle$ — скалярное произведение. Рассмотрим уравнение первого рода типа свёртки на S^{n-1}

$$f(u) = \int_{S^{n-1}} K(\langle u, v \rangle) z(v) d\omega_v, \quad (1)$$

где $d\omega$ — элемент площади.

Теорема 1. Уравнение (1) с полным ядром $K(\langle u, v \rangle) \in L_2[-1, 1]$ имеет аналитическое на сфере S^{n-1} решение $z(u)$ тогда и только тогда, когда функция $f(u)$ является аналитической на S^{n-1} .

Доказательство. Для доказательства аналитичности воспользуемся следующим фактом. Пусть на сфере S^{n-1} задана суммируемая функция $g(u)$, и пусть $g(u) = \sum_{k=0}^{\infty} Y_k(u)$ — её разложение в ряд по сферическим гармоникам. Тогда для того, чтобы функция $g(u)$ была аналитической, необходимо и достаточно выполнение неравенств:

$$|Y_k(u)| \leq c \exp(-\eta k), u \in S^{n-1}, k = 0, 1, 2, \dots$$

с некоторыми постоянными $c > 0, \eta > 0$ [7].

Предположим теперь, что $z(u)$ — аналитическая на S^{n-1} функция и

$$z(u) = \sum_{k=0}^{\infty} Y_k(u), \quad Y_k(u) = \frac{(n+2k-2)\Gamma(n/2-1)}{4\pi^{n/2}} \int_{S^{n-1}} C_k^{(n/2-1)}(\langle u, v \rangle) z(v) d\omega_v$$

— её разложение по системе сферических гармоник $\{Y_k(u)\}$. Тогда разложение функции $f(u)$ в ряд имеет вид:

$$f(u) = \sum_{k=0}^{\infty} \lambda_k Y_k(u),$$

где λ_k — собственные значения ядра $K(\langle u, v \rangle)$. Ввиду полноты ядра все собственные значения ядра отличны от нуля. По формуле Функа-Гекке [1, с. 240]:

$$\begin{aligned} \lambda_k &= \frac{\Gamma(n-2)\Gamma(k+1)}{\Gamma(k+n-2)} \int_{S^{n-1}} K(\langle u, v \rangle) C_k^{(n/2-1)}(\langle u, v \rangle) d\omega_v = \\ &= \frac{|S^{n-2}| \Gamma(n-2)\Gamma(k+1)}{\Gamma(k+n-2)} \int_{-1}^1 K(t) C_k^{(n/2-1)}(1-t^2)^{(n-3)/2} dt, \end{aligned}$$

где $|S^{n-2}| = \frac{2\pi^{(n-1)/2}}{\Gamma((n-1)/2)}$ — площадь поверхности сферы S^{n-2} , $\Gamma(t)$ — гамма-функция, $C_k^{(n/2-1)}(t)$ — многочлены Гегенбауэра.

В статье [9] показано, что собственные значения уравнения (1) с ядром $K(\langle u, v \rangle)$ класса $L_1[-1, 1]$ при $k \rightarrow \infty$ убывают быстрее, чем $k^{-n/2+1}$. Поэтому члены разложения $f(u)$ по сферическим гармоникам убывают по экспоненциальному закону. Следовательно, $f(u)$ — аналитическая на S^{n-1} функция. Аналогично доказывается аналитичность $z(u)$, если $z(u)$ — аналитическая функция. ■

Пусть B — замкнутая выпуклая дважды непрерывно дифференцируемая поверхность в $\mathbb{R}^3$ с положительной кривизной. Через $F(u)$ обозначим площадь ортогональной проекции поверхности B на плоскость с нормальным вектором u , $S(u)$ — площадь освещённой в направлении u части поверхности. Имеют место формулы [2]:

$$F(u) = \frac{1}{2} \int_{S^{n-1}} |\langle u, v \rangle| R_1(v) R_2(v) d\omega_v, \quad S(u) = \int_{S^{n-1}} \chi(\langle u, v \rangle) R_1(v) R_2(v) d\omega_v,$$

где χ — функция Хевисайда и формулы обращения [2, 5, 8]:

$$\begin{aligned} [R_1(u)R_2(u)]^+ &= -\frac{1}{4\pi^2} \frac{d}{dt} \int_{\langle u, v \rangle^2 > t} \frac{(\Delta_S + 2)F(u)|\langle u, v \rangle| d\omega_v}{\sqrt{\langle u, v \rangle^2 - t}} \Big|_{t=0} = \\ &= \frac{1}{2\pi} (\Delta_S + 2)F(u) + \frac{1}{2\pi} \int_0^{2\pi} \frac{[(\Delta_S + 2)\tilde{F}(\gamma, u)]' d\gamma}{\cos \gamma}, \end{aligned}$$

$$\begin{aligned} [R_1(u)R_2(u)]^- &= -\frac{1}{4\pi^2} \int_{\langle u, v \rangle > 0} \frac{\Delta_S S^-(v) d\omega_v}{\langle u, v \rangle} = \\ &= \frac{1}{2\pi} S^-(u) - \frac{1}{2\pi} \int_0^{\pi/2} \tan \gamma \tilde{S}^{--}(\gamma, u) d\gamma, \end{aligned}$$

где $f^\pm(u)$ — чётная и нечётная части функции $f(u)$, Δ_S — поверхностный оператор Лапласа, $\tilde{f}(\gamma, u)$ — среднее значение функции $f(u)$ на окружности $\langle u, v \rangle = \cos \gamma$, $0 < \gamma < \pi$.

Из этих формул следуют оценки для разности произведений $R^{(j)}(u) = R_1^{(j)}(u) \cdot R_2^{(j)}(u)$ главных радиусов кривизны поверхностей B_1 и B_2 [8]:

$$\begin{aligned} \|R^{(2)}(u) - R^{(1)}(u)\|_{C(S^{n-1})} &\leq \\ &\leq c_1 \|F_2(u) - F_1(u)\|_{C^4(S^{n-1})} + c_2 \|S_2(u) - S_1(u)\|_{C^3(S^{n-1})}, \end{aligned} \quad (2)$$

где $c_1 > 0$ и $c_2 > 0$ — некоторые постоянные.

Если B — замкнутая выпуклая аналитическая поверхность с положительной гауссовой кривизной, то её опорная функция $H(u)$ аналитична на S^{n-1} . Для произведения главных радиусов кривизны известна формула

$$R_1(u)R_2(u) = \begin{vmatrix} H_{11}(u) & H_{12}(u) \\ H_{12}(u) & H_{22}(u) \end{vmatrix} + \begin{vmatrix} H_{22}(u) & H_{23}(u) \\ H_{32}(u) & H_{33}(u) \end{vmatrix} + \begin{vmatrix} H_{11}(u) & H_{13}(u) \\ H_{31}(u) & H_{33}(u) \end{vmatrix},$$

где $H_{ij}(u) = \frac{\partial^2 H(x)}{\partial x_i \partial x_j} \Big|_{x=u}$, из которой следует аналитичность $R_1(u)R_2(u)$. Теперь из теоремы 1 получаем:

Следствие 1. Если B — аналитическая поверхность, то функции $F(u)$ и $S(u)$ — аналитические.

Пусть u_0 — фиксированная точка на сфере S^2 ; (θ_0, φ_0) — сферические координаты точки u_0 относительно некоторого полюса. Обозначим через J_δ множество всех тех точек $u \in S^2$, для которых $|\theta - \theta_0| \leq \delta$, $|\varphi - \varphi_0| \leq \delta$, где ϑ, φ — сферические координаты точки u относительно того же полюса и $\delta > 0$ — фиксированное число.

Теорема 2. Пусть B_1 и B_2 — замкнутые выпуклые аналитические поверхности в $\mathbb{R}^3$, гауссовы кривизны которых положительны. Тогда для всякого положительного $\varepsilon \leq 1/2$ в J_δ можно фиксировать $\nu \leq A_1 \left(\ln \frac{1}{\varepsilon}\right)^2$ точек $u_1, u_2, \dots, u_\nu$ таких, что если в этих точках функции $f(u) = F_2(u) - F_1(u)$, $s(u) = S_2(u) - S_1(u)$ и их производные $D^\alpha f(u)$, $|\alpha| \leq 4$, $D^\beta s(u)$, $|\beta| \leq 4$, принимают значения по модулю не превосходящие ε , то при некотором расположении поверхностей B_1 и B_2 разность их опорных функций $H_2(u) - H_1(u)$ удовлетворяет неравенству

$$|H_2(u) - H_1(u)| \leq A_2 \varepsilon^{\frac{1-\gamma}{5}} \left(\ln \frac{1}{\varepsilon}\right)^{\frac{4}{5}}, \quad u \in S^2,$$

где ρ, r ($\rho > r > 1$), $A_1 > 0$, $A_2 > 0$, $\gamma = \frac{2 \ln r}{\ln \rho - \ln r}$ — некоторые постоянные

Доказательство. По следствию функции $F_j(u)$, $S_j(u)$, $j = 1, 2$ — аналитические на S^2 , следовательно, и функции $f_{\alpha_1, \alpha_2}(u) = D^\alpha f(u)$, $|\alpha| = \alpha_1 + \alpha_2 \leq 4$, $s_{\beta_1, \beta_2}(u) = D^\beta s(u)$, $|\beta| = \beta_1 + \beta_2 \leq 3$ также аналитические на S^2 . Пусть $\tilde{f}_{\alpha_1, \alpha_2}(z, \varsigma)$ и $\tilde{s}_{\beta_1, \beta_2}(z, \varsigma)$ — аналитические продолжения функций $f_{\alpha_1, \alpha_2}(u)$, $s_{\beta_1, \beta_2}(u)$ в замкнутую область

$$E_\rho^{\delta, u_0} = E_\rho^{\delta, \theta_0} \times E_\rho^{\delta, \varphi_0},$$

ограниченные в этой области некоторой константой $c > 0$. Здесь $E_\rho^{\delta, \theta_0}$ — эллипс в комплексной плоскости $z = \theta + i\eta$ с фокусами в точках $(\theta_0 \mp \delta, 0)$, и $E_\rho^{\delta, \varphi_0}$ — эллипс в комплексной плоскости $\zeta = \varphi + i\tau$ с фокусами в точках $(\varphi_0 \mp \delta, 0)$ ($\eta = \text{const}$ и $\tau = \text{const}$), причём отношение полусуммы длин осей этих эллипсов к числу $\delta > 0$ равно $\rho > 1$. Возьмём такое достаточно большое ρ ($\delta > 0$ — фиксировано) и r , $1 < r < \sqrt[3]{\rho}$, чтобы замкнутая область E_ρ^{δ, u_0} содержала внутри себя прямоугольник $\{0 \leq \theta \leq \pi, 0 \leq \varphi \leq 2\pi\}$. Применяя к функциям $\tilde{f}_{\alpha_1, \alpha_2}(z, \varsigma)$ и $\tilde{s}_{\beta_1, \beta_2}(z, \varsigma)$ лемму А.Г. Витушкина [3, с. 197], заключаем, что всюду в E_ρ^{δ, u_0} функции $\tilde{f}_{\alpha_1, \alpha_2}(z, \varsigma)$, $\tilde{s}_{\beta_1, \beta_2}(z, \varsigma)$ будут ограничены по модулю величинами

$$M_{\alpha_1, \alpha_2} \varepsilon^{1-\gamma} \left(\ln \frac{1}{\varepsilon}\right)^4, \quad N_{\beta_1, \beta_2} \varepsilon^{1-\gamma} \left(\ln \frac{1}{\varepsilon}\right)^4,$$

где $\gamma = \frac{2 \ln r}{\ln \rho - \ln r}$, а $M_{\alpha_1, \alpha_2} > 0$, $N_{\beta_1, \beta_2} > 0$ — некоторые постоянные, не зависящие от ε и r , причём в силу выбора $r < \sqrt[3]{\rho}$ величина $\varepsilon^{1-\gamma} \left(\ln \frac{1}{\varepsilon}\right)^4 \rightarrow 0$ при $\varepsilon \rightarrow 0$.

Так как $\tilde{f}_{\alpha_1, \alpha_2}(z, \varsigma)$, $\tilde{s}_{\beta_1, \beta_2}(z, \varsigma)$ — аналитические продолжения функций $f_{\alpha_1, \alpha_2}(u)$, $s_{\beta_1, \beta_2}(u)$, то тем более для всех $u \in S^2$

$$|D^\alpha f(u)| \leq M_{\alpha_1, \alpha_2} \varepsilon^{1-\gamma} \left(\ln \frac{1}{\varepsilon}\right)^4, \quad 0 \leq |\alpha| \leq 4, \quad u \in S^2,$$

$$|D^\beta s(u)| \leq N_{\beta_1, \beta_2} \varepsilon^{1-\gamma} \left(\ln \frac{1}{\varepsilon} \right)^4, \quad 0 \leq |\beta| \leq 3, \quad u \in S^2.$$

Пусть Q — любое множество ненулевой меры на сфере S^2 . Так как поверхности B_1 и B_2 — аналитические, то их поверхностные функции $\sigma_j(Q)$, $j = 1, 2$, можно записать в виде [6, с. 18]:

$$\sigma_j(Q) = \int_Q R^{(j)}(u) d\omega_v, \quad R^{(j)}(u) = R_1^{(j)}(u) R_2^{(j)}(u).$$

Поэтому

$$\begin{aligned} |\sigma_2(Q) - \sigma_1(Q)| &= \left| \int_Q [R^{(2)}(u) - R^{(1)}(u)] d\omega_v \right| \leq \text{mes } Q \max |R^{(2)}(u) - R^{(1)}(u)| \leq \\ &\leq 4\pi \|R^{(2)}(u) - R^{(1)}(u)\|_{C(\omega)}. \end{aligned}$$

Следовательно, ввиду неравенства (2) и условий

$$\|F_2(u) - F_1(u)\|_{C^4(S^{n-1})} \leq \varepsilon, \quad \|S_2(u) - S_1(u)\|_{C^3(S^{n-1})} \leq \varepsilon,$$

получаем

$$|\sigma_2(Q) - \sigma_1(Q)| \leq c\varepsilon,$$

где $c > 0$ — некоторая постоянная. Применяя теперь теорему Ю.А. Волкова [4] об устойчивости решения проблемы Минковского, получаем, что при некотором расположении поверхностей B_1 и B_2 разность их опорных функций удовлетворяет неравенству

$$|H_2(u) - H_1(u)| \leq A_2 \varepsilon^{\frac{1-\gamma}{5}} \left(\ln \frac{1}{\varepsilon} \right)^{\frac{4}{5}}, \quad u \in S^2.$$

■

ЛИТЕРАТУРА

1. Бейтмен Г., Эрдейи А. Высшие трансцендентные функции. Том 2. М. : Наука, 1974. 296 с.
2. Бляшке В. Дифференциальная геометрия и геометрические основы теории относительности Эйнштейна. Москва, Ленинград : ОНТИ, 1935. 302 с.
3. Витушкин А.Г. Оценка сложности задачи табулирования. М. : Физматгиз, 1959. 215 с.
4. Волков Ю.А. Устойчивость проблемы Минковского // Вестн. ЛГУ. 1963. № 1. С. 33–43.
5. Погорелов А.В. Четвёртая проблема Гильберта. М. : Наука, 1974. 78 с.
6. Погорелов А.В. Многомерная проблема Минковского. М. : Наука, 1975. 96 с.
7. Соболев С.Л. Введение в теорию кубатурных формул. М. : Наука, 1974. 808 с.

8. Степанов В.Н. Некоторые вопросы геометрии выпуклых поверхностей. Исследование корректности обратных задач и некоторых операторных уравнений // Сб. научн. тр. Новосибирск : ВЦ СО РАН СССР. 1981. С.65–75.
9. Степанов В.Н. Асимптотика собственных значений уравнения первого рода // Омский научный вестник. 2011. № 3(103). С. 41-44.
10. Bourgain J., Lindenstrauss J. Projection bodies // Geometric Aspects of functional analysis (J. Lindenstrauss and V.D. Milman, eds) Lecture Notes in Math., vol. 1317. Springer-Verlag. New York: 1988. P. 250–270.
11. Campi S. Reconstructing a convex surface from certain Measurements of its projections // Bull. Un. Math. Ital. 1986. No. 5B, 6. P. 945–959.
12. Campi S. Recovering a centred convex body from the areas of its shadows: a stability estimate // Ann. Mat. Pura ed appl. 1988. No. 151. P. 289–302.
13. Campi S., Colesanti A., Gronchi P. Convex bodies with extremal volumes having prescribed brightness in finitely many directions // Geom. Dedicata. 1995. No. 57. P. 121–133.
14. Gardner R., Milanfar P. Reconstruction of convex bodies from brightness functions // Discrete and Computational Geometry. January 2003. Vol. 29, Issue 2. P. 279–303.
15. Gardner R., Kiderlen M., Milanfar P. Convergence of algorithms for reconstructing convex bodies and directional measures // The Annals of Statistics. 2006. Vol. 34, No. 3. P. 1331–1374.
16. Goodey P., Groemer H. Stability results for first order projection bodies // Proc. Amer. Math. Soc. 1990. 109, No. 4. P. 1103–1114.
17. Groemer H., Geometric Applications of Fourier Series and Spherical Harmonics. Encyclopedia of Mathematics and its Applications, vol. 61. Cambridge University Press: Cambridge 1996. 329 c.
18. Groemer H. Stability results for convex bodies and related spherical integral transforms // Adv. Math. 1994. No. 109. C. 45–74.
19. Hug D., Schneider R. Stability results involving surface area measures of convex bodies, part II // Rend. Circ. Mat. Palermo (2) Suppl. 2002. No. 70. P. 21–51.
20. Schneider R., Weil W. Über die Bestimmung eines Konvexen Körpers durch die Inhalte seiner Projektionen // Math. Z. 1970. 116, No. 4. P.338–348.
21. Schneider R. Stability in Aleksandrov-Fenchel-Jenssen theorem // Mathematika. 1989. No. 36. P. 50–59.
22. Schneider R. Stable Determination of Convex Bodies From Projections // Monatsh. Math. 2007. No. 150. P. 241–247.

THE EVALUATION OF STABILITY FOR CONVEX SURFACES

V.N. Stepanov

Ph.D. (Phys.-Math.), Associate Professor, e-mail: stpnv@yandex.ru

Omsk State Technical University

Abstract. Necessary and sufficient conditions for the analytic solution of the first kind equation of convolution type on the sphere with the core $K(\langle u, v \rangle) \in L_2[-1, 1]$ are given in the article. Also, some functional analyticity of closed convex surface is

proved. The evaluation for the difference between the support functions of the two closed convex analytic surfaces is obtained by Vitushkin's Lemma, if the deviations of surfaces functionals and their derivatives on a finite set of points are known. Some known results on the stability problem in the definition of a convex surface on the tomographic data are given.

Keywords: the equation of the first kind, analyticity, supporting function, convex surface, stability, Vitushkin's Lemma.

НЕ-ГЁДЕЛЕВСКАЯ МАШИНА ВРЕМЕНИ

А.К. Гуц

профессор, д.ф.-м.н., зав. каф. кибернетики, e-mail: aguts@mail.ru

Омский государственный университет им. Ф.М. Достоевского

Аннотация. Описывается квантовый принцип работы машины времени, не использующий времениподобные гладкие петли в лоренцевом многообразии. Предлагаемая машина времени основана на разрушении интерференции квантовой суперпозиции состояний в суперпространстве Уилера.

Ключевые слова: машина времени, не-гёделевский принцип работы, выделение нужной исторической эпохи, суперпространство Уилера.

Введение

В 1949 году знаменитый логик Курт Гёдель описал принцип работы машины времени в рамках общей теории относительности [1]. Для перехода человека в прошлое используются времениподобные гладкие замкнутые кривые (временные петли), которые может содержать пространство-время, являющееся 4-мерным лоренцевым многообразием.

Все конструкции машины времени, обеспечивающие переход человека в прошлое и изучаемые до настоящего времени, используют принцип Гёделя. В основе идеи машины времени Гёделя лежит *аксиома об абсолютном 4-мерном пространстве-времени*, все события которого, будь то события прошлого, настоящего или будущего, всегда *равно присутствуют* в бытии, присутствуют в пространстве-времени. Поэтому в прошлое можно совершить путешествие, используя идущие туда временные петли.

Существующие проекты *квантовых машин времени* ограничиваются получением экземпляров вещей в том виде, какими они были в прошлом или будут в будущем [13, 14]. Они не используют механизм Гёделя, а используют разрушение квантовой суперпозиции. Однако, что важно отметить, переход в прошлое квантовая машина времени осуществляет в объемлющем пространстве-времени.

В данной статье предлагается новый способ перемещения человека в прошлое, который основан на использовании квантовой интерференции, но осуществляется оно не в пространстве-времени, а в суперпространстве Уилера. При этом прошлым считаются не топологически близкие события в абсолютном пространстве-времени, а некоторая «волна» в суперпространстве Уилера, образующая с некоторыми другими подобными «волнами» интерферирующую квантовую суперпозицию. Поэтому предлагаемый механизм путешествия в прошлое называется *не-гёделевской машиной времени*.

1. Исторические эпохи

Суперпространство Уилера состоит из всевозможных 3-мерных римановых геометрий ${}^{(3)}\mathcal{G} = \langle M^3, h_{\alpha\beta} \rangle$. Соответствующая квантовая теория в Суперпространстве называется квантовой геометродинамикой. Подробно эта теория изложена в [2, 3].

Предлагаемый принцип работы машины времени основывается на следующей аксиоме.

АКСИОМА 1. Мир существует в форме исторических эпох. Каждая историческая эпоха — это волна вероятности 3-геометрий ${}^{(3)}\mathcal{G}$ в суперпространстве Уилера.

Каждая историческая эпоха — это внешний мир, окружающий людей, совокупность бытующих среди людей представлений о внешнем мире; это соответствующие этим представлениям наука, культура и искусство. Это типы вооружения, одежда, мода, транспорт, манера письма, повседневная лексика и т.д.

Примеры.

1. Древний Египет.
2. Древняя Греция.
3. Европа от начала XIX века до начала XX века
4. Европа от начала Первой мировой войны до конца XX века

Историческая эпоха Ω_k — это «замороженное» бытие людей. Изменения в жизни отсутствуют в каждой конкретной исторической эпохе на протяжении всего времени её существования, точнее, всей длительности эпохи. Это гештальт Гёте [5, 6] или гештальт Шпенглера [7].

Исторические эпохи $\Omega_k, k \in K$ реальны, они существуют, они являют различные состояния $|\Omega_k\rangle$, различные грани единого квантового кристалла Менского [8, с. 177], [1, с. 96]:

$$\sum_{k \in K} c_k |\Omega_k\rangle, \quad c_k \in \mathbb{C}, \quad (1)$$

который представляет полное бытие человечества во все времена.

Историческая эпоха — волна стационарности, волна неизменности в абстрактном бесконечномерном историческом пространстве.

Эта волна в квазиклассическом приближении имеет вид

$$\begin{aligned} \Psi(\Omega_k) &\equiv \Psi_k({}^{(3)}\mathcal{G}, \mu, B, e, \sigma, \nu) = \\ &= A_k \left(\begin{array}{c} \text{медленно меняющаяся} \\ \text{амплитудная функция} \end{array} \right) e^{-\frac{i}{\hbar} S_k({}^{(3)}\mathcal{G}, \mu, B, e, \sigma, \nu)}, \end{aligned} \quad (2)$$

где ${}^{(3)}\mathcal{G}$ — геометрия пространства, вмещающего человеческую цивилизацию, пребывающую в конкретной исторической эпохе, μ — её материально-физическое устройство, B — состояние её биосферы на «тверди земной», e —

этническая организация, σ — государственное устройство и, наконец, ν — учёт разумности людей, S_k — действие, удовлетворяющее уравнению Эйнштейна-Гамильтона-Якоби [2, 3].

$$G_{\alpha\beta\gamma\delta} \left(\frac{\delta S_k}{\delta h_{\alpha\beta}} \right) \left(\frac{\delta S_k}{\delta h_{\gamma\delta}} \right) - \sqrt{h} {}^{(3)}R + \mathcal{E}(h_{\alpha\beta}, \mu, B, e, \sigma, \nu) = 0, \quad (3)$$

где $\mathcal{E}(h_{\alpha\beta}, \mu, B, e, \sigma, \nu)$ — член, учитывающий вклад материальных источников μ , окружающей среды (природы) B и полей e , σ и ν [2].

Сами волновые функции (2) являются решением уравнения ДеВитта-Уилера:

$$\left(\alpha\beta\gamma\delta \frac{\delta}{\delta h_{\alpha\beta}} \frac{\delta}{\delta h_{\gamma\delta}} + \sqrt{h} {}^{(3)}R + \mathcal{E}(h_{\alpha\beta}, \mu, B, e, \sigma, \nu) \right) \Psi_k({}^{(3)}\mathcal{G}, \mu, B, e, \sigma, \nu) = 0. \quad (4)$$

Как видим, квантовая геометродинамика не содержит никакого временного параметра, она *вневременна*.

2. Исторические последовательности

Исторические эпохи — это различные волны стационарности, волны вероятности 3-геометрий в суперпространстве. Классическое пространство-время M^4 возникает из определённой совокупности исторических эпох $\{\Omega_k\}_{k \in A}$, т.е.

$$\Psi(M^4) = \sum_{k \in K} c_k \Psi(\Omega_k), \quad c_k \in \mathbb{C} \quad (5)$$

посредством интерференции.

Интерференция исторических эпох даёт в историческом пространстве разные «цепи горных пиков», каждая из которых представляет изменчивую эволюционирующую историческую последовательность, вселенную, в которой течёт объективное время и которая известна под названием пространство-время.

«Цепи горных пиков» возникают для волн-эпох, у которых совпадают фазы, т.е. когда

$$S_k = S_{k'} = \dots = S.$$

Тогда

$$c_k \Psi(\Omega_k) + c_{k'} \Psi(\Omega_{k'}) + \dots =$$

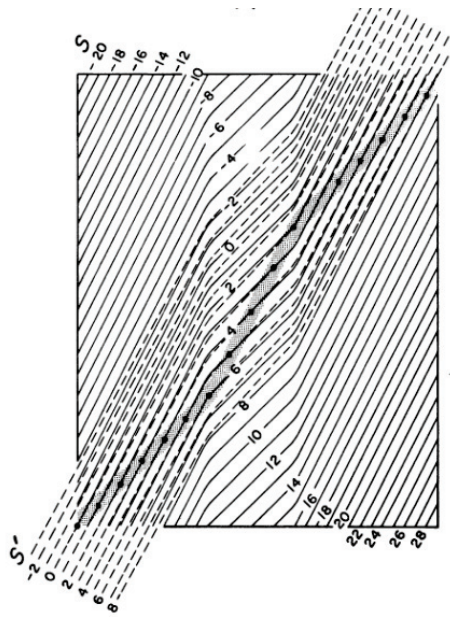


Рис. 1. Появление классического пространства-времени в результате интерференции: числа -20, -18, ..., +28, ... — пронумерованы гребни эпохи α . Пунктирные линии — гребни эпохи Ω_k .

Заштрихованная область есть область интерференции (волновой пакет), в которой усиливаются амплитуды волн-эпох. Чёрные точки отмечают классическую траекторию в историческом пространстве, являющуюся классическим пространством-временем [3]

$$= c_k A_k e^{\frac{i}{\hbar} S_k} + c_{k'} A_{k'} e^{\frac{i}{\hbar} S_{k'}} + \dots = [c_k A_k + c_{k'} A_{k'} + \dots] e^{\frac{i}{\hbar} S}.$$

Формула показывает, что каждая эпоха вносит свой вклад в историю эволюционирующей пространственно-временной вселенной. Из неё видно, как «горные пики» складываются из разных интерферирующих эпох. Благодаря этому втиснутые в единое пространство-время наблюдатели рассуждают о наблюдаемых сменах исторических эпох, помнят своих предков, раскапывают исторические артефакты и прочее. При этом каждый из этих наблюдателей принадлежит конкретной исторической эпохе Ω_k , являющейся состоянием квантовой системы Ω , которая есть не что иное как вселенная, появляющаяся как интерференция в форме пространства-времени (цепи «горных пиков»).

Изменить своё пространство-время можно. Но для этого надо «соскользнуть» из своей «цепи горных пиков». Иначе говоря, изменить свою фазу S_k .

Какова разница между исторической последовательностью и исторической эпохой с точки зрения сущности человеческого бытия? Это различие между разумом и рассудком:

Разум имеет дело со становящимся, рассудок — со ставшим. Первый не беспокоится о вопросе: к чему? Второй не спрашивает: откуда? Разум радуется развитию. Рассудок для использования в дальнейшем желает все закрепить (Гёте, [5, с. 346]).

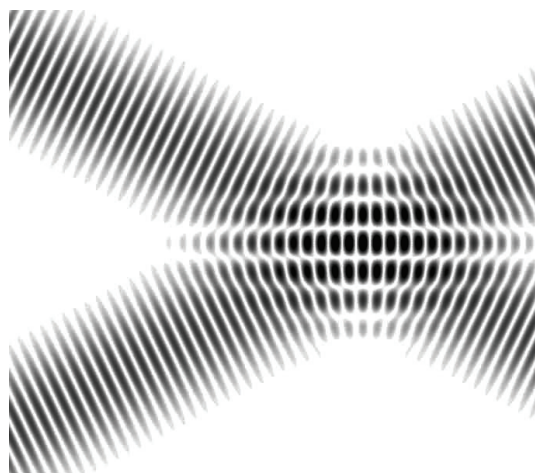


Рис. 2. Интерференция двух плоских волн. Видны параллельные исторические вселенные-реальности = несколько «цепей горных пиков», по выражению Halliwell'a [10, p. 180].

В самом деле, конкретный человек принадлежит конкретной исторической эпохе, и для него сведения о временах былых или грядущих — всего лишь культурная традиция.

3. Не-гёделевская машина времени

Суперпозиция исторических эпох (5) дана нам в форме интерференции, которая являет собой классическое пространство-время M^4 и в котором мы существуем.

Геометрия ${}^{(3)}\mathcal{G}$ исторической эпохи Ω_k знает место своей *временной локализации* в 4-геометрии пространства-времени: "гиперповерхность, проведённая через пространство-время и состоящая из 3-геометрий, может сдвигаться во времени в любой из точек слишком незначительно, чтобы включить какую-либо новую 3-геометрию. Время, определённое таким образом, означает не более и не менее как локализацию 3-геометрии в 4-геометрии. В этом смысле 3-геометрия выступает как «носитель временной информации» [3, p. 37].

Если создать устройство, которое *выделяет* в суперпозиции (5) конкретную историческую эпоху на фоне шума, который состоит из всех остальных исторических эпох, то мы окажемся в выделенной исторической эпохе, скажем в эпохе Древнего Египта. Действительно, известно, что интерференционная картина, создаваемая квантовым объектом, исчезает, если *производится измерение* состояния этого квантового объекта или имеет место декогеренция за счёт взаимодействия со средой. В нашем случае трудно говорить о какой-либо иной среде, чем совокупность сознаний людей, погружённых в пространство-время.

Эпоха Ω_k — это сигнал в суперпространстве. Его надо отделить от всех других сигналов с той же фазой, с которыми он образует наше пространство-время, и выделить из фона всех прочих сигналов с другими фазами.

Иначе говоря, имеем аналог задачи из статистической радиофизики, которую следует решать, разрабатывая соответствующую теорию «пространственно-временного» обнаружения «сигналов» в модельных минисуперпространствах.

Что нужно измерять, чтобы выделить конкретный сигнал (2)? То, от чего он зависит. А зависит он от геометрии пространства ${}^{(3)}\mathcal{G}$, вмещающего человеческую цивилизацию, пребывающей в соответствующей исторической эпохе, от состояния материи μ , от состояния биосферы B , от этнической организации e , от состояния социума σ , и наконец, от сознания людей ν . Последнее, как отмечалось выдающимися физиками Шрёдингером и Вигнером, играет существенную роль в понимании многих квантовых эффектов (кот Шрёдингера, друг Вигнера).

Конечно, выделение, замер исторической эпохи, естественно, должен производиться в ограниченном пространстве V . Именно в этом ограниченном пространстве и запускается машина времени. В области V и проявляется мир исторической эпохи; вне его интерференционная картина остаётся неизменной — там мир продолжает существовать в форме сменяющихся исторических эпох, т.е. как историческая последовательность.

Переход в прошлую эпоху совершается за счёт измерений геометрии (и не только геометрии, поскольку квантовая теория использует ещё и наблюдателей, т.е. моменты сознания людей). Мы очень хорошо понимаем, как измеряется геометрия. В рамках идеологии общей теории относительности измерение геометрии — это измерение гравитационных полей. Каждую историческую эпоху можно рассматривать как гравитационную волну с фиксированными параметрами, аппаратура обнаружения которой более или менее разработана.

Обозначим через $|k\rangle$ состояние исторической эпохи Ω_k с волновой функцией $\Psi_k({}^{(3)}\mathcal{G})$ квантовой системы Ω , описываемой суперпозицией

$$\sum_{k \in K} c_k |k\rangle.$$

Тогда измерение, производимое наблюдателем X с помощью аппарата $\mathcal{A}$ с начальным состоянием A , нацеленное на специфическое значение 3-геометрии ${}^{(3)}\mathcal{G}'$ и присущее только исторической эпохе $\Omega_{k'}$, приводит к новой суперпозиции:

$$\sum_k c_k |k\rangle \otimes |A\rangle \rightarrow \left(\sum_{k, k \neq k_0, k'} c_k |k\rangle \right) \otimes |\tilde{A}\rangle + (c_{k_0} |k_0\rangle \otimes |A_0\rangle + c_{k'} |k'\rangle \otimes |A'\rangle). \quad (5)$$

Мы видим, что две эпохи $\Omega_{k_0}, \Omega_{k'}$ сцепливаются (переплетаются) с аппаратурой (окружением), а все прочие образуют интерференционную квантовую суперпозицию, намечающую классическую Вселенную, в которой отсутствует вклад эпох $\Omega_{k_0}, \Omega_{k'}$.

Пусть наблюдатель X в исторической эпохе Ω_{k_0} создаёт условия, которые осуществляют измерения геометрии (или иной величины), соответствующей другой эпохе $\Omega_{k'}$. Тогда происходит локализации части пространства эпохи $\Omega_{k'}$ в пространстве эпохи Ω_{k_0} наблюдателя X , благодаря тому, что возникает квантовая сцепленность (квантовая корреляция) двух эпох.

Другими словами, мы имеем коллапс пакета (1) в эпоху-реальность Ω_{k_0} , проявляющийся в локализации части пространства эпохи $\Omega_{k'}$ в реальности Ω_{k_0} . Открывается проход в иную эпоху.

Однако симметричным образом, поскольку эпоха $\Omega_{k'}$ — это такая же объективная реальность, каковой является эпоха Ω_{k_0} , а также в силу того, что измерение, производимое наблюдателем эпохи Ω_{k_0} , по сути дела двустороннее, т.е. является *взаимодействием* (благодаря их квантовой сцепленности), происходит локализация части пространства эпохи Ω_{k_0} в эпохе-реальности $\Omega_{k'}$. Грубо говоря, наблюдатель X оказывается в другой исторической эпохе вместе со своей аппаратурой. Мы имеем то, что называется машиной времени. Будем называть её *не-гёделевской*, поскольку не идёт речь о замкнутых временных петлях, придуманных Гёделем.

Закрывается ли проход из одной исторической эпохи в другую? Да, закрывается. Для этого наблюдателю X не нужно даже выключать свою аппаратуру, поскольку согласно эффекту Зенона, непрерывное измерение может привести к тому, что динамика становится тривиальной, по существу, приведёт к исчезновению динамики. Да и сам процесс локализации макрообъектов крайне скоротечен.

Какая геометрия отвечает конкретной исторической эпохе Ω_k ? Поскольку $|\Psi(\Omega_k)|^2$ — это плотность вероятности, то следует считать, что исторической эпохе отвечают наиболее вероятные геометрии ⁽³⁾ $\mathcal{G}$, подсказываемые волновой функцией. Именно на фиксацию этих геометрий и должна быть настроена измерительная аппаратура. Возможно это не просто измерительная аппаратура, а устройства, «*входящие в резонанс*» с геометрией интересующей нас исторической эпохи.

Обратим внимание, что переходы носят вероятностный характер. Другими словами, оказаться в нужной исторической эпохе можно далеко не всегда. Переход более или менее надёжно произойдёт лишь в том случае, когда число $|c_{k'}|^2$ близко к 1.

Поскольку переход совершается только с вероятностью $|c_{k'}|^2$, то в случае малости величины этой вероятности по закону больших чисел он не происходит. Иначе говоря, путешествия во времени крайне редки, и в силу этого, мы не наблюдаем путешественников во времени в нашей эпохе.

Как оценить число $|c_{k'}|^2$? Каждая историческая эпоха — это доминирующая культура, доминирующий суперэтнос. Сколько таких суперэтносов известно в человеческой истории? Древний Китай, Древняя Индия, якуты в XVII веке, племена папуасов на острове Новая Гвинея, описанные Миклухо-Маклаем, — примеры таких культур, таких исторических эпох. Их явно не менее тысячи, поэтому при равномерном распределении исторических эпох $|c_{k'}|^2 \sim 10^{-6}$. Мы имеем явно маловероятные события, говорящие о том, что переходы во времени невозможны. Но вполне возможно, что существуют наиболее вероятные эпохи. Они-то и могут быть местами повального туризма путешественников во времени. Но в таких эпохах к ним относятся, как к недавнему навалу русских в Турцию. Наша же эпоха явно не Анталия. Хотя, и такие поездки рискованны: кому захочется отправиться в путешествие, конечное место назначения

которого точно указать невозможно. Например, вместо того, чтобы комфортно наблюдать Пушкина в Летнем саду можно оказаться в «Парке юрского периода».

Но «настройкой» аппаратуры на нужную геометрию, пожалуй, не обойтись. Во-первых, квантовая механика сильно связана с проблемой сознания [8], а во-вторых, исторические эпохи не пустые геометрические миры, а миры, заселённые людьми, и в силу этого историческая эпоха — это ещё и энергия этносов, политические страсти и прочее. Геометрия такова, какой она нужна людям, а квантовые интерференционные картины разрушаются не только в силу проведённых измерений, но и в силу самих намерений проводить такие измерения.

Но прежде чем говорить о намерениях, т.е. о людях, следует вспомнить о природе, окружающей человека.

Понятие биосферы, которая присутствует в волновой функции, было введено В.И. Вернадским, и с ним связано представление о геохимической энергии. *Геохимическая энергия* любого существа может быть определена по его способности распространиться на расстояние, равное земному экватору, и покрыть поверхность Земли тонкой плёнкой при условии беспрепятственного размножения. Такое условие, естественно, не может быть выполнено, и поэтому геохимическая энергия живого может рассматриваться лишь как некий условный показатель.

Поэтому историческая эпоха может быть охарактеризована определённым уровнем геохимической энергии, отвечающим людям этой эпохи.

Существует также понятие пассионарной энергии в рамках теории этногенеза Л.И. Гумилёва. Пассионарность, *пассионарная энергия* — избыток биохимической энергии живого вещества, порождающий жертвенность людей, часто ради высоких целей. Пассионарность — это непреодолимое внутреннее стремление к деятельности, направленной на изменение своей жизни, окружающей обстановки, статуса-кво. Пассионарность изменяется по количеству значимых исторических событий, приходящихся на фиксированный отрезок времени.

Как характеризовать и как изменять состояние социума или состояния сознания, сказать не можем, хотя определённые соображения в этом направлении имеются.

Ясно, что выделение сигнала Ω_k , а значит и осуществление перехода в прошлое, крайне трудная и, пожалуй, фантастическая задача. В нашей исторической эпохе скорее всего неосуществимая.

4. Заключение

Предложенный принцип работы машины времени опирается на квантовую теорию, обобщающую классическую физику. Главными понятиями квантовой теории являются понятие когерентной суперпозиции возможных состояний физической системы и понятие интерференции входящих в суперпозицию физических состояний системы. Именно они в рамках квантовой гравитации де Витта-Уилера использованы для объяснения механизма перехода в прошлое. В ре-

зультате исчезает необходимость преодолевать парадокс бабушки, поскольку исторические эпохи представляют собой абсолютно независимые реальности.

В 1968 году академик А.Д. Александров поставил проблему изучения физических условий, при выполнении которых становится возможным переход в прошлые исторические эпохи [11, 12]. Но он имел в виду единственный тогда известный механизм — гёделевскую машину времени. Поэтому при решении проблемы Александрова ограничивались изучением лоренцевых многообразий, в которых имелись замкнутые времениподобные гладкие кривые.

Этот путь решения проблемы Александрова привёл к парадоксальной ситуации. То, что сделало возможным сам механизм Гёделя — а это абсолютность пространства-времени — становилось препятствием на пути реализации гёделевской машины времени, поскольку требовалось переделывать то, что считалось абсолютным и неизменным (см. [12]).

Естественным было попытаться оставить общую теорию относительности в её классическом исполнении, т. е. в рамках классической физики, и воспользоваться более совершенной и более глубокой квантовой теорией гравитации.

Идея использовать квантовую интерференцию для переходов во времени впервые была изложена в статье [13, 14]. Но в ней рассуждения основываются на использовании эффекта замедления времени для отражённого фотона при его движении в гравитационном поле и образующего суперпозицию с исходным фотоном, движущимся прямо вне гравитационного поля. Речь идёт об известном эксперименте пропускания фотона через полупрозрачное зеркало. Появляются два фотона, которые составляют квантовую суперпозицию и интерферируют друг с другом. В общем случае рассматривается суперпозиция эволюционирующих различных состояний системы, отличающихся степенью замедления времени, достигнутой за счёт гравитационного поля. Рассматривая ситуацию, когда этих состояний порядка 10000, устанавливается, что средний эффект замедления времени для системы относительно времени внешнего наблюдателя намного больше, чем для каждого отдельного состояния. Измерение состояния системы обнаруживает его с некоторой вероятностью таким, каким оно было в прошлом. Это устройство назовём AARV-машиной по первым буквам фамилий авторов статьи [13].

В какой-то мере, это квантовая вариация машины времени Торна, эффективность которой крайне сомнительна [2]. Но в отличие от машины времени Торна, которую считают устройством для путешествия людей в прошлое, квантовая AARV-машина для этого не предназначена. AARV-машина работает в рамках пространства-времени и остаётся в нем, в отличие от не-гёделевской машины времени, предложенной в данной статье.

Наш проект квантовой машины времени также содержит моменты, которые вызывают возражение и сомнение. Принятая нами аксиома, утверждающая, что пространство-время, вселенная — это результат квантовой интерференции исторических эпох, бесспорна в части, говорящей об интерференции.

Другая часть — это предположение о вечных исторических эпохах. Ясно, что данная гипотеза заменяет гипотезу о вечно существующих событиях прошлого, принадлежащих абсолютному пространству-времени, к которым устрем-

ляется путешественник, управляющий гёделевской машиной времени. Мы отдельные события заменяем совокупностью событий целой исторической эпохи. По сути дела, это лишь более широкое предположение — компактный пучок мировых линий вырывается из единого пространства-времени и рассматривается как нечто целостное и независимое, существующие вечно, стационарно, развивающееся самоподобно и обнаруживаемое нами через интерференцию с другими эпохами в каждом моменте времени исторической последовательности как историческое наследие.

Тем не менее, допущение о, в принципе, вечных исторических эпохах как-то с трудом укладывается в голове. Но... «есть многое в природе, друг Горацио, что и не снилось нашим мудрецам».

ЛИТЕРАТУРА

1. Gödel K. An Example of a New Type of Cosmological Solutions of Einstein's Field Equations of Gravitation // *Rev. Mod. Phys.* 1949. V.21, no.3. P.447-450.
2. Гуц А.К. Физика реальности. Омск: Изд-во КАН, 2012. 424с.
3. Уилер Дж. Предвидение Эйнштейна. М.: Мир, 1970.
4. Гуц А.К. Многовариантная Вселенная и теория исторических последовательностей // *Математические структуры и моделирование*. 2012. № 25. С. 70–80.
5. Гёте И.-В. Избранные философские произведения. М.: Наука, 1964.
6. Гёте И.-В. Избранные сочинения по естествознанию. М.: Изд-во АН СССР, 1957.
7. Шпенглер О. Закат Европы. Очерки морфологии мировой истории. М. 1993.
8. Менский М. Человек и квантовый мир. Фрязино: «Век2», 2005.
9. Лебедев Ю.А. Многоликое мироздание. Эвереттическая аксиоматика. – М.: Фирма ЛеЖе, 2009.
10. Halliwell J.J. Introductory lectures on quantum cosmology // In: *Quantum cosmology and baby universes* / Eds. S. Coleman, J.B. Hartle, T. Pien and S. Weinberg. World Scientific Publishing Co. Pte. Ltd., 1991. P. 159–244.
11. Гуц А.К. Аксиоматики А.Д. Александрова для квантовой механики и теории относительности // *Вестник Новосибирского государственного университета. Серия: математика, механика, информатика*. 2012. Т. 12, № 3. С. 19–30.
12. Гуц А.К. Элементы теории времени. М.: УРСС, 2011.
13. Aharonov Y., Anandan J., Popescu S., Vaidman L. Superposition of Time Evolutions of Quantum System and a Quantum Time Translation Machine // *Physical Review Letters*. 1990. V. 64. P. 2965.
14. Vaidman L. A Quantum Time Machine // *Foundations of Physics*. 1991. V. 21, No. 8. P. 947–958.

NON-GÖDEL TIME MACHINE**A.K. Guts**

Dr.Sc. (Phys.-Math.), Professor, e-mail: aguts@mail.ru

Dostoevsky Omsk State University

Abstract. The new quantum principle of a time machine not using a smooth timelike loops in Lorentz manifolds is described. The proposed time machine is based on the destruction of interference of quantum superposition states in the Wheeler superspace.

Keywords: Time machine, non-Gödel principle, historical epoch, Wheeler superspace.

СПЕЦИФИКА МОДЕЛИРОВАНИЯ ЭВЕРЕТТИЧЕСКИХ ПРОСТРАНСТВ

Ю.А. Лебедев

к.т.н., доцент, e-mail: lebedev@bmstu.ru

Московский государственный технический университет им. Н.Э. Баумана

Аннотация. Эвереттические пространства, как интуитивно-образные модели действительности с точки зрения многомировой интерпретации квантовой механики, до настоящего времени не имеют адекватного математического описания. Это во многом связано с тем, что отсутствует математический аппарат для описания специфики психоидного аспекта действительности. В работе предлагается по сути фрактально-суперпозиционный подход к формулировке основных понятий, необходимых для построения пригодной для конкретных вычислений математической модели. По очевидной причине аналогом ядра фрактала, описывающего психоидность, выбрано фундаментальное понятие квантовой механики — волновая функция.

Ключевые слова: эвереттика, квантовая механика, соотнесённое состояние, физичность, психоидность, эвереттические пространства, мультиверс, альтерверс, моделирование действительности, эвереттиан.

1. Эвереттические пространства

Эвереттические пространства — это области существования соотнесённых состояний физического и психоидного полюсов каждого универса мультиверса. Действительные сущности (соотнесённые состояния) в темпоральных взаимодействиях порождают альтерверсальные структуры универсов. Каждое эвереттическое пространство содержит совокупность всех возможных историй эволюции начального состояния, совместимых с физическими и психоидными законами данного универса.

1.1. Геометризованные эвереттические пространства

Рассмотрим возможные варианты геометризованных представлений эвереттических пространств — множеств отображений соотнесённых состояний целостных универсов. Качество целостности универса позволяет не рассматривать физически разрушительных последствий декогеренции при эвереттических ветвлениях. Как было замечено М.Б. Менским, «декогеренции в этом случае не происходит, потому что у квантового мира как целого нет никакого окружения, которое могло бы вызвать декогеренцию» [1, с. 16].

1.1.1. Формализация понятия «эвереттовский мир». Малые эвереттианы

Ранее была предложена одна из конструкций эвереттических пространств, акцент при построении которой делался на учёт эволюции событийных характеристик бытия [2,3]. В этих работах рассмотрены конструкции пространств, являющихся метапространствами (в гёделевском смысле) по отношению к пространству-времени Минковского.

Поскольку категория времени относится к квантовомеханической характеристике соотнесённого состояния, которое включает индивидуальные для всякого процесса физические и психоидные составляющие, целесообразно рассмотреть формально атемпоральный математический конструкт соотнесённого состояния, содержащий время в неявной форме в виде параметра. Это позволит в дальнейшем рассматривать эволюцию в разных темпоральных системах отсчёта. Возможность «вмонтировать» в эволюцию различные темпоральные конструкции особенно важна для *новых* фундаментальных моделей, поскольку нельзя не согласиться с утверждением А.К. Гуца о том, что мейнстримная «современная теория времени — это сугубо *человеческая* теория времени» [4, с. 44]. Но вряд ли можно полагать, что человечность — единственный источник происхождения времени.

Если к тому же согласиться с утверждением Шекспира о том, что «всё сущее — это некая сцена» («All the world's a stage» [5]), то следует ожидать, что сцена для эвереттического бытия должна иметь весьма причудливую конструкцию.

Прежде всего, введём класс «двойственных сущностей» — малых эвереттианов $e_{v_{i+j}}$:

$$e_{v_{i+j}} = \Psi_{q_i} + \Omega_{p_j}. \quad (1)$$

Здесь:

Ψ_{q_i} — волновая функция квантовой реальности (КвР) универса,

Ω_{p_j} — волновая функция психоидной реальности универса (аналог Ψ_q в психоидной реальности),

q, p, i, j — идентификаторы (натуральные числа).

Величина $e_{v_{i+j}}$ названа эвереттианом, поскольку включает оба полюса эвереттической интерпретации соотнесённого состояния целостного мироздания — физический (Ψ_{q_i}) и психоидный (Ω_{p_j}) — и является представлением «эвереттовского мира» или, что то же самое, ветви «альтерверса по Менскому». Причины, по которым данный эвереттиан назван «малым», будут ясны из дальнейшего.

Сама природа малого эвереттиана квазиатемпоральна, поскольку ее структурные элементы Ψ_{q_i} и Ω_{p_j} описываются параметрами амплитуды и фазы, которые обладают физическим смыслом только вместе с понятием локально ньютоновского времени.

Следует сказать, почему q и p — натуральные числа. Такая конструкция предполагает, что множества состояний как квантовой, так и психоидной реальностей счётны, а множество малых эвереттианов — дискретно. Индексация

квантовых и психоидных волновых функций i и j индивидуализирует их в соответствии с гипотезой о наличии у них памяти. Понятие памяти волновой функции, введённое Эвереттом [6, р. 458], индивидуализирует как квантовые, так и психоидные состояния и позволяет выделять как физические объекты в квантовой реальности, так и психоидные субъекты в реальности психоидной.

Это соответствует «квантовой идеологии мироздания» и является развитием концепции Дж. Барбура о мироздании как «пинакотеке состояний» — хаотическом собрании вечных и неизменных «кадров», на которых запечатлены все возможные в данной ветви мультиверса состояния всех его элементов [7]. Отдельным вопросом является способ упорядочения барбуровской пинакотеки — множества малых эвереттианов.

В качестве возможных операций с эвереттианами допускается только операция сложения. Физическим смыслом сложения малых эвереттианов является создание их суперпозиции.

В результате для множества, включающего «а» эвереттианов, может быть получена максимальная суперпозиция вида:

$$\sum_{i=1}^{i=a} \Psi_{q_i} + \sum_{j=1}^{j=a} \Omega_{p_j} = e_{v_{a_i+a_j}}. \quad (2)$$

Кроме того, суммирование в (2) может проводиться по части малых эвереттианов из множества, содержащего «а» элементов, по множествам из (а-1), (а-2), ..., 2 элементов, причём набор слагаемых в каждом случае может быть различным. Таким образом, множество сумм вида (2) даст новое множество эвереттианов $e_{v_{a_i+a_j}}$.

Эвереттианы, как и комплексные числа, не могут быть упорядочены по критерию «больше — меньше». Единственным критерием упорядочения, имеющим ясный физический смысл, представляется критерий «раньше — позже».

Именно в процессе упорядочения, согласно Барбуру, сознание и порождает время. С эвереттической точки зрения активным элементом при создании любой действительности является психоидность «наблюдателя», поэтому способов упорядочивания, а, следовательно, и «времен», должно быть очень много. Можно предположить, что на низшем уровне психоидности (уровне «косной материи») при создании классических реальностей физического мира (КРФМ) разделяющее квантовые альтернативы сознание (в смысле, обычно употреблявшемся М.Б. Менским [1]) различает сущее на уровне малых эвереттианов (1). На более высоком уровне (уровне жизни) различие ощущается на уровне расширенного множества малых эвереттианов (2). Как эти ощущения синхронизируются, порождая временную упорядоченность того или иного сорта (в том числе и «равномерное математическое время» Ньютона), — отдельная эвереттическая проблема. С содержательной точки зрения простая линейная упорядоченность может трактоваться как нумерологически, так и абстрактно.

Геометрически совокупность малых эвереттианов можно представить в виде множества точек первого квадранта евклидовой плоскости.

Отдельным вопросом является математическая сущность точек Ψ_{q_i} и Ω_{p_j} на

координатных осях этой плоскости. В общем случае всякая «точка» на координатных осях — это индивидуальный квантовый или психоидный мир, а «точка» на плоскости первого квадранта — индивидуальная КРФМ, «кадр пинакотеки Барбура».

Поскольку конкретная структура каждого построенного эвереттического пространства зависит от способа упорядочения элементов Ψ_{q_i} и Ω_{p_j} , который, в свою очередь, определяется выбором стрелы времени для каждого индекса q_i и p_j , построенное пространство является квазиатемпоральным не только в смысле природы *волновых* функций, но и в смысле упорядоченности состояний параметрами памяти входящих в них волновых функций. Такое время определяется объёмом эвереттовской памяти волновых функций Ψ_{q_i} и Ω_{p_j} и какими-то дополнительными правилами построения из них вектора стрелы времени, которые ещё должны быть определены.

Поскольку на такой евклидовой плоскости нет отрицательных значений, то и деления здесь нет, поскольку нет «обратных величин» (которым соответствует степень минус единица).

Нет здесь и нуля. «Природа не терпит пустоты» — нулевой малый эвереттиан соответствует «абсолютному небытию» КРФМ. Геометрически оси Ψ_{q_i} и Ω_{p_j} пересекаются в точке $\{1;1\}$.

В построенной математической модели из четырёх структурных элементов декартового рассечения евклидовой плоскости только один идентифицирован как инструмент описания эвереттического многомирия. Плоскость в целом соответствует модели «внутреннего» и «внешнего» наблюдателей (наблюдателей «от первого лица» и «от третьего лица» в терминологии Дж. Хартля и Т. Хертога [8]).

Данная конструкция конкретизирует условия существования и наблюдения нашего универса «третьим лицом» — внешним наблюдателем, существенная роль которого в квантовой механике обсуждалась ещё Эвереттом [6], но введённым в данном контексте в [9] и, независимо, в [10].

Однако система введённых постулатов выводит внешнего наблюдателя из бытийствования и квантовой, и психоидной, и классической реальностей. Он занимает три четверти евклидовой плоскости эвереттического пространства, но не зафиксирован ни в одной его точке и не взаимодействует с эвереттианами. Возникающая физикалистская аналогия с тёмной энергией может оказаться содержательной при дальнейшем развитии предложенной геометризованной модели эвереттического пространства.

Построенное на таких аксиомах пространство является простейшим квазиатемпоральным эвереттическим пространством.

1.1.2. Эвереттическое умножение. Полные эвереттианы

Введение операции умножения эвереттианов вводит в описание нелинейность, запутывание и порождает новые математические, физические и психоидные сущности.

Принимаем, что в операции «эвереттического умножения» первый сомножи-

тель является активным началом, «источником действия» на второй сомножитель. В связи с этим эвереттическое умножение является действием некоммуникативным и неассоциативным. Дистрибутивность сохраняется и обеспечивает возникновение новых сущностей.

Действительно, в простейшем случае перемножения двух малых эвереттианов $e_{v_{1+1}}$ и $e_{v_{2+2}}$ имеем:

$$(\Psi_{q_1} + \Omega_{p_1}) \cdot (\Psi_{q_2} + \Omega_{p_2}) = \Psi_{q_1} \cdot \Psi_{q_2} + \Psi_{q_1} \cdot \Omega_{p_2} + \Omega_{p_1} \cdot \Psi_{q_2} + \Omega_{p_1} \cdot \Omega_{p_2}. \quad (3)$$

Все слагаемые в правой части уравнения (3) – перепутанные состояния действительностей 1 ($e_{v_{1+1}}$) и 2 ($e_{v_{2+2}}$). В эвереттике такие состояния называются склейками (фузиями) [11, 12]. Тогда:

а) $\Psi_{q_1} \cdot \Psi_{q_2}$ — материальная склейка (проникновение, фузия): воздействие физических сущностей действительности 1 на физические сущности действительности 2.

б) $\Psi_{q_1} \cdot \Omega_{p_2}$ — физико-ментальная склейка (проникновение, фузия): воздействие физических сущностей действительности 1 на психоидные сущности действительности 2.

в) $\Omega_{p_1} \cdot \Psi_{q_2}$ — ментально-физическая склейка (проникновение, фузия): воздействие психоидных сущностей действительности 1 на физические сущности действительности 2.

г) $\Omega_{p_1} \cdot \Omega_{p_2}$ — ментальная склейка (проникновение, фузия): воздействие психоидных сущностей действительности 1 на психоидные сущности действительности 2.

Интересен частный случай уравнения (3), в котором перемножаются *одинаковые* эвереттианы, т.е. рассматривается квадрат эвереттиана $(e_{v_{1+1}})^2$.

В этом случае материальная склейка а) отражает процессы самодействия физических объектов (например, воздействие электрического заряда электрона на самого себя), физико-ментальная склейка б) соответствует представлению воздействия физического объекта на наблюдателя (процесс измерения), ментально-физическая склейка в) представляет «мысленное воздействие» наблюдателя на физический объект (например, телекинез), ментальная склейка г) отражает процесс рефлексии в психологии.

Эвереттическое умножение в общем случае порождает произведения произвольного количества сомножителей вида Ψ_{q_i} и Ω_{p_j} , расположенных в произвольном порядке. Суперпозиции полученных дистрибутивных произведений также будут бытийственными. Трактовка смыслового содержания таких произведений и их суперпозиционных сумм является задачей конкретных эвереттических исследований. В философском плане изучению таких эвереттических объектов посвящены работы А. Костерина [13].

Для описания и идентификации этих объектов должно быть введено специальное конфигурационное гильбертово пространство H_{e_v} , каждая ось которого представляет соответствующий член суперпозиции, но без учёта индексов i и j , а началом координат является точка $\{1;1;\dots;1\}$.

Упорядочивание членов с индексами должно соответствовать правилам, вводимым для упорядочивания малых эвереттианов.

Каждая точка, не лежащая на осях Ψ_{q_i} и Ω_{p_j} в пространстве H_{e_v} , будет являться *полным эвереттианом* $E_{v_{i+j}}$ некоего физико-психоидного объекта — конкретного «эвереттовского мира» с разным «содержанием» физичности и психоидности.

Фантастическое разнообразие полных эвереттианов позволяет надеяться, что после селекции с помощью антропного принципа среди них могут быть выделены миры, физически весьма отличные от нашего универса, но пригодные для существования разума «нашего типа». Такую возможность предвидел и выразил в художественной форме П. Амнуэль в романе «Тривселенная» [14].

1.2. Проблемы введения нуля и дополнительных размерностей в эвереттианы

Введение в рассмотрение нуля (а, следовательно, деления и отрицательных чисел, поскольку $0=k^{-1}$, где $k = 1/0$ по [15]), фактически разрушает предложенную конструкцию. Появляются сущности вида Ψ_{q_i}/Ω_{p_j} и подобные им структуры, а также «реально отрицательные» КвР, ПсР и КРФМ, физический смысл которых совершенно неясен. И с формальной точки зрения разрушается структура эвереттического пространства H_{e_v} с началом координат в точке $\{1;1;\dots 1\}$.

Но вместе с этим появляется и инструмент анализа структуры физических сингулярностей [11, с. 145] с помощью методов инфинитной математики, неизбежность оформления которой в качестве инструмента анализа эвереттических феноменов предсказал П. Амнуэль [16].

Какая геометрическая конструкция может содержать компоненты эвереттиана и ноль, или, что эквивалентно по смыслу, какой может быть структура точки $\{1;1;\dots 1\}$ при введении понятия нуля — это отдельная топологическая задача.

Исходя из целей представляемой конструкции эвереттического пространства, после введения нуля и, как следствие этого, отрицательных значений Ψ_{q_i} , Ω_{p_j} можно рассмотреть и комплексные малые эвереттианы:

$$\Psi_{q_i} + i\Omega_{p_j} = e_{v_{i+ij}}, \quad (4)$$

$$\Omega_{p_j} + i\Psi_{q_i} = e_{v_{j+ii}}, \quad (5)$$

где $i = \sqrt{-1}$.

Это вводит сущностную мнимость, используемую для описания фаз волновых функций в квантовом пространстве, и позволяет надеяться, что в рамках физико-психоидного рассмотрения преимущества манипулирования с комплексными числами облегчат осознание физического смысла понятия фазы, как ключевой характеристики и шрёдингеровской, и психоидной волновых функций.

Очевидно, что Ψ_{q_i} и Ω_{p_j} не исчерпывают всех смысловых характеристик бытия. Наверняка найдутся характеристики («измерения»), для номинации которых может не хватить букв не только греческого алфавита, но и десятков тысяч китайских иероглифов. И сцена бытия, на которой развёртывается его действие, обязательно потребует расширения рассмотренного гильбертова пространства H_{ev} . Придётся вводить, например, новые оси физикалистского класса Ψ_{q_i} для миров «суперсимметричной» и «зеркальной» материй, если их откроют на Большом адронном коллайдере. Но ещё больше сюрпризов ожидает нас на психоидных осях Ω_{p_j} . На какой, например, оси может оказаться категорический императив Канта для космического «морального субъекта» SS 433 Лефевра [17]?

1.3. Алгебраические эвереттические пространства

Все рассмотренные варианты «сцены мирового действия» являются геометризованными конструкциями. Но ведь есть ещё и принципиально другие, более глубокие по отношению к геометрии, алгебраические варианты! И в рамках их рассмотрения, например, группой В.В. Кассандрова уже построена алгебраическая модель миров Уилера-Фейнмана, в которых *все* электроны и позитроны универса являются проявлениями *одной* частицы [18], и модель взаимодействия этой частицы с наблюдателем [19].

Алгебраическим является и подход В.Л. Янчилина к описанию движения [20]. Он учитывает нелокальность квантовых состояний, которые описываются с помощью функции Дирихле:

$$x(t) = x_1, \text{ если } t - \text{рациональное число,}$$

$$x(t) = x_0, \text{ если } t - \text{иррациональное число.}$$

Такой подход к взаимосвязи пространства (x) и ньютоновского времени (t) фактически является атемпоральным представлением, поскольку параметр t входит в него не динамически (у времени в этом рассмотрении нет длительности), а термодинамически. Здесь время — маркер состояния, а не процесса. При этом функцию Дирихле легко можно обобщить так, чтобы сделать не двузначной, а многозначной, даже бесконечнозначной. Например, функция становится десятизначной, если потребовать, чтобы она равнялась какому-то M -ному числу после запятой в десятичном представлении точки t на действительной оси, если эта точка рациональна, и N -ному числу, если она иррациональна [21].

Проблема использования ньютоновского (континуального) времени при описании эвереттических процессов ветвления (дискретные процессы) состоит в том, чтобы установить механизм упорядочивания счётного множества событий на континуальной временной шкале.

2. Обсуждение

Разумеется, приведённые примеры не исчерпывают геометрических и алгебраических возможностей построения «шекспировских сцен» для описания

театра бытия. Философским же является вопрос — все ли они пригодны для постановки каких бы то ни было «действительностей»? Иными словами, может ли Разум «придумать» нечто, лишённое возможности физического бытийствования? Или всё это — и «внешневнутренние состояния», и действительности с числовыми осями масштабов «меньше нуля» и «больше бесконечности», и математика психоидности, в которой «сумма страданий даёт абсурд» [22, с. 113], являются только «игрой ума»? Но говорят, что Н. Бурбаки приписали Эрмиту такую фразу: «Числа, функции и прочие математические понятия подобны зверушкам в зоопарке: ими можно любоваться, но изменить нельзя: они ЕСТЬ!» Сегодня М. Тегмарк заподозрил эрмитовских зверушек не только в бытийственности, но и в физичности, и прозорливо отнёс их к многомириям четвёртого, математического уровня эвереттических многомирий [23].

3. Заключение

Предложенные конструкции эвереттического пространства не являются расчётными математическими моделями, поскольку включают в себя содержательные параметры (прежде всего, связанные с психоидностью бытийствующих структур), которые в настоящее время не имеют количественного описания. Может быть, психоидность и не поддаётся количественным оценкам в том смысле, в котором мы сегодня понимаем категорию количественного. Как бы то ни было, для получения содержательно предсказательных результатов моделирования событий в эвереттических пространствах необходимо «проявить» фрактальную структуру психоидного, опираясь на математическое «проявление» фрактальной структуры физического.

ЛИТЕРАТУРА

1. Менский М.Б. Сознание и квантовая механика. Фрязино : «Век 2», 2011, 320 с.
2. Lebedev Yu.A., Amnuel P.R., Dulphan A.Ya. Infinite-Dimensional Multievents Space-Time of Minkowski and Everett's Axiom of Parallelism // American Journal of Modern Physics. Special Issue: Physics of Time: Theory and Experiment. 2015. V. 4, No. 2-1. P. 1–8.
3. Лебедев Ю.А. Модель бесконечномерного мультисобытийного пространства-времени Минковского и физический смысл эвереттических ветвлений и склеек // Математические структуры и моделирование. 2014. № 4. С. 13–22.
4. Гуц А.К. Элементы теории времени. Омск : Издательство Наследие. Диалог-Сибирь, 2004. 364 с.
5. Shakespeare W. As You Like It, spoken by Jaques in Act II Scene VII.
6. Everett H. "Relative State" Formulation of Quantum Mechanics // Reviews of Modern Physics. 1957. V. 29. P. 454–462.
7. Barbour J. The end of time: the next revolution in our understanding of the universe. London : Weidenfeld & Nicolson, 1999. P. 371.
8. Hartle J., Hertog T. The Observer Strikes Back. arXiv:1503.07205v1 [gr-qc] 24 Mar 2015.

9. Hawking S.W., Hertog T. Populating the Landscape: A Top Down Approach // Phys. Rev. D73. 2006. P. 123527. arXiv:hep-th/0602091.
10. Лебедев Ю.А. Нелинейные семантические аспекты квантовомеханической концепции соотнесённых состояний Х. Эверетта и перспективы развития эвереттики // Математические структуры и моделирование. 2007. Вып. 17. С. 53–71.
11. Лебедев Ю.А. Неоднозначное мироздание. Кострома, 2000. 320 с.
12. Lebedev Y.A., Amnuel P.R., Dulfan A.Y. The Everett axiom of parallelism. arXiv:1304.0310v1 [physics.gen-ph].
13. Костерин А. О возможности взаимовлияния альтерверсов (2015). URL: <http://milkywaycenter.com/everettica/AK120315D.pdf>. И др. статьи на сайте Международного центра эвереттических исследований. URL: <http://www.everettica.org/member.php3?mode=1&m=kost>.
14. Амнуэль П. Тривселенная. Новомосковск, 2004. 366 с.
15. Лебедев Ю. Аш-функция Хевисайда // Alma mater. 1991. № 5. С. 91–97.
16. Амнуэль П. Вселенные: ступени бесконечностей. Иерусалим : изд-во «Млечный путь», 2057. 458 с.
17. Лефевр В.А. Космический субъект. М. : Ин-кварти, 1996. 184 с.
18. Kassandrov V.V., Khasanov I.Sh., Markova N.V. Algebraic dynamics on a single worldline: Vieta formulas and conservation laws. arXiv:1402.6158v2 [math-ph].
19. Kassandrov V.V., Khasanov I.Sh., Markova N.V. Collective Lorentz invariant dynamics on a single "polynomial" worldline. arXiv:1501.01606v1 [physics.gen-ph].
20. Янчилин В.Л. Квантовая нелокальность. М. : Изд-во Красанд, 146 с.
21. Лебедев Ю.А. «Левенгуки» и «Ньютоны» квантового мира // Млечный путь (Иерусалим). 2015. № 1. С. 218–240.
22. Бродский И.А. Письмо генералу Z / Избранные стихотворения. М. : Изд-во «Панорама», 1994. 496 с.
23. Tegmark M. The Mathematical Universe // Foundations of Physics. 2008. V. 38. P. 101–150. arXiv:0704.0646 (2007).

SPECIFICITY OF MODELING OF EVERETTIAN SPACES

Ju.A. Lebedev

Ph.D. (Phys.-Math.), Associate Professor, e-mail: lebedev@bmstu.ru

Bauman Moscow State Technical University

Abstract. Everettian spaces as intuitively shaped models of reality in terms of the manyworlds interpretation of quantum mechanics have not adequate mathematical description. This is largely due to the fact that mathematical apparatus is absent to describe the specifics of psychoid aspect of reality. We propose the fractal superposition approach to the formulation of basic concepts needed to build the mathematical model suitable for particular calculations. For obvious reasons, the fundamental concept of quantum mechanics — the wave function is selected as analogue of fractal kernel describing psychoid.

Keywords: everettics, quantum mechanics, the correlation of states, physicality, psychoid, everettian spaces, multiverse, altervers, modeling reality, everettian.

HOW TO COMPUTE VON NEUMANN-MORGENSTERN SOLUTIONS

Martha Osegueda Escobar

Student, e-mail: mcoseguedaescobar@miners.utep.edu

Vladik Kreinovich

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

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

Abstract. One of the main solutions to a conflict situation is the von Neumann-Morgenstern (vN-M) solution. Intuitively, it is a set of outcomes describing a "social norm", so that (1) if someone tries to propose an outcome outside this set, then we can force this outcome back into this set, and (2) once an outcome within this set is selected, no coalition is interested in switching to a different socially acceptable outcome. There are two main problems with this approach: some situations lack such a solution, and no general algorithm is known for producing this solutions when it exists. In this paper, we show that, if we take into account that in real-life situations, the outcomes are known only with some accuracy, then it becomes possible to algorithmically find the corresponding "approximate" vN-M solutions.

Keywords: cooperative game, von Neumann-Morgenstern solution, algorithmic computability, approximate von Neumann-Morgenstern solution.

Cooperative games and von Neumann-Morgenstern solution. Situations when all participants collaborate with each other is known as a *cooperative game*. One way to describe a cooperative game is to assign, to every subset $S \subseteq N \stackrel{\text{def}}{=} \{1, \dots, n\}$ of the set of all the participants, the value $v(S)$ that describes what players from S can gain if they collaborate between themselves only. Such subsets S are called *coalitions*.

We consider cooperative situations, so if two disjoint coalitions S and S' collaborate, they should be able to gain not less than they would get on their own, i.e., we should have $v(S \cup S') \geq v(S) + v(S')$.

It always makes sense to consider only gains due to collaboration, so if $v(\{i\}) \neq 0$, then we can take $v'(S) = v(S) - \sum_{i \in S} v(\{i\})$ for which $v'(\{i\}) = 0$.

Definition 1. Let n be a positive integer. By a cooperative game, we mean a function that assigns, to each subset $S \subseteq N \stackrel{\text{def}}{=} \{1, \dots, n\}$, a non-negative number $v(S)$ so that

- $v(\{i\}) = 0$ for all i , and
- when $S \cap S' = \emptyset$, then $v(S \cup S') \geq v(S) + v(S')$.

Since everyone is collaborating, the participants together get the value $v(N)$. The question is: what is a fair way to divide this total amount $v(N)$ between n participants, i.e., how to allocate non-negative amounts $x_1, \dots, x_n$ for which $\sum_{i=1}^n x_i = v(N)$. Such allocations are known as *imputations*.

Definition 2. By an imputation, we mean a tuple $(x_1, \dots, x_n)$ of non-negative numbers for which $\sum_{i=1}^n x_i = v(N)$.

In their original book [6] that started game theory, John von Neumann and Oscar Morgenstern considered the following notion of *dominance* between imputations $x \succ y$. We say that an imputation x *dominates* an imputation y if there exists a coalition S for which each player from S gets more money in x than in y ($x_i > y_i$) and which is “reachable” for S – i.e., for which $\sum_{i \in S} x_i \leq v(S)$.

The condition $\sum_{i \in S} x_i \leq v(S)$ means that the coalition S can force a switch from y to x , and the condition $x_i > y_i$ means that this switch is beneficial for all members of the coalition S . So, if both imputations x and y are possible, S will force a switch from x to y .

Definition 3. We say that an imputation x dominates an imputation y (and denote it by $x \succ y$) if $x_i > y_i$ for all $i \in S$ and $\sum_{i \in S} x_i \leq v(S)$.

At first glance, it may seem reasonable to consider the set of all non-dominated imputations as a solution; this set is known as a *core* [3, 6]. Alas, often, this set is empty: we may have both $x \prec y$ (because of one coalition S) and $y \prec x$ (because of another coalition S'). In such games, if we allow all possible imputations, we can potentially switch infinitely many times, never reaching an equilibrium.

To avoid such situations, von Neumann and Morgenstern suggested that we adopt some *social norms* that would limit the set of possible imputations in such a way that no two imputations within this norm dominate each other. The social norm has to be *enforceable* meaning that if someone proposes an imputation which is outside this norm, there should be a coalition that forces a switch to a solution within the norm. The resulting definition is known as a *von Neumann-Morgenstern solution* (or *NM-solution*, for short).

Definition 4. A set C of imputations is called a von Neumann-Morgenstern solution if it satisfies the following two properties:

- if $x, y \in C$, then $x \not\succ y$;
- if $y \notin C$, then there exists an $x \in C$ for which $x \succ y$.

In this case, our decision making consists of two stages:

- first, all participants agree on an appropriate “social norm”, i.e., on an appropriate set of imputations C within which they search for an imputation;

- then, once the set C is selected, the participants select an imputation x from this set.

The two conditions from Definition 4 guarantee that:

- if someone tries to violate an agreement and propose an imputation y outside the set C corresponding to the social norm, then we can force it back into C ;
- second, that once an imputation x is selected, no coalition is interested in switching to a different socially acceptable imputation y .

Our goal is to compute the list of all possible social norms C (or at least compute one social norm C).

Computing von Neumann-Morgenstern solution is a challenge. Originally, von Neumann and Morgenstern proposed their solution as the main solution concept for cooperative games. Unfortunately, the two major challenges emerged. A minor challenge is that there are games which do not have this solution at all. A major challenge is that computing this solution is not easy. As of now, it is not even clear whether there exists an algorithm that can compute such a solution [2, 5].

In the discrete case, it is known that the problem of checking the existence of an NM-solution is NP-hard. Indeed, we can represent the conflict situation as a graph, with outcomes as vertices and $x \succ y$ if and only if there is an edge from x to y . In graph terms, an NM-solution is a *minimum independent dominating set*, or a *kernel*. The problem of checking the existence of such a kernel is NP-complete. This result was first proven in [1]; see also [4] (Problem 9.5.10).

Comment. For a general overview of complexity of different conflict resolution notions, see, e.g., [5] and references therein.

Our idea. A real-life division of the overall sum may involve not just division of money, but rather a division of objects whose price is also only approximately known. Thus, the actual value x_i allocated to each person is also only approximately known. Let us introduce the following definition.

Definition 5. Let $\varepsilon > 0$.

- We say that tuples x and x' are ε -close if $|x_i - x'_i| \leq \varepsilon$ for all i and $\left| \sum_{i \in S} x_i - \sum_{i \in S} x'_i \right| \leq \varepsilon$ for all coalitions S .
- We say that two sets of tuples X and X' are ε -close if every $x \in X$ is ε -close to some tuple from X' , and every $x' \in X'$ is ε -close to some tuple from X .

When we only know ε -approximations $\tilde{x}$ and $\tilde{y}$ to numbers x and y , then we cannot check whether $x > y$. To be more precise, if $\tilde{x} > \tilde{y} + 2\varepsilon$, then $x \geq \tilde{x} - \varepsilon > \tilde{y} + \varepsilon \geq y$ hence $x > y$. Similarly, if $\tilde{x} \leq \tilde{y} - 2\varepsilon$, then we are sure that $x \leq \tilde{x} + \varepsilon < \tilde{y} - \varepsilon \leq y$ and

thus, that $x \not\succ y$. However, if $\tilde{y} - 2\varepsilon < \tilde{x} \leq \tilde{y} + 2\varepsilon$, then we can have both $x > y$ and $x \not\succ y$. So, we can distinguish between “necessarily larger” relation $\tilde{x} > \tilde{y} + 2\varepsilon$ and “possibly larger” relation $\tilde{x} > \tilde{y} - 2\varepsilon$.

When we request that no two imputations from the set C dominate each other, this cannot mean “possibly larger”, since even for $x = y$, the value y_i is possibly larger than x_i . Thus, we need to require “necessarily larger” condition. On the other hand, when we require that every imputation not from C can be forced into C , we cannot use the “necessarily larger” condition, since this way we may eliminate some “forcings”. Thus, we arrive at the following definitions.

Definition 8.

- We say that a tuple x necessarily ε -dominates y , and denote it by $x \succ^\varepsilon y$ if $x_i > y_i + 2\varepsilon$ for all $i \in S$ and $\sum_{i \in S} x_i \leq v(S) - \varepsilon$.
- We say that a tuple x possibly ε -dominates y , and denote it by $x \succ_\varepsilon y$ if $x_i > y_i - 2\varepsilon$ for all $i \in S$ and $\sum_{i \in S} x_i \leq v(S) + \varepsilon$.
- Let X be a set of tuples. We say that a set C is an ε -NM solution for the set X if the following two conditions are satisfied:
 - if $x, y \in C$, then $x \not\succ^\varepsilon y$;
 - if $y \in X - C$, then there exists an $x \in C$ for which $x \succ_\varepsilon y$.

This modification of the original NM-definition takes into account that we only know the values $v(S)$ and x_i only approximately. Such “approximate” NM-solutions can be algorithmically computed in the following sense:

Theorem. *There exists an algorithm that, given a rational-valued game v and positive rational numbers $\varepsilon > \delta > 0$, returns a finite list of finite sets $C_1, \dots, C_M$ such that:*

- each of the sets C_i is an $(\varepsilon + \delta)$ -NM solution, and
- each $(\varepsilon - \delta)$ -NM solution C is δ -close to one of the sets C_i .

Proof. Let us use a grid with step δ to form a finite δ -approximation X' to the set X of all imputations. For any subset C of this finite set X' , we can algorithmically check whether this set is an ε -NM solution for the set X' . Let us show that the list of all such solutions is the desired list $C_1, \dots$

Indeed, one can easily check that if $x \succ^\alpha y$, x' is δ -close to x , and y' is δ -close to y , then $x' \succ^{\alpha-\delta} y'$. Thus, if $x' \not\succ^{\alpha-\delta} y'$, then $x \not\succ^\alpha y$. In particular, if $x' \not\succ^\alpha y'$, then $x \not\succ^{\alpha+\delta} y$.

Similarly, if $x \succ_\alpha y$, x' is δ -close to x , and y' is δ -close to y , then $y' \succ_{\alpha+\delta} x'$.

Thus, if C is an α -NM solution for X , C' is δ -close to C , and X' is δ -close to X , then C' is $(\alpha + \delta)$ -NM solution for X' .

Since each set C_i is an ε -NM solution for X' , and X' is δ -close to X , we conclude that C_i is an $(\varepsilon + \delta)$ -NM solution for the set X .

Vice versa, let C be an $(\varepsilon - \delta)$ -NM solution for the set X . By construction of the set X' , each point from the set X , in particular, each point from C , is δ -close to some point from X' . Let C' denote the set of all the points from X' which are δ -close to some point from C . By definition, this set C' is δ -close to C . Since X' is δ -close to X , we thus conclude that $C' \subseteq X'$ is an ε -NM solution for X' and is, thus, one of the sets C_i . The theorem is proven.

Acknowledgments. This work was supported in part by the National Science Foundation grants HRD-0734825, HRD-1242122, and DUE-0926721.

REFERENCES

1. Chvatal V. On the computational complexity of finding a kernel // Center de Recherches Mathématiques, Université de Montréal, Technical Report CRM-300, 1973.
2. Deng X., Papadimitriou C.H. On the complexity of cooperative solution concepts // Mathematics of Operations Research. 1994. V. 19, No. 2. P. 257–266.
3. Luce R.D., Raiffa H. Games and Decisions: Introduction and Critical Survey. New York : Dover, 1989.
4. Papadimitriou C.H. Computational Complexity. San Diego : Addison Wesley, 1994.
5. Papadimitriou C.H. Algorithms, games, and the Internet // Proceedings of ACM Symposium on Theory of Computing STOC'01, Hersonissos, Crete, Greece, July 6–8, 2001.
6. von Neumann J., Morgenstern O. Theory of Games and Economic Behavior. Princeton, New Jersey : Princeton University Press, 1944.

ВЫЧИСЛЕНИЕ РЕШЕНИЯ ПО НЕЙМАНУ-МОРГЕНШТЕРНУ

Марта Осегеда Эскобар

студент, e-mail: mcoseguedaescobar@miners.utep.edu

Владик Крейнович

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

Техасский университет в Эль Пасо, США

Аннотация. Одним из основных решений конфликтной ситуации является решение по Нейману-Моргенштерну (НМ-решение). Интуитивно понятно, что это набор исходов, описывающих «социальную норму», так что: (1) если кто-то пытается предложить исход не из этого набора, мы можем принудительно включить этот исход в этот набор, и (2) если выбран исход из этого набора, ни одна коалиция не заинтересована в переходе к другому социально приемлемому исходу. Есть две основные проблемы, связанные с этим подходом: в некоторых ситуациях нет такого решения, и не известен ни один общий алгоритм для создания такого решения, когда оно существует. В этой статье, приняв во внимание, что в реальных жизненных ситуациях исходы известны лишь с некоторой точностью,

мы покажем, что становится возможным алгоритмически найти соответствующие «приближенные» НМ-решения.

Ключевые слова: кооперативная игра, решение по Нейману-Моргенштерну, алгоритмическая вычислимость, приближение решения по Нейману-Моргенштерну.

СЕГМЕНТАЦИЯ ИЗОБРАЖЕНИЙ НА ОСНОВЕ АЛГОРИТМА ВЫДЕЛЕНИЯ СООБЩЕСТВ НА ГРАФЕ

С.В. Белим

д.ф.-м.н., профессор кафедры инф. безопасности, e-mail: belimsv@omsu.ru

С.Б. Ларионов

аспирант, e-mail: me@stas-larionov.ru

Омский государственный университет им. Ф.М. Достоевского

Аннотация. В статье предложен алгоритм сегментации изображений, использующий его представление в виде неориентированного взвешенного графа, с последующим выделением сообществ на графе. Каждому сообществу соответствует сегмент изображения. В качестве меры веса дуг графа используется функция от разности интенсивности цветов в пространстве RGB, возрастающая по экспоненциальному закону. Для выделения сообществ построен жадный алгоритм. Для оценки качества разбиения графа на сообщества используется функция модульности, введенная Ньюманом. Проведён компьютерный эксперимент. Показано, что размеры сегментов зависят от значения параметра в функции веса рёбер. Предложенный алгоритм устойчив к импульсным шумам.

Ключевые слова: сегментация изображений, сообщества, графы.

Введение

Под сегментацией принято понимать разбиение изображения на группы пикселей по близким значениям некоторого показателя [1]. Необходимость в сегментации изображения возникает в ряде прикладных задач, таких как обработка медицинских и биологических снимков [2–4], обработка аэрокосмических снимков [5–7], контроль качества поверхности материалов [8], системах технического зрения [9, 10] и др.

На сегодняшний день разработано достаточно много методов сегментации, среди которых можно выделить несколько базовых, на которых основываются большое количество прикладных алгоритмов: метод пороговой сегментации [11, 12], метод центроидного связывания [13], метод водораздела [14] и др.

Основная идея всех алгоритмов сегментации состоит в выделении связной области изображения, яркость пикселей которой различается не более чем на некоторую величину. Данное число принято называть порогом сегментации. Ряд работ направлен на выработку алгоритмов автоматического выбора порога сегментации [15]. Кроме этого, большое распространение получили алгоритмы, основанные на выделении контурных границ с помощью вычисления первой

производной (операторы Собеля и др. [16–19]) или второй производной (оператор Лапласа [16, 17, 20]). В этом случае осуществляется поиск границ, которые в дальнейшем объединяются в контуры. В качестве сегментов выбираются области внутри замкнутых контуров. Также существуют методы, основанные на выращивании областей [21, 22], в которых происходит группировка пикселей в области вокруг «центров кристаллизации». Достаточно много методов основываются на представлении пикселей изображения в виде точек многомерного пространства признаков с последующей кластеризацией [23–30].

Существует ряд методов сегментации, основанных на разрезании графа. Основная идея состоит в сопоставлении пикселям изображения вершин графа. Между всеми вершинами строятся ребра, веса которых показывают меру сходства между вершинами по какой-либо характеристике. Далее граф разрезается на подграфы таким образом, чтобы внутри подграфа веса рёбер были значительно больше весов рёбер, связывающих подграфы. В работе [31, 32] предложен метод нормализованных разрезов. Применение данного метода к задаче выявления дефектов поверхности исследовано в работе [8]. Одним из лучших графовых алгоритмов на сегодняшний день считается GraphCut [33–35]. В данном алгоритме множество вершин состоит из пикселей изображения и двух дополнительных искусственно добавленных вершин — исток и сток. Пользователь в интерактивном режиме задаёт «семена» объекта — несколько пикселей, принадлежащих объекту, и «семена» фона — несколько пикселей, принадлежащих фону. Семена объекта связываются с истоком рёбрами, имеющими бесконечный вес. Семена фона связываются со стоком рёбрами, имеющими бесконечный вес. Для полученного графа, дополненного вершинами истока и стока, ищется минимальный разрез [36], делящий граф на две части. Подграф, содержащий исток, соответствует объекту. Подграф, содержащий сток, соответствует фону. Данный алгоритм работает в интерактивном режиме и позволяет выделять отдельный элемент изображения. Другой подход на основе теории графов, позволяющий проводить полную сегментацию изображения, предложен в работах [37, 38]. Основная идея сегментации состоит в сопоставлении изображению графа с последующей его кластеризацией на основе построения минимального остовного дерева.

По своей постановке задача сегментации изображений близка к проблеме выделения сообществ (community) на графах. В обоих случаях необходимо выделить и сгруппировать элементы, близкие друг к другу по некоторому показателю. На сегодняшний день наблюдается всплеск интереса к проблеме поиска сообществ на графах [39], связанный, прежде всего, с развитием социальных сетей и других сообществ в глобальной сети Интернет. Однако предложенные алгоритмы могут найти применение и в других задачах близких по постановке, при соответствующем представлении данных.

Целью данной статьи является разработка и тестирование алгоритма сегментации, основанного на наиболее распространённом алгоритме поиска сообществ на графах, предложенного в работе [40].

1. Постановка задачи и представление данных

Задача кластеризации изображений состоит в разбиении множества точек изображения на подмножества, «близкие» по цвету. Основная проблема состоит в строгом определении степени близости точек. В большинстве прикладных задач необходимо выделить некоторые предметы на изображении, причём эти предметы человеческий глаз выделяет на общем фоне именно по цветовому признаку. То есть предмет идентифицируется как множество точек близких по цвету друг к другу, но отличающихся от окружающего фона. При этом цвет самого выделяемого предмета может плавно изменяться в широком диапазоне. В дальнейшем, при решении задачи кластеризации, ограничимся поиском только односвязных областей.

Пусть задано изображение, размеры которого составляют N пикселей по горизонтали и M пикселей по вертикали. Если цвет задан в градациях серого, то будем считать, что палитра содержит m цветов, то есть каждый пиксель характеризуется целым числом в интервале $[0, m - 1]$. Если изображение цветное, то будем использовать модель RGB. В этом случае каждому пикселю сопоставляется три целых числа в интервале $[0, m - 1]$. В дальнейшем будем рассматривать общий случай цветного изображения. Более простой случай монохромного изображения может быть получен из него простым обнулением двух из трёх цветовых координат.

Введём пятимерное пространство XYRGB. Построим в этом пространстве множество точек, соответствующих пикселям изображения. Первые две координаты будут определяться геометрическим положением пикселя на исходном изображении, оставшиеся три — его цветовыми характеристиками. В некоторых случаях имеет смысл провести нормировку координат так, чтобы по всем направлениям множество точек было ограничено величинами, сопоставимыми по порядку. Однако в рамках подхода, предлагаемого в данной статье, такое преобразование не является обязательным, так как в формуле для расстояния между точками геометрические и цветовые координаты входят не равноправно. Как будет видно далее, нормировка цветовых координат приводит лишь к переопределению значения параметра, определяемого пользователем. Будем считать, что пикселю, расположенному на изображении в точке с координатами (x, y) , в пространстве XYRGB соответствует точка с координатами $(x, y, r(x), g(x), b(x))$, где $r(x)$ — интенсивность красной составляющей цвета пикселя, $g(x)$ — интенсивность зелёной составляющей цвета пикселя, $b(x)$ — интенсивность синей составляющей цвета пикселя.

Значительное влияние на результат работы алгоритма кластеризации оказывает способ вычисления расстояния между точками в пространстве XYRGB. Величина расстояния между двумя точками должна быть чувствительна к изменению цвета, поэтому необходимо выбирать функцию, быстро растущую при переходе к соседнему пикселю с резко отличающимся цветом. Для координат, определяемых геометрическим положением пикселя на изображении, наоборот выберем зависимость, убывающую с расстоянием, для того, чтобы сгладить накапливание значения расстояния в рамках одного кластера при градиент-

ной заливке. По аналогии с работой [40] будем вычислять расстояние между точками $v_1=(x_1, y_1, r_1, g_1, b_1)$ и $v_2=(x_2, y_2, r_2, g_2, b_2)$ по формуле:

$$d(v_1, v_2) = f(x_1 - x_2, y_1 - y_2) \cdot \exp\left(\frac{1}{\delta} \sqrt{(r_1 - r_2)^2 + (g_1 - g_2)^2 + (b_1 - b_2)^2}\right).$$

Здесь функция $f(x, y)$ — зависит от геометрической удалённости двух точек друг от друга. Её выбор зависит от графа, сопоставляемого изображению на следующем шаге алгоритма. Параметр δ является выбираемым и влияет на скорость роста функции расстояния между двумя точками. Данный параметр является настраиваемым и, выбирая его, можно влиять на то, какие области признаются равномерно закрашенными, а какие нет.

Построим неориентированный взвешенный граф, множество вершин которого совпадает с построенным множеством точек. Соединим рёбрами только вершины, соответствующие соседним пикселям на исходном изображении. Причём соседями будем считать не только пиксели, граничащие с данным по горизонтали и вертикали, но и по диагонали. Таким образом, каждый пиксель, расположенный не на границе изображения, имеет восемь соседей, с которыми он соединён рёбрами. Четыре точки по углам изображения имеют всего по три соседа. Точки, расположенные на границе изображения, имеют по пять соседей. Из построения графа видно, что выбор функции будет слабо влиять на результат, так как мы рассматриваем только соседние пиксели, расстояние между которыми фиксировано. В дальнейшем будем считать $f(x, y) = 1$, что равносильно работе в трёхмерном пространстве RGB.

Задача кластеризации изображения сводится к разбиению построенного графа на подграфы, обладающие определёнными свойствами. Неформально данное свойство подграфа определяется как отношение к одному кластеру на изображении. Одна из формализаций искомых свойств может быть построена на основе поиска сообществ в графе.

Таким образом, кластеризация изображения может быть осуществлена в три этапа. На первом этапе производится представление множества пикселей изображения в виде сначала множества точек пятимерного пространства, а потом взвешенного графа. На втором этапе выполняется разбиение графа на подграфы. На третьем этапе происходит восстановление кластеров изображения, соответствующих подграфам.

2. Алгоритм кластеризации

Как было сказано ранее, кластеризация изображения в нашем представлении свелась к разбиению графа на подграфы. Причём выделение областей одного цвета равносильно выделению подграфов, для которых ребра между собственными вершинами имеют больший вес, чем ребра связывающие с внешними вершинами. Такие подграфы принято называть сообществами (community).

Как было указано во введении, на сегодняшний день разработано достаточно много алгоритмов выделения сообществ в графах. Для кластеризации изображений используем один из них, предложенный Ньюманом [40].

Будем задавать построенный граф с помощью матрицы весов E . Элементы матрицы весов задаются следующим образом. Диагональный элемент E_{ii} показывает вес вершины с номером i (v_i). Элемент E_{ij} ($i \neq j$) показывает величину связи вершины v_i с вершиной v_j . Как показано в работе [40], более удобным является приведённый вид матрицы весов $e = E/m$, где $m = \sum E_{ij}$. В приведённой матрице весов элемент e_{ij} показывает долю веса заданного ребра в общем весе графа. В дальнейшем под матрицей весов будет пониматься именно приведённый вид. Легко увидеть, что $\sum e_{ij} = 1$.

Для выявления связанных структур введём величину модульности (modularity), зависящую от элементов матрицы смешения, и численно определяющую «силу» связности. Модульность задаётся в виде [40]:

$$Q(e) = \sum_{i=1}^{MN} (e_{ii} - p_{ii}),$$

где p_{ii} — «ожидаемая связность». В канонической модели [40] p_{ii} определяется через исходящую степень вершины a_i , и входящую степень вершины b_i :

$$p_{ii} = a_i b_i.$$

В этом случае модульность записывается в виде:

$$Q(e) = \sum_{i=1}^{MN} (e_{ii} - a_i b_i),$$

где

$$a_i = \sum_{j=1, j \neq i}^{MN} e_{ij}, \quad b_i = \sum_{j=1, j \neq i}^{MN} e_{ji}.$$

Определим более строго процедуру образования сообщества в графе. Начнём с алгоритма образования стяжек как преобразования графа G в граф G_1 . Выделим в графе G подграф G' и заменим все входящие в него вершины одной вершиной, при этом вершины подграфа $G \setminus G'$ остаются неизменными. Образованная вершина связана дугами с теми вершинами графа G_1 , с которыми были связаны вершины, вошедшие в стяжку. Вес вершины, вошедшей в стяжку, равен сумме весов вершин и дуг, вошедших в стяжку. Под сообществом будем понимать подграф исходного графа, который при образовании из него стяжки максимизирует модульность графа $Q(e)$. Нашей задачей является такое разбиение исходного графа изображения на сообщества, которое максимизирует $Q(e)$. Каждое такое сообщество представляет собой кластер исходного изображения.

При образовании стяжки изменяется граф и, как следствие, меняется значение функции $Q(e)$. Пусть произошло объединение вершин v_k и v_l в одну вершину v'_k , в результате чего граф G преобразовался в граф G' . Изменение функции модульности будет иметь вид:

$$\Delta Q(e) = Q(e') - Q(e) = 2(e_{kl} - a_k a_l).$$

Таким образом, изменение функции модульности графа может быть легко найдено из параметров графа G .

Разбиение графа на сообщества будем производить с помощью «жадного» алгоритма. Выберем одну из вершин графа v_0 , которая будет служить зерном сообщества C_0 . Пусть на некотором этапе построена часть сообщества C_k , содержащая вершины $\{v_{i1}, \dots, v_{ik}\}$. Рассмотрим точки, являющиеся ближайшими соседями каждой из вершин, входящих в C_k и не включённые в сообщество. Для каждого из ближайших соседей сообщества рассчитаем величину ΔQ_{kl} , показывающую изменение функции модульности при присоединении к C_k вершины v_l . Присоединим к сообществу ту вершину, для которой ΔQ_{kl} максимально, и при этом $\Delta Q_{kl} > 0$. Полученный граф обозначим C_{k+1} , модульность графа увеличим на ΔQ_{kl} и повторим процедуру. Продолжаем алгоритм до тех пор, пока будут находиться вершины, присоединение которых увеличивает модульность графа. Если в некоторый момент такие вершины отсутствуют, то сообщество считается сформированным и выбирается новое зерно, не входящее ни в одно сообщество, для формирования нового сообщества. Алгоритм продолжается до тех пор, пока все вершины не будут включены в какое-нибудь из сообществ. Трудоемкость данного алгоритма $O(M^2N^2)$.

Выбор зерна сообщества зависит от режима работы алгоритма кластеризации изображения. При автоматическом разбиении изображения на кластеры зерно нового сообщества может выбираться случайным образом. При интерактивном режиме работы зерно указывает пользователь.

3. Компьютерный эксперимент

Эксперимент проводился как на искусственных изображениях, так и на фотографиях различных объектов. Рассмотрим результаты работы алгоритма на искусственных изображениях с чёткими контурами и равномерной заливкой (рис.1). Как хорошо видно из рисунка, алгоритм абсолютно точно разделя-

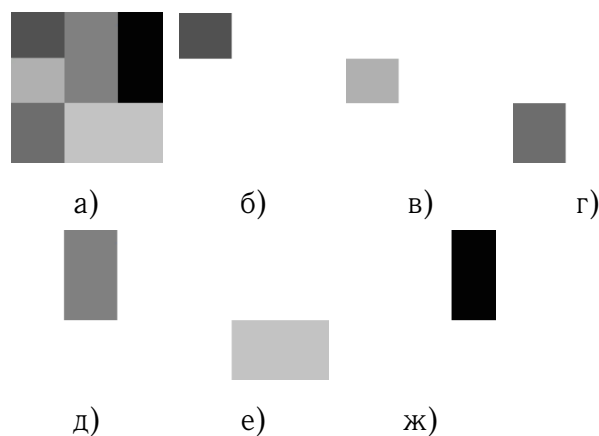


Рис. 1. Результат работы алгоритма по разбиению искусственного изображения с равномерной заливкой на связные области одного цвета: а) исходное изображение, б-ж) выделенные области.

ет изображение на области, причём результат не зависит от выбора значения параметра δ . Рассмотрим результаты работы алгоритма на искусственном изображении с градиентной заливкой (рис. 2). При уменьшении параметра

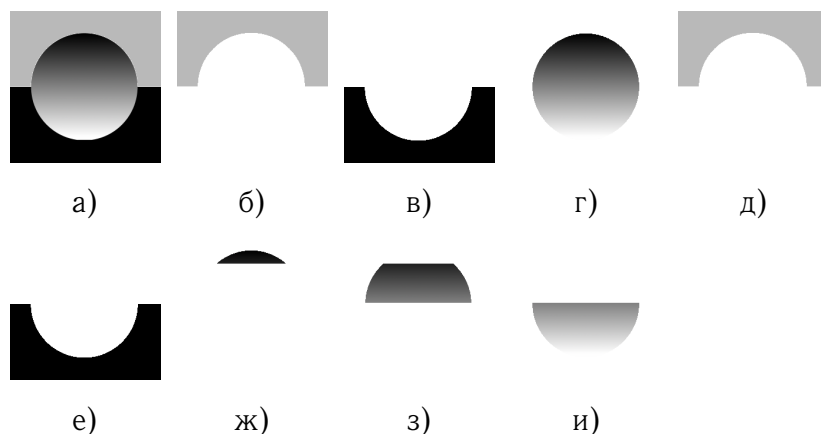


Рис. 2. Результат работы алгоритма по разбиению искусственного изображения с градиентной заливкой на связанные области одного цвета (размеры изображения 256x256): а) исходное изображение, б-г) выделенные области при $\delta = 0.04$, д-и) выделенные области при $\delta = 0.03$.

δ увеличивается количество кластеров, на которые разбивается изображение. Выделим отдельный кластер, например рисунок 3г, и проследим его разбиение при уменьшении параметра δ от значения 0.05 до значения 0.04. Результат представлен на рисунке 4.

Как видно из рисунка 2, при $\delta = 0.04$ внутренний круг с градиентной заливкой определяется как одна область, а при $\delta = 0.03$ он разбивается на четыре части. При уменьшении δ круг с градиентной заливкой разбивается на всё большее количество частей. Следует отметить, что при $\delta > 0.1$ всё изображение воспринимается как один кластер. Таким образом, выбор параметра δ влияет на значение скорости изменения цвета, которое определяется как граница между областями.

На рисунке 3 приведены наиболее крупные кластеры, получаемые на фотографическом изображении «Перцы» при $\delta = 0.05$, общее количество выделяемых кластеров равно 174.

Как видно из сравнения рисунков 3 и 4 для фотографических изображений, как и для искусственных изображений, варьирование параметра δ позволяет выбирать величину перепада цветов, которая считается границей области. На рисунке 5 приведён график зависимости количества кластеров, на которые разбивается изображение «Перцы», в зависимости от параметра δ .

Как видно из рисунка 5, количество кластеров достаточно быстро убывает с ростом значения δ , что свидетельствует о высокой чувствительности предложенного метода к выбору данного параметра. Начиная со значения $\delta = 0.1$, количество кластеров становится равным 1, то есть всё изображение воспринимается как один кластер.

Проведём исследование предложенного алгоритма на устойчивость к наличию случайно повреждённых пикселей. Компьютерный эксперимент состоял в

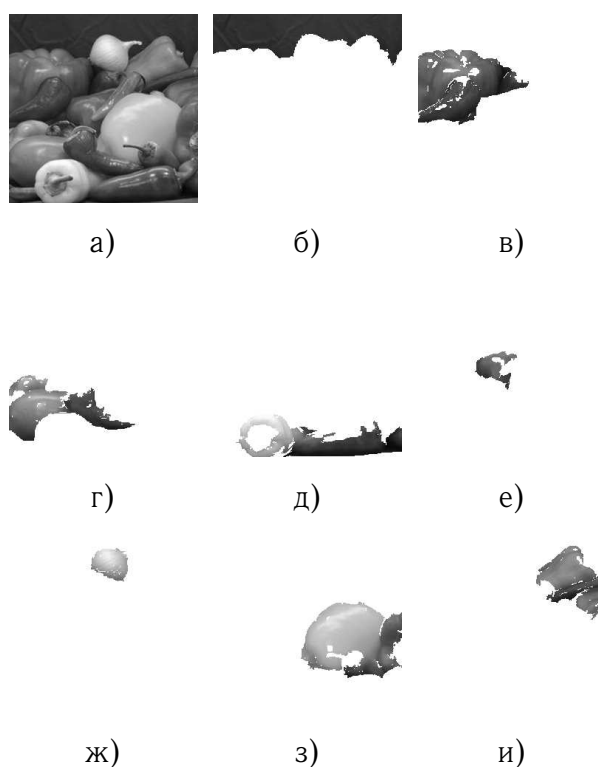


Рис. 3. Результат работы алгоритма по разбиению фотографического изображения «Перцы» при $\delta = 0.05$: а) исходное изображение, б-и) наиболее крупные выделенные области.

том, что p процентов случайно выбранных пикселей изображения заменялись на пиксели со случайным цветом. После чего осуществлялась кластеризация. При кластеризации не учитывались кластеры размером в один пиксель. Результаты эксперимента для изображения «Перцы» при $p = 10\%$ и $\delta = 0.5$ приведены на рисунке 6.

Как видно на рисунке 6, наличие повреждённых пикселей не влияет на выделение крупных кластеров, так подавляющее большинство повреждённых пикселей идентифицируется как кластеры размером один пиксель.

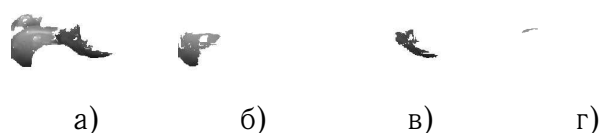


Рис. 4. Разбиение кластера на более мелкие кластеры при переходе от $\delta = 0.05$ к $\delta = 0.04$. а) один из кластеров, выделенных на изображении «Перцы» при $\delta = 0.05$, б-г) наиболее крупные кластеры, на которые разбивается исходный кластер.

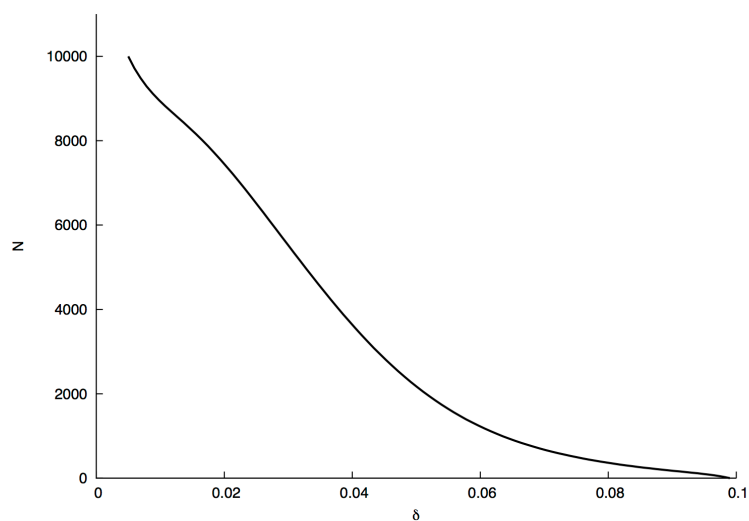


Рис. 5. График зависимости количества кластеров N , на которые разбивается изображение «Перцы», от параметра δ .

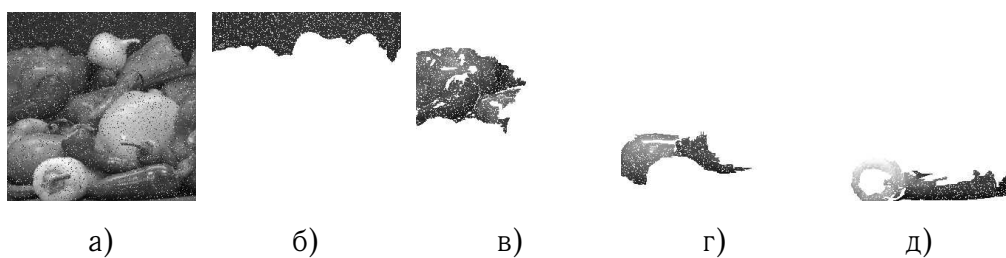


Рис. 6. Результат работы алгоритма по разбиению фотографического изображения «Перцы» повреждённого на 10% при $\delta = 0.05$: а) зашумлённое изображение, б-д) наиболее крупные кластеры.

4. Обсуждение результатов и выводы

Таким образом, предложенный подход к сегментации изображений, состоящий в представлении изображения в виде графа с последующим поиском сообществ (community), даёт достаточно хорошие результаты как на искусственно сформированных графических объектах, так и на фотографиях. Следует отметить, что разработанный алгоритм тестировался нами не только на монохромных изображениях, представленных в статье, но и на цветных рисунках и фотографиях и продемонстрировал хорошие результаты.

В сравнении с другими методами сегментации следует отметить, что предложенный подход использует разность интенсивностей цветов соседних пикселей, характерную для дифференциальных методов сегментации. Однако получаемые кластеры имеют чёткие границы, которые свойственны методам, основанным на абсолютном значении интенсивности, например, методе водораздела. Дополнительной положительной характеристикой предложенного алгоритма является устойчивость к импульсным шумам, которая состоит в малом влиянии испорченных пикселей на внешние границы крупных кластеров.

ЛИТЕРАТУРА

1. Шапиро Л., Стокман Дж. Компьютерное зрение. М. : Бином, 2006. 752 с.
2. Tek H., Kimia B.B. Volumetric Segmentation of Medical Images By Three-Dimensional Bubbles // CVIU – 1997. V. 64(2). P. 246–258.
3. Dogdas B., Shattuck D.W., Leahy R.M. Segmentation of Skull in 3D Human MR Images Using Mathematical Morphology // Proceedings of SPIE Medical Imaging Conference. 2002. V. 4684(180). P. 1553–1562.
4. D.L. Pham, C. Xu and J.L. Prince. Current Methods in Medical Image Segmentation // Annual Review of Biomedical Engineering. 2000. V. 2. P. 315–337.
5. Koepfler G., Lopez C., Morel J.M. A Multiscale Algorithm for Image Segmentation by Variational Method // SIAM J. on Numerical Analysis. 1994. V. 31, No. 1. P. 282–299.
6. Newsam S., Bhagavathy S., Fonseca L., Kenney C., Manjunath B. S. Object based representations of spatial images // Acta Astronautica. 2001. V. 48, No. 5–12. P. 567–577.
7. Siebert A. Segmentation based Image Retrieval // Proc. SPIE 3312, Storage and Retrieval for Image and Video Databases VI. 1997. V. 14. doi:10.1117/12.298444.
8. Цапаев А.П., Кретинин О.В. Методы сегментации изображений в задачах обнаружения дефектов поверхности // Компьютерная оптика. 2012. Т. 36, № 3. С. 448–452.
9. Мошкин В.И., Титов В.С., Якушенков Ю.Г. Техническое зрение роботов / Под общ. ред. Ю.Г. Якушенкова. М. : Машиностроение, 1990. 272 с.
10. Bruce J., Balch T., Veloso M. Fast and in-expensive color image segmentation for interactive robots // Proceedings of the 2000 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '00). 2000. V. 3. P. 2061–2066.
11. Павлидис Т. Алгоритмы машинной графики и обработки изображения. М. : Радио и связь, 1986. 400 с.

12. Прэтт У. Цифровая обработка изображений. Кн. 1. М. : Мир, 1982. 312 с.
13. Анисимов Б.В., Курганов В.Д., Злобин В.К. Распознавание и цифровая обработка изображений. М. : Высшая школа, 1983. 295 с.
14. Vincent L., Soille P. Watersheds in Digital Spaces: An Efficient Algorithm Based on Immersion Simulations // IEEE Trans. on Pattern Analysis and Machine Intelligence. 1991. V. 13, No. 6. P. 583–598.
15. Самойленко Д. Е. Структурная сегментация изображений // Искусственный интеллект. 2004. № 4. С. 512–529.
16. Гонсалес Р., Вудс Р. Цифровая обработка изображений. М. : Техносфера, 2005. 1072 с.
17. Rosenfeld A., Kak A.C. Digital Picture Processing. New York : Academic press, 1976.
18. Робертс Л. Автоматическое восприятие трёхмерных объектов // В кн.: Интегральные роботы. Т. 1. М. : Мир, 1973. С. 162–208.
19. Prewitt J.M.S. Object Enhancement and Extraction // in Picture Processing and Psychopictories. New York : Academic Press, 1970. P. 75–150.
20. Clark J. J. Authenticating Edges Produced by Zero-Crossing Algorithms // IEEE Trans. Pattern Analysis Machine Intelligence. 1989. V. 12, No. 8. P. 830–831.
21. Fu K.S., Mui J.K. A Survey of Image Segmentation // Pattern Recognition. 1981. V. 13, No. 1. P. 3–16.
22. Haralick R.M., Shapiro L.G. Image Segmentation Techniques // Computer Vision, Graphics, Image Processing. 1985. V. 29, No. 2. P. 100–132.
23. Дуда Р., Харт П. Распознавание образов и анализ сцен. М. : Мир, 1976. 511 с.
24. Jain A.K., Dubes R.C. Algorithms for Clustering Data. Prentice Hall, 1988.
25. Matas J. Kittler J. Spatial and Feature Space Clustering: Applications in Image Analysis // 6th Int. Conf. on Computer Analysis and Patterns. Prague, Czech Republic, 1995.
26. Яне Б. Цифровая обработка информации. М. : Техносфера, 2007.
27. Ohta Y., Kanade T., Sakai T. Color Information for Region Segmentation // Computer Graphics, Image Processing. 1980. V. 13. P. 224–241.
28. Pal N.K., Pal S.K. A Review on Image Segmentation Techniques // Pattern Recognition. 1993. V. 26, No. 9. P. 1277–1293.
29. Haralick R.M. Image Texture Survey // Fundamentals in Computer Vision. Cambridge. 1983. P. 145–172.
30. Миронов Б.М., Малов А.Н. Сегментация изображений кластерным методом и алгоритмом случайных скачков: сравнительный анализ // Компьютерная оптика. 2010. Т. 34, № 1. С. 132–137.
31. Форсайт Д.А. Компьютерное зрение. Современных подход. М. : Издательский дом «Вильямс», 2004.
32. Jitendra M., Belongie S., Leung T., Shi J. Contour and Texture Analysis for Image Segmentation // International Journal of Computer Vision. 2001. V. 43, No. 1. P. 7–27.
33. Boykov Y., Jolly M.-P. Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images // Proc. of the International Conference on Computer Vision. 2001. V. 1. P. 105–112.
34. Boykov Y., Jolly M.-P. Interactive organ segmentation using graph cuts // Medical Image Computing and Computer-Assisted Intervention. 2000. P. 276–286.
35. Felzenszwalb P.F., Huttenlocher D.P. Efficient graph-based image segmentation //

- International Journal of Computer Vision. 2004. V. 59(2). P. 167–181.
36. Boykov Y., Kolmogorov V. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2004. P. 1124–1137.
 37. Белим С.В., Кутлунин П.Е. Использование алгоритма кластеризации для разбиения изображения на односвязные области. // Наука и образование: электронное научно-техническое издание. 2015, N. 3, URL: <http://technomag.bmstu.ru/doc/759275.html>, DOI:10.7463/0315.0759275 (Дата обращения 13.06.2016).
 38. Белим С.В., Кутлунин П.Е. Выделение контуров на изображениях с помощью алгоритма кластеризации // Компьютерная оптика. 2015, Т. 39, № 1, С. 119–124.
 39. Fortunato S. Community detection in graphs // Physics Reports 486. 2010. P. 75–174
 40. Newman M.E.J. Mixing patterns in networks // Phys. Rev. E. 2003. V. 67. P. 026126.

SEGMENTATION OF IMAGES ON THE BASIS OF ALGORITHM OF COMMUNITIES SEPARATION ON THE GRAPH

S.V. Belim

Dr.Sc. (Phys.-Math.), Professor, e-mail: belimsv@omsu.ru

S.B. Larionov

Postgraduate Student, e-mail: me@stas-larionov.ru

Dostoevsky Omsk State University

Abstract. In the article we propose the image segmentation algorithm using its representation as an undirected weighted graph, followed by separation of the communities on the graph. Each community corresponds to the image segment. The function of the difference between the intensity of the colors in the space of RGB increasing exponentially is used as the measure of arcs weight. To separate the communities a greedy algorithm is built. The modularity function introduced by Newman is used to assess the quality of the graph separation into communities. Computer experiment is provided. It is shown that segments sizes depend on the parameter value in the edges weight function. The proposed algorithm is robust to impulse noise.

Keywords: segmentation of image, community, graphs.

МЕТОД И АЛГОРИТМ ВЫЧИСЛЕНИЯ ОБЪЁМА ЗАДАННОЙ ДИСКРЕТНО 3D ПОВЕРХНОСТИ С ПОМОЩЬЮ ПОЛИНОМОВ ЛАГРАНЖА

К.Т. Кошеков

профессор, д.т.н., заведующий кафедрой «Энергетики и радиоэлектроники»,
e-mail: kkoshekov@mail.ru

Н.В. Астапенко

докторант специальности 6D075100 «Информатика, вычислительная техника
и управление», e-mail: astankin@mail.ru

Северо-Казахстанский государственный университет имени М. Козыбаева

Аннотация. При хранении габаритных объектов с неровной поверхностью (например, сыпучих) на различных складах возникает задача вычисления объёма. В данной статье предлагается алгоритм расчёта объёма объекта на основе данных, полученных после сканирования поверхности. Данный алгоритм может использоваться в том случае, когда нет иных способов определения объёма. Результаты сканирования представлены в виде значений узлов регулярной сетки. В представленном алгоритме предлагается с помощью полиномов Лагранжа аппроксимировать значения высот объекта между узлами сетки с необходимым шагом и вычислять объём путём сложения прямоугольных параллелепипедов. Прямоугольные параллелепипеды, составляющие объект, имеют высоту, вычисленную с помощью полинома Лагранжа, и площадь основания, равную квадрату шага. Конечно, представленный алгоритм является не оптимизированным и в дальнейшем планируется его доработка. Тем не менее, алгоритм является рабочим и неплохо справляется с небольшими фрагментами. Суммировать объёмы для всех фрагментов объекта не составляет труда. Особая ценность представленного материала в том, что описанный алгоритм до сих пор не рассматривался в литературе.

Ключевые слова: алгоритмы, полином Лагранжа, регулярная сетка, численные методы.

Введение

В разных областях человеческой деятельности возникает задача исследования параметров габаритного объекта, в частности измерения объёмов сыпучих тел. Традиционно используются два метода измерения объёма неровных тел: вычисление объёма тела при помощи ёмкости с водой и вычисление объёма тела через разбиение его на правильные геометрические фигуры. Первый метод очень сложно организовать с габаритными объектами. Реализация второго

представляет собой сложную задачу, если нельзя явно выделить правильные геометрические фигуры. В данной статье предлагается алгоритм вычисления объёма габаритного объекта с неровной поверхностью, заданной дискретными значениями.

1. Метод вычисления объёма поверхности

В результате сканирования объекта мы получаем набор значений высот поверхности, расположенных в узлах регулярной сетки. На рисунке представлен фрагмент объекта размером метр на метр, расстояние между узлами 1 дм (см. рис. 1).

Объём данной фигуры можно представить как сумму объёмов прямоугольных параллелепипедов. Если мы возьмём прямоугольные параллелепипеды в соответствии с размерами регулярной сетки, то получим большую погрешность даже на представленном участке 1 дм × 1 дм. Найти точки, расположенные между узлов сетки, можно с помощью аппроксимации.

Наиболее целесообразно для аппроксимации сложных пространственных поверхностей, заданных координатами узловых точек, применять многомерные полиномы. В отличие от описания поверхности сплайн-функциями, данный метод позволяет исключить колебательный процесс, который возникает в результате совпадения точек поверхности в узловых точках и отсутствия гладкости полученного описания в промежутках между опорными точками поверхности. В данном случае аппроксимация поверхности между двумя точками основывается на знании координат опорной точки и частных производных в данной точке. Это не накладывает требований на гладкость поверхности между опорными точками.

Рассмотрим применение для этих целей многомерных полиномов Лагранжа, зависящих от двух переменных. Определение объёма объекта тогда будет сведено к решению следующих задач:

1. Приближенное описание поверхности полиномами по заданным координатам опорных точек поверхности в системе координат (XYZ) .
2. Определение значений точек поверхности с заданным шагом в соответствии с полученными полиномами Лагранжа.
3. Нахождение суммы объёмов прямоугольных параллелепипедов, установленных в рассчитанных точках.

Интерполяционные полиномы Лагранжа одной переменной позволяют аппроксимировать функцию $y = f(x)$ в декартовой системе координат (XY) , задаваемую координатами опорных точек (x_i, y_i) . Аппроксимированная функция описана в формуле (см. формулу 1).

$$f(x) = P(x) = \sum_{i=0}^n p(x_i) \times \prod_{j=0, j \neq i}^n (x - x_j), \quad (1)$$

где коэффициенты полиномов Лагранжа $p(x_i)$ определяются через значения x_i, y_i в опорных точках. Коэффициент полинома вычисляется по формуле (см. формулу 2).

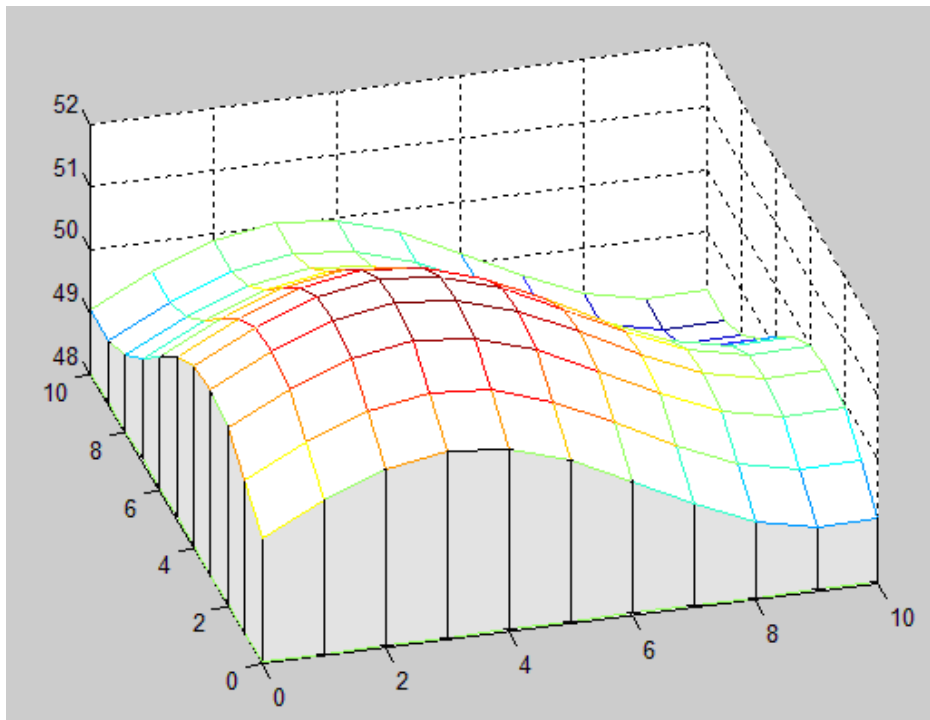


Рис. 1. Фрагмент исследуемого объекта

$$p(x_i) = \frac{y_i}{\prod_{j=0, j \neq i}^n (x - x_j)}, \quad (2)$$

где $i = 0..n$, $j = 0..n$, n — степень полинома. По аналогии с полиномом одной переменной можно описать полином Лагранжа двух переменных для поверхности, представленной в декартовой системе координат (XYZ):

$$z = P(x, y) = \sum_{i=0}^n \sum_{j=0}^m p(x_i, y_j) \times \prod_{n=0, n \neq i}^n (x - x_n) \prod_{m=0, m \neq j}^m (y - y_m), \quad (3)$$

где $i = [0..n]$ — количество опорных сечений поверхности вдоль оси OX ; $j = [0..m]$ — количество опорных сечений поверхности вдоль оси OY ; $p(x_i, y_j)$ — коэффициенты полинома, определяемые через соответствующие координаты опорных точек поверхности. При интерполяции поверхности полиномами двух переменных необходимо, чтобы опорные точки поверхности (узлы интерполяции) образовывали сетку. Наиболее удобной является прямоугольная сетка с равномерным распределением клеток. В этом случае поверхность, представляемая координатами узлов прямоугольной сетки (см. рис. 1), в зависимости от расположения текущих координат (x, y) поверхности последовательно «накрывается» прямоугольником. При этом для более точной аппроксимации необходимо, чтобы текущие координаты поверхности находились в области центра прямоугольника.

Постоянные коэффициенты полинома $p(x_i, y_i)$ определяются для каждого элемента поверхности через координаты известных опорных точек поверхности и определяются по формуле (см. формулу 4).

$$P(x^i, y_j) = \frac{z_{ij}}{\prod_{n=0, n \neq i} \prod_{m=0, m \neq j} (x - x_n)(y - y_m)}. \quad (4)$$

2. Алгоритм расчёта объёма 3D поверхности

Рассмотрим алгоритм работы программы, вычисляющей объем фигуры, заданной дискретными значениями.

Шаг 1. Исходные данные (результат сканирования) представлены двумерным массивом (см. рис. 2).

50.000	50.479	50.841	50.997	50.909	5
50.479	50.959	51.321	51.477	51.389	5
50.841	51.321	51.683	51.839	51.751	5
50.997	51.477	51.839	51.995	51.907	5
50.909	51.389	51.751	51.907	51.819	5
50.598	51.078	51.440	51.596	51.508	5
50.141	50.621	50.983	51.139	51.050	5
49.649	50.129	50.491	50.647	50.559	5
49.243	49.723	50.085	50.241	50.152	4
49.022	49.502	49.864	50.020	49.932	4
49.041	49.521	49.883	50.039	49.950	4

Рис. 2. Фрагмент массива исходных данных

Номер строки соответствует координате x в декартовой системе координат, а номер столбца — координате y . Значения x и y изменяются с шагом 1, что соответствует расстоянию 1 дм между точками съёмки. Размер массива 10×10 , то есть представлен фрагмент площадью 1 м^2 . Таким образом, элементы представленной матрицы соответствуют высоте заданной точки или координате z в декартовой системе координат (XYZ).

Шаг 2. Вычисление элементов массива $p(x_i, y_i)$ реализуется в соответствии с формулой 4. Алгоритм вычисления представлен блок-схемой (см. рис. 3).

На входе в подпрограмму подаётся массив исходных данных $a[n, n]$. В цикле с параметром k определяется знаменатель для очередного коэффициента полинома. В результате коэффициент определяется как отношение соответствующего элемента исходного массива к вычисленному знаменателю tmp .

Шаг 3. Теперь, когда коэффициенты Лагранжа получены, можно вычислить объем фигуры путём сложения прямоугольных параллелепипедов с квадратным основанием заданного размера. Коэффициенты выполняются в соответствии с формулой 3. На рисунке представлена блок-схема подпрограммы расчёта объёма объекта с заданным шагом h , который определяет сторону квадратного основания прямоугольного параллелепипеда (см. рис. 4).

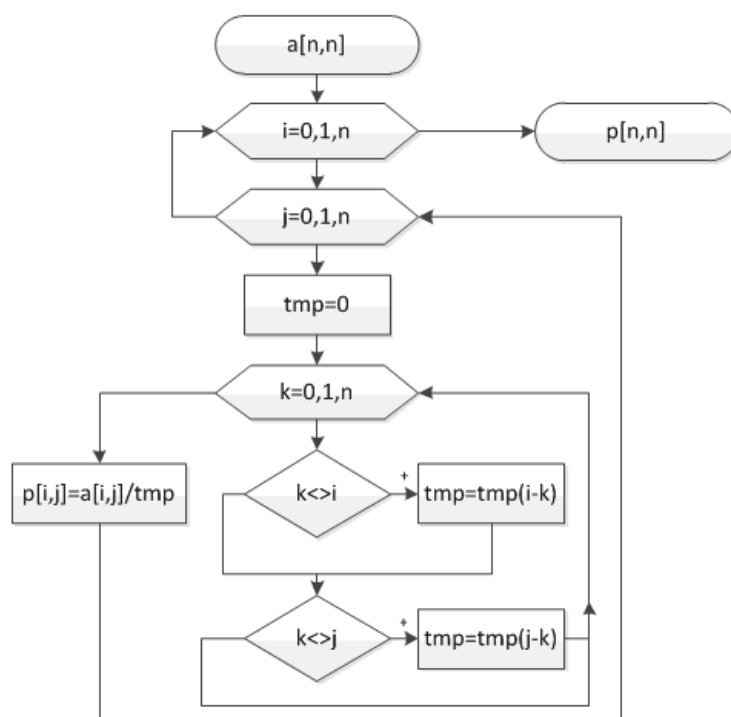


Рис. 3. Блок-схема подпрограммы расчёта коэффициентов полинома Лагранжа

Вычисление сводится к двум этапам:

- определение высоты прямоугольного параллелепипеда;
- добавление объёма очередного параллелепипеда в накопитель.

На входе в подпрограмму подаются размер шага h и матрица коэффициентов полинома Лагранжа $p[n, n]$. Начиная с точки $(0, 0, z_{00})$, определяется величина z_{ij} , где $i = [0..n]$ — значение координаты оси OX , а $j = [0; n]$ — значение координаты по оси OY .

Объём прямоугольного параллелограмма вычисляется традиционно путём умножения высоты на ширину и длину $(h \cdot h \cdot z_{ij})$.

После определения объёма очередного прямоугольного параллелограмма, полученная величина добавляется в накопитель V . В результате перебора всех точек (x, y) на заданном поле с шагом h в переменной V получаем искомый объём.

3. Заключение

Нами были проведены экспериментальные расчёты объёмов с помощью реализованного алгоритма. Для проведения эксперимента были получены данные сканирования четырёх фрагментов с площадями 1 м^2 . Для каждого фрагмента был рассчитан объём поверхности при разных размерах шага.

По полученным в результате проведённого эксперимента данным можно сде-

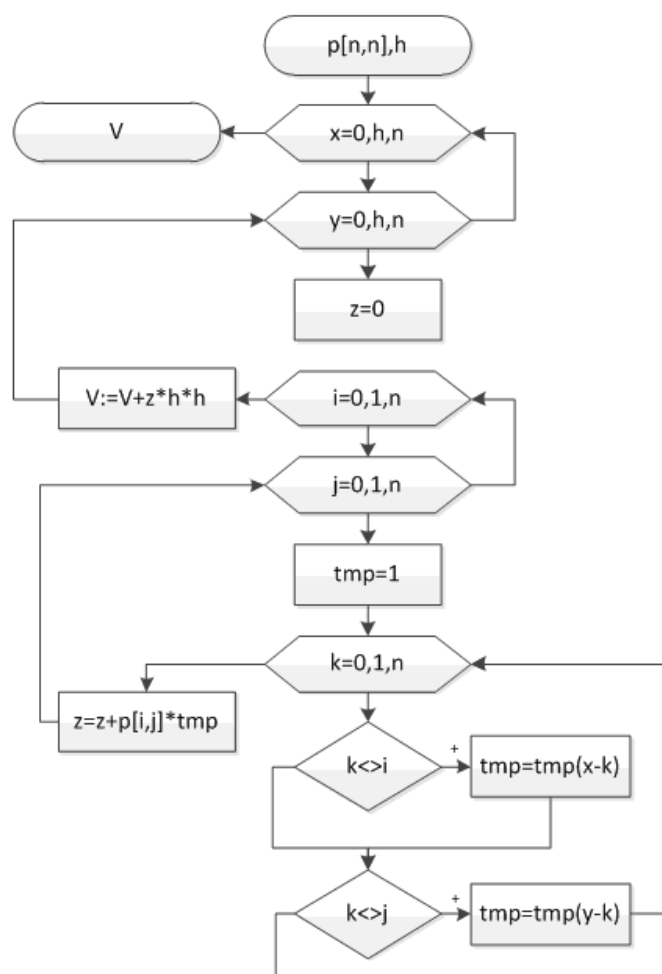


Рис. 4. Блок-схема подпрограммы расчёта объёма объекта с заданным шагом

лать вывод о том, что точность вычислений тем выше, чем меньше шаг и, следовательно, чем больше точек задействовано в вычислениях. Но, к сожалению, большое количество точек увеличивает сложность алгоритма, что заметно при больших массивах данных. В целом, проведённый эксперимент показал, что с помощью данного алгоритма можно рассчитать объём зерна, хранимого на складе. В дальнейшем планируется предпринять действия по уменьшению погрешности вычислений и увеличению быстродействия алгоритма.

ЛИТЕРАТУРА

1. Афонин В., Макушкин В. Учебный курс Интеллектуальные робототехнические системы / Национальный открытый университет «Интуит». URL: <http://www.intuit.ru/> (дата обращения: 30.09.2015).

2. Демидович Б.П., Марон И.А. Основы вычислительной математики: учебное пособие. 6-е изд. М. : Наука, 1966. 672 с.
3. Колдаев В.Д. Численные методы и программирование: учебное пособие. М. : Форум. ИНФРА-М, 2008. 336 с.

**METHOD AND ALGORITHM FOR CALCULATING THE VOLUME
OF DETERMINED DISCRETE 3D SURFACE USING LAGRANGE
POLYNOMIALS**

K.T. Koshekov

Dr.Sc. (Eng.), Professor, Head of the Department "Energy and Electronics",
e-mail: kkoshekov@mail.ru

N.V. Astapenko

Doctoral Specialty 6D075100 "Computer Science, Computer Facilities and Management",
e-mail: astankin@mail.ru

M. Kozybayev North Kazakhstan State University

Abstract. When the overall objects with an uneven surface (e.g., bulk) are stored at various warehouses, there is a problem of calculating the volume. This article describes an algorithm for calculating the volume of the object based on the data obtained after the surface scanning. This algorithm may be used when there is no other method for volume determining. The scan results are presented as the values of the nodes of a regular grid. The presented algorithm uses the Lagrange polynomials to approximate the height values of the object between grid points with the step you need, and calculates the amount by adding cuboids. The cuboids composing object have the height calculated by using the Lagrange polynomial and the base area equal to the square of the step. Of course, the present algorithm is non-optimized and to be developed. However, the algorithm works well with small fragments. To summarize the volumes for all of the fragments of the object is not difficult. The special value of the material presented is that the algorithm has not yet been considered in the literature.

Keywords: algorithms, Lagrange polynomial, regular grid, numerical methods..

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ СИЛЬНОГО СЖАТИЯ ОДНОМЕРНЫХ СЛОЁВ ГАЗА В КОНФИГУРАЦИИ Р. МИЗЕСА

Н.С. Новаковский

аспирант, e-mail: n.s.novakovskiy@yandex.ru

Уральский государственный университет путей сообщения (УрГУПС)

Аннотация. В статье изложены результаты численного исследования задачи сильного сжатия одномерных слоёв газа в конфигурации Р. Мизеса. Приведена математическая постановка задачи в виде двух характеристических задач Коши (ХЗК1 и ХЗК2) и одной задачи Коши для восстановления траектории движения сжимающего поршня в обратном направлении изменения времени. Приводятся алгоритмы численного решения этих двух ХЗК (полностью) при помощи стандартного метода характеристик с предварительным раскрытием особенности. Описан алгоритм восстановления траектории сжимающего поршня. Описан конечно-разностный метод «РОМБ» для численного решения системы уравнений газовой динамики (СУГД) в прямом направлении изменения времени. Представлены результаты численных исследований, в том числе сравнение с точным решением.

Ключевые слова: сильное сжатие газа, характеристическая задача Коши, центрированная волна (ЦВ) сжатия, метод характеристик, конечно-разностный метод «РОМБ».

Введение

Математическое описание процесса безударного изоэнтропического сжатия идеального газа до любого наперёд заданного значения плотности, в том числе до бесконечной плотности представляет интерес в связи с проблемой лазерного термоядерного синтеза [1, 2] (более подробную библиографию см. в [3]).

В [3, 7] разработана математическая теория безударного сильного сжатия идеального газа. В частности, для случая сжатия цилиндрически $\nu = 1$ и сферически $\nu = 2$ симметричных слоев политропного газа с показателем $\gamma > 1$ доказано, что непрерывная состыковка двух течений даёт решение задачи о безударном сильном сжатии до любой наперёд заданной плотности ненулевой массы газа.

В работе [10] предложен алгоритм расчёта безударного сильного сжатия в обратном направлении изменения времени одномерных слоёв первоначально однородного и покоящегося газа с $\rho_0 = 1$ до любой наперёд заданной конечной постоянной плотности $\rho_* > 1$ с последующим восстановлением закона движения поршня, сжимающего слой газа. В работе [11] представлены результаты расчётов вышеупомянутым алгоритмом ряда одномерных задач с различной симметрией. Максимально достигнутое значение плотности в расчётах ЦВ Римана из работы [12] в $3 \cdot 10^4$ раз превышает исходную плотность. При этом среднее значение плотности в ЦВ равно $8 \cdot 10^3$.

Результаты расчётов сильного сжатия двумерных областей приводятся в работе [13]. В работах [14, 15] подробно рассматриваются расчёты автомодельных двумерных и трёхмерных задач сильного сжатия. В работах [12, 14, 15] расчёты проведены в прямом направлении изменения времени, а движение сжимающего поршня происходит в соответствии с точными решениями, явно полученными в более ранних работах [4, 16, 17].

Основная цель выполнения данной работы обусловлена требованиями со стороны физиков-прикладников, которые чётко сформулировал академик Е.Н. Аврорин: «Какие конкретные рекомендации для физического эксперимента дают результаты вашего математического моделирования сильного сжатия?» Эта цель формулируется так: с использованием ранее опубликованного алгоритма построить закон движения сжимающего поршня и расчётами в прямом направлении изменения времени конечно-разностным методом «Ромб» [18] определить возможность сжатия газа с помощью найденного закона движения сжимающего поршня до нужных степеней плотности.

В работе кратко описан алгоритм расчёта в обратном направлении изменения времени безударного сильного сжатия изнутри (то есть при увеличении радиуса сжимаемого слоя) одномерных слоёв первоначально однородного и покоящегося газа в диапазоне от $\rho_0 = 1$ до любой наперёд заданной конечной плотности $\rho > 1$. Представлены тестовые расчёты в обратном направлении изменения времени для плоского случая $\nu = 0$, а также некоторые результаты расчётов сжатия цилиндрических ($\nu = 1$) и сферических ($\nu = 2$) слоёв фиксированной массы до наперёд заданного конечного значения плотности. Приводятся некоторые сведения о построении конечно-разностного метода «Ромб», необходимые для понимания сконструированного алгоритма.

Основным результатом данной работы являются численные расчёты задачи сильного сжатия плоского симметричного слоя в прямом направлении изменения времени конечно-разностным методом «Ромб». Расчёты при использовании численно восстановленного закона движения сжимающего поршня.

1. Математическая постановка одномерной задачи безударного сильного сжатия газовых слоёв

Далее рассматривается изоэнтропическое сжатие одномерного газового слоя изнутри. Подобные течения являются решениями системы уравнений [3]:

$$\begin{cases} \sigma_t + u\sigma_r + \frac{(\gamma-1)}{2}\sigma\left(u_r + \nu\frac{u}{r}\right) = 0, \\ u_t + \frac{2}{(\gamma-1)}\sigma\sigma_r + uu_r = 0. \end{cases} \quad (1)$$

Здесь $\sigma = \rho^{(\gamma-1)/2}$ – скорость звука в газе, ρ – плотность, $\gamma > 1$ – константа в уравнении состояния $p = \rho^\gamma/\gamma$, p – давление; $\mathbf{U} = (\sigma, u)$ – искомые функции. Значения $\nu = 0, 1, 2$ соответствуют случаям плоской, цилиндрической и сферической симметрии.

В случае $\nu = 0$ в (1) $r = x_1$, $u = v_1$ – проекция вектора $\mathbf{V}$ на ось Ox_1 .

В случае $\nu = 1, 2$ $r = \sqrt{\sum_{i=1}^{\nu+1} x_i^2} \geq 0$, а скорость газа u есть проекция вектора $\mathbf{V}$ на радиус-вектор в плоскости x_1Ox_2 при $\nu = 1$ или на радиус-вектор в пространстве переменных x_1, x_2, x_3 при $\nu = 2$ соответственно.

Пусть в некоторой окрестности точки $(t = t_*, r = r_*)$ при $r_* > 0$ задано какое-либо фоновое течение $\mathbf{U} = \mathbf{U}^o(t, r)$, где компоненты вектора $\mathbf{U}^o(t, r)$ – аналитические функции $\sigma^o(t, r)$, $u^o(t, r)$ – являются решением системы (1). На рис. 1 в области 0 находится фоновое течение, а в области 1 требуется построить искомое течение с наперёд заданным распределением одного из газодинамических параметров, которое будет сопряжено с фоновым течением через слабый разрыв.

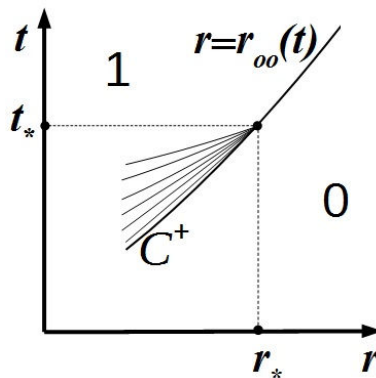


Рис. 1. Области однородного покоя и ЦВ сжатия

Слабым разрывом будет являться звуковая характеристика C^+ , однозначно определяющаяся при решении задачи Коши:

$$\frac{dr}{dt} = u^o(t, r) + \sigma^o(t, r), r(t)|_{t=t_*} = r_*. \quad (2)$$

Из аналитичности фонового течения следует существование и единственность решения данной задачи — аналитической функции $r = r_{oo}(t)$, описывающей закон движения характеристики C^+ . Далее везде считается, что функция $r = r_{oo}(t)$ известна и, следовательно, известны значения газодинамических параметров фонового течения на этой C_0^+ -характеристике:

$$U^o(t, r)|_{r=r_{oo}(t)} = U_{oo}(t),$$

т.е.

$$\sigma|_{C_0^+} = \sigma_{oo}(t), \quad u|_{C_0^+} = u_{oo}(t). \quad (3)$$

Задача (1), (3) является характеристической задачей Коши, для существования и единственности решения которой необходимо задать дополнительное краевое условие.

1.1. Задача о получении наперёд заданного непрерывного распределения плотности

Рассмотрим случай, когда звуковая характеристика

$$C^+ : r = (t - t_*) + r_*$$

разделяет однородный покой и искомое течение. Тогда на ней определены условия:

$$u|_{C^+} = 0, \sigma|_{C^+} = 1.$$

Область определения искомого течения ограничена прямой $t = t_*$ и звуковой характеристикой C_0^+ фонового течения.

Требуется получить в наперёд заданное распределение плотности $\rho|_{t=t_*} = \rho_*(r)$. К системе (1), (3) добавляется следующее краевое условие

$$\sigma|_{t=t_*} = \sigma_*(r) = \rho_*^{(\gamma-1)/2}(r). \quad (4)$$

Предполагается, что наперёд заданное распределение плотности в момент времени $t = t_*$ в точке $r = r_*$ непрерывно стыкуется с плотностью газа фонового течения.

Задача (1), (3), (4) является ХЗК стандартного вида и имеет в окрестности точки $(t = t_*, r = r_*)$ единственное аналитическое решение [3, 9].

Далее предположим, что в момент времени $t = t_*$ необходимо в точке $r = r_*$ получить вертикальное распределение плотности. Тогда вместо условия (4) необходимо задать условие вертикали. Такая задача, далее называемая ХЗК1, имеет единственное аналитическое решение в окрестности точки $(t = t_*, r = r_*)$ [3].

Течение в области 1 (см. рис. 1) имеет особенность в точке $(t = t_*, r = r_*)$ и описывается ЦВ Римана в случае плоской симметрии ($\nu = 0$)

$$\begin{aligned} u(t, r) &= \frac{2}{\gamma-1}\sigma - \frac{2}{\gamma-1}\text{const}, \\ \frac{r-r_*}{t-t_*} &= \frac{\gamma+1}{\gamma-1} - \frac{2}{\gamma-1}\text{const} \end{aligned} \quad (5)$$

или её обобщением при $\nu = 1, 2$ (см. [3]).

Далее предполагаем, что искомое течение является составным. Область определения решения состоит из двух частей: нижнего треугольника (обозначен цифрой 1 на рис. 2) — области определения обобщения ЦВ Римана и верхнего треугольника (обозначен цифрой 2 на рис. 2) — области определения течения, имеющего в момент $t = t_*$ наперёд заданное распределение $\sigma = \sigma^*(r)$, например, постоянное — $\sigma^*(r) = \sigma^* = \text{const}$.

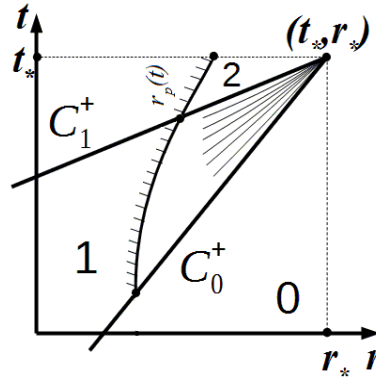


Рис. 2. Области определения ХЗК1 и ХЗК2

Характеристика C_1^+ определяется решением соответствующей задачи Коши, которая имеет единственное аналитическое решение, поскольку ХЗК1 имеет аналитическое решение. На этой характеристике значения параметров газа являются аналитическими функциями

$$\sigma = \sigma_1(t), \quad u = u(t, \sigma_1(t)) \equiv u_1(t), \quad (6)$$

определяемыми из решения ХЗК1. В области 2 ставится новая ХЗК, называемая ХЗК2. К системе (1) добавляются условия на характеристике C_1^+ и краевые условия

$$\begin{aligned} \sigma(t, r)|_{r=r_1(t)} &= \sigma_1(t), \\ u(t, r)|_{r=r_1(t)} &= u(t, \sigma_1(t)), \\ \sigma(t, r)|_{t=t_*} &= \sigma_*(r), \quad \sigma_*(r_*) = \sigma_1(t_*). \end{aligned} \quad (7)$$

Аналитическая функция $r_1(t)$ считается известной, для неё выполняется соотношение

$$\frac{dr_1(t)}{dt} = u_1(t) + \sigma_1(t). \quad (8)$$

Функция $\sigma = \sigma_*(r)$ считается аналитической, известной и связанной соотношением $\sigma_*(r) = [\rho_*(r)]^{(\gamma-1)/2}$ с наперёд заданным в момент $t = t_*$ требуемым распределением плотности $\rho = \rho_*(r)$. Тогда ХЗК2 (1), (8), (7) имеет в окрестности точки (t_*, r_*) единственное аналитическое решение [3, 9].

1.2. Задача Коши для восстановления траектории поршня

Требуется найти закон движения поршня, безударно сжимающего однородный покоящийся газ массой m_* с плотностью $\rho = 1$ до плотности $\rho = \rho(t_*, r) = \rho_* = \text{const}$, постоянной в сжатом слое шириной $d_* = m_*/\rho_*$ (рис. 2). Эта траектория определяется при решении в обратном направлении изменения времени следующей задачи Коши:

$$\begin{cases} \frac{dr}{dt} = u(t, r), \\ r|_{t=t_*} = r_* - d_*. \end{cases} \quad (9)$$

В (9) в правой части уравнения стоит функция $u(t, r)$, определяющаяся при решении ХЗК2.

Теоремы, доказанные в [3], утверждают, что существует ненулевая масса газа m_* , расположенная в области между траекторией движения поршня $r = r_p(t)$ и прямой $r = r_*$, которую можно безударно сжать до любой наперёд заданной плотности $\rho_*(r)$. Но эти теоремы не определяют максимально возможное значение m_* , а следовательно и максимальную ширину слоя d_* .

2. Алгоритм численного решения одномерной задачи сильного сжатия газа методом характеристик в обратном направлении изменения времени

Приведём далее алгоритм, основанный на стандартном методе характеристик с пересчётом [19] и ранее подробно изложенный в [3, 10]. Здесь приводятся лишь узловые моменты алгоритма и схема расчёта.

1. С помощью введения инвариантов Римана система (1) сводится к системе двух обыкновенных дифференциальных уравнений (ОДУ). К системе присоединяются дифференциальные уравнения для характеристик из (2), а также учитываются начальные условия (3) и краевое условие.

2. Система ОДУ решается отдельно в области 1 и в области 2.

2.1. Алгоритм построения численного решения в областях 1 и 2

Расчёт сетки в нижнем треугольнике (область 1 на рис. 2) и вычисление значений параметров газа в её узлах происходят в обратном направлении изменения времени. Схема расчёта проиллюстрирована на рис. 3.

А. В точке (r_*, t_*) интервал $[1, \sigma_*]$ изменения $\sigma[1, \sigma_*]$ разбивается на N промежутков, здесь $\sigma_* = \rho_*^{(\gamma-1)/2}$. Для каждого значения $\sigma_i = 1 + i\Delta\sigma$, $\Delta\sigma = \frac{\sigma_* - 1}{N}$, $i = 0, 1, \dots, N$, используя формулу обобщённой ЦВ Римана [3] при $t = t_*$, получаем значения скорости u_i .

В. Из точки (r_*, t_*) выпускаются N штук C^+ -характеристик.

С. Задаём число шагов по времени M , таким образом определяется шаг по времени $\Delta t = t_*/M$.

Д. Делается шаг Δt в обратном направлении изменения времени вдоль характеристики C_0^+ (обозначена на рис. 3 как C_{01}^+ , у которой вторая цифра в индексе обозначает принадлежность характеристики к области 1). Из точки на этой характеристике, соответствующей моменту времени $(t_* - i \cdot \Delta t)$ выпускается C_i^- -характеристика до пересечения с линией C_1^+ . Здесь и далее $i = 1, \dots, M$ — номер шага по времени.

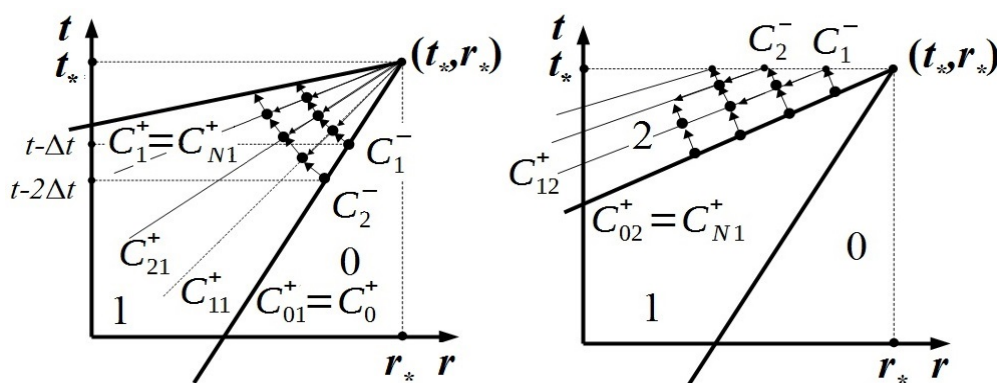


Рис. 3. Расчёт нижнего (слева) и верхнего (справа) треугольников

Е. Используя разностные аналоги уравнений (4) и значения газодинамических параметров в начальных точках соответствующих характеристик, получаем газодинамические параметры в точке пересечения.

Ф. Повторяем процедуру, выпуская C_i^- -характеристику из полученной точки пересечения (см. рис. 3). После того, как найдена точка пересечения с C_N^+ -характеристикой, переходим к пункту С, делая новый шаг по времени. Если достигнуто значение $t = 0$, значит решение в области 1 построено.

Расчёт сетки в верхнем треугольнике (область 2 на рис. 2) также ведётся в обратном направлении изменения времени. Схема расчёта полностью аналогична нижнему треугольнику, с той лишь разницей, что характеристики $C_1^+, C_2^+, \dots$ выпускаются с прямой $t = t_*$, на которой известно значение $\sigma = \sigma_*$.

2.2. Алгоритм построения траектории сжимающего поршня

Построение траектории движения сжимающего поршня в виде $r = r_p(t)$, как и построение характеристической сетки, происходит в обратном направлении изменения времени, т. е. при $t \leq t_*$.

А. Траектория $r = r_p(t)$ движения сжимающего поршня строится исходящей из точки $(t = t_*, r = r_* - d_*)$, где d_* определяет ширину уже сжатого слоя газа. Из этой точки выпускаем прямую, определяемую разностным аналогом задачи Коши (9).

В. Находим точку пересечения этой прямой с одной из линий характеристической сетки, построенной в соответствии с вышеописанными алгоритмами. В найденной точке пересечения скорость газа u находится линейной интерполяцией по значениям в ближайших узлах. Пусть найденная точка пересечения

имеет координаты $(t = \tilde{t}, r = \tilde{r})$. Выпускаем из неё прямую, определяемую разностным аналогом задачи Коши

$$\begin{cases} \frac{dr}{dt} = u|_{t=\tilde{t}, r=\tilde{r}} \\ r|_{t=\tilde{t}} = \tilde{r} \end{cases}$$

до пересечения этой прямой со следующей линией построенной характеристической сетки.

С. Продолжаем вычисления по этому алгоритму до тех пор, пока либо траектория поршня не пересечёт характеристику C_0^+ нижнего треугольника, либо не выйдет за пределы области, в которой существует решение.

Д. В последнем случае следует уменьшить ширину сжатого слоя d_* и вернуться к началу алгоритма.

Завершая вторую часть статьи, необходимо сделать некоторые выводы:

1. Задача в обратном направлении изменения времени с математической точки зрения поставлена правильно, в том числе существует единственное решение для двух ХЗК и существует единственная траектория сжимающего поршня, дающая в момент времени $t = t_*$ требуемое распределение плотности для нужной массы газа.
2. Закон движения сжимающего поршня численно восстановлен, т.е. возможно сформулировать конкретные рекомендации к физическому эксперименту.
3. Требуется решение другой начально-краевой задачи в прямом направлении изменения времени при заданном законе движения сжимающего поршня. Необходимость дальнейшего исследования обусловлена физической сутью изучаемой проблемы. Требуется ответить на вопрос: переводит ли, при возрастании времени, сжимающий по полученному закону поршень исходную массу газа в нужное физическое состояние, при котором достигается необходимое значение плотности.

3. Конечно-разностный метод «Ромб» для численного решения одномерных уравнений газовой динамики

Рассмотрим одномерную систему уравнений газовой динамики в лагранжевых переменных.

$$\begin{aligned} \frac{\partial v}{\partial t} &= \frac{\partial u}{\partial m}, \\ \frac{\partial u}{\partial t} &= -\frac{\partial p}{\partial m}, \\ \frac{\partial E}{\partial t} + v \frac{\partial(pu)}{\partial x} &= 0, \end{aligned} \tag{10}$$

где $t, m, \rho, v = \frac{1}{\rho}, \varepsilon, p$ — соответственно: время, массовая лагранжева координата, плотность, удельный объём, удельная внутренняя энергия, давление, $E = \varepsilon + \frac{u^2}{2}$ — удельная полная энергия.

Система замыкается уравнением состояния идеального газа $p = (\gamma - 1)\rho c_v T$, $\varepsilon = c_v T$, где c_v, T — теплоёмкость при постоянном объёме и температура газа соответственно.

К системе (10) присоединяются граничные и начальные условия в следующем виде

$$\begin{aligned} u|_{x=x_l}(t) &= u_p(t), \\ u|_{x=x_r}(t) &= 0, \\ u(x, 0) &= 0, \end{aligned} \quad (11)$$

где x_l, x_r — левая (left) и правая (right) границы отрезка, на котором решается система (10). В момент времени $t = 0$ $x_l = r_p(0), x_r = r_*$. $u_p(t)$ — скорость сжимающего поршня, полученная в соответствии с алгоритмом из пункта 2.2.

Применив для (10) неявную аппроксимацию и введя шаг по времени τ , получим:

$$\begin{aligned} \frac{v^{n+1} - v^n}{\tau} &= \left(\frac{\partial u}{\partial m} \right)^{n+1}, \\ \frac{u^{n+1} - u^n}{\tau} &= - \left(\frac{\partial p}{\partial m} \right)^{n+1}, \\ \frac{E^{n+1} - E^n}{\tau} + v^{n+1} \left(\frac{\partial(pu)}{\partial x} \right)^{n+1} &= 0. \end{aligned} \quad (12)$$

Систему (12) будем решать, разбивая исходный отрезок $[x_l, x_r]$ на N ячеек (пронумеруем узлы полученной сетки $x_0 = x_l, x_1, x_2, \dots, x_N = x_r$). Граничные условия на левой и правой границах запишем так:

$$\begin{aligned} \alpha_0 p_0^n + \beta_0 u_0^n &= r_p(t^n), \\ \alpha_N p_N^n + \beta_N u_N^n &= \varphi_N. \end{aligned} \quad (13)$$

Система (12) нелинейная, для её разрешения линеаризуем давление и удельную внутреннюю энергию по Ньютону, введём итерационный процесс. После введения через соотношения связи новых неизвестных (давления P и скорости U , определяемых в узлах пространственной сетки) и ряда преобразований получим СЛАУ относительно P, U с четырёхдиагональной матрицей, которая решается методом, аналогичным методу потоковой прогонки [20].

Затем определяются новые координаты $x^{\mu+1}$, плотность $\rho^{\mu+1}$, ячейечные скорости $u^{\mu+1}$, удельная внутренняя энергия $\varepsilon^{\mu+1}$, из УРС восстанавливаем давление в ячейке $p^{\mu+1}$ на новой итерации $\mu + 1$. Подробный вывод формул и обоснование метода изложены в [18].

Определив все величины на текущей итерации по нелинейности p , проверяем условие сходимости итераций

$$|p_{i+1/2}^{\mu+1} - p_{i+1/2}^{\mu}| \leq \epsilon_{con}(1 + p_{i+1/2}^{\mu}). \quad (14)$$

Если условие (14) выполнено во всех ячейках, то итерации заканчиваются. Далее переходим на новый временной шаг.

$$u(t^0) = 0; v(t^0) = 1; E(t^0) = 1/(\gamma(\gamma - 1)). \quad (15)$$

4. Результаты расчётов

В настоящей статье было описано несколько различных численных алгоритмов. Сначала будут приведены результаты расчётов в обратном направлении изменения времени. Эти результаты приведены для демонстрации работоспособности и адекватности алгоритмов. Затем будут приведены результаты расчётов в направлении возрастания времени методом «Ромб».

4.1. Расчёты в обратном направлении изменения времени

Одним из поводов для этой работы стала возросшая с момента публикации [11] вычислительная мощность компьютеров. Изложенные выше алгоритмы были реализованы в виде программы для РС на языке C++. Далее приводятся в том числе и результаты на более подробных сетках, чем использованы в работе [11].

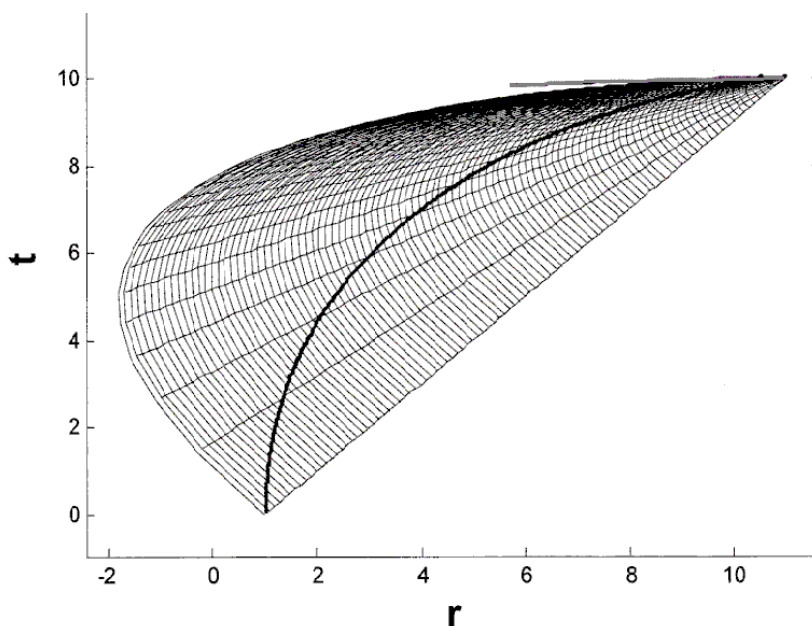


Рис. 4. Пример характеристической сетки

Для проверки работоспособности программ сначала были посчитаны задачи сильного сжатия для случая плоской симметрии, когда известно точное решение [5, 6]. Рассчитанные поля совпали с этим точным решением с точностью, близкой к машинному нулю. На рис. 4 приведён пример рассчитанной характеристической сетки и траектории поршня. По оси абсцисс отложена координата r траектории частицы газа, полученная в процессе решения, по оси ординат — время t . Выделенная прямая в верхней части рисунка — характеристика C_0^+ верхнего треугольника, которая пересекает траекторию поршня в момент времени $(t_* - 0.0001)$.

В следующей таблице приводятся результаты сравнения четырёх численно построенных решений с точным решением для плоской симметрии ($\nu = 0$).

Таблица 1. Результаты тестовых расчётов

γ	m_*	ρ_{max}	$\ c_{ex} - c_{num}\ _C$	$\ u_{ex} - u_{num}\ _C$
1.4	10	10^4	$1.66355818 \cdot 10^{-12}$	$8.37019342 \cdot 10^{-12}$
1.4	100	10^5	$2.25801919 \cdot 10^{-11}$	$1.12819975 \cdot 10^{-10}$
5/3	10	10^4	$4.13891143 \cdot 10^{-13}$	$1.24167343 \cdot 10^{-12}$
5/3	100	10^5	$1.81159976 \cdot 10^{-10}$	$5.43906253 \cdot 10^{-10}$

Варьировались следующие параметры: γ — показатель политропы идеального газа, m_*, ρ_{max} — масса сжимаемого слоя и максимальная плотность слоя газа. Были посчитаны и приведены в таблице нормы разности численно-полученных и точных значений скорости звука и скорости газа. Используются обозначения: $c_{ex}, c_{num}, u_{ex}, u_{num}$ — точное (exact) и численное (numerical) значения скорости звука и скорости газа, $\|\vec{F}\|_C = \max_i(|F_i|)$ — норма.

Построенная численно траектория сжимающего поршня также удовлетворительно согласуется с точным решением [3]. Поточечная норма разности убывает обратно пропорционально возрастанию числа точек в расчёте. В таблице 2 приведены нормы ошибки при сгущении сетки для случая $\gamma = 1.4, d_* = 10, \rho_{max} = 10^4$.

Таблица 2. Результаты тестовых расчётов траектории сжимающего поршня

$N \times M$	250x250	500x500	1000x1000	2000x2000
$\ r_{ex} - r_{num}\ _C$	$4.41842 \cdot 10^{-3}$	$1.09307 \cdot 10^{-3}$	$2.71721 \cdot 10^{-4}$	$6.78031 \cdot 10^{-5}$

Основываясь на данных из таблиц 1 и 2, можно сделать вывод об адекватном описании предложенным алгоритмом решения задачи о сильном сжатии одномерного газового слоя в случае плоской симметрии.

Далее приводятся результаты численных экспериментов в случае $\nu = 1$ и $\nu = 2$, для которых не существует точного решения в аналитическом виде.

В качестве критерия точности полученного решения задачи мы, вслед за [3, 10], выбрали следующий: масса несжатого слоя шириной d_0 равна массе

сжатого слоя шириной d_* , т. е. проверялось равенство масс в начальный и конечный моменты сжатия газа. Кроме того, нами была посчитана работа поршня по сжатию газа и проведено сравнение с изменением полной энергии в сжимаемой системе.

Таблица 3. Результаты численных экспериментов

ν	γ	t_*	m_*	ρ_{max}	NxM	Δm , %
0	1.4	10.01166	10	10^4	500x500	0.117
1	1.4	1.04242	10	10^4	500x500	0.131
2	1.4	0.44934	10	10^4	500x500	0.132
0	5/3	10.04016	10	10^4	500x500	0.402
1	5/3	1.03799	10	10^4	500x500	0.502
2	5/3	0.45187	10	10^4	500x500	0.516
0	1.4	100.0202	100	10^5	2000x2000	0.02
1	1.4	4.4574	100	10^5	1000x1000	0.103
2	1.4	1.5411	100	10^5	1000x1000	0.092
0	5/3	100.1122	100	10^5	2000x2000	0.112
1	5/3	4.5588	100	10^5	1000x1000	0.844
2	5/3	1.5656	100	10^5	1000x1000	0.735

Такая проверка выполнялась и в промежуточные моменты сжатия. В приведённых в таблице 3 данных различие масс Δm сжатого и несжатого газа колеблется от 0.02% до 0.85%, относительная погрешность сохранения полной энергии не превышает сотых долей процента. Точность можно повышать, уменьшая Δt и увеличивая N и M .

По представленным результатам можно сделать вывод, что предложенный алгоритм решения задачи сильного сжатия строит решение и адекватно описывает процесс сжатия с использованием найденной траектории сжимающего поршня. Времена счёта на РС не превышают нескольких минут для задач с десятками тысяч расчётных точек.

4.2. Результаты расчётов в прямом направлении изменения времени методом «Ромб»

Для реализации основной цели исследования необходимо использовать построенную траекторию движения сжимающего поршня как граничное условие при расчётах в прямую. Для этого был запрограммирован надёжный безусловно-устойчивый конечно-разностный метод «Ромб». Ниже излагаются первые результаты представляемого подхода. Для проверки работоспособности избранного метода при расчётах задачи сильного сжатия была выбрана одна из задач в плоскосимметричной постановке: $\gamma = 1.4$, $d_* = 10$, $\rho_{max} = 10^4$.

Сначала задача решается в обратном направлении изменения времени. Затем из траектории сжимающего поршня выбираются левая $x_l = r_p(t_0)$ и правая $x_r = r_p(t_*)$ границы. На границах ставятся условия: $u_l = u_p(t)$, $u_r = 0$. Начальные условия $u_0 = 0$, $v_0 = \frac{1}{\rho_0} = 1$, $\varepsilon_0 = \frac{1}{\gamma(\gamma - 1)}$. Далее решается начально-краевая задача, при этом на каждом временном шаге левое граничное условие интерполируется, т.к. значение скорости и координаты сжимающего поршня известны только в фиксированные моменты времени.

Таблица 4. Погрешности метода «Ромб» при расчёте плоскосимметричной задачи

t_{end}	$\ u_{ex} - u_{num}\ _C$	$\ \sigma_{ex} - \sigma_{num}\ _C$	$\frac{\ \sigma_{ex} - \sigma_{num}\ _C}{\ \sigma_{num}\ _C} \cdot 100\%$
$0.5 \cdot t_{ex}$	0.0018	$3.6304 \cdot 10^{-4}$	0.031
$0.8 \cdot t_{ex}$	0.0043	$8.8103 \cdot 10^{-4}$	0.0673
$0.9 \cdot t_{ex}$	0.0088	0.0018	0.12
$0.99 \cdot t_{ex}$	0.0835	0.0170	0.78

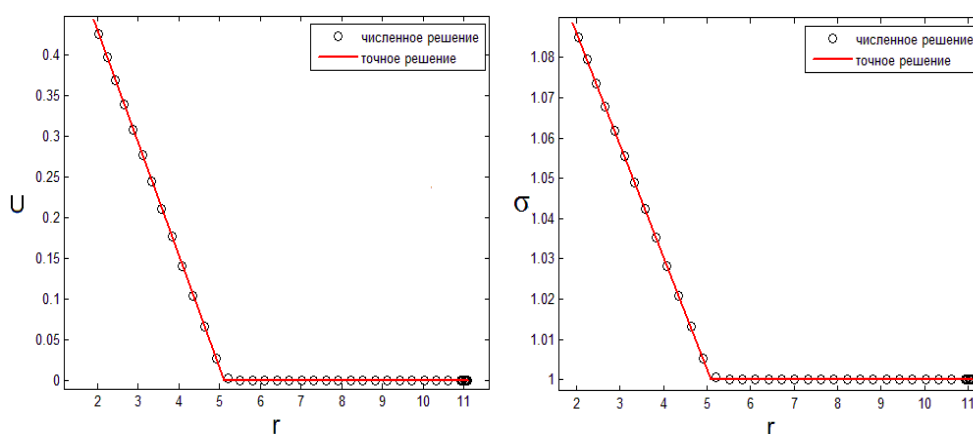


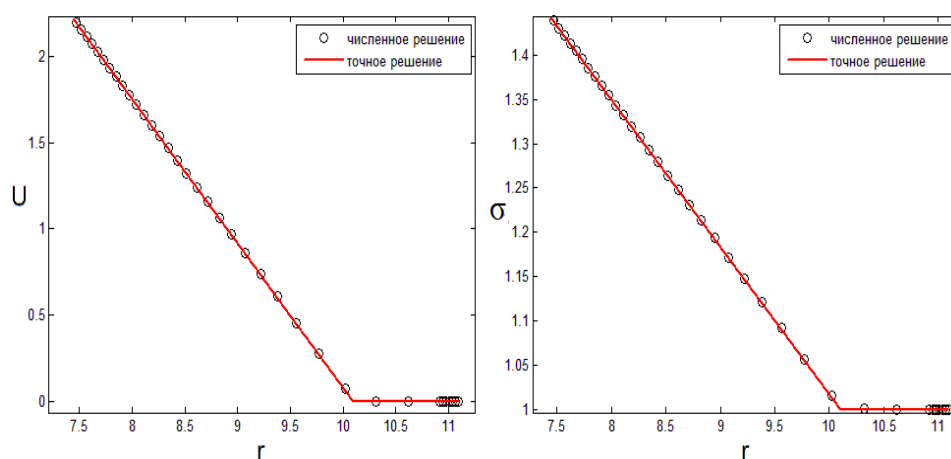
Рис. 5. Скорость и скорость звука на момент времени $t = 0.6t_*$

На рисунках 5–6 представлены результаты расчёта вышеупомянутой плоскосимметричной задачи на различные моменты времени.

В таблице 4 представлены погрешности расчётов при варьировании момента окончания счёта t_{end} . Используются обозначения: u_{ex}, σ_{ex} — точные значения скорости и скорости звука (ЦВ Римана), u_{num}, σ_{num} — значения скорости и скорости звука, полученные численно.

Из рисунков 5–6 и таблицы 4 можно сделать вывод, что метод «Ромб» хорошо описывает распространение центрированной волны сжатия по покоящемуся газу до момента времени $t_{end} = 0.999t_*$

Из таблицы 5 видно, что в вышеприведённые моменты времени сжатия существенной массы газа не происходит. Требуется считать дальше. На следующем рисунке представлены результаты расчётов до более поздних момен-

Рис. 6. Скорость и скорость звука на момент времени $t = 0.9t_*$ Таблица 5. Время до сжатия части массы до требуемой плотности в долях от t_*

ν	Время до сжатия части массы до требуемой плотности, доля от t_*				
	$0.1 \cdot m_*$	$0.3 \cdot m_*$	$0.5 \cdot m_*$	$0.7 \cdot m_*$	$0.9 \cdot m_*$
0	0.999980525	0.9999889	0.9999921	0.9999953	0.9999984257
1	0.999983714	0.9999918762	0.9999918762	0.999996381	0.9999988571
2	0.9999899	0.9999921	0.9999944	0.9999966	0.9999989

тов времени. Из рисунка 7 видно, что на левой границе отрезка формируется область газа, сжатого до требуемой плотности. Точное решение — сплошная линия, численное — пунктир.

Стоит отметить, что в районе состыковки ЦВ с однородным покоем значительным становится «размазывание» волны, что сказывается на точности результатов. Более того, численное возмущение, приходящее на правую границу, вынуждает отодвигать её, т.к. иначе погрешность расчёта в области правой границы делает решение бессмысленным. Следовательно, необходимо каким-то образом модифицировать алгоритм для более аккуратного описания течения в моменты времени, близкие к t_* . Планируется несколько путей модификации: либо разбиение исходного пространственного отрезка на две счётные области, либо исходное неравномерное разбиение отрезка интегрирования для более детального описания интересующей области.

Выводы

Представленные результаты позволяют сделать некоторые выводы.

1. Мы получили закон сжатия исходной массы газа до заданной плотности, решая задачу в обратном направлении изменения времени. Показана работоспособность предложенного алгоритма расчёта одномерных задач

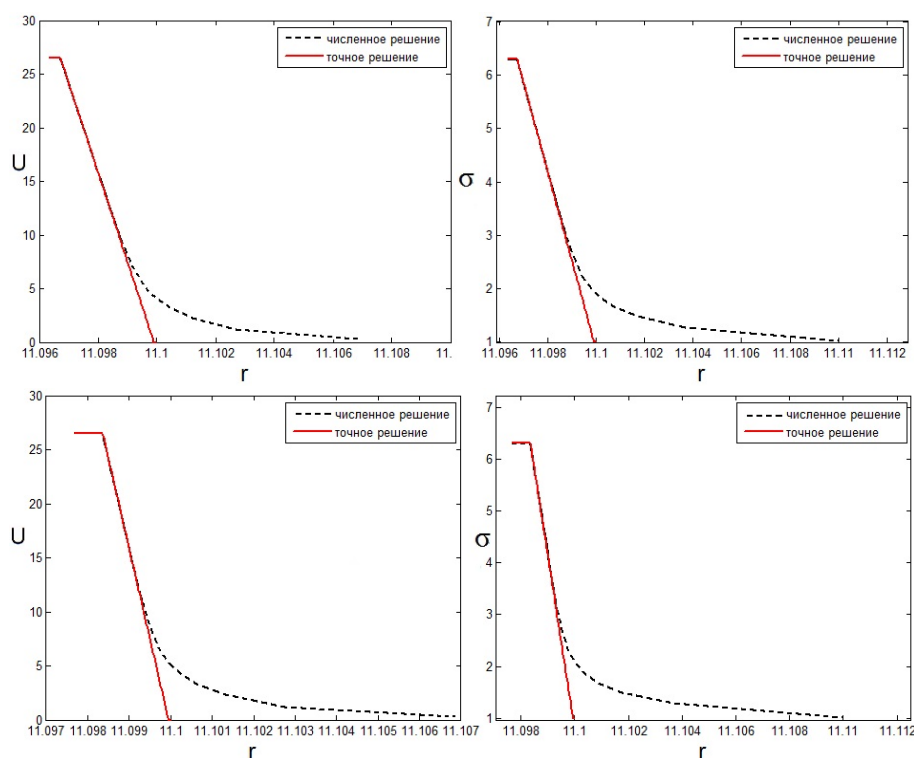


Рис. 7. Скорость и скорость звука на при $t = 0.99999t_*$ (вверху) $t = 0.999995t_*$ (внизу)

сильного сжатия. Точность и скорость расчётов в обратном направлении изменения времени позволяет восстанавливать искомую траекторию сжимающего поршня за 1-2 минуты. Таким образом, рекомендации по проведению физического эксперимента могут быть сформулированы. Однако решение задачи сильного сжатия в обратном направлении изменения времени с явным раскрытием особенности в точке (r_*, t_*) достаточно сложно и не является общепринятым.

2. Реализован и оттестирован безусловно устойчивый конечно-разностный метод «Ромб» для решения одномерной СУГД в лагранжевых переменных.
3. Проведены расчёты плоскосимметричной задачи сильного сжатия газа при возрастании времени методом «Ромб». Результаты показывают, что численный алгоритм с хорошей точностью восстанавливает решение вплоть до момента времени $0.999 \cdot t_*$. Дальнейшие расчёты демонстрируют, что при приближении к особенности в решении, когда производные в точке r_* начинают быстро расти, точность расчёта ухудшается, однако, в решении реализуется требуемое физическое состояние исходной массы газа.

Автор выражает своему научному руководителю профессору С.П. Баутину признательность за внимание, помощь и поддержку.

ЛИТЕРАТУРА

1. Забабахин Е.И., Забабахин И.Е. Явления неограниченной кумуляции. М. : Наука, 1988. 173 с.
2. Накколс Дж.Г. Осуществимость инерциально-термоядерного синтеза // Успехи физ. наук. 1984. Т. 143, № 3. С. 467–482.
3. Баутин С.П. Математическое моделирование сильного сжатия газа. Новосибирск : Наука, 2007. 308 с.
4. Риман Б. О распаде плоских волн конечной амплитуды // Соч. М.-Л. : ОГИЗ, 1948. С. 376–395.
5. Станюкович К.П. Неустановившиеся движения сплошной среды. М. : Гос. изд-во техн.-теор. лит-ры, 1955. 804 с.
6. Мизес Р. Математическая теория течений сжимаемой жидкости. М. : Изд. иностранной литературы, 1961. 588 с.
7. Баутин С.П. Математическая теория безударного сильного сжатия идеального газа. Новосибирск : Наука, 1997. 160 с.
8. Баутин С.П. Асимптотические законы безударного сильного сжатия квазиодномерных слоев газа // Прикл. матем. и механика. 1999. Т. 63, Вып. 3. С. 415–423.
9. Баутин С.П. Характеристическая задача Коши и её приложения в газовой динамике. Новосибирск : Наука, 2009. 367 с.
10. Баутин С.П., Николаев Ю.В. Об одном методе расчёта безударного сильного сжатия одномерных слоёв газа // Вычислительные технологии. 2000. Т. 5, № 4. С. 3–12.
11. Николаев Ю.В. О численном решении задачи безударного сильного сжатия одномерных слоёв газа // Вычислительные технологии. 2001. Т. 6, № 2. С. 104–108.
12. Анучин М.Г. Влияние теплопроводности на неограниченное безударное сжатие плоского газового слоя // ПМТФ. 1998. Т. 39, № 4. С. 25–32.
13. Рощупкин А.В. Исследование некоторых характеристических задач Коши, возникающих при решении неоднородных задач безударного сильного сжатия газа // Вычислительные технологии. 2002. Т. 7 (совм. вып.), Ч. 4. С. 96–103.
14. Артемьев А.Ю., Делов В.И., Дмитриева Л.В. и др. Численное моделирование безударного неограниченного сжатия газа в переменных Лагранжа по методике Д. // ВАНТ. Сер. Мат. моделирование физ. процессов. 1995. Вып. 4. С. 42–47.
15. Сараев В.А. Решение задач о неограниченном безударном сжатии газа в областях специальной формы по программам комплекса МИМОЗА // ВАНТ. Сер. Мат. моделирование физ. процессов. 1998. Вып. 4. С. 3–16.
16. Сучков В.А. Истечение газа в вакуум на косой стенке // Прикл. матем. и механика. 1963. Т. 27, Вып. 4. С. 739–740.
17. Сидоров А. Ф. Некоторые оценки степени куммуляции энергии при плоском и пространственном сжатии газа // Доклады АН. 1991. Т. 318, № 3. С. 548–552.
18. Гаджиев А. Д., Писарев В. Н. Неявный конечно-разностный метод «Ромб» для численного решения уравнений газовой динамики с теплопроводностью. // Ж. выч. матем. и матем. физ. 1979. Т. 19. N. 5. С. 1288–1303.
19. Рождественский Б. Л., Яненко Н. Н. Системы квазилинейных уравнений и их приложения к газовой динамике. М.: Наука, 1968. 529 с.
20. Дегтярев Л. М., Фаворский А. П. Поточный вариант метода прогонки. // Ж. выч. матем. и матем. физ. 1968. Т. 8. N. 3. С. 679–684.

ONE-DIMENSIONAL MATH MODELING OF IDEAL GAS STRONG COMPRESSION IN R. MISES CONFIGURATION

N.S. Novakovskiy

Postgraduate Student, e-mail: n.s.novakovskiy@yandex.ru

Ural State University of Railway Transport (USURT)

Abstract. The article describes results of one-dimensional gas layers strong compression problem's math modeling. Mathematical formulation of the problem includes two characteristic Cauchy problems (cCp1 and cCp2) and one Cauchy problem for compression piston's trajectory reconstruction with time decrease. The article provided numerical algorithm based on typical characteristic method for solving these problems. Algorithm for compression piston's trajectory reconstruction is also described. Finite-difference method "ROMB" for solving system of gas dynamics equations while time increasing is described too. The main results of numerical simulations are shown in graphs and tables.

Keywords: gas strong compression, the characteristic Cauchy problem, centered wave of compression, characteristic's method, finite difference method "ROMB".

ДИФФЕРЕНЦИАЛЬНЫЕ ИГРЫ В ЭКОЛОГИИ ЧЕЛОВЕКА И В СОЦИОЛОГИИ

А.К. Гуц

профессор, д.ф.-м.н., заведующий кафедрой кибернетики ОмГУ, e-mail: aguts@mail.ru

Л.А. Володченкова

доцент, к.б.н., e-mail: volodchenkova2007@yandex.ru

Омский государственный университет им. Ф.М. Достоевского

Аннотация. Показано, что динамика уровня здоровья человека в экологии или уровень доверия людей к власти можно описывать как дифференциальную игру и, следовательно, находить оптимальные равновесные ситуации (оптимальные стратегии).

Ключевые слова: экология человека, динамика уровня здоровья, социология, доверие к власти, дифференциальная игра, оптимальные стратегии.

Введение

В экологии и в социологии мы легко находим примеры ситуаций, в которых наблюдаются две противоборствующие, конфликтующие стороны. Чаще всего ни одна из сторон не способна обеспечить себе «полную победу». В жизни всегда приходится искать компромиссные решения, результатом которых являются в общем-то удовлетворительные или оптимальные для обеих сторон ситуации. Для поиска таких оптимальных ситуаций создана математическая теория игр, в которой противоборствующие стороны называются игроками, а под оптимальной ситуацией понимается надлежащий выбор оптимальных или равновесных стратегий, которых придерживаются игроки, управляя тем самым ходом игры.

В данной статье показано, как можно применить теорию дифференциальных игр к нахождению и удержанию оптимальных ситуаций, называемых в теории дифференциальных игр оптимальными управлениями или равновесиями.

1. Описание модели здоровья человека

В рамках медицинской модели здоровья степень здоровья человека может быть охарактеризована достаточно большим числом количественных показателей, которые получают при проведении различных анализов (кровяное давление, температура тела, количество эритроцитов, сахар в крови и т. д.). К этим показателям следует добавить различные показатели, используемые другими моделями здоровья человека.

Пусть величины x_j , $j = 1, 2, \dots, N$ — совокупность всевозможных показателей здоровья человека.

Введём *интегральный показатель здоровья человека*, имеющий вид

$$x = \sum_{j=1}^N w_j x_j,$$

где w_j — вес показателя x_j , т. е. его вклад (доля) в интегральный показатель. Значения показателя x в момент времени t обозначаем как $x(t)$. Это число принимается нами как степень здоровья человека.

Показатель имеет нижнюю границу — число Z_0 . Человек считается здоровым в момент времени t , если сумма его показателей $x(t) \geq Z_0$, и болеющим, если $x(t) < Z_0$.

Очевидно, что такой подход является крайне упрощённым, но любая модель здорового человека есть определённое упрощение, которое может быть со временем усложнено.

Здоровье людей в конкретном регионе во многом определяется *действием долговременного вредоносного фактора риска* — $k_{\text{ВФР}}$, который является неустранимым фактором. Это радиоактивный фон местности, некачественная вода в колодцах, реке, озере и др. Данный фактор мы не рассматриваем как фактор управления.

Внешние управляющие факторы в нашей задаче, оказывающие влияние на здоровье человека, это:

1) v — *неблагополучная медико-санитарная ситуация* (временный, переменный фактор, который может быть устранён: задымлённость при лесных пожарах, ядовитые сбросы в реки и др.);

2) u — *принятие мер по преодолению неблагоприятной медико-санитарной ситуации* (лечение, профилактика) u .

В [1, 2] было выведено дифференциальное уравнение, описывающее динамику интегрального показателя здоровья человека $x(t)$:

$$\frac{dx}{dt} = -\frac{\partial}{\partial x} V(x, u, v, k_{\text{ВФР}}), \quad (1)$$

где

$$V(x, u, v, k_{\text{ВФР}}) = \frac{k_0}{5} x^5 + u x^3 + v x^2 + k_{\text{ВФР}} x. \quad (2)$$

Отметим, что хорошее здоровье людей характеризуется неравенством $x > Z_0$, ухудшение — неравенством $x < Z_0$; действие долговременного вредоносного фактора риска — неравенством $k_{\text{ВФР}} < 0$, наличие неблагоприятной медико-санитарной ситуации в регионе — неравенством $v < 0$, принятие мер по преодолению неблагоприятной медико-санитарной ситуации (лечение) — неравенством $u > 0$.

Функция V , заданная выражением (2), описывает катастрофу «ласточкин хвост» [1, 2].

Равновесные состояния

$$\frac{\partial}{\partial x} V(x, u, v, k_{\text{ВФР}}) = 0 \quad (3)$$

данной динамической системы были изучены в [1, 2].

2. Экология человека как дифференциальная игра

Поскольку фактор $k_{\text{ВФР}}$ мы не рассматриваем как управляющий, то перепишем уравнение (1) в виде:

$$\begin{cases} \frac{dx}{dt}(t) = -\frac{\partial}{\partial x} W(x(t), u(t), v(t)), \\ x(t_0) = x_0, \end{cases} \quad (4)$$

где

$$W(x, u, v) = V(x, u, v, k_{\text{ВФР}}), \quad k_{\text{ВФР}} < 0.$$

Нас интересует, какая пара управлений $u(t), v(t)$ является в некотором смысле оптимальной? Фактически это означает ситуацию, соответствующую реальности: трудно предотвратить вредоносное действие фактора $v(t) < 0$, как и трудно добиться желательного уровня мер, предотвращающих неблагоприятные медико-санитарные ситуации, и обеспечить необходимое лечение пострадавших $u(t) > 0$.

Реальная жизнь демонстрирует, что даже если руководитель предприятия, производящего периодические вредные выбросы в атмосферу и в водоёмы, вполне понимает, как это плохо отражается на здоровье населения, тем не менее отсутствие средств на очистительные сооружения, на модернизацию оборудования вынуждает его санкционировать вредоносные выбросы. Подобным же образом экологическим учреждениям и экологическим организациям часто трудно преодолеть бюрократические препятствия на пути внедрения нужных природозащитных мероприятий, которые связаны как с отсутствием нужных средств, так с подкупом тех, от кого зависит обеспечение таких мероприятий.

На языке математической теории игр это означает, что у нас есть два игрока 1 и 2, первый из которых борется за здоровье людей, а второй создаёт вредоносную окружающую среду.

Для каждого игрока надо выбрать подходящие к региональной ситуации платёжные функции, имеющие вид

$$J_i(x, u, v) = \int_{t_0}^T F_i(t, x(t), u(t), v(t)) dt + h_i(x(T)), \quad (i = 1, 2),$$

и критерий оптимальности, определяющий выбор управлений $\tilde{u}(t) \in U_1, \tilde{v}(t) \in U_2$, адекватных сложившейся ситуации.

2.1. Примеры критериев оптимальности управления

Например, таковым является следующая форма *принципа минимакса*: $J_1 = J_2$ и ищем управления $\tilde{u}(t, \tilde{v}(t)) \in U_1, \tilde{v}(t) \in U_2$ такие, что

$$J(t_0, x_0, \tilde{u}(t, \tilde{v}(t)), \tilde{v}(t)) = \inf_{v(t)} \sup_{u(t)} J(t_0, x_0, u(t), v(t)) \quad (5)$$

в предположении, что игрок 1 знает, какое управление $\tilde{v}(t)$ выбрал игрок 2 [3, с. 46]. Пара $\tilde{u}(t, \tilde{v}(t)), \tilde{v}(t)$ называется *оптимальной парой стратегий*.

Фактически это означает, что экологи хорошо ознакомлены с действиями чиновником и руководителей загрязняющих среду предприятий и пытаются всячески добиться высокого значения интегрального показателя здоровья человека. Но действия чиновников, директоров предприятий, их владельцев, преступников и прочее объективно ведут к снижению усилий экологов. Особенно это видно, если взять такую платёжную функцию

$$J(t_0, u, v) = x(T).$$

Перед игрой берётся некоторая точка $x_1 > Z_0$, называемая терминальной частью границы пространства $[x_0, x_1]$. Когда $x(t)$ достигает её, т. е. $x(T) = x_1$, игра оканчивается [4, с. 49].

В случае, когда оба игрока не имеют информации об используемых противником стратегиях, то в качестве критерия оптимальности можно взять принцип *седловой точки* [3, с. 49]: ищутся управления $\tilde{u}(t), \tilde{v}(t)$ такие, что

$$J(t_0, x_0, \tilde{u}(t), \tilde{v}(t)) = \inf_{v(t)} \sup_{u(t)} J(t_0, x_0, u(t), v(t)) = \sup_{u(t)} \inf_{v(t)} J(t_0, x_0, u(t), v(t)). \quad (6)$$

Пара управлений $(\tilde{u}(t), \tilde{v}(t))$ — это *оптимальные управления*.

В книге [3] излагаются алгоритмы отыскания оптимальных управлений для задач (5) и (6). При реализации этих алгоритмов приходится решать очень сложные задачи, которые далеко не всегда приводят к успеху.

Пара управлений $(\bar{u}_1(-), \bar{u}_2(-)) \in U_1 \times U_2$ представляет *равновесие Нэша*, если

$$\forall u_1 \in U_1 [J_1((t, x, \bar{u}_1(-), \bar{u}_2(-)) \leq J_1((t, x, u_1(-), \bar{u}_2(-))),$$

$$\forall u_2 \in U_2 [J_2((t, x, \bar{u}_1(-), \bar{u}_2(-)) \leq J_2((t, x, \bar{u}_1(-), u_2(-))).$$

Равновесие Нэша — это стратегии игроков, которые стараются учитывать интересы противника и стремятся к компромиссу.

В [5, р. 109] даётся необходимое условие существования равновесия Нэша, сводящееся к аналогу принципа максимума Понтрягина.

Различные достаточные условия существования оптимальных стратегий и равновесий для дифференциальных игр даны в книге [6].

2.2. Существование равновесий Нэша

Если игрок формирует «свое» управляющее воздействие в виде только функции времени $u(t)$ на всю продолжительность игры, то $u(t)$ — это *программное управление* игрока. Ранее мы называли его, используя термин «управление». Однако игрок может выбирать своё управление в зависимости от того, в каком положении x в момент времени t находится система. В таком случае игрок конструирует управляющее воздействие в виде функции $u(t, x)$, зависящей уже от позиции $\{t, x\}$, и для $u(t, x)$ используется термин *позиционное управление* игрока [9]. Часто пишут просто $u(x)$.

Приведём два примера, когда ищутся равновесия Нэша, являющиеся позиционными управлениями.

1. Игра с ненулевой суммой. Для дифференциальной игры N -игроков

$$\begin{aligned}\frac{dx}{dt} &= f(x) + \sum_{j=1}^N g_j(x)u_j, \quad f(0) = 0, \\ x &\in \mathbb{R}, \quad u_j \in \mathbb{R}, \\ J_i(x, u_1, \dots, u_N) &= \int_0^{+\infty} [Q_i(x) + \sum_{j=1}^N R_{ij}(u_j)^2] dt, \quad (i = 1, \dots, N), \\ Q_i &> 0, \quad R_{ii} > 0, \quad R_{ij} \geq 0,\end{aligned}$$

существование равновесий Нэша

$$J_i(u_1^*, u_2^*, u_i^*, \dots, u_N^*) \leq J_i(u_1^*, u_2^*, \dots, u_{i-1}^*, u_i, u_{i+1}^*, \dots, u_N^*), \quad \forall u_i, \quad i \in N, \quad (7)$$

сводится к крайне сложной задаче отыскания положительно определённого решения $V_i(x) > 0$ нелинейного уравнения Гамильтона-Якоби

$$\begin{aligned}(V_i)'_x(x)f(x) + Q_i(x) - \frac{1}{2}(V_i)'_x \sum_{j=1}^N [g_j(x)]^2 (R_{jj})^{-1} (V_j)'_x + \\ + \frac{1}{4} \sum_{j=1}^N R_{ij} [g_j(x)]^2 (R_{jj})^{-1} [(V_j)'_x]^2 = 0,\end{aligned} \quad (8)$$

по которому строится равновесие Нэша [8, Theorem 10.4-2]:

$$u_i^*(x) = u_i(V_i(x)) = -\frac{1}{2} R_{ii} g_i(x) (V_i)'_x, \quad i \in N. \quad (9)$$

Равновесие Нэша в данном случае означает, что если каждый игрок пытается в одностороннем порядке изменить свою стратегию управления, в то время как политика остальных игроков остаётся неизменной, то он имеет худший результат (большой проигрыш).

В нашем случае $N = 2$, и

$$f(x) = k_0 x^4 + k_{\text{ВФР}}, \quad g_1(x) = 3x^2, \quad g_2(x) = 2x,$$

и при $R_{11} = R_{22} = 1, R_{12} = R_{21} = 0$ уравнения Гамильтона-Якоби имеют вид:

$$\begin{aligned} Q_1 + (V_1)'_x f(x) - \frac{1}{4}[g_1(x)]^2[(V_1)'_x]^2 - \frac{1}{2}[g_2(x)]^2(V_1)'_x(V_2)'_x &= 0, \\ Q_2 + (V_2)'_x f(x) - \frac{1}{4}[g_2(x)]^2[(V_2)'_x]^2 - \frac{1}{2}[g_1(x)]^2(V_1)'_x(V_2)'_x &= 0. \end{aligned} \quad (10)$$

Примем, что $k_{\text{ВФР}} = 0$, т.е. в регионе отсутствует долговременный вредоносный фактор риска. Тогда имеем уравнения Гамильтона-Якоби в виде:

$$\begin{aligned} Q_1 + (V_1)'_x k_0 x^4 - \frac{9}{4}x^4[(V_1)'_x]^2 - 2x^2(V_1)'_x(V_2)'_x &= 0, \\ Q_2 + (V_2)'_x k_0 x^4 - x^2[(V_2)'_x]^2 - \frac{9}{2}x^4(V_1)'_x(V_2)'_x &= 0. \end{aligned} \quad (11)$$

Проигрышные функции имеют вид:

$$\begin{aligned} J_1(x, u, v) &= \int_0^{+\infty} [Q_1(x) + u^2] dt, \\ J_2(x, u, v) &= \int_0^{+\infty} [Q_2(x) + v^2] dt. \end{aligned} \quad (12)$$

Нетрудно проверить, что уравнения (11) выполнены, если

$$\begin{aligned} V_1(x) = V_2(x) &= \frac{1}{2}x^2, \\ Q_1(x) &= \frac{9}{4}x^6 + 2x^4 - k_0 x^5 = x^4 \left(\frac{9}{4}x^2 - k_0 x + 2 \right), \\ Q_2(x) &= \frac{9}{4}x^6 + x^4 - k_0 x^5 = x^4 \left(\frac{9}{4}x^2 - k_0 x + 1 \right). \end{aligned} \quad (13)$$

Все эти функции положительно определённые, если $0 < k_0 < 3$. Поэтому по теореме 10.4-2 из [8] имеем равновесие Нэша

$$u^* = -\frac{3}{2}x^3, \quad v_2^* = -x^2, \quad (14)$$

найденное по формулам (4).

2. Игра с нулевой суммой. Равновесия Нэша

$$J(x(0), u^*, v) \leq J(x(0), u^*, v^*) \leq J(x(0), u, v^*), \quad \forall u, v,$$

для игры с нулевой суммой для уравнения

$$\frac{dx}{dt} = f(x) + g(x)u + k(x)v, \quad f(0) = 0,$$

с функцией выигрыша/проигрыша

$$J(x(0), u, v) = \int_0^{+\infty} [h^2(x) + Ru^2 - \gamma v^2] dt,$$

$$h^2(x) \geq 0, \quad R, \gamma > 0,$$

исследуются в [7, 8]. Решение игры будет найдено, если будет найдено положительно определённое решение $V(x) > 0$ нелинейного уравнения Гамильтона-Якоби-Айзекса

$$h^2 + V'_x \cdot f(x) - \frac{1}{4R}(V'_x)^2[g(x)]^2 + \frac{1}{4\gamma^2}(V'_x)^2[k(x)]^2 = 0,$$

$$V(0) = 0,$$

при двух ещё дополнительных условиях [8, Theorem 10.2-2]. Однако сделать это крайне сложно.

Равновесия, являющиеся позиционными управлениями, в таком случае задаются формулами:

$$u^* = u(V(x)) = -\frac{1}{2R}g(x)V'_x,$$

$$v^* = v(V(x)) = \frac{1}{2\gamma^2}k(x)V'_x.$$

В нашем случае при $k_{\text{ВФР}} = 0$

$$h^2 + V'_x \cdot k_0 x^4 - \frac{9}{4R}(V'_x)^2 x^4 + \frac{1}{\gamma^2}(V'_x)^2 x^2 = 0. \quad (15)$$

К сожалению, нам не удалось найти положительно определённого решения $V(x) > 0$ уравнения (15) (для $h(x) \geq 0, h(0) = 0$). Похоже, решения игры с нулевой суммой в форме равновесия не существует. Впрочем, в какой-то мере, так и должно быть, поскольку нулевая сумма говорит нам, что выигрыш экологов в точности есть проигрыш чиновников. Вряд ли так должно быть в правовом обществе. Более естественной в данном случае является игра с ненулевой суммой, а для неё равновесие было найдено.

3. Модель уровня доверия населения к власти

Уравнение (4) можно использовать для описания такого чисто социального явления, как доверие населения к власти¹.

¹Нетрудно убедиться, что вывод этого уравнения в данном случае повторяет вывод уравнения в случае экологии человека (см. [1, 2]).

При этом долговременно действующий вредоносный фактор $k_{\text{ВФР}}$ — это экономическая ситуация в регионе (уровень зарплаты, безработица, дороговизна питания и т.д.). Сам фактор, конечно следует переименовать: неблагоприятная экономическая ситуация — $k_{\text{НЭС}}$.

Вместо управляемого фактора «наличие неблагоприятной медико-санитарной ситуации в регионе» $v < 0$ следует рассматривать *ошибки правящей в регионе элиты*, такие, как произвол полиции, коррупция чиновника, плохие дороги и пр. Обозначение v для данного управления сохраняем.

Наконец, вместо управляемого фактора «принятие мер по преодолению неблагоприятной медико-санитарной ситуации (лечение)» u — вводим другой управляемый фактор, означающий *действия оппозиции как политической, так и различных общественных организаций*. Обозначение u для данного управления сохраняем.

Для описания динамики *уровня доверия населения к власти* $x(t)$ естественно рассмотреть аналог уравнения (4):

$$\frac{dx}{dt} = -\frac{\partial}{\partial x} W(x, u, v), \quad (16)$$

где

$$W(x, u, v) = V(x, u, v, k_{\text{НЭС}}), \quad k_{\text{НЭС}} < 0.$$

Мы можем теперь, опираясь на данное уравнение, пытаться отыскать оптимальные стратегии для следующей задачи минимакса

$$\min_{v(t)} \max_{u(t)} x(T)$$

или максимина

$$\max_{u(t)} \min_{v(t)} x(T).$$

Однако для социологии особое значение имеет выявление компромиссных ситуаций, когда противоборствующие стороны начинают учитывать интересы друг друга. В теории игр компромиссы — это равновесия Нэша.

Результат, полученный в § 2.2, позволяет заявить, что задача определения уровня доверия населения к власти в случае отсутствия неблагоприятной экономической ситуации — $k_{\text{НЭС}} = 0$ и рассматриваемая как дифференциальная игра с ненулевой суммой, допускает равновесия Нэша (2) в форме позиционно-го управления (8) с выигрышными функциями (9), (13).

ЛИТЕРАТУРА

1. Гуц А.К., Володченкова Л.А. Катастрофы типа «ласточкин хвост» в экологии человека // Математические структуры и моделирование. 2009. Вып. 19. С.68–77.
2. Гуц А.К., Володченкова Л.А. Кибернетика катастроф лесных экосистем. Омск : Изд-во КАН, 2012. 220 с.
3. Пацюков В.П. Дифференциальные игры при различном информировании игроков. М. : Советское радио, 1976. 200 с.

4. Айзекс Р. Дифференциальные игры. М. : Мир, 1967. 489 с.
5. Yong J. Differential games: a concise introduction. University of Central Florida, 2015.
6. Вайсборд Э.М., Жуковский В.И. Дифференциальные игры нескольких лиц и их приложения. М. : Советское радио, 1980. 304 с.
7. Vamvoudakis K.G., Lewis F.L. Online solution of nonlinear two-player zero-sum games using synchronous policy iteration // Int. J. Robust and Nonlinear Control. 2012. V. 22. P. 1460–1483.
8. Lewis F.L., Vrabie D.L., Syrmos V.L. Optimal Control. John Wiley & Sons, Inc., 2012.
URL: <http://www.uta.edu/utari/acs/FL%20talks/CDC%20Orlando%202011-%20online%20synch%20PI.pdf>.
9. Тынянский Н.Т., Жуковский В.И. Дифференциальные игры с ненулевой суммой (кооперативный вариант) // Итоги науки и техн. Сер. Мат. анализ. 1979. Т. 17. С. 3–112.

DIFFERENTIAL GAMES IN HUMAN ECOLOGY AND SOCIOLOGY

A.K. Guts

Dr.Sc. (Phys.-Math.), Professor, e-mail: aguts@mail.ru

L.A. Volodchenkova

Ph.D. (Biology), Associate Professor, e-mail: volodchenkova2007@yandex.ru

Dostoevsky Omsk State University

Abstract. It is shown that the dynamics of the human health level in ecology or the level of people's trust to the authorities can be described as a differential game and, therefore, the optimal equilibrium situations (best optimal strategies) can be find.

Keywords: human ecology, dynamics of the health level, sociology, trust to government, differential game, optimal strategies.

ИНТЕЛЛЕКТУАЛЬНАЯ ГЕОИНФОРМАЦИОННАЯ СИСТЕМА АРХЕОЛОГИЧЕСКИХ ОБЪЕКТОВ

С.В. Белим

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

Д.М. Бречка

доцент, к.т.н., e-mail: dbrechkawork@yandex.ru

Т.А. Горбунова

доцент, к.и.н., e-mail: GorbunovaTA@omsu.ru

И.Б. Ларионов

доцент, к.т.н., e-mail: LarionovIB@omsu.ru

И.В. Шмидт

доцент, к.и.н., e-mail: ShmidtIV@omsu.ru

Омский государственный университет им. Ф.М. Достоевского

Аннотация. Приведено описание геоинформационной системы археологических объектов ArGIS. Выделены объекты и их свойства. На основе свойств археологических объектов построено пространство признаков. К археологическим объектам как точкам пространства признаков применены методы кластеризации и методы построения ассоциативных правил. Реализованы модули визуализации результатов обработки данных. По результатам обработки данных происходит также выгрузка данных в текстовом виде. Предложенная система позволяет автоматически выявлять закономерности в расположении и датировке археологических фактов.

Ключевые слова: геоинформационная система, археологические объекты, Data Mining.

Введение

В археологии объекты исследования — археологические памятники и подземный материал — имеют территориальную привязку. Поэтому одним из основных методов исследования является метод картографии. Ещё недавно археологические карты составлялись вручную или с использованием таких графических редакторов как CorelDraw и Photoshop. В настоящее время в археологии все большее распространение получают географические информационные системы (ГИС), которые стали революционной технологией, выходящей за пределы простого составления карт благодаря своему мощному аналитическому инструментарию. Это комплексные системы для сбора, хранения, поиска, анализа и представления пространственных данных любого типа [1]. ГИС позволяют обрабатывать большие массивы данных, применять методы не инвазивного археологического исследования с использованием данных дистанционного зондирования, устанавливать взаимосвязи между археологическими памятниками

и окружающим ландшафтом, прогнозировать обнаружение новых археологических памятников на основе пространственного анализа и многое другое. Кроме того, благодаря ГИС появляется возможность эффективного менеджмента историко-культурного наследия.

Археологические исследования с применением ГИС являются относительно молодым направлением в археологии. Наиболее активно они развиваются в течение последних 20–25 лет, и за это время существенно расширилась тематика исследований в данной области, выросло их количество, сформировалось два основных направления применения ГИС в археологии. Первое может быть обозначено как «управление культурным наследием» («Cultural resource management», CRM), второе — как археология ландшафта («Landscape archaeology»).

Широкое применение теоретические и методические разработки геоинформационного подхода к археологическим исследованиям нашли в региональных археологических проектах, большинство которых сфокусировано на реконструкциях и анализе древних культур и ландшафтов или же территориальном распределении отдельных культурных феноменов, например, образцов древнего искусства [2]. В качестве примеров использования ГИС в археологии можно также отметить такие проекты, как АИС «Археограф», «Археологические памятники Красноярского края» [3], «Археологические памятники Кисловодской котловины» [4, 5], «Палеолит Северной Азии» [6].

Широкое распространение ГИС получили в археологических исследованиях в Германии [7–9], Великобритании [10], Италии [11, 12], Испании [13], Греции [14] и других странах.

Мировыми лидерами в области археологических ГИС сегодня являются ArcGIS от ESRI (США), MapInfo от Pitney Bowes Business Insight (США), GeoMedia от Intergraph (США), Bentley MicroStation от Bentley Systems (США). Среди российских производителей конкуренцию им могут составить продукты компании «Интегро» (Уфа) — ИнГео и Центра Геоинформационных Исследований Института Географии РАН — ГеоГраф (GeoGraph). Большое распространение получают открытые ГИС, такие как ILWIS (Голландия), MapWindow GIS (США), GRASS GIS (США) и, особенно Quantum GIS (международный проект) и SAGA (Германия). Однако большинство открытых ГИС требуют серьёзной доработки для использования в археологических целях.

Однако на сегодняшний день практически отсутствуют работы по интеллектуальному анализу данных, размещённых в ГИС. Тем не менее использование методов анализа данных в качестве поддержки археологических исследований позволит выявить новые закономерности, которые окажут помощь в исторических исследованиях.

1. Описание ArGIS

Объектами геоинформационной системы являются археологические памятники. Каждый памятник характеризуется набором атрибутов: координаты многоугольника, ограничивающего контуры археологического объекта; наимено-

вание археологического объекта; список работ, выполнявшихся на объекте; документы с результатами выполненных работ; классификационные признаки объекта; степень исследованности объекта. В списке работ, выполнявшихся на объекте, указывается год проведения работ, руководитель работ и виды выполненных работ. Классификационные признаки включают в себя эпоху, к которой относится памятник, вид объекта и др.

Геоинформационная система включает в себя как достаточно простой функционал просмотра и редактирования информации, так и возможность выявления закономерностей методами Data Mining. К простому функционалу можно отнести функции поиска объектов по координатам, степени исследованности, наименованию и классификационным признакам. В качестве дополнительной функции поиска возможна фильтрация объектов в выделенной географической области.

Добавление и редактирование объектов возможно в двух режимах. В стационарном режиме выполняется заполнение полей формы. В мобильном режиме происходит автоматическая привязка по координатам. Также в мобильном режиме возможна загрузка фотографий с мобильного устройства.

Модуль обработки карт даёт возможность привязки как распространённых карт (Bingo, Google, Yandex), так и исторических карт. При загрузке исторических реализована возможность сопоставления с другими картами и взаимная корреляция объектов.

2. Построение ассоциативных правил

Для анализа взаимосвязи различных археологических объектов и их свойств используются методы построения ассоциативных правил. Наибольшее распространение метод построения ассоциативных правил получил в экономических науках. Ассоциативные правила позволяют выявлять взаимосвязь между объектами на основе статистических данных. Для построения ассоциативных правил необходимо выделить транзакции. В качестве одной транзакции будем использовать запись об археологическом объекте. Поля в записи об объекте будут играть роль товаров. Причём выбираются только те поля, которые могут быть одинаковыми для любого объекта:

1. Вид объекта.
2. Годы открытия и постановки на учёт.
3. Принадлежность объекта к эпохе.
4. Принадлежность объекта к культуре.
5. Степень разработанности.
6. Виды работ.
7. Результаты раскопок.

Для каждого набора значений F_i определяется уровень поддержки $Supp(F_i)$.

$$Supp(F_i) = \frac{D(F_i)}{N},$$

где $D(F_i)$ — число транзакций, содержащих набор F_i , N — общее количество транзакций. Из полученных наборов строим ассоциативные правила для каждого встречающегося значения признака c_j в наборе F_i :

$$\{c_1, c_2, \dots, c_{j-1}, c_{j+1}, \dots, c_k\} \rightarrow c_j.$$

Для каждого из ассоциативных правил необходимо рассчитать величину, называемую достоверностью:

$$\text{conf}(F_i, c_j) = D(F_i)/D(c_j), \quad (j = 1, \dots, k),$$

где $D(F_i)$ — количество транзакций, содержащих набор F_i , $D(c_j)$ — число транзакций, содержащих значение c_j . Достоверность лежит в интервале от 0 до 1.

Оставляя ассоциативные правила с поддержкой не ниже некоторого значения Supp_{min} , выбираем наиболее часто встречающиеся сочетания признаков. Из оставленных ассоциативных правил отбираем то, которое имеет наибольшую достоверность.

В системе ArgGIS был реализован поиск ассоциативных правил с подмножеством различной длины. Длины наборов пользователь задаёт самостоятельно.

3. Кластеризация данных

Процедура кластеризации позволяет выделить группы объектов, близких по некоторому набору признаков. Для выполнения кластеризации необходимо выделить пространство признаков. Каждому объекту сопоставляется точка в пространстве признаков. Введём следующие координаты объектов:

1. Год выполнения работ (x_1).
2. Вид объекта (x_2).
3. Эпоха, к которой относится объект (x_3).
4. Степень разработанности (x_4).
5. Культура, к которой относится объект (x_5).
6. Географические координаты (x_6, x_7).

Для всех координат в пространстве признаков важным является не абсолютное значение, а степень близости с другими объектами. Год выполнения работ будем отсчитывать от некоторой начальной даты исследований на данной территории. В качестве значения координаты вида объекта будем использовать его номер в полном списке возможных видов объектов. На результатах кластеризации будет сказываться способ упорядочивания видов в полном списке. Чтобы снизить данное влияние, необходимо использовать метод классификации видов объектов, позволяющий проводить однозначное упорядочивание.

Эпохи, к которым относятся объекты, упорядочиваются по возрастанию. В качестве координаты x_3 выбирается номер эпохи, к которой относится объект.

Степень разработанности численно определяется с помощью метода экспертных оценок. Проводится попарное сравнение степени разработанности объектов. После чего строится матрица парных сравнений и определяется удельный вес каждого объекта, который используется как значение координаты x_4 .

Для характеристики культурной составляющей объекта используется тот же подход, что и для эпохи.

Географические координаты (x_6, x_7) задаются как широта и долгота центра масс многоугольника, задающего границы объекта.

Для устранения доминирования одной координаты над остальными необходимо провести масштабирование и привести интервалы значений всех координат к отрезку $[0, 1]$. Меру близости двух точек в пространстве координат будем вычислять как евклидово расстояние между ними.

Для кластеризации объектов используем метод построения минимального остовного дерева. Для этого сопоставим множеству археологических объектов полносвязный взвешенный граф. Вершинами графа служат точки в пространстве признаков. Вес ребра определяется расстоянием между точками. Кластеризация сводится к разбиению графа на подграфы. Для данного графа ищем минимальное остовное дерево, используя алгоритм Прима. Найдём среднее значение и дисперсию длин рёбер минимального остовного дерева. Разобьём дерево на поддеревья. Для этого удалим наиболее длинные ребра. Каждое поддерево будет отдельным кластером. Для выбора удаляемых ребер может быть использован один из двух подходов. В первом случае удаляем ребра, длина которых больше среднего значения более чем в p раз ($p > 1$). Во втором случае удаляем $k\%$ самых длинных рёбер ($0 < k < 100$). Варьирование параметрами p и k позволяет менять размеры кластеров. Возможна кластеризация в любом подпространстве полного пространства признаков. В этом случае не учитывается влияние некоторых параметров.

Визуализация результатов кластеризации состоит в отображении на карте объектов, входящих в различные кластеры, разными цветами.

4. Статистический анализ

При анализе исторических процессов на основе археологических объектов важную роль играет информация о распределении признаков объектов на карте. Наиболее важным фактором является частота встречаемости того или иного признака. Для нахождения распределения частот всю исследуемую территорию необходимо разбить на области.

Важным вопросом является выбор формы и размеров образуемых областей. Будем определять площадь областей в зависимости от общего количества археологических объектов N . Проведём разбиение так, чтобы в одну область попало в среднем n объектов. Таким образом, общее количество областей будет не менее N/n . Наложим на карту территории прямоугольную сетку. Найдём минимальное число m такое, что $m^2 \geq N/n$. Пусть максимальное расстояние с севера на юг между точками исследуемой территории равно d_y . Максимальное расстояние с запада на восток обозначим d_x . Вся территория может быть вписана в прямоугольник $d_x \times d_y$. Построим прямоугольную сетку с шагом по горизонтали d_x/m , а по вертикали d_y/m . Частоту встречаемости признаков археологических объектов будем вычислять по клеткам наложенной сетки.

Выбор параметра n осуществляет пользователь. Значение n влияет на размер ячеек сетки. Исследования могут быть реализованы методом последовательных приближений. Выбор большого n позволяет выделить крупные области, представляющие интерес. Уменьшая n можно получить более детальную информацию.

Визуализация результатов частотного анализа осуществляется с помощью цветовой раскраски сетки. Частотное распределение одного признака визуализируется градациями одного цвета. Палитра каждого цвета содержит $m + 1$ оттенков в интервале $[0, m]$. Отображая одновременно частотное распределение нескольких признаков, можно визуально выявить зависимости между ними.

5. Заключение

Таким образом, представление базы археологических объектов в виде географической информационной системы позволяет не только визуализировать данные, но и проводить их автоматизированный анализ. Взаимосвязь между свойствами объектов выявляется с помощью ассоциативных правил. Дальнейшему ручному анализу подлежат ассоциативные правила с высоким уровнем достоверности. Такие правила позволяют обнаружить новые закономерности в хорошо разработанных объектах. Также появляется возможность прогнозировать новые результаты для недостаточно разработанных объектов. Кластеризация позволяет сгруппировать схожие объекты. Визуализация результатов кластеризации играет роль предварительной обработки информации для дальнейшего ручного анализа.

Разработанная ГИС вместе с модулями интеллектуального анализа данных представляет собой мощный инструмент поддержки археологических исследований.

Система тестировалась для археологических объектов Омской области. Были выявлены закономерности, связанные с процессом исследования объектов. Также удалось выявить ряд закономерностей расположения объектов, относящихся к различным эпохам и культурам.

ЛИТЕРАТУРА

1. Gaffney V., Stancic Z. GIS Approaches to Regional Analysis: A Case Study of the Island of Hvar, Ljubljana, Yugoslavia. Znanstveni institut Filozofske fakultete, 1991.
2. Engur — history and culture. URL: http://engur.ru/ru_RU/map-prehistoric-art (дата обращения: 13.06.2016).
3. Артемьев Е.В., Дроздов Н.И., Зайцев Н.К., Шапарев Н.Я., Якубайлик О.Э., Шамамов А.В. Создание геоинформационной системы «Археологические памятники Красноярского края» // Вычислительные технологии. 1998. Т. 3, № 5. С. 5–10.
4. Афанасьев Г.Е., Савенко С.Н., Коробов Д.С. Древности Кисловодской котловины. М. : Мир науки, 2004. 240 с.

5. Гарбузов Г.П. Археология ландшафта и геоинформатика: теоретические аспекты взаимоотношений [Электронный ресурс]. / Археология и геоинформатика. М. : АГИС, ИА РАН, 2007. Вып. 4. (CD-ROM).
6. Деревянко А.П., Холушкин Ю.П., Бердников Е.В., Воронин В.Т. ГИС «Палеолит Северной Азии» // Информационные технологии в гуманитарных исследованиях. Вып. 6. Новосибирск : Ред.-изд. Центр НГУ. 2003. С. 21–29.
7. Geldener R., Kreienbrink F. Erfassung ortsgebundener Denkmale und Fundstellen // Ausgrabungen in Sachsen. 2010. No. 2. Arbeits- und Forschungsberichte zur sochsische Bodendenkmalpflege. Beiheft 21. P. 189–198.
8. Grunwald S. Das ergab aber ein so buntes wenig eindrucksvolles Bild. Zu den Anfängen der archeologischen Kartographie in Deutschland (1870–1914) // EAZ. 2012. No. 53. P. 5–34.
9. Stuble H. Braunkohlen- und Trassenarchaologie: eine Herausforderung mit Tradition // Ausgrabungen in Sachsen. 2010. No. 2. Arbeits- und Forschungsberichte zur sochsische Bodendenkmalpflege. Beiheft 21. P. 67–82.
10. Conolly J., Lake M. Geographical Information Systems in Archaeology. Cambridge Manuals in Archaeology. Cambridge : Cambridge University Press. 2006.
11. Moscati P. GIS application in Italian archeology // Archeologia e Calcolatori. 1998. No. 9. P. 191–236.
12. Tantillo M.D. GIS application in archaeological site of Solunto // XXI International CIPA Symposium, 01–06 October 2007, Athens, Greece.
13. Mejuto J., Castao J.G., Rodriguez-Caderot G. GIS Techniques in Archaeology: An Archaeoastronomical Approach // Archaeology, New Approaches in Theory and Techniques, Dr. Imma Ollich-Castanyer (Ed.), InTech, 2012.
14. Bevan A., Conolly J. GIS and Archaeological Survey Data: Four Case Studies in Landscape Archaeology from the Island of Kythera (Greece) // Postprint of a 2004 paper in Journal of Field Archaeology. 2004. No. 29. P. 123–138.

INTELLECTUAL GEOGRAPHIC INFORMATION SYSTEM OF ARCHAEOLOGICAL OBJECTS

S.V. Belim

Dr.Sc.(Phys.-Math.), Professor, e-mail: sbelim@mail.ru

D.M. Brechka

Ph.D. (Eng.), Associate Professor, e-mail: dbrechkawork@yandex.ru

T.A. Gorbunova

Ph.D. (Hist.), Associate Professor, e-mail: GorbunovaTA@omsu.ru

I.B. Larionov

Ph.D. (Eng.), Associate Professor, e-mail: LarionovIB@omsu.ru

I.V. Schmidt

Ph.D. (Hist.), Associate Professor, e-mail: ShmidtIV@omsu.ru

Dostoevsky Omsk State University

Abstract. The description of the geographic information system of archaeological objects ArGIS is provided. The objects and their properties are selected. Based on

the properties of archaeological objects the feature space is built. Clustering methods and methods of construction of association rules are applied to the archaeological objects as points of the feature space. The modules of visualization of data processing results are implemented. According to the data processing results, the data unloading in a text format occurs. The proposed system can automatically identify patterns in the distribution and dating of archaeological facts.

Keywords: GIS, archaeological objects, Data Mining.

ПОШАГОВАЯ ИНСТРУКЦИЯ НАПИСАНИЯ SHELL NAMESPACE EXTENSION

Д.С. Сиберт

студент, e-mail: dimaaasik.s@gmail.com

Омский государственный университет им. Ф.М. Достоевского

Аннотация. В статье приведена развёрнутая инструкция написания Shell Namespace Extensions, включающая создание пустого приложения и набора необходимых функций файловой системы.

Ключевые слова: пошаговая инструкция, windows, shell, namespace, extension.

Введение

Существуют популярные приложения Dropbox, Google Drive, Yandex.Диск и т.д., которые предоставляют возможность работать с файлами и директориями, расположенными на удалённых серверах, одному или нескольким пользователям.

Некоторые из этих инструментов встраиваются в операционную систему и позволяют работать со своим содержимым через её графическую оболочку. В семействе операционных систем Windows организовать взаимодействие пользователя с данными через GUI можно, написав расширение пространства имён оболочки (Shell Namespace Extension, далее NSE). Оно состоит из двух компонент:

- Менеджер данных.
- Интерфейс между менеджером данных и оболочкой Windows.

Первая делается полностью по усмотрению разработчика. Чтобы сделать вторую, необходимо реализовать набор интерфейсов: IShellFolder2, IPersistFolder2, IShellView, IEnumIDList, IDropTarget, IExplorerCommandProvider, IEnumExplorerCommand, IContextMenu, IShellExtInit, IObjectWithSite. К ним будет обращаться оболочка Windows за обработкой запросов пользователя. Кроме того, необходимо реализовать функции, стандартные для DLL: DllMain, DllGetClassObject, DllCanUnloadNow.

Информацию о написании NSE можно найти в MSDN [1], книге Visual C++ Windows Shell Programming [2], цикле статей [3]. Некоторые шаги могут быть упрощены:

- Графическое представление директории.
- Перечисление содержимого виртуальной директории.

Некоторые шаги реализации не представлены в этих источниках, а именно:

- Переименование директории\файла.

- Удаление директории\файла.
- Создание новой директории.
- Drag and drop элементов файловой системы в NSE и элементов внутри NSE.

В тексте статьи освещены все эти пункты.

Написание пустого расширения

В целях упрощения задачи можно воспользоваться готовым примером расширения из примеров Windows SDK версии 7.1 [4]. После установки SDK пример будет находиться по пути C:\ProgramFiles\Microsoft SDKs\Windows\v7.1\Samples\winui\shell\shellextensibility\explorerdataprovder. В примере используются актуальные, на момент написания статьи, версии классов, подходящие как для 32, так и для 64 разрядных систем. В расширении частично реализованы дополнительные функции, такие как контекстное меню (IContextMenu), панель инструментов (IExplorerCommandProvider), перечисление элементов (IEnumIDList), категоризация элементов (ICategoryProvider). Для пустого расширения в них нету необходимости. Рассмотрим только базовый класс и регистрацию в системе.

Основной класс NSE CFolderViewImplFolder (CFolder в листингах кода) находится в файле ExplorerdataProvider.cpp. Он реализует интерфейсы IShellFolder2, IPersistFolder2. Реализация отвечает за внутреннюю работу корневой и вложенных директорий, инициализации обработки панели инструментов, контекстного меню, графического представления и др. Каждой директории соответствует свой экземпляр такого класса.

Для работы пустого расширения необходимо реализовать только часть методов:

1. Методы интерфейса IUnknown (является базовым классом для IShellFolder2): QueryInterface, AddRef, Release. Они служат для того, чтобы получать определённый интерфейс объекта, увеличивать счётчик ссылок, уменьшать счётчик ссылок и уничтожать объект.

```
HRESULT CFolder::QueryInterface(REFIID riid, void **ppv)
{
    static const QITAB qit[] =
    {
        QITABENT(CFolder, IShellFolder),
        QITABENT(CFolder, IShellFolder2),
        QITABENT(CFolder, IPersist),
        QITABENT(CFolder, IPersistFolder),
        QITABENT(CFolder, IPersistFolder2),
        { 0 },
    };
    return QISearch(this, qit, riid, ppv);
}
```

```
ULONG CFolder::AddRef()
{
    return InterlockedIncrement(&m_cRef);
}

ULONG CFolder::Release()
{
    long cRef = InterlockedDecrement(&m_cRef);
    if (0 == cRef)
    {
        delete this;
    }
    return cRef;
}
```

2. Метод `IShellFolder::BindToObject`. Создает новый экземпляр класса директории для вложенной директории. Конструирует её PIDL и передаёт ссылку системе. Напомню, что в пространстве имён оболочки Windows каждый элемент уникально идентифицируется объектом PIDL. Это указатель на список структур, которые содержат полный путь до элемента в пространстве имён. При написании своего расширения разработчику обязательно надо будет осуществлять действия над этими структурами, такие как: создание, удаление, конкатенация, клонирование и другие. Во всех источниках, которые мне попадались, авторы реализовывали специальный класс – менеджер структур PIDL, который выполнял вышеописанные действия. Вместо этого можно воспользоваться функциями, которые предоставляет API Windows. Эти функции имеют префикс `IL*` и находятся в библиотеке `Shell32.lib`. Например, `ILClone`, которая создаёт полную копию некоторой структуры PIDL. Для корневой директории PIDL создаётся системой на основе заполненных при регистрации расширения для регистра данных и передаётся экземпляру класса директории с помощью метода `IPersistFolder::Initialize`.

```
HRESULT CFolder::BindToObject(
    PCUIDLIST_RELATIVE pidl, IBindCtx *pbc,
    REFIID riid, void **ppv)
{
    *ppv = NULL;
    HRESULT hr = S_OK;
    if (SUCCEEDED(hr))
    {
        CFolder* pCFolder = new (std::nothrow) CFolder();
        hr = pCFolder ? S_OK : E_OUTOFMEMORY;
        if (SUCCEEDED(hr))
        {
            PITEMID_CHILD pidlFirst = ILCloneFirst(pidl);
            hr = pidlFirst ? S_OK : E_OUTOFMEMORY;
            if (SUCCEEDED(hr))
```

```

    {
        PIDLIST_ABSOLUTE pidlBind =
            ILCombine(m_pidl, pidlFirst);
        hr = pidlBind ? S_OK : E_OUTOFMEMORY;
        if (SUCCEEDED(hr))
        {
            hr = pCFolder->Initialize(pidlBind);
            if (SUCCEEDED(hr))
            {
                PCUIDLIST_RELATIVE pidlNext = ILNext(pidl);
                if (ILIsEmpty(pidlNext))
                {
                    hr = pCFolder->QueryInterface(riid, ppv);
                }
                else
                {
                    hr = pCFolder->BindToObject(
                        pidlNext, pbc, riid, ppv);
                }
            }
            CoTaskMemFree(pidlBind);
        }
        ILFree(pidlFirst);
    }
    pCFolder->Release();
}
return hr;
}

```

3. Метод IPersistFolder::Initialize. Для корневой директории вызывается системой, а для вложенных внутри обработчика BindToObject, для инициализации экземпляра класса директории её уникальным идентификатором PIDL.

```

HRESULT CFolder::Initialize(PCIDLIST_ABSOLUTE pidl)
{
    m_pidl = ILCloneFull(pidl);
    return m_pidl ? S_OK : E_FAIL;
}

```

4. Метод IShellFolder::CreateViewObject. Вызывается системой для получения объекта, который затем будет использован для взаимодействия с директорией.

```

HRESULT CFolder::CreateViewObject(
    HWND hwnd, REFIID riid, void **ppv)

```

```

{
    *ppv = NULL;
    HRESULT hr = E_NOINTERFACE;
    if (riid == IID_IShellView)
    {
        SFV_CREATE csfv = { sizeof(csfv), 0 };
        hr = QueryInterface(IID_PPV_ARGS(&csfv.pshf));
        hr = SHCreateShellFolderView(
            &csfv, (IShellView**)ppv);
        csfv.pshf->Release();
    }
    return hr;
}

```

Для отображения директории в GUI используется реализация интерфейса IShellView.

5. Метод IShellFolder2:: GetDetailsOf. Он вызывается системой для получения более детальной информации о конкретном элементе директории. Даже для пустого расширения, для корректной работы, требуется верно обработать запрос с параметром iColumn равным 0.

```

HRESULT CFolder::GetDetailsOf(PCITEMID_CHILD pidl,
    UINT iColumn, SHELLDETAILS *pDetails)
{
    PROPERTYKEY key;
    HRESULT hr = S_OK;
    pDetails->cxChar = 24;
    WCHAR szRet[MAX_PATH];
    if (!pidl)
    {
        switch (iColumn)
        {
            case 0:
            {
                key = PKEY_ItemNameDisplay;
                pDetails->fmt = LVCFMT_LEFT;
                hr = StringCchCopy(
                    szRet, ARRAYSIZE(szRet), L"Name");
                break;
            }
            default:
            {
                hr = E_FAIL;
                break;
            }
        }
    }
}

```

```

    else if (SUCCEEDED(hr))
    {
        hr = S_OK;
    }
    if (SUCCEEDED(hr))
    {
        hr = StringToStrRet(szRet, &pDetails->str);
    }
    return hr;
}

```

6. Метод IPersist::GetClassID. Он вызывается системой, чтобы запросить GUID, сгенерированный для всего расширения.

```

DEFINE_GUID(CLSID_FolderViewImpl,
    0xba16ce0e, 0x728c, 0x4fc9, 0x98, 0xe5,
    0xd0, 0xb3, 0x5b, 0x38, 0x45, 0x97);

HRESULT Folder::GetClassID(CLSID *pClassID)
{
    *pClassID = CLSID_FolderViewImpl;
    return S_OK;
}

```

7. Необходима функция, которая будет создавать экземпляр класса корневой директории.

```

HRESULT CFolder_CreateInstance(REFIID riid, void **ppv)
{
    *ppv = NULL;
    CFolder* pFolder = new (std::nothrow) CFolder();
    HRESULT hr = pFolder ? S_OK : E_OUTOFMEMORY;
    if (SUCCEEDED(hr))
    {
        hr = pFolder->QueryInterface(riid, ppv);
        pFolder->Release();
    }
    return hr;
}

```

8. Метод IShellFolder::EnumObjects должен возвращать значение S_FALSE. Все остальные, присутствующие в интерфейсах методы, должны возвращать значение E_NOTIMPL.

Графическое представление директории

Для отображения директории необходимо реализовать интерфейс IShellView. Система запрашивает его при помощи вызова метода

CFolder::CreateViewObject(HWND hwnd, REFIID riid, void **ppv) с riid равным IID_IShellView. В зависимости от требований, можно сделать какой угодно внешний вид, в большинстве материалов по написанию расширений делают свою реализацию этого интерфейса. Если необходимо эмулировать стандартную работу с файловой системой, лучше всего воспользоваться функцией SHCreateShellFolderView(const SFV_CREATE *pcsfv, IShellView **ppsv). На выходе получаем объект, который отобразит директорию в обычной, для Windows форме (этот метод используется для пустого расширения из пункта выше). Непосредственно для отображения элементов в директории реализации IShellView необходимы данные о списке элементов, их типах, порядке и другие. Для этого системой будут вызываться методы интерфейсов IShellFolder и IShellFolder2: EnumObjects, CompareIDs, GetAttributesOf, GetDefaultColumn, GetUIObjectOf, GetDefaultColumnState, GetDetailsEx, GetDetailsOf, MapColumnToSCID.

При запросе интерфейса IShellView во входном параметре hwnd находится корректный уникальный идентификатор окна расширения. Его необходимо сохранить во внутреннюю переменную класса и в дальнейшем использовать для инициализации обновления окна, при помощи посылки сообщения.

```
SendMessage(m_hWnd, 0x111, 0x7103, 0);
```

Перечисление содержимого виртуальной директории

Чтобы отобразить список всех элементов в корневой и дочерних директориях, система вызывает метод IShellFolder:: EnumObjects.

```
HRESULT CFolder::EnumObjects(HWND hwnd, DWORD grfFlags,
    IEnumIDList **ppenumIDList)
{
    HRESULT hr;
    CEnumIDList *penum = new (std::nothrow) CEnumIDList();
    hr = penum ? S_OK : E_OUTOFMEMORY;
    if (SUCCEEDED(hr))
    {
        hr = penum->Initialize();
        if (SUCCEEDED(hr))
        {
            hr = penum->QueryInterface(
                IID_PPV_ARGS(ppenumIDList));
        }
        penum->Release();
    }
    return hr;
}
```

Ей необходимо передать указатель на реализацию интерфейса IEnumIDList, которая является итератором по всем элементам конкретной директории. Для каждого элемента система запрашивает информацию об атрибутах вызовом

метода `IShellFolder::GetAttributesOf`. Атрибутами определяется тип элемента (файл или папка), возможность переименования, возможность удаления и другие.

```
HRESULT CFolder::GetAttributesOf(UINT cidl,
    PCITEMID_CHILD_ARRAY apidl, ULONG *rgfInOut)
{
    HRESULT hr = S_OK;
    //Случай, когда система запрашивает
    //информацию об одном объекте
    if (1 == cidl)
    {
        //Прежде, чем выставлять флаги,
        //необходимо проанализировать apidl[0]
        //для того чтобы понять, какие свойства имеет объект
        //эти данные можно либо запросить у сервера,
        //либо хранить прямо в PIDL
        DWORD dwAttribs = 0;
        dwAttribs |= SFGAO_FOLDER;
        dwAttribs |= SFGAO_DROPTARGET;
        dwAttribs |= SFGAO_CANRENAME;
        dwAttribs |= SFGAO_CANDELETE;
        dwAttribs |= SFGAO_HASSUBFOLDER;
        *rgfInOut &= dwAttribs;
    }
    return hr;
}
```

В момент инициализации реализации `IEnumIDList` может понадобиться строковое представление пути к текущей директории. Его можно получить по `PIDL` с помощью функции `SHGetNameFromIDList(PCIDLIST_ABSOLUTE, SIGDN, PWSTR)`.

Переименование директории\файла

При использовании стандартной реализации `IShellView`, полученной с помощью функции `IShellFolder::SHCreateShellView`, переименование элементов обрабатывается классом директории. Каждому объекту, который можно будет переименовывать, необходимо выставить атрибут `SFGAO_CANRENAME` в методе `IShellFolder::GetAttributesOf`. В контекстном меню для этого элемента, появится пункт «переименовать». Также инициировать процесс переименования можно будет повторным нажатием на выделенный элемент. Если имя элемента будет изменено, то при подтверждении будет вызвана функция `IShellFolder::SetNameOf`, одними из параметров которой будут `PIDL` элемента и его новое имя. В теле этой функции необходимо обработать переименование.

Удаление директории\файла

Элементам директории можно выставлять атрибут SFGAO_CANDELETE в методе IShellFolder::GetAttributesOf, но даже при том, что в контекстном меню появляется активируемый пункт удаления, точки входа для обработки удаления, как в случае с переименованием, я не нашёл. Пришлось реализовывать не совсем стандартный способ.

У NSE существует два типа контекстных меню: для директории и для элементов директории. Контекстное меню для элементов директории является общим для всего расширения. Оно реализует интерфейсы IContextMenu, IShellExtInit. Подобно основному классу расширения, создание экземпляра класса меню оборачивается вызовом IClassFactory::CreateInstance. Для меню создаются соответствующие записи в регистре. Помимо основных команд, таких как переименование и открытие, реализуемых системой, контекстное меню можно расширять собственными командами. Удаление элемента будет одной из таких команд.

Для добавления и обработки своих команд необходимо реализовать функции IContextMenu::QueryContextMenu, IContextMenu::InvokeCommand, IShellExtInit::Initialize. QueryContextMenu - добавляет пользовательские пункты в меню.

```
HRESULT CContextMenu::QueryContextMenu(HMENU hmenu,
    UINT indexMenu, UINT idCmdFirst,
    UINT /*idCmdLast*/, UINT /*uFlags*/)
{
    WCHAR szMenuItem[80];
    //g_hInst - глобальная переменная,
    //которая инициализируется в DllMain
    LoadString(g_hInst, IDS_DELETE,
        szMenuItem, ARRAYSIZE(szMenuItem));
    //MENUVERB_DELETE - числовой идентификатор пункта меню,
    //для последующих пунктов можно брать на единицу больше
    InsertMenu(hmenu, indexMenu++, MF_BYPOSITION,
        idCmdFirst + MENUVERB_DELETE, szMenuItem);
    //тут добавляем следующие пункты...

    return MAKE_HRESULT(SEVERITY_SUCCESS, 0, (USHORT)(2));
}
```

Initialize - система вызывает функцию и передаёт через неё объект, содержащий информацию об элементе, для которого вызвано контекстное меню.

```
HRESULT CContextMenu::Initialize(
    PCIDLIST_ABSOLUTE /* pidlFolder */,
    IDataObject *pdtobj, HKEY /* hkeyProgID */)
{
    //Освобождаем старый объект
    if (m_pdtobj)
```

```

{
    m_pdtobj->Release();
    m_pdtobj = NULL;
}
//Запоминаем новый объект
m_pdtobj = pdtobj;
if (pdtobj)
{
    pdtobj->AddRef();
}
return S_OK;
}

```

InvokeCommand - выполнение команды над объектом, полученным при вызове функции IShellExtInit::Initialize.

```

HRESULT CContextMenu::InvokeCommand(
    LPCMINVOKECOMMANDINFO picl)
{
    HRESULT hr = E_INVALIDARG;
    UINT uID;
    //Получаем числовой идентификатор пункта меню,
    //способ получения зависит от реализации,
    //можно анализировать название команды
    if (SUCCEEDED(_GetCommandId(picl, &uID)) && _pdtobj)
    {
        switch (uID)
        {
            case MENUVERB_DELETE:
            {
                IShellItemArray *psia;
                hr = SHCreateShellItemArrayFromDataObject(
                    _pdtobj, IID_PPV_ARGS(&psia));
                IShellItem *psi;
                //Получаем объект с данными об элементе директории
                hr = psia->GetItemAt(0, &psi);
                if (SUCCEEDED(hr))
                {
                    int nDlgRes = MessageBox(
                        NULL, L"Sure?", L"Deleting!", MB_OKCANCEL);
                    if (nDlgRes == IDOK) {
                        //Обрабатываем удаление
                        //Посылаем сообщение перерисовки окну NSE
                        SendMessage(picl->hwnd, 0x111, 0x7103, 0);
                    }
                    psi->Release();
                    psia->Release();
                }
            }
            break;

```

```

    }
  }
}
return hr;
}

```

Создание новой директории

Создавать новые директории можно с помощью контекстного меню для директории. Для этого необходимо создать класс, реализующий интерфейс `IContextMenu`. Добавление пунктов в это меню происходит в функции `IContextMenu::QueryContextMenu`, точно так же, как и в предыдущем пункте. Обработка команды в функции `IContextMenu::InvokeCommand` выглядит иначе.

```

HRESULT FolderContextMenu::InvokeCommand(
    LPCMINVOKECOMMANDINFO pici)
{
    HRESULT hr = E_INVALIDARG;
    UINT uID;
    //Получаем числовой идентификатор пункта меню,
    //способ получения зависит от реализации,
    //можно анализировать название команды
    if (SUCCEEDED(_GetCommandId(pici, &uID)) && _pdtobj)
    {
        switch (uID)
        {
            case MENUVERB_CREATE:
            {
                m_pFolder->_OnCreateFolder();
                break;
            }
        }
    }
    return hr;
}

```

Инициализация экземпляра класса меню происходит в методе `IShellFolder::CreateViewObject`.

```

HRESULT CFolder::CreateViewObject(
    HWND hwnd, REFIID riid, void **ppv)
{
    *ppv = NULL;
    HRESULT hr = E_NOINTERFACE;
    //...
    if (riid == IID_IContextMenu)
    {
        FolderContextMenu* p_folderContextMenu =
            new (std::nothrow) FolderContextMenu(this);
    }
}

```

```

    hr = p_folderContextMenu ? S_OK : E_OUTOFMEMORY;
    if (SUCCEEDED(hr))
    {
        hr = p_folderContextMenu->QueryInterface(riid, ppv);
        p_folderContextMenu->Release();
    }
}
//....
return hr;
}

```

В конструктор класса меню передаётся ссылка на экземпляр класса директории. Это сделано, чтобы вызывать метод `_OnCreateFolder()`, который будет непосредственно обрабатывать создание новой директории.

Drag and drop элементов файловой системы в NSE и элементов внутри NSE

Для того чтобы NSE начало поддерживать эти операции, необходимо будет реализовать интерфейс `IDropTarget`. Он содержит методы, в которых будут обрабатываться перетаскивание файлов и отображение нужных эффектов.

Необходимо обработать два случая перетаскивания файлов. Если «бросить» элементы на пустое место текущей открытой директории, то будет вызван метод `IShellFolder::CreateViewObject` с `REFIID` равным `IID_IDropTarget`.

```

HRESULT CFolder::CreateViewObject(
    HWND hwnd, REFIID riid, void **ppv)
{
    *ppv = NULL;
    HRESULT hr = E_NOINTERFACE;
    //...
    if (riid == IID_IDropTarget)
    {
        IUnknown *pFolder;
        this->QueryInterface(IID_IUnknown, (void**) &pFolder);
        MyDropTarget *pMyDropTarget =
            new MyDropTarget(pFolder, NULL);
        pMyDropTarget->AddRef();
        hr = pMyDropTarget->QueryInterface(
            IID_IDropTarget, ppv);
        pMyDropTarget->Release();
    }
    //...
    return hr;
}

```

Если «бросить» элементы на элемент-директорию в текущей открытой директории, то будет вызван метод `IShellFolder::GetUIObjectOf(HWND`

hwndOwner, UINT cidl, PCUITEMID_CHILD_ARRAY apidl, REFIID riid, UINT *rgfReserved, void **ppv), при этом riid будет равен IID_IDropTarget, а первым и единственным элементом массива apidl будет PIDL целевой директории.

```
HRESULT CFolder::GetUIObjectOf(HWND hwnd, UINT cidl,
    PCUITEMID_CHILD_ARRAY apidl, REFIID riid,
    UINT * /* prgfInOut */, void **ppv)
{
    *ppv = NULL;
    HRESULT hr = E_NOINTERFACE;
    //...
    if (riid == IID_IDropTarget)
    {
        IUnknown *pFolder;
        this->QueryInterface(IID_IUnknown, (void**)&pFolder);
        MyDropTarget *pMyDropTarget =
            new MyDropTarget(pFolder, apidl[0]);
        pMyDropTarget->AddRef();
        hr = pMyDropTarget->QueryInterface(
            IID_IDropTarget, ppv);
        pMyDropTarget->Release();
    }
    //...
    return hr;
}
```

Для простоты методы IDropTarget::DragEnter, IDropTarget::DragOver, IDropTarget::DragLeave могут в любых случаях возвращать S_OK, при этом DragEnter и DragOver в выходной параметр DWORD *pdwEffect записывают значение DROPEFFECT_MOVE.

Основная обработка происходит в методе IDropTarget::Drop.

```
STDMETHOD(Drop)(LPDATAOBJECT pDataObj, DWORD dwKeyState,
    POINTL /*pt*/, LPDWORD pdwEffect)
{
    *pdwEffect = DROPEFFECT_NONE;
    IShellItemArray *psia;
    HRESULT hr = SHCreateShellItemArrayFromDataObject(
        pDataObj, IID_PPV_ARGS(&psia));
    IShellItem *psi;
    DWORD nFiles;
    std::list<TCHAR*> files;
    if (SUCCEEDED(hr))
    {
        hr = psia->GetCount(&nFiles);
        if (SUCCEEDED(hr))
        {
            for (DWORD i = 0; i < nFiles; ++i)
            {
```

```

        hr = psia->GetItemAt(i, &psi);
        if (SUCCEEDED(hr))
        {
            PIDLIST_ABSOLUTE pidl;
            hr = SHGetIDListFromObject(psi, &pidl);
            wchar_t *szFileDropped = new wchar_t[MAX_PATH];
            SHGetNameFromIDList(pidl,
                SIGDN_DESKTOPABSOLUTEEDITING, &szFileDropped);
            files.push_back(szFileDropped);
        }
    }
}

IDropHandler *pDropHandler;
this->m_pFolder->QueryInterface(IID_IDropHandler,
    (void**) &pDropHandler);
pDropHandler->DoDrop(files, this->m_subfolder);
while(!files.empty())
{
    TCHAR *file = files.back();
    CoTaskMemFree(file);
    files.pop_back();
}
this->m_pFolder->Release();
return S_OK;
}

```

Экземпляр класса реализации IDropTarget на вход конструктора принимает указатель на экземпляр класса директории типа IUnknown, и PIDL элемента-директории (если процесс перетаскивания завершился на нём). Это сделано, чтобы вызывать метод IDropHandler::DoDrop, где IDropHandler - это специальный интерфейс, который реализовывает класс директории.

```

#pragma once
#include <list>
#include <vector>
//{{052979AD-F172-4318-A828-9A6AC20FC403}
static const GUID IID_IDropHandler = { 0x52979ad, 0xf172,
    0x4318, { 0xa8, 0x28, 0x9a, 0x6a, 0xc2, 0xf, 0xc4, 0x3 } };

class IDropHandler
{
public:
    virtual void DoDrop(std::list<TCHAR*> files,
        LPCITEMIDLIST subfolder) = 0;
};

```

Реализацию NSE со всеми вышеописанными элементами можно найти по ссылке <https://github.com/DmitrySibert/virtual-folder>.

ЛИТЕРАТУРА

1. Creating Shell Data Source Objects and Extending the Shell Namespace // msdn.microsoft.com : информ.-справочный портал. [Электронный ресурс]. URL: [https://msdn.microsoft.com/en-us/library/windows/desktop/ff521656\(v=vs.85\).aspx](https://msdn.microsoft.com/en-us/library/windows/desktop/ff521656(v=vs.85).aspx) (дата обращения: 04.03.2016).
2. Esposito D. Visual C++ Windows Shell Programming. Wrox Press Ltd., 1998.
3. Dunn M. The Complete Idiot's Guide to Writing Shell Extensions. Part I // www.CodeProject.com: информ.-справочный портал. 2006. 15 марта. [Электронный ресурс]. URL: <http://www.codeproject.com/Articles/441/The-Complete-Idiot-s-Guide-to-Writing-Shell-Extensions> (дата обращения: 04.03.2016).
4. Microsoft Windows SDK for Windows 7 and .NET Framework 4 // msdn.microsoft.com : информ.-справочный портал. [Электронный ресурс]. URL: <https://www.microsoft.com/en-us/download/confirmation.aspx?id=8279> (дата обращения: 04.03.2016).

STEP BY STEP GUIDE TO WRITING SHELL NAMESPACE EXTENSION

D.S. SibertStudent, e-mail: dimaasik.s@gmail.com

Dostoevsky Omsk State University

Abstract. The article provides detailed guide for writing Shell Namespace Extension, including creating a blank application and a set of necessary file system functions.

Keywords: guide, windows, shell, namespace, extension.

FUTURE INTERNET ARCHITECTURE: CLEAN-SLATE VS EVOLUTIONARY DESIGN

A. Taran

Postgraduate Student, e-mail: alina.taran.v@gmail.com

D.N. Lavrov

Associate Professor, Ph.D. (Eng.), e-mail: dmitry.lavrov72@gmail.com

Dostoevsky Omsk State University

Abstract. Long ago developed the Internet has a range of the problems such as IP's narrow waist, security needs, availability, routing scalability, support of mobility and multihoming. Many appeal to the known "clean slate" approach to resolve these challenges facing Internet today. In this paper we consider two concepts of future Internet architecture: clean slate and evolutionary research. In part about clean slate we mention the main ideas, design goals and several of examples in this area. Then we describe the Internet design process through biological metaphor and consider the pros and cons of two ways of research.

Keywords: Internet, computer network, evolutionary design.

Introduction

As it is known, the Internet was developed as reliable data transmission system for the United States Department of Defense, then this network was used with academic purpose in different scientific areas and after that came to broader market becoming a social phenomenon. It was unpredictable in the early '70 that the Internet take up such a great part of our world. Internet has entered into areas of people's life such as business, economy, military, communication. This changes can be traced by fact that a number of connected hosts has grown from less than 200 in the 80s to millions in 2014. Also the nature of Internet traffic has been changed according to new requirements of society. It has led to the appearance of Web, file sharing, video streaming and other new applications. Furthermore, do not only software companies such as Microsoft, Oracle, SAP and others take advantages of the Internet, but other major sectors like the automotive industry, the entertainment industry, banks, insurance companies.

Such broad possibilities for application affect the huge success of the Internet, but it is also a cause of many difficulties and diversity of the system, which grew out from sending e-mails in the university circles to interconnecting the world. Moreover, developers of services and applications should respond as quickly as possible to the growing needs of users throughout various areas of industries. Due to these requirements of users, the Internet is faced with few challenges which are mentioned below.

First, a lack of security in the Internet is the biggest problem. It is not enough to add security to each layer or individual protocol used in the Internet, because security in each component don't give the security in common system. Fishing, spam, spyware, worms and viruses are gaps in security, problems faced by average people. It might be said that these problems are not responsibility of the network. But if we speak about a future Internet as a trusted and cooperative system, the goal of "true" security should be stated. That's why a defensible position of the network role in supporting the endhost security should be claimed, and a consistent division of responsibility between the network and the endhost system should be proposed.

There is one more issue faced by internet service providers (ISP) constantly. The ISPs should provide a service which complies with user requirements of the Internet's crucial role in both business and private life, in terms of reliability, resilience, and availability. If a number of users or applications in systems have an intention to communicate and these systems are allowed to communicate by the policies of the interconnecting networks, then they should be able to do so. This aim seems to be quite simple, but it is not. It is not so easy to remedy any barrier to availability ranging from transient routing instability to denial of service attacks.

Also the problem of flexibility, extensibility and capability of supporting innovations in Internet architecture can be mentioned. Also this problem is well known as "narrow waist" of IP: today the Internet is primarily used for content distribution, but not just for host-to-host communication. Such architecture, with network layer (IP) in the middle and various design above and below it, implements all the functionality necessary for global interconnectivity, but in the same time it is rigid and makes it hard to change middle layer to adapt for future unanticipated and large-scale changes in network behaviour.

Another unresolved problem is network management. There is the lack of tools for recognition which network items are being used by which users and which applications are currently using which specific network components. We still do not understand how to set up and control network for reliable operation, easy management, debugging, and still scales well.

The Internet is experiencing a new stage of development — a shift from PC-computing to mobile computing. Mobility has become an important component of the future Internet. It is also very difficult to add mobility to the current Internet architecture as the current Internet naming system is based on the host address, typically the IP address. To achieve scalability of routing, an address hierarchy is used which imposes a structure on the host addresses that relates to its location within the Internet. Thus putting mobility as the norm is one more big task of the present.

There are two essential approaches to resolve these challenges. The incremental or evolutionary research means that an investigated system is changed incrementally, is moved from one state to another supporting old potential and adding new opportunities. The second, clean-slate approach, means that every new state of system is developed from scratch, providing old possibilities on new principles.

Many researchers believe that it is impossible to solve the problems mentioned above in the current internet architecture. So a lot of papers has appeared in the direction of rethinking the fundamental assumptions and design decisions, and starting from scratch. The research community nowadays intends to focus not on the incremental method, but on a clean-slate approach attempts to find appropriate solution for all these issues. We will consider and give an evaluation of both approaches.

1. Clean-slate approach

As were said above, the aim of clean-slate approach is not to be limited by the existing architecture. Proponents of this approach intend to use benefits of current technologies, opportunities and demands to suggest a brand-new one which will be easily manipulated, easy integrated and light-exposed to changes due to new requirements. As the Internet has been used more then 30 years now, probably it is reached the point where people are unwilling or unable to experiment on the current architecture. This approach has good state of mind: out-of-the-box thinking, the design of alternative network architectures, exploring of the architecture, so that it was more likely to be integrated into the network environment. Contribute to development of this approach proceeds breakthroughs in technologies such as fast packet optical components, wireless networks, fast packet forwarding hardware, and significant computational resources [5].

It is noteworthy that a few attractive and valuable enhancements, like multicast, Mobile IP, IPSec, QoS mechanism or IPv6, have not been widely integrated in spite of theoretical overall excellence [5]. Probable reason is the risk to replace a complex system that already works, no short-term benefits for adopters, difficulty of updating the equipment base. Therefore, some part of the community believes that this type of research is complementary to the indispensable evolutionary approach. Further, we review some clean-slate projects which are forward-looking or have already yielded fruit: GENI, named data networking, and network virtualization.

1.1. Global Environment for Network Innovations (GENI)

The National Science Foundation (NSF) in the United States has initiated the research program which has strong focus on defining an architecture for the Future Internet, and it plans to encourage research teams to reach solidarity on broad architectural themes. For providing an open, large-scale, realistic experimental facility for evaluating new network architectures, the Global Environment for Network Innovations (GENI) was initiated by NSF. The map of GENI is shown in Figure 1.

GENI not only supports existing projects on a dedicated backbone network infrastructure, but also guides other infrastructure platforms to participate in the federation — the device control framework to provide these participating networks with users and operating environments, to observe, measure, and record the resulting experimental outcomes [7]. That is why it is different from other testbeds

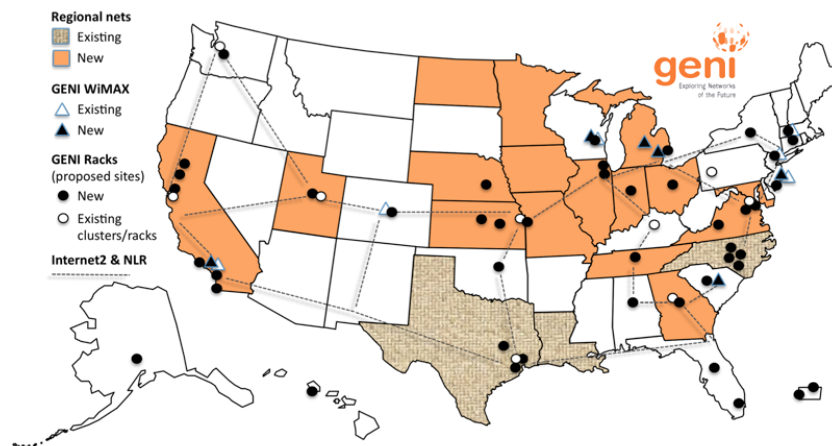


Figure 1. The GENI map

because there are no limits on the network architectures, services, and applications to be evaluated. Thus it allows clean-slate designs to experiment with real users under real conditions.

The main goal of GENI is to construct multiple virtualized slices out of the substrate for resource sharing and experiments. It contains two key points:

- Physical network substrates that are expandable building block components
- A global control and management framework that collects the building blocks together into a connected object

So that in GENI testbeds there are two main branches: the first is deploying a prototype testbed federating different small and medium ones together, the second is running observable, controllable, and recordable experiments on it. Concerning these different branches there are several researching fields, such as GENI experiment workflow and service working group; the control framework; campus/operation, integration, and security area; and instrumentation and management. All in all by delivering instrumentation and measurement support, which is implemented as initial entity of GENI, open and extensible testbed for experimentation with new Internet architecture is provided.

1.2. Named Data Networking (NDN)

The Named Data Networking was funded by NSF in September 2010 as one of the four projects under NSF with participation from about 13 universities and research institutes from the United States, France, China. The main goal of this project is reconstructing current layer network architecture following stretched initial assumptions of the Internet, which often cause challenges and confrontation with user's needs.

This new architecture is focused on the fact, that usage of Internet has shifted from end-to-end packet delivery to a content-centric model. The old one is increasingly limited and makes difficult to conform to IP's requirement to communicate by discovering and specifying location. Figure 2 is an illustration of the Named Data Networking model to build another "narrow waist" around content instead of IP.

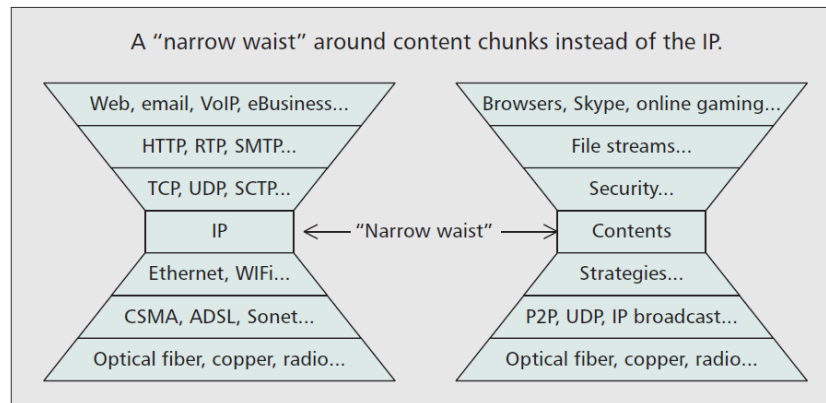


Figure 2. The new "narrow waist" of NDN (right) compared to the current Internet (left) [7]

The NDN architecture assumes naming data instead of their locations, and in this way data become a first-class entity of this structure. So it allows to develop a new Internet architecture that can rely on strengths — and address weaknesses — of the Internet's current host-based, point-to-point communication architecture in order to naturally provide emerging needs of communication. Transfer to a new "what"-model from "where"-model can resolve the technical challenges that must be addressed in a future Internet architecture: routing scalability, fast forwarding, trust models, network security, content protection and privacy, and fundamental communication theory [7].

Furthermore, the concept of encryption also changes. Instead of trying to secure the transmission channel or data path through encryption, Named Data Networking decouples trust in data from trust in hosts and tries to secure the content by naming the data through a security-enhanced method. What is more, such architecture enables to optimize the traffic on the network's part and multiple copies of the same data would not be sent between endpoint on the network again.

1.3. Network virtualization

Let us consider the research which represents a success of clean-slate thinking. Network virtualization has long been a challenge of the network research community. The aim of the network virtualization is that multiple isolated logical networks, each with potentially different addressing and forwarding mechanisms, share the same physical infrastructure. In order to better understand the principle, let us turn to the analogy with computer virtualization. Reason for the success of

computer virtualization lies in an abstraction of the underlying hardware. Rather the computer virtualization layer has a hardware abstraction that allows slice and share of resources among the guest operating systems. So each operating system thinks that it has its own private hardware. Thus there is a hardware abstraction, which allows to quickly respond to changes both above and below the virtualization layer and allows the system be more competitive due to its flexibility.

Also network virtualization guarantees the improving resource allocation, permits operators to checkpoint their network before changes, and provides competing customers to share the same equipment in a controlled and isolated scope [10].

Moreover, virtual networks also claim to provide a safe and realistic environment to deploy and evaluate experimental "clean slate" protocols in production networks.

Let us give an example of project in network virtualization area. The OpenFlow initiative is an innovative concept to providing opportunities to test new network architecture which was initiated by Stanford Clean Slate project. Protocol OpenFlow is based on an Ethernet switch, with an internal flow-table, and a standardized interface to add and remove flow entries. The structure of these tables is different between vendors, but it has a certain common set of functions. The OpenFlow provides the tools for experimenters to act on the flow tables to change the way the router forwards packets of certain flows [9]. So it allows researchers to run experiments on heterogeneous switches and can be used to create virtual network.

2. Evolutionary approach

The other approach is evolutionary research. The goals of evolutionary Internet research are to understand requirements and behaviour of the current Internet, to clarify existing problems, and to resolve them with two essential limitations: backward compatibility and incremental distribution.

We can understand the processes in computer network through the biological metaphor. Let us make arrangement what we mean by "species" and "environment" in terms of computer network. By species we mean a system which is based on a conceptual organization or architecture (e.g., datagram internetworking), uses certain protocols (e.g., IP, TCP, 802.11), is implemented on various technologies (e.g., electronic routers, optical transmission), and it supports several applications (e.g., Web, P2P) [8]. This system is surrounded by an "environment": a big number of users which interact between them, service needs, outside danger, economic circs, and external conditions. As in biology, network-environment changes constantly, permanently and what is more important uncertainly over the time. There is one good example of this unpredictability — network security. Historically this component was not a big issue. In those times when Internet was designed people did not think about such type of problem, because as it was said Internet was intended for other purposes.

So nature has opportunistic mechanism not to let species die in permanently changing environment — evolution. Biological evolution is based on three factors: generic heritage, mutation and selection. These three basic mechanism can be sim-

ilarly illustrated in Internet evolution. To begin with, generic heritage is referred to property of backward compatibility. Future generations of network's architecture, protocols, backend technologies keep up mostly because of their predecessors.

Second, mutation is referred to transformation of network-species components such as architecture, protocols, technologies or applications. By this process there is no creation of new species. When the external conditions have changed, then nature calls for evolution process and predecessor is pushed by evolution variates itself. Of course, there is a difference between network variation and biological in that the first is not random. That is why we have to cope on their own, the scientific community or one individual can create plenty of highly probable survivable variations. Still we can not predict the best of this mutation due to a non-static environment, we can make a conditions for nature and let environment select the most competitive mutation. So it is our responsibility to provide the most diverse variations of predecessor.

Finally, environment chooses one mutation via natural selection. Each mutation is associated with a certain degree of validity i.e., the ability of the current network to satisfy the needs of the environment in that period of time. Certainly as in world of nature, in Internet researches wins the mutation that is associated with the highest validity; in other words, that mutation which has lower costs to deploy and expansion.

3. Evaluation

During the last years Internet architecture design was "clean-slate" oriented. The goal of clean-slate design is to make a new architecture of Internet significantly better, design it from scratch [1]. The author is convinced that it is too ambitious aim which hardly can be successful in reality. There are many reasons to believe that.

Firstly, economical reason plays a major role. If we have got two opportunities: to use something new or use something that works badly, but is still working, then we likely choose the second option. Costs of integration of new product far exceed the benefit in the short run. Furthermore, it is hardly to estimate where in new technology can be a weak point and which problems can we get through deploying this technology in real world. So it is too hard for new technology to replace the working one. It should be a "really big urge" for deploying new architecture. The problem is not the lack of experimentation, but the fact that new technology is not so competitive as previous ones.

Another argument of proponents of clean-slate approach is "ossification" current Internet architecture [6]. As we said above, the middle layer of protocol stack is a narrow place of network architecture: everything goes through it. We can look at it from the other side. Internet architecture should relate a variety of constantly appearing link-layer technologies with a diversity of services and applications. The central protocols of the architecture is a background, which must be stable and expand very slowly to support innovations at other layers of the architecture. In terms of biology, such stable background is Gene Regulatory Networks, which

were based about 510 million years ago and have not substantially advanced since then. Despite the different morphology of human body parts, this GRNs are a component part of them [4]. So evolutionary core delivers a stable basis on which the higher levels may be developed.

What we can do now is stop thinking of the current Internet architecture as an artefact and to start thinking of it as an expanding ecosystem which is affected by many areas of science. The author believes that a new architecture model can not come out of the blue. It is always hard work which is learned by mistakes of the past. We can not abandon its past, that is a natural choice of a mutation that occurs in nature. Researches can be focused on measuring and understanding of current problems. Our goal is to help the evolution by creating great variety of different ideas which firstly can resolve this challenge and secondly can be adopted by current Internet with backward compatible and incrementally deployable. As we involved in this evolution, we should continue to direct and bend the ideas of researches which collaborate and communicate.

There is a successful example of such collaboration from the mathematical field. There is a list of twenty-three problems in mathematics as well-known Hilbert's problems which were published by German mathematician David Hilbert in 1900. He presented ten of the problems at the Paris conference of the International Congress of Mathematicians and published the complete list of 23 problems a pair years later. In his report he formulated the problems which were the most significant for mathematics at the beginning of XX century. No one before him did not set out such titanic task. There were a lot of different directions in math, and it was very difficult for one person to cover all of its sections. But Gilbert was broad-minded: he worked in almost all existing fields of mathematics and in many of them made significant results. Based on the success of co-working mathematical colleagues (19 problems were fully or partly resolved), Clay Mathematics Institute stated seven problems in mathematics (The Millennium Prize Problems) in 2000. A correct solution to any of the problems results in a US \$ 1,000,000 prize being awarded by the institute. Our aim in investigations of future Internet technologies is to continue pointing out an issues for current situation and expanding the areas of investigation by cooperation of researchers. What also can be done is to make such impact for good science which requires relevance to the real world, to formulate problems in current Internet and create a good environment for cooperation of people to make this good science.

Conclusion

The author believes that there is the one opportune way in reality — evolutionary Internet research. Unfortunately, evolutionary approach is frequently considered as lightweight and small-minded, as a range of "hacks". The biggest difference between clean-slate and evolutionary research is to take into consideration the existing Internet. The main goal of evolutionary research is to design protocols in the context of existing environment and ecosystem, but not in a vacuum. As well-known mathematician Walter Savitch said, "In theory there is no difference

between theory and practice; in practice there is.” A brand-new protocol or service can not be so competitive and robust as already working analogues. Also it neither considers all links between variety of all players in the Internet environment, nor displays the spectrum and synergy of all technologies and application in the real Internet.

Besides, we know very well one concept which already works successfully in practice. Nature has the good mechanism to adapt in new conditions of reality — evolutionary process. The evolution of species can be identified by three natural mechanisms, such as genetic heritage, mutation and selection, which help to survive in radically changing environments without needing a clean-slate restart [3]. Going back to the concept of evolution of the Internet, evolutionary research can be viewed as the creation of survivable mutation. We are able to produce multiple of manifold solutions which can compete then in the real Internet. The author believes that we can not state what will be important and necessary in 10 years. We can not predict the future, we do not even truly know the present. What we can do now is to invest and support the various of ideas and innovation depending on real world to find the optimal solution which meets the requirements of the present.

REFERENCES

1. Bellovin S.M., Clark D.D., Perrig A., Song D. A clean-slate design for the next-generation secure internet. 2006.
2. Papadopoulos J.D.C. NDN Testbed. Accessed January 15, 2015.
3. Dovrolis C. What would Darwin think about clean-slate architectures? // *ACM SIGCOMM Computer Communication Review*, 38(1):29–34, 2008.
4. Dovrolis C., Streelman J.T. Evolvable network architectures: What can we learn from biology? // *ACM SIGCOMM Computer Communication Review*, 40(2):72–77, 2010.
5. Feldmann A. Internet clean-slate design: what and why? // *ACM SIGCOMM Computer Communication Review*, 37(3):59–64, 2007.
6. Pan J., Jain R., Paul S., So-In C. MILSA: A new evolutionary architecture for scalability, mobility, and multihoming in the future internet // *Selected Areas in Communications*, *IEEE Journal on*, 28(8):1344–1362, 2010.
7. Pan J., Paul S., Jain R. A survey of the research on future internet architectures // *Communications Magazine*, *IEEE*, 49(7):26–36, 2011.
8. Jennifer Rexford and Constantine Dovrolis. Future Internet architecture: clean-slate versus evolutionary research // *Communications of the ACM*, 53(9):36–40, 2010.
9. Roberts J. The clean-slate approach to future internet design: a survey of research initiatives // *Annals of telecommunications*, 64(5):271–276, 2009.
10. Sherwood R., Gibb G., Yap K.-K., Appenzeller G., Casado M., McKeown N., Parulkar G. Flowvisor: A network virtualization layer. 2009.

БУДУЩЕЕ АРХИТЕКТУРЫ ИНТЕРНЕТ: С «ЧИСТОГО ЛИСТА» ПРОТИВ ЭВОЛЮЦИОННОЙ РАЗРАБОТКИ

А.В. Таран

аспирант, e-mail: alina.taran.v@gmail.com

Д.Н. Лавров

к.т.н., доцент, e-mail: dmitry.lavrov72@gmail.com

Омский государственный университет им. Ф.М. Достоевского

Аннотация. Давно разработанный Интернет накопила множество проблем, таких как нехватка IP адресов, растущие потребности в области безопасности, доступности, масштабируемости маршрутизации, мобильности. Многие призывают к разработке «с чистого листа» для разрешения этих проблем, с которыми сталкивается работа Интернет в настоящее время. В этой работе мы рассматриваем две концепции будущего Интернет-архитектуры: «с чистого листа» и эволюционного развития. В части, посвящённой разработке с нуля, мы опишем основные идеи, цели проектирования и несколько примеров. Затем мы опишем процесс проектирования Интернет через биологические метафоры и рассмотрим все достоинства и недостатки двух способов развития.

Ключевые слова: Интернет, компьютерные сети, эволюционное развитие.

ADAPTIVE VIDEO MULTICAST STREAMING OVER MULTIRATE WIRELESS LANs

A. Taran

Postgraduate Student, e-mail: alina.taran.v@gmail.com

D.N. Lavrov

Ph.D. (Eng.), Associate Professor, e-mail: dmitry.lavrov72@gmail.com

Dostoevsky Omsk State University

Abstract. Video multicast services over wireless network has become widespread nowadays. Multicasting over wired networks approach can not be directly transfer and assumed to wireless network. In this paper the overview of such problems of wireless multicast is considered. The main factors for wired multicast for video streaming are discussed. An analysis and an overview of existing mechanism for multicast video streaming are given.

Keywords: multicast, video streaming, wireless network.

Introduction

The video streaming has become one of the popular application driving wireless LAN nowadays. But multicast protocols over wireless network have some challenges. Noise, multipath, interference, and mobility of hosts cause rapid changes of channel condition in wireless environment [11].

The most growing network technologies in wireless communication is IEEE 802.11 wireless LANs (WLANs). But there are two main problems, which are faced effective deployment of multicast services over WLANs. Firstly, in the IEEE 802.11 standards it is determined, that multicasting is a simple broadcasting mechanism, where the Acknowledgement (ACK) frames are not used. Such nonavailability of feedback mechanism of course makes a very strong affect on the reliability of the service, which will be provided to the user. The second problem is that according to the IEEE 802.11a/b standard all frames with multicast and broadcast Receiver Address (RA) should be transmitted at one of the rates included in the basic rate set. But in the standard it is not defined how the best rate can be chosen. Also it is not specified how the rate can be adopted according to the spectrum of the possible current wireless channel conditions during the time [11].

In the second section the small summation of the technologies in video streaming is presented. The focus of the paper is the multicast technology. The first goal of this paper is to consider existing problems in wireless video streaming and find out what makes it difficult to extend the actual wired approach. The second aim is to make overview of existing mechanisms for multicast video streaming and to make an analysis of pros and cons of them.

1. Video Streaming and Variants of Implementing of Technology

In all streaming media is multimedia that will be transferred from provider and is constantly received by and displayed by an end-user [15]. The common form of streamed media can be presented by the Internet television. The video streaming has become especially popular since 2000s. The main reasons for that are increasing of network bandwidth and increased access to the Internet for average user. This method of processing of delivering media now is also a competitive approach, such as downloading.

There are some protocols which support streaming media. One of them is User Datagram Protocol (UDP) where the streaming data is sent as series of small packets. This protocol hasn't got guarantee of the delivery, so it is make it difficult to detect loss or corruption and recover data. Also, if it happens that the data is lost, the stream can be dropped out.

Also there are reliable protocols which guarantee correct delivery of each bit in the media stream, such as Transmission Control Protocol (TCP). However, if there is a large amount of errors by connection or by confirmation of received information, then the transmitted information can become irrelevant. It can also cause significant delays in the transmission of information on the time spent on sending corrupted data. About such delays, which can be significant critical for the real-time multimedia communication, it will be mentioned in third subsection in the next section.

Unicast protocols send from a server to each node an individual copy of the media stream. These not-scalable protocols also can be burdened for network equipment.

Multicast Protocols, as were said above, are developed to decrease the load on servers if there is a large number of customers getting a streaming multimedia.

2. Multicast

Firstly, it should be mentioned what is multicast commonly means. Multicast is a type of group communication where there is a one sender who transfer the information to a group of destination nodes simultaneously [12]. Multicast in computer networking can be implemented at the Internet layer with the method of IP multicast which sends IP datagrams to a set of interested receivers in a single transmission [13].

Some applications, such as distance learning, sending email, radio, video on demand, video conferencing, support multicast. In unicast network an individual connection with each received node is set even when there is consumption of one resource on the common route. In multicast communication a source node sends only one copy of the data for the common route to recipients who has a subscription. The advantages of this approach are that, if it has added new users in subscription, then there is no need to increase bandwidth for a common route to consumers. Accordingly, the load on the intermediate equipment will be

reduced [5].

2.1. Characteristics of Wired and Wireless Multicast

Wired multicast, where there is a fixed number of users, can be extended also to a wireless environment where there are changing numbers of mobile users. But direct transfer of this approach to the wireless multicast can cause some difficulties (see Table 1). For example, the bandwidth, which is available in the two directions of any given wireless link, can be unequal. Furthermore, they can be unidirectional. That can influence transferred information which is sent between two mobile nodes, and it can lead to some problems. Inefficient multicast tree/mesh, loss of packets, incorrect routing, even the discarding of multicast packets — all can take place and effect certain consequences [10].

2.2. Requirements on Wireless Multicast Video Streaming

The world of wireless multicast application is very broad, so requirements, which wireless multicast serves must fulfill, are very different. In the paper of Varchney six application categories are given: interactive games, mobile and locational advertising, mobile distance education, proactive service management, product recommendation systems, management, product location and search [10]. Also, the six groups of multicast requirements are present which are characterised for these types of application. From this table it is possible to extract requirements which suit not just mobile nodes, but also common applications and to the focus of this paper — video streaming.

So, for video streaming high bandwidth and very low delay are very important, such as a few hundred ms. Secondly, service interruptions due to intermittent connectivity or brief disconnectivity significantly affects the users' overall experience. Thirdly, reliability and QoS requirements are very significant. But requirements, such as security and privacy, are not critical for multicast video streaming.

2.3. Real-Time Multimedia Communication and It's Requirements

Real-Time multimedia communication is a special sort of multicast streaming [8]. Real-time voice and video data must be synchronous and ideally must be processed without delay. Of course, in reality it is not possible to process without delay, but the most important challenge for real-time communication is to minimize this delay.

The voice and video data streams can be considered as a sequence of samples which have finite size. Also samples must be generated, transmitted and received at fixed time intervals which must not be gone over this time limit. So, there are four types of such time constraints: *sampling*, *packetization*, *network transmission* and *presentation* delay. The delay between the generation samples of the stream object is called sampling delay. The packetization delay is the time interval which is required to generate a sample and transfer it to the network. The time, that

Table 1. Qualitative comparison of wired and wireless multicast [10]

Issue	Current “wired” multicast	Wireless and mobile multicast	Possible ways to support wireless and mobile multicast
Type of links	Symmetrical and fixed characteristics Broadcast links in LANs	Possibly asymmetrical and/or unidirectional links of varying performance and point-to-point links in cellular and PCS	Design of new protocols to handle route asymmetry and unidirectional links without reverse-path information (possible history and prediction-based schemes)
Bandwidth	Plentiful	Limited and variable amount	Protocols to adapt membership management and routing updates to the amount of bandwidth available and user mobility
Topology	Fixed	Fixed in infrastructure-based, dynamic in ad hoc networks	Protocols for both fixed and changing topology by “sensing” topological changes
Loss of packets	Infrequent (< 1%)	Frequent and variable (1%-30% based on links)	Error control with possible retransmission from neighboring user(s)
Membership changes	Only when a user leaves or joins a group	Also when a user moves to another location	Protocols with reduced overhead for managing membership
Routing	Fixed routing structure throughout the multicast session	Routing structure subject to change due to user mobility	Protocols that could dynamically adapt the routing to current structure and available resources
Security Issues	Less complex due to fixed users and wired links	More complex due to wireless links and possible use of broadcasting	Encryption and security techniques in routing and membership management
Quality of service	Individual routes can use RSVP	Due to user mobility, RSVP may cause excessive overhead	Design of new protocols for “soft” QoS under varying link conditions and mobility
Reliability	Possible use of a transport-layer protocol (such as the Multicast File Transfer Protocol)	More complex due to wireless links and user mobility: possible unwanted interaction of protocols at transport and link layers	Design of new protocols that could allow one or more different retransmission schemes at one or more protocol layers

needs to transmit the sample over network, is called network transmission delay. Finally, presentation delay is the time which is requisite for buffering the sample before presenting it to the user. The last three delays also can be called by the common term — transmission delay.

In a local area network environment the most dominant delay is the packetization delay [8]. Also, the choice of the sample size plays important role in magnitude of this delay. That is why the size of one sample should be chosen so that the packetization delay would be acceptable.

As it was said above, the big problem by real-time communication is a maxi-

mum allowable streaming delay. And a significant different real-time communication from generic streaming is that a late packet is also not useful as a lost packet. This means that is no sense to use acknowledgement packets to detect lost packet by user. It happens because retransmitted packet often come out of timeout which is appropriate for real-time communication.

3. The Problems of Multicast Video Streaming in WLAN

As it was said in the third section, wired multicast for video streaming can be also extended to a wireless environment, but there are certain challenges which will be described in details further. The main reason of it is the transfer to wireless network adds some network performance characteristics. Variable data rates, packet loss, and multicast unreliability include in new consent and do not settle in the traditional approaches to guaranteed quality of service (QoS). So, in following sub-sections these characteristics are described [3].

3.1. Variable Data Rate

The first important difference between wired and wireless environment is the data rate of transmission in wireless network which changes over time. Also, the data rate in WLAN in opposite to wired network depends on the distance of the client from the access point (AP). For example, if a wired connection in the second type of network is operating at 100 Mb/s today, it will operate at 100 Mb/s also the next day. The one of consequences of such data rates is the throughput of individual video flows and the capacity of the overall network which changes with time [3].

The traditional QoS approach of bandwidth reservation and admission control do not suit such variable throughput and capacity. For example, the client can operate at 54 Mb/s, and the requested video stream rate is 10 Mb/s. The system recognizes that the necessary airtime for the new stream can be accommodated, and so accepts the stream. If after that the client goes away from the access point, the data rate of the client drops to 6 Mb/s. So the client data rate become smaller as by requested stream, and the video stream can not be supported anymore. Thus, sending video over a wireless network can be considered as sending video over the public Internet, due to widely varied user's experience and throughput over time.

3.2. Packet Loss

Another problem of wireless network in comparison to a wired is the relative unreliability of the underlying Layer 2 transport. So-called unreliability is that in WLAN environment much more packets are lost than wired.

The first reason of packet loss is collisions: two devices, which are connected with WLAN, attempt to transmit at the same time. There is a prevention measure, the "listen-before-talk" medium access method, against such collisions in a shared half-duplex medium which WLAN uses. But all of them can not be avoided. This

situation can be escalated if non-wireless devices work in the same band as WLAN devices. Most of them do not communicate, following the “listen-before-talk” algorithm, and so collisions can be faced.

The second reason of packet loss is that wireless transmissions suffer from short-term signal loss or fades. Fades can appear due to some physical properties of objects, for example due to absorption from intervening objects in the environment or reflections of waves in the environment accidentally causing signal cancellation.

The third factor of packet loss is that in WLAN systems are trying different rates for the best transmission rate. So, by hunting of the best, some packets can be lost during the search process.

If all these reasons will be summarized (combination of collisions, fades, and data rate selection), an underlying packet error rate (PER) can give 5 percent. For the error compensation, in wireless network the retransmission mechanism can be used. Packets, which are not successfully received and acknowledged, can be resent. This mechanism often helps to reduce the final packet loss rate (PLR) to less than 0.1 percent. The other side of the coin is that these retransmissions cause jitter¹ and take network throughput, both of which can influence QoS. And even after work of retransmission mechanism, the final PLR can stay still much higher than is typically observed on wired connections.

3.3. Multicast Unreliability

For the wireless multicast traffic the underlying packet error rate plays an important role. In wireless network there is no according retransmission mechanism for multicast with multiple receivers. So the PLR for multicast traffic stays equal to the PER. As a result, the final packet loss rate for wireless multicast traffic can be about 5 percent. In worst case, it can be very big problem for video where loss of even a single packet can result in an error that spreads for many video frames. For this reason, it is quite normal situation for multicast video applications, which work on a wired network, to fail completely when they operate on a wireless network.

Variable data rate, packet loss, multicast unreliability — each of these factors can strongly influence a video, so the application must handle by itself how to manage this situation and decries the impact of these three factors.

4. Overview and Analysis of Existing Dynamic Rate Adaptation Mechanism for Multicast Video Streaming

In this section we present some new proposals to liquidate mentioned problems. All these proposals concern a popular method of Dynamic Rate Adaptation [1]. To understand the backgrounds and foundation of rate adaptation mechanisms designed for multicast transmission, we also demonstrate pioneer rate adaptation mechanisms in IEEE 802.11 for unicast transmission.

¹Jitter is the deviation from true periodicity of a presumed periodic signal in electronics and telecommunications, often in relation to a reference clock source [14]

4.1. Rate adaptation mechanisms in IEEE 802.11

The first proposal and widely used rate adaptation protocol in commercial is a result of Kamerman and Monteban — Auto Rate Fallback (ARF) [6]. The main idea of ARF-protocol is when signal to noise ratio (SNR) decreases, then an access point tries to recover by decreasing the bandwidth. The process of increasing and decreasing of the rate by access point take place as follows. If the access point successfully receives a certain number of packets (10), then it switches to a higher rate. It decreases a rate if a failure occurs right after rate increase. That means that the access point switches to a lower rate if there is two consecutive failures of 10-sequence transmissions. This mechanism was implemented in a lot of numbers of commercial products, and the feedback of this approach is positive, but also has got a big minus. Because of the static-thresholds, it can not be adapted well to varying condition in wireless networks.

To take this drawback away, Lacage et al. have proposed dynamic-threshold method Adaptive ARF (AARF) [7]. The authors got threshold-based mechanism as in ARF as basis, but they do not use the fixed number of the threshold. The main idea of this mechanism is to calculate binary exponential threshold at runtime to make the mechanism better reflect the channel conditions. For the calculation of new higher rate, the last number of consecutive successful transmission must be multiplied by two. The advantage of this mechanism is that the period between successive failed attempts to use a higher rate is increased. Also it is made in fewer failures and retransmissions, so the overall throughput is improved. As shows practice, the AARF-mechanism is relatively efficient, but it cannot be used in multicast transmission. The reason of that is that in implementation of this protocol acknowledgement and retransmission which are disabled in multicast are used.

Holland et al. proposed another popular protocol — Receiver-Based Auto Rate (RBAR) [4]. The main aim of this approach is to optimize the performance in wireless networks using rate-adaptation mechanism at MAC layer. The authors use some kind of communication mechanism — Request/Clear To Send(RTS/CTS). This channel probing mechanism is enabled in order to get/send feedback from receiver. RTS frame is always sent out before any transmission by sender. This frame will be obtained by the receiver who computes the SNR of the frame. Then the receiver sends back the transmission rate which is got from a mapping table of SNRs and rates. And then this new transmission rate will be used by sender in the next transmissions in CTS. In this implementation the headers of the RTS and CTS have been modified for the purposes. The first disadvantage of this protocol is that the SNR which is used for calculation of new transmission rate, is a physical parameter that does not always correlate well with human perception. The second disadvantage of RBAR is unavailability for multicast transmissions, because RTS/CTS are unusable in those cases.

4.2. Rate adaptation in wireless multicast

Basalamah et al. developed Rate Adaptive Multicast protocol (RAM) which is based on similar idea of using RTS/CTS in RBAR for channel evaluation and selection of rate. The RTS frame is used by multicast receivers to estimate channel condition, and based of it, the receiver send back transmission rate for sender to use in CTS. If the multicast node do not receive the data frame correctly, this node must send a NACK (Not Acknowledge). Frame sequence field is added to RTS for enhancing the throughput and it is used by member of the multicasting to control which data frame is received: a new frame or retransmission of a previously successfully received frame. If a frame is marked as retransmission frame, then a member will not participate in the multicast transmission. Because of aforementioned technique, this field helps overall to reduce the number of retransmission. As a disadvantage can be mentioned that the protocol makes use of RTS/CTS, NACK and retransmission which are disable in multicast. Also, due to retransmission and additionally field in RTS, it is received many frames which are already exist.

To solve the problem of oversaturated feedback, the Leader-based Rate Adaptive Multicasting for Wireless LANs (LM-ARF) protocol was proposed by Choi et al [2]. This protocols, which is based on ARF, uses the same data rate adaptation mechanism, but also uses leader-based feedback. One of the receiving stations is chosen as leader, and it is responsible for sending ACKs on behalf of the stations which are participating in multicast communication. If the station, which is not marked as leader, cannot receive a multicast frame, it must send a negative acknowledgement (NAK) to request retransmission. The access point keeps fairness between unicast flows by uniform distribution of window size the same way as that of a unicast transmission. In this protocol a new frame type is added — CTS-to-Self frame which serves for guarantee of the channel access and announcement of the transmission of a multicast frame. This protocol has a lot of advantages such as fairness, reliability, effective and good performance, but there is also a disadvantage as in ARF — static-threshold approach and other drawbacks of ARF as well.

Villalón et al. proposed Auto Rate Selection for Multicast (ARSM) protocol which instead of using RTS/CTS uses multicast channel probe operation (MCPO) [11]. This operation must be processed before multicast traffic will be sent: access point must send out the multicast probe frame. The user, which have the lowest SNR, must reply to the access point by multicast response. After that, the access point select the multicast data rate which based on type of response: explicit, implicit, and no feedback. Also, in this protocol there is implemented mechanism to avoiding collision: multicast users select backoff timer according to their SNR value.

Another effective protocol was proposed by Park et al. : SNR-based Auto Rate for Multicast (SARM) [9]. Authors are taking into account user perception in this protocol, and the transmission rate will be adapted according to node's SNR which has the worst channel condition. SNR references are mapped in one table where the each element must be higher than 30 to represent good quality, for

each transmission rate. The practice is showed that the wireless channel can be used more efficiently if multicast transmission rate will be changed based on SNR values. To solve the feedback problem, it is proposed to make a channel probing mechanism and after that send notification to the AP. To avoid collision during feedback transmission to the access point, as in ARSM, it uses a backoff timer based on the SNR.

For better representation, the protocols, which were considered above, are summarized in Table 2 below.

Table 2. Summary of Rate Adaptation Protocols

	Protocol	Threshold	Feedback
Unicast	ARF	static	ACK
	AARF	dynamic	ACK
	RBAR	static	RTS/CTS
Multicast	RAM	static	RTS/CTS, NACK
	ARSM	static	Channel probing
	LM-ARF	static	Leader-based
	SARM	static	Channel probing

Conclusion

In recent years, the interest to video application over wireless network rises up. As was mentioned in section two, multicasting is an relative effective technology of data transition to a group of users. The reason for that is that the data is transferred to a group of destination nodes simultaneously, i.e in a single transmission. However, the high packet loss ratio and bandwidth spectrum of wireless channels produce a problem for video multicast over wireless networks [1].

In this paper we give an overview of the mechanisms for multicast video streaming which solve these problems by a method of dynamic rate adaptation, and mechanisms for unicast which were used as foundation for the next researches in multicast direction.

REFERENCES

1. Alay Ö., Korakis T., Wang Y., Panwar S.. Dynamic rate and FEC adaptation for video multicast in multi-rate wireless networks // *Mobile Networks and Applications*, 15(3):425–434, 2010.
2. Choi S., Choi N., Seok Y., Kwon T., Choi Y. Leader-based rate adaptive multicasting for wireless LANs // *Global Telecommunications Conference, 2007. GLOBECOM'07. IEEE*, pp. 3656–3660.

3. Inc. Cisco Systems. Optimizing Enterprise Video Over Wireless LAN White Paper. URL: http://www.cisco.com/c/en/us/products/collateral/wireless/5508-wireless-controller/white_paper_c11-577721.html, 2014. [Online; accessed 23-March-2016].
4. Holland G., Vaidya N., Bahl P. A rate-adaptive MAC protocol for multi-hop wireless networks // Proceedings of the 7th annual international conference on Mobile computing and networking, pp. 236–251. ACM, 2001.
5. Kapoor D. What's the difference between real-time processing and stream processing? URL: <https://www.quora.com/Whats-the-difference-between-real-time-processing-and-stream-processing>, 2012. [Online; accessed 20-March-2016].
6. Kamerman A., Monteban L. WaveLAN®-II: a high-performance wireless LAN for the unlicensed band // Bell Labs technical journal, 2(3):118–133, 1997.
7. Lacage M., Manshaei M.H., Turletti T. IEEE 802.11 rate adaptation: a practical approach // Proceedings of the 7th ACM international symposium on Modeling, analysis and simulation of wireless and mobile systems, pp. 126–134. ACM, 2004.
8. Nicolaou C. An architecture for real-time multimedia communication systems // Selected Areas in Communications, IEEE Journal on, 8(3):391–400, 1990.
9. Park Y., Seok Y., Choi N., Choi Y., Bonnin J.-M. Rate-adaptive multimedia multicasting over IEEE 802.11 wireless LANs // Consumer Communications and Networking Conference, 2006. CCNC 2006. 3-rd IEEE, V. 1, pp. 178–182.
10. Varshney U. Multicast over Wireless Networks // Commun. ACM, 45(12):31–37, December 2002.
11. Villalon J., Cuenca P., Orozco-Barbosa L., Seok Y., Turletti T. Cross-layer architecture for adaptive video multicast streaming over multirate wireless LANs // Selected Areas in Communications, IEEE Journal on, 25(4):699–711, 2007.
12. Inc Wikimedia Foundation. Multicast. URL: <https://en.wikipedia.org/wiki/Multicast/>, 2015. [Online; accessed 23-March-2016].
13. Inc Wikimedia Foundation. IP Multicast. URL: https://en.wikipedia.org/wiki/IP_multicast, 2016. [Online; accessed 23-March-2016].
14. Inc Wikimedia Foundation. Jitter. URL: <https://en.wikipedia.org/wiki/Jitter>, 2016. [Online; accessed 29-March-2016].
15. Inc Wikimedia Foundation. Streaming media. URL: https://en.wikipedia.org/wiki/Streaming_media, 2016. [Online; accessed 23-March-2016].

АДАПТИВНАЯ ГРУППОВАЯ ПЕРЕДАЧА ВИДЕО ЧЕРЕЗ БЕСПРОВОДНЫЕ ЛОКАЛЬНЫЕ СЕТИ С РАЗНОЙ СКОРОСТЬЮ ПЕРЕДАЧИ

А.В. Таран

аспирант, e-mail: alina.taran.v@gmail.com

Д.Н. Лавров

к.т.н., доцент, e-mail: dmitry.lavrov72@gmail.com

Омский государственный университет им. Ф.М. Достоевского

Аннотация. Услуги групповой рассылки видео через беспроводные сети сейчас широко распространены. Подходы, используемые для групповой передачи данных в проводных сетях, не могут быть напрямую перенесены в беспроводные среды. В данной работе рассматриваются проблемы групповой рассылки в сетях беспроводного доступа, основные механизмы проводной групповой передачи видео потоков, даётся обзор и анализ существующих механизмов групповой передачи видео потоков.

Ключевые слова: групповые рассылки (multicast), видео потоки, беспроводные сети.

ИДЕНТИФИКАЦИЯ ЛИЧНОСТИ ПО БИОМЕТРИЧЕСКИМ ХАРАКТЕРИСТИКАМ ЖЕСТОВ НА СЕНСОРНОМ ЭКРАНЕ

К.И. Лейфер

студент, e-mail: leyfer.kirill@gmail.com

Д.Н. Лавров

к.т.н., доцент, e-mail: dmitry.lavrov72@gmail.com

Омский государственный университет им. Ф.М. Достоевский

Аннотация. В статье проводится исследование возможных способов идентификации пользователя по жестам, вводимым им на сенсорном экране смартфона.

Ключевые слова: биометрия, идентификация, сенсорный экран, смартфон.

Введение

В наши дни смартфоны, оснащённые сенсорными дисплеями, получили повсеместное распространение. Сенсорный экран в смартфонах применяется с 1992 года, и за это время он превратился в довольно точное и отзывчивое устройство ввода. В большинстве современных смартфонов тачскрин способен распознавать до 10 одновременных касаний с точностью до одного пикселя, а так же позволяет определять площадь каждого из касаний в широком диапазоне. Всё это позволяет отслеживать с высокой точностью каждый жест пользователя, оставляемый им на сенсорном экране. Можно предположить, что жесты, вводимые пользователем на сенсорном экране, могут обладать неким набором уникальных для каждого человека особенностей, то есть содержать в себе набор некоторых биометрических признаков. Сенсорный экран, исходя из перечисленных выше особенностей, вполне способен зафиксировать эти признаки, выступая в роли биометрического сенсора.

Поиск таких биометрических характеристик и построение на их основе системы идентификации пользователей представляют большой интерес с точки зрения информационной безопасности и будут подробно рассмотрены в данной статье.

Для поиска биометрических характеристик в сенсорных жестах необходимо накопить некоторое количество тестовых данных. Эту задачу можно возложить на клиент-серверное приложение. Клиентская часть должна устанавливаться на устройства тестовых пользователей и отслеживать все жесты, вводимые на сенсорном экране. Эта часть должна собирать всю необходимую статистику не создавая пользователю никаких неудобств, а затем автоматически отправлять

собранные данные на сервер. Серверная часть проекта должна сохранять всю статистику в базе данных.

После сбора данных необходимо провести исследование полученной информации и поиск биометрических признаков, позволяющих построить систему биометрической идентификации пользователей. Затем необходимо построить данную систему и протестировать её работу.

При разработке системы использовались руководства [1, 2]. Работа опиралась на ряд выводов, сделанных в [3], и проводилась в рамках программы исследований по биометрии кафедры компьютерных технологий и сетей факультета компьютерных наук ОмГУ им. Ф.М. Достоевского [4–18]. В перечисленных работах исследуются статические и динамические биометрические характеристики. Наибольшую сложность вызывают динамические характеристики: идентификация по голосу, по походке, клавиатурный почерк и др. Жесты на сенсорном экране также относятся к динамическим характеристикам.

1. Описание и принцип работы проекта TouchLogger

Клиентская часть

Разработанное клиент-серверное приложение получило название TouchLogger. Клиентская часть представляет собой android-приложение, которое устанавливается на смартфоны с включённым полным доступом к устройству. Сбор данных клиентской частью осуществляется из любого приложения, в котором работает пользователь, будь то клиент социальной сети, экран блокировки или виртуальная клавиатура. Пользователь может сгенерировать довольно большой объём данных, поэтому их отправка на сервер осуществляется только при наличии соединения с Интернетом по Wi-Fi. Перед отправкой собранные данные помещаются во временное хранилище на накопителе устройства, и как только устанавливается Wi-Fi соединение, все накопленные данные автоматически отправляются на сервер, и временное хранилище очищается.

Стоит учитывать, что жесты сенсорного ввода могут содержать любую информацию о пользователе: от данных кредитных карт до личной переписки. Поэтому, несмотря на небольшой масштаб исследования, стоит предусмотреть некоторые механизмы защиты информации. Для этого в проекте TouchLogger используется следующий криптографический протокол:

$$A \rightarrow B : E_{K_o}(K), E_K(M),$$

где M — открытое сообщение, K_o — открытый ключ алгоритма RSA, K — сеансовый ключ, A — клиент, B — сервер.

Клиент, сгенерировав сеансовый ключ K , шифрует им своё сообщение, а сам ключ K шифрует с помощью заранее известного всем клиентам открытого ключа K_o . Зашифрованное сообщение вместе с зашифрованным ключом клиент отправляет на сервер. Сервер с помощью секретного ключа K_c расшифровывает

сеансовый ключ и уже с его помощью расшифровывает сообщение и помещает его в базу данных.

Итак, собираемые данные передаются по сети в зашифрованном виде и используются только в исследовательских целях. Однако тестовый пользователь всё равно может счесть эти меры сохранности личных данных недостаточными и найти какой-либо способ избежать отправки части данных на сервер. Бороться с этим невозможно, поэтому самый лучший выход из данной ситуации — внедрить в клиентскую часть функцию «паузы», которая будет приостанавливать сбор данных по желанию пользователя.

Серверная часть

Серверная часть написана на языке JavaScript с использованием платформы Node.js. Данная платформа позволяет в довольно короткий срок разработать несложный сервер, что и стало основной причиной её выбора. В качестве СУБД была выбрана MongoDB. Эта СУБД представляет документо-ориентированный подход к хранению данных, иначе называемый NoSQL (в противопоставление традиционным SQL базам данных). Выбор был продиктован необходимостью хранения большого количества вложенных данных, а также тем, что документ в MongoDB можно напрямую создать из JavaScript объекта или из документа в формате JSON. Последняя особенность также значительно упростила и ускорила разработку серверной части.

2. Анализ полученных данных

В сборе пользовательских данных приняли участие 19 пользователей, установивших на свои устройства клиентскую часть проекта. Экраны у большей части смартфонов имели диагональ 5 дюймов и разрешение 1920x1080 пикселей. В результате за 14 дней тестирования в сумме было получено около 200000 жестов.

Классификация данных

Для того, чтобы проанализировать полученные данные, необходимо классифицировать их некоторым образом с целью дальнейшего исследования отдельных классов. В этой работе данные были классифицированы следующим образом:

1. по количеству касаний в одном жесте:
 - а) однопальцевые
 - б) многопальцевые (multitouch)
2. по типу касания:

- а) «тачи» (от англ. «touch» — прикосновение) — простое нажатие на экран. Для этих жестов характерно небольшое (часто — нулевое) расстояние, проделанное пальцем от момента касания до окончания жеста.
- б) «свайпы» (от англ. «swipe» — происхождение термина неизвестно, означает быстрый «смахивающий» жест на сенсорном экране). Для таких жестов характерно большое расстояние, проделанное пальцем за время жеста.

Для разбиения всех однопальцевых жестов на тачи и свайпы необходимо вычислить расстояние, пройденное пальцем в течение жеста. Затем необходимо найти пороговое значение этого расстояния так, чтобы все жесты, длина которых меньше этого значения, можно было отнести к тачам, а остальные — к свайпам. Очевидно, подобная задача стояла и перед разработчиками ОС Android, поэтому в исходных кодах можно отыскать величину этого порогового значения, она составляет 4 dp (dp — density-independent pixels, «пиксели, не зависящие от плотности» — абстрактная единица измерения длины, эквивалент одного пикселя при плотности экрана 160 точек на дюйм). На экране смартфона с разрешением 1920x1080 и диагональю 5 дюймов 4 dp примерно соответствует 12 пикселям. Итак, все жесты длиной более 12 пикселей будем считать свайпами, остальные — тачами.

Имея пороговое значение, классифицируем все жесты на одно- и многопальцевые, а первые разделим на тачи и свайпы. Построим процентное соотношение этих жестов (см. рис. 1).

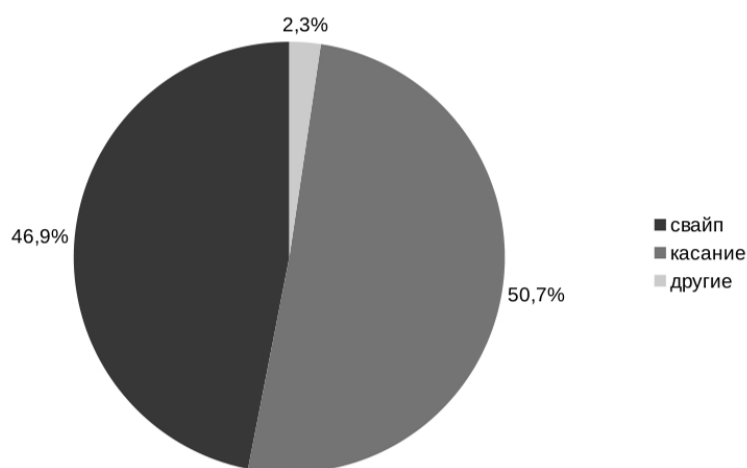


Рис. 1. Распределение жестов на тачи, свайпы и многопальцевые жесты

Как видно из графика, многопальцевые (мульти-тач) жесты составляют менее 3% от общего количества событий ввода, что оставляет для анализа при-

близительно по 300 таких жестов с каждого пользователя. Этого явно недостаточно для полноценного исследования, поэтому в дальнейшем будем анализировать только тачи и свайпы.

Анализ распределения тачей по площади касания

Для каждого события ввода известно большое количество различных параметров, в том числе площадь касания. Это число от 0 до 1. У каждого пользователя распределение этого значения должно отличаться в силу различной толщины пальцев и различной частоты использования разных пальцев для ввода данных. Проверим эту гипотезу, построив распределение всех тачей по значению средней площади касания, по площади касания в начале тача и в конце тача. Пример полученных распределений для двух разных пользователей приведён на рис. 2.

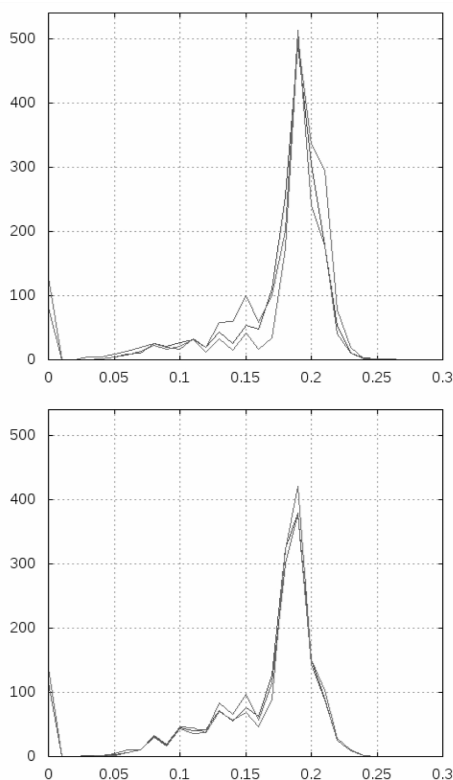


Рис. 2. Распределение тачей по площади касания

По оси X отложено значение площади касания, по оси Y — количество жестов. Три разных линии на графике представляют распределение средней площади касания, площади касания в начале и конце жеста. График отражает незначительные отличия в распределении площади касания среди разных пользователей, соответственно, ни одна из трёх характеристик не является уникальной для пользователей, следовательно, построение на их основе системы

идентификации невозможно.

Анализ тачей по продолжительности

У разных пользователей продолжительность касаний, зависящая от многих параметров, может значительно отличаться. Проверим это гипотезу, построив распределение тачей по продолжительности касания. На рис. 3 приведён пример такого распределения для четырёх пользователей на основании данных, полученных за некоторый промежуток времени.

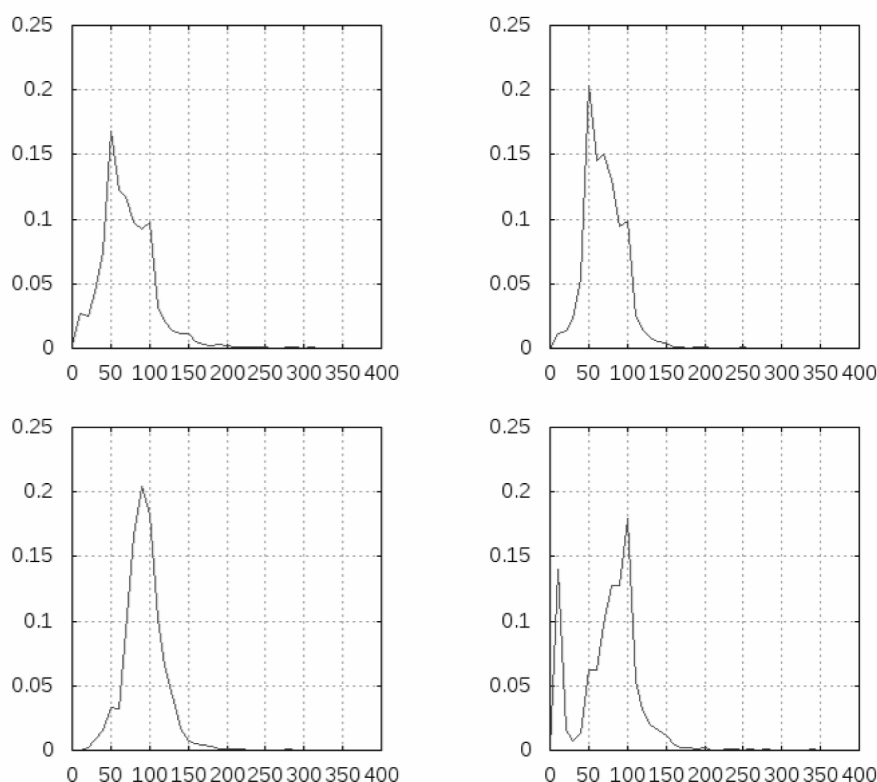


Рис. 3. Распределение тачей по продолжительности

На этом рисунке по оси X отложена продолжительность касаний в миллисекундах, по оси Y — доля жестов данной продолжительности. На первый взгляд, формой графики значительно отличаются друг от друга, а значит, распределение тачей по продолжительности может быть использовано в качестве биометрической характеристики в системе биометрической идентификации. Однако, если построить графики распределения этой характеристики для одного и того же пользователя за разные промежутки времени, они так же будут значительно отличаться формой. Отсюда следует, что несмотря на наличие некой уникальности этой характеристики, система идентификации, построенная на её основе, покажет довольно плохие результаты.

Анализ распределения свайпов по расстоянию

Другая характеристика, которая может быть уникальной для пользователей — это распределение свайпов по их длине. Для ускорения расчётов будем считать, что длина свайпа — это расстояние между точкой касания и точкой, где пользователь отпустил палец. То есть в расчётах будем использовать не длину пути, а расстояние. Такое упрощение обосновано тем, что распределение свайпов по расстоянию практически не отличается от распределения по длине пути, но в то же время вычисляется значительно быстрее.

Итак, построим распределение свайпов по расстоянию. На рис. 4 приведён пример графиков распределения этой величины для четырёх разных пользователей.

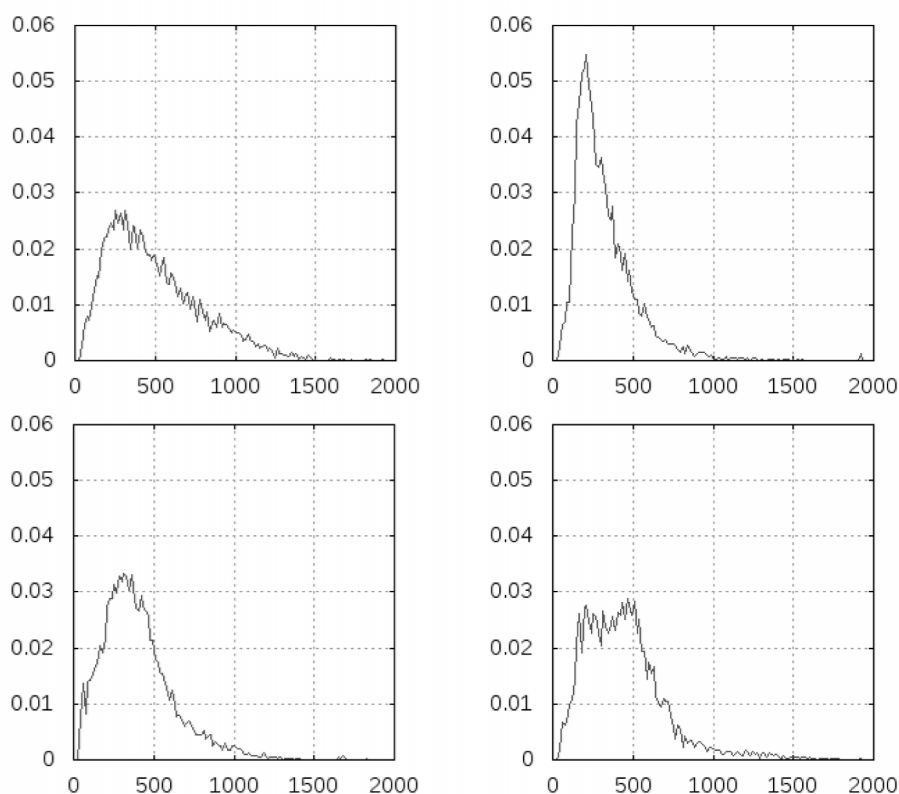


Рис. 4. Распределение свайпов по расстоянию

Графики отличаются по форме, более того, при выборке данных для одного и того же пользователя за разные промежутки времени форма графика распределения изменяется незначительно. Это говорит о том, что данная характеристика является уникальной для каждого пользователя и по ней можно построить систему биометрической идентификации. Однако не стоит ожидать от системы, построенной по данной характеристике, выдающихся результатов. На рисунке выше видно, что отличия в форме графиков незначительные, а для большого количества пользователей станут ещё менее различимыми.

Анализ распределения свайпов по продолжительности

Ещё одна характеристика, представляющая интерес — это распределение свайпов по продолжительности в миллисекундах. Построим распределение этой величины и рассмотрим несколько примеров такого распределения на рис. 5.

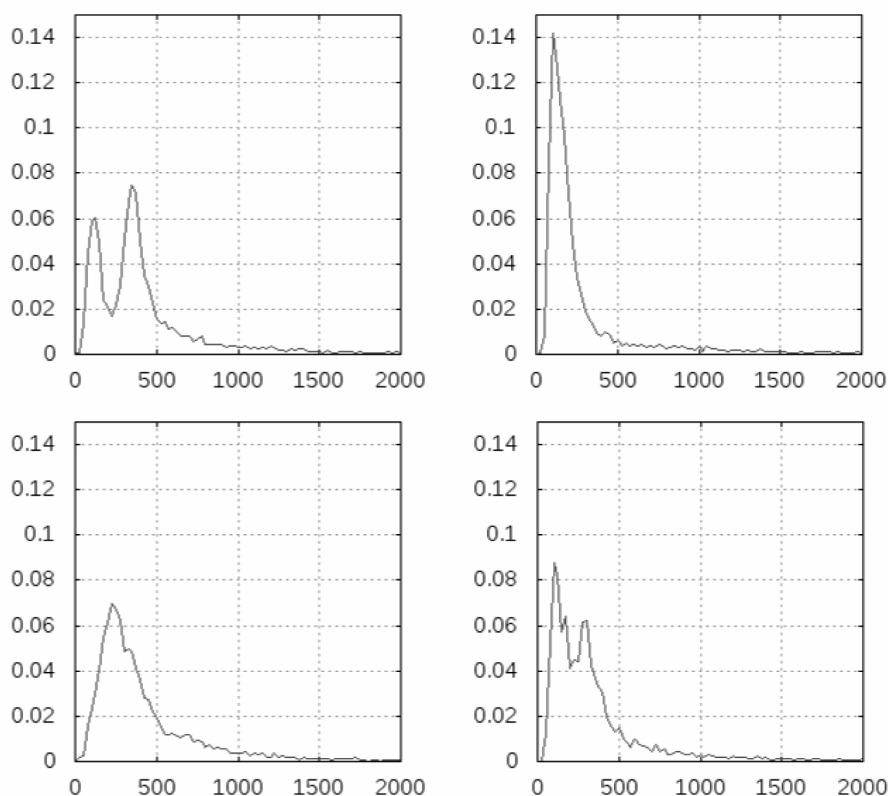


Рис. 5. Распределение свайпов по продолжительности

Данная характеристика демонстрирует на графиках гораздо более сильное разнообразие, которое может обеспечить хороший процент успешных идентификаций пользователя по этим признакам. Более того, распределение данных, полученных в разные промежутки времени для одного и того же пользователя, будет практически идентичным. Значит, эту характеристику также можно использовать для построения системы идентификации.

Другие способы анализа данных

Кроме построения распределения некоторой величины существуют и другие способы анализа пользовательских данных. На рис. 6 представлено распределение точек по координатам.

Если применить к этим данным некий алгоритм кластеризации, то для каждого пользователя можно будет построить эталонный набор кластеров точек касания, с которыми затем можно будет сравнивать поступающие на вход тестовые данные.

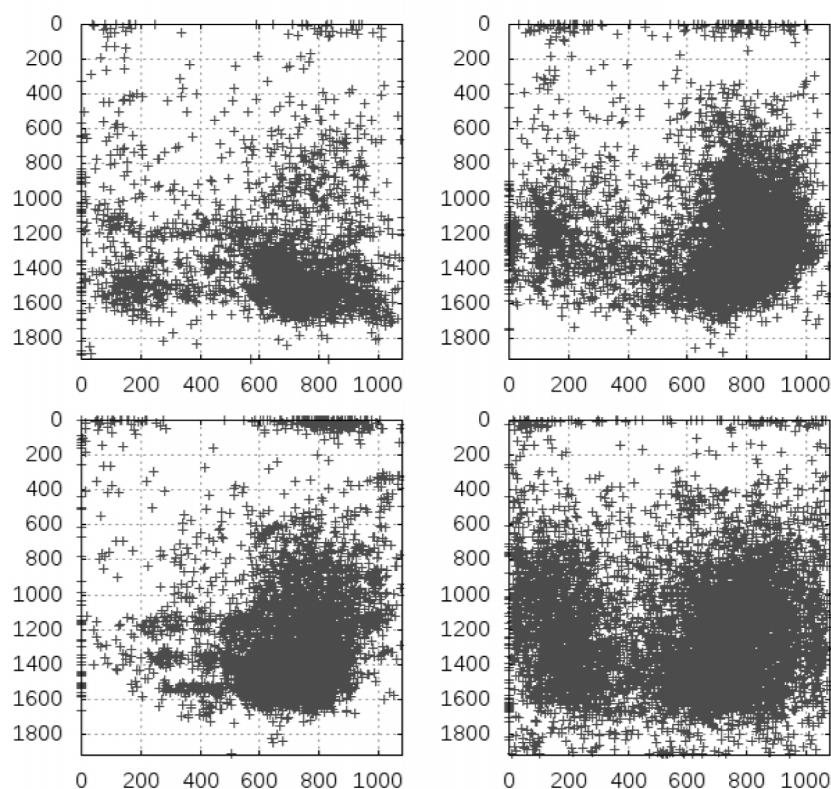


Рис. 6. Распределение точек по координатам

На рис. 7 для четырёх различных пользователей показаны графики другой характеристики, для которой можно построить эталон подобным образом. На этом графике по оси X отмечена продолжительность свайпа в мс, а по оси Y — его длина в пикселях. Таким образом, одна точка на графике некоторым образом отражает скорость одного свайпа.

Перечисленные выше способы анализа данных представляют большой интерес и обязательно будут исследованы в следующих работах.

3. Построение системы идентификации

Итак, в предыдущих пунктах были рассмотрены различные характеристики, извлечённые из тестовых данных. Самыми многообещающими оказались распределения свайпов по продолжительности и длине, поэтому построим две системы идентификации по каждой из этих характеристик и сравним процент успешных идентификаций для этих систем.

Для формирования эталонов выберем для каждого пользователя первые по времени поступления свайпы в количестве 1500 штук. Для полученных выборок построим распределения по продолжительности свайпов и по их длине. Таким образом, для каждого пользователя будут сформированы эталоны по двум признакам. Теперь необходимо построить массив тестовых данных. Для

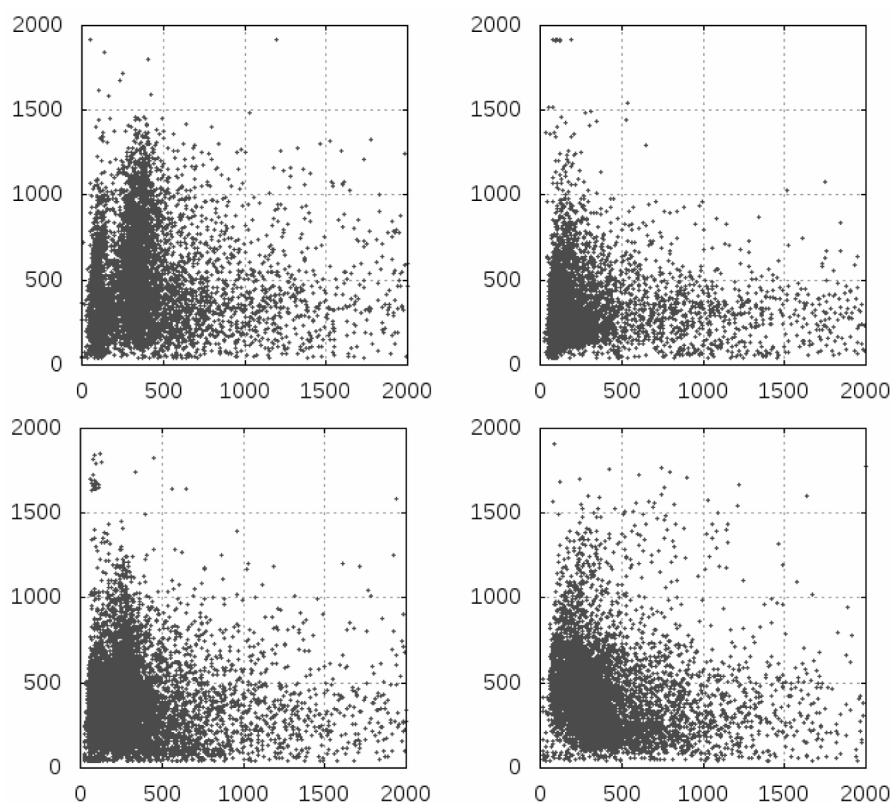


Рис. 7. Распределение свайпов по скорости

этого все оставшиеся события ввода для каждого из пользователей разобьём на выборки по 200 свайпов. Построим распределения нужных характеристик для каждой из полученных выборок. Получим набор тестовых данных, которые затем будем сравнивать с эталонами.

Распределение зависимости количества свайпов от некоторого параметра можно представить в виде вектора. Рассмотрим это на примере. Пусть у пользователя было взято 1500 свайпов. Посмотрим, как много свайпов имеют длину от 0 до 199 пикселей, от 200 до 399 пикселей и так далее и запишем результат в таблицу 1. Количество свайпов, находящееся в определённых пределах, можно нормировать, разделив на общее количество свайпов. Эти данные представлены в третьей строке таблицы 1.

Теперь в третьей строке записан в точности вектор, характеризующий распределение свайпов по длине в пикселях. Однако разные смартфоны обладают экранами различного разрешения, и это отличие также необходимо устранить при идентификации пользователей. Для этого нормирование проводилось не только по количеству свайпов, но и по их длине в пикселях. Длина каждого свайпа делилась на максимально возможное значение этого параметра ($\sqrt{w^2 + h^2}$, где w — ширина экрана в пикселях, h — высота экрана в пикселях), поскольку разрешения экранов смартфонов, с которых приходили тестовые данные, также были известны. Ещё одно отличие примера выше от построенной

Таблица 1. Таблица распределения свайпов по длине

Длина свайпа, px	от 0 до 199	от 200 до 399	от 400 до 599	от 600 до 799	от 800 до 999	от 1000 до 1199
Кол-во свайпов данной длины	453	581	252	131	70	13
Доля свайпов данной длины	0,3	0,39	0,17	0,09	0,05	0,01

системы идентификации заключается в размере векторов. В примере выше шаг составляет 200 пикселей, что даёт вектор из 6 элементов. В реальной работе при построении вектора распределения свайпов по длине шаг составлял 0,02, соответственно, эталонные и тестовые векторы состояли из 50 элементов. Векторы распределения свайпов по продолжительности имели шаг в 5 миллисекунд, что при исследовании свайпов продолжительностью до 2 секунд дало векторы размером в 400 элементов. Также для каждого вектора, будь то тестовый или эталонный, известен идентификатор того пользователя, чьи данные были использованы для построения этого вектора. Это необходимо для тестирования системы идентификации, так как позволяет понять, принадлежат ли два разных вектора одному и тому же пользователю.

Тестирование системы идентификации

После формирования эталонов и тестовых векторов запускается алгоритм идентификации пользователей. Алгоритм для каждого вектора из массива тестовых данных находит наиболее близкий эталонный вектор. Если тестовый вектор и ближайший к нему эталон принадлежат одному и тому же пользователю, идентификация признаётся успешной, иначе считается, что произошла ошибка второго рода. В конце вычисляется процент успешных идентификаций для каждого из исследуемых параметров.

В итоге, были получены следующие результаты:

1. Количество успешных идентификаций по распределению длины свайпов: 51%
2. Количество успешных идентификаций по распределению продолжительности свайпов: 93%

Это неплохие результаты для динамической биометрической характеристики. Однако текущая версия алгоритма не может отличить зарегистрированных пользователей от незарегистрированных, то есть не учитывает того, что в системе могут находиться тестовые данные пользователей, для которых не существует эталонов. Для того, чтобы это исправить, необходимо ввести некоторое пороговое значение и модифицировать алгоритм таким образом, чтобы он регистрировал ошибки первого и второго рода.

Пороговым значением в данном случае будет максимально допустимое расстояние между тестовым и эталонным векторами. Если наименьшее расстояние между тестовым вектором и всеми эталонами превысит пороговое значение, пользователь, чьи данные сформировали тестовый вектор, будет признан незарегистрированным в системе. Если же на самом деле для данного пользователя существует эталонный вектор, будет сгенерирована ошибка первого рода. Если минимальное расстояние не превысит порогового значения, считается, что пользователь в системе зарегистрирован и алгоритм попытается идентифицировать пользователя в системе так же, как это было описано выше, до введения в алгоритм порога.

При тестировании модифицированного алгоритма для формирования тестовых данных были использованы в том числе данные пользователей, которые не смогли отправить достаточное количество жестов для формирования эталонов и поэтому в предыдущей версии алгоритма не рассматривались. Это позволило сформировать дополнительные тестовые данные заведомо незарегистрированных в системе пользователей. Модифицированный алгоритм запускался только для распределения свайпов по продолжительности в миллисекундах, так как эта характеристика показала лучшие результаты в предыдущем тесте. Пороговый коэффициент был выбран таким образом, чтобы на тестовой базе данных количество ошибок первого и второго рода совпадало. В результате при выбранном пороговом коэффициенте 1,35 количество успешных идентификаций составило 76% от общего количества тестовых данных, количество ошибок первого и второго рода, соответственно, по 12%.

Снижение эффективности алгоритма при введении в систему незарегистрированных пользователей показывает необходимость поиска более совершенных биометрических характеристик в событиях ввода, чему будут посвящены следующие работы.

Заключение

В результате проделанной работы было реализовано клиент-серверное приложение для сбора событий ввода с сенсорного экрана и отправки этих событий на сервер. Сформирована тестовая база из более чем 200000 жестов, собранных с 19 пользователей. Проведён анализ большого числа характеристик данных жестов на предмет их уникальности среди пользователей. Выявлены характеристики, на основе которых может быть построена система биометрической идентификации. Реализована и протестирована система биометрической идентификации на основе распределения свайпов по длине и по продолжительности. Получены удовлетворительные результаты тестирования.

ЛИТЕРАТУРА

1. Майер Р. Android 4. Программирование приложений для планшетных компьютеров и смартфонов. Эксмо, 2013. 816 с.

2. Yaghmour K. Embedded Android. Porting, Extending and Customizing. O'reilly, 2013. 412 p.
3. Angulo J., Wästlund E. Exploring Touch-screen Biometrics for User Identification on Smart Phones [Электронный ресурс] URL: <http://dl.ifip.org/db/conf/primelife/primelife2011/AnguloW11.pdf> (дата обращения: 02.11.2015)
4. Шарун И.В., Лавров Д.Н. Проектирование и разработка программного обеспечения для моделирования идентификации по геометрии руки // Математическое и компьютерное моделирование [Электронный ресурс] : материалы III Международной научной конференции (Омск, 12 ноября 2015 г.) / [отв. за вып. И. П. Бесценный]. – Электрон. текст. дан. – Омск : Изд-во Ом. гос. ун-та, 2015. 1 электрон. опт. диск (CD-ROM); 12 см. С. 151-152.
5. Лазаренко П.Н., Лавров Д.Н. Оценка алгоритмов скелетизации в рамках исследований по сегментации кровеносных сосудов // Математическое и компьютерное моделирование [Электронный ресурс] : материалы III Международной научной конференции (Омск, 12 ноября 2015 г.) / [отв. за вып. И. П. Бесценный]. Электрон. текст. дан. Омск : Изд-во Ом. гос. ун-та, 2015. 1 электрон. опт. диск (CD-ROM); 12 см. С. 153–155.
6. Гришенкова Н.П., Лавров Д.Н. Обзор методов идентификации человека по радужной оболочке глаза // Математические структуры и моделирование. Омск : Ом. гос. ун-т, 2014. №1(29). С. 43–64.
7. А.В. Альтергот, О.А. Вишнякова, Лавров Д.Н. Формат данных на основе XML-разметки для обмена биометрическими образцами // Математические структуры и моделирование. Омск : Омский государственный университет, 2013. Т. 28, № 2. С. 60–67.
8. Альтергот А.В., Вишнякова О.А., Долгополов В.П., Лавров Д.Н. Спецификатор бинарных данных для XML-формата обмена биометрическими данными (тезисы) печатная Математическое и компьютерное моделирование : сборник материалов научной конференции (Омск, 18 октября 2013 г.). Омск : Изд-во Ом. гос. ун-та, 2013. С. 49–52.
9. Вишнякова О.А., Казанцева А.Г., Лавров Д.Н. Проектирование мультибиометрической базы данных для компьютерного моделирования распознавания личности // Вестник Омского университета. 2013. N.4. С. 207–212.
10. Вишнякова О.А., Лавров Д.Н. Формат обмена данными в системе сбора и обработки биометрических образцов // Международная научно-практическая конференция «Информационные ресурсы в образовании». Нижневартовск : НГГУ, 17-19 апреля 2013. С. 146–148
11. Первушин Е.А., Лавров Д.Н. Алгоритм учёта влияния параметров речевого сигнала в задаче распознавания дикторов // Вестник Омского университета. 2011. N.4. С. 153–156.
12. Первушин Е.А., Лавров Д.Н. Алгоритм извлечения признаков речевого сигнала во временной области для задачи распознавания дикторов // Вестник Омского университета. 2011. №.2. С. 182–185.
13. Первушин Е.А., Лавров Д.Н. Использование нормированных кадров сигнала в задаче распознавания по походке // Математические структуры и моделирование. Омск : Омский государственный университет им. Ф.М. Достоевского. 2011. Вып.23. С. 38–42.
14. Казанцева А.Г., Лавров Д.Н. Распознавание личности по походке на основе

- wavelet-параметризации показаний акселерометров // Математические структуры и моделирование. Омск : Омский государственный университет им. Ф.М. Достоевского, 2011. Вып. 23. С. 31–37.
15. Вишнякова О.А., Лавров Д.Н. Подходы к задаче идентификации диктора (научная статья) (РИНЦ). печатная Математические структуры и моделирование: Под ред. А.К.Гуца. Омск : Омский государственный университет им. Ф.М.Достоевского, 2011. Вып. 22. С. 48–54.
 16. Первушин Е.А., Лавров Д.Н. Система идентификации диктора на основе выделения информативных участков речевого сигнала // Информационные технологии и автоматизация управления. Матер. II межвуз. науч.-практ. конф. ОмГТУ 06-09 апреля 2010 года. / Под общ. науч. ред. В.Н. Задорожного. Омск : Изд-во ОмГТУ, 2010. С. 188-189.
 17. О.А. Вишнякова, Лавров Д.Н. Идентификация речевого сигнала на базе вейвлет-преобразования // Информационные технологии и автоматизация управления: матер. межвуз. науч.-практ. конф. ОмГТУ 2-24 апреля 2009 года. / Под общ. науч. ред. В.Н. Задорожного. Омск : Изд-во ОмГТУ, 2009. С. 167-168.
 18. Венедиктова Е.В., Лавров Д.Н. Идентификация голоса по спектральным характеристикам с помощью линейного классификатора // Математические структуры и моделирование : Под ред. А.К.Гуца. Омск : ОмГУ, 2008. Вып. 18. С. 108–115

IDENTIFICATION OF PERSON BASED ON BIOMETRIC CHARACTERISTICS OF TOUCH SCREEN GESTURES

K.I. Leyfer

Student, e-mail: leyfer.kirill@gmail.com

D.N. Lavrov

Ph.D. (Eng.), Associate Professor, e-mail: dmitry.lavrov72@gmail.com

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

Abstract. The article presents a study of possible ways to identify a user by his touch gestures on the screen of smartphone.

Keywords: gesrutes, identification, screen, smartphone.

МОДИФИКАЦИЯ МЕТОДА КУТТЕРА-ДЖОРДАНА-БОССЕНА СКРЫТОГО ХРАНЕНИЯ ИНФОРМАЦИИ В ИЗОБРАЖЕНИЯХ ФОРМАТА JPEG

А.Е. Дизер

студент, e-mail: ahemoh@gmail.com

Е.С. Дизер

аспирант, e-mail: e.s.dizer@gmail.com

Т.М. Опарина

старший преподаватель, e-mail: oparina2007@yandex.ru

Омский государственный университет им. Ф.М. Достоевского

Аннотация. В данной статье представлена модификация метода Куттера-Джордана-Боссена, целью которой было повышение устойчивости стеганоконтейнеров к различным видам атак, увеличение количества скрываемой информации в графическом файле и корректность считывания скрытой информации.

Ключевые слова: стеганография, стеганоконтейнер, метод Куттера-Джордана-Боссена.

Введение

Стеганографическая система или стегосистема — совокупность средств и методов, которые используются для формирования скрытого канала передачи информации. Обобщённая модель такой системы представлена (см. рис. 1).

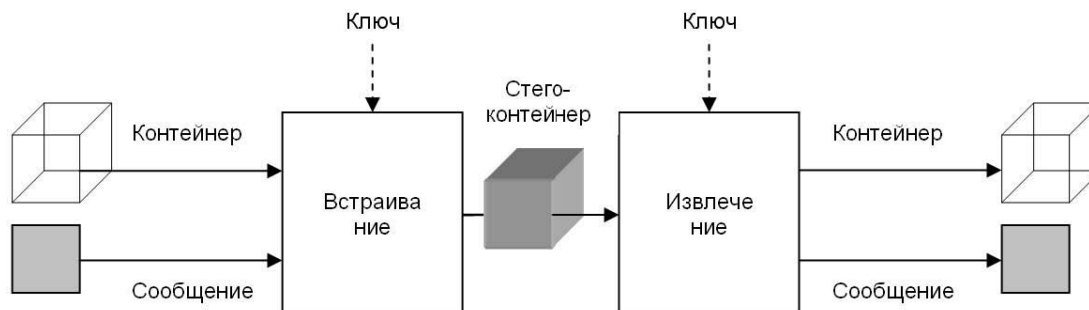


Рис. 1. Обобщенная модель стегосистемы

При построении стегосистемы должны учитываться следующие положения:

1. Противник имеет полное представление о стеганографической системе и деталях её реализации. Единственной информацией, которая остаётся неизвестной, является ключ, который позволяет установить факт наличия и получения скрытого сообщения.
2. Если противник каким-либо образом устанавливает факт наличия скрытого сообщения, то он не должен иметь возможность прочитать подобные сообщения из других данных до тех пор, пока ключ хранится в тайне.
3. Потенциальный противник должен быть лишён технических или каких-либо иных преимуществ в раскрытии содержимого тайных сообщений.

Далее будем использовать следующие понятия:

Контейнер — любая информация, которую можно использовать для сокрытия секретных сообщений. Пустой контейнер — контейнер без встроенного сообщения, а стегоконтейнер — со встроенной информацией.

Стеганографический канал или стегоканал — канал передачи скрытых сообщений.

Стегоключ — секретный ключ, который необходим для сокрытия секретного сообщения. Допускается использование нескольких стегоключей.

Согласно [1] все алгоритмы внедрения скрытой информации можно разделить на группы:

1. Алгоритмы, работающие с самим цифровым сигналом.
2. Алгоритмы, накладывающие скрытую информацию (изображения, звука, иногда текста) поверх оригинала.
3. Алгоритмы, связанные с особенностями форматов файлов.

Из группы методов сокрытия данных в графических файлах выберем и рассмотрим для реализации метод Куттера-Джордана-Боссена [4]. Для встраивания информации в контейнер используется одно из свойств зрительной системы человека, которое заключается в том, что восприимчивость человека к изменениям яркости синего цвета по сравнению с красным и зелёным — меньше всего, поэтому для встраивания скрытой информации методом Куттера-Джордана-Боссена используется синий цвет заданного контейнера-изображения.

Функции встраивания и извлечения в методе Куттера-Джордана-Боссена не симметричны. Следовательно, хотя правильное распознавание бита сообщения является высоковероятным, но не является стопроцентным, поэтому для уменьшения вероятности ошибок извлечения было предложено в процессе встраивания каждый бит повторять несколько раз (многократное встраивание). Следует отметить, что увеличение количества повторений существенно уменьшает вероятность возникновения ошибок при извлечении скрытого сообщения, но также это негативно отразится на вместимости стегоконтейнера.

1. Анализ существующих модификаций метода Куттера-Джордана-Боссена

В работах [2] и [3] были предложены модификации метода, которые направлены на преодоление проблем переполнения на этапе встраивания бита данных [2] и уменьшение вносимых в контейнер изменений [3]. Однако остаётся актуальным ряд недостатков, один из которых – это вероятностный характер чтения внедрённой информации из графического изображения. Это связано с тем, что встраивание бита данных происходит путём изменения синей составляющей цвета. При этом есть большая вероятность того, что изменённой яркости будет недостаточно для того, чтобы верно считать данный бит. Такая ситуация может возникнуть при встраивании бита «1» в тёмный пиксель на светлом фоне, либо встраивании «0» в светлый пиксель на тёмном фоне. При последующем чтении данного бита в большинстве случаев результат не будет корректным.

Если же попытаться изменить яркость синей составляющей настолько, чтобы при чтении результат был правильным, то это приведёт к сильному изменению цвета пикселя, и он будет выделяться на общем фоне (см. рис. 2).

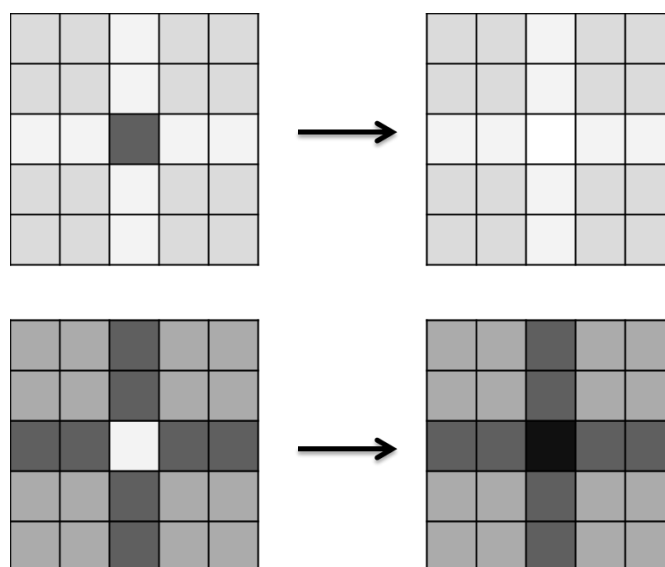
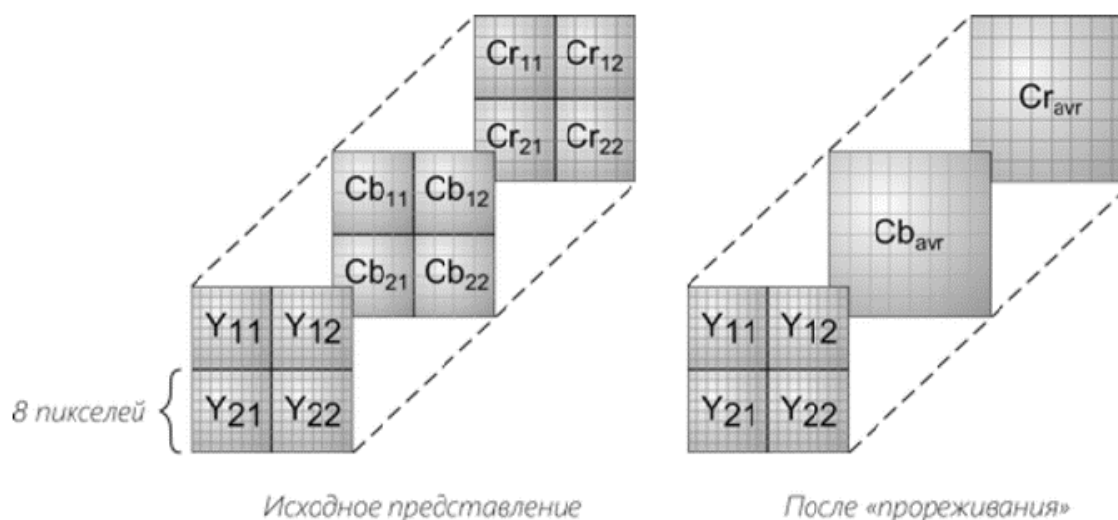


Рис. 2. Сильное изменение яркости пикселя с целью корректности считывания скрытой информации

Второй проблемой изменения только синей составляющей цвета является алгоритм сжатия JPEG (см. рис. 3). При преобразовании RGB $\rightarrow$ YCbCr для каналов Cb и Cr, которые отвечают за цвет, может выполняться прореживание [5]. Например, каждому из 4 блоков яркостного канала Y ставятся в соответствие средние значения каналов Cb и Cr [6].

Рис. 3. Преобразование RGB $\rightarrow$ YCbCr

2. Модификация метода Куттера-Джордана-Боссена на основе трёх составляющих цвета

Для того чтобы внедрённая информация была более устойчива к JPEG сжатию, а также сами изменения оставались менее заметными для восприятия, в данной работе предлагается использовать все три составляющие цвета, которые образуют Y :

$$Y_{x,y}^* = 0,299R_{x,y}^* + 0,587G_{x,y}^* + 0,114B_{x,y}^*.$$

Обратимся к первому требованию к стеганосистеме, описанному выше: изменения должны быть не заметны при визуальном контроле. Существует эмпирический закон Вебера-Фехнера, который позволяет определить степень заметности изменения яркости для человеческого глаза:

$$p = \ln \frac{Y + \varepsilon}{Y},$$

где p — контрастность, Y — яркость пикселя, ε — величина изменения яркости. В соответствии с законом Вебера-Фехнера, пороговая контрастность — это постоянная величина, которая равна 0,02–0,05 в зависимости от условий наблюдения. Пороговая контрастность при изменении яркости на $\varepsilon = 1$ и $\varepsilon = 2$ каждого канала пикселя в отдельности и яркости всего пикселя в целом показана на (см. рис. 4).

В модификации метода пиксель изменяется таким образом, чтобы его яркость отличалась от средней яркости его фона в пределах этого закона. От-

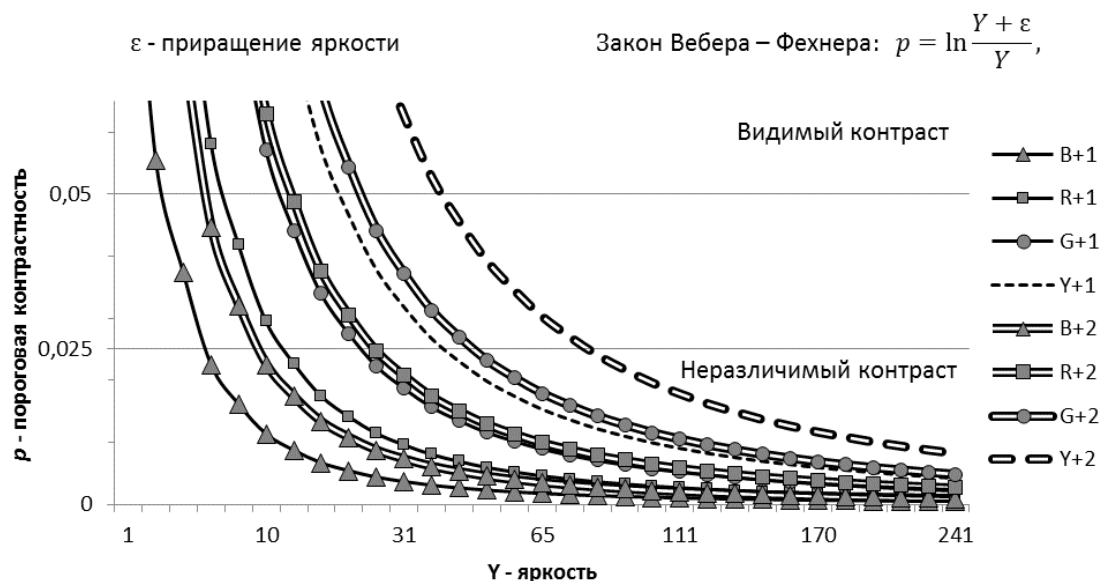


Рис. 4. Зависимость изменения видимости от приращения яркости

дельно стоит рассмотреть случай, когда пиксель сильно контрастирует со своим фоном, к примеру, в области контуров. То есть разница между яркостью изменённого пикселя и его фона превышает некоторое значение Δ :

$$|Y - Y_{cp}| > \Delta.$$

Модификация яркости такого пикселя приводит к сильной его заметности, следовательно, такие области необходимо пропускать. Второе требование к стеганосистеме — это устойчивость к искажениям. Алгоритм JPEG использует сжатие с потерями. Даже если оригинальное изображение без изменений сохранить алгоритмом JPEG, произойдёт потеря информации в виде незначительного изменения оттенка пикселей. Чтобы избежать потери скрытых битов, в модификации метода Куттера-Джордана-Боссена яркость изменённого пикселя должна отличаться от фона не менее чем на некоторую величину δ . Таким образом, яркость изменённого пикселя должна удовлетворять условию:

$$\begin{cases} |Y - Y_{cp}| > \Delta, \\ \left| k * \ln \frac{Y + \varepsilon}{Y_{cp}} \right| < p. \end{cases}$$

В связи с этим необходимо дополнить данные условия проверкой (см. рис. 5):

Если яркость изменённого пикселя $\Delta + \delta < |Y - Y_{cp}| < \Delta + 2\delta$, то данный пиксель лежит в пограничной области между контрастным и подходящим для встраивания, поэтому яркость данного пикселя необходимо изменить так, чтобы она удовлетворяла условию $|Y - Y_{cp}| > \Delta + 2\delta$ и данный пиксель однозначно

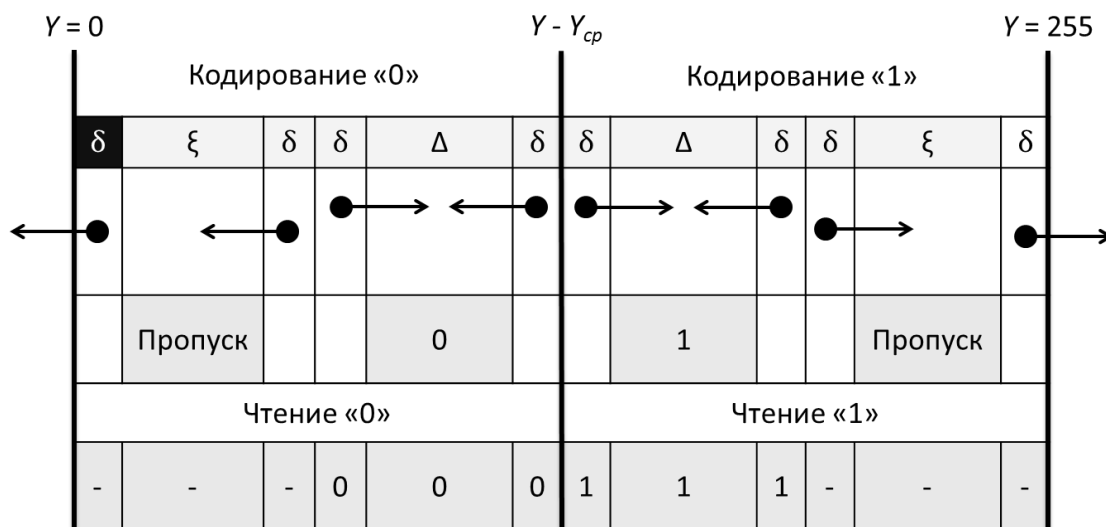


Рис. 5. Кодирование и чтение бита информации

трактовался как слишком контрастный.

При приближении яркости пикселя к границам чёрного ($Y = 0$) и белого ($Y = 255$) цвета значение яркости данного пикселя явно устанавливается как 0 или 255 соответственно. Это позволяет устранить возможные погрешности при работе алгоритма. Таким образом, значение разницы между яркостью пикселя и фона будет лежать либо в области Δ и означать успешную запись бита, либо в области ξ и означать, что данный пиксель не подходит для записи информации. При чтении область контрастных пикселей расширяется и задаётся как $|Y - Y_{cp}| > \Delta + 2\delta$, аналогично область информационных пикселей расширяется до $|Y - Y_{cp}| < \Delta + 2\delta$, что позволяет учитывать возможное изменение яркости при работе алгоритма сжатия JPEG.

Заключение

В отличие от оригинального метода, данная модификация позволяет достичь практически стопроцентной корректности считывания всех битов за счёт использования всех трёх составляющих его цвета, а также исключения из расчётов пикселей, имеющих большую разность яркости по сравнению с окрестностью. К тому же модификация позволяет расширить размер стеганоконтейнера за счёт уменьшения избыточности вносимой информации и увеличить стойкость к JPEG сжатию. Помимо этого модифицируемый алгоритм сохраняет преимущества оригинального метода Куттера-Джордана-Боссена, а именно повышенная устойчивость к различным видам атак, таким как удаление информации из наименее значимых битов, зашумление и кадрирование, а также сохраняет высокую устойчивость к частотному детектированию.

ЛИТЕРАТУРА

1. Васина Т.С. Обзор современных алгоритмов стеганографии // Наука и образование, 2012. № 4. URL: <http://technomag.bmstu.ru/doc/370605.html> (дата обращения: 20.09.2015).
2. Защелкин К.В. Усовершенствование метода скрытия данных Куттера-Джордана-Боссена // МНПК «Современные информационные и электронные технологии», 2013. С. 214–216.
3. Фомин Д.В. Модификация метода скрытия информации Куттера-Джордана-Боссена // Вестник Амурского государственного университета, 2014. Вып. 65 : Сер. Естеств. и экон. науки. С. 58–62.
4. Kutter M., Jordan F., Bossen F. Digital Signature of color image using amplitude modulation // Proc. of the SPIE Storage and Retrieval for Image and Video Databases V. 1997. Vol. 3022. P. 518–526.
5. Digital compression and coding of continuous-tone still images. Requirements and guidelines // ITU-T Recommendation T. 81, 1992. 28 p.
6. Douglas A. Kerr. Chrominance Subsampling in Digital Images. 2012. Issue 3. URL: <http://dougkerr.net/Pumpkin/articles/Subsampling.pdf> (дата обращения 10.10.2015).

**THE MODIFICATION OF THE KUTTER-JORDAN-BOSSSEN METHOD
OF THE LATENT INFORMATION STORAGE IN JPEG IMAGES****Alexander Dieser**Student, e-mail: ahemoh@gmail.com**Katherina Dieser**Postgraduate Student, e-mail: e.s.dizer@gmail.com**T.M. Oparina**Senior Tutor, e-mail: oparina2007@yandex.ru

Dostoevsky Omsk State University

Abstract. This article presents a modification of the Kutter-Jordan-Bossen method aimed at improving the sustainability of stego files to different types of attacks, an increased amount of hidden information in graphics file and the correct reading of the hidden information.

Keywords: steganography, stego file, Kutter-Jordan-Bossen method.

DDOS-АТАКИ КАК ДИФФЕРЕНЦИАЛЬНАЯ ИГРА

Т.В. Вахний

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

А.К. Гуц

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

Омский государственный университет им. Ф.М. Достоевского

Аннотация. Для нахождения возможных равновесных ситуаций при DDoS-атаках на компьютерные системы предлагается использовать теорию дифференциальных игр.

Ключевые слова: DDoS-атаки, защита компьютерной системы, дифференциальная игра, равновесие Нэша.

Введение

DDoS-атаки являются распространённым способом нанесения ущерба компьютерным системам. Как известно, DDoS-атака предполагает истощение злоумышленниками ресурсов, например, Web-сервера или канала связи путём коллективной отправки со своих компьютеров бессмысленных вредоносных запросов. При этом легитимные пользователи не могут получить доступ к предоставляемым системой ресурсам (серверам), либо такой доступ затруднён. Целью такой атаки является доведение компьютерной системы до отказа в обслуживании.

Полностью защититься от DDoS-атак невозможно, так как совершенно надёжных систем в настоящее время не существует. Здесь также большую роль играет человеческий фактор, потому что любая ошибка системного администратора, неправильно настроившего маршрутизатор, может привести к весьма плачевным последствиям. Однако, несмотря на все это, на настоящий момент существует масса как аппаратно-программных средств защиты, так и организационных методов противостояния [1, 2].

В данной работе DDoS-атаки рассматриваются как дифференциальная игра двух игроков — хакера и администратора, первый из которых управляет трафиком τ , а второй — производительностью p компьютерной системы. Устанавливается наличие особого типа оптимального управления (τ^*, p^*) , известного в теории игр под названием равновесие Нэша.

1. DDoS-атака как дифференциальная игра

В работе [3] DDoS-атаки описаны с помощью дифференциального уравнения

$$\frac{dx(t)}{dt} = [(p - p_0) - x^2(t)] \cdot x(t) + (\tau - \tau_0), \quad (1)$$

где:

$x(t)$ — число откликов в момент времени t компьютерной системы на внешние запросы, востребованные при обработке получаемых системой пакетов.

p — средняя скорости обработки входящих пакетов с учётом её падения или увеличения в зависимости от объёма занятых ресурсов: чем больше загружены ресурсы, тем меньше скорость обработки входящих пакетов. Другими словами, это производительность компьютерной системы, а p_0 — «типичная» характерная для данной системы величина производительности;

τ — величина входящего трафика; τ_0 — «типичная» характерная для системы «нормальная» величина трафика.

В уравнении (1) отражено требование, что увеличение трафика требует нарастания числа откликов на запросы.

Функционирующая компьютерная система способна справляться с ежедневным характерным трафиком τ_0 с определённым запасом надёжности системы. Один из способов DDOS-атаки состоит в том, чтобы добиться переполнения компьютерной системы с помощью такого большого количества пакетов, которое невозможно обработать. Очевидно, что при таком способе атаки наблюдается резкое возрастание входящего трафика. Увеличение трафика требует для его обработки увеличения свободных ресурсов системы.

Естественно поискать некоторое равновесие, которое может установиться при DDoS-атаках, если ресурсы хакера наращивать трафик не беспредельно, а компьютерная система имеет достаточно высокий уровень производительности.

Для отыскания такого равновесия (τ^*, p^*) воспользуемся теорией дифференциальных игр [2, 4], имея ввиду под равновесием равновесия Нэша.

2. Алгоритм нахождения равновесий Нэша

Естественно рассматривать игру с ненулевой суммой, поскольку выигрыши хакера и администратора слабо связаны.

Если игрок формирует «своё» управляющее воздействие в виде только функции времени $u(t)$ на всю продолжительность игры, то $u(t)$ — это *программное управление* игрока. Ранее мы называли его, используя термин «управление». Однако игрок может выбирать своё управление в зависимости от того, в каком положении x в момент времени t находится система. В таком случае игрок конструирует управляющее воздействие в виде функции $u(t, x)$, зависящей уже от позиции $\{t, x\}$, и для $u(t, x)$ используется термин *позиционное управление* игрока [5]. Часто пишут просто $u(x)$.

Мы будем искать позиционное управление, позиционное равновесие Нэша, когда хакер и администратор выбирают оптимально возможные доступные им управляющие параметры в каждой позиции $\{t, x\}$ в каждый момент времени.

Для дифференциальной игры N -игроков

$$\begin{aligned}\frac{dx}{dt} &= f(x) + \sum_{j=1}^N g_j(x)u_j, \quad f(0) = 0, \\ x &\in \mathbb{R}, \quad u_j \in \mathbb{R}, \\ J_i(x, u_1, \dots, u_N) &= \int_0^{+\infty} [Q_i(x) + \sum_{j=1}^N R_{ij}(u_j)^2] dt, \quad (i = 1, \dots, N), \\ Q_i &> 0, \quad R_{ii} > 0, \quad R_{ij} \geq 0,\end{aligned}$$

существование равновесий Нэша

$$J_i(u_1^*, u_2^*, u_i^*, \dots, u_N^*) \leq J_i(u_1^*, u_2^*, \dots, u_{i-1}^*, u_i, u_{i+1}^*, \dots, u_N^*), \quad \forall u_i, \quad i \in N, \quad (2)$$

сводится к крайне сложной задаче отыскания положительно определённого решения $V_i(x) > 0$ нелинейного уравнения Гамильтона-Якоби

$$\begin{aligned}(V_i)'_x(x)f(x) + Q_i(x) - \frac{1}{2}(V_i)'_x \sum_{j=1}^N [g_j(x)]^2 (R_{jj})^{-1} (V_j)'_x + \\ + \frac{1}{4} \sum_{j=1}^N R_{ij} [g_j(x)]^2 (R_{jj})^{-1} [(V_j)'_x]^2 = 0,\end{aligned} \quad (3)$$

по которому строится равновесие Нэша [4, Theorem 10.4-2]:

$$u_i^*(x) = u_i(V_i(x)) = -\frac{1}{2} R_{ii} g_i(x) (V_i)'_x, \quad i \in N. \quad (4)$$

Равновесие Нэша в данном случае означает, что если каждый игрок пытается в одностороннем порядке изменить свою стратегию управления, в то время как политика остальных игроков остаётся неизменной, то он имеет худший результат (большой проигрыш).

В нашем случае $N = 2$, игрок 1 — это администратор, игрок 2 — это хакер и

$$f(x) = -x^3, \quad g_1(x) = x, \quad g_2(x) = 1,$$

и при $R_{11} = R_{22} = 1, R_{12} = R_{21} = 0$ уравнения Гамильтона-Якоби имеют вид:

$$\begin{aligned}Q_1 + (V_1)'_x f(x) - \frac{1}{4} [g_1(x)]^2 [(V_1)'_x]^2 - \frac{1}{2} [g_2(x)]^2 (V_1)'_x (V_2)'_x &= 0, \\ Q_2 + (V_2)'_x f(x) - \frac{1}{4} [g_2(x)]^2 [(V_2)'_x]^2 - \frac{1}{2} [g_1(x)]^2 (V_1)'_x (V_2)'_x &= 0.\end{aligned} \quad (5)$$

Полагая, что

$$V_1(x) = V_2(x) = \frac{1}{2} x^2,$$

получаем уравнения Гамильтона-Якоби в виде

$$\begin{aligned} Q_1 - x^4 - \frac{1}{4}x^4 - \frac{1}{2}x^2 &= 0, \\ Q_2 - x^4 - \frac{1}{4}x^2 - \frac{1}{2}x^4 &= 0. \end{aligned} \quad (6)$$

Следовательно,

$$\begin{aligned} Q_1 &= \frac{5}{4}x^4 + \frac{1}{2}x^2 > 0, \\ Q_2 &= \frac{3}{2}x^4 + \frac{1}{4}x^2 > 0. \end{aligned} \quad (7)$$

(Эти функции положительно определённые). Поэтому по теореме 10.4-2 из [4] имеем равновесие Нэша

$$p^* = p_0 - \frac{1}{2}x^2, \quad \tau^* = \tau_0 - \frac{1}{2}x, \quad (8)$$

найденное по формулам (4). Выигрышные/проигрышные функции приобретают следующий вид:

$$\begin{aligned} J_1(x, p, \tau) &= \int_0^{+\infty} [Q_1(x) + (p - p_0)^2] dt, \\ J_2(x, p, \tau) &= \int_0^{+\infty} [Q_2(x) + (\tau - \tau_0)^2] dt. \end{aligned} \quad (9)$$

Если подставить (8) в (1) и проинтегрировать получаемое дифференциальное уравнение, то найдём оптимальное число откликов в момент времени t компьютерной системы на внешние запросы, востребованные при обработке получаемых системой пакетов:

$$x^2 = \frac{1}{Ce^t - 3},$$

где C — константа интегрирования.

3. Заключение

Мы показали, что между хакером и администратором возможно установление равновесия Нэша, идеология которого состоит в том, что каждая сторона считается с другой. Конечно, трудно надеяться, что хакер придерживается столь гуманной психологии, но учитывать, что его атаки, если они длительны или в значительной мере разрушительны, способствуют успешному проведению ответных мер со стороны жертвы, ему приходится.

Проведение дифференциальных игр и вычисление равновесий полезно с точки зрения определения степени надёжности компьютерной системы. Равновесия устанавливаются, если система способна сопротивляться. Если равновесий много, то в распоряжении администратора оказывается спектр порогов сопротивления, состоящих из пар (τ^*, p^*) , дающих критические значения трафика и соответствующих значений производительности системы.

ЛИТЕРАТУРА

1. Гуц А.К., Вахний Т.В. Теория игр и защита компьютерных систем: Учебное пособие. Омск : Изд-во ОмГУ, 2013. 160 с.
2. Гуц А.К., Володченкова Л.А. Дифференциальные игры в экологии человека и в социологии // Математические структуры и моделирование. 2016. № 3(39). С. 109–117.
3. Гуц А.К., Лавров Д.Н. Описание DDoS-атаки с помощью катастрофы «сборка» // Математические структуры и моделирование. 2013. № 1(27). С. 42–45.
4. Lewis F.L., Vrabie D.L., Syrmos V.L. Optimal Control. John Wiley & Sons, Inc., 2012. URL: <http://www.uta.edu/utari/acs/FL/%20talks/CDC/%20Orlando/%202011-%20online/%20synch/%20PI.pdf>
5. Тынянский Н.Т., Жуковский В.И. Дифференциальные игры с ненулевой суммой (кооперативный вариант) // Итоги науки и техн. Сер. Мат. анализ. 1979. Т. 17. С. 3–112.

DDOS-ATTACKS AS A DIFFERENTIAL GAME

T.V. Vahniy

Candidate of Mathematics, docent, e-mail: vahniytv@mail.ru

A.K. Guts

Doctor of Mathematics, Professor, e-mail: guts@omsu.ru

Dostoevsky Omsk State University

Abstract. Method of finding of the most optimal strategies for computer protection from DDoS-attacks is given based on differential game theory.

Keywords: DDoS-attacks, protection of computer system, differential game, Nash's balance.

Авторам

Редакция журнала «Математические структуры и моделирование» предлагает авторам ознакомиться с данными правилами и придерживаться их при подготовке рукописей, направляемых в журнал.

Общие положения

К публикации принимаются рукописи объемом не более 16 страниц.

Рукопись сопровождается краткой аннотацией на русском и английском языках (объемом от 100 до 250 слов).

Авторам необходимо предоставить следующую информацию на русском и английском языках:

- название статьи;
- список авторов с указанием
 - фамилии, имени и отчества,
 - учёного звания,
 - учёной степени,
 - должности,
 - места работы или учёбы,
 - действующего адреса электронной почты;
- аннотация (абстракт);
- список ключевых слов.

Автор также указывает УДК (универсальный десятичный код) статьи. Его можно подобрать по тематике статьи в справочнике <http://msm.univer.omsk.su/udc/>.

Рукопись статьи представляется в редакцию по электронной почте в двух форматах pdf и tex. Статья должна быть набрана с использованием макропакета LATEX2e и стиля msmb.cls, предоставляемого редакцией <http://msm.univer.omsk.su/files/msmb.zip>. Рекомендуется установить компилятор MiKTeX, так как именно им пользуются в редакции.

Отклонения в оформлении рукописи от приведённых правил позволяют редколлегии принять решение о снятии статьи с публикации. Статья может быть отклонена по причинам несоответствия тематике журнала или в связи с низким уровнем качества научного исследования.

Требования к оформлению

При подготовке статьи следует использовать класс (стиль) msmb.cls и шаблон-заготовку для текущего номера. В шаблоне приведены все наиболее типичные примеры оформления формул, рисунков, таблиц, разделов, библиографических ссылок.

В статье запрещается переопределять стандартные команды и окружения.

Нумеруемые формулы необходимо выделять в отдельную строку.

Нумерация только арабскими цифрами в порядке возрастания с единицы. Нумеровать следует только те формулы, на которые в тексте имеются ссылки.

Запрещается использовать в формулах буквы русского алфавита. Если без них никак не обойтись, то следует использовать команду `\mbox{...}`.

Все рисунки и таблицы должны иметь подпись, оформленную с помощью команды `\caption{...}`.

Файлы с рисунками необходимо представить в формате PDF или EPS (использовать редакторы векторной графики типа InkScape, Adobe Illustrator или Corel Draw).

Используйте стандартные команды переключения на готический, каллиграфический и ажурный шрифты: `\mathfrak`, `\mathcal` и `\mathbb`.

Не допускается заканчивать статью рисунком или таблицей.

В списке литературы обязательно указание следующих данных: для книг — фамилии и инициалы авторов, название книги, место издания, издательство, год издания, количество страниц; для статей — фамилии и инициалы авторов, название статьи, название журнала, год издания, том, номер (выпуск), страницы начала и конца статьи (для депонированных статей обязательно указать номер регистрации).

Кавычки в русском тексте («абвгд») должны быть угловыми, в английском прямыми верхними кавычками ("abcdefg" или "abcdefg").

Обязательна расшифровка сокращений при первом вхождении термина. Например: ... искусственный интеллект (ИИ)...

Порядок рецензирования

Первичная экспертиза проводится главным редактором (заместителем главного редактора). При первичной экспертизе оценивается соответствие статьи тематике журнала, правилам оформления и требованиям, установленным редакцией журнала к научным публикациям.

Все статьи, поступившие в редакцию научного журнала «Математические структуры и моделирование», проходят через институт рецензирования.

Рецензент выбирается главным редактором журнала из числа членов редколлегии или ведущих специалистов по профилю данной работы.

Рецензенты уведомляются о том, что присланные им рукописи являются частной собственностью авторов и относятся к сведениям, не подлежащим разглашению. Рецензентам не разрешается делать копии статей для своих нужд.

Срок для написания рецензии устанавливается по согласованию с рецензентом.

Рецензия должна раскрывать актуальность представленного материала, степень научной новизны исследования, определять соответствие предлагаемого к публикации текста общему профилю издания и стиль изложения.

Рецензент выносит заключение о возможности опубликования статьи: «рекомендуется», «рекомендуется с учётом исправления замечаний, отмеченных рецензентом» или «не рекомендуется». В случае отрицательной рецензии редакция направляет автору мотивированный отказ, заверенный главным редактором или его заместителем.

В случае несогласия с мнением рецензента автор статьи имеет право представить аргументированный ответ в редакцию журнала. Статья может быть направлена на повторное рецензирование, либо на согласование в редакционную коллегию.

При наличии в рецензии рекомендаций по исправлению и доработке статьи автору направляется текст рецензии с предложением учесть их при подготовке нового варианта статьи или аргументированно (частично или полностью) их опровергнуть. Доработанная (переработанная) автором статья повторно направляется на рецензирование и рассматривается в общем порядке. В этом случае датой поступления в редакцию считается дата возвращения доработанной статьи.

После принятия редколлегией решения о допуске статьи к публикации автор информируется об этом и указываются сроки публикации.

Оригиналы рецензий хранятся в редакции в течение пяти лет.

Памятка для перевода должностей, учёных степеней и званий на английский язык

Профессор = Professor

Доцент = Associate Professor

Старший преподаватель = Senior Tutor

Преподаватель = Tutor

Ассистент = Assistant

Аспирант = Postgraduate Student или Ph.D. Student

Соискатель = Ph.D. Doctoral Candidate

Студент = Student

д.ф.-м.н. = Dr.Sc. (Phys.-Math.)

к.ф.-м.н. = Ph.D. (Phys.-Math.)

д.т.н. = Dr.Sc. (Eng.)

к.т.н. = Ph.D. (Eng.)

Инженер-программист = Software Engineer

Старший/младший научный сотрудник = Senior/Junior Scientist Researcher

Электронная почта для отправки статей

lavrov@omsu.ru — зам. главного редактора, выпускающий редактор
Д.Н. Лавров.

Научный журнал

Математические структуры и моделирование

№3(39)

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 19.09.2016. Формат 60 × 84 1/8.

Усл. печ. л. 22,08. Тираж 100 экз. Заказ № 126.

Отпечатано на полиграфической базе издательства ОмГУ им. Ф.М. Достоевского
644077, г. Омск, пр. Мира, 55А

ISSN 2222-8772



9 772222 877005

16039 >

