МИНИСТЕРСТВО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
ОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
КАФЕДРА МАТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ
КАФЕДРА ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ

МАТЕМАТИЧЕСКИЕ СТРУКТУРЫ и МОДЕЛИРОВАНИЕ

Выпуск 8

Издание ОмГУ Омск 2001

УДК 007, 51, 53, 536, 512, 513, 517, 519, 530, 531, 532, 611, 612, 616, 681

Математические структуры и моделирование: Сб. научн. тр. / Под ред. А.К. Гуца. – Омск: Омск. гос. ун-т, 2001. – Вып. 8. – 178 с.

ISBN 5 – 8239 – 0078 – 3

Сборник составлен из статей преподавателей и аспирантов математического факультета Омского государственного университета и других вузов.

Для научных работников, аспирантов и студентов старших курсов.

Адрес научной редакции

Россия, 644077, Омск - 77, пр. Мира, 55 А
Омский государственный университет
математический факультет
Кафедра математического моделирования

E-mail: guts@univer.omsk.su
zhihalkina@math.omsu.omskreg.ru
lavrov@univer.omsk.su

ISBN 5 – 8239 – 0078 – 3

МАТЕМАТИЧЕСКИЕ СТРУКТУРЫ и МОДЕЛИРОВАНИЕ

В серии сборников публикуются статьи, в которых излагаются результаты исследований по фундаментальной и прикладной математике и размышления, касающиеся окружающей нас природы и общества.

Публикуются также статьи по философии и истории математики, по методике преподавания математики.

Объекты исследования должны быть представлены в форме некоторых математических структур и моделей.

Сборник является реферируемым. Рефераты статей публикуются в журналах "Zentralblat für Mathematik" (Германия) и "Mathematical Reviews" (США).

Электронная версия сборника представлена в сети «Интернет» по адресу:

`http://cmm.univer.omsk.su`

Сборник издается на коммерческие средства кафедры математического моделирования Омского государственного университета.

Кафедра готова к сотрудничеству в издании сборника. Наш E-mail:

`cmm@univer.omsk.su`

Подробную информацию можно найти на Web-сервере кафедры:

`http://cmm.univer.omsk.su`

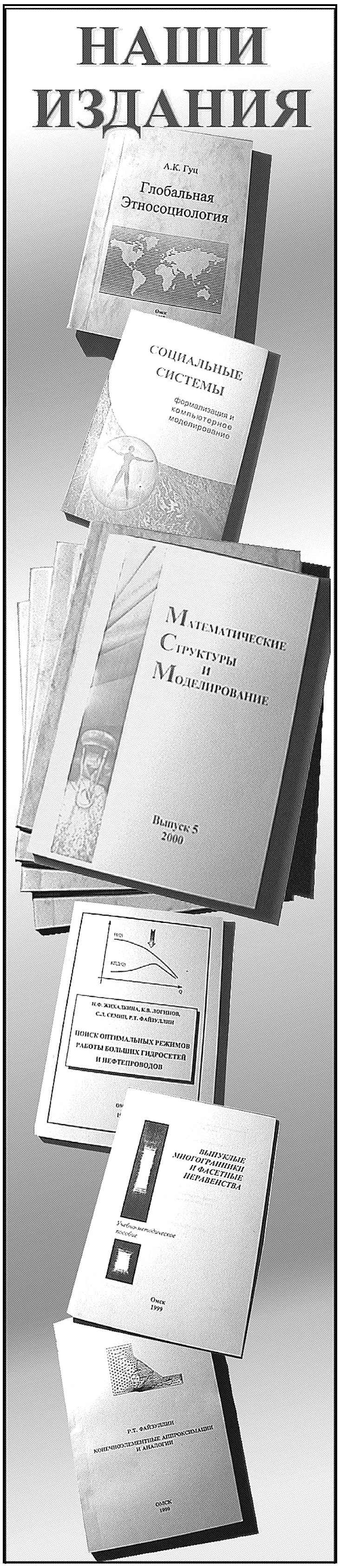
СОДЕРЖАНИЕ

Фундаментальная математика

Моделирование

Теоретическая физика

продолжение на след. странице



Математическая социология

Информационные технологии и искусственный интеллект

Экономическая математика

Преподавание математики

Математические
структуры и моделирование
2001, вып. 8, с. 5–9

УДК 513.813

ЛИНЕЙНАЯ СВЯЗНОСТЬ КАК КЛАСС ЭКВИВАЛЕНТНОСТИ ИЗОМОРФИЗМОВ КАСАТЕЛЬНЫХ ПРОСТРАНСТВ

Ю.Ф. Борисов

In this paper a definition of linear connection on smooth manifold is given.

В статье предлагается определение линейной связности как класса эквивалентности изоморфизмов касательных пространств.

Введение

Рассмотрим гладкое многообразие M^n. В каждой точке X имеется касательное векторное пространство T^n_X. Возьмем две различные точки $X_1 \neq X_2$ и соответствующие им касательные пространства $T^n_{X_1}$ и $T^n_{X_2}$. Известно, что они изоморфны как линейные векторные пространства. Можно ли как-нибудь сравнивать векторы, исходящие из разных точек, т.е. принадлежащие $T^n_{X_1}$ и $T^n_{X_2}$? Близкий вопрос: можно ли отождествить два вектора из разных касательных пространств?

Мы желаем найти способы, позволяющие в окрестности U каждой точки X устанавливать изоморфизм между касательными пространствами в точках, принадлежащих этой окрестности, и касательным пространством в точке X_0, определенный дифференциально, т.е. с точностью до бесконечно малых. Иначе говоря, поскольку вектор – это смещение, т.е. класс эквивалентных путей, то речь идет о равенстве двух смещений в близких точках с точностью до бесконечно малых.

1. Определение линейной (аффинной) связности в точке гладкого многообразия

Пусть имеется гладкое многообразие M^n класса C^m $(m \geq 2)$ и $X_0 \in M^n$ – точка, U – координатная окрестность точки X_0. Допустим, что в окрестности U задано поле изоморфизмов f_X пространств T^n_X на $T^n_{X_0}$, где $X \in U$,

$$T^n_X \xrightarrow[f_X]{\text{на}} T^n_{X_0}.$$

E-mail: borisov@math.nsc.ru

Институт математики СО РАН, г. Новосибирск

Статья подготовлена к печати преподавателями кафедры математического моделирования на основе записей лекций Ю.Ф. Борисова, сделанных в 1968 г.

Иначе говоря, мы указываем правило, по которому вектор в точке X отождествляется с вектором из $T^n_{X_0}$. Наложим на поле изоморфизмом f_X следующие условия:

(i) $X = X_0$, то f_X – тождественный изоморфизм.

(ii) Если $\{x^i\}$ локальные координаты в U, то изоморфизм f_X может быть записан аналитически следующим образом: он записывается как преобразование, сопоставляющее локальным координатам вектора $\overrightarrow{\xi} \in T^n_X$ (в данной локальной карте) локальные координаты вектора $\overrightarrow{\xi_0} \in T^n_{X_0}$; $\overrightarrow{\xi_0} = f_X(\overrightarrow{\xi}) \in T^n_{X_0}$. Такое преобразование задается при фиксировании локальной карты несобственной матрицей n-го порядка, коэффициенты которой зависят от локальной карты, т.е. являются функциями n переменных; полагаем, что независимо от выбора локальной карты элементы указанной матрицы непрерывно дифференцируемы в точке $(x^1, ..., x^n)$, соответствующей точке X.

Совсем не очевидно существование таких полей в том смысле, что условия (i), (ii) справедливы для всех локальных карт. Поэтому покажем, что если (ii) выполнено в некоторой локальной карте, действующей в окрестности U, то оно выполняется в любой другой локальной карте, действующей в U.

Каждая локальная карта K, действующая в U, сопоставляет полю f_X, $X \in U$ матричное поле $\|a^i_j(x^1, \ldots, x^n)\|$, $(x^1, \ldots, x^n) \overset{K}{=} X$, задающее изоморфизм в K вида

$$f_X : \xi^i_0 = a^i_j(x^1, \ldots, x^n)\xi^j, \tag{1}$$

где ξ^j – координаты в K вектора $\overrightarrow{\xi} \in T^n_X$, а ξ^j_0 координаты в K вектора $\overrightarrow{\xi_0} = f_X(\overrightarrow{\xi}) \in T^n_{X_0}$. Допустим, что в U введена новая локальная карта K'. Пусть полю f_X в K' сопоставляется новое матричное поле $\|a^{i'}_{j'}(x^{1'}, \ldots, x^{n'})\|$. Тогда аналогично (1)

$$\xi^{i'}_0 = a^{i'}_{j'}(x^{1'}, \ldots, x^{n'})\xi^{j'}.$$

Имеем

$$\xi^{i'}_0 = \Big(\frac{\partial x^{i'}}{\partial x^\alpha}\Big)_{X_0}\xi^\alpha_0 = \Big(\frac{\partial x^{i'}}{\partial x^\alpha}\Big)_{X_0} a^\alpha_j(x^1, \ldots, x^n)\xi^j =$$

$$= \Big(\frac{\partial x^{i'}}{\partial x^\alpha}\Big)_{X_0}\Big(\frac{\partial x^j}{\partial x^{\beta'}}\Big)_X \xi^{\beta'} a^\alpha_j(x^1, \ldots, x^n) = a^{i'}_{\beta'}(x^{1'}, \ldots, x^{n'})\xi^{\beta'}.$$

Таким образом, ввиду произвольности $\xi^{\beta'}$ получаем, что справедлива формула

$$a^{i'}_{\beta'}(x^{1'}, \ldots, x^{n'}) = \Big(\frac{\partial x^{i'}}{\partial x^\alpha}\Big)_{X_0}\Big(\frac{\partial x^j}{\partial x^{\beta'}}\Big)_X a^\alpha_j(x^1, \ldots, x^n). \tag{2}$$

Из этой формулы видно, что в силу того, что наше многообразие принадлежит классу дифференцируемости C^m $(m \geq 2)$, непрерывная дифференцируемость $a^{i'}_{j'}$ доказана (в точке $(x^{1'}, \ldots, x^{n'}) \overset{K'}{=} X_0$). Тем самым независимость условия (ii) от выбора локальной карты в U установлена.

Пусть имеются два поля изоморфизмов – f_X, заданное в окрестности U точки X, и g_X, заданное в окрестности V той же точки X, удовлетворяющие (i) и (ii). Будем говорить, что поля f_X и g_X имеют в точке X_0 совпадение 1-го порядка, если для любой локальной карты в окрестности $W = U \bigcap V$ соответствующие этим изоморфизмам матричные поля $\|a^i_j\|$ и $\|b^i_j\|$ в точке X_0 совпадают вместе с 1-ми производными. Иначе говоря, поля совпадают в 1-м порядке, если

$$a^i_j = b^i_j = \delta_{ij}$$

и

$$\left(\frac{\partial a^i_j}{\partial x^k}\right)_{X_0} = \left(\frac{\partial b^i_j}{\partial x^k}\right)_{X_0}.$$

Очевидно, что это отношение совпадения, определенное на полях изоморфизмов, является отношением эквивалентности.

Определение 1. *Линейной связностью в точке* $X_0 \in M^n$ называется любой класс эквивалентности полей изоморфизмов, удовлетворяющих условиям (i) и (ii), относительно отношения совпадения в 1-м порядке.

Таким образом, линейная связность Γ в точке X_0 есть поле изоморфизмов касательных пространств (с точностью до эквивалентности относительно совпадения полей изоморфизмов в 1-м порядке).

2. Коэффициенты линейной связности

Теорема 1. *Каковы бы ни были система, состоящая из n^3 вещественных чисел Γ^i_{jk} $(i, j, k = 1, \dots, n)$, и локальная карта в окрестности U точки X_0, существует, и притом единственная, линейная связность Γ в точке X_0, обладающая свойством: если $A \in \Gamma$ – поле изоморфизмов, заданное на U, которому в данной локальной карте соответствует матричное поле $\|a^i_j(x^1, \dots, x^n)\|$, то имеет место равенство*

$$\left(\frac{\partial a^i_j}{\partial x^k}\right)_{X_0} = \Gamma^i_{jk}.$$

Числа Γ^i_{jk} называются координатами линейной связности в данной локальной карте, или символами Кристоффеля 2-го порядка.

Доказательство. Существование такого поля A тривиально, т.к. в данной локальной карте U матричное поле можно представить следующим образом

$$a^i_j(x^1, \dots, x^n) = \delta^i_j + \Gamma^i_{jk} \cdot (x^k - x^k_0). \tag{3}$$

Докажем единственность. Допустим, что существует связность Γ^*, удовлетворяющая тем же условиям. Покажем, что $\Gamma^* = \Gamma$. Так как Γ – класс эквивалентных объектов, то достаточно установить наличие общего элемента, т.е. что $\Gamma^* \cap \Gamma \neq \emptyset$. Возьмем элемент $B \in \Gamma^*$. Поле B задано в координатной окрестности V точки

X_0. Рассмотрим сужение $B_1 = B|_W$, где $W = V \cap U$ поля B. Имеем $B_1 \in \Gamma^*$. Покажем, что $B_1 \in \Gamma$. Заметим, что если $\|b^i_j(x^1, ..., x^n)\|$ матричное поле, задающее B_1 в карте K (исходной), то

$$\left(\frac{\partial b^i_j}{\partial x^k}\right)_{X_0} = \Gamma^i_{jk}.$$

Вместе с тем поле $A \in \Gamma$ задается в K матричным полем $\|a^i_j(x^1, ..., x^n)\|$ вида (3), для которого

$$\left(\frac{\partial a^i_j}{\partial x^k}\right)_{X_0} = \Gamma^i_{jk}.$$

Поэтому в K имеет место равенство

$$\left(\frac{\partial b^i_j}{\partial x^k}\right)_{X_0} = \left(\frac{\partial a^i_j}{\partial x^k}\right)_{X_0}. \tag{4}$$

Если покажем, что аналогичное равенство имеет место в любой карте K', действующей в W, то этим будет показана эквивалентность B_1 и A, т.е. будет установлено, что $B_1 \in \Gamma$. Вспомним формулу (2) предыдущего параграфа

$$a^{i'}_{j'} = \left(\frac{\partial x^{i'}}{\partial x^\alpha}\right)_{X_0} \left(\frac{\partial x^\beta}{\partial x^{j'}}\right)_X a^\alpha_\beta.$$

Имеем

$$\left(\frac{\partial a^{i'}_{j'}}{\partial x^{k'}}\right)_{X_0} = \left(\frac{\partial x^{i'}}{\partial x^\alpha}\right)_{X_0} \left[\left(\frac{\partial^2 x^\beta}{\partial x^{j'} \partial x^{k'}}\right)_{X_0} \left(a^\alpha_\beta\right)_{X_0} + \left(\frac{\partial x^\beta}{\partial x^{j'}}\right)_{X_0} \left(\frac{\partial a^\alpha_\beta}{\partial x^\gamma} \frac{\partial x^\gamma}{\partial x^{k'}}\right)_{X_0}\right].$$

Заметим, что

$$(a^\alpha_\beta)_{X_0} = \delta^\alpha_\beta.$$

Поэтому предыдущее равенство примет вид

$$\left(\frac{\partial a^{i'}_{j'}}{\partial x^{k'}}\right)_{X_0} = \left(\frac{\partial x^{i'}}{\partial x^\beta}\right)_{X_0} \left(\frac{\partial^2 x^\beta}{\partial x^{j'} \partial x^{k'}}\right)_{X_0} + \left(\frac{\partial x^{i'}}{\partial x^\alpha}\right)_{X_0} \left(\frac{\partial x^\beta}{\partial x^{j'}}\right)_{X_0} \left(\frac{\partial x^\gamma}{\partial x^{k'}}\right)_{X_0} \left(\frac{\partial a^\alpha_\beta}{\partial x^\gamma}\right)_{X_0} \tag{5}$$

Совершенно аналогично

$$\left(\frac{\partial b^{i'}_{j'}}{\partial x^{k'}}\right)_{X_0} = \left(\frac{\partial x^{i'}}{\partial x^\beta}\right)_{X_0} \left(\frac{\partial^2 x^\beta}{\partial x^{j'} \partial x^{k'}}\right)_{X_0} + \left(\frac{\partial x^{i'}}{\partial x^\alpha}\right)_{X_0} \left(\frac{\partial x^\beta}{\partial x^{j'}}\right)_{X_0} \left(\frac{\partial x^\gamma}{\partial x^{k'}}\right)_{X_0} \left(\frac{\partial a^\alpha_\beta}{\partial x^\gamma}\right)_{X_0}. \tag{6}$$

Следовательно, сравнивая (5), (6) с учетом (4), получаем

$$\left(\frac{\partial a^{i'}_{j'}}{\partial x^{k'}}\right)_{X_0} = \left(\frac{\partial b^{i'}_{j'}}{\partial x^{k'}}\right)_{X_0}.$$

Теорема доказана. ■

Доказанная теорема говорит нам, что если выбрана локальная карта, то мы можем установить взаимно однозначное соответствие между всевозможными линейными связностями в точке X_0 и n^3 вещественных чисел Γ^i_{jk}, получая при этом отображение множества линейных связностей в пространство $\mathbb{R}^{n^3}$.

Числа Γ^i_{jk}, как уже говорилось, называются координатами линейной связности в данной локальной карте, или коэффициентами Кристоффеля 2-го рода. Легко получить формулу преобразования координат линейной связности в зависимости от преобразования координат различных локальных карт. Эта формула следует из (5) (или (6)) и имеет вид

$$\Gamma^{i'}_{j'k'} = \frac{\partial x^{i'}}{\partial x^{\beta}} \frac{\partial^2 x^{\beta}}{\partial x^{j'} \partial x^{k'}} + \frac{\partial x^{i'}}{\partial x^{\alpha}} \frac{\partial x^{\beta}}{\partial x^{j'}} \frac{\partial x^{\gamma}}{\partial x^{k'}} \Gamma^{\alpha}_{\beta\gamma}.$$

Пусть Γ^i_{jk} коэффициенты линейной связности Γ в локальной карте K в U. Тогда для любого поля изоморфизмов из Γ соответствующее ему поле матриц имеет вид

$$\|\delta^i_j + \Gamma^i_{jk}(x^k - x^k_0) + \varepsilon^i_{jk}(x^k - x^k_0)\|,$$

где $\varepsilon^i_{jk} \to 0$ при $n \to \infty$. Нетрудно видеть, что это простое разложение элементов матрицы $\|a^i_j(x^1, \ldots, x^n)\|$ в ряд Тейлора. Иначе говоря, матрицы, задающие поле изоморфизмов, принадлежащих данной связности, отличаются от матриц вида

$$\|\delta^i_j + \Gamma^i_{jk}(x^k - x^k_0)\| \tag{7}$$

на величину, меньшую, чем $\max\limits_{1 \le k \le n} |x^k - x^k_0|$. При этом матрицы (7) однозначно находятся в любой локальной карте.

Математические
структуры и моделирование
2001, вып. 8, с. 10–14

УДК 512.817

BOUNDED SHORTENING IN COXETER COMPLEXES AND BUILDINGS

G.A. Noskov

We prove that the standard generating set for a Coxeter group of finite rank satisfies "falsification by a fellow traveller property" in the sense of W. Neumann and M. Shapiro. In particular this implies that the geodesic words in standard generators form a regular language. Similar property established for buildings and this implies rationality of growth series for certain groups acting on buildings.

Introduction

W. Neumann and M. Shapiro have shown that any geometrically finite hyperbolic group G has a generating set A so that the geodesic words in generators A form a regular language and the growth function is rational (Theorem 4.3 in [7]). This is done by using a criterion which essentially goes back to [4], namely, that any word in generators A which is not geodesic has a close neighbour which is shorter. In [7] this criterion is called "falsification by a fellow traveller" (FFT–property). Clearly this property can be defined for any graph with a graph metric. We prove that FFT property holds in the Cayley graph of a Coxeter group with respect to a standard generating set. We prove that the dual graph of any building satisfies FFT property. This implies that that the growth function of the groups acting simply transitively on the chambers of building is rational with respect to some generating set.

1. FFT–property

Definitions. Let Γ be a connected graph. A path joining the vertex x to the vertex y is a map $p : \{0, 1, 2, \ldots, n_p\} \to \text{Vert } \Gamma, n_p \in \mathbb{N}$, where $p(0) = x, p(y) = y$ and all subsequent pairs of vertices are incident. For convenience, we often consider p as an ultimately constant map from $\mathbb{N}$ to Vert Γ, making p stopped after the moment n_p. We say that the edge paths with the same extremities δ-fellow travel for $\delta \in \mathbb{N}$ if the distance $d(w(t), v(t))$ never exceeds δ. We say that Γ has the falsification by fellow traveller property (or FFT– property) if there is a δ such that for any non-geodesic edge path in Γ there exists a shorter path with the same value that δ-fellow travels

E-mail: noskov@private.omsk.su

Omsk Branch of Institute of Mathematics, and Mathematisches Institute der Heinrich-Heine-Universitat Duesseldorf

This research was supported by a DMV grant Gr 627–11 and ZFB 343 of Bielefeld University

it. Let G be a finitely generated group and A a finite set and $a \mapsto \bar{a}$ a map of A to a monoid generating set $\bar{A} \subset G$. As is usual, A^* denotes the free monoid on A and the natural projection $A^* \to G$ is denoted $w \mapsto \bar{w}$. Any subset L of A^* which surjects onto G is called a normal form for G. The Cayley graph $\mathcal{C}_A(G)$ is the directed graph with vertex set G and a directed edge from g to $g\bar{a}$ for each $g \in G$ and $a \in A$; we give this edge a label a. We require that $\bar{A} = \bar{A}^{-1}$.

The following proposition explains our interest to the FFT–property ([7], Prop. 4.2).

Proposition. *If A has the falsification by fellow traveller property then the growth function of G with respect to A is rational.*

2. Walls in Coxeter complexes

We recall some basic definitions about Coxeter systems. For more about them see [Hi] or [Bo].

Definitions. A pair (W, S) is called a Coxeter system (of finite type) if W is a group with a finite subset S such that W has the presentation

$$\langle s : s \in S | (ss')^{m_{ss'}} = 1 \text{ when } m_{ss'} < \infty \rangle$$

where $m_{ss'} \in \{1, 2, 3, \ldots, \infty\}$ is the order of ss', and $m_{ss'} = 1$ if and only if $s = s'$. Let $\mathcal{C} = \mathcal{C}(\mathcal{W})$ be a Cayley graph of a Coxeter group W with respect to the standard generating system S. We call the edge $w \xrightarrow{s} ws$ to be inverse to the edge $ws \xrightarrow{s} w$ and we will call the pair of mutually inverted edges by a combinatorial edge and denote it by $\{w, ws\}$. The group W acts on the left on $\mathcal{C}$ by isometries. For any involution $w \in W$ we define its wall H_w as the set of all edges, inverted by w. The edge path $e_1 e_2 \cdots e_n$ is said to cross the wall H if at least one of its edges belongs to H. Each wall H separates $\mathcal{C}$ into two connected components H^+, H^- which are full subgraphs and each edge path connecting the vertices from different components crosses H. Two walls H_u, H_v are parallel if the element uv is of infinite order.

We can now state the Parallel Wall Theorem.

Theorem 1. *There exists a constant $K > 0$ such that for each point vertex x of $\mathcal{C}$ and for each wall H distance at least K from x, there exists another wall H' parallel to H which separates x from H. Moreover this wall can be chosen parallel to H.*

Clearly, the Parallel Wall Theorem implies the Separating Wall Theorem stated below.

Theorem 2. *There exists a constant $K > 0$ such that for each point $x \in \mathcal{C}$ and any wall H distance at least K from x, there exists another wall H' which separates x from H. Moreover this wall can be chosen parallel to H.*

In [1] the dominance relation is defined on the set of roots of (W, S). The dominance can be translated to geometry of walls as follows. Each wall H

divides the Cayley graph into two halfspaces H^+, H^- – we choose as the positive halfspace H^+ those one which does not contained 1. Then the dominance relation on the set of walls is just the containment relation on the set of positive halfspaces. It is shown in [2] that both assertions above and moreover they are equivalent to the following Finiteness Theorem which is proven in [1].

Theorem 3. *For any finitely generated Coxeter group the set of maximal positive halfspaces (relative to the containment) is finite.*

3. Falsification in Coxeter groups

We recall the construction of a Coxeter complex Σ for a Coxeter system (W, S) [3], [5]. By a special subgroup of W we mean a subgroup $\langle T \rangle$, generated by a proper subset $T \subset S$. The vertices of Σ are in one one correspondence with the left cosets of maximal special subgroups. More generally, the k-simplices of Σ are the left cosets of the special subgroups of rank $|S|-1-k$. In particular, the top-dimensional simplices (=chambers) are the cosets of the trivial subgroup of W, that is the elements of W. The codimension one simplexes (=panels) are the cosets of cyclic special subgroups. The incidence relation between the simplices is given by the containment relation between cosets. For example

$$v \langle s \rangle \text{ is the panel of } w \Longleftrightarrow w = vs \text{ or } w = v,$$

thus $v \langle s \rangle$ is the panel of exactly two chambers: v and vs. Define the dual graph of Σ with the set of all chambers W as vertices and panels as the edges – the ends of the edge are the chambers, adjacent along the corresponding panel. Thus the chambers v, w are adjacent iff they have the panel $u \langle s \rangle$ in common that is $w = vs$ or $v = ws$. We conclude that $\mathcal{C}$ is the modified Cayley graph of W with respect to generating system S [5]. The modification consists of the identification the edge $w \xrightarrow{s} ws$ with its inverse $ws \xrightarrow{s} w$. W acts simplicially on the Coxeter complex and this induces the standard action on the Cayley graph.

The edge paths in the Cayley graph are in the one one correspondence with the nonstuttering galleries in the Coxeter complex. The notion of the wall in the Cayley graph, being translated into the Coxeter complex, means the standard notion of the wall there – namely replacing the edges of Cayley wall by the corresponding panels we get the Coxeter wall.

Theorem 4. *The standard generating set S of any Coxeter group (W, S) satisfies the falsification by a fellow traveller property.*

Proof. We have to prove that any gallery $\Gamma = C_1 C_2 \cdots C_n$ in $\mathcal{C}$ which is not geodesic has a uniformly closed neighbour which is shorter. Take a subgallery $\Gamma' = C_i \cdots C_j$ which is not geodesic but any proper subgallery of which is already geodesic. It is well known fact that in a Coxeter complex the gallery is geodesic iff it crosses each wall at most twice, see e. g. [3]. Hence there is a wall H which is crossed by the subgallery Γ' at least twice. Indeed Γ' crosses H exactly twice since if some proper subgallery of Γ' would cross H twice then it would not be geodesic, contrary to the choice of Γ'. It follows that the chambers C_i and C_j lie on the same side of H, say H^-, and the subgallery Γ'' of Γ' obtained by deleting C_i and C_j lies on the another side, say H^+.

Γ' *lies in a* K*– neighbourhood of* H *where* K *is a constant in a Separating Wall Theorem.*

Suppose not, then there is a chamber $C \in \mathcal{C}$ at a distance greater than K from H. By the Parallel Wall Theorem there is the wall H_1 separating C from H. In particular, H_1 is contained in H^+. Let H_1^+ be those halfspace of H_1 which is contained in H^+. Then $C \subset H_1^+$ and $C_{i+1}, C_{j-1} \subset H_1^-$. Hence Γ'' have to cross H' at least twice and thus is not geodesic - contradiction.

Let $H = H_w$ for some reflection $w \in W$ and consider the gallery $w\Gamma''$. Clearly, it has the same origin and the end as Γ' does. But it is shorter than Γ' since it does not contain C_i, C_j. We assert that $w\Gamma''$ δ–fellow travels Γ' for $\delta = 2K + 2$. This immediately follows from the fact that $d(x, wx) \leq 2K$.

4. Falsification in buildings

Given a building Δ, there is a metric on the set of chambers of Δ, and we will want a path metric space which reflects this metric [6]. To do this we let Δ' be the graph dual to Δ. That is to say, the vertices of Δ' are the barycenters of the chambers of Δ. Two such vertices are connected by an edge when they lie in chambers with a common face. As usual, Δ' is metrized considering each edge as isometric to the unit interval. Non-stuttering galleries of Δ correspond to edge paths in Δ'. The decomposition of Δ into apartments induces a decomposition of Δ' into apartments which are isometric as labelled graphs to the Cayley graph of (W, S), the Coxeter system of Δ.

Theorem 5. *The dual graph of any locally finite building satisfies FFT property.*

Proof. Thus we have to prove that any nonstuttering gallery $\Gamma = C_1C_2\cdots C_n$ in $\mathcal{C}$ which is not geodesic has a uniformly closed neighbour which is shorter. Take a subgallery $\Gamma' = C_iC_{i+1}\cdots C_j$ which is not geodesic but any proper subgallery of which is already geodesic. In particular $\Gamma'' = C_iC_{i+1}\cdots C_{j-1}$ is geodesic. Let Σ be an apartment, containing both C_i and C_{j-1}. Any apartment Σ in any building Δ is convex in a sense that any geodesic gallery in Δ with both extremities in Σ is entirely contained in Σ [3], IY,4. In particular Γ'' is entirely contained in Σ. We let $\rho_{\Sigma,C}$ be the canonical retraction onto Σ centered at chamber C. (See, for example [Brown, IV.3].) It can be characterized as the unique chamber map $\Delta \to \Sigma$ which fixes C pointwise and maps every apartment containing C isomorphically onto Σ. We use the fact that ρ does not increase distance. Take $\rho = \rho_{\Sigma,C_{j-1}}$. Note that C_j is not folded by ρ onto C_{j-1} since these two chambers are contained in some apartment which mapped by ρ isomorphically onto Σ. Hence the length of the gallery $\rho(\Gamma) = C_iC_{i+1}\cdots C_{j-1}\rho(C_j)$ is the same as that of Γ and clearly Γ and $\rho(\Gamma)$ 1 -fellow travel each other. Thus it is enough to prove that $\rho(\Gamma)$ can be boundedly shortened. But this is proven above for the Coxeter complex.

Corollary. *Suppose Δ is a building whose apartments are isomorphic to the Coxeter complex Σ of a Coxeter system and that G is a finitely generated group which acts simplicially and simply transitively on the chambers of Σ. Let A be a generating set of G consisting of elements moving the fixed base vertex of X distance one apart relative to a graph metric on a 1-skeleton of Δ. Then A satisfies the falsification by fellow traveller property. In particular, the set of A–geodesic words forms a regular language and the growth function of G with respect to A is rational.*

Литература

1. Brink B., Howlett R.B. *A finiteness property and an automatic structure for Coxeter groups* // Math. Ann. 1993. V.296, N.2. P.179–190.
2. Brink B., Howlett R. *Parallel Wall Theorem* – Preprint, University of Sydney, 1998.
3. Brown K. S. *Buildings.* Graduate texts in mathematics. Springer-Verlag, 1989.
4. Cannon J.W. *The combinatorial structure of cocompact discrete hyperbolic groups* // Geom. Dedicata. 1984. V.16, N.2. P.123-148.
5. Cooper D., Long D. D., Reid A. W. *Infinite Coxeter groups are virtually indicable* // Proc. Edinb. Math. Soc., II. 1998. V.41, N.2. P.303-313.
6. Cartwright D.I., Shapiro M. *Hyperbolic buildings, affine buildings, and automatic groups* // Michigan Math. J. 1995. V.42, N.3. P.511-523.
7. Neumann W.D., Shapiro M. *Automatic structures, rational growth, and geometrically finite hyperbolic groups* // Invent. Math. 1995. V.120, N.2. P.259-287.

Математические
структуры и моделирование
2001, вып. 8, с. 15–17

УДК 513.813

ЛОРЕНЦЕВА ФУНКЦИЯ РАССТОЯНИЯ НА ЦИЛИНДРИЧЕСКИХ МНОГООБРАЗИЯХ

А.Н. Романов

In this article we will study the behaviour of Lorentzian distance function on cylindrical spaces, and will consider the reasons, bringing about origin of infinite importances of this function.

Целью данной работы является изучение поведения лоренцевой функции расстояния на пространствах, являющихся цилиндрами. А именно будут рассмотрены два примера лоренцевых многообразий и проведено исследование того, как ведёт себя определённая на них лоренцева функция расстояния. Все определения и обозначения соответствуют принятым в работе [1].

Пример 1. Пусть (M, g)– пространство-время, где $M = \mathbb{R}^1 \times S^1$– цилиндр, а метрика задаётся равенством:

$$ds^2 = -d\theta dt + f(-dt^2 + d\theta^2)$$

$$\|g\| = \begin{pmatrix} g_{tt} & g_{\theta t} \\ g_{t\theta} & g_{\theta\theta} \end{pmatrix} = \begin{pmatrix} -f & -\frac{1}{2} \\ -\frac{1}{2} & f \end{pmatrix},$$

где $\theta \in [0, 2\pi)$ – угловая координата; t – «вертикальная» координата времени, а $f = f(t) \geq 0$ – гладкая функция, зависящая от одного аргумента (от времени t), со следующими свойствами: $f = 0$ при $t = 0$ и $f \to +\infty$ при $t \to \pm\infty$. Данное пространство-время является хронологическим, так как в нём нет замкнутых времениподобных кривых,однако оно не является причинным: кривая $\gamma(\theta) = (0, \theta)$ представляет собой замкнутую причинную (изотропную) кривую. Метрика имеет некоторую «особенность» при $t = 0$ – она-то и создаёт замкнутую изотропную кривую, то есть имеет место нарушение причинности. В областях $\{t < 0\}$ и $\{t > 0\}$ замкнутых причинных кривых нет.

Покажем, что данное пространство-время допускает бесконечные значения лоренцевой функции расстояния. Возьмём точки $a = (-1, 0)$ и $b = (1, 0)$ и посчитаем лоренцеву длину какой-нибудь причинной кривой $\gamma[t, \theta(t)]$, их соединяющей, учитывая, что касательный вектор этой кривой в базисе пространства T_xM имеет координаты: $\dot\gamma = [1, \dot\theta]$:

$$L(\gamma) = \int_{-1}^{1} \sqrt{-g(\dot\gamma, \dot\gamma)}dt = \int_{-1}^{1} \sqrt{-(g_{tt} + 2\dot\theta g_{t\theta} + \dot\theta^2 g_{\theta\theta})}dt =$$

E-mail: romanow@univer.omsk.su
Омский государственный университет

$$= \int_{-1}^{1} \sqrt{-(-f - \dot{\theta} + \dot{\theta}^2 f)} dt = \int_{-1}^{1} \sqrt{(-f)\dot{\theta}^2 + \dot{\theta} + f}\, dt$$

Рассматривая последнее подкоренное выражение $P = (-f)\dot{\theta}^2 + \dot{\theta} + f$ как квадратный трёхчлен от $\dot{\theta}$, можно найти его максимум:

$$P_{max} = f + \frac{1}{4f} \quad \text{при} \quad \dot{\theta} = \frac{1}{2f}$$

Заметим, что кривой γ_{ab} с касательным вектором $\dot{\gamma} = [1, \dot{\theta} = 1/2f]$ существовать не может, так как $f = 0$ при $t = 0$. Однако $\dot{\theta}$ может принимать значения, сколь угодно близкие к значению $\dot{\theta} = 1/2f$. Поэтому, учитывая определение лоренцева расстояния, получаем следующее:

$$d(a, b) = sup\ L(\gamma_{ab}) = \int_{-1}^{1} \sqrt{P_{max}} dt =$$

$$= \int_{-1}^{1} \sqrt{f + \frac{1}{4f}} dt \geq \int_{-1}^{1} \sqrt{0 + \frac{1}{4f}} dt = \frac{1}{2} \int_{-1}^{1} \frac{1}{\sqrt{f}} dt$$

Используя разложение функции f в ряд Тейлора в окрестности точки $t = 0$, покажем, что последний интеграл является расходящимся.

$$f(t) = f(0) + f'(0)t + f''(0)\frac{t^2}{2} + f'''(0)\frac{t^3}{6} + ...$$

Используя свойства функции f, получаем:

$$f(t) = f''(0)\frac{t^2}{2} + f'''(0)\frac{t^3}{6} + ...$$

Для того, чтобы показать,что интеграл $\int_{-1}^{1} 1/\sqrt{f} dt$ расходится, составим отношение:

$$\frac{t^\alpha}{\sqrt{f}} = \frac{1}{\sqrt{\frac{1}{2} f''(0) t^{2-2\alpha} + \frac{1}{6} f'''(0) t^{3-2\alpha} + ...}}$$

Для сходимости интеграла необходимо выполнение неравенства: $\alpha < 1$. Однако в этом случае при $t \to 0$ получаем:

$$\frac{t^\alpha}{\sqrt{f}} \longrightarrow \infty$$

Таким образом, интеграл $\int_{-1}^{1} 1/\sqrt{f} dt$ является расходящимся, и поэтому лоренцевы длины причинных кривых, идущих из a в b, неограниченны. Отсюда следует, что лоренцево расстояние между точками a и b равно бесконечности:

$$d(p, s) = \sup\ L(\gamma_{ps}) = \infty$$

■

Пример 2. Рассмотрим пространство-время (M, g), где

$$M = \mathbb{R} \times S^1 \times S^1 = \{(t, y, z) \in \mathbb{R}^3\}$$

$$(t, y, z) = (t, y, z + 1) \text{ и } (t, y, z) = (t, y + 1, z + a).$$

Число $a-$ иррациональное. Лоренцева метрика задаётся следующим образом:

$$ds^2 = -dtdy + (ch\, t - 1)^2(-dt^2 + dy^2) + dz^2$$

Данное пространство-время содержит захваченные причинные кривые, то есть имеет место явление захвата. В то же время оно является причинным, то есть в нём нет замкнутых причинных кривых (см. [2, с.217]). Однако поведение метрики вблизи области $\{t = 0\}$ приводит к возникновению бесконечных значений лоренцевой функции расстояния.

Для доказательства этого возьмём точки $a = (-1, 0, 0)$ и $b = (1, 0, 0)$ и рассмотрим причинные кривые, лежащие в гиперповерхности $\{z = 0\}$. Так как $z = 0$, то для вычисления длин таких причинных кривых используется уже рассмотренная ранее метрика:

$$ds^2 = -dtdy + f(-dt^2 + dy^2)$$

Здесь функция $f = (ch\, t - 1)^2\,(\geq 0)$ обладает всеми свойствами, описанными в предыдущем примере: $f = 0$ при $t = 0$ и $f \to +\infty$ при $t \to \pm\infty$.

Таким образом, справедлив предыдущий результат: лоренцевы длины причинных кривых, идущих из a в b, неограниченны и лоренцево расстояние между точками a и b равно бесконечности:

$$d(p, s) = \sup\; L(\gamma_{ps}) = \infty$$

Данное пространство-время не является устойчиво причинным, то есть малые изменения метрики могут приводить к замкнутым причинным кривым.

■

Используя технику построения лоренцевых многообразий, основанную на применении искривлённого лоренцева произведения (см. [1]), подобные пространства, допускающие бесконечные лоренцевы расстояния, можно получать достаточно легко. Во всех них причиной бесконечности расстояния будет как раз особый вид функции f и, соответственно, особое поведение метрики в определённых областях пространства-времени.

Литература

1. Бим Дж., Эрлих П. *Глобальная лоренцева геометрия.* М.: Мир, 1985.
2. Хокинг С., Эллис Дж. *Крупномасштабная структура пространства-времени.* М.: Мир, 1977.

Математические структуры и моделирование
2001, вып. 8, с. 18–21

УДК 517.382:517.58

О НЕКОТОРЫХ ИНТЕГРАЛАХ ТЕОРИИ ПОЛУПРОВОДНИКОВ

С.Д. Симонженков

A method for numerical evaluation of some integrals appearing in semiconductor theory by means of the confluent hypergeometric functions is described.

Речь идет о приближенном вычислении интегралов

$$A_n(p,x) = \frac{x^n}{\Gamma(p+1)} \int\limits_0^\infty e^{-t} t^p (t^n + x^n)^{-1}\, dt, \tag{1}$$

$$B_n(p,x) = \frac{x^{2n}}{\Gamma(p+1)} \int\limits_0^\infty e^{-t} t^p (t^n + x^n)^{-2}\, dt, \tag{2}$$

где $p > -1, x > 0, n = 1, 2, \ldots$. Такие интегралы часто встречаются в анализе и приложениях. Например, при $n = 1, 2, 3$ они использовались и табулировались в задачах теории полупроводников [1–3]. Для $n = 1, 2$ известны их аналитические представления в виде некоторых специальных функций (см., например, [4], формулы $2.3.6.9 - 15, 2.3.7.8 - 13$), однако для практических вычислений такие представления не всегда приемлемы ввиду громоздкости и разнородности.

С помощью интегрирования по частям нетрудно проверить, что

$$B_n(p,x) = \left(1 - \frac{p+1}{n}\right) A_n(p,x) + \frac{p+1}{n} A_n(p+1,x),$$

поэтому достаточно уметь находить интегралы (1). В данной работе их предлагается вычислять на основе функции

$$I(p,z) = \frac{z}{\Gamma(p+1)} \int\limits_0^\infty \frac{e^{-t} t^p}{t+z}\, dt, \ \ |\arg z| < \pi$$

следующим образом. При $x = const$ разложим дробь $x^n/(t^n + x^n)$ на сумму простейших дробей вида $a_k/(t + z_k)$ с некоторыми комплексными a_k, z_k. Окажется, что будет иметь место равенство

$$\frac{x^n}{t^n + x^n} = \frac{1}{n} \sum_{k=0}^{n-1} \frac{z_k}{t + z_k}, \ \ z_k = x \exp\left(\frac{2k+1-n}{n} \pi i\right),$$

Омский государственный педагогический университет

поэтому

$$A_n(p, x) = \frac{1}{n} \sum_{k=0}^{n-1} I(p, z_k). \tag{3}$$

Так как интегралы $I(p, z), I(p, \bar{z})$ комплексно сопряженные, то в этой сумме на самом деле участвуют лишь действительные части слагаемых:

$$A_n(p, x) = \frac{1}{n} \sum_{k=0}^{n-1} ReI(p, z_k).$$

Способ вычисления интегралов $I(p, z_k)$ зависит от параметров x, p, n. При $p = 0, 1, 2, ...$ имеет место равенство

$$I(p, z) = ze^z E_{p+1}(z),$$

где $E_{p+1}(z)$ – стандартные интегральные показательные функции, методы вычисления которых достаточно хорошо разработаны. Известна, например, техника Гаучи [5], описание которой можно найти в справочнике Люка [6, с.106-107]. Поэтому в данной работе рассматривается случай, когда p не является целым числом.

Пусть $p \neq 0, 1, 2, ...$ Предлагаются представления

$$I(p, z) = z^{p+1} U(p + 1; p + 1; z), \tag{4}$$

$$I(p, z) = \frac{z}{p}[F(1; 1 - p; z) - z^p e^z \Gamma(1 - p)], \tag{5}$$

где F, U – вырожденные гипергеометрические функции соответственно первого и второго рода. Равенство (4) выгодно использовать при больших x, когда точки z_k далеки от разреза комплексной плоскости вдоль отрицательной вещественной полуоси. В реальных приложениях интегралов (1), (2) параметр n обычно невелик (не более 4), поэтому основную роль играет x. Мы предлагаем применять (4) при $x \geq 4$. Если $x < 4$, то выгоднее использовать (5), при этом в силу соотношения

$$I(p + 1, z) = \frac{z}{p + 1}[1 - I(p, z)]$$

достаточно ограничиться областью $-1 < p < 0$.

Таким образом, вычисление интегралов (1), (2) сводится к вычислению гипергеометрических функций $F(a; c; z)$, $U(a; c; z)$. В расчетах на ЭВМ их нахождение удобно осуществлять с помощью соответствующих процедур. Чаще всего используются конечно-разностные методы, когда F, U рассматриваются как начальные значения минимальных решений $\{f_n\}$ соответствующих разностных уравнений вида

$$y_{n+1} + a_n y_n + b_n y_{n-1} = 0, \;\; b_n \neq 0 \;\forall n \geq 1. \tag{6}$$

Тогда при наличии дополнительного соотношения

$$\sum_{n \geq 0} c_n f_n = S \neq 0 \tag{7}$$

f_0 (а следовательно, и искомая функция) может быть найдено устойчивой обратной рекурсией на основе, например, алгоритма Миллера. Поэтому приведем соответствующие коэффициенты в (6) и (7).

Пусть a, c, z – комплексные числа, причем $a, c \neq 0, -1, -2, ..., z \neq 0$. Последовательность

$$f_n = \frac{(-z)^n (a)_n}{n!(c)_n} F(a+n; c+n; z), \ n \geq 0$$

образует минимальное решение уравнения (6) с коэффициентами

$$a_n = -\frac{n+c-1-z}{n+1}, \ b_n = -\frac{z(n+a-1)}{n(n+1)},$$

при этом в (7) $c_n = S = 1$. Здесь, как и ниже, $(a)_n$ означает символ Похгаммера.

Пусть теперь

$$a, a+1-c \neq 0, -1, -2, ...; z \neq 0, \ |\arg z| < \pi.$$

Тогда последовательность

$$f_n = z^a (a)_n (a+1-c)_n U(a+n; c; z)/n!$$

является минимальным решением уравнения (6) с коэффициентами

$$a_n = -\frac{2n+2a-c+z}{n+1}, b_n = \frac{n-1+a}{n}\frac{n+a-c}{n+1},$$

равенство (7) имеет место при $c_n = S = 1$.

В качестве примера рассмотрим вычисление интегралов

$$A_2(p, \frac{\pi}{2}t^2), \ p = \pm\frac{1}{2}, \ t = \frac{1}{2}, 1, \frac{5}{4}, \frac{10}{7}, \frac{5}{3}, 2.$$

В целях контроля используются равенства

$$A_2(-\frac{1}{2}, \frac{\pi}{2}t^2) = \pi t f(t), \ A_2(\frac{1}{2}, \frac{\pi}{2}t^2) = \pi t^3 g(t), \tag{8}$$

где

$$f(t) = \left[\frac{1}{2} - S(t)\right] \cos\frac{\pi}{2}t^2 - \left[\frac{1}{2} - C(t)\right] \sin\frac{\pi}{2}t^2,$$

$$g(t) = \left[\frac{1}{2} - C(t)\right] \cos\frac{\pi}{2}t^2 + \left[\frac{1}{2} - S(t)\right] \sin\frac{\pi}{2}t^2$$

– вспомогательные функции для вычисления интегрального косинуса и синуса $C(t)$, $S(t)$. Функции $f(t)$, $g(t)$ затабулированы, их значения брались из таблицы 7.8 справочника [7]. Результаты вычислений согласно (3)-(5) приведены ниже.

t	$x = \frac{\pi}{2}t^2$	$A_2(-\frac{1}{2}, x)$	$A_2(\frac{1}{2}, x)$
1/2	0,39269908	0,62707	0,21422
1	1,5707963	0,87931	0,60936
5/4	2,4543692	0,93027	0,77335
10/7	3,2057067	0,95209	0,81124
5/3	4,3633230	0,97020	0,87434
2	6,2831854	0,98385	0,92744

Значения гипергеометрических функций в (4),(5) вычислялись по алгоритму Миллера в нелинейной версии Гаучи с погрешностью 10^{-5}. Результаты, представленные в таблице, совпадают в пределах такой точности с контрольными согласно (8).

В заключение несколько замечаний по поводу используемых фактов. Равенство (4) — это следствие из интегрального представления для функции U; см., например, [7], 13.2.5. Равенство (5) вытекает из (4), связи между функциями F и U и того факта, что $F(a; a; z) = e^z$. О методах вычисления специальных функций на основе линейных рекуррентных соотношений см. [6, гл.12].

Литература

1. Dingle R., Doreen A., Roy K. *The integrals*

$$A_p(x) = (p!)^{-1}\int\limits_0^\infty \varepsilon^p(\varepsilon + x)^{-1}e^{-\varepsilon}\,d\varepsilon, B_p = (p!)^{-1}\int\limits_0^\infty \varepsilon^p(\varepsilon + x)^{-2}e^{-\varepsilon}\,d\varepsilon$$

and their tabulation // Appl. Sci. Res. 1956. V.6. №4. P.144-153.

2. Dingle R., Doreen A., Roy K. *The integrals*

$$C_p(x) = (p!)^{-1}\int\limits_0^\infty \varepsilon^p(\varepsilon^2 + x^2)^{-1}e^{-\varepsilon}\,d\varepsilon, D_p(x) = (p!)^{-1}\int\limits_0^\infty \varepsilon^p(\varepsilon^2 + x^2)^{-2}e^{-\varepsilon}\,d\varepsilon$$

and their tabulation // Appl. Sci. Res. 1956. V.6. №4. P.155-161.

3. Dingle R., Doreen A., Roy K. *The integrals*

$$E_p(x) = (p!)^{-1}\int\limits_0^\infty \varepsilon^p(1 + x\varepsilon^3)^{-1}e^{-\varepsilon}\,d\varepsilon, F_p(x) = (p!)^{-1}\int\limits_0^\infty \varepsilon^p(1 + x\varepsilon^3)^{-2}e^{-\varepsilon}\,d\varepsilon$$

and their tabulation // Appl. Sci. Res. 1956. V.6. №4. P.245-252.

4. Прудников А.П. и др. *Интегралы и ряды. Элементарные функции.* М.: Наука, 1981. 800 с.
5. Gautschi W. *Recursive computation of certain integrals* // J. Asoc. Comput. Mach. 1961. V.8. P.21-40.
6. Люк Ю. *Специальные математические функции и их аппроксимации.* М.: Мир, 1980. 608 с.
7. *Справочник по специальным функциям* / Под ред. М.Абрамовица и И.Стиган. М.: Наука, 1979. 830 с.

Математические структуры и моделирование
2001, вып. 8, с. 22–27

УДК 517.53:517.947.42

О СУЩЕСТВОВАНИИ ЭКСТРЕМАЛЬНЫХ КВАЗИКОНФОРМНЫХ В СРЕДНЕМ ОТОБРАЖЕНИЙ

Ю.Ф. Стругов

In this article proof existence of extremal mappings of circular domain that are quasi-conformal in the mean with free values on another component is presented.

Введение

Введем основные термины и определения. $D,\ D^*$ — ограниченные области в R^n; $x = (x_1, \ldots, x_n)$—произвольная точка в R_n; $dist(A, B)$ — евклидово расстояние между множествами A и B; $diam A$ — диаметр множества A; $m_n(A) = |A|$ — n-мерная мера Лебега множества A; $\Gamma(A, B; D)$ — семейство кривых, соединяющих множества A и B в области D; $M(\Gamma(A, B; D))$ — n-мерный модуль семейства кривых Γ [1, гл.2]; $f(x) = (f_1(x), f_2(x), \ldots, f_n(x))$ — гомеоморфное отображение области D на область D^*, $f^{-1}(y)$ — обратное отображение области D^* на область D.

В точках дифференцируемости отображения $f(x)$ определены величины: $\nabla f(x) = f'(x)$ – матрица Якоби отображения $f(x)$ в точке x; $J(x, f) = \det f'(x)$ – якобиан отображения $f(x)$;

$$|\nabla f(x)|^2 = \sum_{i,j=1}^{n} \left(\frac{\partial f_i}{\partial x_j}\right)^2.$$

Величина

$$H_I(x, f) = \frac{|J(x, f)|}{l(x, f)^n}$$

называется локальной внутренней характеристикой квазиконформности отображения $y = f(x)$ в точке x, где $J(x, f) \neq 0$; здесь обозначено

$$l(x, f) = \min_{|h|=1} |f'(x)h|.$$

Обозначим $J^i_j(x, f)$ – алгебраическое дополнение элемента $\frac{\partial f_i}{\partial x_j}$ определителя $J(x, f)$. Определим функционалы

$$\mathcal{H}(f) = \frac{1}{|D|}\int\limits_D H_I(x, f)dx + \frac{1}{|D^*|}\int\limits_{D^*} H_I(y, f^{-1})dy,$$

E-mail: strugov@univer.omsk.su
Омский государственный университет

$$\mathcal{F}(\nabla f; D, D^*) = a\int\limits_D |\nabla f(x)|^p dx + b\int\limits_D \left(\sum_{i,j=1}^{n}\left(\frac{J_j^i(x,f)}{J(x,f)}\right)^2\right)^{\frac{q}{2}} dx,$$

где $p \geqslant n$, $q \geqslant n$, и коэффициенты

$$a = \frac{|D|^{\frac{p-n}{n}}}{2n^{\frac{p}{n}}|D^*|^{\frac{p}{n}}}, \quad b = \frac{|D^*|^{\frac{q}{n}}}{2n^{\frac{q}{n}}|D|^{\frac{q+n}{n}}}.$$

Обозначим $\mathcal{M}_{p,q}(D, D*)$ класс всех гомеоморфизмов $f : D \to D*$, таких, что $f \in W_n^1(D)$, $f^{-1} \in W_n^1(D*)$, на которых функционал $\mathcal{F}(\nabla f) < \infty$. Класс всех гомеоморфных отображений, на которых $\mathcal{H}_I(f) < \infty$, обозначим $\mathcal{M}_0(D, D^*)$.

Отображения этого класса дифференцируемы почти всюду, прямые и обратные отображения обладают N-свойством, для них справедливы классические формулы замены переменных и правило дифференцирования сложных функций.

1. Теоремы существования

Теорема 1. *Пусть класс отображений $\mathcal{M}_{p,q}(D, D^*)$, $\frac{1}{p} + \frac{1}{q} \leq \frac{1}{n-1}$, не пуст. Тогда в нем существует отображение $f_0 : D \to D*$ такое, что*

$$\mathcal{F}(\nabla f_0) \leq \mathcal{F}(\nabla f)$$

для всех $f \in \mathcal{M}_{p,q}(D, D^*)$.

Доказательство. Пусть $f_k, k = 1, 2, \dots$, произвольная последовательность, минимизирующая функционал $\mathcal{F}(\nabla f)$ в классе $\mathcal{M}_{p,q}(D, D^*)$. Тогда последовательность $\mathcal{F}(\nabla f_k)$ ограничена сверху некоторой константой M_1. В силу теоремы 2 [2, с.167] для всех номеров m справедливы неравенства

$$\mathcal{H}_I(f_m) \leq (a^{-1}\mathcal{F}(\nabla f^m))^{\frac{n}{p}}\frac{|D|^{\frac{p-n}{n}}}{|D^*|} + |D|^{-1}a^{\frac{-q}{p+q}}b^{\frac{-p}{p+q}}\mathcal{F}(\nabla f_m) \leq M < \infty,$$

где $M = M(n, p, q, |D|, |D^*|, M_1)$ — некоторая постоянная величина. Известно, что из всякой бесконечной последовательности отображений $f_m : D \to D^*$ с равномерно ограниченным функционалом $\mathcal{H}_I(f_m)$ можно извлечь равномерно внутри D сходящуюся подпоследовательность. При этом предельное отображение либо постоянно, либо является гомеоморфизмом области D на область D^* ([3, следствие 2.3.9, с.71]), принадлежащим классу $\mathcal{M}_0(D, D^*)$. Выберем из минимизирующей последовательности сходящуюся подпоследовательность. Если эта минимизирующая подпоследовательность равномерно в D сходится к гомеоморфизму $f_0 \in \mathcal{M}_0(D, D^*)$, то по теореме 3 [2, с.174]

$$\mathcal{F}(\nabla f_0) \leq \lim_{m\to\infty} \mathcal{F}(\nabla f_m),$$

и теорема в этом случае доказана.

Осталось показать, что минимизирующая последовательность не может сходиться к постоянному отображению. Действительно, пусть $f_m \to c$ при $m \to \infty$.

Так как последовательность отображений ограничена в пространстве $W_n^1(D)$, то ([4, с.94]) для любой непрерывной функции φ финитной в области D

$$\int\limits_D \varphi(x) J(x, f_m) dx \to 0$$

при $m \to \infty$. Пусть $B^n(a,r)$ – шар, замыкание которого содержится в D. Выберем финитную функцию φ такой, что $0 \leq \varphi \leq 1$, $\varphi(x) = 1$ для всех $x \in B^n(a,r)$. Тогда, применяя неравенство Гельдера, получим

$$0 < |B^n(f,r)| < \int\limits_D \varphi(x) dx \leq \left(\int\limits_D \frac{\varphi(x) dx}{J(x, f_m)^{\frac{q}{n}}} \right)^{\frac{n}{n+q}} \left(\int\limits_D \varphi(x) J(x, f_m) dx \right)^{\frac{q}{n+q}}.$$

Из этого неравенства и равенства $\lim_{m\to\infty} \int\limits_D \varphi(x) J(x, f_m) dx = 0$ следует, что

$$\lim_{m\to\infty} \int_D \frac{\varphi(x) dx}{J(x, f_m)^{\frac{q}{n}}} = +\infty.$$

Далее, применяя неравенство (5) из теоремы 2 ([2, с.167]), найдем

$$\int\limits_D \frac{\varphi(x) dx}{J(x, f_m)^{\frac{q}{n}}} < \int\limits_D \frac{dx}{J(x, f_m)}^{\frac{q}{n}} \leq \int\limits_D \frac{H_I(x, f_m)^{\frac{q}{n}}}{J(x, f_m)^{\frac{q}{n}}} \leq b_{-1} \mathcal{F}(\nabla f_m).$$

Таким образом, $\mathcal{F}(\nabla f_m) \to \infty$, а это противоречит тому, что f_m– минимизирующая последовательность. Следовательно, предельное отображение не может быть постоянным, и тем самым теорема доказана. ■

Пусть γ – некоторый континуум из области D, а γ^* – континуум из области D^*. Ниже будем предполагать, что в области D^* найдется континуум γ^*, для которого класс отображений $\mathcal{M}_{p,q}(D\backslash\gamma, D^*\backslash\gamma^*) \neq \emptyset$. Зафиксируем области D, D^* и континуум γ. Докажем, что существует континуум γ^* и гомеоморфное отображение $f_0 : D \to D^*$ такое, что

$$\mathcal{F}(\nabla f_0; D\backslash\gamma, D^*\backslash\gamma^*) = \inf_{\gamma^*} \mathcal{F}(\nabla f; D\backslash\gamma, D^*\backslash\gamma^*).$$

Лемма 1. *Пусть $f_m : D\backslash\gamma \to D^*\backslash\gamma_m$ – минимизирующая последовательность. Тогда из нее можно извлечь подпоследовательность, которая сходится равномерно в области D к некоторому непрерывному отображению*

$$f : D\backslash\gamma \to R^n, f \in W_p^1(D\backslash\gamma).$$

Доказательство. [5, лемма 1, с.69] ■

Лемма 2. *Отображение f непостоянно.*

Доказательство. Допустим, что $f = c$. Пусть U и V окрестности континуума γ такие, что $\overline{U} \subset D$, $\overline{V} \subset U$. Тогда $\overline{U}\backslash V$ – компактное множество в области $D\backslash\gamma$, и на нем $f_m \to c$ при $m \to \infty$. Следовательно, $diam\gamma_m \to 0$ при $m \to \infty$, и поэтому меры Лебега $|D^*\backslash\gamma_m| \geq \delta > 0$ для всех номеров m. Отсюда следует, что коэффициенты a, b (стоящие перед интегралами в функционалах $\mathcal{F}(\nabla f_m; D\backslash\gamma, D^*\backslash\gamma_m)$ и зависящие от номеров m) равномерно ограничены сверху и отделены от нуля. Поэтому из равенства

$$\lim_{m\to\infty} \int\limits_{D\backslash\gamma} \frac{dx}{J(x, f_m)^{\frac{q}{n}}} = \infty,$$

также как и в доказательстве теоремы 1, вытекает неограниченность последовательности функционалов $\mathcal{F}(f_m; D\backslash\gamma, D_*\backslash\gamma_m)$. Полученное противоречие доказывает лемму. ■

Лемма 3. $\liminf_{m\to\infty} |D\backslash\gamma_m| = d > 0$.

Доказательство. Так как отображение f непостоянно, поэтому из полунепрерывности интегралов Дирихле будем иметь

$$\liminf_{m\to\infty} \int\limits_{D\backslash\gamma} |\nabla f_m|^p dx \geq \int\limits_{D\backslash\gamma} |\nabla f|^p dx > 0.$$

Отсюда и из ограниченности последовательности функционалов следует ограниченность сверху последовательности коэффициентов $a = a(m)$ перед интегралами Дирихле. Из ограниченности последовательности коэффициентов следует утверждение леммы. ■

Из топологии известно, что из ограниченной последовательности континуумов γ_m можно выбрать подпоследовательность, которая в метрике Хаусдорфа сходится к некоторому континууму γ^*.

Лемма 4. *Если $x \in D\backslash\gamma$, то $f(x) \in D^*\backslash\gamma^*$, причем $\gamma^* \subset D^*$ и $diam\ \gamma^* > 0$.*

Доказательство. Если $f(x) \in \partial D^*$, то по лемме 2.3.2 [3, с.64] отображение f постоянно, что противоречит лемме 2. Поэтому $f(x) \in D^*$. Если $\partial D^* \bigcap \gamma^{ast} \neq \emptyset$, то найдется $x \in D\backslash\gamma$ такое, что $f(x) \in \partial D^*$, и поэтому отображение постоянно. Полученное противоречие доказывает, что $\gamma^* \subset D^*$. Если $f(x) \in \gamma^*$ и $diam\gamma^* > 0$, то опять по лемме 2.3.2 из [3] получим, что отображение f постоянно. Если γ^* есть точка, то выберем две точки $a, b \in D\backslash\gamma$ такие, что $f(a) \neq f(b)$. Обозначим через E континуум, соединяющий в $D\backslash\gamma$ точку a и γ, F континуум, соединяющий точку b и γ, причем континуумы выбираем непересекающимися. Обозначим $\Gamma(E, F; D\backslash\gamma)$ семейство всевозможных кривых, соединяющих в области $D\backslash\gamma$ континуумы E и F. Пусть Γ^*_m— образ семейства Γ при отображении f_m. Тогда для модулей $M(\Gamma^*_m)$ справедливы оценки

$$M(\Gamma^*_m) \leq \int\limits_{D\backslash\gamma} \rho^n(x) H_I(x, f_m) dx \leq d(E, F)^{-n} \int\limits_{D\backslash\gamma} H_I(x, f_m) dx \leq M < \infty,$$

где $\rho(x) = d(E,F)^{-1}$– допустимая метрика для семейства кривых Γ ([6]). С другой стороны, $\lim_{m\to\infty} M(\Gamma_m^*) = \infty$ [1], так как $d(f_m(E), f_m(F)) \to 0$ при $m \to \infty$. Полученное противоречие доказывает невырожденность континуума γ^*. ■

Лемма 5. *Если* $x_1, x_2 \in D\backslash\gamma, x_1 \neq x_2$, *то* $f(x_1) \neq f(x_2)$.
Доказательство. [5, лемма 4, с.70]. ■

Лемма 6. *Пусть* $U \supset \gamma_*, \overline{U} \subset D^*$ *– произвольная открытая окрестность континуума* γ_*. *Тогда из последовательности обратных отображений* f_m^{-1} *можно выбрать подпоследовательность, которая равномерно в* $D^*\backslash\overline{U}$ *сходится к некоторому непрерывному отображению* $h : D^*\backslash\overline{U} \to R^n, h \in W_n^1(D^*\backslash\overline{U})$, *при этом а) отображение* h *непостоянно; б) если* $y \in D^*\backslash\overline{U}$, *то* $h(y)$*– внутренняя точка множества* $h(D^*\backslash\overline{U})$*; c) Если* $y_1, y_2 \in D^*\backslash\overline{U}, y_1 \neq y_2$, *то* $h(y_1) \neq h(y_2)$. ■

Доказательство. Семейство обратных отображений, также как и в лемме 1, равномерно ограниченно и локально равностепенно непрерывно. Поэтому из нее можно извлечь сходящуюся подпоследовательность к непрерывному отображению $h \in W_n^1$. Так как $h(f(x)) = x$ для любого x такого, что $f(x) \in D^*\backslash\overline{U}$, то отображение h непостоянно. Действительно, выражение $h(f(x)) - x = (h(f(x)) - f_m^{-1}(f(x))) + (f_m^{-1}(f(x)) - f_m^{-1}(f_m(x)))$, начиная с некоторого номера, определено и стремится к нулю. Утверждения б), с) доказываются аналогично леммам 4-5.■

Лемма 7. *Предельное отображение* f *есть отображение области* $D\backslash\gamma$ *на область* $D^*\backslash\gamma^*$.
Доказательство. [5, лемма 6]. ■

Лемма 8. *Отображение* f *является гомеоморфизмом области* $D\backslash\gamma$ *на область* $D^*\backslash\gamma^*$ *. Причем* $f \in W_p^1(D\backslash\gamma), f^{-1} \in W_n^1(D^*\backslash\gamma^*)$.
Доказательство. [5, лемма 7]. ■

Лемма 9. *Пусть* $f_m : D\backslash\gamma \to D^*\backslash\gamma_m^*$*– минимизирующая подпоследовательность из леммы 8. Тогда*

$$\mathcal{F}(\nabla f; D\backslash\gamma, D^*\backslash\gamma^*) \leq \liminf_{m\to\infty} \mathcal{F}(\nabla f_m; D\backslash\gamma, D^*\backslash\gamma_m^*).$$

Доказательство. [7, теорема, с.34], [2, теорема 3, с.174]. ■

Теорема 2. *Существуют континуум* $\gamma^* \subset D^*$ *и гомеоморфизм* $f_0 : D\backslash\gamma \to D^*\backslash\gamma^*$ *такие, что*

$$\mathcal{F}(\nabla f_0; D\backslash\gamma, D^*\backslash\gamma^*) = \inf_{\gamma^*} \mathcal{F}(\nabla f; D\backslash\gamma, D^*\backslash\gamma^*).$$

Доказательство. Из лемм 8,9, повторяя рассуждения доказательства теоремы [7], мы получим утверждение теоремы 2. ■

Литература

1. Сычев А.В. *Модули и пространственные квазиконформные отображения.* Новосибирск: Наука, 1983. С.152.
2. Стругов Ю.Ф. *О существовании гомеоморфных решений систем уравнений эллиптического типа* // Групповые и метрические свойства отображений. Межвузовский сборник научных трудов.Новосибирск, 1995. С.164–180.
3. Стругов Ю.Ф. *Квазиконформные в среднем отображения и экстремальные задачи. Ч.1.* – М.,1994.–153 с. Деп. в ВИНИТИ 05.12.94. N.2786 – В 94.
4. Решетняк Ю.Г. *Пространственные отображения с ограниченным искажением.* Новосибирск: Наука, 1982. с.279.
5. Стругов Ю.Ф., Гарифуллина Е.В. *О компактности семейств квазиконформных в среднем отображений со свободными значениями на границе* // Омский научный вестник. 1999. N.8. С.68-71.
6. Стругов Ю.Ф., Сычев А.В. *Различные классы пространственных отображений, квазиконформных в среднем* // Алгебра и математический анализ. Новосибирск, 1990. С.104-125.
7. Стругов Ю.Ф., Гарифуллина Е.В. *О существовании экстремального отображения кольцевой области со свободными значениями на одной граничной компоненте* // Омский научный вестник. 1999. N.9. С.34-37.

Математические структуры и моделирование
2001, вып. 8, с. 28–33

УДК 519.6

УСКОРЕНИЕ МЕТОДОВ РЕШЕНИЯ СИСТЕМ ЛИНЕЙНЫХ УРАВНЕНИЙ

А.В. Кондратьев

In this article some applications of accelleration methods for finite element problems is presented.

1. Методы решения систем линейных уравнений

В настоящее время известно много численных систем, в которых возникает необходимость решения систем линейных уравнений большой размерности. Возникает задача выбора наилучшего метода решения системы линейных уравнений. Нами была рассмотрена возможность использования ряда прямых (Гаусса, Холецкого, LU-разложение) и итерационных (Зейделя, верхней релаксации, сопряженных градиентов) методов. Для удобства анализа приведем краткое описание методов:

Метод Гаусса. Система:

$$\sum_{j=1}^{n} a_{ij} = b_i, \quad i = \overline{1,n} \tag{1}$$

приводится к треугольному виду:

$$\sum_{j=1}^{n} a_{ij} = b_i; \quad i = \overline{1,n} \tag{2}$$

Для этого x_i, $i = \overline{1, n-1}$ исключаем из уравнений $j = i + \overline{1,n}$, вычитая из j-го уравнения i-ое, умноженное на $g = \frac{a_{ji}}{a_{ii}}$;

$$b_j = b_j - gb_i; \quad a_{jk} = a_{jk} - ga_{ik}; \quad k = \overline{i+1, n} \tag{3}$$

Затем вычисляем:

$$x_n = \frac{b_n}{a_{n,n}} \tag{4}$$

а далее

$$x_{n-i} = (b_{n-i} - \sum_{m}^{n} a_{m-i,j} x_j)/a_{n-i,n-i} \tag{5}$$

$j = n - i + 1$, $i = 1, n-1$

Сибирская государственная автомобильно-дорожная академия (СибАДИ)

Метод Холецкого: Если в системе $[A]\overrightarrow{x} = \overrightarrow{b}$ матрица $[A]$ симметрична и положительно определена, то $\exists$ нижняя треугольная матрица $[L]$ такая, что

$$[A] = [L][L^T]. \tag{6}$$

Тогда

$$[L][L^T]\overrightarrow{x} = \overrightarrow{b}, \tag{7}$$

а исходную систему можно заменить двумя системами:

$$[L]\overrightarrow{y} = \overrightarrow{b}, [L^T]\overrightarrow{x} = \overrightarrow{y} \tag{8}$$

с треугольными матрицами. Матрицу $[L]$ можно найти из соотношений: $l_{11} = \sqrt{a_{11}},\ l_{i1} = a_{1i}/l_{11},\ i = 2, n$

$$l_{ii} = \sqrt{a_{ii} - \sum_{k=1}^{i-1} l_{ik}^2},\ l_{i1} = a_{1i}/l_{11},\ i = \overline{2, n} \tag{9}$$

$$l_{ji} = \frac{a_{ji} - \sum\limits_{k}^{i-1} l_{ik} l_{jk}}{l_{ii}},\ j = \overline{i+1, n}.$$

Этот метод получил широкое распространение при решении больших разреженных систем.

LU-разложение: Систему $[A]\overline{x} = \overline{b}$ заменяют парой систем:

$$[L]\overline{y} = \overline{b};\quad [U]\overline{x} = b, \tag{10}$$

где [U] – верхняя треугольная матрица, образующаяся в прямом ходе метода Гаусса, а [L] – нижняя треугольная матрица, на диагонали которой единицы, в $l_{ij} = g$ (см. метод Гаусса) $j > i$. Этот метод можно считать обобщением метода Холецкого на случай несимметричных матриц.

Однако, как показали расчеты, при большой размерности системы, более 2 тыс. неизвестных и при плохой обусловленности матрицы, прямыми методами не всегда получается получить решение с высокой точностью. Для получения решения в таких случаях мы предлагаем собственную процедуру, заключающуюся в следующем: после решения системы прямым методом полученный вектор решений поступает в качестве первоначального заброса для решения системы итерационным методом.

Приведем описание этих методов:

Метод Зейделя: Из системы

$$\sum_{j=1}^{n} a_{ij} x_i = b_i;\ i = \overline{1, n} \tag{11}$$

выражаем x_i, $i = \overline{1, n}$ из i-го уравнения

$$a_{ii} x_i = b_i - \sum_{j \neq i} a_{ij} x_j = (b_i - \sum a_{ij} x_j) + a_{ii} x_i \tag{12}$$

(в правой части старое значение x_i, в левой – новое).

$$x_i = x_i + \frac{s_i}{a_{ii}}; \tag{13}$$

где

$$s_i = b_i - \sum_{j=1}^{n} a_{ij} x_j \tag{14}$$

невязка i-го уравнения.

Если матрица $[A]$ симметрична и положительно определена, то описанная процедура есть не что иное, как метод покоординатного спуска при минимизации функции:

$$f(x) = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j - \sum_{i=1}^{n} b_i x_i, \tag{15}$$

следовательно, сходится к решению системы.

Метод верхней релаксации: Если x, y -две последовательные итерации метода Зейделя, то можно подобрать число $1 < \omega < 2$, (зависящее от спектра матрицы), такое, что процесс:

$$x = x + \omega(y - x) \tag{16}$$

сходится существенно быстрее метода Зейделя (в отдельных случаях до 30 раз).

Метод сопряженных градиентов:

Если при решении системы $[A]x = b$ с симметричной, положительно определенной матрицей $[A]$ (или, что то же самое при минимизации функции $f(x) = \frac{1}{2}([A]x, x) - (b, x)$, спускаться не по координатам, а по направлениям $P_1, \ldots, P_n$ таким, что $(P_i, [A]P_j) = 0$, то можно показать, что такая процедура должна сходиться к решению не более, чем за n шагов (практически требуемая точность достигается существенно быстрее).

Приведем выражение для нахождения $P_i, i = 1, n$ и вычисления шагов вдоль этих направлений

$$r_0 = b - [A]x_0, \quad P_0 = r_0; \tag{17}$$

где x_0 – вектор начального заброса, r_0 – вектор невязок.

Пусть на некотором шаге k

$$\lambda_k = \frac{(\overrightarrow{r_k}, \overrightarrow{P_k})}{(\overrightarrow{P_k}, [A]\overrightarrow{P_k})}, \tag{18}$$

тогда

$$\overrightarrow{x_{k+1}} = \overrightarrow{x_k} + \lambda_k \overline{P_k} \tag{19}$$

$$\overrightarrow{r_{k+1}} = \overrightarrow{b} - [A]\overrightarrow{x_{k+1}} = \overrightarrow{r_k} - \lambda_k [A]\overrightarrow{P_k}, \tag{20}$$

и если ввести

$$\beta_k = -\frac{(\overrightarrow{r_{k+1}}, [A]\overrightarrow{P_k})}{(\overrightarrow{P_k}, [A]\overrightarrow{P_k})}, \tag{21}$$

то $\overrightarrow{P_{k+1}}$ вектор сопряженного направления.

$$\overrightarrow{P_{k+1}} = \overrightarrow{r_{k+1}} + \beta_k \overrightarrow{P_k}. \tag{22}$$

Нетрудно показать, что $(\overrightarrow{r_i}, \overrightarrow{r_j}) = 0$, откуда следует сформулированное выше утверждение о сходимости метода, а также $(\overrightarrow{r_j}, \overrightarrow{P_j}) = 0; i > j$ откуда $([A]\overrightarrow{P_i}, \overrightarrow{P_j}) = 0$.

2. Методы ускорения итерационных процессов

Для широкого класса итерационных процессов можно добиться значительного увеличения скорости сходимости применением экстраполяции Эйткина [1, 2]. Приведем краткое описание этой процедуры. Пусть

$$\overline{x}^{n+1} = \overline{g}(\overline{x}^n) \tag{23}$$

– некоторый итерационный процесс, сходящийся к точке $\overline{x}^*$, тогда

$$\overline{x}^* = \overline{g}(\overline{x}^*). \tag{24}$$

Пусть

$$\overline{e}^n = \overline{x}^n - \overline{x}^*, \tag{25}$$

где $\overline{e}^n$ невязка n-го шага. Далее, разложив $\overline{x}^{n+1}$ в окрестности $\overline{x}^*$ в ряд Тейлора, получим:

$$\overline{x}^{n+1} = \overline{g}(\overline{x}^n) = \overline{g}(\overline{x}^*) - g_{i,k}(\overline{x}^*)(x_k^n - x_k^*) + ... \tag{26}$$

где $g_{i,k}(\overline{x}^*) = ||J||$ -матрица Якоби функции g.

Тогда:

$$\overline{x}^{n+1} = \overline{g}(\overline{x}^k) + ||J||\overline{e}^n, \tag{27}$$

откуда

$$\overline{e}^{n+1} \approx \lambda \overline{e}^n, \tag{28}$$

где λ – максимальное собственное число матрицы $||J||$. Соотношение (28) показывает, что итерационный процесс будет расходиться. То есть невязка будет нарастать, если $|\lambda| > 1$, и наоборот, если $|\lambda| < 1$, то процесс будет сходиться. Для оценки значения максимального числа введем два вектора:

$$\overline{Z} = \overline{e}^{n+2} - \overline{e}^{n+1} = \overline{x}^{n+2} - \overline{x}^* - \overline{x}^{n+1} + \overline{x}^* = \overline{x}^{n+2} - \overline{x}^{n+1}, \tag{29}$$

$$\overline{y} = \overline{e}^{n+1} - \overline{e}^n = \overline{x}^{n+1} - \overline{x}^* - \overline{x}^n + \overline{x}^* = \overline{x}^{n+1} - \overline{x}^n. \tag{30}$$

Из соотношений (29) и (30) следует:

$$\overline{Z} = \overline{e}^{n+2} - \overline{e}^{n+1} = \lambda \overline{e}^{n+1} - \lambda \overline{e}^n = \lambda(\overline{e}^{n+1} - \overline{e}^n). \tag{31}$$

Подставим в (31) соотношение (30) и получим

$$\overline{Z} = \lambda \overline{y}. \tag{32}$$

Умножим выражение (32) скалярно справа и слева на $\overline{y}$, тогда

$$\lambda = \frac{(\overline{Z}, \overline{y})}{(\overline{y}, \overline{y})}. \tag{33}$$

Выразим через λ и $\overline{Z}$ значение вектора $\overline{e}^{n+2}$ из соотношений (28) и (29):

$$\overline{e}^{n+2} = \frac{\lambda \overline{Z}}{\lambda - 1}. \tag{34}$$

Далее, используя выражение (25), получим:

$$\overline{x}^{*} = \overline{x}^{n+z} - \overline{e}^{n+z}. \tag{35}$$

Выполняя такие поправочные операции после каждых двух итераций процесса, мы получим увеличение скорости сходимости различных методов.

Например, метода Ньютона-Канторовича в 2,5 раза, метода Зейделя для больших разреженных систем в 8 раз.

В работе австралийского математика Лаузера природа такого ускорения получила интересное объяснение [3].

Пусть $\overline{\Phi}$ -собственный вектор матрицы $||J||$, соответствующий максимальному собственному числу λ.

Пусть

$$h_i(\overline{x}) = g_i(\overline{x}) + \frac{\lambda}{1-\lambda}\Phi_i \Phi_e (g_e(\overline{x}) - x_e). \tag{36}$$

Тогда $h_i(\overline{x}^*) = \overline{x}^*$, следовательно, итерационный процесс сходится в той же точке, что и g

$$h_i(\overline{\Phi}) = \lambda \Phi_i + \frac{\lambda}{1-\lambda}\Phi_i \Phi_e (\lambda \Phi_e - \Phi_e) = \lambda \Phi_i - \lambda \Phi_i = 0. \tag{37}$$

Если $\overline{\Phi}^{*}$ – собственный вектор, соответствующий собственному числу $\lambda^* \neq \lambda$, то:

$$h_i(\overline{\Phi}^{*}) = \lambda^* \Phi_i^* + \frac{\lambda}{1-\lambda}\Phi_i \Phi_e (\lambda \Phi_e^* - \Phi_e) = \lambda^* \Phi_i^* + \frac{\lambda}{1-\lambda}\Phi_i (\Phi_e \Phi_e^*)(\lambda^* - 1) = \lambda^* \Phi_i^*, \tag{38}$$

откуда видно, что максимальное собственное число матрицы Якоби функции h меньше, чем для g, следовательно, h процесс сходится быстрее, чем исходный. Другими словами, срезается то направление, по которому процесс ведет себя особенно «плохо», таким образом можно не только ускорять сходящиеся процессы, но и добиваться сходимости расходящихся. Если в качестве приближения $\overline{\Phi}$ взять $\frac{\overline{Z}}{||Z||}$ (см. 29), то получим:

$$h_i(\overline{x}) = g_i(\overline{x}) + \frac{\lambda}{1-\lambda}\frac{Z_i}{||Z||} Z_e (g_e(\overline{x}) - x_e). \tag{39}$$

Подставляя вместо $\overline{x}$ - $\overline{x}^{n+1}$, получим

$$h_i(\overline{x}) = x_i^{n-2} + \frac{\lambda}{1-\lambda}\frac{Z_i}{||Z||}(Z_e * Z_e) = x_i^{n-2} - \lambda \frac{Z_i}{\lambda - 1}. \tag{40}$$

Сравнивая формулы (40) модификации Лаузера и (34) экстраполяции Эйткина, можно увидеть, что с точки зрения численной реализации, они тождественны. Ряд примеров применения описанной выше процедуры приведен в работе [3].

ВЫВОДЫ

1. Разработана методика и программы расчета для ускорения различных методов решения систем линейных уравнений, надежность методики подтверждена совпадением с аналитическими решениями. На основе использования данных процедур значительно снижены число операций и затраты машинного времени при решении задач расчета напряженно-деформированного состояния методом конечных элементов.
2. В ряде случаев при моделировании итерационных процедур имели место вообще расходящиеся процессы, однако после применения итерационного уточнения с использованием экстраполяции Эйткина удавалось из расходящегося процесса получить сходящийся, что существенно расширяет область применения данных процедур.
3. Для повышения точности расчетов после прямого метода решения системы линейных уравнений применено итерационное уточнение с использованием экстраполяции Эйткина.

Литература

1. Aitken A.C. *On Bernoulli's numerical Solution of algebraic equation* // Proc. Roy. Soc. Edinburgh. 1926. V.46. P.513-532.
2. Aitken A.C. *Studies in practical mathematics* // Proc. Roy. Soc. Edinburgh. 1926. V.46. P.1112-1123.
3. Lawther R. *Modifications of iterative processes for improved convergence characteristics* // International Journal for Numerical Methods in Engineering. 1980. V.15. P.1149-1159.

Математические структуры и моделирование
2001, вып. 8, с. 34–39

УДК 007:681.518.2

СПОСОБЫ ПРЕДСТАВЛЕНИЯ ДАННЫХ В ЭКСПЕРТНЫХ СИСТЕМАХ

В.А. Маренко

The typical models of knowledge representation in expert systems are considered in the present article. The description of symantic nets is given in terms of fuzzy sets.

Для реализации полезной системы, основанной на знаниях, необходимо соответствующее их представление. Сложная интеллектуальная система, способная делать логические выводы на основании знаний в конкретной предметной области и обеспечивающая решение специфических задач, – это экспертная система. Ее необходимо наделить функциями, позволяющими решать задачи, которые в отсутствие эксперта – специалиста в конкретной предметной области невозможно правильно решить. Требования к экспертным системам таковы:

- ориентация на решение реальных и достаточно сложных задач;
- использование знаний, связанных с конкретной предметной областью;
- приобретение знаний от эксперта, то есть обучаемость системы;
- наделение системы способностями эксперта.

В экспертных системах знания о предметной области отделены от других типов знаний, таких как общие знания о том, как решать задачи, или знания о том, как взаимодействовать с пользователем. На рисунке 1 показана базовая структура экспертной системы и ее основные элементы [1]:

- Механизм представления знаний в конкретной предметной области и управления ими — база знаний.
- Механизм, который на основании знаний, имеющихся в базе знаний, способен делать логические выводы — механизм логических выводов.
- Важнейшим структурным элементом является механизм получения знаний от эксперта — модуль приобретения знаний, который служит для поддержки базы знаний и ее пополнения.
- Механизм, который не только способен давать заключение, но и представлять различные комментарии, прилагаемые к этому заключению, объяснять его мотивы — модуль советов и объяснений.
- Пользовательский интерфейс необходим для правильной передачи ответов пользователю.

E-mail: marenko@iitam.omsk.net.ru

Омский филиал Института математики им. Соболева СО РАН

Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (код проекта 01-01-00303).

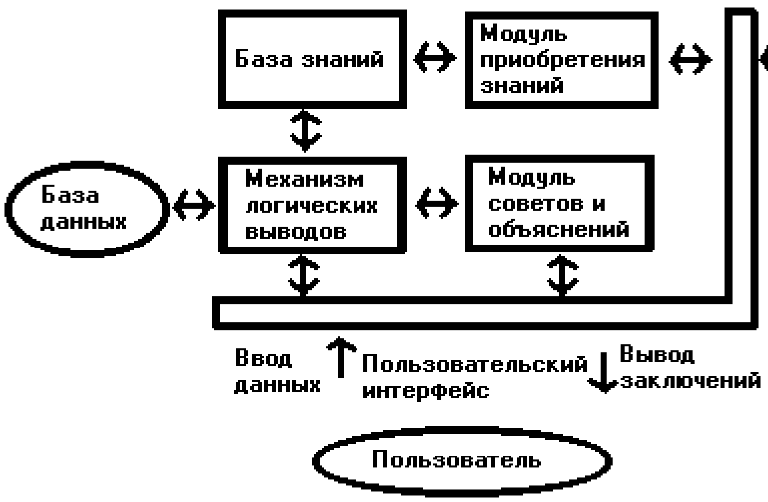


Рис. 1. Пример структуры экспертной системы

В реальных экспертных системах функции отдельных структурных элементов могут быть усилены или расширены. Важную роль играет механизм объяснений — и не только для пользователя системы, но и для эксперта, предоставляющего знания в конкретной предметной области. С его помощью эксперт определяет, как работает система, и может точно выяснить, как используются представленные им знания.

В системах, основанных на концепциях искусственного интеллекта, знания представлены в конкретной форме, имеющаяся база знаний позволяет их легко определять, модифицировать и пополнять. Категория знаний, основанных на собственном опыте специалиста в данной предметной области, накопленных в результате многолетней практики, играет решающую роль в повышении эффективности экспертных систем.

Типичные модели представления знаний:

- логическая модель;
- продукционная модель;
- модель, основанная на использовании фреймов;
- модель семантической сети.

Отличительными чертами логических моделей являются единственность решения и возможность реализации формально точных определений и выводов.

В продукционных системах знания представлены отдельными фрагментами-продукциями, которые можно представить в виде $A \to B$, где A — условие, B — следствие. Связь продукций во время работы обеспечивается динамически функционирующими логическими цепочками, и каждое полученное следствие может стать новым условием. Продукции используются для представления слабо связанных и слабо структурированных данных.

Фреймовая модель представляет собой систематизированную в виде единой теории психологическую модель памяти человека и его сознания [2]. Фрейм (рамка) — это единица представления знаний, детали которой при необходимости могут быть изменены согласно текущей ситуации. В этой иерархической структуре можно выделить: уникальное имя для класса объектов; имена компонентов данного объекта; имена отношений, связывающие объект и компоненты; значения компонентов, которые могут различаться у объектов, описываемых данным фреймом. Во время работы системы при обмене сообщениями между различными фреймами осуществляется или межфреймовый обмен информацией или передача управления другому фрейму.

Семантическая сеть — это один из наиболее удобных способов представления знаний, так как обеспечивает стабильный метод анализа смысла предложений и указывает сходства в смысле предложений, тесно связанных, но обладающих разной структурой. Семантические сети [3] состоят из точек, называемых узлами, и связывающих их дуг, описывающих отношения между ними. Узлы в семантической сети соответствуют объектам, концепциям или событиям. Подобная ассоциативная структура называется плоскостью, описываемые узлы — вершинами типа, а связанные с ними соответствующие отдельные слова, определяющие время, место, качество и т. д., — вершинами лексем. Заменив вершину типа на элемент, а вершину лексемы — на свойство, данные, основанные на фактах, можно представить с помощью структур трех типов: элементы, свойства и указатели.

При построении семантической сети отсутствуют ограничения на сложность сети и число связей элементов и свойств. Поэтому функции, обозначенные указателями, можно упорядочить с помощью теории множеств. Суть порождающего метода состоит в том, что сначала задается небольшое множество, затем на основе этого множества порождается новое множество. Для такого порождения задаются определенные способы, которые обычно аксиоматизируются. В терминах теории обработки информации правила порождения, создающие новое множество из уже известного, есть элементы структуризации данных. Определенное таким образом и аксиоматизированное множество будет теперь достаточно широким для полного математического описания системы. Следовательно, использование в качестве механизма структуризации данных порождающих правил позволяет построить систему обработки информации с четким теоретическим фундаментом. Одним из методов изучения множеств без уточнения их границ является теория нечетких множеств, которая успешно применяется ныне для описания моделей представления знаний.

Рассмотрим более подробно представление знаний с помощью семантических сетей [4]: семантическая сеть, соответствующая модели конкретной предметной области, может быть представлена двойкой следующего вида:

$$S = \{D, U\},$$

где D – множество документов, U – множество дуг, связывающих документы.

Каждая дуга показывает связь двух документов. Документ D_i семантической сети представляется следующим образом:

$$D_i = \{I, P, U_{D_i}\},$$

где I – содержание документа, P – множество понятий, описываемых документом, U_{D_i} – множество отношений между понятиями и содержанием документа определяется множеством двоек вида:

$$U_{D_i} = \{T, U^d_{IP}\},$$

где T – тип отношения для понятия и информационной части документа, определяющий два вида зависимости: t_1 – понятие является входным для документа – и t_2 – понятие является выходным для документа. U^d_{IP} – нечеткое подмножество, которое показывает степень зависимости информационной части документа и понятия.

$$U^d_{IP} = \left(\mu_{\tilde{U}^d_{IP}}(P_j, I) \mid P_j \in D_i,\ I \in D_i\right),$$

$j = 1, \ldots, n$, где $\mu_A(x)$ – значение функции принадлежности элемента x из универсума X нечеткому подмножеству $\tilde{A}$ множества X; P_j – понятие, принадлежащее документу D_i; n – количество понятий в документе.

Документ D_i соответствует объекту $\tilde{D}_i$ с неопределенными и фиксированными атрибутами:

$$\tilde{D}_i = \{I_i, P_1, \ldots, P_n, t_p, \{\mu_{\tilde{D}_i}(I_i, P_1),$$

$$\ldots, \mu_{\tilde{D}_i}(I_i, P_n)\}\},$$

где I_i – информационная часть i-го документа, $P_1, \ldots, P_n$ – понятия, принадлежащие i-му документу, t_p – тип отношения для понятий, $\mu_{\tilde{D}_i}(I_i, P_i)$ – отношение близости понятия P_i и информационной части I_i.

Зависимость между узлами семантической сети строится на основе взаимосвязи между понятиями документа, поэтому вводится нечеткое отношение, определяющее близость понятий:

$$U_{P_{ij}} = \mu_s(P_i, P_j),$$

где s принадлежит S, и на этой основе формируется нечеткое подмножество:

$$U_p = \{P_i, P_j, \mu_s(P_i, P_j) \mid P_i \in P, P_j \in P, i, j = 1, \ldots, N\},$$

где N – количество понятий в базе знаний.

Структура семантической сети показана на примере модели предметной области, которая описывает электромагнитную совместимость радиоэлектронных средств.

Документами в ней являются главы книги В. И. Владимирова и др. «Электромагнитная совместимость радиоэлектронных средств и систем», информационными частями – тексты глав, а понятиями – ключевые слова. Модель предметной области S (рисунок 2) состоит из следующего: Документ D_1. Информационная часть (I_1) – «Причины возникновения проблемы электромагнитной

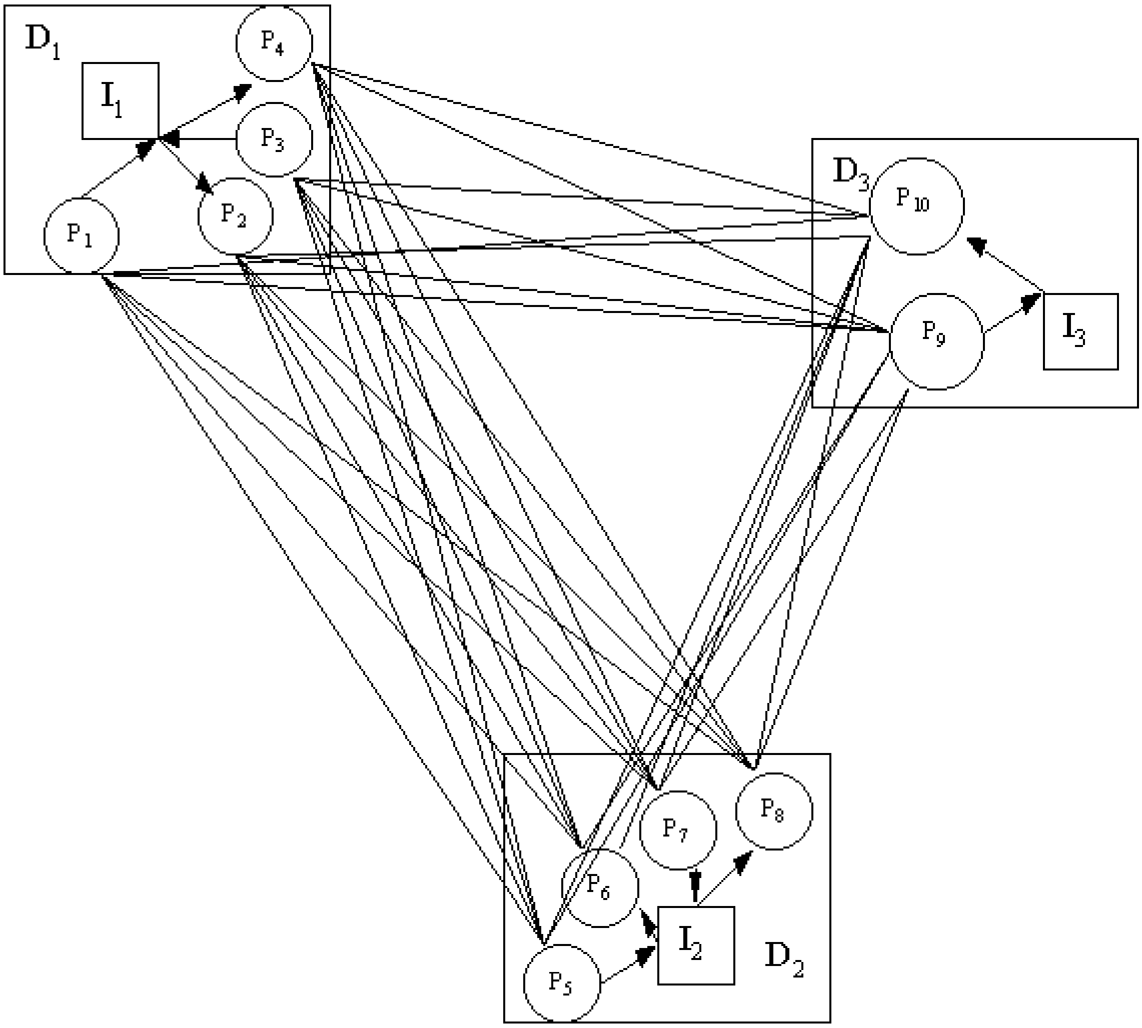


Рис. 2. Модель предметной области. Прямые линии показывают отношения (U_{ij}) между документами.

совместимости». Понятия ($P_1 - P_4$): техническое средство, диапазон радиочастот, излучение, радиопомеха. Документ D_2. Информационная часть I_2 – «Технические характеристики радиоэлектронных средств, влияющие на их электромагнитную совместимость». Понятия ($P_5 - P_8$): рабочая частота, излучение, диаграмма направленности антенны, частотная избирательность. Документ D_3. Информационная часть I_3 – «Основные пути обеспечения электромагнитной совместимости радиоэлектронных средств и систем». Понятия (P_9, P_{10}): техническое решение, стандарт.

Основным преимуществом представления базы знаний в виде семантических сетей является то, что семантические сети более других способов соответствуют современным представлениям об организации долговременной памяти человека. А их описание при помощи нечетких множеств позволяет при определении понятий учитывать субъективные мнения отдельных индивидуумов [5].

Литература

1. Уэно Х. и др. *Представление и использование знаний.* М.: Мир, 1989. 220 с.
2. Осуга С. *Обработка знаний.* М.: Мир, 1989. – 293 с.
3. Уотермен Д. *Руководство по экспертным системам.* – М.: Мир, 1989. – 388 с.
4. Чапейкин А. О. *Использование нечеткой логики в описании предметной области* // Вычислительные машины, комплексы и сети: Межвуз. сб. научн. тр. Рязань: РГРТА, 1998. — С. 136–140.
5. Гаврилова Т. А., Хорошевский В. Ф. *Базы знаний интеллектуальных систем.* – СПб: Питер, 2000. - 384 с.

Математические структуры и моделирование
2001, вып. 8, с. 40–43

УДК 517.938

МЕТОД ФИКТИВНЫХ ПОТОКОВ

Б.К. Нартов

The new method of the optimization of the Smooth dynamic Systems of initial characteristics is discribed.

Представляемый метод оптимизации вектора начальных характеристик динамической системы был предложен нами в работе [1] для конкретной задачи оптимального управления траекториями подвижных объектов, характеристики которых ухудшаются в результате взаимодействия с объектами противника. В настоящем сообщении в основном излагаются результаты [2], где удалось распространить метод фиктивных потоков на весь класс гладких систем и существенно расширить его предметную область.

Рассмотрим гладкую управляемую систему

$$\dot{P}_k(t) = f_k(P_1(t), \ldots, P_N(t), x(t), t),\ 1 \le k \le N, \tag{1}$$

где $x(t)$ – вектор управления. Задав далее начальные условия, интервал управления $(0, T)$ и множество управлений X, запишем задачу

$$J(x^*) = \inf_{x \in X} J(x), \tag{2}$$

где

$$J(x) = F(P_1(T), \ldots, P_N(T)), \tag{3}$$

где F – непрерывно дифференцируема. Решаемая далее задача состоит в отыскании оптимальных в смысле (1)-(3) значений $P_k(0)$ при типичном ограничении $\sum_{k=1}^{N} P_k(0) = C$, где C – заданное число.

Рассмотрим предварительно динамическую систему

$$\dot{Z}_k(t) = -U_k(t) \cdot Z_k(t) + N^{-1} \sum_{i=1}^{N} U_i(t) \cdot Z_i(t),\ 1 \le k \le N, \tag{4}$$

с заданными начальными условиями $Z_k(0)$, где $U_k(t)$ – заданные непрерывные функции.

E-mail: nartov@iitpm.omsk.net.ru

Омский филиал Института математики им. С.Л.Соболева СО РАН

Работа выполнена при финансовой поддержке РФФИ (проект 01-01-00303 – теоретическая часть, проект 01-07-90003 – верифицирующие эксперименты).

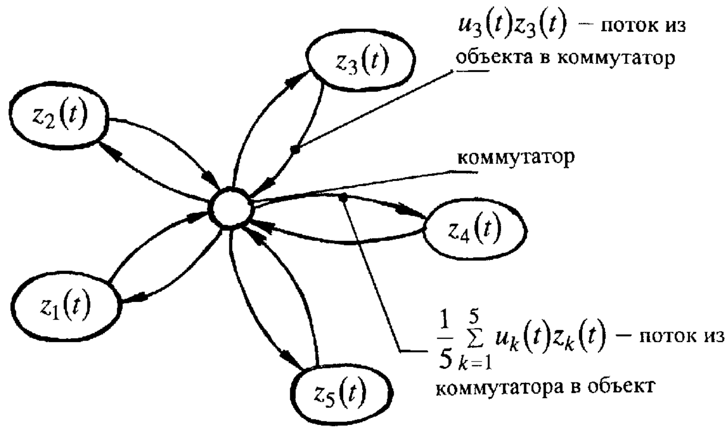


Рис. 1. Коммутатор потоков характеристик.

Обозначим

$$U(t) = \{U_k(\tau),\ 1 \le k \le N,\ 0 \le \tau \le t\};$$

$$U^+(t) = \{U_k(\tau)|U_k(\tau) \ge 0,\ 1 \le k \le N,\ 0 \le \tau \le t\};$$

$$Z(t) = \{Z_k(t),\ 1 \le k \le N\};$$

$$Z^+(t) = \{Z_k(t)|Z_k(t) \ge 0,\ 1 \le k \le N\};$$

$$Z^-(t) = \{Z_k(t)|Z_k(t) < 0,\ 1 \le k \le N\}.$$

Если $Z_k(t)$ есть характеристики некоторых объектов, то система уравнений (4) описывает управляемый «коммутатор» потоков характеристик из объекта в объект (рис.1).

При этом (4) обладает следующими полезными свойствами:

1. При любых начальных условиях $Z(0)$ произвольное управление $U(t)$ сохраняет начальный ресурс C, то есть

$$\sum_{k=1}^{N} Z_k(t) \equiv \sum_{k=1}^{N} Z_k(0) = C;$$

2. Любое управление $U^+(t)$ отображает произвольный набор неотрицательных начальных характеристик в некоторый набор неотрицательных конечных характеристик, то есть

$$U^+(t) : Z^+(0) \to Z^+(t),$$

и, более того, для любых двух наборов Z^{1+} и Z^{2+} с совпадающими суммами характеристик и любого t найдется

$$U^+(t) : Z^{1+}(0) \to Z^+(t) = Z^{2+}.$$

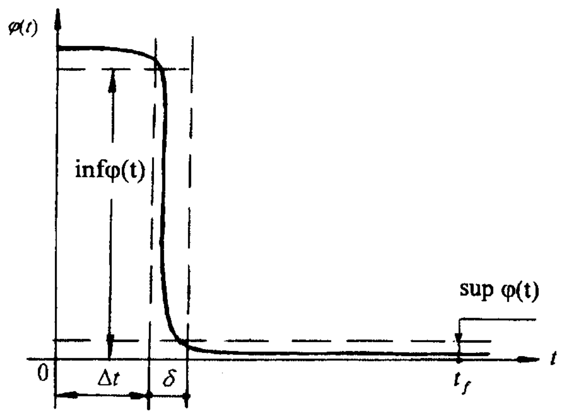


Рис. 2. График функции $\varphi(t)$.

3. Любое управление $U^+(t)$ отображает произвольный набор отрицательных начальных характеристик в некоторый набор отрицательных конечных характеристик, то есть

$$U^+(t) : Z^-(0) \to Z^-(t),$$

и, более того, для любых двух наборов Z^{1-} и Z^{2-} с совпадающими суммами характеристик и любого t найдется

$$U^+(t) : Z^{1-}(0) \to Z^-(t) = Z^{2-}.$$

Рассмотрим теперь следующую модификацию исходной системы уравнений (1):

$$\dot{\tilde{P}}_k(t) = f_k(\tilde{P}_1(t), \ldots, \tilde{P}_N(t), x(t), t) - U_k \tilde{P}_k(t) + N^{-1} \sum_{i=1}^{N} U_i(t) \tilde{P}_i(t), \ 1 \le k \le N, \quad (5)$$

с начальными условиями и интервалом управления $(0, T)$ из задачи (1)-(3) и множеством управлений $\tilde{X} = X \otimes U$, где

$$U = \{U^+(t) | 0 \le U_k(t) \le \varphi(t), \ 1 \le k \le N, \ 0 \le t \le T\},$$

где $U_k(0)$ заданы; $\varphi(t)$ – заданная непрерывно дифференцируемая функция. При этом $\inf \varphi(t)$ на интервале $(0, \, \Delta t)$, $\sup \varphi(t)$ на интервале $(\Delta t + \delta, \, T)$, Δt и δ – регулируются параметрами $\varphi(t)$ независимо (рис.2). Конкретный вид $\varphi(t)$ вполне произволен.

Рассмотрим далее для системы (5) задачу

$$J(\tilde{x}^*) = \inf_{\tilde{x} \in \tilde{X}} J(\tilde{x}), \quad (6)$$

где

$$J(\tilde{x}) = F(\tilde{P}_1(T), \ldots, \tilde{P}_N(T)), \tag{7}$$

см. (2), (3). Используя теперь гладкость исходной системы (1) и исходного функционала качества (3) и свойства вспомогательной системы (4), можно показать, что при исходном ограничении $P_k(0) \geq 0$, $1 \leq k \leq N$, или $P_k(0) < 0$, $1 \leq k \leq N$, для любой исходной задачи вида (1)-(3), соответствующей вспомогательной задачи (5)-(7) и любого $\varepsilon > 0$ найдется такая $\varphi(t)$, что

$$|P_k^*(0) - \tilde{P}_k(\Delta t + \delta)| < \varepsilon, \ 1 \leq k \leq N, \tag{8}$$

где $P_k^*(0)$ – искомые оптимальные начальные характеристики исходной задачи (1)-(3); $\tilde{P}_k(\Delta t+\delta)$ – характеристики из решения задачи (5)-(7), взятые в момент $\Delta t + \delta$. Случай несовпадающих знаков $P_k(0)$ требует более сложной модификации исходной задачи.

Замечание 1. В широком классе задач оптимального управления, например в задачах планирования конфликтов, часть начальных характеристик в (1) является неоптимизируемыми параметрами. В этом случае в системе (5) вспомогательной задачи (5)-(7) модифицируются лишь дифференциальные уравнения характеристик, начальные значения $P_k(0)$ которых могут быть оптимизированы.

Замечание 2. Задачу оптимального размещения начальных ресурсов можно поставить и для неуправляемой гладкой системы, переписав функционал качества (3) в виде

$$J(x) = F(P_1(T), \ldots, P_N(T)) = \tilde{F}(P_1(0), \ldots, P_N(0), T).$$

Представленный метод тривиально распространяется и на этот класс задач.

Замечание 3. Хотя решение вспомогательной задачи (5)-(7) и позволяет получить оптимальные начальные характеристики $P_k(0)$ исходной задачи (1)-(3) за один шаг, вопрос о цене дополнительных ограничений и численного представления функции $\varphi(t)$ в реализующем алгоритме вспомогательной задачи остается открытым.

Замечание 4. Расширение системы (4) позволяет записать и решать в виде задачи оптимального управления некоторые транспортные задачи линейного программирования. В этом случае матрица перевозок модифицируется в течение $(0, T)$ при $\varphi(t) = \text{const}$.

Литература

1. Nartov B.K. *Conflict of Moving Systems.* – AMSE Press, France, 1994. – 87 p.
2. Нартов Б.К., Чуканов С.Н. *Модели траекторного управления.* - Омск: Изд-во ОмГУ, 2001. – 95 с.

Математические
структуры и моделирование
2001, вып. 8, с. 44–51

УДК 517.977

О МЕТОДЕ УПРУГИХ ФУНКЦИЙ

Б.К. Нартов

In the paper the new method‘s possibilities of the formalization of the path (trajectory‘s) problems in the form of the optimal control problems are presented.

1. Введение

Разработку метода упругих функций стимулировала известная задача планирования поиска на плоскости неподвижных точечных целей с заданными функциями плотностей распределения вероятностей [4]. Долгое время не удавалось свести эту задачу к классическим задачам условной оптимизации. Основную проблему представлял аналитический учет пересечений и самопересечений полос поиска – непереборные решения строились лишь для частных случаев или получались при дополнительных ограничениях на управление поисковыми единицами. Предложенный нами в [1–3] подход решает эту проблему и позволяет формально просто записать общий случай планирования поиска в виде задачи оптимального управления.

Существенно, что предметная область найденного подхода оказалась много шире задач поиска стационарных объектов. Например, в задачах преследования целей с заданными траекториями метод упругих функций позволяет аналитически запоминать моменты первых касаний целей объектами-преследователями и сколь угодно точно «затормаживать» движения коснувшихся пар. Таким образом, удается избавиться от дифференциальных уравнений траекторий объектов-преследователей, обойти комбинаторные трудности и свести общий случай исходной задачи к классической задаче условной оптимизации. В контексте представляемого метода удалось формализовать в виде стандартных задач оптимального управления и некоторые задачи дискретной оптимизации, например задачу коммивояжера, интерпретируемую предварительно как «задачу преследования неподвижных целей».

Ниже мы опускаем проведенные доказательства и описания верифицирующих экспериментов, ограничиваясь по возможности наглядными примерами, представляющими три контрастных класса задач траекторного управления.

E-mail: nartov@iitpm.omsk.net.ru

Омский филиал Института математики им. С.Л.Соболева СО РАН

Работа выполнена при финансовой поддержке РФФИ (проект 01-01-00303 – теоретическая часть, проект 01-07-90003 – верифицирующие эксперименты).

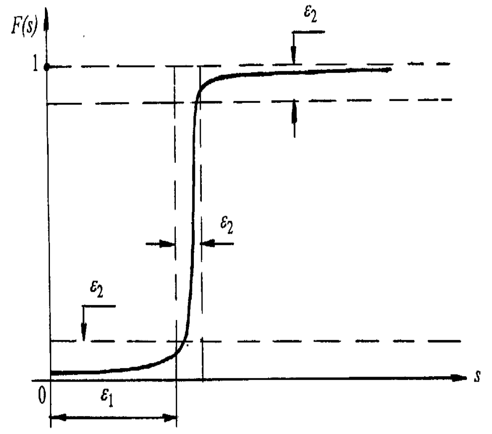


Рис. 1. Вспомогательная функция $F(s)$

2. Преследование подвижных целей

Задача 1 (простейший случай). Задан временной интервал $[0,\, t_f]$, траектория цели $\bar{a}(t)$, $0 \le t \le t_f$, в евклидовом пространстве E^3, начальная координата объекта-преследователя $\bar{x}(0)$, кинематические и пространственные ограничения на его возможные траектории $\bar{x}(\cdot)$, выражающиеся в системе неравенств, связывающих функцию $\bar{x}(t)$ и ее производные до второй включительно.

Требуется формализовать задачу вычисления траектории $\bar{x}^*(\cdot)$, оптимальной в смысле задачи

$$t_\zeta \to \inf, \tag{1}$$

где t_ζ – момент захвата цели, а именно минимальное решение неравенства $|\bar{x}(t) - \bar{a}(t)| < \varepsilon_1$, принадлежащее $[0,\, t_f]$, где ε_1 задано.

Решение. Обозначив $\bar{r}(t) = \bar{x}(t) - \bar{a}(t)$, введем (рис.1) первую вспомогательную функцию $F(s(t))$, $s(t) = |\bar{r}(t)|$, удовлетворяющую трем условиям:

1. $F(s)$ дифференцируема.

2. $\begin{cases} 0 \le F(s) \le \varepsilon_2 & \text{при } 0 \le s \le \varepsilon_1, \\ 1 - \varepsilon_2 \le F(s) < 1 & \text{при } s \ge \varepsilon_1 + \varepsilon_2. \end{cases}$

3. Значения ε_1 и ε_2 регулируются параметрами функции $F(s)$ независимо (в смысле неравенств $0 < \varepsilon_2 < \varepsilon_1$ и $0 < \varepsilon_1 < \varepsilon_3$, где ε_3 – заданное число).

Конкретный вид $F(s)$, удовлетворяющей этим условиям, вполне произволен.

Попытаемся теперь подобрать дифференцируемую функцию $\tilde{F}(s(\cdot), t)$, которая для любого интервала $[0,\, t] \subset [0,\, t_f]$ сколь угодно точно (регулировкой параметров функций $\tilde{F}$ и F) приближает значения функции $F_1(s(\cdot), t) = \inf_{0 \le \tau \le t} F(s(\tau))$ при условии $\tilde{F}(s(\cdot), 0) = 1$.

Можно показать, что простейшая удовлетворяющая этим требованиям функция $\tilde{F}(s(\cdot), t)$ является решением дифференциального уравнения

$$\dot{\tilde{F}} = \alpha \frac{F - \tilde{F}}{F}, \tag{2}$$

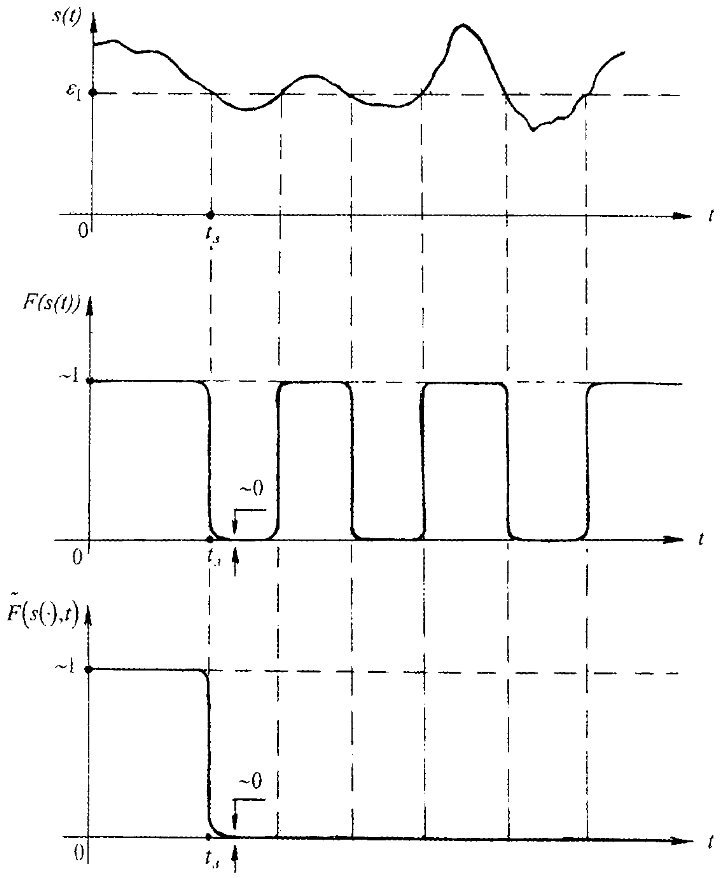


Рис. 2. К построению функции $\tilde{F}$.

где α – регулируемый положительный параметр.

Физически уравнение (2) обусловливает следующие свойства функции $\tilde{F}(s(\cdot), t)$:

$\tilde{F}$ «достаточно быстро» убывает по t при $F < \tilde{F}$;

$\tilde{F}$ «достаточно медленно» возрастает по t при $F > \tilde{F}$.

Введем для функции $\tilde{F}$ естественный термин «упругая» функция, а для функции F – «трафаретная» функция.

Обратимся теперь к рис.2, иллюстрирующему преобразование $s(t) \to F(s(t)) \to \tilde{F}(s(\cdot), t)$, и запишем приближенное (см.ниже) решение исходной задачи в виде

$$\int\limits_0^{t_f} \tilde{F}(s(\cdot), t)\, dt \to \inf . \tag{3}$$

Решая уравнение (2) с начальным условием $\tilde{F}(s(\cdot), 0) = 1$, получим

$$\tilde{F}(s(\cdot), t) = \left[1 + \alpha \int\limits_0^t \exp \left(\alpha \int\limits_0^\tau \frac{d\Theta}{F(s(\Theta))} \right) d\tau \right] \exp \left(-\alpha \int\limits_0^t \frac{d\tau}{F(s(\tau))} \right). \quad (4)$$

Подставляя (4) в (3), получаем классическую задачу вариационного исчисления. Фактически в (3) используется следующая интерпретация «упругой» функции $\tilde{F}$:

$$\frac{d\tilde{t}}{dt} = \tilde{F}(s(\cdot), t), \quad (5)$$

где $\tilde{t}$ – новая переменная – «управляемое» время ($\tilde{t}(0) = 0$). Можно доказать, что для любых параметров и начальных условий задачи 1 всегда найдутся такие конкретные параметры функций F и $\tilde{F}$, что функция $\tilde{t}(t_f)$ из решения (3) с заданной точностью приблизит момент захвата цели t_ζ (если, разумеется, захват вообще возможен – в противном случае $\tilde{t}(t_f)$ с заданной точностью совпадает с t_f).

Практически случай одной цели и одного преследователя, разумеется, легко разрешим стандартными средствами и рассмотрен здесь лишь в качестве иллюстрации. Однако предлагаемый подход позволяет формализовать и многие общие случаи задач оптимального преследования, например

Задачу 2. Задан временной интервал $[0, t_f]$, траектории целей $\bar{a}_i(t)$, $0 \le t \le t_f$, $1 \le i \le M$, в евклидовом пространстве E^3, начальные координаты объектов-преследователей $\bar{x}_j(0)$, $1 \le j \le N$, кинематические и пространственные ограничения на их возможные траектории $\bar{x}_j(\cdot)$.

Требуется формализовать задачу вычисления траекторий $\bar{x}_j(\cdot)$, $1 \le j \le N$, оптимальных в смысле минимизации суммарного времени жизни целей на интервале $[0, t_f]$. Если цель захвачена каким-либо объектом-преследователем (см. задачу 1), ее время жизни равно времени захвата, в противном случае время жизни цели совпадает со значением t_f. При этом объект-преследователь, захвативший какую-либо цель, вычеркивается из дальнейшего рассмотрения.

Заметим, что схему решения задачи 1 нельзя тривиально распространить на общий случай. Необходимо, например, исключить следующие версии:

— «атаку» одной цели многими объектами-преследователями;

— последовательные «атаки» многих целей одним объектом-преследователем.

Учет подобных ограничений в функционале качества [3] существенно усложняет реализующий алгоритм.

3. Поиск стационарных объектов

Исходные данные рассматриваемых в этом разделе задач таковы. Односвязной области $G \subset E^2$ достоверно принадлежат K неподвижных целей с известными функциями плотностей распределения вероятностей $f_k(x)$, $1 \le k \le K$, $x = (x_1, x_2) \in [G]$ где $[G]$ – замыкание области G. В замыкании $[G]$ произвольным образом расположены в начальный момент времени $t = 0$ N поисковых единиц (ПЕ). При движении каждая ПЕ является центром окружности радиуса a, заметающей в области G полосу шириной $2a$ – полосу поиска; попавшая в полосу цель считается обнаруженной (рис. 3).

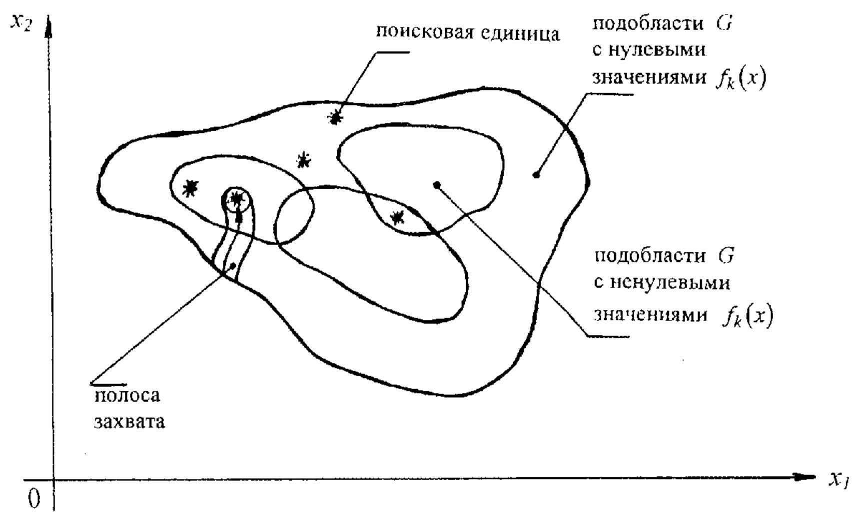


Рис. 3. Исходные данные в задачах поиска.

Введем следующие обозначения:

$\xi_i(t) = (\xi_1^i(t),\, \xi_2^i(t))$ – траектория i-й ПЕ, $1 \le i \le N$, $0 \le t \le t_f$, t_f – время поиска;

$\tilde{K} = \tilde{K}(\{\xi_i(t) | 0 \le t \le t_f,\, 1 \le i \le N\},\, t)$ – случайная величина – количество целей, обнаруженных на интервале времени $[0, t]$, соответствующее стратегии поиска $u = \{\xi_i(t) | 0 \le t \le t_f,\, 1 \le i \le N\}$.

Считая реальные состояния поиска на заданном интервале $(0,\, t_f)$ неизвестными, сформулируем следующую задачу: при заданных кинематических и пространственных ограничениях на управления ПЕ вычислить стратегию поиска $u^* = \{\xi_i^*(t) | 0 \le t \le t_f,\, 1 \le i \le N\}$, максимизирующую математическое ожидание количества целей $M[\tilde{K}(u,\, t_f)]$, обнаруживаемых за время поиска t_f.

Рассмотрим функцию

$$f(x) = \sum_{k=1}^{K} f_k(x).$$

Очевидно

$$\int\limits_G f(x)\, dS = K.$$

Обозначим $u(t) = \{\xi_i(\tau) | 0 \le \tau \le t,\, 1 \le i \le N\}$, $u(t_f) = u$. В каждый момент времени $t \in (0,\, t_f)$ множеству траекторий $u(t)$ соответствует объединение полос поиска ПЕ $G_1(u(t))$. Обозначим теперь $G(u(t)) = G \setminus G_1(u(t))$ и $\Delta G_1(u(t)) = G_1(u(t + \Delta t)) \setminus G_1(u(t))$.

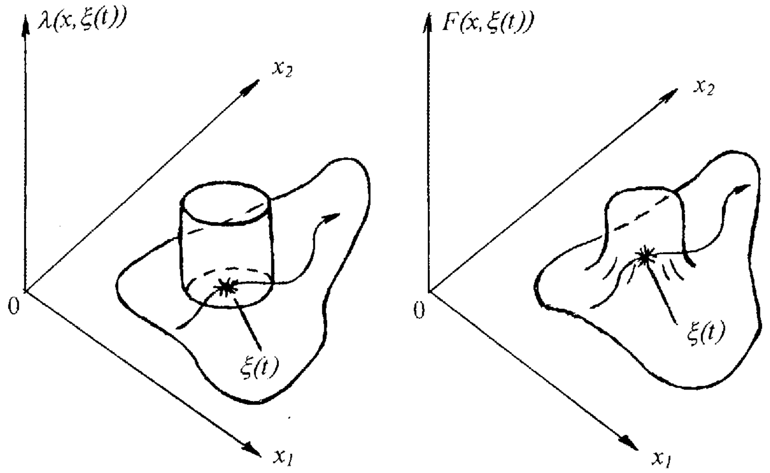


Рис. 4. Функции λ и F.

Отсюда и из определения функции $f(x)$ и математического ожидания $M[\tilde{K}(u,\,t)]$ следует:

$$\Delta M[\tilde{K}(u,\,t)] = \int\limits_{\Delta G_1(u(t))} f(x)\,dS. \tag{6}$$

Теперь исходную задачу можно записать в виде

$$\int\limits_{G_1(u(t_f))} f(x)\,dS \to \max. \tag{7}$$

Из (6) очевидно, что при планировании слепого поиска необходим учет пересечений и самопересечений полос поиска ПЕ, составляющий основную проблему формализации (7) в виде задачи оптимального управления.

Предлагаемое ниже решение основано на описанном в предыдущем разделе механизме «упругих» функций $\tilde{F}$, задаваемых, однако, в другом виде.

Определим предварительно на области G вспомогательную функцию λ, связанную с координатами ПЕ,

$$\lambda(x,\,\xi(t)) = \begin{cases} 1, & \text{если } |\xi(t)-x| < \alpha, \\ 0, & \text{если } |\xi(t)-x| \ge \alpha. \end{cases}$$

Зададим далее дифференцируемую колоколообразную функцию $F(x,\,\xi(t))$, сколь угодно точно приближающую значения λ. Конкретный вид F вполне произволен (рис.4.)

Рассмотрим теперь функцию $\tilde{F}(x,u,t)$, являющуюся решением дифференциального уравнения

$$\dot{\tilde{F}} = \alpha \sum_{i=1}^{N} \frac{F_i - \tilde{F}}{1 - F_i}, \tag{8}$$

где α – регулируемый положительный параметр; $\tilde{F}(x,u,0)=0$.

Можно показать, что для любых начальных условий исходной задачи всегда найдутся такие F_i и F, что к моменту окончания поиска t_f функция $\tilde{F}$ реализует над областью поиска G дифференцируемый профиль, повторяющий движения ПЕ. При этом высота профиля с заданной точностью будет равна единице над просмотренными областями, в том числе и над областями пересечений и самопересечений полос поиска, и с заданной точностью равна нулю вне просмотренных областей. Это полезное свойство $\tilde{F}$ позволяет формализовать исходную задачу поиска (7) в виде

$$J(u)=\int\limits_G f(x)\tilde{F}(x,u,t_f)\,dS\to\max, \tag{9}$$

то есть получить функционал с дифференцируемым интегрантом, учитывающим любое возможное наложение полос поиска ПЕ.

Фактически относительно (9) утверждается следующее.

Для любых параметров и начальных условий исходной задачи и любого $\varepsilon>0$ найдутся такие F_i и $\tilde{F}$, что (9) формализует задачу (7) для некоторого набора плотностей вероятностей $\tilde{f}_k(x)$, $1\le k\le K$, из интервала $|f_k(x)-\tilde{f}_k(x)|<\varepsilon$, где $f_k(x)$ – исходные плотности вероятностей.

В [3] по подобной схеме мы формализовали и некоторые задачи управления поиском в реальном масштабе времени, когда известны моменты обнаружения и координаты обнаруживаемых целей. Заметим здесь, что вопросы о цене представления $\tilde{F}$ в реализующем алгоритме и устойчивости траекторий ПЕ относительно точности приближения $f_k(x)$ требуют отдельного обсуждения.

4. Дискретная оптимизация

Используя результаты раздела 2, интерпретируем задачу коммивояжера как «задачу оптимального преследования неподвижных целей», а именно как задачу минимизации времени захвата последней из N неподвижных целей одним неуничтожаемым объектом-преследователем.

Обозначим:

$\bar{x}_i\in E^3$, $i=\overline{1,N}$ – координата i-го пункта;

$\bar{x}(t)$ – координата коммивояжера, где $\bar{x}(0)$ задана;

F и $\tilde{F}$ – трафаретная и упругая функции – в том виде, в котором они определены в модели оптимального поиска в разделе 3 (случай одной поисковой единицы);

L – длина минимального маршрута.

Запишем задачу

$$\int\limits_0^{t_f}\prod_{i=1}^{N}\tilde{F}(\bar{x}_i,\bar{x}(\cdot),t)\,dt\to\sup \tag{10}$$

с ограничением на управление $|\dot{\bar{x}}(t)|\le V$, $t\in[0,\,t_f]$.

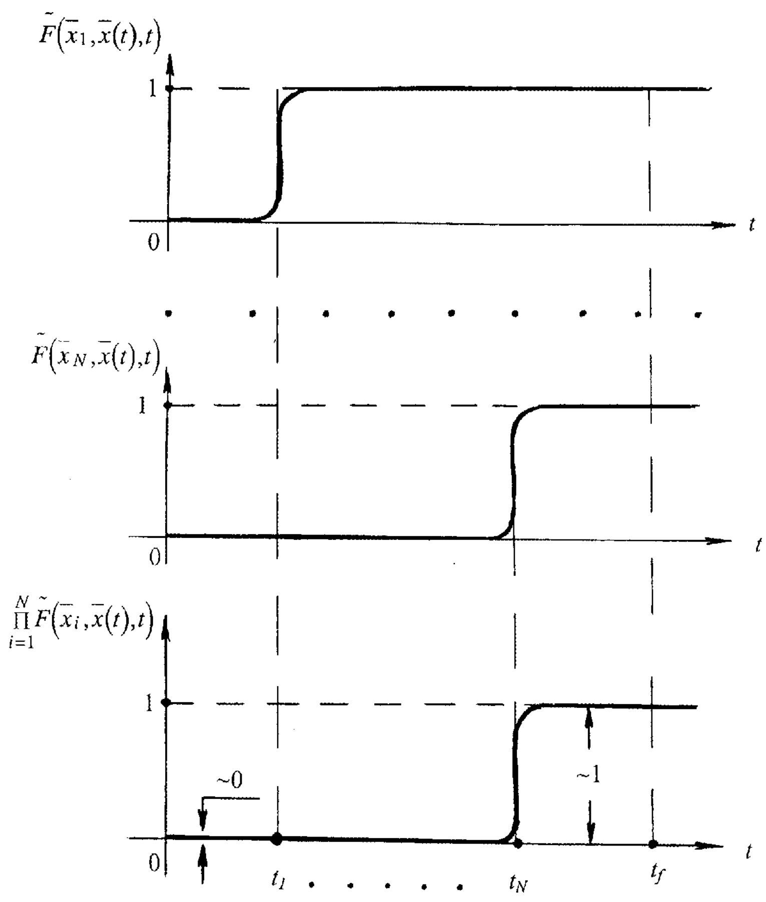


Рис. 5. К задаче коммивояжера. $t_1, \ldots, t_N$ – моменты ε-касаний пунктов $1, \ldots, N$.

Далее легко показать, что при условии $Vt_f > L$ для любого $\varepsilon > 0$ найдутся такие F и $\tilde{F}$ и ограничение на $|\dddot{x}(t)|$, что решение (10) обеспечивает ε-близкий и, с точностью до $N\varepsilon$, оптимальный обход пунктов (рис. 5).

Литература

1. Nartov B.K. *Conflict of Moving Systems.* AMSE Press, France, 1994. – 87 p.
2. Нартов Б.К. и др. *Конфликт сложных систем. Модели и управление.* М.: Изд-во МАИ, 1995. – 120 с.
3. Нартов Б.К., Чуканов С.Н. *Модели траекторного управления:* Монография. Омск: ОмГУ, 2001. – 95 с.
4. Хеллман О. *Введение в теорию оптимального поиска.* М.: Наука, 1985. – 254 с.

Математические структуры и моделирование
2001, вып. 8, с. 52–55

УДК 519.6:611.716.4:
616.716.4-089.28-
053:612.015

МАТЕМАТИЧЕСКИЕ МОДЕЛИ ИНТАКТНОГО МОЛЯРА И МОЛЯРА ПОСЛЕ ГЕМИСЕКЦИИ

В.М. Семенюк, А.В. Артюхов, А.В. Сырцова, А.К. Гуц

A part of the crown and one of molar-roots can be used as a base for dentures. We have applied morphological findings and computer modeling in order to mathematically motivate the use the remained root and a part of the crown.

Осложненный кариес или травма могут вызывать разрушение одного из корней двукорневых зубов, тогда как второй корень не был подвержен разрушению. Возможно развитие патологического процесса в костной ткани челюсти вокруг пораженного корня, в то же время костная ткань челюсти вокруг второго корня остается интактной (здоровой). В этом случае врачам-стоматологам приходится решать вопрос о полном удалении зуба. Между тем удаление зуба вызывает резорбцию костной ткани альвеолярного отростка и повышение функциональных нагрузок на пародонт соседних зубов. Для профилактики этих осложнений часть коронки и один из корней многокорневого зуба при определенных условиях можно сохранить и использовать в качестве опоры различных конструкций зубных протезов.

Чтобы математически обосновать использование оставшихся корня и части коронки зуба, мы применили данные морфологических исследований и компьютерное моделирование. На основе метода конечных элементов с помощью пакета прикладных программ «COSMOSM 2.5» построили компьютерную модель двукорневого зуба (интактного и после гемисекции) вместе с окружающей челюстной костью с целью исследования на прочность при внешних нагрузках.

Суть использованных методов заключается в следующем:

1. На изъятых интактных первых и вторых больших коренных зубах (молярах) человека получали параметры коронки и корня.
2. Строили компьютерную модель зуба (здорового зуба и зуба после гемисекции) вместе с окружающей его челюстной костью.
3. С помощью компьютерного пакета прикладных программ «COSMOSM 2.5» моделировали деформирование зуба и окружающей челюстной кости под влиянием заданной нагрузки, получали полную картину распределения напряженных состояний здорового зуба и зуба после гемисекции.

E-mail: gmaart@other.omsu.omskreg.ru, guts@univer.omsk.su
Омская государственная медицинская академия, Омский государственный университет

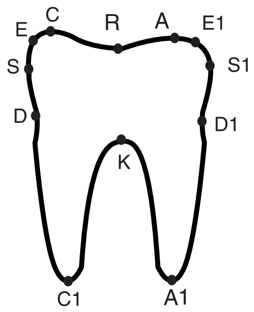


Рис. 1. Модель интактного зуба.

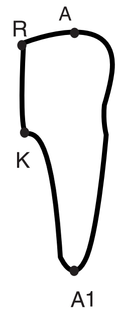


Рис. 2. Модель зуба после гемисекции.

Для упрощения расчетов на данном этапе рассматривалась плоская модель, то есть, по существу, мы пытались увидеть то, что происходит внутри зуба в мысленно выделенном плоском сечении, проходящем через геометрическую ось зуба. Измерение параметров коронок и корней проводились с точностью до 0,1 мм. На рис.1 представлен ряд точек. Измерение расстояний между ними позволяет создать двухмерную плоскую модель моляра.

Усредненные результаты измерения и необходимые физические показатели даны в табл.1 и табл.2.

Таблица 1. Средние арифметические параметры измеренных зубов (в мм)

AA1	CC1	DD1	A1C1	RK	EE1	SS1
18,9	19,9	9,9	4,2	8,5	10,6	11,6

Таблица 2. Физические показатели костной ткани и дентина

	Плотность, $lb \cdot s^2/in^4$	Модуль Юнга, lb/in^2
Кость	$0.168 - 0,187 \cdot 10^{-3}$	$2,03 - 2,75 \cdot 10^6$
Дентин	$0.168 - 0,187 \cdot 10^{-3}$	$2,00 - 2,81 \cdot 10^6$

Были построены математические модели здорового зуба и зуба после гемисекции (рис.2).

Появление обширных зон в корне и коронке зуба с напряжениями, превосходящими предел прочности ($\sigma_B \sim 20 - 40$ н/мм^2) дентина, трактуется как ситуация, ведущая к их разрушению. Накладывалось условие нулевого граничного перемещения (жесткого закрепления). Нагрузка бралась точечная, сосредоточенная в трех узлах (имеющая одну и ту же величину в каждом узле).

Мы смогли наглядно (на экране монитора) увидеть, как происходит деформирование зуба и окружающей челюстной кости под влиянием заданной нагрузки, а также увидели полную картину распределения (в виде цветных зон) напряженных состояний зуба.

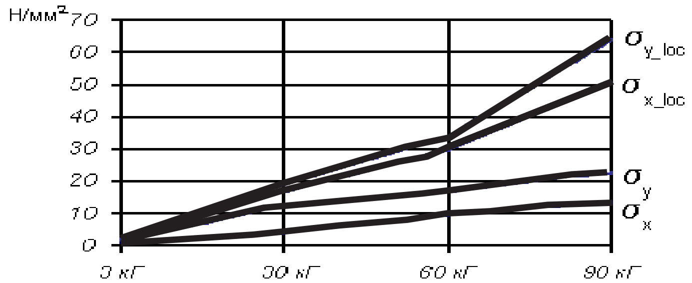


Рис. 3. Напряжения при вертикальной несимметричной нагрузке для модели здорового зуба.

Все это позволило провести серию компьютерных экспериментов, позволяющих определить напряженно-деформированные состояния здорового зуба и зуба после гемисекции. Самый неблагоприятный и реальный вид нагрузки – вертикальная несимметричная. В случае вертикальной нагрузки в 3 кГ для модели «здорового зуба» полученные напряжения по порядку величины согласуются с результатами, полученными ранее [3]. При вертикальной несимметричной нагрузке в 30 кГ возникающие напряжения на большей части зуба возрастают, но еще не попадают в критическую зону (для разрыва это напряжения от 20 до 42 Н/мм2). Если нагрузки приближаются к 60 кГ, то появляются участки с напряжениями, которые уже попадают в критическую зону (рис.1). При вертикальной несимметричной нагрузке в 90 кГ напряжения на больших участках дентина попадают в зону разрушения (где $\sigma_x, \sigma_y > 42$Н/мм^2).

Локально, на очень малых участках корня возникающие напряжения по вертикали и по горизонтали попадают в критическую зону уже при приближении нагрузки к 30 кГ (рис. 1), в зону разрушения – при нагрузке выше 30 кГ. Аналогичная картина наблюдается и при нагрузке под углом 45° к оси зуба.

При несимметричной вертикальной нагрузке в 3 кГ для модели зуба после гемисекции (рис.2) возникают напряжения большие, чем при этой же нагрузке на здоровый зуб, но они еще не попадают в критическую зону. Если нагрузка приближается к 30 кГ, то в дентине появляются зоны с критическими напряжениями (рис.2). При наличии дефектов (микротрещин) в дентине оставшаяся часть коронки подвержена разрушению в области этих дефектов. Увеличивая нагрузку до 60кГ и далее, мы можем наблюдать дальнейшее увеличение зон с критическим напряжением, появляются зоны разрушения.

На основании проведенных исследований можно сформулировать следующие выводы:

- данные морфологических исследований могут служить основой для компьютерного моделирования двухкорневого зуба (интактного, после гемисекции);
- зуб после гемисекции может не только удерживаться в альвеоле челюсти, но и воспринимать дополнительную нагрузку за счет резервных сил

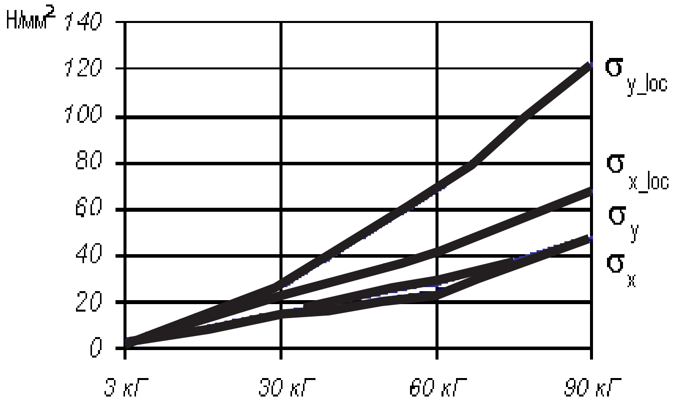


Рис. 4. Напряжения при вертикальной несимметричной нагрузке для модели зуба после гемисекции.

пародонта;

– в определенных пределах эти силы можно использовать при создании ортопедических конструкций.

Литература

1. Князева М.Б. *Подготовка зубов и пародонта к применению металлокерамических протезов*: Автореф. дис. канд. мед. наук: 14.00.21/ АО «Стоматология».-М.,1997.-16 с.
2. Воробьев В.А. *Выбор конструкции зубных протезов и имплантантных систем на основе программного математического моделирования при лечении больных с различными дефектами зубных рядов*: Автореф. дис.: д-ра мед. наук. Омск, 1997
3. Гуц А.К., Капотина Т.Н., Панова Н.И., Семенюк В.М., Файзуллин Р.Т., Яковлев К.К. *Математическое обоснование к использованию корней фронтальных зубов, разрушенных ниже уровня десны, под штифтовые конструкции.* Деп. в ВИНИТИ 21.06.95. N 1790 - И95. 22с.
4. Партон В.З. *Механика разрушения: от теории к практике.* М.:Наука, 1990.
5. Сегерлинд Л. *Применение метода конечных элементов.* М., 1979.
6. Маркин В.А. *Прогнозирование осложнений при использовании металлокерамических протезов с помощью метода математического моделирования* : Автореф. дис. канд. мед. наук: 14.00.21. Моск. гос. медико-стоматолог. ун-т, 1999. -24 с.
7. Капотина Т.Н., Семенюк В.М., Яковлев К.К., Гуц А.К., Панова Н.И. *Математическое обоснование к использованию культевой штифтовой вкладки с «воротничком» при разрушении корней зубов ниже уровня десны* // Вестник Омско- го университета. 1996. N.2. С.17-19.
8. Семенюк В.М., Гуц А.К., Капотина Т.Н., Вагнер В.Л. *Выносливость опорно-удерживающего аппарата зуба, восстановленного литой культевой штифтовой конструкцией, к вертикальной нагрузке* // Перспективы развития современной стоматологии: проблемы уральского региона. материалы конференции стоматологов 20-22 мая 1997 г., Екатеринбург. – Уральская гос. мед. академия, Екатнринбург, 1997. С.211-213

Математические структуры и моделирование
2001, вып. 8, с. 56–65

УДК 53.072+536.425

ОПРЕДЕЛЕНИЕ КРИТИЧЕСКИХ ПАРАМЕТРОВ ЭФФЕКТИВНОГО ГАМИЛЬТОНИАНА СЛАБО НЕУПОРЯДОЧЕННОЙ ТРЕХМЕРНОЙ МОДЕЛИ ИЗИНГА

В.Н. Бородихин, А.Н. Вакилов, В.В. Прудников

For the first time the critical parameters of the effective Hamiltonian for weakly disordered three-dimensional Ising model are determined by Monte Carlo simulation.

1. Введение

Исследование критического поведения неупорядоченных систем с замороженными дефектами структуры представляет большой теоретический и экспериментальный интерес. Это обусловлено тем, что большинство реальных твердых тел содержат замороженные дефекты структуры, присутствие которых влияет на характеристики систем и, в частности, может существенно сказываться на поведении систем при фазовых переходах. С другой стороны, центральной концепцией теории фазовых переходов и критических явлений является принцип универсальности, т.е. независимость термодинамических характеристик различных систем при фазовых переходах от различий в значениях мелкомасштабных параметров и разделение всех систем на небольшое число классов универсальности в зависимости от пространственной размерности системы и симметрии его параметра порядка. В случае неупорядоченных систем до сих пор остался невыясненным вопрос: являются ли такие характеристики критического поведения, как безразмерные амплитуды взаимодействия флуктуаций параметра порядка, и критические показатели универсальными, т.е. не зависящими от концентрации дефектов структуры вплоть до порога перколяции, или существует линия фиксированных точек для значений амплитуд взаимодействия, определяющая непрерывное изменение критических показателей с концентрацией.

Исследования показали [1], что присутствие замороженных точечных дефектов (например, примеси немагнитных атомов в ферро- или антиферромагнитных материалах) изменяет критические свойства систем, теплоемкость которых в однородном состоянии испытывает расходимость в критической точке с

Омский государственный университет
Исследования поддержаны грантами РФФИ (№00-02-16455) и Минобразования РФ (Е00-3.2-43).

показателем $\alpha > 0$. Данному критерию удовлетворяют только системы, эффективный гамильтониан которых вблизи критической точки изоморфен модели Изинга.

Теоретико-полевое описание критического поведения слабо неупорядоченной модели Изинга, проведенное непосредственно для трехмерных систем ($d = 3$) в высокопетлевых порядках приближения теории (5-ти петлевом [2] и 6-ти петлевом [3]) с применением методов суммирования получаемых асимптотических рядов, позволило с наибольшей доступной к настоящему времени точностью получить значения безразмерных амплитуд взаимодействия флуктуаций параметра порядка g_R и u_R в неподвижной фиксированной точке ренормгрупповых преобразований модели, задающих ее критические свойства. Так, в соответствии с [3] $g* = 36,72(32)$, $u* = 11,89(30)$. Данные значения отражают главную особенность критического поведения, характеризующегося аномально сильным взаимодействием флуктуаций параметра порядка. Поэтому возникает вопрос о сходимости рядов теории возмущения и насколько адекватно результаты применения методов суммирования могут соответствовать реальным критическим характеристикам системы. Вычисление критических параметров $g*$ и $u*$ эффективного гамильтониана непертурбативным способом представляет большой интерес.

Экспериментальные исследования неупорядоченных систем, таких как кристаллические смеси одноосных Изинго-подобных антиферромагнетиков (FeF_2, MnF_2) с немагнитными материалами (ZnF_2), показывают существенное отличие значений критических показателей для неупорядоченных систем от соответствующих показателей однородных систем. В частности, для материалов $Fe_xZn_{1-x}F_2$ были измерены значения критических показателей $\nu = 0,70(2)$, $\gamma = 1,34(6)$, а для материалов $Mn_xZn_{1-x}F_2$ $\nu = 0,715(35)$, $\gamma = 1,364(76)$. При этом многочисленные экспериментальные оценки демонстрируют независимость значений критических показателей от концентрации примесей (см. приведенные в [3] таблицы измеренных показателей из широкого ряда экспериментальных работ). Поскольку возможности теоретического подхода ограничены описанием слабо неупорядоченных систем, проверка данных результатов методом компьютерного моделирования имеет большое значение.

2. Модель

Известно, что исходная неупорядоченная модель в критической области термодинамически эквивалентна O(m) симметричной модели Гинзбурга- Ландау-Вильсона, определяемой эффективным гамильтонианом:

$$H = \int d^dx \left[\frac{1}{2}(\partial_\mu\varphi(x))^2 + \frac{m_0^2}{2}\varphi(x)^2 + \frac{V(x)}{2}\varphi^2 + \frac{g_0}{4!}\varphi(x)^4\right], \qquad (1)$$

где φ(x) – поле m-компонентного параметра порядка, $V(x)$ – примесный потенциал, $m_0^2 \sim T - T_c(p)$, $T_c(p)$ – критическая температура разбавленного магнетика, зависящая от концентрации спинов p, g_0 – положительная константа.

После применения процедуры репличного усреднения по гауссовски распределенному потенциалу случайного поля примесей гамильтониан принимает вид:

$$H_{repl} = \int d^dx \left[\frac{1}{2} \sum_{\alpha=1}^{n} (\partial_\mu \varphi_\alpha(x))^2 + \frac{m_0^2}{2} \sum_{\alpha=1}^{n} \varphi_\alpha(x)^2 + \frac{g_0}{4!} \sum_{\alpha=1}^{n} (\varphi_\alpha(x)^2)^2 - \right.$$
$$\left. - \frac{u_0}{4!} \left(\sum_{\alpha=1}^{n} (\varphi_\alpha(x))^2 \right)^2 \right], \qquad (2)$$

где индекс α нумерует реплики (образы) однородной составляющей в гамильтониане неупорядоченной модели (1), а дополнительная вершина u_0, возникшая в (2), задает эффективное взаимодействие флуктуаций параметра порядка через поле дефектов. В пределе $n \to 0$ данная модель термодинамически эквивалентна исходной неупорядоченной модели.

После применения процедуры перенормировки — усреднения по мелкомасштабным флуктуациям поля φ(x) и последующего масштабного преобразования, сопровождающегося перенормировкой поля параметра порядка в Z раз, устраняющей расходимость модели в критической точке, гамильтониан принимает вид:

$$H_R = \int d^dx \left[\frac{1}{2} \sum_{\alpha=1}^{n} (\partial_\mu \varphi_{\alpha R}(x))^2 + \frac{m_R^2}{2} \sum_{\alpha=1}^{n} \varphi_{\alpha R}(x)^2 + \frac{m_R g_R}{4!} \sum_{\alpha=1}^{n} (\varphi_{\alpha R}(x)^2)^2 - \right.$$
$$\left. - \frac{m_R u_R}{4!} \left(\sum_{\alpha=1}^{n} (\varphi_{\alpha R}(x))^2 \right)^2 \right], \qquad (3)$$

где индекс R обозначает перенормированные величины, $\varphi \sim Z_\varphi^{0.5} \varphi_R$, $Z_\varphi = \chi m_R^2$ — нормировочный множитель, χ – восприимчивость, $u_R = 3v\chi^2/\xi^d$, ξ – корреляционная длина, связанная с перенормированной массой соотношением $m_R^2 = 1/\xi^2$.

Процедура ренормгрупповых преобразований модели характеризуется наличием предельной неподвижной точки (фиксированной точки) в пространстве безразмерных амплитуд взаимодействия флуктуаций параметра порядка g_R и u_R, которая задает ее критические свойства и позволяет определить критические показатели для основных термодинамических и корреляционных функций системы. Как отмечалось выше, в шестипетлевом приближении [3] $g* = 36,72(32)$, $u* = 11,89(30)$.

3. Численная методика и результаты

Значения эффективных амплитуд взаимодействия флуктуаций могут быть получены методами компьютерного моделирования путем вычисления различных корреляционных функций или моментов функций распределения для параметра порядка. Монте-Карло-результаты в критической области для нетривиальной фиксированной точки однородной модели Изинга [4], [5] находятся в хорошем соответствии с результатами теоретико-полевого подхода.

В настоящей работе впервые методом Монте-Карло определены координаты фиксированной точки эффективного гамильтониана (3) для концентрации спинов $p = 0,95$.

Рассматривается трехмерная модель неупорядоченной спиновой системы в виде кубической решетки с размерами L с наложенными периодическими граничными условиями. С узлами решетки связаны спины σ_i, принимающие значения ± 1, и немагнитные атомы примеси (пустые узлы с $\sigma_i \equiv 0$). Данная система описывается гамильтонианом:

$$H = \frac{1}{2}\sum_{i,j} J_{ij} p_i p_j \sigma_i \sigma_j, \qquad (4)$$

где J_{ij} – константа обменного ферромагнитного взаимодействия, p_i – случайная переменная, описываемая функцией распределения

$$P(p_i) = p\delta(p_i - 1) + (1 - p_i)\delta(p_i), \qquad (5)$$

с $p = 1 - c$, где c – концентрация примеси.

Примесь равномерно распределяется по всей системе, и ее положение фиксируется для отдельной примесной конфигурации в процессе моделирования системы. Концентрация спинов определяется суммированием абсолютных значений спинов по всем узлам решетки

$$p = \frac{1}{L^3}\sum_{i=1}^{L^3} |\sigma_i|.$$

При создании спиновой конфигурации со случайно распределенными примесями в решетке возникают несвязанные геометрические кластеры узлов со спинами. При концентрации спинов p больших порога спиновой перколяции p_c в системе практически всегда существует спиновый кластер, характеризующийся общей связностью (протеканием) с одной грани решетки на другую грань, и какое-то количество изолированных кластеров, содержащих относительно небольшое число спинов. В пределе бесконечно большого размера решетки вклад в магнитные характеристики системы будут давать только скоррелированные спины бесконечного перколяционного кластера, поэтому будет разумным при вычислении критических характеристик не учитывать вклад от узлов, не имеющих связи с перколяционным кластером. Такая процедура позволяет уменьшить «шум» от спинов кластеров конечного размера.

Для распределения спинов с заданной концентрацией p по узлам решетки использовался алгоритм выращивания перколяционного кластера Хаммерсли-Лиса-Александровица [6]. Практические детали реализации данного алгоритма следующие. В центре кубической решетки размещается затравочный спин. Шесть соседних узлов образуют «периметр» затравочного спина. Случайным образом выбирается узел из «периметра». Затем с вероятностью p этот узел занимается спином, а его соседи добавляются в «периметр». В противном случае узел остается свободным (примесным). Чтобы узлы решетки оставались свободными с вероятностью $1 - p$, данный узел больше не проверяется. Если узел уже занят спином, то определяется, нет ли новых непроверенных узлов «периметра». Процедура повторяется до тех пор, пока не будут просмотрены все узлы периметра.

В основе компьютерного моделирования статистических процессов лежит метод Монте-Карло, суть которого заключается в использовании случайных чисел для машинной имитации вероятностных распределений. В данной работе для получения последовательных спиновых конфигураций был применен однокластерный алгоритм Вольфа, который по сравнению с традиционным для метода Монте-Карло алгоритмом Метрополиса позволяет получать значительно менее скоррелированную последовательность спиновых конфигураций. Известно, что время корреляции между двумя состояниями системы с линейным размером L вблизи критической температуры ведет себя как $\tau \simeq L^z$. При этом для алгоритма Метрополиса показатель $z_{Metropolis} \simeq 2$, а для алгоритма Вольфа $z_{Wolf} \simeq 0,5$.

Примененный в работе вариант алгоритма Вольфа состоял в следующем. В системе случайным образом выбирался спин и переворачивался. Затем рассматривались ближайшие соседи спина, и если они сонаправлены с этим «центральным» спином (неперевернутым), то с вероятностью $1 - \exp(-2\beta)$, где $\beta = 1/T$ они также переворачивались, а их координаты заносились в стек. После того, как были проверены соседние спины, спин, координаты которого были занесены в стек последними, выбирается «центральным», и вся процедура затем повторяется. Процесс повторяется до тех пор, пока стек не окажется пустым, что соответствует полному перевороту кластера Вольфа. По данному алгоритму реализуется марковский процесс и с соответствующей вероятностью генерируются конфигурации спинов.

Для уменьшения корреляций спиновых конфигураций вычисление намагниченности и других термодинамических величин осуществлялось через три переворота кластера Вольфа, что условно можно назвать одним Монте-Карло-шагом.

В самом начале процесса все спины полагались сонаправленными (что соответствует состоянию системы при $T = 0$). Процедуре установления термодинамического равновесия в системе, соответствующего температуре T, отводилось 10^4 шагов Монте-Карло.

Поскольку моделируемая система являлась неупорядоченной, кроме усреднения по спиновым конфигурациям проводилось усреднение по различным примесным конфигурациям. В данной работе использовалось 100 примесных конфигураций. Для проведения статистического усреднения каждой примесной конфигурации сопоставлялось 10^5 спиновых конфигураций или Монте-Карло-шагов.

Для каждой примесной конфигурации вычислялись корреляционная длина ξ и восприимчивость χ по формулам [7]:

$$\xi = \frac{1}{2\sin(\pi/L)}\sqrt{\frac{\chi}{F} - 1}, \tag{6}$$

где L – размер системы, $\chi = \frac{1}{V}\langle\sigma^2\rangle$, $F = \frac{1}{V}\langle\Phi\rangle$,

$$\Phi = \frac{1}{3}\left(\left|\sum_x \sigma_x \exp\left(\frac{2\pi i x_1}{L}\right)\right|^2 + \left|\sum_x \sigma_x \exp\left(\frac{2\pi i x_2}{L}\right)\right|^2 + \right. \tag{7}$$

$$\left. + \left|\sum_x \sigma_x \exp\left(\frac{2\pi i x_3}{L}\right)\right|^2\right), \tag{8}$$

$x : (x_l, x_2, x_3)$ с последующим усреднением по всем примесным конфигурациям.

Эффективные амплитуды взаимодействия флуктуаций параметра порядка u_R и g_R в гамильтониане (3) вычислялись при усреднении по всей совокупности примесных конфигураций следующих выражений:

$$g_R = 3\left(\frac{L}{\xi}\right)^3\left(1 - \frac{\overline{\langle\sigma^4\rangle}}{3\overline{\langle\sigma^2\rangle}^2}\right), \tag{9}$$

$$u_R = \frac{3\chi^2}{\xi^d}\frac{\overline{\langle\sigma_\alpha^2\sigma_\beta^2\rangle} - \overline{\langle\sigma^2\rangle}^2}{\overline{\langle\sigma^2\rangle}^2}, \tag{10}$$

где $\langle ... \rangle$ означает статистическое усреднение по шагам Монте-Карло, а черта – усреднение по примесным конфигурациям. Индексы α и β характеризуют спиновые конфигурации для различных реплик неупорядоченной системы размера L, моделируемых одновременно при одной и той же температуре и отличающихся различными начальными конфигурациями.

Для системы со спиновой концентрацией $p = 0,95$ измерения проводились при температурах $T = 4,275; 4,285; 4,295; 4,315; 4,335$ для решеток с различными L. Мы использовали значение критической температуры фазового перехода в решетке бесконечного размера $T_c = 4,2571$, найденное в работе [8]. В табл. 1 приведены найденные значения корреляционной длины ξ, восприимчивости χ и вершин g и u при всех вышеперечисленных температурах для решеток с размерами L, указанными в таблицах.

Из приведенных в таблице данных видно, что чем дальше температура системы отстоит от T_c, тем для меньших размеров решетки L измеряемые величины достигают своего максимального асимптотического значения. Так, для ближайшей к критической температуре $T = 4,275$ $L_{max} = 100$, тогда как для $T = 4,335$ $L_{max} = 45$. Это связано с уменьшением флуктуаций параметра порядка и определяющего флуктуационного вклада в измеряемые величины при удалении от критической температуры.

В конечной системе не может проявиться настоящий фазовый переход. Тем не менее можно ожидать, что если $\xi(T)$ меньше линейного размера L системы, то конечная система будет правильно передавать свойства бесконечной системы при применении методики конечно-мерного скейлинга [5], которая позволяет осуществить экстраполяцию результатов для решетки размера L на систему бесконечного размера для каждой отдельной температуры. Алгоритм процедуры

Таблица 1. Зависимость физических величин от размеров системы L для различных температур

T	L	ξ	χ	g	u
4,335	45	$6,58 \pm 0,02$	$203,9 \pm 0,36$	$30,531 \pm 2,98$	$2,157 \pm 1,54$
	40	$6,4 \pm 0,026$	$194,67 \pm 0,414$	$29,93 \pm 2,74$	$3,1 \pm 1,83$
	30	$6,241 \pm 0,036$	$189,134 \pm 0,57$	$29,31 \pm 2,56$	$6,316 \pm 2,65$
	20	$5,753 \pm 0,005$	$169,32 \pm 0,78$	$25,2 \pm 1,42$	$15,335 \pm 4,16$
4,315	50	$8,08 \pm 0,027$	$304,563 \pm 0,52$	$30,668 \pm 3,18$	$2,431 \pm 1,8$
	45	$7,9 \pm 0,03$	$290,74 \pm 0,567$	$30,087 \pm 2,86$	$3,093 \pm 2,007$
	40	$7,554 \pm 0,03$	$271,176 \pm 0,55$	$29,72 \pm 2,36$	$3,37 \pm 2,09$
	30	$7,305 \pm 0,045$	$262,26 \pm 0,84$	$28,175 \pm 1,64$	$8,58 \pm 3,4$
	20	$6,45 \pm 0,052$	$213,736 \pm 0,91$	$22,84 \pm 0,94$	$14,82 \pm 4,386$
4,295	65	$10,646 \pm 0,04$	$522,64 \pm 0,99$	$30,72 \pm 3,53$	$3,94 \pm 2,7$
	60	$10,468 \pm 0,007$	$509,02 \pm 0,68$	$30,528 \pm 2,02$	$6,031 \pm 1,81$
	50	$10,322 \pm 0,009$	$503,635 \pm 0,86$	$29,515 \pm 1,46$	$10,02 \pm 2,35$
	40	$9,978 \pm 0,01$	$482,76 \pm 1,05$	$27,439 \pm 0,96$	$16,61 \pm 2,915$
	30	$9,21 \pm 0,011$	$426,74 \pm 1,16$	$23,513 \pm 0,59$	$26,04 \pm 3,66$
	20	$7,547 \pm 0,01$	$300,19 \pm 0,88$	$18,583 \pm 0,3$	$27,29 \pm 3,7$
4,285	80	$13,319 \pm 0,06$	$809,29 \pm 1,794$	$30,544 \pm 3,9$	$6,486 \pm 3,68$
	70	$13,133 \pm 0,072$	$803,214 \pm 2,24$	$29,795 \pm 3,24$	$9,51 \pm 4,77$
	60	$13,049 \pm 0,05$	$800,37 \pm 2,453$	$29,335 \pm 2,21$	$12,77 \pm 5,3$
	50	$12,626 \pm 0,1$	$766,61 \pm 3,237$	$27,006 \pm 1,94$	$24,76 \pm 7,42$
	40	$11,868 \pm 0,11$	$696,263 \pm 3,46$	$24,005 \pm 1,38$	$34,02 \pm 8,7$
	30	$10,685 \pm 0,1$	$585,66 \pm 3,01$	$19,887 \pm 0,85$	$36,94 \pm 8,93$
4,275	100	$19,034 \pm 0,091$	$1652,31 \pm 3,934$	$31,452 \pm 2,68$	$10,667 \pm 5,655$
	90	$18,8 \pm 0,09$	$1618,31 \pm 4,9$	$29,54 \pm 2,31$	$11,019 \pm 5,73$
	80	$18,6 \pm 0,048$	$1612,87 \pm 4,25$	$28,46 \pm 2,35$	$16,83 \pm 6,49$
	70	$18,046 \pm 0,12$	$1539,91 \pm 5,175$	$27,29 \pm 1,49$	$21,68 \pm 8,22$
	60	$17,488 \pm 0,02$	$1483,25 \pm 3,77$	$24,925 \pm 2,86$	$39,81 \pm 6,27$
	50	$16,54 \pm 0,15$	$1361,27 \pm 6,825$	$21,633 \pm 1,1$	$48,78 \pm 12,42$
	40	$14,854 \pm 0,146$	$1129,97 \pm 6,01$	$18,72 \pm 0,897$	$52,783 \pm 12,5$

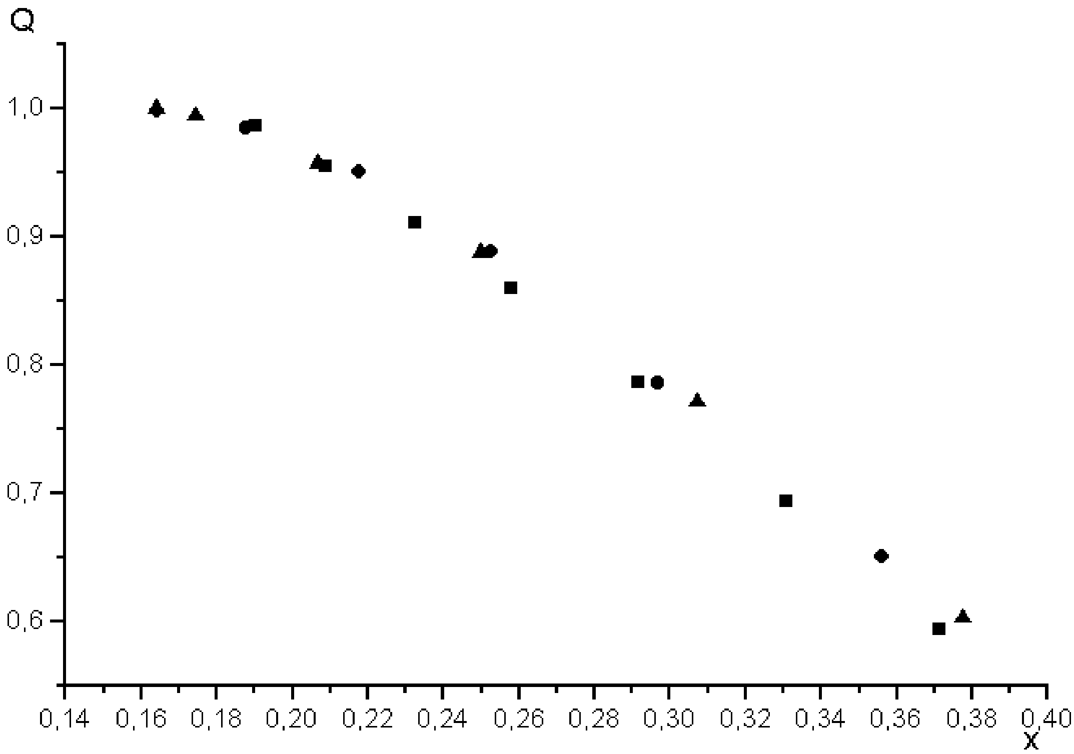


Рис. 1. Зависимость функции Q от $x = \xi/L$ для разных температур. Квадратные значки соответствуют $T = 4,275$, круглые $T = 4,285$, треугольные $T = 4,295$.

конечно-мерного скейлинга состоит в следущем. Для заданной температуры T измеряется значение величины $A_{L,T}$ и $x = \xi_{L,T}/L$ при увеличивающемся L. Это позволяет определить значение L_{max}, при котором величина $A_{L,T}$ становится не зависящей от размеров системы. Зная L_{max}, полученные значения $A_{L,T}(x)$ для каждого L можно аппроксимировать с помощью специально подобранной функции $Q(x)$ так, чтобы они не зависели от размеров системы, по формуле $A(T) = A_L(T)/Q(x(L,T))$. В соответствии с рекомендациями работы [5] функция $Q(x)$ была выбрана в виде $Q(x) = 1+c_1x+c_2x^2+c_3x^3+c_4x^4$ с подбираемыми коэффициентами c_i по методу наименьших квадратов. Функция $Q(x)$ характеризуется универсальностью для значений температур в критической области. Это позволяет использовать ее для нахождения физических величин при температурах T, столь близких к T_c, при которых L_{max} может быть и недостижима, но известны величины $A_{L,T}(x)$ и $x = \xi_{L,T}/L$ при $L < L_{max}$. Графики функций $Q(x)$ (рис.1), построенные для различных температур, демонстрируют эту универсальность.

В таблице 2 приведены результирующие зависимости значений корреляционной длины ξ, восприимчивости χ и вершин g и u от температуры, найденные

Таблица 2. Зависимость физических величин от температур в критической области

T	ξ	χ	g	u
4,275	$19,034 \pm 0,1$	$1652,26 \pm 5,73$	$31,455 \pm 2,3$	$10,706 \pm 3,766$
4,285	$13,31 \pm 0,09$	$807,558 \pm 2,997$	$30,549 \pm 2,46$	$6,514 \pm 2,485$
4,295	$10,664 \pm 0,01$	$529,02 \pm 1,132$	$30,72 \pm 1,574$	$4,125 \pm 1,135$
4,315	$8,09 \pm 0,04$	$304,62 \pm 0,8$	$30,65 \pm 2,31$	$2,438 \pm 1,32$
4,335	$6,59 \pm 0,036$	$204,05 \pm 0,59$	$30,529 \pm 2,54$	$2,18 \pm 1,09$

с помощью процедуры конечно-мерного скейлинга.

Из выявленной температурной зависимости перенормированных вершины g_R и u_R могут быть выделены их критические значения $g*, u*$ на основе скейлинговых зависимостей данных вершин

$$g_R(t) = g * (1 + at^{\theta}), \tag{11}$$

$$u_R(t) = u * (1 + bt^{\theta}) \tag{12}$$

с $t = T - T_c(p)$, $\theta = \nu\omega$, задаваемых ν – критическим показателем для корреляционной длины и ω – критическим показателем, характеризующим поправки к скейлингу. В соответствии с результатами работы [3] $\theta = 0,17(10)$. Проведенная аппроксимация g_R и u_R по выражениям (11), (12) позволила получить следующие значения критических амплитуд взаимодействия флуктуаций параметра порядка для неупорядоченной системы с $p = 0,95$ в фиксированной точке $g* = 31,2382 \pm 2,273$, $u* = 9,874 \pm 3,41$. Сопоставление полученных значений с теоретико-полевыми указывает на их хорошее согласие в пределах статистических погрешностей результатов.

Из данных по температурной зависимости восприимчивости и корреляционной длины нами были выделены значения критических показателей для соответствующих величин: $\gamma = 1,40 \pm 0,04$, $\nu = 0,71 \pm 0,02$. Сопоставление полученных значений с теоретико-полевыми значениями показателей [3] $\gamma = 1,342(10)$ и $\nu = 0,6837(53)$ показывает, что полученные нами значения являются слегка завышенными, хотя и находятся в достаточно хорошем соответствии со значениями показателей, найденными в экспериментальных исследованиях (см. [3]).

В заключении отметим, что распространение развитой в настоящей работе методики на исследование критического поведения неупорядоченных систем с большими значениями концентрации примесей и выделение для них координат фиксированной точки, а также определение значений критических показателей позволит ответить на фундаментальный вопрос об универсальности критического поведения неупорядоченных систем.

Литература

1. Harris A.B. *Effect of random defects on the critical behaviour of Ising models* // J. Phys. C. 1974. V.7, N6. P.1671-1692.
2. Pakhnin D.V., Sokolov A.I. *Five-loop renormalization-group expansion for the three-dimensional n-vector cubic model and critical exponents for impure Ising sistems* // Phys. Rev. B. 2000. V.61. P.15130.
3. Pelissetto A., Vicari E. *Randomly dilute spin models: a six-loop field-theoretic study* // Phys. Rev. B. 2000. V.62. P.6393.
4. Tsypin M.M. *Effective potential for scalar field in three dimensions: Ising models in the ferromagnetic phase* // Phys. Rev. B. 1997. V.55. P.8911.
5. Kim J.-K. *Critical renormalized coupling constants in the symmetric phase of the Ising models* // J. Phys. A: Math. Gen.2000. V.33. P.2675.
6. Гулд Х., Тобочник Я.К. *Компьютерное моделирование в физике: В 2 ч.* М.: Наука, 1989.
7. Salas J. Sokal A. *Exact Finite-Size-Scaling corrections to the critical two-dimensional Ising model on a torus* // J. Stat. Phys. 2000. V.98. P.551.
8. Heuer H.-O. *Monte Carlo simulation of disrdered 3 - dimensional Ising systems* // Europhys. Lett. 1990. V.12. P.551.

Математические структуры и моделирование
2001, вып. 8, с. 66–75

УДК 530.12:531.51

СПИНОРНЫЕ ДУХИ, ТЕНЕВЫЕ ЭЛЕКТРОНЫ И МУЛЬТИВЕРС ДОЙЧА

Е.В. Палешева

In this article the new solution of the Einstein-Dirac's equations is presented. There are ghost spinors, i.e. the stress-energy tensor equal to zero and the current of these fields is non-zero vector. Last the ghost neutrinos was found. These spinors and shadow particles of Deutsch are identified. And in result the ghost spinors have a phisical interpretation and shadow electrons as another shadow particles the field equations are got.

Введение

В общей теории относительности в случае равенства нулю правых частей уравнений Эйнштейна без космологической постоянной говорят о пустом пространстве. В данном случае под пустотой пространства понимается отсутствие в нём какой-либо материи. Но возникает вопрос: если как раз последней и порождается гравитационное поле, то откуда оно появляется при её исчезновении? Ответ на него достаточно прост: дело в том, что следует различать такие два, на первый взгляд, взаимозаменяющих друг друга понятия, как вещество и материя, под которой в первую очередь необходимо понимать нечто порождающее данную структуру мира. При этом материя не обязана иметь какой бы то ни было энергии, что как раз и происходит при занулении тензора энергии-импульса. Наличие же вещества, наоборот, характеризуется тензором энергии-импульса, отличным от нуля. Фактическим примером материи, которая в нашем мире проявляет свойство объекта с нулевыми энергией и импульсом, т.е. в предложенной терминологии не является веществом, служат спинорные духи. Об их существовании можно говорить в свете результатов, полученных в данной статье, а также в работах [3, 4, 6, 7, 9].

Окружающий нас мир таит в себе множество загадок, и мы всё время хотим понять природу пространства, в котором протекает наша жизнь. Так и Дэвид Дойч делает попытку логически согласованно объяснить явление интерференции квантовых частиц и приходит к выводу о существовании параллельных миров, во всей своей совокупности представляющих мультиверс [9]. Точнее говоря, предположение о наличии в нашем пространстве-времени теневых фотонов, отождествлённых им с фотонами иной вселенной, приводит его к законченному обоснованию появления соответствующей интерференционной

E-mail: m82palesheva@math.omsu.omskreg.ru
Омский государственный университет

картины. Но вообще-то в данном случае теневые частицы обладают свойством тел с нулевым тензором энергии-импульса – это непосредственно следует из хорошо известных опытов – и поэтому их существование должно быть физически обоснованно. Результаты, описанные в статье, а именно проведённая параллель между духами квантовых частиц и соответствующими теневыми частицами, в точности приводят к такому обоснованию.

Ранее уже были найдены решения уравнений Эйнштейна-Дирака, соответствующие нейтринным духам в случае гравитационного поля с плоской симметрией [3, 6], с цилиндрической симметрией [4], а также в случае волнового гравитационного поля [7]. При этом пространство-время является искривлённым. В [7] указано кроме этого на наличие нейтринных духов и в плоском протранстве-времени. Пространство-время, рассмотренное в этой работе, также является плоским.

1. Описание геометрии пространства-времени и соответствующих спинорных полей

Рассмотрим самогравитирующую спинорную материю, описываемую системой уравнений Эйнштейна-Дирака [5, с.142]:

$$R_{ik} - \frac{1}{2} g_{ik} R = \kappa T_{ik}, \tag{1}$$

$$i\hbar\gamma^k \left(\frac{\partial\psi}{\partial x^k} - \Gamma_k \psi \right) - mc\psi = 0. \tag{2}$$

Гравитационное поле задаётся метрикой

$$ds^2 = dx^{0\,2} + 2e^{x^0} dx^0 dx^3 - dx^{1\,2} - dx^{2\,2}, \tag{3}$$

которая является плоской, так как соответствующий тензор Римана R^i_{klm} равен нулю. Поэтому зануляется и тензор энергии-импульса, который в случае спинорной материи представляется системой уравнений [1, с.381]

$$T_{ik} = \frac{i\hbar c}{4} \left\{ \psi^* \gamma^{(0)} \gamma_i \left(\frac{\partial\psi}{\partial x^k} - \Gamma_k \psi \right) - \left(\frac{\partial\psi^*}{\partial x^k} \gamma^{(0)} + \psi^* \gamma^{(0)} \Gamma_k \right) \gamma_i \psi + \right.$$

$$\left. + \psi^* \gamma^{(0)} \gamma_k \left(\frac{\partial\psi}{\partial x^i} - \Gamma_i \psi \right) - \left(\frac{\partial\psi^*}{\partial x^i} \gamma^{(0)} + \psi^* \gamma^{(0)} \Gamma_i \right) \gamma_k \psi \right\}, \tag{4}$$

где ψ – биспинор, а символ * означает эрмитово сопряжение. Кроме этого,

$$\Gamma_k = \frac{1}{4} g_{ml} \left(\frac{\partial \lambda_r^{(s)}}{\partial x^k} \lambda^l_{(s)} - \Gamma^l_{rk} \right) s^{mr},$$

$$s^{mr} = \frac{1}{2} \left(\gamma^m \gamma^r - \gamma^r \gamma^m \right), \quad \gamma^k \equiv \lambda^k_{(i)} \gamma^{(i)},$$

здесь $\lambda^k_{(i)}$ – i-й вектор тетрады, а $\gamma^{(i)}$ – четырёхрядные матрицы Дирака, для которых имеется следующее представление, записанное с помощью двухрядных матриц Паули [8, с.73]:

$$\gamma^{(0)} = \begin{bmatrix} I & 0 \\ 0 & -I \end{bmatrix}, \quad \gamma^{(\alpha)} = \begin{bmatrix} 0 & \sigma_\alpha \\ -\sigma_\alpha & 0 \end{bmatrix},$$

$$\sigma_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \sigma_2 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \sigma_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Метрический тензор пространства-времени может быть выражен через тетраду векторов следующим образом [2, с.373]:

$$ds^2 = \eta_{ab} \left(\lambda^{(a)}_i dx^i \right) \left(\lambda^{(b)}_k dx^k \right).$$

В нашем случае положим

$$\eta_{ab} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}.$$

Тогда для поля (3) находим

$$\begin{aligned} \lambda^{(0)}_i &= (1,0,0,e^{x^0}), \quad \lambda^i_{(0)} = (1,0,0,0), \\ \lambda^{(1)}_i &= (0,1,0,0), \qquad \lambda^i_{(1)} = (0,1,0,0), \\ \lambda^{(2)}_i &= (0,0,1,0), \qquad \lambda^i_{(2)} = (0,0,1,0), \\ \lambda^{(3)}_i &= (0,0,0,e^{x^0}), \quad \lambda^i_{(3)} = (-1,0,0,e^{-x^0}). \end{aligned}$$

В итоге, используя матрицы Дирака, вычисляем

$$\gamma^0 = \begin{bmatrix} I & -\sigma_3 \\ \sigma_3 & -I \end{bmatrix}, \gamma^1 = \begin{bmatrix} 0 & \sigma_1 \\ -\sigma_1 & 0 \end{bmatrix}, \gamma^2 = \begin{bmatrix} 0 & \sigma_2 \\ -\sigma_2 & 0 \end{bmatrix}, \gamma^3 = e^{-x^0} \begin{bmatrix} 0 & \sigma_3 \\ -\sigma_3 & 0 \end{bmatrix}$$

и

$$\gamma_0 = \begin{bmatrix} I & 0 \\ 0 & -I \end{bmatrix}, \gamma_1 = \begin{bmatrix} 0 & -\sigma_1 \\ \sigma_1 & 0 \end{bmatrix}, \gamma_2 = \begin{bmatrix} 0 & -\sigma_2 \\ \sigma_2 & 0 \end{bmatrix}, \gamma_3 = e^{x^0} \begin{bmatrix} I & -\sigma_3 \\ \sigma_3 & -I \end{bmatrix}.$$

При этом

$$s^{01} = \begin{bmatrix} i\sigma_2 & \sigma_1 \\ \sigma_1 & i\sigma_2 \end{bmatrix}, s^{02} = \begin{bmatrix} -i\sigma_1 & \sigma_2 \\ \sigma_2 & -i\sigma_1 \end{bmatrix}, s^{03} = e^{-x^0} \begin{bmatrix} 0 & \sigma_3 \\ \sigma_3 & 0 \end{bmatrix},$$

$$s^{12} = \begin{bmatrix} -i\sigma_3 & 0 \\ 0 & -i\sigma_3 \end{bmatrix}, s^{13} = e^{-x^0} \begin{bmatrix} i\sigma_2 & 0 \\ 0 & i\sigma_2 \end{bmatrix}, s^{23} = e^{-x^0} \begin{bmatrix} -i\sigma_1 & 0 \\ 0 & -i\sigma_1 \end{bmatrix}.$$

Тогда получаем, что $\Gamma_1 = \Gamma_2 = \Gamma_3 = 0$ и

$$\Gamma_0 = \frac{1}{2} \begin{bmatrix} 0 & \sigma_3 \\ \sigma_3 & 0 \end{bmatrix}.$$

2. Спинорные духи

Как уже говорилось, в настоящее время известен ряд решений уравнений Эйнштейна-Дирака с нулевыми массой и тензором энергии-импульса. В этом случае говорят о нейтринных духах. Вообще говоря, уравнение Дирака (2) является уравнением поля для спинорной материи, при этом постоянная m интерпретируется как масса частицы. А если рассмотреть решения соответствующих уравнений с таким же условием на тензор энергии-импульса, но при этом положить $m \neq 0$? Оказывается, такие решения существуют. Но в этом случае мы должны внести некоторые уточнения относительно введённой Дираком в уравнения (2) постоянной m. Этот вопрос будет рассмотрен ниже, а пока что будем полагать, что m не является массой в привычном для нас смысле этого слова. Это всего лишь некоторая постоянная, характеризующая рассматриваемые частицы.

2.1. Нейтринные духи

Будем рассматривать случай, в котором спинорное поле описывает нейтрино, т.е. тогда в уравнении (2) мы должны положить $m = 0$. При этом считаем, что

$$\frac{\partial\psi}{\partial x^1} = \frac{\partial\psi}{\partial x^2} = \frac{\partial\psi}{\partial x^3} = 0, \quad \frac{\partial\psi}{\partial x^0} = \alpha\psi,$$

где α – вещественная константа. Например, такому условию удовлетворяет следующий спинор:

$$\psi = \begin{bmatrix} u_0 \\ u_1 \\ u_2 \\ u_3 \end{bmatrix} e^{\alpha x^0 + \beta}, \tag{5}$$

здесь u_i, β – произвольные комплексные постоянные. Тогда уравнение Дирака (2) в нашем случае можно записать в виде

$$\left(\gamma^0 - \gamma^0 \Gamma_0\right) \psi = 0.$$

Или, вычисляя, получаем условие, представленное в матричной форме:

$$\begin{bmatrix} I & -\sigma_3 \\ \sigma_3 & -I \end{bmatrix} \psi = 0.$$

Отсюда находим, что

$$u_0 = u_2, \quad u_1 = -u_3. \tag{6}$$

Для тензора энергии-импульса (4) с учётом условий, наложенных на спинорное поле ψ, имеем:

$$
\begin{aligned}
T_{00} &= -\frac{i\hbar c}{2}\alpha\psi^*\gamma^{(0)}\left\{\gamma_0\Gamma_0+\Gamma_0\gamma_0\right\}\psi\\
T_{01} &= -\frac{i\hbar c}{4}\alpha\psi^*\gamma^{(0)}\left\{\gamma_1\Gamma_0+\Gamma_0\gamma_1\right\}\psi\\
T_{02} &= -\frac{i\hbar c}{4}\alpha\psi^*\gamma^{(0)}\left\{\gamma_2\Gamma_0+\Gamma_0\gamma_2\right\}\psi\\
T_{03} &= -\frac{i\hbar c}{4}\alpha\psi^*\gamma^{(0)}\left\{\gamma_3\Gamma_0+\Gamma_0\gamma_3\right\}\psi\\
T_{11} &= T_{12}=T_{13}=T_{22}=T_{23}=T_{33}=0.
\end{aligned}
$$

После вычислений получаем:

$$
\begin{aligned}
T_{00} &= T_{03}=0\\
T_{01} &= -\frac{i\hbar c}{4}\alpha\left(\bar{u_0},\bar{u_1},-\bar{u_2},-\bar{u_3}\right)\begin{bmatrix}-i\sigma_2 & 0\\ 0 & i\sigma_2\end{bmatrix}\psi,\\
T_{02} &= -\frac{i\hbar c}{4}\alpha\left(\bar{u_0},\bar{u_1},-\bar{u_2},-\bar{u_3}\right)\begin{bmatrix}-i\sigma_1 & 0\\ 0 & i\sigma_1\end{bmatrix}\psi.
\end{aligned}
\tag{7}
$$

В результате, подставляя в (7) спинор (5), удовлетворяющий условию (6), находим, что $T_{01}=T_{02}=0$.

Таким образом, для рассмотренного спинорного поля, описывающего нейтрино, $T_{ik}\equiv 0$, т.е. получено решение уравнений Эйнштейна-Дирака, соответствующее нейтринному духу, потому что ему отвечает ненулевой 4-ток:

$$
j^{(k)}=\left(2({u_0}^2+{u_1}^2)e^{2\alpha x^0+2\beta},0,0,2({u_0}^2+{u_1}^2)e^{2\alpha x^0+2\beta}\right),
$$

который, как известно, вычисляется по формуле:

$$
j^{(k)}=\lambda_i^{(k)}\psi^*\gamma^{(0)}\gamma^i\psi. \tag{8}
$$

2.2. Спинорные духи

Ввиду того что $\Gamma_1=\Gamma_2=\Gamma_3=0$, уравнение Дирака (2) принимает вид:

$$
i\hbar\left(\gamma^k\frac{\partial\psi}{\partial x^k}-\gamma^0\Gamma_0\psi\right)-mc\psi=0, \tag{9}
$$

или после некоторых преобразований получаем

$$
\begin{bmatrix}I & -\sigma_3\\ \sigma_3 & -I\end{bmatrix}\frac{\partial\psi}{\partial x^0}+\begin{bmatrix}0 & \sigma_1\\ -\sigma_1 & 0\end{bmatrix}\frac{\partial\psi}{\partial x^1}+\begin{bmatrix}0 & \sigma_2\\ -\sigma_2 & 0\end{bmatrix}\frac{\partial\psi}{\partial x^2}+e^{-x^0}\begin{bmatrix}0 & \sigma_3\\ -\sigma_3 & 0\end{bmatrix}\frac{\partial\psi}{\partial x^3}-
$$

$$
-\frac{1}{2}\begin{bmatrix}-I & \sigma_3\\ -\sigma_3 & I\end{bmatrix}\psi=-i\frac{mc}{\hbar}\psi.
$$

В результате, произведя ряд несложных вычислений, приходим к системе следующего вида:

$$\begin{cases} u_{3,1} + u_{1,1} - i(u_{3,2} + u_{1,2}) + e^{-x^0}(u_{2,3} + u_{1,3}) = -i\dfrac{mc}{\hbar}(u_0 - u_2) \\ u_{0,0} - u_{2,0} + iu_{1,2} - u_{1,1} - e^{-x^0}u_{1,3} + \dfrac{1}{2}u_0 - \dfrac{1}{2}u_2 = -i\dfrac{mc}{\hbar}u_2 \\ u_{2,1} - u_{0,1} + i(u_{2,2} - u_{0,2}) - e^{-x^0}(u_{3,3} - u_{0,3}) = -i\dfrac{mc}{\hbar}(u_1 + u_3) \\ -u_{1,0} - u_{3,0} - iu_{0,2} - u_{0,1} + e^{-x^0}u_{0,3} - \dfrac{1}{2}u_1 - \dfrac{1}{2}u_3 = -i\dfrac{mc}{\hbar}u_3, \end{cases}$$

где $u_{i,k}$ означает дифференцирование по k. Теперь предположим, что $u_0 = u_2$, $u_1 = -u_3$, тогда остаётся условие

$$\begin{cases} (u_2 + u_1)_{,3} = 0 \\ -u_{1,1} + iu_{1,2} - e^{-x^0}u_{1,3} = -i\dfrac{mc}{\hbar}u_2 \\ (u_3 - u_0)_{,3} = 0 \\ -u_{0,1} - iu_{0,2} + e^{-x^0}u_{0,3} = -i\dfrac{mc}{\hbar}u_3. \end{cases}$$

Далее, накладывая дополнительные ограничения таким образом, чтобы $u_0 = -u_1 = u_2 = u_3 = u$, после некоторых преобразований находим:

$$\begin{cases} \dfrac{\partial u}{\partial x^1} = 0 \\ -i\dfrac{\partial u}{\partial x^2} + e^{-x^0}\dfrac{\partial u}{\partial x^3} = -i\dfrac{mc}{\hbar}u. \end{cases}$$

Полагая теперь $\partial u/\partial x^3 = 0$, получаем, что

$$\frac{\partial u}{\partial x^2} = \frac{mc}{\hbar}u,$$

а на $\partial u/\partial x^0$ никаких ограничений нет. При этом

$$u = e^{\frac{mc}{\hbar}x^2 + \alpha(x^0)},$$

или в общем виде имеем следующее спинорное поле:

$$\psi = \begin{bmatrix} 1 \\ -1 \\ 1 \\ 1 \end{bmatrix} e^{\frac{mc}{\hbar}x^2 + \alpha(x^0)}. \tag{10}$$

Теперь хотим, чтобы спинор (10) оказался духом, а это эквивалентно условию $T_{ik} \equiv 0$. После дополнительных предположений о том, что $Im\,\alpha(x^0) = 0$, где функция α из (10), получаем зануление тензора энергии-импульса рассмотренного спинорного поля. Таким образом, поле (10) такое, что $Im\,\alpha(x^0) = 0$, описывает спинорных духов, если же $m = 0$, то это нейтринные духи, являющиеся дополнением к уже найденным в предыдущем пункте.

Подставляя полученные результаты в (8), находим, что плотность тока отлична от нуля:

$$j^{(k)} = \left(4e^{2\frac{mc}{\hbar}x^2+2\alpha(x^0)}, 0, 0, 4e^{2\frac{mc}{\hbar}x^2+2\alpha(x^0)}\right),$$

следовательно, в пространстве (3) существуют потоки частиц с нулевыми энергией и импульсом.

3. Спинорные духи в пространстве-времени Минковского

В описанных выше результатах говорилось о существовании спинорных духов в плоском пространстве-времени, неевклидова структура которого вызвана неинерциальностью системы отсчёта. Но будут ли спинорные духи существовать и в пространстве Минковского? Возьмём соответствующую метрику пространства в виде:

$$ds^2 = 2dx^0dx^1 - dx^{2^2} - dx^{3^2}, \tag{11}$$

она является частным случаем пространства, рассмотренного в работе [7], в которой говорилось о существовании нейтринных духов для данного гравитационного поля. Используя результаты работы [7], можно заключить, что для зануления тензора энергии-импульса спинорной материи в приведённом пространстве достаточно наложить следующие ограничения:

$$u_0 = -u_3,\ u_1 = -u_2,\ u_k = \bar{u}_k. \tag{12}$$

При этом уравнения Дирака (2) для метрики (11) преобразуются к выражению:

$$\gamma^k \frac{\partial\psi}{\partial x^k} = -i\frac{mc}{\hbar}\psi.$$

Положив дополнительно к условиям (12)

$$\frac{\partial\psi}{\partial x^1} = \frac{\partial\psi}{\partial x^3} = 0,$$

находим следующий спинор:

$$\psi = \begin{bmatrix} e^{-\frac{mc}{\hbar}x^2+\alpha(x^0)} \\ e^{\frac{mc}{\hbar}x^2+\alpha(x^0)} \\ -e^{\frac{mc}{\hbar}x^2+\alpha(x^0)} \\ -e^{-\frac{mc}{\hbar}x^2+\alpha(x^0)} \end{bmatrix}, \tag{13}$$

где, как и раньше, $Im\,\alpha(x^0) = 0$. При этом ток имеет выражение

$$j^{(k)} = \left(4e^{2\alpha(x^0)}\cosh\left(2\frac{mc}{\hbar}x^2\right), -4e^{2\alpha(x^0)}, 0, 4e^{2\alpha(x^0)}\sinh\left(2\frac{mc}{\hbar}x^2\right)\right).$$

4. Параллельные вселенные

В работе [9] описывается ряд экспериментов, в результате которых делается попытка понять природу интерференции света. Ключевым пунктом объяснения полученных экспериментальных данных является предположение о существовании теневых фотонов, откуда Дойч приходит к выводу о разбиении единого мультиверса на совокупность различных реальностей, или, другими словами, параллельных миров. Логически построенная последовательность рассуждений и соответственно выводов, именно относительно наличия множества параллельных вселенных, является чётко слаженной. Хотя в общей сложности есть и некоторая необоснованность в сформулированных результатах относительно взаимодействия миров между собой. В частности сложно согласиться с выводом о взаимодействии частицы лишь со своими собственными теневыми частицами. В действительности это всего лишь желаемое предположение, которое ниоткуда не следует. Кроме этого есть ещё один минус в предложенном объяснении: если теневой фотон воздействует на реальный, причём таким образом, что данное воздействие отражается на результатах эксперимента – а именно на интерференционной картине, причём непосредственным образом – то должны быть уравнения, описывающие это взаимодействие. Да и предположение о существовании таких фотонов могло бы оказаться неверным. Действительно, сам факт того, что никакие датчики не могли зафиксировать наличие теневого фотона, а также прямая зависимость от него интерференционной картины приводят к выводу о равенстве нулю его энергии и в результате этого тензора энергии-импульса. Всё это, казалось бы, говорит о невозможности подобной ситуации. Но для полного уяснения представленного положения давайте отложим в сторону фотоны. Ведь изначально задача стояла в понимании интерференционной природы вообще квантовых частиц. Просто Дойчем был рассмотрен случай с фотонами. Поскольку интерференция квантовых частиц протекает одинаково, то вывод, сделанный в [9] относительно параллельных вселенных, можно обобщить и, например, на спинорные поля. Но опять же для этого просто необходимы уравнения, описывающие теневые спинорные поля, последние, как уже говорилось выше, в силу нулевой энергии имеют и нулевой тензор энергии-импульса. Но как раз это теперь и не составляет затруднений! Во-первых, рассмотрим пространство $M = \{U_i\}$, по своей структуре представляющее мультиверс Дойча. Здесь каждое $U_i \in M$ представляет вселенную в привычном для нас смысле этого слова. В дальнейшем отождествим теневые частицы Дойча с духами соответствующих частиц. Пусть теперь наше плоское пространство $V \in M$ в точке x касается соответственно нетождественного с нашим, но уже искривлённого пространства-времени $U \in M$. При этом в U существует некоторая спинорная материя с массой, отличной от нуля, что является вполне законным предположением. Далее, поскольку вселенные U и V касаются в точке x, то в окрестности этой точки в пространстве V мы можем рассмотреть решение уравнения Дирака, для которого $m \neq 0$. Теперь мы можем проинтерпретировать эту константу как массу соответствующей частицы в мире U, т.е. частица с полуцелым спином и ненулевой массой в искривлённом пространстве-времени, при его соприкосновении с некоторым плоским миром, проявляет в последнем свойства частицы-духа. А так как тензор энергии-импульса для спинорных духов тождественно равен нулю, то соответствующие поля не оказывают никакого воздействия на геометрию мира V. В свете описанных положений у нас не возникает и противоречия с известным законом, связывающим энергию и массу: если энергия равна нулю, то и масса должна быть равна нулю. Постоянная в уравнении Дирака для духов не есть масса частицы в реальной вселенной! Теперь мы к тому же вполне корректно можем говорить и о нулевом импульсе рассматриваемой частицы.

Но остаётся один ещё не учтённый факт. В теории Дэвида Дойча говорится о том, что интерференция возникает за счёт отталкивания реальной частицы теневой частицей. А в этой работе происходит отождествление последней с частицей-духом. В физике единой вселенной соответствующее отталкивание происходит за счёт передачи импульса. А каким образом аналогичное перемещение частицы в пространстве происходит в физике мультиверса? Ведь импульс частицы-духа по определению равен нулю?! Но нельзя забывать о ненулевом токе соответствующих полей, он, как и импульс частицы имеет, направление скорости. Приведённым положением хочется сказать о том, что, по всей видимости, не только в физике мультиверса, но и в физике единой вселенной две частицы меняют направление своего движения после столкновения как за счёт импульса, так и за счёт тока. Просто влияние тока на соответствующее изменение имеет достаточный порядок малости по сравнению с влиянием импульса, несущегося частицами, и поэтому пренебрегается. В физике же мультиверса единственной составляющей, способной произвести подобные изменения, является ток, так как импульс частиц-духов тождественно равен нулю. Но, вообще говоря, это всего лишь предположение, возможно, впоследствии будут найдены соответствующие уравнения.

Таким образом, мы получаем, во-первых, физическую интерпретацию полей-духов – это соответствующие поля в параллельных вселенных, а во-вторых, физическое обоснование теневых частиц – найдены уравнения, их описывающие. Более того, мы теперь к тому же получили возможность сделать вполне законный переход к утверждению о существовании параллельных вселенных. Действительно, поскольку все тела состоят из атомов, а атомы из электронов, нейтронов и протонов, которые описываются уравнением Дирака, то наличие в пространстве духов для этих частиц влечёт наличие духов и для каждого тела, т.е. получаем множество параллельных миров.

Заключение

В данной работе были найдены спинорные духи и дана их физическая интерпретация при опоре на параллельные миры. Но кроме этого важен факт того, что отныне материя и вещество не являются синонимами. Так называемые духи как раз и являются материей, но не являются веществом, т.е. это потоки, не обладающие характеристиками последнего. В результате теневые электроны Дойча отождествляются с электронными духами, а также теневым спинорным полям ставятся в соответствие спинорные духи.

Возвращаясь к вопросу относительно фотонов, следует заметить, что теневые фотоны не могут описываться уравнениями Максвелла в отличие от реальных фотонов. Это связано с тем, что равенство нулю тензора энергии-импульса автоматически влечёт отсутствие электромагнитного поля в рассматриваемом пространстве. Но здесь на самом деле нет никакого противоречия! Ведь реальный фотон является не чем иным, как переносчиком некоторой энергии, а энергия теневого равна нулю. Просто уравнения Максвелла не захватывают теневые фотоны, как это происходит в случае со спинорными полями. В отличие от уравнений электромагнитного поля уравнения Дирака описывают не только частицы с полуцелым спином, но и спинорные духи. По всей видимости, для вывода уравнений поля для теневых фотонов надо исходить из аналогичного соображения для спиноров. В соответствующем уравнении Дирака постоянная m была проинтерпретирована как масса частицы в параллельном нашему искривлённом мире, и она не является массой частицы в нашем мире. Аналогичным образом надо поступить и с теневым электромагнитным полем.

Автор выражает благодарность А.К. Гуцу и Р.Т. Файзулину за обсуждение полученных результатов, а также за замечания, сделанные в ходе работы. В результате чего в окончательный вариант был внесён ряд полезных поправок.

Литература

1. Бриль Д., Уилер Дж. *Новейшие проблемы гравитации.* М.: ИЛ, 1961.
2. Ландау Л.Д., Лифшиц Е.М. *Теория поля.* М.: Наука, 1973.
3. Davis T.M., Ray J.R. *Ghost neutrinos in plane-symmetric spacetimes* // J. Math. Phys. 1975. V.16. №1. P.75–79.
4. Davis T.M., Ray J.R. *Neutrinos in cyllindrically-symmetric spacetimes* // J. Math. Phys. 1975. V.16. №1. P.80–81.
5. Гололобова А.С., Кречет В.Г., Лапчинский В.Г. *Теория относительности и гравитация.* М.: Наука, 1976.
6. Pechenick K.R., Cohen J.M. *New exact solution to the Einstein-Dirac equations* // Phys. Rev., D 19, №6. P.1635–1640. 1979.
7. Гуц А.К. *Новое решение уравнений Эйнштейна-Дирака* // Известия вузов. Физика. 1979. №8. с.91-95.
8. Боголюбов Н.Н., Ширков Д.В. *Введение в теорию квантованных полей.* М.: Наука, 1984.
9. Дойч Д. *Структура реальности.* Ижевск: НИЦ «Регулярная и хаотическая динамика», 2001.

Математические
структуры и моделирование
2001, вып. 8, с. 76–90

УДК 530.12:531.51

ТЕОРЕТИКО-ТОПОСНАЯ МОДЕЛЬ МУЛЬТИВЕРСА ДОЙЧА

А.К. Гуц

The Deutsch multiverse is collection of parallel universes. In this article a formal theory of the Deutsch multiverse and a topos-theoretic model of multiverse are given. For this the Lawvere-Kock Synthetic Differential Geometry and models for smooth infinitesimal analysis are used.

Введение

В книге Дэвида Дойча [1] излагается эскиз структуры физической реальности, которая представляет собой совокупность взаимодействующих параллельных вселенных, называемой *мультиверсом*, правильное описание которого, как считает Дойч, возможно лишь в рамках квантовой теории.

Наша цель – оставаясь в рамках математического аппарата 4-мерной общей теории относительности, описывающей Вселенную как конкретное 4-мерное лоренцево многообразие $\langle R^4,\ g^{(4)}\rangle$, называемое *пространством-временем*, предоставить возможность учитывать наличие параллельных, т.е. других вселенных, являющихся самыми различными 4-мерными псевдоримановыми многообразиями, за счет любого необходимого произвольного увеличения размерности особого Гиперпространства, объемлющего все вселенные. Более того, Гиперпространств должно быть сколь угодно много; геометрия, топология, размерность Гиперпространств должны быть сколь угодно различными, чтобы всегда можно было найти бесчисленное число вселенных, сколь угодно подобных нашей, и одновременно должно существовать сколь угодно много вселенных, совершенно непохожих на мир, в котором мы живем.

Структура физической реальности должна учитывать прихоть мыслящего существа видеть ее во всевозможных мыслимых формах, располагая при этом весьма скудным исследовательским инструментарием, основой которого должны быть теория относительности и квантовая механика.

Следут особо подчеркнуть, что мы не намерены переходить к многомерным теориям типа Калуцы-Клейна. Нет, ни в коем случае. Подчеркиваем, что основой теории мультиверса должна быть 4-мерная метрика $g^{(4)}$.

E-mail: guts@univer.omsk.su

Омский государственный университет

Работа выполнена при финансовой поддержке РФФИ (проект 01-01-00303 – теоретическая часть).

Нетрудно понять, что поставленная нами цель требует иного взгляда на общую теорию относительности, поскольку мы собираемся совместить несовместимые вещи. Тем не менее, выход находится при обращении к интуиционистскому взгляду на риманову геометрию. Отказываясь от закона исключенного третьего, можно построить теорию, включающую как классический вариант общей теории относительности, так и множество других ее многомерных обобщений.

1. Формальная теория мультиверса

Теорию мультиверса следует строить как формальную теорию $\mathcal{T}$, максимально похожую на общую теорию относительности, т.е. как теорию одной 4-мерной вселенной, а параллельные вселенные должны появиться при построении моделей формальной теории.

Основой формальной теории $\mathcal{T}$ может послужить так называемая Синтетическая дифференциальная геометрия (СДГ) Ловера-Кока [2]. Как известно, из-за того, что принимаемая СДГ аксиома Кока-Ловера несовместима с законом исключенного третьего, нельзя построить модель этой теории в категории теории множеств Кантора **Set**.

Отказ от закона исключенного третьего приводит нас к интуиционистской логике, которой мы должны придерживаться при развитии теории мультиверса, опираясь на СДГ. Место теоретико-множественных моделей формальной теории мультиверса должны занять теоретико-топосные модели. Последние хотя и обладают, в общем случае, внутренней интуиционистской логикой, развиваются в рамках двузначной классической логики. Это позволяет математику иметь дело с привычными объектами, правда, в рамках очень сложных конструкций, каковыми являются топосы.

Основным для СДГ Кока-Ловера является замена поля действительных чисел $\mathbb{R}$ на коммутативное кольцо R. В идеале хотелось бы, чтобы оно удовлетворяло следующим аксиомам [1]:

(A1) $\langle R, +, \cdot, 0, 1\rangle$ – коммутативное кольцо.

(A2) R локальное кольцо, т.е.

$$0 = 1 \Longrightarrow \bot$$
$$\exists y\ (x \cdot y = 1) \exists y\ (1 - x) \cdot y = 1.$$

(A3) $\langle R, <\rangle$ – действительное упорядоченное локальное кольцо, т.е. $<$ – транзитивное отношение, совместимое с кольцевой структурой в том смысле, что

$$(a)\ 0 < 1,\ (0 < x \ \&\ 0 < y \Longrightarrow 0 < x + y \ \&\ 0 < x \cdot y),$$
$$(b)\ \exists y (x \cdot y = 1) \Longleftrightarrow (0 < x \vee x < 0),$$
$$(c)\ 0 < x \Longrightarrow \exists y (x = y^2)\ \text{(евклидовость)}.$$

[1] Мы приводим только часть аксиом. Другие аксиомы см. в [7, Гл.VII].

(A4) $\leq$ – предпорядок, совместимый с кольцевой структурой, т.е. рефлексивное и транзитивное отношение, и

$$(a)\ 0 \leq 1,\ (0 \leq x\ \&\ 0 \leq y \Longrightarrow 0 \leq x + y\ \&\ 0 \leq x \cdot y),\ 0 \leq x^2,$$
$$(b)\ (x - \text{нильпотент, т.е. } x^n = 0) \Longrightarrow 0 \leq x.$$

(A5) $<$ и $\leq$ – совместимы, т.е.

$$(a)\ x < y \Longrightarrow x \leq y,$$
$$(b)\ x < y\ \&\ y \leq x \Longrightarrow \perp.$$

(A6) (Аксиома Кока-Ловера). Для любого

$$\forall (f \in R^D) \exists ! (a, b) \in R \times R \forall d \in D (f(d) = a + b \cdot d),$$

где $D = \{x \in R : x^2 = 0\}$.

Как показано в [2], аксиома (A6) несовместима с законом исключенного третьего.

(A7) (Аксиома интеграла).

$$\forall f \in R^{[0.1]} \exists ! g \in R^{[0.1]} (g(0) = 0\ \&\ \forall x \in [0,1]\ (g'(x) = f(x)),$$

где $[0,1] = \{x \in R : 0 \leq x\ \&\ x \leq 1\}$ и $g'(x)$ – это единственное b такое, что $\forall d \in D (g(x + d) = g(x) + b \cdot d)$.

Используется символическая запись

$$g(x) = \int_0^1 f(t)dt.$$

(A8) $\forall x \in [0,1]\ 0 < f(x) \Longrightarrow 0 < \int_0^1 f(x)dx.$

(A8$'$) $\forall x \in [0,1]\ 0 \leq f(x) \Longrightarrow 0 \leq \int_0^1 f(x)dx.$

(A9) (Существование обратной функции).

$$\forall f \in R^R\ \forall x \in R\ \ (f'(x) - \text{обратимо} \Longrightarrow$$
$$\Longrightarrow \exists \text{ открытые } U, V (x \in U\ \&\ f(x) \in V\ \&\ f|_U : U \to V - \text{биекция})).$$

(A10) $N \subset R$, т.е. $\forall x \in N\ \exists y \in R (x = y)$.

(A11) R – архимедово, т.е. $\forall x \in R\ \exists n \in N (x < n)$.

(A12) (Аксиомы Пеано).

$$\begin{aligned}&0 \in N\\ &\forall x \in R\ (x \in N \Longrightarrow x + 1 \in N)\\ &\forall x \in R\ (x \in N\ \&\ x + 1 = 0 \Longrightarrow \bot).\end{aligned}$$

Кольцо R дополнительно к обычным действительным числам из $\mathbb{R}$ располагает элементами, называемыми *инфинитезималами* и входящими в «множества»

$$D = \{d \in R : d^2 = 0\}, ..., D_k = \{d \in R : d^{k+1} = 0\}, ...,$$

$$\mathbb{\Delta} = \{x \in R : f(x) = 0,\ f \in m_0^g\},$$

где $m_{\{0\}}^g$ – идеал функций, имеющих нулевой росток в 0 [2], причем

$$D \subset D_2 \subset ... \subset D_k \subset ... \subset \mathbb{\Delta}.$$

В рамках изложенной аксиоматики можно построить [3,4] риманову геометрию для 4-мерных (формальных) многообразий $\langle R^4, g^{(4)}\rangle$, являющуюся основой для эйнштейновской теории гравитации.

Мы постулируем, что *мультиверс – это 4-мерное пространство-время, описываемое с помощью СДГ, т.е. является формальным лоренцевым многообразием $\langle R^4, g^{(4)}\rangle$, для которого выполняются уравнения Эйнштейна, представленные в традиционном виде*:

$$R_{ik}^{(4)} - \frac{1}{2} g_{ik}^{(4)} (R^{(4)} - 2\Lambda) = \frac{8\pi G}{c^4} T_{ik}. \tag{1}$$

Решением этих уравнений будет 4-метрика $g^{(4)}$.

На формальном уровне физические следствия таких предположений не так заметны, как математические. Поэтому необходимо обратиться к моделям формальной теории. Наиболее исследованными являются так называемые хорошо адаптированные модели вида $\mathbf{Set}^{\mathbb{L}^{op}}$, содержащие как полную подкатегорию категорию гладких многообразий $\mathcal{M}$.

2. Гладкие топосные модели мультиверса

Пусть $\mathbb{L}$ – это дуальная категория для категории конечно порожденных C^∞-колец. Она называется *категорией локусов* [7]. Объектами категории $\mathbb{L}$ являются все те же конечно порожденные C^∞-кольца, а морфизмами – обращенные морфизмы категории конечно порожденных C^∞-колец. Принято во избежание путаницы объекты (локусы) категории $\mathbb{L}$ обозначать как ℓA, где A – C^∞-кольцо. Следовательно, $\mathbb{L}$-морфизм $\ell A \to \ell B$ – это C^∞-гомоморфизм $B \to A$.

[2]Иначе говоря, исчезающих в некоторой окрестности точки 0.

Конечно порожденное C^∞-кольцо ℓA изоморфно кольцу вида $C^\infty(\mathbb{R}^n)/I$ (для некоторого натурально числа n и некоторого конечно порожденного идеала I).

Категория $\mathbf{Set}^{\mathbb{L}^{op}}$ является топосом. Мы рассмотрим топос $\mathbf{Set}^{\mathbb{L}^{op}}$ как модель формальной теории мультиверса. Важно отметить, что модель $\mathbf{Set}^{\mathbb{L}^{op}}$ обладает патологическими свойствами: многие из аксиом (A1)-(A12) не выполняются в $\mathbf{Set}^{\mathbb{L}^{op}}$. Например, оказывается, что *гладкая прямая* R, будучи коммутативным кольцом с единицей 1, не является при этом даже локальным кольцом, т.е. нарушается аксиома (A2). Более того, R не обладает свойством архимедовости (аксиома (A11)).

Можно рассматривать в качестве моделей топосы $\mathcal{F}, \mathcal{G}$ и $\mathcal{Z}$ и многие другие [7, Appendix 2]. Для них выполнены все аксиомы (A1)-(A12) (см. [7, с.300]). Однако работа с топосом $\mathbf{Set}^{\mathbb{L}^{op}}$ позволяет быстрее ознакомиться с излагаемой теорией мультиверса, не усложняя изложение математическими конструкциями.

На языке Дойча переход к конкретной модели формальной теории – это порождение *виртуальной реальности* [3]. Физическая реальность, воспринимаемая нами и названнная Дойчем *мультиверсом* [4], также является виртуальной реальностью, созданной нашим мозгом [1, с.140]. Более того, «виртуальная реальность, основанная на неправильных законах, и *есть* наш единственный источник получения знаний! ... А поскольку наши концепции и теории (будь они врожденные или приобретенные) никогда не совершенны, все наши передачи на самом деле неточны. То есть, они дают нам ощущение среды, которая значительно отличается от среды, в которой мы действительно находимся» [1, с.140].

Модель мультиверса – это *генератор виртуальной реальности*, который обладает определенным *репертуаром сред*, которые он создает и в которые мы погружаемся. Поясним, как это происходит.

При интерпретации $i : \mathbf{Set}^{\mathbb{L}^{op}} \models \mathcal{T}$ формальной теории $\mathcal{T}$ мультиверса в топосе $\mathbf{Set}^{\mathbb{L}^{op}}$ объектам теории, например кольцу R, степени R^R и т.д., ставятся в соответствие объекты топоса, т.е. функторы $F = i(R)$, $F^F = i(R^R)$ и т.д. Отображениям, например $R \to R$, $R \to R^R$, – морфизмы топоса $\mathbf{Set}^{\mathbb{L}^{op}}$, т.е. естественные преобразования функторов – $F \to F$, $F \to F^F$.

Наконец, при интерпретации языка формальной теории мультиверса необходимо приписать элементам кольца R «элементы» функторов $F \in \mathbf{Set}^{\mathbb{L}^{op}}$. Иначе говоря, нужно проинтерпретировать отношение $r \in R$. Это сделать не так просто потому, что функтор F определен на категории локусов $\mathbb{L}$; его переменной (аргументом) является произвольный локус ℓA, а значением множество $F(\ell A) \in \mathbf{Set}$. Выход из затруднения заключается в определении *обобщенных элементов* $x \in_{\ell A} F$ функтора F.

Обобщенным элементом $x \in_{\ell A} F$, или *элементом x функтора F в стадии ℓA*, называется элемент $x \in F(\ell A)$.

---

[3]Это предположение автор услышал от А.А.Звягинцева.

[4]Multiverse – много (multi-) вселенных (universe); причем universe – одна (uni) вселенная. Другой не мыслили, и это отразилось в языке.

Теперь можно сопоставить элементу $r \in R$ обобщенный элемент $i(r) \in_{\ell A} F$. Но, как видим, таких элементов столько сколько локусов. При переходе к модели $\mathbf{Set}^{\mathbb{L}^{op}}$ происходит «размножение» элемента r. Он начинает существовать в бесконечном числе вариантов $\{i(r) : i(r) \in_{\ell A} F, \ell A \in \mathbb{L}\}$.

Важно отметить, что поскольку 4-метрика $g^{(4)}$ – это элемент объекта $R^{R^4 \times R^4}$, то «интуиционистская» 4-метрика начинает существовать в бесконечном числе классических вариантов $i(g)^{(4)} \in_{\ell A} i(R^{R^4 \times R^4})$. Обозначим каждый такой вариант как $i(g)^{(4)}(\ell A)$.

Для упрощения изложения будем далее иметь дело с объектами модели $\mathbf{Set}^{\mathbb{L}^{op}}$. Другими словами, будем писать $g^{(4)}(\ell A)$ вместо $i(g)^{(4)}(\ell A)$.

Нетрудно понять, что каждый вариант $g^{(4)}(\ell A)$ классической 4-метрики удовлетворяет «своему» уравнению Эйнштейна [4]

$$R_{ik}^{(4)}(\ell A) - \frac{1}{2} g_{ik}^{(4)}(\ell A)[R^{(4)}(\ell A) - 2\Lambda(\ell A)] = \frac{8\pi G}{c^4} T_{ik}(\ell A). \tag{2}$$

Причем не исключено, что физические константы G, c также могут меняться от варианта к варианту.

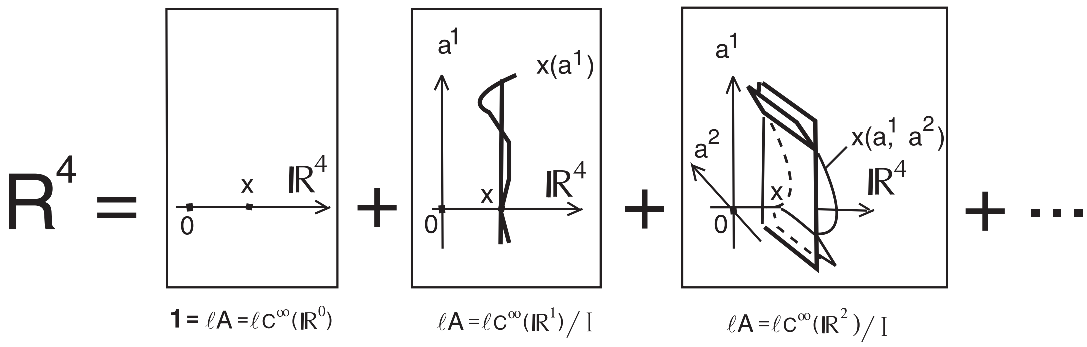


Рис. 1. Физическая (виртуальная) реальность $\mathbf{R}^4$ как сумма многомерных гиперпространств (сред), расслоенных на параллельные 4-мерные вселенные, соответствующих различному «вычислению» реальности.

Прежде чем пойти дальше, укажем на существование вложения Ионеды (Yoneda)

$$y : \mathbb{L} \hookrightarrow \mathbf{Set}^{\mathbb{L}^{op}},$$

$$y(\ell A) = Hom_{\mathbb{L}}(-, \ell A).$$

Примем, что кольцо R интерпретируется как функтор $y(\ell C^\infty(\mathbb{R}))$, т.е.$i(R) = y(\ell C^\infty(\mathbb{R}))$. Будем далее писать ℓA вместо $y(\ell A)$ и опустим символ i. Тогда имеем

$$R(-) = \ell C^\infty(\mathbb{R})(-) = Hom_{\mathbb{L}}(-, \ell C^\infty(\mathbb{R})).$$

Аналогично

$$R^{R^4 \times R^4}(\ell A) = Hom_{\mathbb{L}}(\ell A, R^{R^4 \times R^4}) = Hom_{\mathbb{L}}(\ell A \times (R^4 \times R^4), R) =$$

$$= Hom_{\mathbb{L}}(\ell C^\infty(\mathbb{R}^m)/I \times \ell C^\infty(\mathbb{R}^4) \times \ell C^\infty(\mathbb{R}^4), \ell C^\infty(\mathbb{R})) =$$

$$= Hom_{\mathbb{L}^{op}}(\ell C^\infty(\mathbb{R}), C^\infty(\mathbb{R}^m)/I \otimes_\infty C^\infty(\mathbb{R}^4) \otimes_\infty C^\infty(\mathbb{R}^4)) =$$

$$= Hom_{\mathbb{L}^{op}}(C^\infty(\mathbb{R}), C^\infty(\mathbb{R}^{m+8})/(I,\{0\})) = Hom_{\mathbb{L}}(\ell C^\infty(\mathbb{R}^{m+8})/(I,\{0\}), \ell C^\infty(\mathbb{R})),$$

где $\ell A = \ell C^\infty(\mathbb{R}^m)/I$, $\otimes_\infty$ – символ копроизведения C^∞-колец, и при вычислении использованы формулы

$$C^\infty(\mathbb{R}^n) \otimes_\infty C^\infty(\mathbb{R}^k) = C^\infty(\mathbb{R}^{n+k}),$$

$$\frac{\ell A \to \ell C^{\ell B}}{\ell B \times \ell A \to \ell C}.$$

Отсюда следует, что при $\ell A = \ell C^\infty(\mathbb{R}^m)$

$$g^{(4)}(\ell A) = [g \in_{\ell A} R^{R^4 \times R^4}] \equiv g^{(4)}_{ik}(x^0,...,x^3,a)dx^i dx^k, \quad a = (a^1,...,a^m) \in \mathbb{R}^m.$$

Дополним метрику $g^{(4)}_{ik}(x^0,...,x^3,a)$ до (4+m)-метрики в пространстве $\mathbb{R}^{4+m}$

$$g^{(4)}_{ik}(x^0,...,x^3,a)dx^i dx^k - da^{1^2} - ... - da^{m^2}. \tag{3}$$

Получаем $(4+m)$-мерную геометрию.

Символически процедуру получения многомерных вариантов геометрии, порождаемых интуиционистской 4-геометрией $g^{(4)}$, можно представить в виде формальной суммы

$$g^{(4)} = c_0 \cdot \underbrace{[g^{(4)} \in_{\mathbf{1}} R^{R^4 \times R^4}]}_{\text{4-мерная геометрия}} + c_1 \cdot \underbrace{[g^{(4)} \in_{\ell C^\infty(\mathbb{R}^1)} R^{R^4 \times R^4}]}_{\text{5-мерная геометрия}} + ...$$

$$... + c_{n-4} \cdot \underbrace{[g^{(4)} \in_{\ell C^\infty(\mathbb{R}^{n-4})} R^{R^4 \times R^4}]}_{\text{n-мерная геометрия}} + ...,$$

где коэффициенты c_m берутся из поля комплексных чисел.

Поскольку стадий несчетное число, то вместо суммы следует писать интеграл

$$g^{(4)} = \int\limits_{\mathbb{L}} \mathcal{D}[\ell A] c(\ell A)[g^{(4)} \in_{\ell C^\infty(\mathbb{R}^{n-4})} R^{R^4 \times R^4}]. \tag{4}$$

Используем обозначения квантовой механики [5]:

$$g^{(4)} \to |g^{(4)}\rangle, \quad [g^{(4)} \in_{\ell C^\infty(\mathbb{R}^{n-4})} R^{R^4 \times R^4}] \to |g^{(4)}(\ell A)\rangle.$$

Тогда (4) перепишется в виде

$$|g^{(4)}\rangle = \int\limits_{\mathbb{L}} \mathcal{D}[\ell A] c(\ell A)|g^{(4)}(\ell A)\rangle. \tag{5}$$

[5] Дираковские обозначения: $|P\rangle = \psi(\xi)\rangle \equiv \psi(\xi)$; в данном случае $\psi(\xi)$ – это $g^{(4)}$ (представитель состояния $|P\rangle$), а $|P\rangle$ – это $|g^{(4)}\rangle$ [5, с.111-112].

Таким образом, формальная 4-геометрия Кока-Ловера $\langle R^4, g^{(4)}\rangle$ есть сумма бесконечного числа классических многомерных псевдоримановых геометрий (гиперпространств), которые расслаиваются посредством фиксации $a = a_0$ на 4-мерные параллельные вселенные. Геометрические свойства параллельных вселенных могут, как показано в [9, 10], существенно различаться даже в рамках одной стадии ℓA. О природе, смысле коэффициентов $c(\ell A)$ поговорим ниже в §5.

Здесь как раз уместно вспомнить о средах виртуальносй реальности, которые должны возникать при обращении к модели мультиверса, в данном случае к модели $\mathbf{Set}^{\mathbb{L}^{op}}$, являющейся генератором виртуальной реальности. Нетрудно понять, что обобщенные элементы $|g^{(4)}(\ell A)\rangle$ – это метрики конкретной среды (=гиперпространство) с «номером» ℓA. Другими словами, обращение к изучению любого объекта теории в стадии ℓA есть не что иное, как переход к одной из сред, входящих в репертуар генератора виртуальной реальности $\mathbf{Set}^{\mathbb{L}^{op}}$.

3. Космология Дойча-Гёделя

В качестве примера мультиверса рассмотрим космологическое решение Гёделя [6]:

$$g^{(4)}_{ik} = \alpha^2 \begin{pmatrix} 1 & 0 & e^{x^1} & 0 \\ 0 & -1 & 0 & 0 \\ e^{x^1} & 0 & e^{2x^1}/2 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \quad (g^{(4)})^{ik} = \frac{1}{\alpha^2} \begin{pmatrix} -1 & 0 & 2e^{-x^1} & 0 \\ 0 & -1 & 0 & 0 \\ 2e^{-x^1} & 0 & -2e^{-2x^1} & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \tag{6}$$

Эта метрика удовлетворяет уравнениям Эйнштейна (1) с тензором энергии-импульса пылевой материи

$$T_{ik} = c^2 \rho u_i u_k,$$

при условии, что

$$\frac{1}{\alpha^2} = \frac{8\pi G}{c^2}\rho, \quad \Lambda = -\frac{1}{2\alpha^2} = -\frac{4\pi G\rho}{c^2}. \tag{7}$$

Если теперь положить

$$\alpha = \alpha_0 + d, \Lambda = \Lambda_0 + \lambda, \rho = \rho_0 + \varrho, \tag{8}$$

где $d, \lambda, \varrho \in D$ – инфинитезималы, и подставить в (7), то имеем

$$\frac{1}{(\alpha_0 + d)^2} = \frac{1}{\alpha_0^2} - \frac{2d}{\alpha_0^3} = \frac{8\pi G}{c^2}(\rho_0 + \varrho),$$

$$2\Lambda_0 + 2\lambda = -\frac{1}{\alpha_0^2} + \frac{2d}{\alpha_0^3}, \quad \Lambda_0 + \lambda = -\frac{4\pi G\rho_0}{c^2} - \frac{4\pi G\varrho}{c^2}.$$

Предположим, что $\alpha_0, \Lambda_0, \rho_0$ связаны соотношениями (7). Тогда из предыдущих равенств находим связь между инфинитезиамалами

$$\lambda = -\frac{4\pi G}{c^2}\varrho, \quad d = -\frac{4\pi G a_0^3}{c^2}\varrho.$$

При интерпретации в гладком топосе $\mathbf{Set}^{\mathbb{L}^{op}}$ инфинитезимал $\varrho \in D$ в стадии $\ell A = C^\infty(\mathbb{R}^m)/I$ представляется классом гладких функций вида $\varrho(a) mod\ I$, где $[\varrho(a)]^2 \in I$ [7, с.77].

Рассмотрим состояние мультиверса Гёделя, точнее, мультиверса Дойча-Гёделя в стадии $\ell A = \ell C^\infty(\mathbb{R})/(a^4)$ [6]. Очевидно, что можно взять инфинитезимал вида $\varrho(a) = a^2$. Мультиверс в этой стадии является 5-мерным гиперпространством, слои которого, задаваемые уравнением $a = a_0$, – параллельные вселенные (среды) $R^4(\ell A)$ с метрикой $g^{(4)}(\ell A) = g^{(4)}_{ik}(x, a)$, заданной формулами (6) с учетом (8). Плотность материи $\rho = \rho_0 + \varrho(a)$ начнет расти от классического значения $\rho_0 \sim 2 \cdot 10^{-31}$ *г/см*^{3} до $+\infty$ при $a \to \pm\infty$. Начинает неограниченно расти до $-\infty$ и космологическая постоянная. Все это говорит о том, что параллельные вселенные могут иметь физические свойства, совершенно отличные от свойств нашей Вселенной.

В стадии $\ell A = \ell C^\infty(\mathbb{R})/(a^2)$ $\varrho(a) = a$ и $\rho = \rho_0 + \varrho(a) \to -\infty$ при $a \to -\infty$, т.е. становится физически неинтерпретируемой, поскольку не ясно, что представляет собой «экзотическая» материя с отрицательной плотностью.

Наконец, в стадии $\mathbf{1} = \ell C^\infty(\mathbb{R})/(a)$ все $\varrho(a) = d(a) = \lambda(a) = 0$, т.е. имеем дело с классической вселенной Гёделя.

4. Квантовые свойства геометрии параллельных вселенных

В излагаемой теории мультиверса естественным образом переносятся идеи квантовой геометродинамики Уилера. Так, формула для амплитуды вероятности перехода от 3-геометрии $g^{(3)}$ физического пространства к 3-геометрии $h^{(3)}$ принимает вид «двойного» интеграла Фейнмана по траекториям, которыми являются различные 4-геометрии $g^{(4)}$:

$$\langle g^{(3)} | h^{(3)} \rangle = \int\limits_{\mathbb{L}} \mathcal{D}[\ell A] \int\limits_{g^{(3)}(\ell A)}^{h^{(3)}(\ell A)} \mathcal{D}[g^{(4)}(\ell A)] e^{\frac{i}{\hbar} S[g^{(4)}(\ell A)]},$$

где

$$S[g^{(4)}(\ell A)] = \kappa_m(\ell A) \int\limits_{\mathbb{R}^{4+m}} \sqrt{-\det ||g^{(4)}(\ell A)||} R^{(4)}(\ell A) d^4x d^m a$$

– действие в пространстве $\langle \mathbb{R}^{4+m}, g^{(4)}(\ell A) \rangle$.

Как видим, в действительности интеграл Фейнмана по траекториям $g^{(4)}$ – это бесконечное число интегралов по $(4+m)$-мерным траекториям $g^{(4)}(\ell A)$ вида (3).

Повторяя вычисления Уилера, можно оценить квантовые флутуации 4-метрики $g^{(4)} \to g^{(4)} + \Delta g^{(4)}$, не вносящие искажение в интерференционную картину, задаваемую интегралами по траекториям.

[6] Через $(f_1, ..., f_k)$ обозначается идеал кольца $C^\infty(\mathbb{R}^n)$, порожденный функциями $f_1, ..., f_k \in C^\infty(\mathbb{R}^n)$, т.е. имеющий вид $\sum_{i=1}^k g_i f_i$, где $g_1, ..., g_k \in C^\infty(\mathbb{R}^n)$ – произвольные гладкие функции.

При предположении, что при флуктуациях $\det ||g^{(4)}(\ell A)|| \sim 1$, получаем для искомых флуктуаций в $(4+m)$-мерной области с размерами $L^4 \times L_1^m$

$$\Delta g^{(4)}(\ell A) \sim \frac{L^*}{L}\left(\frac{T}{L_1}\right)^{\frac{m}{2}}, \tag{9}$$

где

$$L^* = \sqrt{\frac{G\hbar}{c^3}} \sim 10^{-33} \text{см}$$

– планковская длина и принято, что $\kappa_m(\ell A) \sim c^3/(\hbar G T^m)$, где T [см] – величина, характеризующая «размеры» дополнительных измерений.

Из (9) вытекает, что при $L \sim L^*, L_1 \sim T$ все флуктуации $\Delta g^{(4)}(\ell A) \sim 1$, т.е. становятся существенными. Геометрия и топология «пенятся» на уровне микромира.

Как показано в [13, 14], флуктуации могут иметь место и на макроскопических расстояниях или отрезках времени. Это возможно за счет высших измерений, которые появляются за счет рассмотрения мультиверса в различных стадиях ℓA, т.е. различных состояний (сред) $R^4(\ell A)$ мультиверса.

5. Электроны-двойники

Дойч предположил, что параллельная вселенная образуется за счет *теневых* элементарных частиц, сопровождающих каждую *реальную* частицу. Реальные частицы «мы можем увидеть или обнаружить с помощью приборов, тогда как вторые (теневые – А.Г.) – неосязаемы (невидимы): их можно обнаружить только косвенно через их воздействие на видимые» частицы [1, с.48]. «Между реальными и теневыми фотонами не существует особой разницы: каждый фотон осязаем в одной Вселенной и не осязаем во всех параллельных Вселенных».

Уравнение Дирака в СДГ

$$i\hbar\gamma^{(k)}\frac{\partial\psi}{\partial x^k} - mc\psi = 0 \tag{10}$$

в пространстве-времени Минковского, т.е. в мультиверсе Дойча-Минковского M^4 с метрикой, записанной в виде

$$ds^2 = dx^{0^2} - dx^{1^2} - dx^{2^2} - dx^{3^2} \tag{11}$$

имеет, например, следующее решение

$$\psi(x) = \begin{pmatrix} 1 \\ 1 \\ -1 \\ 1 \end{pmatrix} e^{\frac{mc}{\hbar}x^2 + g(x^3+x^0) + i\theta\cdot f(x^3+x^0)}, \tag{12}$$

которое при $\theta\cdot f(x^3+x^0)=const$ являтся спинорным духом [7], т.е. имеет нулевой тензор-энергии импульса поля $\psi(x)$

$$T_{ik}=\frac{i\hbar c}{4}\left\{\psi^*\gamma^{(0)}\gamma_{(i)}\frac{\partial\psi}{\partial x^k}-\frac{\partial\psi^*}{\partial x^k}\gamma^{(0)}\gamma_{(i)}\psi+\psi^*\gamma^{(0)}\gamma_{(k)}\frac{\partial\psi}{\partial x^i}-\frac{\partial\psi^*}{\partial x^i}\gamma^{(0)}\gamma_{(k)}\psi\right\}. \quad (13)$$

Спинорный дух, как видим, описываемый спинорым *полем* ψ, является призрачным, поскольку не обладает ни энергией, ни импульсом. Уместно вспомнить замечание Эйнштейна на соотношение электромагнитного поля и световых квантов (фотонов). «Эйнштейн считал, что поле «прокладывает путь» световым квантам. Эти поля определяют вероятность найти в системе квант, который переносит вдоль заданного пути энергию и импульс. Сами же поля, поскольку они призрачны, не обладают ни энергией, ни импульсом» [15, с.71-72].

Поскольку духи как спинорное поле не имеют энергии и импульса, то они *не могут фиксироваться приборами*. Они неосязаемы. Именно поэтому Е.В.Палешева предложила [16] отождествлять спинорные духи с теневыми частицами Дойча.

Решению ψ можно сопоставить [8] дираковский ket-вектор $|\Psi\rangle$, представленный в виде суммы [9]

$$|\Psi\rangle=\int\limits_{\mathbb{L}}\mathcal{D}[\ell A]a(\ell A)|\Psi(\ell A)\rangle. \quad (14)$$

Естественно трактовать $\psi=|\Psi\rangle$. Тогда $\psi^*\psi=\langle\Psi|\Psi\rangle$ – плотность вероятности электрона и

$$\int\limits_{R^4}\psi^*\psi d^4x=\int\limits_{R^4}\langle\Psi|\Psi\rangle d^4x=1. \quad (15)$$

Полагая, что

$$\langle\Psi|=\int\limits_{\mathbb{L}}\mathcal{D}[\ell B]a^*(\ell B)\langle\Psi(\ell B)|.$$

Поэтому

$$1=\int\limits_{R^4}\langle\Psi|\Psi\rangle d^4x=\int\limits_{\mathbb{R}^4}d^4x\int\limits_{\mathbb{L}}\mathcal{D}[\ell B]\int\limits_{\mathbb{L}}\mathcal{D}[\ell A]a^*(\ell B)a(\ell A)\langle\Psi(\ell B)|\Psi(\ell A)\rangle=$$

$$=\int\limits_{\mathbb{L}}\mathcal{D}[\ell B]a^*(\ell B)\int\limits_{\mathbb{L}}\mathcal{D}[\ell A]a(\ell A)\left(\int\limits_{\mathbb{R}^4}d^4x\langle\Psi(\ell B)|\Psi(\ell A)\rangle\right)=$$

$$=\int\limits_{\mathbb{L}}\mathcal{D}[\ell B]a^*(\ell B)\int\limits_{\mathbb{L}}\mathcal{D}[\ell A]a(\ell A)\delta(\ell B-\ell A)=\int\limits_{\mathbb{L}}\mathcal{D}[\ell B]a^*(\ell B)a(\ell B),$$

[7] Данное решение найдено Е.В.Палешевой.

[8] См. примечание 5.

[9] Приводимая формула и придаваемый ей в этой статье смысл имеет прямое отношение к эвереттовской трактовке квантовой механики [8].

где положили (как логическое продолжение равенства (15), что

$$\int\limits_{\mathbb{R}^4} d^4x \langle \Psi(\ell B)|\Psi(\ell A)\rangle = \delta(\ell B - \ell A),$$

$$\int\limits_{\mathbb{L}} \mathcal{D}[\ell B] f(\ell B)\delta(\ell B - \ell A) = f(\ell A).$$

Следовательно,

$$\int\limits_{\mathbb{L}} \mathcal{D}[\ell A] a^*(\ell A) a(\ell A) = 1.$$

и вполне разумно допустить, что $a^*(\ell A)a(\ell A)$ – это квадрат модуля амплитуды вероятности стадии ℓA, характеризующий вероятность наблюдения электрона в стадии ℓA мультиверса M^4.

Такой вывод позволяет трактовать $c^*(\ell A)c(\ell A)$, где $c(\ell A)$ – комплексные коэффициенты в разложении (5) 4-метрики мультиверса $\langle R^4, g^{(4)}\rangle$, как вероятность (точнее, квадрат модуля амплитуды вероятности) того, что мультиверс находится в состоянии $|g^{(4)}(\ell A)\rangle$ [10].

Пусть в выражении для спинорного поля (12) число $\theta = 1 - \varepsilon$, где ε инфинитезимал, т.е. $\epsilon \in \mathbb{\Delta} = \{x \in \mathbf{R} | f(x) = 0,\ f \in m^g_{\{0\}}\}$, $m^g_{\{0\}}$ идеал функций, имеющих нулевой росток в 0.

Если $\epsilon \in \mathbb{\Delta}$, то ε в стадии $\ell C^\infty(\mathbb{R}^n)/I$ задается функцией $\varepsilon(a), a \in \mathbb{R}^n$ такой, что для для любой $\phi \in m^g_{\{0\}} \quad \phi(\varepsilon(a)) \in I$ [7, с.77].

Имеем

$$\phi(\varepsilon(a)) = \phi(\varepsilon(0)) + \sum_{|\alpha|=1}^{\infty} \frac{1}{\alpha!} D^\alpha(\phi \circ \varepsilon)(0) a^\alpha =$$

$$= \phi(\varepsilon(0)) + \sum_{|\alpha|=1}^{\infty} \frac{1}{\alpha!} \left(\sum_{|\beta|=1}^{|\alpha|} D^\beta \phi(\varepsilon(0)) P_\beta(\varepsilon(0)) \right) a^\alpha, \qquad (16)$$

где α, β – мультииндексы и P_β – некоторые полиномы.

В стадии $\ell C^\infty(\mathbb{R}^n) \quad \phi(\varepsilon(a)) \in I = \{0\}$ для любой $\phi \in m^g_{\{0\}}$. Поэтому из (16) следует, что прежде всего $\phi(\varepsilon(0)) = 0$, и, следовательно, $\varepsilon(0) = 0$. Кроме этого

$$\sum_{|\beta|=1}^{|\alpha|} D^\beta \phi(\varepsilon(0)) P_\beta(\varepsilon(0)) = 0.$$

Но для любой $\phi \in m^g_{\{0\}} \quad D^\beta \phi(0) = 0$. Поэтому $\epsilon(a)$ произвольная функция, удовлетворяющая условию $\varepsilon(0) = 0$.

[10]Метрика – это гравитационное поле, определяющее геометрию и в определенной мере топологию пространства-времени. Поэтому естественно отождествлять состояние (среду) мультиверса $|R^4(\ell A)\rangle$ в стадии ℓA (см., например, рис.1) с состоянием 4-метрики $|g^{(4)}(\ell A)\rangle$.

Возвращаясь к полю (12), примем, что $\theta(a) = 1 - \varepsilon$, где

$$\varepsilon(0) = 0, \ \ \varepsilon(a) > 0 \text{ при } a \neq 0, \text{ и } \epsilon = 1 \text{ при } ||a|| \geq r_0,$$

а f некоторая не равная тождественно нулю функция. Тогда в стадии $\ell A = \ell C^\infty(\mathbb{R}^n)$ имеем

$$\theta(a) = 1 - \varepsilon(a) = \begin{cases} 0 & \text{при } ||a|| \geq r_0, \\ > 0 & \text{при } ||a|| < r_0. \end{cases}$$

Следовательно, в стадии $\ell A = \ell C^\infty(\mathbb{R}^n)$ поле ψ не является спинорным духом в нашей Вселенной ($a = 0$) и во вселенных с $||a|| < r_0$, но – дух в параллельных вселенных, для которых $\varepsilon(a) \geq r_0$. Можно взять число r_0 столь малым, что вселенные, «помеченные» параметром a с $||a|| < r_0$, в силу квантового вспенивания топологии и геометрии, должны рассматриваться как одна вселенная (r_0 – «толщина» вселенной). Это означает, что поле ψ – это реальная частица в нашей Вселенной и теневые частицы-двойники во всех других вселенных.

Если же взять $\theta \in \mathbb{A}$ так, что

$$\theta(a) > 0 \text{ при } ||a - a_0|| < r_0 \text{ и } \theta(a) = 0 \text{ при } ||a|| > r_0,$$

где $a_0 \neq 0$ и $r_0 < ||a_0||$, то поле ψ в стадии $\ell C^\infty(\mathbb{R}^n)$ не является спинорным духом во вселенной $a = a_0$, имеющей «толщину» r_0, и является духом, т.е. теневой частицей-близнецом, во всех других вселенных, включая нашу Вселенную ($a = \ 0$).

При этом в стадии $\mathbf{1} = \ell C^\infty(\mathbb{R}^0) = \ell C^\infty(\mathbb{R})/(a^1) \quad \theta \cdot f(x^3 + x^0) \mod \{a^1\} = f(x^3 + x^0)$. Это означает, что мы имеем дело с обычной частицей, несущей энергию и импульс.

6. Фотонные духи и фотоны-двойники

Как известно, плоская монохроматическая электромагнитная волна описывается волновым уравнением

$$\frac{1}{c}\frac{\partial \vec{\mathbf{A}}}{\partial t} = \Delta \vec{\mathbf{A}}$$

и имеет, например, следующий вид

$$\vec{\mathbf{A}} = \vec{\mathbf{A_0}} e^{i(\vec{k}\vec{x} - \omega t)}.$$

Електрическая и магнитная напряженности волны равны

$$\vec{\mathbf{E}} = i|k|\vec{\mathbf{A}}, \quad \vec{\mathbf{H}} = i[k \times \vec{\mathbf{A}}]. \tag{17}$$

Для тензора энергии-импульса волны имеем

$$T^{ij} = \frac{Wc^2}{\omega^2} k^i k^j,$$

где

$$W = \frac{\vec{\mathbf{E}}^2}{4\pi}$$

– плотность энергии волны.

Из приведенных формул видно, что если сделать подстановку $\vec{\mathbf{A}} \to d\vec{\mathbf{A}}$, где $d \in D$, то получим

$$\vec{\mathbf{E}} \to d\vec{\mathbf{E}} \Longrightarrow \vec{\mathbf{E}}(\ell C^\infty(\mathbb{R})/(a^2)) \neq 0 \text{ при } a \neq 0,$$

тогда как $W \to d^2W = 0$ и, следовательно, $T_{ik} \equiv 0$, т.е. имеем фотонный дух во всех вселенных мультиверса, наблюдаемый в виде электромагнитной волны, не несущей ни энергии, ни импульса во всех мирах, кроме мира с $a = 0$, где ее просто нет.

Рассмотрим теперь число $\vartheta \in R$. Пусть в стадии $\ell C^\infty(\mathbb{R}^n)/I$ оно задается классом функций $\vartheta(a) \mod I$, где

$$\vartheta(a) = e^{-k|a|^2} - 1, \quad k > 0. \tag{18}$$

Пусть электромагнитное поле

$$\vec{\mathbf{E}} = i\vartheta|k|\vec{\mathbf{A}}, \quad \vec{\mathbf{H}} = i\vartheta[k \times \vec{\mathbf{A}}], \quad \vec{\mathbf{A}} \neq 0$$

получается из (17) подстановкой $\vec{\mathbf{A}} \to \vartheta\vec{\mathbf{A}}$.

Тогда

$$\vec{\mathbf{E}}(\ell C^\infty(\mathbb{R})/(\vartheta^2)) \neq 0,$$

но

$$T^{ij} = \frac{Wc^2}{\omega^2}k^ik^j(\ell C^\infty(\mathbb{R})/(\vartheta^2)) \mod (\vartheta^2) = 0.$$

Иначе говоря, в стадии (среде) $\ell C^\infty(\mathbb{R})/(\vartheta^2)$ во всех вселенных наблюдаюся фотоны-двойники, не несущие ни энергии, ни импульса, т.е. являющиеся фотонными духами.

7. Виртуальные реальности как топосные модели формального мультиверса

Поскольку «множество действительных чисел» R в $\mathbf{Set}^{\mathbb{L}^{op}}$ не обладает многими привычными свойствами обычных действительных чисел из $\mathbb{R}$, то, пребывая в средах этого генератора виртуальной реальности, мы должны были наблюдать неожиданные или непривычные факты и явления. Некоторые из них были описаны в данной статье.

Топос $\mathbf{Set}^{\mathbb{L}^{op}}$, как уже говорилось, не единственная допустимая модель для формальной теории $\mathcal{T}$. Обращение к другим моделям, другим генераторам виртуальной реальности приведет нас к знакомству с другими возможными реальностями, но трудно сказать, какая из них ближе к той, которая носит название *окружающая нас физическая реальность*.

Литература

1. Дойч Д. *Структура реальности.* Ижевск: НИЦ «Регулярная и хаотическая динамика», 2001.
2. Kock A. *Synthetic Differential Geometry.* Cambridge Univ. Press, 1981.
3. Guts A.K., Grinkevich E.B. *Toposes in General Theory of Relativity.* – Los Alamos E-print paper: gr-qc/9610073 (1996). - http://xxx.lanl.gov/abs/gr-qc/9610073
4. Гуц А.К. *Интуиционистская теория пространства-времени* // Международная геометрическая школа-семинар памяти Н.В.Ефимова. Тезисы докладов. Абрау-Дюрсо. 27 сентября - 4 октября 1996 года.- Ростов-на Дону,1996.- С.87-88.
5. Дирак П. *Принципы квантовой механики.* М.: Наука, 1979.
6. Gödel K. *An Example of a New Type of Cosmological Solution of Einstein's Field Equations of Gravitation* // Rev. Mod. Phys. 1949. V.21, No.3. P.447-450.
7. Moerdijk I., Reyes G.E. *Models for Smooth Infenitesimal Analysis.* Springer-Verlag, 1991.
8. *Квантовая механика Эверетта* // Сайт в Интернет http://www.univer.omsk.su/omsk/Sci/Everett.
9. Guts A.K., Zvyagintsev A.A. *Interpretation of intuitionistic solution of the vacuum Einstein equations in smooth topos.* – Los Alamos E-print Paper: gr-qc/0001076 (2000).
10. Гуц А.К., Звягинцев А.А. *Решение почтивакуумных уравнений Эйнштейна в синтетической дифференциальной геометрии Кока-Ловера* // Математические структуры и моделирование. 2000. Вып.6. С.115-127.
11. Гуц А.К., Звягинцев А.А. *Интуиционистская логика и сигнатура пространства-времени* // Логика и приложения. Международ. конференция, посвящ. 60-летию Ю.Л.Ершова. Тезисы докладов. – Новосибирск: Ин-т дискрет. мат-ки и информатики, 2000. С.38-39.
12. Гуц А.К. *Многозначная логика и многовариантный мир* // Логика и приложения. Международ. конференция, посвящ. 60-летию Ю.Л.Ершова. Тезисы докладов. – Новосибирск: Ин-т дискрет. мат-ки и информатики, 2000. С.36-37.
13. Guts A.K. *Interaction of the Past of parallel universes.* - Los Alamos E-print Paper: physics/9910037 (1999).
14. Гуц А.К. *Модели многовариантной истории* // Математические структуры и моделирование. 1999. Вып.4. С.5-14.
15. Белокуров В.В., Тимофеевская О.Д., Хрусталев О.А. *Квантовая телепортация – обыкновенное чудо.* Ижевск: *R&C Dynamics*, 2000.
16. Palesheva E.V. *Ghost spinors, shadow electrons and the Deutsch Multiverse.* – Los Alamos E-print paper: gr-qc/0108017 (2001).

Математические структуры и моделирование
2001, вып. 8, с. 91–101

УДК 530.145

Посвящается памяти моего научного руководителя В.С. Вербанова

ВОЛНЫ ВЕРОЯТНОСТИ В СТОХАСТИЧЕСКИХ ПРОЦЕССАХ

С.В. Демин

In double-slit experiment in random-walk of the macroparticles waves of probability are discovered. A preferred values of energy of these particles corresponds to them. Describing these phenomenas, author tries to apply laws of quantum mechanics, replacing constant $\hbar$ by the parameter, which is a characteristic of the physical system.

1. Введение

В последние годы в физике активизировалось обсуждение проблем интерпретации квантовой механики. Способствовало этому экспериментальное подтверждение предсказанных еще в первой половине прошлого века парадоксов, из которых, быть может, наиболее известен парадокс Эйнштейна-Подольского-Розена [1]. Из множества современных работ, посвященных этой теме, можно привести, например, статью [2]. Зарегистрированные в недавних экспериментах эффекты противоречат и постулатам теории относительности, и обыденному здравому смыслу, что проявляется, в частности, в нарушении принципа причинности[1]. Однако считается, что это противоречие, по крайней мере, не выходит за рамки законов микромира, несмотря на макроскопический характер наблюдаемых эффектов, и что сверхсветовая передача сигналов, а также передача сигналов обратно во времени с помощью квантовых корреляций невозможна [3]. По мнению большинства физиков, способностями к такому «сверхъестественному общению» наделены только микрочастицы, и эти их способности не могут быть использованы в создаваемых нами технических устройствах.

Парадоксы, иллюстрирующие проявления нелокальности[2] через регистрацию неклассических (квантовых) корреляций пространственно-временных переменных [5,6], вытекают из ортодоксальной интерпретации квантовой механики. Именно поэтому многие пытаются решить возникшие проблемы, предлагая

E-mail: asup_man@xl.ru

[1]Будущее не может влиять на прошлое.

[2]Нелокальность - влияние двух удаленных измерительных приборов друг на друга, обычно связывается с нарушением неравенств Белла [4].

обратиться к альтернативным интерпретациям. Однако ни теоретическая дискуссия наподобие той, что велась между Эйнштейном и Бором, ни экспериментальное подтверждение в общем-то давно известных парадоксов, как видно, не в состоянии существенно повлиять на ситуацию, сложившуюся на сегодня в физике. Пожалуй, к каким-либо кардинальным изменениям в этой науке могло бы привести появление новых опытных данных, не вписывающихся в устоявшиеся физические концепции. Возможно, описанные в статье эксперименты указывают на ту область, в которой следует искать такие данные.

Толчком к опытам, которые были начаты автором статьи много лет назад, послужил крамольный, с точки зрения ортодоксальной физики, вопрос: нарушается ли принцип макроскопической причинности? Влияет ли будущее состояние макрообъекта при каких-то условиях на его состояние в прошлом? Именно результаты опытов, которые можно было рассматривать в качестве утвердительного ответа на этот вопрос [3], заставили автора статьи искать у макрообъектов свойства, напоминающие квантово-волновые свойства микрочастиц, что привело к наблюдению явлений, подобных интерференции в двухщелевом эксперименте, а также к регистрации «дискретных» (или, правильнее, предпочтительных?) энергетических состояний макросистемы, соответствующих наблюдаемым «интерференционным» явлениям.

2. Постановка задачи

Обнаружить закономерности, подобные квантово-механическим, в движении макрообъектов проще всего, изучая случайные процессы. К этому выводу приводят элементарные рассуждения. Во-первых, волновые свойства частиц предполагают либо принципиальное отсутствие траектории их движения (что трудно себе наглядно представить, но на чем настаивает большинство физиков), либо хоть какую-то ее неопределенность, имеющую место, например, при случайном движении. Во-вторых, чтобы нарушения принципа причинности, которые могут возникать в силу временной нелокальности протекающих процессов, не приводили к неразрешимым логическим парадоксам, всякое такое нарушение должно находить разумное объяснение. Допустим, причина некоего явления находится в будущем. Если это явление можно рассматривать как флуктуацию в каком-либо стохастическом процессе, то подобное объяснение вполне удовлетворяет и требованию разумности, и классическому пониманию реальности. Правда, в этом случае появляются противоречия с некоторыми общепринятыми положениями устоявшейся научной парадигмы, касающимися повторяемости результатов и независимости их от времени и места проведения опыта. Но в этой статье мы лишь слегка коснемся этого большого вопроса.

Вообще говоря, строго детерминированное движение, которое описывается законами классической механики, является идеализацией, и в природе в любом реальном процессе можно найти элементы случайности. Для эксперимента по поиску подобия квантовых закономерностей в стохастических процессах было

[3] В данной статье содержание этих опытов не описано.

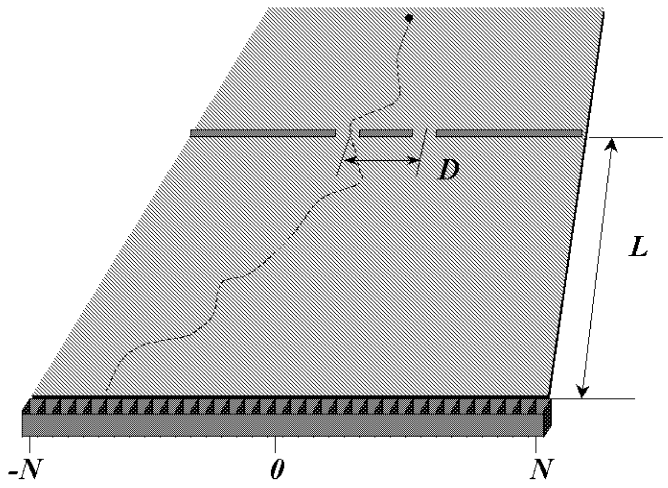


Рис. 1. Наклонная доска для исследования волновых явлений в стохастических процессах

выбрано блуждающее движение шариков по наклонной плоскости – некий аналог известной доски Гальтона (см. рис.1). Шарики, поверхность которых не является идеально сферической, скатывались по плоскости, наклоненной под определенным углом, по случайной траектории, расположенной внутри некоторого конуса, после чего попадали в расположенные в нижней части плоскости ячейки. При этом распределение шариков по ячейкам, как известно, должно подчиняться закону Гаусса (см. рис.2). Характер случайных отклонений от теоретически рассчитанной кривой нормального распределения составил основной предмет изучения в данной работе.

3. Экспериментальные данные

Для моделирования явлений, подобных интерференционным, в средней части наклонной плоскости устанавливалось препятствие с двумя узкими отверстиями, через которые скатывающиеся шарики могли попадать на нижнюю половину плоскости (см. рис.1). Расстояние между отверстиями D в разных опытах составляло 32, 40 и 22 мм. В первом опыте через отверстия прошли 1860 шариков, во втором – 2561, в третьем – 9697. Распределение этих шариков по ячейкам аппроксимировалось нормальным распределением:

$$f_a(n) = \frac{A}{\sqrt{2\pi}\sigma} e^{-\frac{(na-\mu)^2}{2\sigma^2}},$$

где n – номер ячейки от $-N$ до N, σ – стандартное отклонение, a – ширина ячейки.

После этого были найдены разности между экспериментальными значениями и значениями, рассчитанными в результате аппроксимации (см. гист.F на рис.3 и 4):

$$F(n) = f_e(n) - f_a(n).$$

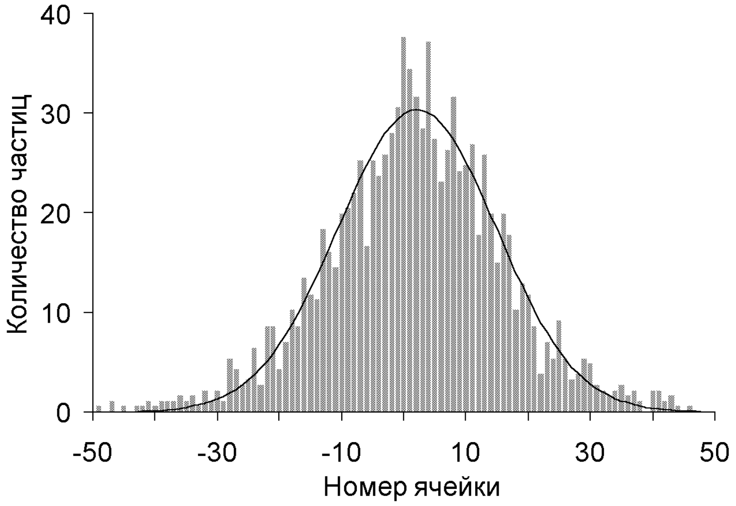


Рис. 2. Пример распределения частиц, скатившихся по наклонной доске. Сплошной линией показана аппроксимация распределением Гаусса.

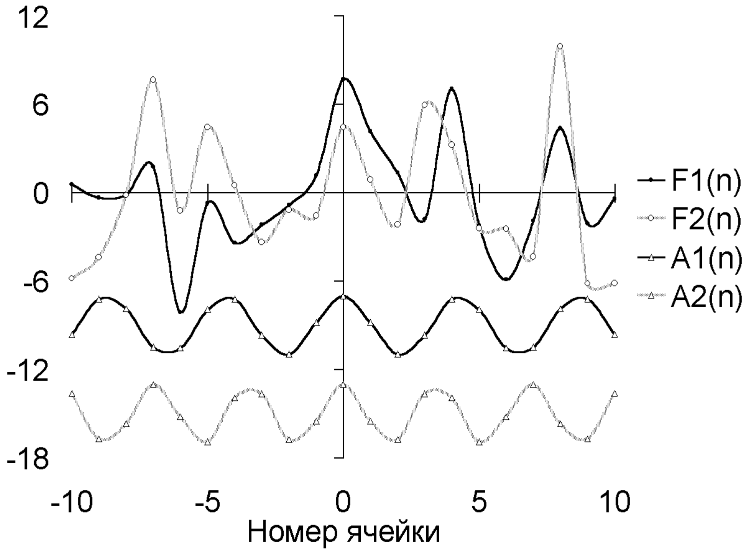


Рис. 3. Отклонение реального распределения от распределения Гаусса при расстоянии между отверстиями 32мм – $F1(n)$ и 40мм – $F2(n)$. Кривые $A1(n)$ и $A2(n)$ показывают теоретически рассчитанное положение максимумов для интерференции волн длиной 2мм.

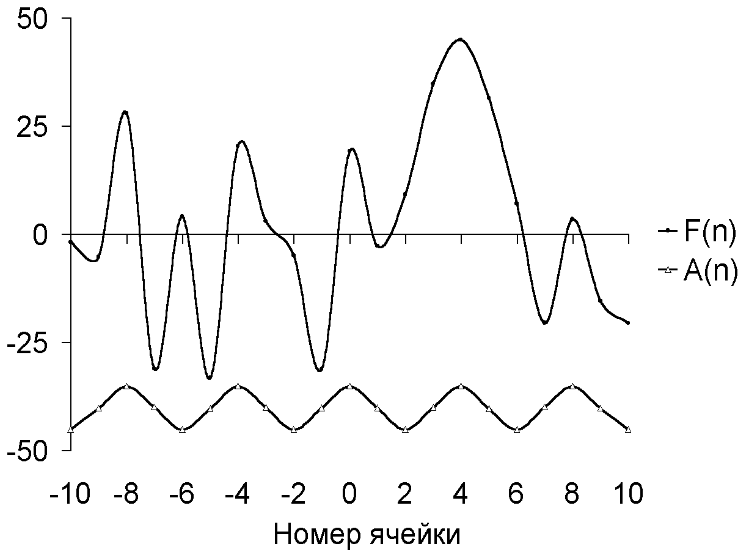


Рис. 4. Отклонение реального распределения от распределения Гаусса при расстоянии между отверстиями 22 мм во втором эксперименте – $F(n)$. Кривая $A(n)$ – теоретически рассчитанное положение максимумов для интерференции волн длинной 1,04 мм.

Поскольку полученные отклонения от нормального распределения выглядели более или менее периодическими, была сделана попытка оценить длину волны, которая при прохождении через соответствующие отверстия давала бы интерференционную картину с аналогичным положением максимумов и минимумов. Для этого к полученным рисункам были добавлены графики функций, положение экстремумов которых точно соответствует расчетному при интерференции волн той или иной длины. Амплитуда этих функций вычислялась по формуле (3).

$$A(n,\lambda) \sim \cos\left(\frac{2\pi}{\lambda}\left(\sqrt{\left(na+\frac{D}{2}\right)^2+L^2}-\sqrt{\left(na-\frac{D}{2}\right)^2+L^2}\right)\right),$$

где λ – длина волны, n – номер ячейки от $-N$ до N, a – ширина ячейки, D – расстояние между щелями, L – расстояние от препятствия со щелями до ячеек.

Какое значение длины волны при этом выбрать, первоначально определялось простым подбором. Затем с целью подтверждения правильности сделанного выбора был вычислен коэффициент ковариации между функциями, моделирующими интерференцию при всевозможных длинах волн $A(n,\lambda)$, и функциями $F(n)$, рассчитанными из реальных гистограмм распределения. Зависимость полученного коэффициента ковариации от длины волны λ в разных опытах показана на рис.5. Для того чтобы оценить распределение кинетической энергии шариков, замерялось время их прохождения от одной отметки до другой. Зная расстояние между отметками и массу шариков, легко рассчитать их среднюю скорость и далее кинетическую энергию. На представленных на рис.6 и 7 гистограммах показано распределение количества шариков по их скорости при различной величине наклона плоскости.

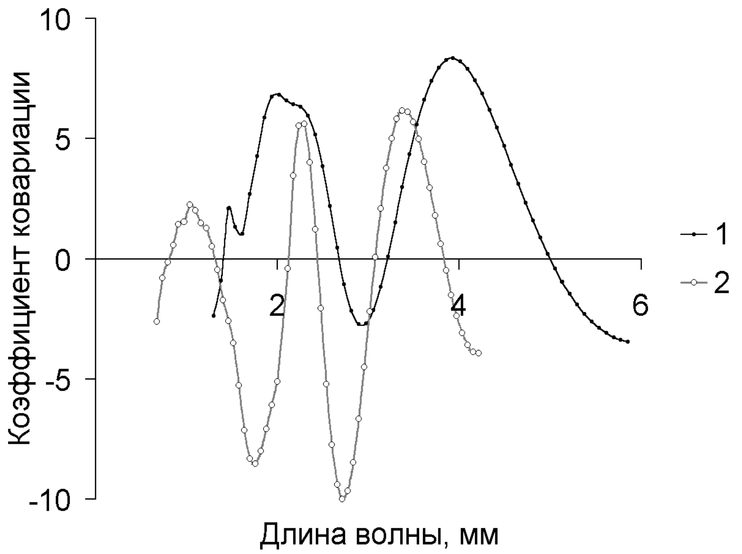


Рис. 5. Определение вероятной длины волны при двухщелевой интерференции по коэффициенту ковариации между экспериментально полученной и теоретически рассчитанной зависимостями. Данные по двум экспериментам. Линии сглажены.

4. Обсуждение

Даже беглый взгляд на представленные диаграммы вызывает желание спросить: а какова достоверность полученных результатов? И соответственно: а имеют ли место вообще эти результаты и есть ли предмет для разговора? Не стоило ли проявить большую аккуратность при постановке опытов или хотя бы просто увеличить количество измерений? Корректно ли вообще делать какие-либо выводы на такого рода материале, тем более выводы, претендующие на фундаментальность?

Возможно, в основе всей данной работы лежит ответ именно на этот вопрос.

Механика Ньютона прекрасно описывает движение макроскопических тел. Наивно ожидать, что результаты простейших экспериментов, подобных описанным выше, придут с ней в открытое противоречие. Такого рода противоречия наверняка были бы замечены много ранее. Следует, однако, заметить, что эта механика имеет дело исключительно с детерминистским аспектом движения. Даже описание случайных процессов в классической физике в конечном итоге сводится к строгому определению границ, в которые должна попасть та или иная физическая величина. Само выражение «погрешность измерения» говорит об отношении физиков ко всему тому, что происходит в мире за рамками этих строгих границ. Считается аксиомой то, что шум, сопровождающий любой эксперимент, не может нести в себе никакой полезной информации, и главная задача экспериментатора в том, чтобы избавиться от этого шума, произведя, например, достаточное количество измерений.

Эта работа написана для тех, кто готов заглянуть за кулисы законов Ньютона, зайти на территорию, которая всегда в науке считалась запретной. На самом деле, готовы ли мы представить себе законы природы, которые бы действовали

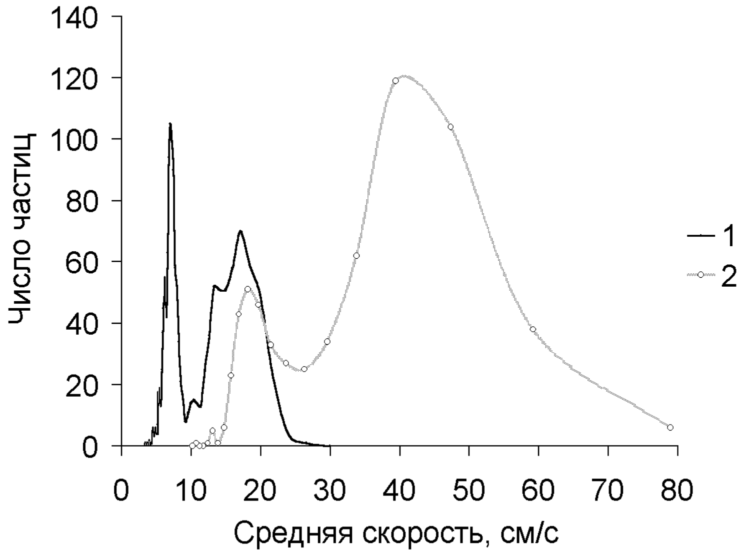


Рис. 6. Распределение частиц диаметром 2мм по скоростям в двух экспериментах с различной величиной наклона плоскости.

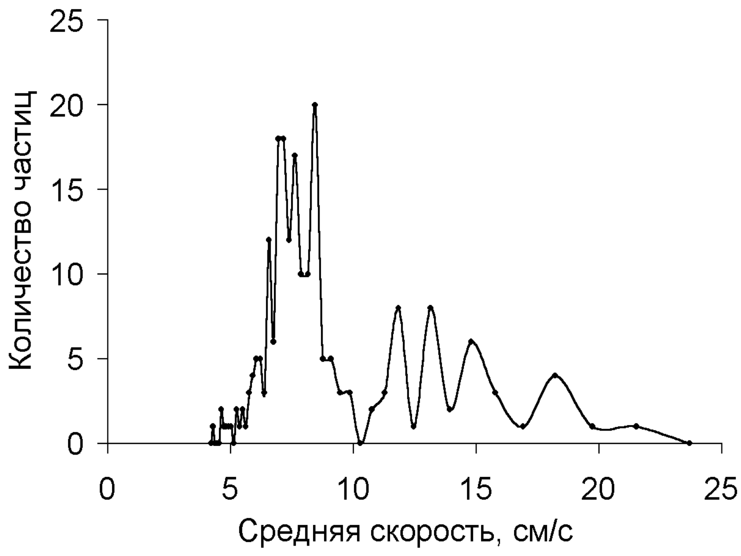


Рис. 7. Распределение количества частиц по скоростям при минимальном наклоне плоскости

исключительно на уровне погрешности измерений, сколько бы этих измерений ни было произведено. Такое допущение сразу приводит к необходимости признания макроскопической временной нелокальности протекающих физических процессов, потому что повторные измерения должны быть связаны между собой таким образом, чтобы при их суммировании достоверность результата не могла бы выходить за некий предел. Многим это покажется противоречащим и опыту, и здравому смыслу. С другой стороны, нелокальности и нарушения принципа причинности не противоречат квантовым законам, которые мы попытаемся применить к макроскопическим процессам.

Разумеется, речь здесь может идти лишь о внешнем сходстве с известными законами квантовой механики или, в крайнем случае, о разных проявлениях неких общих закономерностей. Прежде всего квантовые законы подразумевают наличие волновых свойств у частиц. Применение известной формулы (1), связывающей волновые характеристики частицы с ее энергией, дает для макротел бесконечно малые значения длин волн.

$$E = \hbar\omega. \tag{1}$$

Это объясняется тем, что слишком мало значение постоянной Планка, входящей в данную формулу. Подобие квантовым закономерностям, которое здесь будет предложено, состоит в том, чтобы, сохранив общий вид зависимостей, аналогичных (1), поставить на место постоянной h некий произвольный параметр z, который будет являться квантовой характеристикой системы взаимодействующих макротел. Тогда, используя соотношение де Бройля, можно записать:

$$\lambda = \frac{z}{mv}.$$

После этого остается обнаружить проявления волновых свойств макротел по классическим тестам на дифракцию или интерференцию и, сопоставив вычисленную длину волны со значениями импульса использованных в эксперименте объектов, оценить величину параметра z.

Как же выглядят, с этой точки зрения, приведенные выше экспериментальные данные? Прежде всего хочется сказать, что, как и ожидалось, никакой принципиальной разницы в результатах опытов со ста, тысячью или десятью тысячами измерений замечено не было. В двухщелевом эксперименте, с одной стороны, с первых же измерений наблюдались отклонения от нормального распределения, которые выглядели более или менее периодическими, напоминая картину интерференционных максимумов и минимумов, а с другой стороны, по мере накопления данных эти отклонения постепенно уменьшались, впрочем, так окончательно и не исчезая. Некоторые из симметричных локальных экстремумов на гистограммах были очень устойчивы и соответствовали довольно значительным по величине отклонениям от значений, полученных в результате аппроксимации экспериментальной гистограммы колоколообразной кривой нормального распределения. При изменении расстояния между щелями экстремумы первого и второго порядка в основном вели себя именно так, как и положено по законам интерференции: чем ближе щели, тем больше расстояние между максимумами. В этой статье приводятся не самые убедительные в этом отношении данные, потому что ее объем не позволяет привести все полученные результаты.

Трудно делать какие-либо выводы на основании качественных результатов, представленных на рис.3 и 4. Можно ли считать относительную периодичность следования пиков на гистограммах достаточно убедительным свидетельством того, что мы имеем дело с неким подобием интерференционных явлений? Скорее всего нет.

Иное отношение к этим результатам складывается при взгляде на зависимость коэффициента ковариации между моделирующей интерференцию функцией $A(n, \lambda)$ и экспериментальной кривой $F(n)$ от длины волны λ, по которой строилась моделирующая функция (см. рис.5). Можно предположить, что максимальные значения коэффициента ковариации соответствуют длинам волн различных макрочастиц, присутствующих в потоке. Тогда, исходя из формулы (1), следовало бы ожидать соответствующего «квантования» значений энергии этих частиц.

Пожалуй, именно построение распределения скатывающихся частиц по их средней скорости дало наиболее неожиданные результаты. Внешне движение шариков выглядело совершенно хаотичным, и трудно было себе представить, по какой причине их средняя скорость могла бы иметь два ярко выраженных предпочтительных значения (см. рис.6). И даже, если предположить, что это явилось следствием того, что форма шариков была более близка к эллипсоиду, чем к идеальному шару, то и в этом случае остается совершенно непонятным, почему наиболее гладкая кривая получилась при наибольшей скорости. Казалось бы, уменьшение угла наклона плоскости, приводящее к тому, что траектория движения шариков становилась менее прямолинейной и, соответственно, более случайной, а их скорость в любой точке траектории могла измениться произвольным образом, должно было размыть кривую распределения их по скоростям. В действительности при этом на кривой распределения появились дополнительные пики, которые при достаточно большом числе измерений (1500 частиц) стали совершенно отчетливыми и, вне всякого сомнения, достоверными, далеко выходящими за пределы случайных отклонений (см. гист.1 на рис.6). Еще более ярко это выразилось при минимально возможном наклоне плоскости (см. рис.7).

С другой стороны, такое поведение физической системы хорошо укладывается в первоначально высказанное предположение о том, что квантовые закономерности проявляются в макромире исключительно в стохастических процессах. Чем больше элемент случайности, тем ближе поведение системы к тому, что мы привыкли наблюдать в микромире. К этому позитивному взгляду на полученные результаты можно добавить и попытку оценить величину параметра z, стоящего в формуле (5). В приведенных опытах использовались шарики радиусом 1 мм и массой 6 мг. С учетом этого был рассчитан параметр z при различных длинах волн и предположительно соответствующих им скоростях частиц. В таблице 1 приведены результаты этих расчетов.

Видно, что параметр, стоящий на месте постоянной Планка, действитель-

Таблица 1.

Номер эксперимента	Длина волны, см (рис.5)	Скорость в пиках, см/с (рис.6)	Параметр z, г см/с^2
1	0,2	17,14	0,021
	0,4	8,28	0,020
2	0,104	42	0,026
	0,227	18	0,025
	0,338	13	0,026

но может быть характеристикой физической системы, в которой происходят случайные процессы. То, что его величина получилась примерно одинаковой для двух длин волн, зарегистрированных в первом опыте, и трех длин волн во втором, говорит в пользу того, что сопоставление предпочтительных значений импульса частицы с этими длинами волн не является случайным.

5. Заключение

Простота поставленных опытов позволяет надеяться на то, что выводы, которые из них следуют, сделаны не на пустом месте и не являются следствием тех или иных недостатков, которые, конечно же, имели место в силу слабого материального обеспечения экспериментов. После просмотра всего представленного материала остается впечатление, что, несмотря на малоправдоподобность следующих из него выводов, вероятность того, что он отражает реально существующие закономерности, все же больше вероятности того, что в его основе лежат многократно повторенные совпадения и незамеченные ошибки.

Другая сторона вопроса состоит в том, что волновые свойства макроскопического объекта, который движется пусть по случайной, но все же совершенно определенной траектории, должны иметь какой-то физический смысл. Разумеется, можно в очередной раз сказать, что подчиненность стохастического движения макротел некоторой абстрактной волновой функции является новым странным свойством мира, которое следует принимать как есть, не обращая внимания на то, можно ему дать наглядную интерпретацию или нет. Однако этот трюк, давший однажды положительные результаты, вряд ли удастся применить к макроскопическим процессам. Описывать законы движения никогда и никем не видимых элементарных частиц – это одно, а вводить нематериальную волновую функцию, которой непостижимым образом вдруг окажутся подчинены все случайности нашей обыденной жизни, – совсем другое.

Если позволить себе немного пофантазировать, то в качестве альтернативного подхода, основанного на том отношении к физической реальности, которому в свое время симпатизировал Эйнштейн, можно предложить модель мира, имеющего дополнительное временное измерение, в направлении которого не происходит свободного переноса вещества и энергии, а взаимодействие осуществляется только через волны вероятности. В рамках этой модели видимая траектория частицы представляет собой одну из проекций движения многомерного объекта на четырехмерную пространственно-временную область, в которой производится измерение. Можно сказать, что по другим возможным траекториям данная частица или, вернее, ее продолжение движется в пространствах, смещенных во времени относительно нашего пространства на тот или иной интервал. Все возможные пути частицы, описываемые волновой функцией, таким образом, становятся реальными, существующими в скрытых или ненаблюдаемых измерениях.

И пусть это пока всего лишь фантастическая гипотеза. Главное, что она дает хоть сколько-нибудь приемлемое объяснение и описанным результатам, и тем, что остались за рамками данной статьи, например, противоречащей здравому смыслу реакции физической системы на воздействие, которое еще только произойдет через пару часов. Отбросить эти результаты как недостоверные и случайные проще всего, но гораздо труднее найти путь к истине.

Литература

1. Einstein A., Podolsky B., and Rosen N. // Phys. Rev. 1935. V.47, P.777; УФН. 1936. V.16. P.440.
2. Widom A., SrivastavaY.N., Sassaroli E. *Correlation between future and past photon events.* // Phys. Lett. A. 1995. V.203, N.5–6. P.255–259.
3. Scherer H., Busch P. *Problem of signal transmission via quantum correlation and Einstein incompleteness in quantum mechanics.* // Phys. Rev. A. 1993. V.47, N.3. P.1647–1651.
4. Bell J.S. // Physics. 1964. V.1. P.195.
5. Lamoreaux S.K. *Review of the experimental tests of quantum mechanics.* // Int. J. Mod. Phys. A. 1992. V.7, N.27. P.6691–6762.
6. Shih Y.H., Sergienko A.V., Rubin M.H. *Einstein-Podolsky-Rosen state for space-time variables in a two-photon interference experiment.* // Phys. Rev. A. 1993. V.47, N.2. P.1288–1293.

Математические структуры и моделирование
2001, вып. 8, с. 102–106

УДК 53.01.91

О ВОЗМОЖНОСТИ ПЕРЕРАБОТКИ СЫРЬЯ СЛОЖНОГО СОСТАВА В ВЫСОКОЧАСТОТНОЙ ИНДУКЦИОННОЙ ПЛАЗМЕ

О.Т. Данилова

High-frequency induction plasma characteristics have been investigated by flow visualisation technique with high-speed camera shooting. The heating process of heterogeneouses particles injected into a plasma jet is analysed.

Данная работа проводилась с целью разработки основ плазменной технологии и оборудования для получения порошков оксидов металлов из отходов систем промышленной водоподготовки в плазме высокочастотного индукционного разряда. Отличительной чертой такой технологии является способность воспринять в качестве сырья материал, находящийся в состоянии порошка, или рационально трансформировать в порошкообразную форму сырье, находящееся в недиспергированном состоянии. Кроме того, эта технология позволяет производить разложение любых опасных примесей без выброса в атмосферу, является безреагентной, позволяет устранить такие широко распространенные переделы, как нейтрализация, осаждение, сушка, прокалка, солевой сброс.

Плазменное выделение УДП оксидов металлов состоит из следующих стадий:

1) генерация плазмы необходимого состава в требуемом диапазоне температур и давлений;
2) ввод реагентов — веществ в твердом или жидком состояниях и обеспечение необходимого времени их контакта с плазмой;
3) вывод целевого продукта из зоны реакции.

Основным элементом любой плазменной установки является источник питания – высокочастотный генератор, в котором происходит преобразование тока промышленной частоты в ток высокой частоты, используемый для питания индуктора плазмотрона. Второй элемент – плазмотрон, основным назначением которого является оптимальное поглощение колебательной мощности ВЧ-генератора с минимальными потерями. Таким образом, для определения оптимальных режимов технологического процесса следует рассчитать параметры ВЧИ-разряда с целью определения способов управления его температурой и обозначить конструктивные изменения ВЧ-генератора.

E-mail: danilova@phys.omskreg.ru
Омский государственный университет

Для представления модели технологического процесса был выбран нескинированный (с толщиной скин-слоя $> 0,5\ R_0$) высокочастотный индукционный (ВЧИ) разряд контрагированной формы с температурой плазмы более низкой, чем в используемом обычно скинированном разряде. Контрагированный разряд более технологичен, так как наряду с возможностью повышения эффективности его использования за счет понижения температуры он позволяет значительно повысить ресурс работы реактора из-за меньших тепловых нагрузок на стенки камеры, а геометрия плазмы в нем удобна для проведения плазмохимических процессов непосредственно в зоне генерации плазмы.

Температура ВЧИ-разряда для заданного диапазона температур определяется из условия баланса энергий [1]: выделяемой за счет поглощения плазмой электромагнитной энергии S_0 и отводимой теплопроводностью W. Для случая большого скин-слоя S_0 и W определяются выражениями:

$$S_0 = \frac{4\pi\sigma f^2 r^2 (I_o n)^2}{c^2},$$

$$W = \frac{8\pi\lambda k T^2}{eU},$$

где f – частота ВЧ-поля; σ – электропроводность плазмы; I_0 -число ампер-витков на единицу длины индуктора; r – радиус разряда; λ – теплопроводность

Расчеты показывают, что в режиме генерации ВЧИ-разряда с большой глубиной скин-слоя возможны температуры от 3400 K до 6700 K. Однако при таких температурах для эффективного поглощения мощности ВЧ-генератора нужно увеличивать объем разряда, чего можно добиться, применяя длинный индуктор. Разряд в этом случае должен вытягиваться и переходить в контрагированную форму. Для поддержания такого нескинированного разряда необходимо решить две задачи:

1) найти способ стабилизации разряда в неустойчивом состоянии;
2) согласовать ВЧ-генератор с достаточно длинным индуктором, позволяющим получать разряд в контрагированной форме и вкладывать в него большие энергии.

Работа ВЧ-генератора на такую нагрузку, как плазма, предъявляет к устройству ряд специфических требований:

- в генераторах должна быть предусмотрена широкая регулировочная характеристика, обеспечивающая изменение мощности от номинальной почти до нуля;
- необходимо обеспечение устойчивой работы в режиме холостого хода (т.е. до возбуждения разряда) и автоматическое получение заданного режима после его возбуждения;
- для целого ряда случаев требуется получение значительной напряженности магнитного поля в заданном диапазоне температур;
- генераторы должны удовлетворять требованиям стабильности частоты в пределах разрешенных узких радиоканалов.

Выполнение перечисленных требований обеспечивается выбором схемы высокочастотной части автогенератора. В настоящей работе исследовались режимы генераторов, работающие с самовозбуждением, мощностью 40 и 60 кВт и рабочими частотами 0,44; 5,28; 13,56 МГц.

Расчет частоты генерации основывается на выполнении двух критериев:

- условие баланса фаз;
- условие баланса амплитуд: $K\beta \geq 1$, где K – коэффициент усиления, $a\beta$– коэффициент обратной связи.

Первое условие сводится к выполнению соотношения

$$Z_{ak} + Z_{ag} + Zgk = 0,$$

где Z_{ak}, Z_{ag}, Z_{gk} — соответственно комплексные сопротивления, действующие на участках анод-катод, анод-сетка и сетка-катод.

Условие баланса амплитуд фактически означает, что $\beta > 1$, так как обычно $K \gg 1$. С учетом этого условие баланса может быть записано как

$$\frac{Z_{gk}}{Z_{ak}} > 0.$$

Алгоритм расчета ВЧ-генератора на заданную рабочую частоту сводится к следующему:

1) задаются реактивные сопротивления элементов схемы ВЧ-генератора;
2) рассчитываются частота генерации и коэффициент обратной связи;
3) корректируются значения реактивных сопротивлений для получения требуемой частоты;
4) по результатам расчета вносятся изменения в конструкцию генератора.

Поскольку при корректировке значений реактивных сопротивлений имеется некоторый произвол, в ходе расчета и настройки генератора учитывались требования, вытекающие из анализа результатов численного моделирования ВЧИ-разряда. Суть этих требований заключается в следующем. Для получения устойчивого режима генерации нескинированного разряда необходимо обеспечить поглощение подводимой ВЧ-мощности областью относительно прозрачной для ВЧ-поля плазмы. А это возможно лишь за счет увеличения объема этой области, т.е. при фиксированном радиусе разрядной камеры – за счет увеличения длины плазменного столба. Следовательно, необходимо увеличивать длину индуктора, а значит, и его индуктивность. Поэтому критерием, устраняющим произвол в выборе реактивных сопротивлений, является условие максимального значения индуктивность анодного контура автогенератора. В таблице 1 приведены результаты расчетов параметров эквивалентных схем ВЧ-генераторов с разными рабочими частотами.

Исследования режимов схем генераторов показали устойчивость генерации, возможность изменения выходной мощности и напряжения на индукторе за счет регулирования коэффициента обратной связи путем изменения емкости C_{oc}.

Идентификация исходного сырья методом рентгеноспектрального анализа выявила присутствие в шламе до 46 процентов $Fe(OH)_3$, не имеющей кристаллической структуры, а также наличие оксидов Si, Ti, Ca, K и др.

Таблица 1.

Мгц	K_{oc}	L_a	L_{oc}	L_e	C_a	C_{ca}	C_{oc}
0,44	-	55	80	14	2200	65	1500
5,28	0,15	6,5	2.2	-	65	65	2200
13,56	1,7	1,3	1	-	30	65	200

Термодинамический анализ полученных данных показал, что для системы шлам - плазма конденсация оксидов железа наблюдается при температуре $\sim 3000K$, в интервале температур $2200 - 3000K$ основной конденсированной фазой является FeO. При температурах ниже $2200K$ образуются конденсированные фазы Fe_2O_3 и Fe_3O_4. Очевидно, что для получения порошков оксидов в ультрадисперсном виде реагенты необходимо нагревать до температуры $3000 - 4000K$, а закалку производить при температуре $2000K$.

При обработке дисперсных материалов в плазменных струях возникает проблема нагрева частиц, объединяющих в своем объеме два или несколько веществ с различными теплофизическими свойствами. Нагрев частиц сырья из целевых продуктов в общем случае является нестационарным тепловым процессом и включает две стадии передачи тепла: конвективной теплоотдачи от газовой фазы к поверхности частицы и передачи тепла от поверхности частицы к ее центру путем нестационарной теплопроводности. Условие расплавления частиц исходного материала в модели определяется из ряда обычных допущений [2], а также на основании имеющихся расчетных и экспериментальных данных, что нагрев и разгон частиц происходит на начальном участке струи [2]:

$$\frac{T_g \lambda_g^2}{\mu_g (1+r)(1+\alpha) v_g} \geq \frac{d_p^2 D}{1092 Nu^2 \pi R_0^2 (l_k + 6(R_0 - R))},$$

где T_g – это температура плазмообразующего газа (ПГ);
λ_g – теплопроводность ПГ;
μ_g – коэффициент кинематической вязкости;
v_g – расход ПГ;
r – степень ионизации ПГ;
α – степень диссоциации ПГ;
d_p – диаметр частицы;
l_k – расстояние от места подачи порошка до среза сопла;
R_0 – радиус сопла плазмотрона;
R – расстояние от оси трубки до траектории движения частицы;
D – константа, зависящая от комплекса теплофизических свойств частицы, как-то, характер компонентов и их соотношения.

Для оценки минимальной электрической мощности плазматрона, необходимой для полного расплавления материала частицы в плазменной струе, используется аппроксимирующее выражение:

Таблица 2.

Вещество	Исходное сырье	Целевой продукт
$Fe(OH)_3$	46%	-
Fe_2O_3	-	25%
Fe_3O_4	-	49%
SiO_2	20%	3%
TiO_2	1,5%	1,2%
CaO	2,5%	2%
K_2O	7%	3%
MgO	8%	4%
Na_2O	5%	3%

$$P = 2,99\frac{v_g^{1,5} d_p D^{0,5}}{\eta R^{1,25}(l_k + 6(R_0 - R))^{0,6}},$$

где η –тепловой к.п.д. плазмотрона.

Согласно расчетам, время полного расплавления исходных частиц по всему объему происходит за времена порядка $10^{-3}-10^{-4}c$. Полученные плазмохимическим способом порошки представляют собой субмикронные частицы, имеющие сферическую или сфероидальную форму, особенно характерную для процессов с принудительным быстрым охлаждением продукта, когда скорость фазового перехода происходит очень быстро и частицы не успевают принять характерную для кристаллического состояния форму. В таблице 2 представлены результаты элементного анализа в условных процентах исходного сырья и целевого продукта.

Полученные данные хорошо согласуются с результатами термодинамического расчета. Содержание различных примесей объясняется присутствием последних в конденсированной фазе при температуре закалки $\sim\ 2200K$. Таким образом, можно сделать вывод, что применение в ресурсосберегающих технологиях по переработке металлсодержащих отходов плазмохимического способа позволяет получать порошки на основе оксидов металлов в ультрадисперсной форме, причем модельные представления процесса хорошо согласуются с экспериментальными данными.

Литература

1. Райзер Ю.П. *Физика газового разряда.* М.:Наука, 1987.
2. Юшков В.И., Борисов Ю.С. // *Труды УралНИИ Чермет.* 1971. Т.2. С.301

Математические
структуры и моделирование
2001, вып. 8, с. 107–114

УДК 51-77

МУЛЬТИ-АГЕНТНАЯ МОДЕЛЬ ИМПУЛЬСНОГО ВЛИЯНИЯ НА ПРОЦЕСС ГЕНДЕРНОГО ВЗАИМОДЕЙСТВИЯ В ОБЩЕСТВЕ

В.В. Коробицын, Ю.В. Фролова

In the paper the simulation model of gender contact of agents in artificial society is presented . We are studying the process of the male's and female's interaction based on gender relations. The females radiate the charm pulses. Responding to these pulses, the males choose a female. The computer models were realized on $\mathcal{SWARM}$. Under the computer simulation with various initial data, we reveal three types of agent interaction.

Введение

Какими должны быть мужчина и женщина и какие роли они должны играть в обществе? Пожалуй, это один из главных вопросов, возникающий при изучении гендерных взаимодействий в современном обществе. Согласно сложившимся представлениям в обществе, «настоящий мужчина» предстаёт личностью творческой, профессиональной, знающей, способной принимать решения и одерживать победы в одиночку. Его действия изменяют окружающий мир. Он самодостаточен. «Настоящая женщина» призвана сопровождать «настоящего мужчину», являться дополнительной наградой за его победы. Она предстаёт существом ограниченным, зависимым, домашним. Ей не надо быть умной и творческой личностью, а надо иметь пышные блестящие волосы, стройную фигуру, привлекательную походку. А когда благодаря этим качествам мужчина найден, ей надо следить за семейным уютом, стирать, готовить, лечить так, чтобы он оставался доволен. Он — субъект действия, творец, величие которого дополнительно подчёркнуто умением вовремя проинструктировать и поощрить представительницу слабого во всех отношениях пола. Она — объект созерцания, исполнитель, ждущий внимания, указаний и поощрений [3].

Сложившиеся в нашей культуре гендерные стереотипы об образах мужчин и женщин дают огромный толчок к исследованию общения мужчин и женщин с людьми своего и противоположного пола. Система общения людей имеет информационную структуру, представимую в виде коммуникационной модели общества. Методом исследования социальных явлений и процессов на моделях является компьютерное моделирование (Computer Simulation).

E-mail: korobits@univer.omsk.su, frolova@univer.omsk.su
Омский государственный университет

1. Описание процесса общения под влиянием излучаемых импульсов

Задачей предлагаемой компьютерной модели является попытка выявить основные моменты коммуникативных процессов в общении женщин и мужчин на основе сложившихся патриархальных образов. Данная «игра» провоцирует, искушает, дает возможность переустраивать мир по своему желанию, и эта воплощенная картинка желания крайне интересна для исследования. Вся интрига протекает между женщиной-обольстительницей и мужчиной-завоевателем.

Мужчины и женщины – равноценные люди, которые в то же время различаются по определенным признакам, называемым гендерными особенностями. Женщины обладают лишь им одним присущими механизмами влияния (такие, как красота, очарование и т.п.) на поведение мужчин, имеющих индивидуальную систему восприятия.

Как мужчина, так и женщина обладают собственным *личностным пространством.* Личностное пространство любого человека характеризуется *коммуникативным потенциалом*, отражающим состояние процесса общения. Коммуникативный потенциал — это характеристика возможностей человека, которые и определяют качество его общения. Его уровень зависит от общительности, характера и прочности контактов, устанавливаемых с другими людьми. Межличностное притяжение можно рассматривать как условие и результат совместимости двух лиц в определенных условиях взаимодействия.

Согласно проведенным социологическим исследованиям, при гендерном общении люди ориентированы на отношения доверия, понимания, поддержки, уважения, подражания, равно как и открытой осознанной борьбы преимущественно внутри собственного пола. Взаимодействие же между полами в значительной мере построено для мужчин на «восхищении», «импульсах» и «эмоциональной поддержке», для женщин отчасти на потребности практической и эмоциональной поддержки со стороны мужчин [1].

Воздействие, с помощью которого осуществляется притяжение мужчин и женщин, основано на импульсах женского обаяния. Притяжение, являясь одной из составляющих межличностной привлекательности, в основном связано с потребностью человека быть вместе с другим конкретным человеком. Женское влияние обладает различной *степенью интенсивности*, определяющей силу воздействия. Мужчина представлен в модели в виде существа, окруженного со всех сторон излучаемыми импульсами, но он способен совершить выбор, реализуя возможности индивидуальной свободы передвижения. Посредством совершаемого выбора образуется *поле «гендерного» взаимодействия* как суперпозиция двух личностных пространств. Чтобы различать способы воздействий женщин на мужчин, введено понятие *частоты воздействия*, являющейся индивидуальной характеристикой женщины. Мужчины, со своей стороны, реагируют не на всякие импульсы, а лишь на те, которые имеют частоту из его индивидуального диапазона частот. Таким образом, реализуется разнообразие мужских взглядов на женскую красоту и привлекательность.

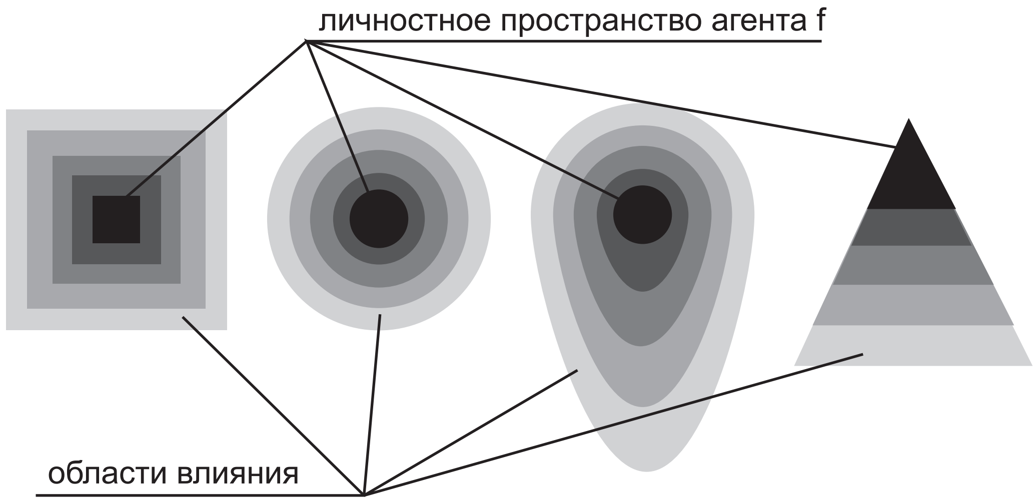


Рис. 1. Типы областей влияния

2. Построение математической модели общения

Будем рассматривать взаимодействие двух групп агентов F и M. Элементы будем обозначать буквами $f_i \in F$ и $m_j \in M$, где индексы i и j меняются в пределах от 1 до n_f и n_m соответственно. Числами n_f и n_m обозначим количество агентов из соответствующей группы.

Взаимодействие агентов происходит в пределах некоторой области P, которую мы будем называть *полем взаимодействия*. Пусть поле будет прямоугольной областью на плоскости со склеенными краями, то есть образует двумерный тор. Введем в данной области целочисленные координаты $(x, y), x = 0, 1, ..., X, y = 0, 1, ..., Y$. Числа X и Y задают размер области P. Каждому агенту f_i и m_j припишем соответствующие координаты $(x_i, y_i) \in P$ и $(x_j, y_j) \in P$.

Всякий агент f_i имеет область $O_i \subset P$ влияния на агентов из M. Области O_i могут иметь различные формы, но обязательно должны включать точку (x_i, y_i). Примеры областей приведены на рис.1. Влияние агентов f_i на m_j оказывается посредством поля взаимодействия Q, которое формируется благодаря областям влияния O_i. Поле Q есть отображение $Q : P \to I \times W$, где $I = [0, I_{max}]$ — интервал изменения интенсивности воздействий, $W = [\omega^{min}, \omega^{max}]$ — интервал изменения частоты воздействий. Поле Q формируется из объединения полей $Q_i : P \to I \times W$, где

$$Q_i(x, y) = \begin{cases} (I_i, \omega_i) & \text{при} \quad x = x_i, \quad y = y_i, \\ (C_{I_i}, \omega_i) & \text{при} \quad (x, y) \in O_i, \\ (0, 0) & \text{при} \quad (x, y) \notin O_i. \end{cases}$$

Здесь I_i — максимальное значение интенсивности агента f_i, C_{I_i} — промежуточное значение из интервала $[0, I_i]$, ω_i — частота воздействия агента f_i. В каждой точке (x_t, y_t) поле Q формируется по следующему правилу. Находим

$$Q_{max} = \max_{i=1,...,n_f} Q_i^1(x_t, y_t),$$

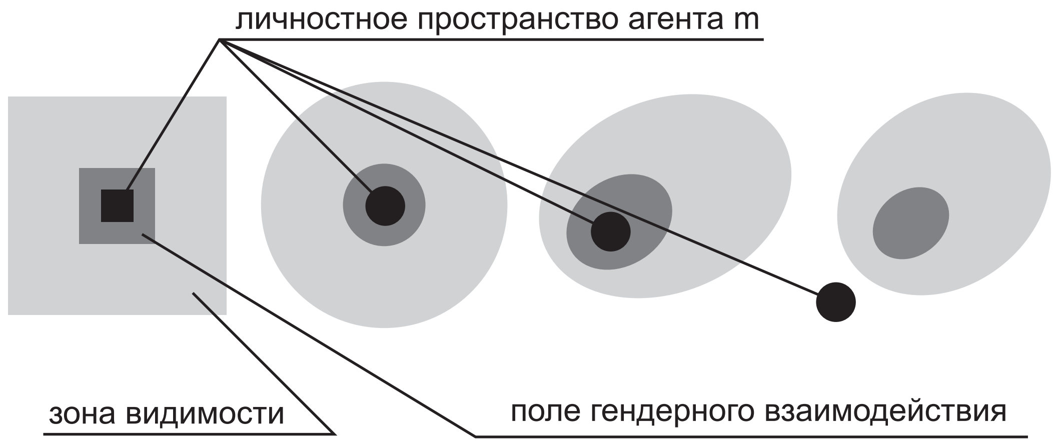


Рис. 2. Типы областей влияния

где $Q_i^1(x_t, y_t)$ — первая координата $Q(x_t, y_t)$, соответствующая интенсивности воздействия. Пусть номер i_0 будет соответствовать номеру агента f_i, при котором достигается максимум, тогда

$$Q(x_t, y_t) = (Q_{max}, \omega_{i_0}),$$

здесь ω_{i_0} — частота воздействия агента f_{i_0}.

Каждый агент m_j имеет способность просматривать поле Q в пределах некоторой зоны $Z_j \subset P$, которую будем называть *зоной видимости*. Эта зона может иметь различные формы и в общем случае может не включать точку (x_j, y_j). Кроме того, внутри области Z_j зададим область взаимодействия $T_j \subset Z_j$, определяющую точки из P, в которых возможно взаимодействие между агентом m_j и агентами из F. Пример таких областей изображен на рис.2.

Просматривая поле Q в пределах области Z_j, агент m_j реагирует только на те воздействия, которые имеют частоту в пределах заданного для него интервала $W_j = [\omega_j^{min}, \omega_j^{max}]$, и игнорирует все другие воздействия.

Правило перемещения агентов по полю. Движение агентов в поле P задается следующим образом. Координаты (x_i, y_i) агента f_i в начальный момент задаются случайным образом и в процессе моделирования не меняются. Со временем меняется область влияния O_i и интенсивность I_i. Каким образом это происходит, будет описано ниже.

Агенты m_j двигаются в направлении, соответствующем максимуму интенсивности воздействия поля Q внутри области Z_j, с частотой воздействия из интервала W_j. Для всех точек $(x, y) \in Z_j$ берем те из них, для которых $Q^2(x, y) \in W_j$. Обозначим все эти точки как множество S_j. Найдем среди точек из S_j точку, где достигается максимум интенсивности

$$Q_j^* = \max_{(x,y)\in S_j} Q^1(x, y).$$

Точкой (x^*, y^*) обозначим точку, где достигается максимум.

Если $(x^*, y^*) \in T_j$, то агент m_j способен взаимодействовать с f_i и поэтому не двигается. Точка максимума в данном случае должна соответствовать точке, где находится агент f_i. Для обеспечения данного условия необходимо, чтобы величина интенсивности поля Q, равная I_i, достигалась только в точке (x_i, y_i) и, кроме того, область T_j была строго внутри Z_j.

Если $(x^*, y^*) \notin T_j$, то агент m_j должен сдвинуться в направлении максимума. Движение должно быть задано направлением от центра области T_j к точке (x^*, y^*). За один такт времени осуществляется шаг в заданном направлении фиксированной длины s_j, приписанной данному агенту m_j. Таким образом осуществляется процесс приближения агента m_j к агенту f_i.

Правило общения агентов. Если в результате перемещения по полю агента m_j в область его возможного гендерного взаимодействия попадает агент f_i, то осуществляется процесс общения. Каждый агент f_i и m_j имеет соответствующий параметр H_i и H_j, характеризующий уровень коммуникативного потенциала. Он отражает «удовлетворенность» агента общением. В результате общения агента m и агента f меняется уровень коммуникативного потенциала обоих агентов. Изменение H_i со временем определяется по следующему правилу

$$H_i(t+1) = H_i(t) + \alpha N_i - \beta,$$

здесь α — скорость увеличения H_i, N_i — количество агентов m_j, взаимодействующих с f_i, β — скорость убывания H_i при отсутствии взаимодействия. Аналогично определяется изменение H_j

$$H_j(t+1) = H_j(t) + \alpha N_j - \beta.$$

Величина N_j взаимодействующих агентов f_i с m_j определяется количеством агентов, попавших в область T_j и имеющих частоту $\omega_i \in W_j$.

Уровень коммуникативного потенциала агента f_i оказывает влияние на изменение степени интенсивности излучаемых им импульсов воздействия. В процессе взаимодействия, согласно состоянию H_i агента f_i, меняется его интенсивность воздействия I_i по следующему правилу

$$I_i = \frac{I^0}{H_i + 1},$$

где I^0 — величина максимально возможной интенсивности.

Результат процесса общения характеризуется коэффициентом гендерного взаимодействия $\mathcal{G}$

$$\mathcal{G} = \frac{C_f}{n_f} : \frac{C_m}{n_m},$$

где C_f, C_m — количество взаимодействующих агентов из группы F и группы M соответственно. Они показывают, какое количество агентов принимают участие во взаимодействиях в данный момент времени. Таким образом, коэффициент $\mathcal{G}$ определяет соотношение долей взаимодействующих агентов группы F и M.

Перейдем к имитации построенной модели общения, воспользовавшись идеями мульти-агентного компьютерного моделирования [2]. Далее в тексте для обозначения агентов будем использовать запись «агент f» и «агент m». Мы опускаем индексы i и j, это означает, что речь идет о любом агенте заданной группы.

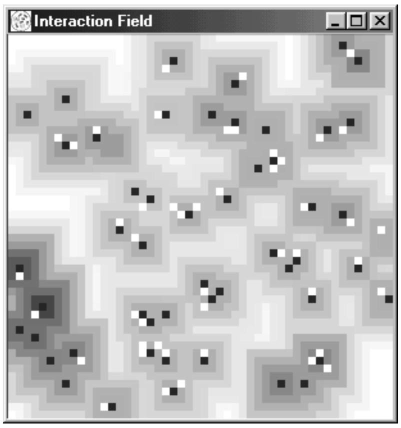


Рис. 3. Поле взаимодействия

3. Компьютерная имитация общения агентов

Для проектирования модели и проведения компьютерного эксперимента используется мульти-агентная система моделирования $\mathcal{SWARM}$, дающая возможность визуально наблюдать за ходом эксперимента. На экран выводится анимационная картинка, отображающая искусственную жизнь агентов, графики различных усредненных функций, например уровень коммуникативного потенциала или коэффициент гендерного общения.

При первоначальном запуске программы-модели появляется пользовательская панель управления, позволяющая переключать режимы работы (остановка, непрерывное и пошаговое развитие, сохранение любого этапа исследования), панель начальных данных модели, где отображаются коэффициенты и начальные значения параметров, которые исследователь может установить. К начальным данным модели относятся: размер поля взаимодействия $X \times Y$, количество агентов n_f и n_m, скорости изменения коммуникативного потенциала α и β, размер зоны видимости агента m, радиус области влияния агента f, интервал изменения частоты воздействия $[\omega_{min}, \omega_{max}]$, величина максимально возможной интенсивности I_0.

В основном окне программы (рис.3) отображается поле взаимодействия, на котором располагаются и перемещаются агенты, а также уровень интенсивности областей влияния. Агенты окрашены в разные цвета: белый — агенты m, черный — агенты f. Степень излучаемых импульсов отражена цветовой гаммой: чем темнее цвет, тем больше интенсивность излучения. Таким образом, мы можем выявить агентов f, имеющих наибольшую область влияния по окружающему его «ореолу».

Имитацию процесса взаимодействия можно наблюдать как в непрерывном, так и в пошаговом режиме. В любой момент времени можно узнать величину коммуникативного потенциала агентов, вызвав вспомогательное графическое окно данных для каждого агента. Изменяя в данном окне значение параметров, мы имеем возможность переместить любого агента, поменять его характеристики.

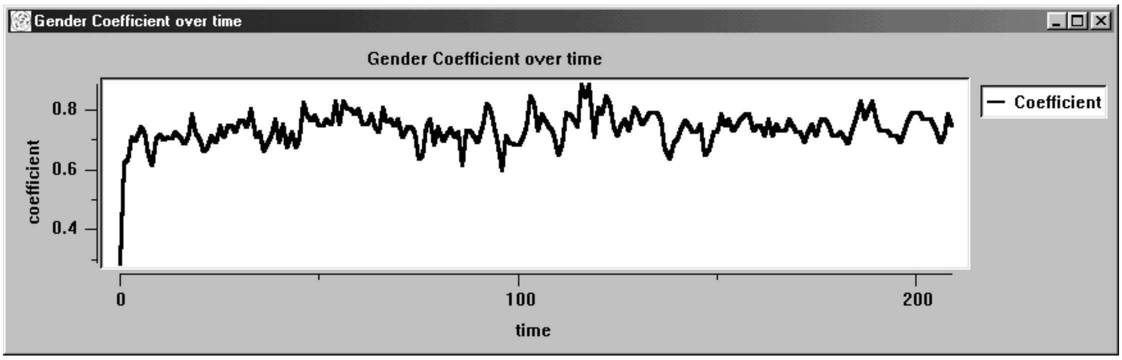


Рис. 4. График изменения значения коэффициента гендерного взаимодействия

Моделирование велось при разных начальных данных. Меняя параметры модели, мы получали различные результаты моделирования. Таким образом, можно было выявить влияние тех или иных исходных данных на процесс общения агентов.

В качестве примера приведем один из компьютерных экспериментов, в котором исследуется поведение 53 агентов f и 47 агентов m. Такое соотношение количества агентов соответствует процентному соотношению женщин и мужчин в реальном обществе. На рис.3 изображено начальное распределение агентов. Поле состоит из 2500 клеток сетки с размерами 50×50. Каждая клетка имеет параметр, характеризующий степень интенсивности влияния импульса. Максимальное значение интенсивности достигается в клетке, где находится агент f, и интенсивность падает по мере отдаления клетки от агента f. Таким образом, интенсивность меняется от I_0 (в эксперименте $I_0 = 100$) до нуля. Значение, равное нулю, означает отсутствие влияния в этой клетке. Кроме того, в эксперименте установлены следующие параметры. Значение скорости увеличения коммуникативного потенциала $\alpha = 0,015$, скорость уменьшения $\beta = 0,01$. Зона видимости агента m, область влияния агента f и область влияния агента m имеют форму квадрата со сторонами 11, 17 и 3 клетки соответственно.

Опишем процесс моделирования в динамике. Агент m двигается по полю согласно правилу перемещения и попадает в область влияния некоторого агента f. Далее он движется в направлении максимального значения интенсивности излучения. Когда агент f попадает в область гендерного взаимодействия агента m, осуществляется процесс общения между агентами. При взаимодействии агентов происходит изменение их характеристик согласно правилу общения.

В результате серии взаимодействий коммуникативный потенциал агента f увеличивается, а интенсивность влияния уменьшается. Поэтому агент m возобновляет поиск агента f с высокой степенью интенсивности излучения. Таким образом, процесс взаимодействия повторяется.

На рис.4 представлен график изменения значения коэффициента гендерного общения $\mathcal{G}$, который колеблется в интервале $[0,6;\ 0,9]$. При таких заданных параметрах получаем, что мужчины более активны, чем женщины в поиске партнера для общения.

При компьютерном моделировании с различными начальными данными, выявилось три типа взаимодействия агентов. Эти типы характеризуются коэффициентом насыщения общением $\mathcal{K}$, который определяется отношением скоростей изменения коммуникативного потенциала $\mathcal{K} = \beta/\alpha$. Согласно принимаемым значениям коэффициента $\mathcal{K}$ выделены следующие типы общения:

- первый тип «стабильный» ($\mathcal{K} \approx 1$);
- второй тип «краткосрочный» ($\mathcal{K} \ll 1$);
- третий тип «мультиконтактный» ($\mathcal{K} \gg 1$).

Стабильный тип общения возникает, когда один агент f взаимодействует с одним агентом m длительное время. Краткосрочный тип взаимодействия реализуется при взаимодействиях агентов в течение одного такта времени. Третий тип возникает, когда один агент f общается с несколькими агентами m одновременно.

Заключение

Анализируя результаты проведенного исследования можно сделать следующие выводы:

- особенность моделирования гендерного общения состоит в выделении двух типов агентов, то есть разделение людей на мужчин и женщин. Поэтому моделирование гендерных отношений отличается от моделирования межличностных взаимодействий наличием конкуренции между агентами внутри своей группы и осуществлением поиска партнера для общения среди членов другой группы (другого пола);
- хотя агенты постоянно двигаются и число взаимодействий изменяется в искусственном обществе, но значение основной характеристики процесса взаимодействия остается на среднем уровне. В нашей модели такой характеристикой является коэффициент гендерного взаимодействия. Из наблюдения динамики значения этого коэффициента мы получили следующий результат. Коэффициент $\mathcal{G}$ колеблется около 0.75. Это означает, что процент взаимодействующих женщин меньше процента взаимодействующих мужчин, несмотря на то что количество женщин больше. В этом случае мужчины играют лидирующую роль по количеству взаимодействий.

Литература

1. Арутюнян М. *Гендерные отношения в семье* // Материалы Первой Российской летней школы по женским и гендерным исследованиям «ВАЛДАЙ-1996». М.: МЦГИ, 1997. С. 133-134.
2. Гуц А.К., Коробицын В.В., Лаптев А.А., Паутова Л.А., Фролова Ю.В. *Социальные системы. Формализация и компьютерное моделирование.* Омск: ОмГУ, 2000.
3. Кон И.С. *Меняющиеся мужчины в изменяющемся мире* // http://konigor.hypermart.net/publ018_4.html

Математические структуры и моделирование
2001, вып. 8, с. 115–120

УДК 51-77

ГЕНДЕР КАК «ИНТРИГА»: ВОЗМОЖНОСТИ КОМПЬЮТЕРНОГО МОДЕЛИРОВАНИЯ В ИССЛЕДОВАНИИ ГЕНДЕРНЫХ СИСТЕМ

Л.А. Паутова, Ю.В. Фролова

The purpose of this paper was to use the possibilities of computer simulation for investigation of gender systems. Defining the gender as an intrigue allows easy to proceed to constructing the model of gender interaction in society. Submersion in world of artificial gender relations allows to understand real relations. Two simulation models of gender interactions are considered.

Введение

Компьютерное моделирование является особым знаком времени, неотделимой частью «духа эпохи». Имитация общения (chat и email), имитация работы с каталогом (система Word Microsoft), имитация чтения книги (работа с электронными публикациями в Internet), имитация хождения по музею и магазинам (виртуальные музеи и магазины) — все это стало нормой повседневного и научного образа жизни. Компьютерное моделирование, обладая своими особенностями, все же воспроизводит характеристики виртуализации мира:

- «нематериальность воздействия (изображаемое производит эффекты, характерные для вещественного);
- условность параметров (объекты искусственны и изменяемы);
- эфемерность (свобода входа/выхода обеспечивает возможность прерывания и возобновления существования)» [1, с.19].

Компьютерная симуляция имеет большие возможности в познании общества (исследование индивида, взаимодействия индивидов, семьи, социальных групп и организаций, этносов, социальных систем). Вместе с тем данный метод обладает и «гендерной сенситивностью» [2, с.15]. «Мыслить гендер», используя компьютерное моделирование, – проблематичное и даже рискованное занятие (игра в симулякры может чрезмерно отдалить от реальности). Однако, как показывает обращение к гендерной проблематике, компьютерные модели гендерных систем, воспроизводя и «играя» в реальные отношения, помогают разрабатывать собственно социологический подход.

E-mail: pautova@univer.omsk.su, frolova@univer.omsk.su
Омский государственный университет

Таблица 1. Гендер как интрига vs компьютерная модель

Гендер как интрига	**Компьютерная модель**
• Участники;	• Искусственные агенты;
• Обстоятельства, сопровождающие события;	• Правила;
• События;	• Взаимодействие агентов;
• Фон, поясняющий события;	• Искусственная среда;
• Оценка участников событий;	• Взаимодействие агентов со средой;
• Информация, соотносящая интригу с событиями.	• Изменения в среде.

1. Особенности метода компьютерного моделирования

Замещая реальные гендерные отношения компьютерной моделью, мы исходили из представления о гендере как об особого рода *интриге*.

> Как пишет И.И. Халеева, «Как в интриге, так и в дискурсе исходная структура имеет вид последовательности элементарных пропозиций, связанных между собой различными логическими отношениями. Элементами и дискурса, и интриги являются излагаемые события, их участники и так называемые «... не-события», то есть: а) обстоятельства, сопровождающие события; б) фон, поясняющий события; в) оценка участников событий; г) информация, соотносящая дискурс (или интригу) с событиями» [3].

Определение гендера в качестве интриги позволяет легко перейти к построению модели, поскольку модель есть игровая имитация, воспроизводящая близкие к интриге параметры (см. таблицу 1).

Однако, замещая и воспроизводя реальные события, модель упрощает и формализует реальность. Из богатой, «многокрасочной» социальной жизни исследователь рассматривает только некоторые аспекты. Подобное отбрасывание отдельных элементов и процессов имеет свою цель — создатель модели должен добиться четкости и логичности. Способность к логическому выводу и систематическому оперированию понятиями дает большие возможности. Язык математики «лишен двусмысленности и более точен, чем естественный язык, он позволяет исследовать скрытый смысл тончайших различий в формулировках, который плохо доступен исследованию посредством естественного языка» [4]. Формализация позволяет преодолеть чересчур свободные допущения естественного языка, исследовать скрытые связи между явлениями, рассматривать множество вариантов. Интуиция исследователя не в состоянии выдать все возможные нетривиальные и непредвиденные выводы. Формализация дает такую возможность. Имитируя социальную жизнь, она в состоянии учитывать множество вариантов и сценариев.

2. Варианты компьютерного моделирования гендерных отношений

Итак, компьютерное моделирование – это взаимодействие агентов искусственной жизни (общества) в среде. Разворачивание гендерной интриги в компьютерном мире может иметь много вариантов в зависимости от эмпирических данных, используемой социальной теории и т.д. Простейший вариант — модель с условным названием *«Влияние ресурсообеспеченности мужчин на поведение женщин»*. Для реализации проекта используется мульти-агентная система моделирования $\mathcal{SWARM}$ [5].

Исследовательская компьютерная модель представляет собой взаимодействие мужчин и женщин (агентов разного цвета и конфигурации) в среде, характеризующейся распределением некоторого ресурса («запах денег»), который способствует созданию союза (рис. 1, слева).

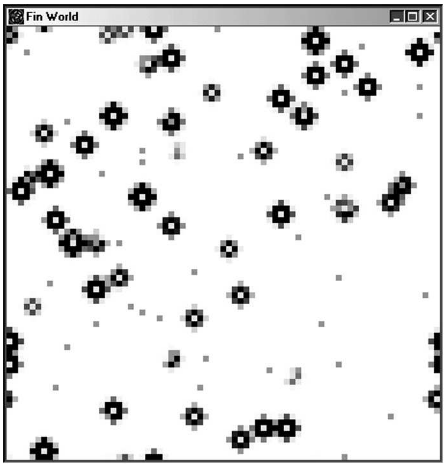
Каждый агент-мужчина является источником ресурса, меняющегося во времени. Случайным образом задается положение агентов-мужчин, их первоначальный запас ресурса и коэффициент естественного расхода наличного ресурса. С учетом величины капитала агентов-мужчин происходит распределение ресурса в окружающем их пространстве, подобно теплу, исходящему от живого мужчины. Начальная позиция агентов-женщин в среде задается также случайным образом. С течением времени координаты агентов-мужчин не изменяются, а агенты-женщины двигаются по определенному правилу.

Главным фактором поведения агентов-женщин является поле ресурса, образованное за счет капитала (дохода) агентов-мужчин. Агенты-женщины передвигаются, ориентируясь по вектору, указывающему на «запасы» ресурса и направляясь к агенту-мужчине с целью потребления ресурса. В результате моделирования мы имеем дело с вполне определенной картиной, демонстрирующей последствия локального взаимодействия агентов. Агент-женщина не покидает агента-мужчину, если запас его ресурса постоянно подпитывает «голод» агента-женщины. Так как первоначальное распределение ресурса ограничено, то агенты-женщины должны стремиться к самым богатым мужчинам.

С течением времени величина капитала изменяется под влиянием поведения агентов. Если в клетке находится агент-мужчина, то в ней происходит естественный рост и расход ресурса. Агент-женщина поглощает определенное количество ресурса в клетке, где она находится. Происходит дальнейшее распределение («расплывание») ресурса. Вследствие регулярного изменения поля ресурса, окружающего агентов-мужчин, агенты-женщины находятся в постоянном движении. Итогом моделирования является эволюция образования семей.

Экспериментально выявлены три типа поведения агентов и соответствующие типы союзов. Моногамный, парный и полигамный. Они зависят от потребностей женщин в ресурсах и от возможностей мужчин обеспечить их этими ресурсами.

Данная модель, безусловно, упрощает реальные отношения, конструируя их в жесткой бихевиористской схеме. Однако использование компьютерной имитации на конечной стадии эмпирического социологического исследования мо-



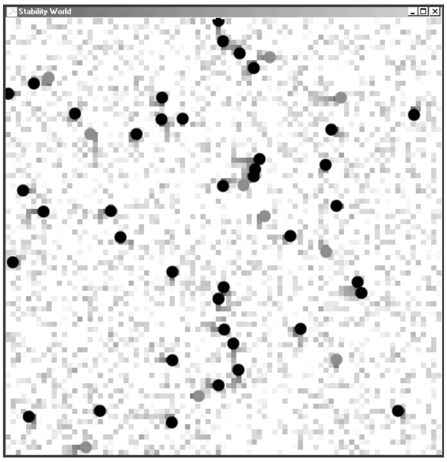


Рис. 1. Распределение агентов (светлые и черные точки), действующих в поле ресурса

жет открыть неизвестные значимые переменные (выявить варианты поведения агентов, условия сохранения или нарушения гендерной асимметрии, характер стабилизации/дестабилизации процессов и т.п.).

Другая предлагаемая модель имеет название *«Адаптивное поведение в условиях кризиса»* (мульти-агентная система $\mathcal{SWARM}$) [5]. При построении модели использовались данные исследования О.М. Здравомысловой и М.Ю. Арутюнян [6].

Гипотеза исследования состоит в том, что основной семейной стратегией в период экономического кризиса становится стратегия выживания. Первоначально в модели задается докризисная ситуация, характеризующаяся семейно-ориентированным образом жизни членов семьи, предполагающим принятие как мужчинами, так и женщинами традиционных ролей.

Семья в модели состоит из агентов-мужчин и агентов-женщин, имеющих определенную идентичность. Первоначально агенты располагаются случайным образом. В одних семьях агент-женщина имеет профессиональную ориентацию, в других женщины ориентированы на семейную жизнь. Вокруг агентов разбросаны источники с ресурсом дохода семьи. Через некоторое время происходит то увеличение, то уменьшение в источниках дохода (рис. 1, справа). Задача агентов состоит в поддержании равновесия между доходом и потреблением ресурсов в семье.

В период роста доходов большая часть агентов-женщин ориентированы на семью. Но приходит время, когда резкое сокращение ресурсов в источниках доходов ведет к наступлению экономического кризиса в семье. В этом случае агентам необходимо сориентироваться и выбрать наиболее приемлемую для их семьи стратегию адаптивного поведения. Это может быть:

1) поиск более денежной работы (агенты перемещаются ближе к наиболее богатым местам расположения ресурсов);
2) снижение притязаний (агенты вынуждены в связи с острой нехваткой доходов сократить свои расходы);
3) смена идентичности: агент-женщина с высокой степенью персональной

активности может изменить на время кризиса семейную ориентацию на профессиональную.

Если стратегия адаптивного поведения выбрана успешно, в семье устанавливается баланс роста и потребления ресурса.

Полученные результаты говорят не столько об удельном весе той или иной стратегии, сколько о тенденциях поведения семей в стремлении преодолеть кризисную ситуацию. Выявленное в исследовании различие стратегий показывает, что основным внутренним резервом, который пытается использовать семья, является возможность женщины сменить семейную ориентацию на профессиональную. Это дает дополнительный доход семье и способность преодолеть экономический кризис.

3. Заключение

Погружение в мир искусственных гендерных отношений — захватывающий эксперимент, развивающий воображение исследователя. Автор модели особым образом «разоблачает» гендерную систему в компьютерном пространстве. Он и играет стереотипами (вспомним модель «Влияние ресурсообеспеченности мужчин на поведение женщин»), и одновременно «разбивает» их в длинной серии компьютерных экспериментов. Как у любого моделирования, в компьютерном моделировании существует множество проблем разного характера. Однако основное обвинение в адрес симуляций такого рода касается упрощения и искажения реальности. Предлагая возможности компьютерного моделирования, исследователи не стремятся проанализировать, объяснить, предсказать социальную реальность в духе классической научности.

> Как пишет Д.В. Иванов, «наука сейчас — это не предприятие по поиску истины, а род языковых игр, состязаний в манипулировании моделями научного дискурса <...> Наука становится перманентным процессом построения альтернативных моделей. Вследствие этого возросла роль воображения, фантазии, парадоксальности мышления в той сфере, где ранее их предавали анафеме, где ранее референцией к реальности строго задавались пределы приращения знания» [1, с.19].

Воображение исследователя позволяет манипулировать реальными гендерными моделями, имитируя на экране гендерную утопию (или антиутопию). Этот процесс отчасти напоминает деконструкцию гендера в Интернете и чате, где можно «играть в гендер», выбирая гендерные роли и имитируя гендерный дисплей. Киберпространство чата — это определенная степень гендерной свободы. Да, гендерная реальность и виртуальность здесь не совпадают, однако многократное повторение в киберпространстве определенных гендерных правил, вероятно, способно изменить реальную гендерную ситуацию [7]. Создаваемый компьютерный мир гендерных моделей не имеет цели объяснить или, тем более, изменить реальность. Это скорее провокация социологического мышления, «интрига познания», играющая и задающая вопросы. Игра, которая, пожалуй, имеет право на существование в гендерной науке.

Литература

1. Иванов Д.В. *Виртуализация общества.* СПб., 2000.
2. Здравомыслова Е.А. , Темкина А.А. *Социология гендерных отношений и гендерный подход в социологии* // Социологические исследования. 2000. № 11.
3. Халеева И.И. *Гендер как интрига познания* // Гендерный фактор в языке и коммуникации. Сборник научных трудов. Вып. 446. М., 1999. С.7-14.
4. Мангейм Дж. Б., Рич Р.К. *Политология. Методы исследования.* М., 1997.
5. Гуц А.К., Коробицын В.В., Лаптев А.А., Паутова Л.А., Фролова Ю.В. *Математическое моделирование социальных систем.* Омск: ОмГУ, 2000.
6. Здравомыслова О.М. Арутюнян М.Ю. *Российская семья на европейском фоне.* М., 1998.
7. Кузнецов В. *Алиса в стране виртуальных чудес* // Женщина и визуальные знаки (Москва). 2000. С.182-183.

Математические структуры и моделирование
2001, вып. 8, с. 121–125

УДК 681.518

ОПЫТ ЭКСПЛУАТАЦИИ МОДУЛЯ «АВТОЗАЧИСЛЕНИЕ» ИНФОРМАЦИОННО-АНАЛИТИЧЕСКОЙ СИСТЕМЫ «АБИТУРИЕНТ» ОМГУ

Е.А. Костюшина

In the paper the experience using the algorithm of automation the process of enrolment entrants is described.

В статье [1] было приведено описание постановки задачи «Автозачисление» и алгоритма информационно-аналитической системы (ИАС) «Абитуриент», который использовался в ходе приемной кампании 2000 года в Омском государственном университете. Там же отмечалось, что приведенный алгоритм хотя и значительно ускорил процедуру зачисления, но не позволил полностью автоматизировать этот процесс. Учитывая опыт приемной кампании 2000 года, в 2001 году был использован комплексный подход к зачислению, в результате функция «Автозачисление» стала частью подсистемы «Зачисление» ИАС «Абитуриент». На сегодня подсистема «Зачисление» состоит из следующих модулей:

- модуль автоматического зачисления на бюджетные места, включая все виды наборов (общий, целевой, целевой областной администрации), результат – списки абитуриентов, рекомендованных к зачислению;
- модуль составления протокола приемной комиссии о внесении изменений в списки рекомендованных к зачислению в связи с исправлением технических ошибок, допущенных техническим секретариатом, и/или отказом абитуриента от бюджетного места, результат – списки зачисленных абитуриентов, протокол приемной комиссии;
- модуль зачисления абитуриентов на внеплановые места, результат – списки зачисленных на внеплановые места.

Подробно остановимся на модуле автоматического зачисления. В [1] была описана проблема «полупроходников»: если в ходе зачисления на специальность возникала группа абитуриентов с полупроходным баллом, работа алгоритма останавливалась и далее зачисление продолжалось уже вручную. Именно эта проблема и не позволила в свое время полностью автоматизировать процесс зачисления. Было выдвинуто предположение, что решение задачи по прогнозированию проходного балла разрешит проблему «полупроходников» и, как

E-mail: kostush@univer.omsk.su
Омский государственный университет

следствие, позволит полностью автоматизировать процесс зачисления. Однако опыт показал, что от задачи по прогнозированию проходного балла, как самостоятельной функции, можно отказаться, так как был предложен и реализован новый подход к задаче. Изменения, внесенные в алгоритм, позволили проводить зачисление абитуриентов не только на места из общего набора, как это было раньше, но и сразу же на все виды наборов, принятых на специальностях.

Общий подход к задаче «Автозачисление» остался прежним – весь процесс зачисления разбивается на два этапа, выполнение которых можно разделить во времени. На первом (подготовительном) этапе происходит отбор заявлений абитуриентов, которые допущены к участию в конкурсе, далее они разбиваются по **конкурсным позициям** *(позиция, на которую происходит зачисление, введение этого термина позволяет из иерархического списка специальностей, который зависит от формы обучения и вида приема, перейти к линейному списку)*. На следующем шаге внутри каждой позиции происходит выделение групп **равных абитуриентов** *(абитуриенты, которые имеют равные характеристики при зачислении (суммарный балл, балл по профильному предмету и т.п.), набор характеристик определяется приемной комиссией и может быть изменен)*, а затем линейное упорядочивание таких групп. Второй этап – собственно этап зачисления.

Для того чтобы появилась возможность осуществлять зачисление на все виды приема, в 2001 г. было принято решение приемной комиссии о введении **сквозного рейтинга позиций** *(сквозная нумерация конкурсных позиций, заявленных абитуриентом, в зависимости от его приоритетов)*. Во время приемной кампании-2001 понятие «группа абитуриентов с полупроходным баллом» заменено на «полупроходная группа». Во время зачисления сразу же учитывались дополнительные характеристики абитуриента, добавляющие ему вес при зачислении, поэтому группа абитуриентов с равным суммарным баллом разбивалась на подгруппы, зачисление происходило именно по этим подгруппам.

Кроме того, положительные результаты зачисления были достигнуты за счет изменения критерия окончания работы алгоритма, а также условия зачисления абитуриента. Все эти изменения отмечены на блок-схеме (см. рис.1).

Отметим еще одно нововведение – этап зачисления был разбит на итерации и добавлена возможность его выполнения в двух режимах: автоматическом и диалоговом. В диалоговом режиме можно остановить выполнение функции на любой итерации, а затем при необходимости продолжить с нее или начать работу алгоритма сначала.

В ходе эксплуатации оказалось, что разбиение на итерации удобно использовать при выполнении дополнительного набора на те конкурсные позиции, на которых прогнозировался отток абитуриентов, рекомендованных к зачислению, обычно это позиции с заочной формой обучения. На этом этапе возникла задача, тесно связанная с проблемой зачисления, – прогнозирование оттока с конкурсных позиций рекомендованных к зачислению абитуриентов.

Как уже отмечалось выше, процесс автозачисления был разбит на два этапа. Затраты машинного времени на подготовительный этап оказались довольно велики (в 3-4 раза больше, чем собственно этап зачисления), так как здесь вы-

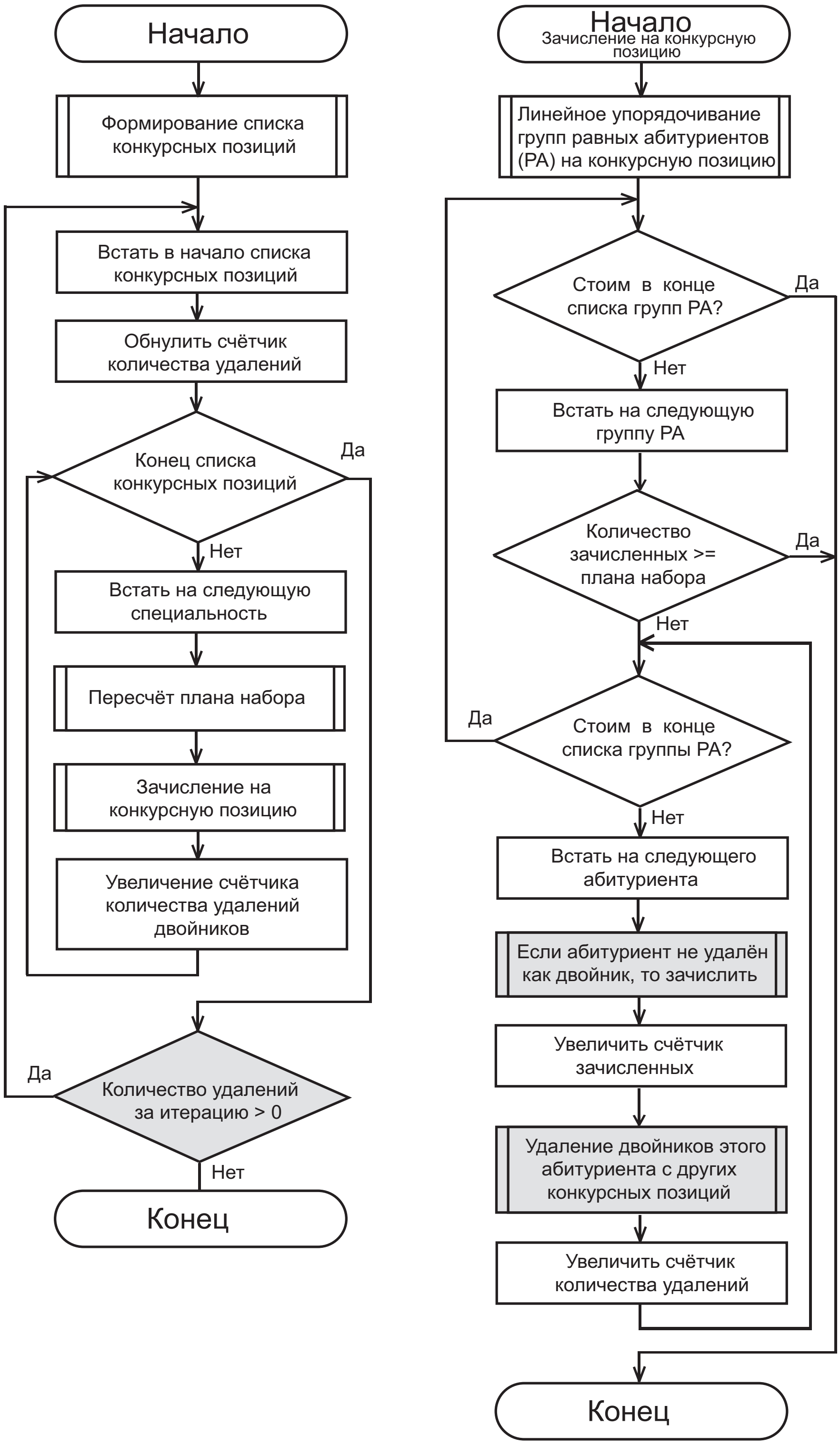


Рис. 1. Укрупненная блок-схема алгоритма «Автозачисление»

полняется операция сортировки, требующая $N \log_2 N$ операций, где $N = 18415$. Отметим, что этап зачисления на компьютере Celeron 330/128 требует всего 13 мин. 35 сек. машинного времени. В случае аварийного отключения компьютера во время выполнения второго этапа данные, полученные на подготовительном этапе, не будут потеряны. После восстановления устойчивой работы компьютера зачисление может быть продолжено со второго этапа. Кроме того, процесс автозачисления можно выполнять несколько раз с разными входными параметрами, при этом повторной реализации подготовительного этапа не требуется. Разделение по этапам позволяет значительно экономить время, что очень важно во время аварийных отключений, так как результаты зачисления должны быть получены в определенные временные рамки. Также могут быть довольно быстро получены результаты зачисления при различных ограничениях, анализ которых позволяет ускорить процесс приема решений приемной комиссией.

Сегодня абитуриенты все более активно используют возможность подачи заявлений на несколько специальностей не только внутри одного ВУЗа, но и в разные ВУЗы города. Приемные комиссии попадают в ситуации, когда абитуриента зачисляют на бюджетное место, а он забирает документы и уходит в другой ВУЗ. Это является довольно большой проблемой. Одним из вариантов ее решения является межвузовское зачисление, идея которого была предложена заместителем председателя приемной комиссии, первым проректором ОмГУ М.В.Хорошевским. Существует предположение, что алгоритм, применяемый в ходе приемной кампании 2001 г. в ОмГУ, позволяет проводить такое зачисление, однако здесь возникают некоторые проблемы, требующие особого внимания.

Обозначим эти проблемы:

1. Проблема единой идентификации абитуриентов. Каждый ВУЗ имеет свою собственную систему идентификации абитуриентов. Необходимо разработать алгоритм перехода от различных систем идентификации к единой системе, при выполнении которого происходил бы поиск абитуриентов, подавших заявления в несколько ВУЗов, а затем им присваивался бы уникальный код вне зависимости от ВУЗа.
2. Сквозной межвузовский рейтинг. Необходимо выработать административное решение о том, как абитуриент при подаче документов может указать свои приоритеты: или абитуриент нумерует ВУЗ, соответственно своим приоритетам, а внутри каждого из них уже выставляет приоритеты конкурсным позициям, или он сразу же выставляет сквозную нумерацию межвузовских конкурсных позиций. Решение данной проблемы опять позволит из иерархического списка ВУЗы-специальность перейти к линейному списку конкурсных межвузовских позиций.
3. Устойчивость работы алгоритма в условиях приемных кампаний других ВУЗов. В ВУЗах Омска приняты разные правила приема, например может существовать система преимуществ, отличная от ОмГУ, или вообще отсутствовать, соответственно может оказаться, что группы «равных абитуриентов» могут оказаться довольно большими, и как результат – зачисление на специальности пройдет с превышением плана приема. Здесь необходимо провести исследование, чтобы выяснить, какие факторы мо-

гут оказывать влияние на результаты зачисления.

Решение этих проблем позволит разработать единые требования, которые должны предъявляться к спискам абитуриентов, выдержавших испытания. Такие списки будут составляться ВУЗами самостоятельно, в зависимости от принятой у них системы проведения приемной кампании. Но здесь же возникает задача об автоматизации переноса списков, представляемых ВУЗами, в ИАС «Абитуриент» ОмГУ. Опыт эксплуатации показал, что изменения, внесенные в алгоритм, позволили полностью автоматизировать процесс зачисления абитуриентов. Отметим, что выполнение задачи на любом промежуточном этапе приемной кампании позволяет получить проходной балл на специальности по результатам сдачи вступительных испытаний абитуриентов к этому моменту. Кроме того, были описаны некоторые дополнительные задачи, связанные с зачислением. Также обозначены некоторые перспективы дальнейшего применения алгоритма функции «Автозачисление».

Литература

1. Костюшина Е.А. *Автоматизированная подсистема «Абитуриент». Формализация алгоритма функции «Автозачисление»* // Математические структуры и моделирование. 2001. Вып.7. С. 157-161

Математические структуры и моделирование
2001, вып. 8, с. 126–131

УДК 681.518

О КОНЦЕПЦИИ ИНДЕКСАЦИИ ИНФОРМАЦИОННЫХ РЕСУРСОВ СЕТИ ИНТЕРНЕТ

И.А. Земсков

In this article two methods building a full-text index for the Web are described.

Введение

Настоящая статья посвящена обсуждению путей повышения эффективности поисковых средств сети Интернет. В ней делается попытка обобщить идеи и соображения, возникшие у автора на основе личного опыта работы с поисковыми средствами сети Интернет, а также на основе многочисленных публикаций на эту тему.

В качестве объекта исследования выбрана система индексирования информационных ресурсов в Интернет или т.н. индексирующий робот. Автор сознательно ограничивает область рассмотрения, вычленяя из всех проблем, связанных с поиском информации в Интернет, проблему индексации, так как, на его взгляд, эта проблема была незаслуженно лишена внимания исследователей.

Целью настоящей статьи является анализ функционирования индексирующих роботов поисковых систем. Это позволит вплотную подойти к вопросу создания математических моделей систем индексирования информационных ресурсов и прогноза дальнейших условий их функционирования. Также целью статьи является описание возможных проблем при технической реализации предлагаемых нововведений.

1. Описание процесса индексации

Концепция работы индексирующих роботов, а их насчитывается уже порядка 270 [1], не менялась со времён изобретения первых поисковых систем Интернет. До недавнего времени подобный факт можно было оставлять без внимания, но бурный рост сети Интернет вносит свои коррективы в расстановку приоритетов решаемых задач. Согласно [2] число серверов World Wide Web, а значит, и документов в этой распределенной информационной системе Интернет удваивается каждые 60 дней. Однако происходит не только удвоение количества ресурсов, но и изменение уже существующих. А это означает, что такие системы, как

E-mail: zemskov@univer.omsk.su
Омский государственный университет

Altavista и Lycos, обязаны не только обновлять свои поисковые индексы непрерывно, но и должны постоянно увеличивать зону охвата индексируемых ресурсов. В настоящее время практически все крупные поисковые системы этого типа как в России, так и за границей имеют своих индексирующих роботов, которые работают постоянно и всё же не поспевают за стремительным развитием обрабатываемой среды. Следует отметить, что время «повторной» индексации ресурса с каждым разом увеличивается. Это, в свою очередь, может приводить к «утере» адекватного отображения реального состояния ресурсов.

В современной концепции процесса индексации мы можем выделить два типа узлов сети Интернет, принимающих непосредственное участие в работе:

узел-донор — на нём расположена коллекция веб-страниц, предназначенная для индексации.

узел-обработчик — на нём происходит три процесса: первоначальная обработка скачанных с узла-донора веб-страниц; процесс построения индекса и его сохранения; выполнение поисковых запросов на основе полученного индекса.

Отбрасывая процесс выполнения поискового запроса, можно сказать, что процесс индексации протекает в две стадии: во-первых, узел-обработчик «скачивает» с узла-донора его коллекцию веб-страниц, а во-вторых, на узле-обработчике происходит обработка полученной коллекции для получения т.н. поискового массива (индекса).

Основное внимание исследователей направлено на изучение и оптимизацию процессов, происходящих на узле-обработчике. Например, в [3] рассматриваются четыре наиболее популярные меры близости, используемые в информационно-поисковых системах Интернет для ранжирования найденных документов. Большая работа по оптимизации протекающих процессов была проделана в работе [4]. Её авторы предложили разделить узел-разработчик на два узла: узел-индексатор и узел-поисковик.

Однако, независимо от разработчиков, неизменными остаются два момента в процессе индексации: во-первых, для того чтобы начать строить индекс, нужно «скачать» коллекцию веб-страниц на узел-обработчик; во-вторых, начальным этапом обработки становится избавление от т.н. «мусора» в виде HTML-кода (что считать «мусором», зависит от алгоритма построения индекса, но как минимум это HTML-код).

Акцентирование внимания на этих моментах приводит к формулированию целого ряда важных вопросов:

1. Какой объём «мусора» скачивает к себе робот?
2. Сколько времени и других собственных ресурсов тратит узел-обработчик на обработку «мусора»?
3. Как посчитать то количество системных ресурсов (например: вычислительная мощность процессоров, ОЗУ, дисковое пространство, ширина канала связи с Интернет), которое потребуется узлу-обработчику для поддержания в актуальном состоянии своего поискового массива, построенного по коллекциям веб-страниц максимально возможного числа узлов-доноров сети Интернет?

Из поставленных вопросов становится очевидным тот факт, что узел-обработчик становится источником бесполезной нагрузки на сеть Интернет, т.к. некоторая часть (её размер ещё предстоит выяснить) его работы является обработкой «мусора», который за ненадобностью впоследствии выбрасывается.

2. Новая концепция индексации

Совсем избавиться от затрат на обработку «мусора» мы не в состоянии, т.к. он служит нам той «оболочкой» для информации, которая делает Интернет Интернетом. Но нам вполне по силам взаимовыгодно разделить эти затраты между всеми членами Интернет-сообщества (имеются в виду только владельцы узлов-доноров и узлов-обработчиков). Или, другими словами, с учётом предыдущих рассуждений становится очевидным желание перераспределить функции, выполняемые в процессе индексации, между узлами-донорами и узлом-обработчиком.

Опишем узлы, участвующие в процессе, с учётом новых функций:

узел-донор — на нём, помимо коллекции веб-страниц, расположен модуль-робот, отвечающий за выполнение предварительной обработки веб-страниц и последующей передачи результатов обработки на узел-обработчик для построения поискового массива;

узел-обработчик — на нём происходит три процесса: приём обработанных на узле-доноре веб-страниц, фактически приём «чистой» информации, лишённой «мусора»; процесс построения индекса и его сохранения; выполнение поисковых запросов на основе полученного индекса.

Рассмотрим подробнее процессы, происходящие на узле-доноре:

1. Процесс инициализации — происходит в начальный момент функционирования самой коллекции веб-страниц.
2. Обнаружение «цели» для обработки — под целью понимается веб-страница. Может быть два вида целей: веб-страница ранее не была обработана для последующей передачи узлу-обработчику и веб-страница претерпела изменение содержимого, но не изменила своих «координат». Для осуществления этой функции может потребоваться организация локального индекса обработанных веб-страниц. Стоит заметить, что в момент процесса инициализации локальный индекс пуст. Техническая реализация процесса обнаружения «цели» может быть основана на двух моделях поведения: пассивной или активной. Пассивная модель заключается в простом «прослушивании» ответов на внешние запросы к коллекции веб-страниц и нахождении т.н. «цели» для обработки. В защиту этой модели поведения можно сказать то, что таким образом мы проиндексируем всю действительно «интересную» информацию, располагающуюся на узле (так как, если к ней обратились, значит она интересна). Активная модель поведения базируется на алгоритмах поведения существующих индексирующих роботов. Данная модель, по мнению автора, имеет больше минусов, чем плюсов. Например, из основных минусов можно выделить следующий: она создаёт неоправданно большую дополнительную вычислительную нагрузку на системные ресурсы узла-донора благодаря своим сложным эвристическим

алгоритмам нахождения «цели» обработки в коллекции веб-страниц, которые потребуется постоянно обновлять в связи с развитием технологий, поддерживаемых при создания веб-страниц.

3. Обработка найденной «цели» – обработка может заключатся как в простом избавлении веб-страницы от определённого набора HTML-тегов, так и в более детальной проработке, в зависимости от последующих потребностей индексирующего узла-обработчика.
4. Передача на узел-обработчик локального индекса, всего или только некоторой части. Другими словами, на индексирующий узел-обработчик должна передаваться не вся веб-страница, а только её информационное наполнение, представленное в нужном для индексации виде, плюс служебная информация (например, помимо информационного наполнения страницы, индексирующему процессу нужны «координаты» веб-страницы).

Дополнительное пояснение требуется термину «локальный индекс»: роль локального индекса может выполнять как минимум локальная (для узла-донора) коллекция обработанных веб-страниц вместе с сопоставленными им координатами реальных веб-страниц. Под координатами веб-страниц может пониматься URL адрес веб-страницы. Существование локального индекса видится принципиально важным, так как, во-первых, с помощью него удаётся принять решение о нахождении новой веб-страницы в коллекции узла-донора, избегая при этом обращения к узлу-обработчику за дополнительной информацией; во-вторых, становится возможным (практически мгновенно) отследить изменение в уже проиндексированных страницах и сообщить о характере изменений на узел-обработчик. Тем самым становится возможным изменить стандартное значение термина «повторная индексация», т.к. теперь под ним будет пониматься не повторная обработка всей коллекции веб-страниц, а выборочная обработка новых или изменившихся веб-страниц коллекции узла-донора.

Согласие с целесообразностью организации локального индекса на основе этих предположений влечёт за собой возникновение вопроса о его размере, т.е. вопроса о том количестве ресурсов узла-донора, которое нужно выделить на поддержку локального индекса. И хотя количество и тип ресурсов ещё предстоит изучить, но уже сейчас можно предположить одно из возможных направлений минимизации дискового пространства, занимаемого индексом. Например, на узле-доноре достаточно хранить не полную версию веб-страницы после её обработки, а только некий код, полученный на основе избавленного от «мусора» содержимого веб-страницы, однозначно идентифицирующий её содержимое. Вполне достаточным может оказаться использование кодирования «в одну сторону» и последующего сравнения получившегося кода с уже имеющимся для этой веб-страницы. Однако нужно иметь в виду, что кодирование должно заметным образом уменьшать размер обрабатываемой информации, т.к. иначе его использование лишено смысла.

Рассмотрение вопроса организации локального индекса, с точки зрения программной реализации, наталкивается на ряд вопросов по обеспечению безопасности (или, другими словами, ограничения доступа к закрытой информации) конфиденциальной информации, возможно находящейся в коллекции веб-страниц.

Немного слов скажем о передаче локального индекса узла-донора на узел-обработчик. В новой ситуации, вызванной перемещением функций первоначальной обработки веб-страниц на узел-донор, узел-обработчик сохранил только две ресурсоёмкие функции — это построение поискового массива (индекса) и выполнение поисковых запросов. Но взамен утерянной функции первоначальной обработки он приобрёл функцию приёма локальных индексов узлов-доноров. Сама по себе функция не является источником большой вычислительной нагрузки, но в условиях большого количества узлов-доноров она может стать серьёзной проблемой организации стабильной работы узла-обработчика. Можно предложить несколько подходов к решению обозначенной проблемы. Первый подход заключается в том, чтобы описать механизм «отложенной передачи локального индекса», который по своей сути является ожиданием освобождения ресурсов узла-обработчика в случае их «занятости». Второй подход реализуется выделением необходимого количества системных ресурсов на основе прогноза возможной максимальной загруженности узла-обработчика. Оба подхода можно изучить с помощью соответствующей имитационной модели очереди с отказом в обслуживании.

С точки зрения технической реализации процесса взаимодействия узла-донора и узла-обработчика, будет интересным получить ответы на следующие вопросы:

1. Какой из двух узлов должен стать инициатором взаимодействия? Узел-донор до сих пор был пассивным участником, но теперь имеет смысл рассмотреть возможность его активизации, т.к. только ему известен момент актуального обновления локального индекса.
2. Возможно ли обойтись в реализации процесса взаимодействия только средствами протокола HTTP или HTTPS? Ответ на этот вопрос во многом будет зависеть от ответа на предыдущий вопрос.
3. Какие меры нужно предпринять, чтобы исключить возможность проведения любых несанкционированных действий по отношению к обоим узлам? В вопросе кроется большая проблема защиты рассматриваемой системы от хакерских атак. Решение этой проблемы во многом будет влиять на получение путёвки в жизнь предлагаемой технологии индексации.

Заключение

Многие направления будущих исследований формулировались по ходу статьи, но первоочередную цель сформулируем ещё раз:

> чтобы подтвердить или опровергнуть выгоду от реализации выдвинутой гипотезы об экономической целесообразности перенесения части индексирующего робота на узел-донор, следует создать имитационную модель взаимодействия узлов-доноров и узла-обработчика по старой концепции индексации, а также имитационную модель взаимодействия узлов-доноров и узла-обработчика по новой концепции индексации, описанной в этой статье.

Литература

1. The Web Robots Pages. – `http://www.robotstxt.org/wc/active/html/index.html`
2. Попов И.И., Храмцов П.Б. *Распределение частоты встречаемости терминов для линейной модели информационного потока* // НТИ. 1991. Сер.2. №2. С.23-26.
3. Budi Yuwono, Dik L.Lee. *Search and Ranking Algorithms for Locating Resources on the World Wide Web* // In Proceedings of the Forth International Conference on the World Wide Web. New York. November. 1995.
4. Sergey Melnik, Sriram Raghavan, Beverly Yang, and Hector Garcia-Molina *Building a Distributed Full-Text Index for the Web.* – `http://www-diglib.stanford.edu/cgi-bin/get/SIDL-WP-2000-0140`
5. Храмцов П.Б. *Моделирование и анализ работы информационно - поисковых систем Internet* // Открытые Системы. 1996. №6.

Математические структуры и моделирование
2001, вып. 8, с. 132–140

УДК 519.71

КРИТЕРИЙ БЛИЗОСТИ ДОКУМЕНТОВ И КЛАСТЕРИЗАЦИЯ

О.Г. Чанышев

In this article the algorithm for automatic clustering is presented. This algorithm is based on original model of the real text. Automatically extracted «dominant lexems» are using for automatic clustering of non-grupped beforehand sets of documents.

Введение

Общей проблемой области автоматического анализа естественноязыковых текстов (ЕЯ-текстов), к которой относятся задача автоиндексирования и тесно связанная с ней [1, стр. 341] задача автоматической тематической классификации, является проблема понимания текста системой искусственного интеллекта. Современные промышленные автоиндексирующие и автоклассифицирующие системы [2–4] обладают высоким быстродействием, эргономными пользовательскими интерфейсами. Например, в продукт LinguistX компании Inxight Software входят усовершенствованные средства обработки естественного языка от поисковых механизмов до средств распознавания рукописного текста, включая автоматическое реферирование, извлечение информации и морфологический анализ [5]. С целью развития классических методов кластеризации достаточно широко используются искусственные нейронные сети [6]. Однако прогресс в этой области имеет скорее технологический характер, что А.С. Нариньяни констатировал (несколько эмоционально) следующим образом: «Массированное, продолжавшееся несколько десятков лет наступление в области автоматической обработки текста захлебнулось. По отношению исходных планов и надежд оно окончилось достаточно очевидным провалом.» [7]. Так или иначе, принципиальный вопрос о приемлемой теории текста, на которой должны базироваться методы автоматического анализа текста, остается открытым.

Алгоритмы классификации относятся к обширному классу алгоритмов распознавания образов [8, Глава 9], [9, Глава 4], в основе которых лежит гипотеза компактности: «реализации одного и того же образа обычно отражаются в признаковом пространстве в геометрически близкие точки, образуя «компактные» сгустки» [10, стр. 29]. Методы классификации текстов едва ли не исчерпывающе представлены в монографии Дж. Солтона (Gerard Salton) [1, Глава 8]

E-mail: chanysh@iitam.omsk.net.ru
Омский филиал Института математики СО РАН

вплоть до физической идеи определения кластеров «путем коллапсирования пространства с помощью гравитационного притяжения».

Представляемый в настоящей статье метод автоматической кластеризации ЕЯ-текстов основан на «ассоциативной модели реального текста» [12, 13]. С операционной точки зрения, он относится к «порождающим методам классификации по принципу снизу вверх, при котором все объекты первоначально считаются несгруппированными» [1, стр. 242]. Основное отличие от геометрического подхода заключается в принципиальной несимметричности используемой меры тематической близости, поскольку требование «равной похожести» части и целого семантически не представляется естественным.

1. Доминанты, тематическая близость и кластеризация документов

Ассоциативная модель рассматривает текст как задание тотального графа предметной области списками смежностей лексем – предложениями. Текст не нормализуется, и не рассматриваются лексемы, принадлежащие заданному стоп-множеству. Из оставшихся учитываются только «независимые» лексемы, для любой пары которых существуют минимум два предложения, в которых они встречаются отдельно. В качестве меры важности лексемы используется «ассоциативная мощность» (Ψ), совпадающая с частотой (ω) только в случае задания графа бинарными списками смежности. При этом, если для низко- и среднечастотных лексем можно положить $\Psi \approx Const \times \omega$, то для высокочастотных эта зависимость существенно не монотонна (из $\omega_j > \omega_i$ не следует $\Psi_j \geq \Psi_i$). Анализ ранговых распределений Ципфа-Мандельбродта для независимых лексем позволил ввести понятие критического значения ассоциативной мощности для выделения наиболее важных (доминантных) лексем, объем которых не превышает 0.04 от объема словаря текста.

Хорошо известно, что из двух «подзадач» задачи распознавания образов: выбора множества признаков объекта и распознавания на основе выбранного множества – наиболее трудно формализуемой и в этом смысле наиболее сложной является первая [11]. Эксперименты по автоматическому реферированию текстов показали, что независимые лексемы связи адекватно представляют текст, но их слишком много для попарного сравнения каждого документа с каждым. А это предусматривает алгоритм кластеризации в случае, когда не используется никакая другая вспомогательная информация. Поэтому в качестве признакового множества решено было использовать доминанты документов.

2. Тематическая близость документов и кластеризация

Пусть

$D_N = (d_1, d_2, \ldots, d_N)$ – произвольное множество документов (при этом через d_i будем обозначать как сами документы, так и их идентификаторы),

$L_i^d = (l_1^i, l_2^i, \ldots, l_j^i, \ldots, l_{n_i}^d)$ – множество доминантных лексем (доминант) i-го документа,

$\Psi_i^d = (\psi_1^i, \psi_2^i, \ldots, \psi_j^i, \ldots, \psi_{n_i}^d)$ – множество ассоциативных мощностей доминант;

$i = (1, 2, \ldots, N),\ j = (1, 2, \ldots, n_i)$.

Каждое из Ψ_i^d и L_i^d частично упорядочено по убыванию ψ так, что из $j_1 < j_2$ следует $\psi_{j_1}^k >= \psi_{j_2}^k$.

Для учета роли одинаковых доминант в различных документах перейдем от ψ_j к рангу r_j – номеру группы с одинаковыми значениями ψ_j. И в качестве «веса» (w_j^i) доминанты возьмем значение, обратное рангу:

$$w_j^i = \frac{1}{r_j^i}. \tag{1}$$

Таким образом, каждый $d_i \in D_N$ представляется векторами

$$L_i^d,\ W_i^d = (w_1^i, w_2^i, \ldots, w_j^i, \ldots, w_{n_i}^i).$$

Если $l_m^k = l_n^p$, то $w_m^k = w_n^p$ только при m=n.

Пусть $R^{k,m} = L^k \cap L^m \neq \emptyset$.

В литературе (например [14]) неоднократно отмечалось, что для решения вопроса, следует ли то или иное слово рассматривать в качестве поискового термина, необходимо учитывать контекст (cluster), в котором данное слово появляется и который, в свою очередь, может быть представлен множеством других слов.

Несмотря на «доминантность» лексем множества $R^{k,m}$, пересечение по одному слову не гарантирует тематическую близость документов. Однако именно в силу того, что из полного словаря текста отобраны доминантные лексемы, наиболее точно представляющие тему, требование $N_R > 1$, эквивалентное требованию учета контекста, может оказаться достаточным для целей определения тематической близости и автоматической кластеризации текстов.

Тогда близости документов $b_{k,m}$ и $b_{m,k}$ определяются следующим образом:

$$b_{k,m} = \sum_i^{N_R} w_i^k,\ b_{m,k} = \sum_i^{N_R} w_i^m,\ N_R > 1. \tag{2}$$

Пусть

$$B^k = (b_{k,m_1}, \ldots, b_{k,m_i} \ldots),\ m_i \neq k, \tag{3}$$

– списки близости документов $d_k \in D_N$ к другим документам множества D_N, частично упорядоченные по убыванию значений b_{k,m_i}.

Отбросив в (3) все b_{k,m_i} меньшие первых максимальных и заменив b_{k,m_i} на соответствующие идентификаторы d_{k,m_i} для каждого k, получим списки документов, максимально близких k-ым, или списки «центроидов»:

$$L^{k,max} = (d_{k,m_1}, d_{k,m_2}, \ldots, d_{k,m_i} \ldots), \tag{4}$$

$$k, m_i \in (1, 2, \ldots, N),$$

причем d_{k,m_i} в (4) обозначает идентификатор m_i - го документа, максимально близкого к k-му.

Построим начальный структурированный список кластеров, каждый элемент которого состоит из: m-го центроида, списка L_m k-ых документов с максимальными $b_{k,m}$ и числа элементов списка (списка элементов m-го кластера – C_m):

$$K = ((d_1^c, C_1, L_1), (d_2^c, C_2, L_2), \ldots (d_i^c, C_i, L_i), \ldots) \tag{5}$$

$$L_i = (d_{i,1}, d_{i,2}, \ldots).$$

Элементы списка K частично упорядочены по убыванию значения C_i.

Для получения итогового списка кластеров необходимо объединить все L_i и L_j, $i < j$, такие, что $d_j^c \in L_i$.

Таким образом, список K есть итоговое разбиения множества D_N на подмножества $(d^c \cup L_i)$, если $\forall (i < j), d_j^c \notin L_i$.

3. Эксперимент

Ниже приводятся результаты экспериментов по проверке сепарационных возможностей метода. Влияние других факторов (например размера документов) не исследовалось. Все документы взяты из Internet'а и проиндексированы системой ГИОС. Документы, не имеющие связей с другими, помещаются в кластер «разное».

В качестве контрольных, однозначно принадлежащих фиксированной предметной области, выбраны два курса лекций по СУБД. Ниже приводится содержание частей для того, чтобы дать представление читателю об их тематике.

Курс I. С.Д. Кузнецов. Введение в СУБД, 9 частей.

(Рубрикация дана точно по электронной публикации)

Часть 1. СИСТЕМЫ УПРАВЛЕНИЯ БАЗАМИ ДАННЫХ #01/95.

1. Численные и информационные прикладные системы.

2. Файловые системы.

3. Области применения файлов.

Часть 2. СИСТЕМЫ УПРАВЛЕНИЯ БАЗАМИ ДАННЫХ #02/95

4. Потребности информационных систем.

5. Что есть СУБД в целом – функции и структура.

6. Да, были средства (управления базами данных) в наше время...

Часть 3.СИСТЕМЫ УПРАВЛЕНИЯ БАЗАМИ ДАННЫХ #03/95.

Глава 4. Реляционный подход к организации баз данных, или Теория и Интуиция.

Глава 5. Базисные средства манипулирования реляционными данными, или на чем базируются языки запросов.

Часть 4. СИСТЕМЫ УПРАВЛЕНИЯ БАЗАМИ ДАННЫХ #04/95.

Глава 6. Проектирование реляционных БД на основе принципов нормализации и семантическое моделирование баз данных.

Часть 5. СИСТЕМЫ УПРАВЛЕНИЯ БАЗАМИ ДАННЫХ #01/96.

Глава 6. System R: более чем удачный эксперимент.

Часть 6. СИСТЕМЫ УПРАВЛЕНИЯ БАЗАМИ ДАННЫХ #02/96.

Глава 7. Ingres: откуда пошли открытые СУБД.

Глава 8. Базы данных: и куда же все складывается?

Часть 7. СИСТЕМЫ УПРАВЛЕНИЯ БАЗАМИ ДАННЫХ #03/96.

Глава 9. Может ли толпа людей пройти через узкую дверь и не слишком наломать бока, или Управление транзакциями в системах баз данных.

Глава 10. Надежно можно жить только имея запасы, или Журнализация изменений БД.

Часть 8. СИСТЕМЫ УПРАВЛЕНИЯ БАЗАМИ ДАННЫХ #04/96.

Глава 11. В любое царство вводят толмачи.

Глава 12. Традиционные социальные методы в компьютерных технологиях, или СУБД в архитектуре клиент-сервер.

Глава 13. Мы не одни в этом мире, или Распределенные базы данных.

Часть 9. СИСТЕМЫ УПРАВЛЕНИЯ БАЗАМИ ДАННЫХ #05-06/96.

Глава 14. Что день грядущий нам готовит?

Глава 15. Каждому субъекту свой объект.

Глава 16. Рулить - это от слова «действовать по правилам».

Курс II. Ладыженский. СУБД - коротко о главном.

Часть 1. СИСТЕМЫ УПРАВЛЕНИЯ БАЗАМИ ДАННЫХ #01/95.

Введение.

Раздел 1. Реляционная база данных - основные понятия.

Часть 2. СИСТЕМЫ УПРАВЛЕНИЯ БАЗАМИ ДАННЫХ #02/95.

Раздел 2. Сервер базы данных.

Часть 3. СИСТЕМЫ УПРАВЛЕНИЯ БАЗАМИ ДАННЫХ #03/95.

Раздел 3. Обработка распределенных данных.

Часть 4. СИСТЕМЫ УПРАВЛЕНИЯ БАЗАМИ ДАННЫХ #04/95.

Раздел 4. Обработка транзакций.

Раздел 5. Средства защиты данных в СУБД.

Заключение.

Литература.

Прежде всего было установлено, что, объединенные в два текста (части каждого курса собраны в соответствующий текст), они составляют один кластер.

Эти тексты, разбитые на части (всего 13), помещались в документальные среды различной тематики. Результаты кластеризации приведены ниже.

Число документов и тематическая характеристика относятся к документам среды.

3.1. Два вышеуказанных курса лекций по СУБД по частям (всего 13 текстов)

Результат. Полное разделение на два кластера, каждый из которых содержит части соответствующего курса.

Субъективная оценка «отлично».

3.2. Отдельные статьи по СУБД (всего10)

Результат. В данном случае имеет смысл полностью привести составы кластеров.

Кластер 1.

Джим Грей. Управление данными: прошлое, настоящее и будущее. СИСТЕМЫ УПРАВЛЕНИЯ БАЗАМИ ДАННЫХ #03/98.

БД - достижения и перспективы на пороге XXI столетия. Под ред. Ави Зильбершатца, Майка Стоунбрейкера и Джеффа Ульмана. СИСТЕМЫ УПРАВЛЕНИЯ БАЗАМИ ДАННЫХ #03/96.

С.Д. Кузнецов. Введение в СУБД. Часть2.

С. Д. Кузнецов. Введение в СУБД. Часть 1.

А.З. Ишмухаметов, В.В. Лукин. Организация словаря данных в предметно-ориентированных программных оболочках. СИСТЕМЫ УПРАВЛЕНИЯ БАЗАМИ ДАННЫХ #01-02/98.

Э. Ларсен, Дж. Олкин, М.Портер. Oracle Media Server: предоставление потребителям интерактивного доступа к данным мультимедиа. СИСТЕМЫ УПРАВЛЕНИЯ БАЗАМИ ДАННЫХ #01/95

Кластер 2.

К. В. Ахтырченко, В. В. Леонтьев. Распределенные объектные технологии в информационных системах. СИСТЕМЫ УПРАВЛЕНИЯ БАЗАМИ ДАННЫХ #05-06/97.

К. В. Ахтырченко. Применение технологии Corba при построении распределенных информационных систем. СИСТЕМЫ УПРАВЛЕНИЯ БАЗАМИ ДАННЫХ #01-02/98.

Кластер 3.

С.Д. Кузнецов. Введение в СУБД. Часть 5.

С.Д. Кузнецов. Введение в СУБД. Часть 6.

С.Д. Кузнецов. Введение в СУБД. Часть 7.

С.Д. Кузнецов. Введение в СУБД. Часть 8.

С.Д. Кузнецов. Введение в СУБД. Часть 3.

С.Д. Кузнецов. Введение в СУБД. Часть 9.

Джон М. Смит, Диана К. Смит. Абстракции баз данных: агрегация и обобщение. СИСТЕМЫ УПРАВЛЕНИЯ БАЗАМИ ДАННЫХ #02/96.

Петер Пин-Шен Чен. Модель «сущность-связь» - шаг к единому представлению о данных. СИСТЕМЫ УПРАВЛЕНИЯ БАЗАМИ ДАННЫХ #03/95

С.Д. Кузнецов. Введение в СУБД. Часть 4.

Кластер 4.

Ладыженский. СУБД - коротко о главном. Часть 3.

Ладыженский. СУБД - коротко о главном. Часть 2.

Ладыженский. СУБД- коротко о главном. Часть 4.

Б.А.Позин. Современные средства программной инженерии для создания открытых прикладных ИС. СИСТЕМЫ УПРАВЛЕНИЯ БАЗАМИ ДАННЫХ #01/95

С.Д. Кузнецов. Введение в информационные системы. СИСТЕМЫ УПРАВЛЕНИЯ БАЗАМИ ДАННЫХ #02/97

Ладыженский. СУБД - коротко о главном. Часть 1.

Комментарий. Отлично. Первый кластер, по сути, введение в тему. Содержание 2-го говорит само за себя. Третий кластер – ядро, основное содержание темы СУБД. 4-й кластер, как и лекции Ладыженского в целом, имеет выраженную «ИС-доминанту».

3.3. Психология (всего текстов 25)

Результат: полное разделение на 3 кластера – психология и два кластера лекций.

Комментарий. Отлично.

3.4. Публицистика различной тематики (всего текстов 30)

Результат. 7 кластеров и «разное». Два отдельных кластера лекций.

Комментарий. Отлично.

3.5. Три различных курса лекций по философии (всего текстов 45)

Е.К. Дулуман. Философия (7). Лекции по истории натурфилософии (28). А.Н. Суворова. Введение в современную философию (10).

Результат. 6 кластеров. Два отдельных кластера лекций по СУБД.

Комментарий. Отлично.

3.6. Семь рассказов и повесть А.П.Чехова «Дама с собачкой»

Результат. 3 кластера плюс «разное». Лекции Ладыженского по-прежнему составляют отдельный кластер, *а в кластер лекций Кузнецова попадает «Дама с собачкой».*

Комментарий. Тройка. Дальнейший анализ показывает, что повесть Чехова попадает последней в список близости девятой части лекций Кузнецова из-за пересечения по доминантам «время» и «памяти».

3.7. Психология плюс «Дама с собачкой» (всего текстов 26)

Результат. Три кластера. Повесть Чехова попадает в кластер «психология». Лекции составляют два отдельных кластера.

Комментарий. Отлично. В данном случае «Дама...» попадает по месту.

3.8. Публицистика различной тематики (всего текстов 30)

Результат. Семь кластеров плюс «разное». Контрольные тексты составляют два отдельных кластера.

Комментарий. Отлично.

4. Обсуждение результатов и выводы

Представленный метод кластеризации демонстрирует высокое качество тематической сепарации текстов, что, в свою очередь, говорит о перспективности подхода, положенного в основание ассоциативной модели. По-видимому, описанный алгоритм кластеризации можно эффективно использовать для снижения доли нерелевантных документов при поиске по образцу, а также для построения субклассов после первичной классификации документов на основе заданных тезаурусов предметных областей. В основу тезаурусов могут быть положены доминантные лексемы.

Тем не менее, как показывает случай с повестью А.П. Чехова, алгоритм не гарантирует 100% тематической однородности кластеров.

Литература

1. Солтон Дж. *Динамические библиотечно-информационные системы.* М.: Мир, 1979.
2. Кузина И. *Новое поколение поисковых машин.*
   – http://koi.www.osp.ru/cw/1997/32/opensys/01.html
3. Керстеттер Д. *Новая лингвистическая технология повышает точность поиска* // Компьютерная неделя. N47 (121) от 2/12/1997
4. Крейнес М.Г. *Смысловой поиск и индексирование текстовой информации в электронных библиотеках: информационная технология «ключи от текста»* // Электронные библиотеки. 1999. Т 2, Выпуск 3.
5. Эссик К. *Документ - это еще не информация* // Computerworld Россия. 1998 № 25.
6. Петухов Д.А., Heuser U., Babanine A., Rosenstie W. *Применение нейронных сетей для кластеризации документов.*
   –http://oasis.peterlink.ru/ dap/nneng/nn−article.html
7. Нариньяни А.С. *Автоматическое понимание текста – новая перспектива.* Сайт РосНИИ искусственного интеллекта: http://www.rriai.org.ru
8. Кузин Л.Т. *Основы кибернетики.* М.: Энергия, 1979.
9. *Искусственный интеллект.* - В 3-х кн. Кн.2. Модели и методы: Справочник / Под ред. Д.А. Поспелова. М.: Радио и связь, 1990.
10. Загоруйко Н.Г. *Прикладные методы анализа данных и знаний.* Новосибирск: Изд-во Ин-та математики, 1999.
11. Бонгард М.М. *Проблема узнавания.* М.: Физматгиз, 1967.
12. Чанышев О.Г. *Ассоциативная модель естественноязыкового текста* // Вестник Омского университета. 1977. Вып. 4. С.17-20.

13. Чанышев О.Г. *Ассоциативная модель реального текста и ее применение в процессах автоиндексирования* // Труды Седьмой национальной конференции по искусственному интеллекту с международным участием КИИ'2000. - Москва: Изд-во Физико-математической литературы, 2000. С. 430-438.
14. Bookstein A.,S Klein . T. *Clumping Properties of Content-Bearing Words* // JASIS. 1998. №2.

Математические структуры и моделирование
2001, вып. 8, с. 141–145

УДК 519.86:336.2.027

К ВОПРОСУ СОЧЕТАНИЯ СТИМУЛИРУЮЩЕГО ВОЗДЕЙСТВИЯ НА НАЛОГОПЛАТЕЛЬЩИКОВ И ВНЕЭКОНОМИЧЕСКОГО ПРИНУЖДЕНИЯ ДЛЯ ПОВЫШЕНИЯ ПОСТУПЛЕНИЯ В БЮДЖЕТНУЮ СИСТЕМУ

Г.В. Песчанских

In this article the approach to the choice of balance between some abilities to improve of level tax and financial discipline are presented. It is an economical incitement and force get on taxpayer. The author built the mathemetical model that are demonstrated the functional correlation between reduction of taxes and raise of working efficiency of legal and control orgfns on first steps of reforms in tax sphere.

В период активного реформирования экономики уклонение от уплаты налогов, перемещение капиталов в «теневой» сектор экономики приобрели массовый характер. Налоговые правонарушения все более отличаются осмысленностью и организованностью исполнения. Государство, в лице его органов, предпринимает активные меры по противодействию этим явлениям. На начальном этапе реформ предпочтение отдавалось силовым методам воздействия. В настоящее время все более используются методы экономического стимулирования законопослушного поведения — понижение ставок налогов.

Общепризнано, что усиление налогового давления способно повысить поступления в бюджет только на короткий период времени. В ближнесрочном периоде это привлечет дополнительные средства от законопослушных налогоплательщиков, т.е. создаст дисбаланс налогового бремени в пользу правонарушителей. В долгосрочной перспективе законопослушные налогоплательщики могут не выдержать конкуренции и прекратят существование, налоговая база сократится, снизится общее поступление средств в бюджетную систему. В то же время это может стимулировать отток капиталов в «теневой» сектор экономики.

При понижении ставки такой дисбаланс не создается, налоговое бремя распределяется более равномерно, но только в долгосрочной перспективе. При этом в течение некоторого периода времени после снижения ставок налогоплательщики «по инерции» будут скрывать свой доход. В итоге, пока принципы законопослушания не войдут в практику повседневной жизни, государство будет

Региональная служба информационно-технологического обеспечения Управления ФСНП России по Свердловской области

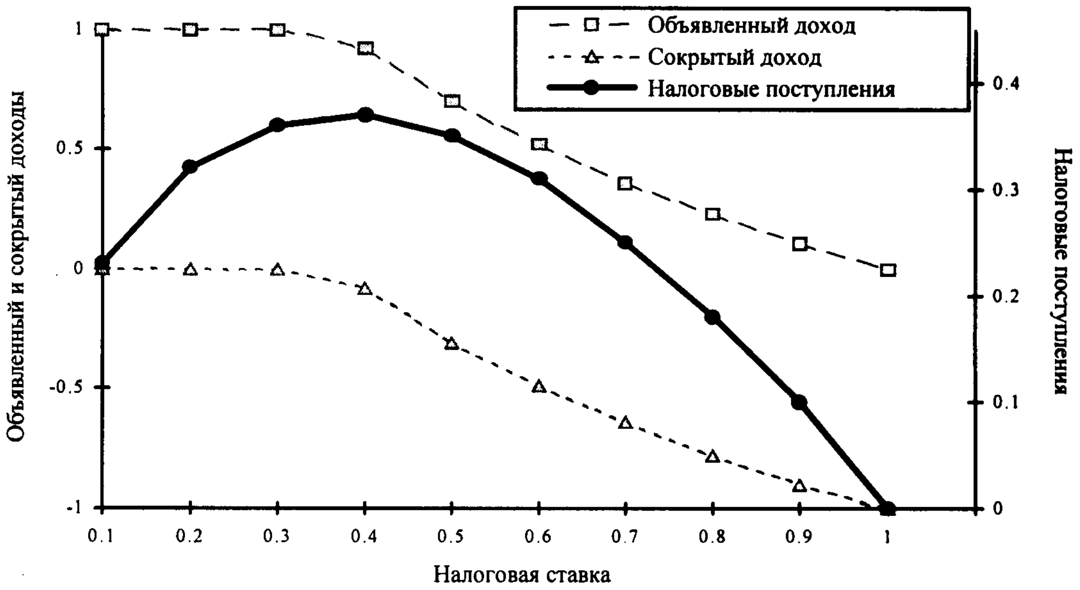


Рис. 1. Зависимость объявленных и сокрытых доходов от величины налоговой ставки

недополучать существенные суммы. Возникшую проблему покрытия затрат государства во время проведения таких реформ можно компенсировать путем активизации работы государственных органов.

В связи с этим встает вопрос об оптимальном количественном соотношении подходов, основанных на экономическом стимулировании и силовом принуждении. Необходим комплексный подход к решению данной проблемы, мобилизация и согласование между собой всех возможностей для противодействия налоговой преступности.

Для формирования оценочных соотношений используем кривую Лафера. Согласно Лаферу, при ставке более $\tau = 0,368$ [1] начинается процесс сокрытия доходов от налогообложения, падают налоговые сборы (Рис. 1).

Дадим некоторые количественные оценки того, как связано понижение ставок налогов с повышением эффективности работы государственных органов (контролирующих и правоохранительных). В качестве основного условия потребуем, чтобы при таком изменении доходы государственной казны остались неизменными.

Рассмотрим кривую Лафера с учетом работы государственных органов по выявлению сокрытого дохода и возвращению ее в легальную экономику. Все приведенные ниже рассуждения будут касаться ставок налога выше оптимальной: $0,368 \le \tau < 1$.
Объявленный доход:

$$Y(\tau) = Y_0 |\ln(\tau)|. \tag{1}$$

Сокрытый доход:

1. Без учета работы государственных органов по его легализации – $Y_c(\tau)$:

$$Y_c(\tau) = Y_0 - Y(\tau) = Y_0\Big[1 - |\ln(\tau)|\Big]. \tag{2}$$

2. С учетом работы государственных органов по его легализации – $Y_c^s(\tau)$:

$$Y_c^s(\tau) = Y_c(\tau)(1-\gamma). \tag{3}$$

Здесь γ – производительность работы государственных органов по возвращению сокрытого дохода в легальный сектор экономики:

$$\gamma = \frac{\begin{array}{c}\{\text{Часть скрытых доходов,выявленная}\\ \text{государственными органами и возвращенная в}\\ \text{легальную экономику}\}\end{array}}{\{\text{Общая сумма скрытых доходов}\}}. \tag{4}$$

Поступления в бюджет:
1. Максимально возможные – $T_m(\tau)$:

$$T_m(\tau) = Y_0\tau = Y(\tau)\tau + Y_c(\tau)\tau. \tag{5}$$

2. Реальные, без учета работы государственных органов – $T(\tau)$:

$$T(\tau) = Y(\tau)\tau = Y_0\tau - Y_c(\tau)\tau. \tag{6}$$

3. Реальные, с учетом работы государственных органов – $T^s(\tau)$:

$$T^s(\tau) = Y_0\tau - Y_c^s(\tau)\tau = Y_0\tau - Y_c(\tau)\tau(1-\gamma). \tag{7}$$

Потери бюджета:
1. Без учета работы государственных органов – $T_c(\tau)$:

$$T_c(\tau) = Y_c(\tau)\tau = Y_0\tau - Y(\tau)\tau. \tag{8}$$

2. С учетом работы государственных органов – $T_c^s(\tau)$:

$$T_c^s(\tau) = Y_c^s(\tau)\tau = Y_c(\tau)\tau(1-\gamma). \tag{9}$$

При переходе от ставки τ_1 к ставке τ_2 какое-то время налогоплательщики продолжают скрывать полученный ими доход, т.е.:

$$Y_c(\tau_1) = Y_c(\tau_2). \tag{10}$$

Поступления в бюджет изменяются следующим образом:
1. Были до изменения ставки $(\tau = \tau_1) - T_1^s(\tau)$:

$$T_1^s(\tau) = Y_0\tau_1 - Y_c(\tau_1)\tau_1(1-\gamma_1). \tag{11}$$

2. Стали после изменения ставки $(\tau = \tau_2) - T_2^s(\tau)$:

$$T_2^s(\tau) = Y_0\tau_2 - Y_c(\tau_1)\tau_2(1-\gamma_2). \tag{12}$$

Здесь τ_1 и τ_2 – старая и новая ставки налога, а γ_1 и γ_2 - старая и новая эффективность работы государственных органов соответственно.

Для сохранения после перехода на новую ставку налога доходов бюджета на прежнем уровне должно выполняться следующее равенство:

$$T_1^s(\tau) = T_2^s(\tau). \tag{13}$$

Подставляя сюда данные выше выражения, можно получить следующие равенства:

$$(\tau_1 - \tau_2) - \Big[1 - |\ln(\tau_1)|\Big][\tau_1(1-\gamma_1) - \tau_2(1-\gamma_2)] = 0, \tag{14}$$

$$\frac{(\tau_1 - \tau_2)}{\tau_1(1-\gamma_1) - \tau_2(1-\gamma_2)} = 1 - |\ln(\tau_1)|, \tag{15}$$

$$\tau_1\left[1 - (1-\gamma_1)\Big(1 - |\ln(\tau_1)|\Big)\right] = \tau_2\left[1 - (1-\gamma_2)\Big(1 - |\ln(\tau_1)|\Big)\right]. \tag{16}$$

Из этих равенств следует, что если при переходе на новую ставку налога эффективность работы государственных органов не изменяется $(\gamma_1 = \gamma_2)$, то единственной возможностью не допустить снижения поступлений в бюджет является полное отсутствие сокрытия как до, так и после преобразования: $Y_c(\tau_1) = Y_c(\tau_2) = 0$.

Таким образом, формулы (14),(15),(16) в разной форме выражают один и тот же факт: сохранение доходов казны при понижении ставки налогов невозможно без одновременного повышения эффективности работы государственных органов. Иначе, на первоначальном этапе, это может вызвать падение доходов консолидированного бюджета.

Выражения (14),(15),(16) позволяют количественно оценить повышение эффективности работы государственных органов, необходимое для сохранения доходов бюджета на прежнем уровне. Соотношение (17) показывает, насколько можно снизить налоговое бремя (τ_2) при повышении эффективности работы государственных органов с γ_1 на γ_2, если в настоящее время действует ставка налога τ_1.

$$\tau_2 = \tau_1 \frac{1 - (1-\gamma_1)\Big(1 - |\ln(\tau_1)|\Big)}{1 - (1-\gamma_2)\Big(1 - |\ln(\tau_1)|\Big)}. \tag{17}$$

Можно поставить вопрос иначе. Насколько необходимо повысить результативность работы государственных органов (с γ_1 на γ_2) для того, чтобы можно было без потерь для бюджета снизить ставку налогов с τ_1 до τ_2.

Суть этого соотношения (17) понятна на интуитивном уровне – снижение поступлений от законопослушных налогоплательщиков должно быть восполнено выведенными из «тени» средствами правонарушителей. Однако соотношение (17) позволяет связать в единую систему экономические стимулы к работе в правовом поле и правоохранительное давление на правонарушителей, причем на уровне количественных оценок.

Количественные оценки «теневого» сектора экономики известны. Оптимальные ставки налогов определены. Полученное соотношение (17) дает возможность спрогнозировать требуемое увеличение результативности работы государственных органов (в частности органов налоговой полиции) в период реформирования налоговой системы. В итоге это позволит более обоснованно ставить задачи государственным органам, определять их численность и объемы необходимого финансирования.

Литература

1. Королев Е.А. *Организационно-экономический механизм управления потоками налоговых платежей:* Диссертация на соискание степени кандидата экономических наук (на правах рукописи). Екатеринбург: Институт экономики УрО РАН, 1998. 171 с.

Математические структуры и моделирование
2001, вып. 8, с. 146–155

УДК 519.86:336.2.027

ВОЗМОЖНОСТЬ МОДЕЛИРОВАНИЯ ДЕЯТЕЛЬНОСТИ ОРГАНОВ НАЛОГОВОЙ ПОЛИЦИИ И НАПРАВЛЕНИЯ ИСПОЛЬЗОВАНИЯ ЕГО РЕЗУЛЬТАТОВ

Г.В. Песчанских

In this article the peculiarities of the tax police organs that help modeling it's work are presented. It is "technological" correlation between ordered steps. The directeds of use this approach for solve of legal problems in the tax police area, optimal utilization of resourses are demonstrated.

В период активного реформирования экономики России широкое распространение получили налоговые преступления. Объем «теневого» сектора экономики исчисляется несколькими десятками процентов от ВВП. Можно упомянуть о таких негативных явлениях, как вывоз капиталов за рубеж, фиктивные платежи в бюджет, «лжеэкспорт» и неправомерное возмещение НДС [1]. Рост налоговой преступности стал основной причиной создания специализированного правоохранительного органа – налоговой полиции, который призван противодействовать негативным явлениям в налоговой сфере.

Годы реформ показали, что налоговая преступность не является статичной. Она развивается вместе с экономикой государства. Это ставит перед органами налоговой полиции, иными правоохранительными и контролирующими ведомствами задачу маневра силами и средствами адекватно обстановке, сложившейся в налоговой сфере. В данной работе представлены подходы к моделированию деятельности органов налоговой полиции, которые позволяют выделить приоритетные направления работы, переориентировать на них личный состав налоговой полиции и определить эффективность их воздействия на налоговую преступность.

Сделаем ряд предварительных замечаний.

Во-первых, налоговая преступность носит латентный, скрытый характер. В отличие от общеуголовных после налоговых преступлений не остаются видимые следы и о них не пишут заявление в милицию. Это затрудняет объективную оценку истинных масштабов налоговой преступности в стране. Однако в отличие от органов МВД это обстоятельство предоставляет некоторую степень свободы налоговым полицейским при выборе ими направлений своей работы. Состояние и тенденции развития налоговой преступности могут быть оценены

Региональная служба информационно-технологического обеспечения Управления ФСНП России по Свердловской области

по внешним данным, которые носят косвенный характер, и по результатам воздействия на нее, т.е. на основе рассмотрения результатов своей работы. Эта особенность налоговой преступности создает возможность для активного выбора направлений работы и, следовательно, потребность в прогнозировании результатов деятельности органов налоговой полиции на основе их моделирования.

Во-вторых, следует упомянуть о необходимых условиях для такого моделирования. Цель органов налоговой полиции должна быть формализована в виде конкретных, измеримых и достижимых показателей. Сделать это однозначным образом на основании закона, регламентирующего их деятельность, не так просто [2].

Согласно действующему законодательству органы налоговой полиции являются правоохранительным органом и частью сил по обеспечению экономической безопасности государства. Т.е. цели органов налоговой полиции присущ финансово-правовой дуализм. Результатом их деятельности может быть как поступление средств в бюджет, так и установление, обвинение виновных лиц и осуждение их [3]. Финансовые и правоохранительные результаты деятельности налоговой полиции не всегда являются одновременно достижимыми.

Цель налоговой полиции может быть формализована как максимизация поступлений в бюджет либо как максимально возможное доведение выявленных налоговых преступлений до суда. Возможны комбинации этих целевых установок. В задачи данной работы не входит обсуждение наиболее оптимального представления цели органов налоговой полиции. Будем полагать, что математическим представлением цели органов налоговой полиции является максимум доведения до суда выявленных налоговых преступлений.

$$\text{MAX}\left(\frac{\text{ПРЕСЕЧ}^{\text{ННЗ}}}{\text{ВЫЯВ}^{\text{ННЗ}}}\right), \tag{1}$$

где: $\text{ВЫЯВ}^{\text{ННЗ}}$ – количество полученных сообщений о фактах и признаках нарушения налогового законодательства (ННЗ);

$\text{ПРЕСЕЧ}^{\text{ННЗ}}$ – количество сообщений о фактах и признаках ННЗ, доведенных до предъявления обвинительного заключения (или осуждения виновных лиц).

В соответствии с действующим законодательством задачи органов налоговой полиции сформулированы как выявление, предупреждение и пресечение налоговых преступлений. Выражение (1) устанавливает желаемое соотношение между количественными показателями этих задач. Чтобы определить механизм влияния на их соотношение, необходимо установить связь между этими задачами.

Исходной точкой в работе органов налоговой полиции, как и любого другого правоохранительного органа, являются полученные сообщения о фактах и признаках преступлений, отнесенных к его компетенции. В ходе последующей работы эти сообщения могут в большей мере привести к возмещению причиненного бюджету ущерба или к возбуждению уголовного дела и осуждению виновных лиц. Т.е. задача руководства территориальных органов налоговой полиции сориентировать имеющиеся силы и средства на выявление нарушений налогового законодательства, которые обеспечат достижение поставленной цели (выполнение выражения (1)).

Возможности для маневра силами и средствами заложена, во-первых, в технологичности работы органов налоговой полиции. Между получением сигнальной информации и пресечением ННЗ лежит ряд этапов: оперативная проверка, документирование правонарушения, правовая оценка и предварительное следствие. Во-вторых, налоговая преступность и сигнальная информация имеют внутреннюю структуру: по схемам нарушений, по видам налогов, по секторам экономики и т.п. Суть предлагаемого подхода к моделированию деятельности налоговой полиции заключается в том, что он должен позволить получать на начальных этапах работы сведения о тех ННЗ, которые с наибольшей вероятностью могут дать желаемый конечный результат.

Следует заранее предостеречь, что в связи с этим может возникнуть проблема выбора между «тянущей» и «толкающей» технологиями работы, когда при наличии нескольких последовательно осуществляемых видов деятельности (оперативно-розыскная, документально-проверочная и уголовно-процессуальная) приоритет отдается одному из них. При «толкающей» технологии — это начальное звено. Особое внимание на начальном этапе уделяется количеству и качеству промежуточных результатов: количеству первичной информации, оперативных проверок и т.п. В дальнейшем остается только обеспечить ее «прохождение» к следующим звеньям технологической цепи. Позитивным при таком подходе является то, что количество первичных и промежуточных результатов (исходной информации, дел оперативной проверки и т.п.) стимулирует количественный рост итоговых результатов на конечных стадиях работы с ней: количество возбужденных и оконченных уголовных дел, суммы нанесенного ущерба и т.п. Отрицательным является то, что количество этой информации может быть несоизмеримо с возможностями подразделений на конечных стадиях работы с выявленными правонарушениями. В итоге – снижение качества, возрастание количества «брака» в работе. Например, увеличение количества материалов, возвращенных на доследование, падение сумм возмещенного ущерба.

При «тянущей» технологии, наоборот, в приоритетном положении находятся подразделения, замыкающие технологическую цепочку. Доминирование следователей в этом процессе стимулирует рост качества конечного результата, снижение степени «брака» в работе (снижение количества уголовных дел, возвращенных на дополнительное расследование, прекращенных по реабилитирующим основаниям, хороший процент возмещения ущерба и т.д.). В то же время это, как правило, приводит к количественному снижению показателей работы, например к уменьшению количества возбужденных уголовных дел. Кроме того, это может привести к работе оперативников «под заказ» следователей. Такой подход может ориентировать оперативный состав на выявление простых нарушений. Налоговые правонарушения, которые не допускают двоякого толкования, как правило, являются не самыми крупными и не самыми социально опасными.

Оптимальным является поддержание баланса между этими видами деятель-

Таблица 1. Методика вычисления вероятности получения основных результатов оперативно-служебной деятельности

Абсолютные значения для ряда последовательных мероприятий		Вероятность получения конечных результатов	
		Возбуждено уголовных дел	Окончено уголовных дел
Получено первичных информаций	А	Г/А $\cdot$ 100%	Д/А $\cdot$ 100%
Заведено подборок материалов	Б	Г/Б $\cdot$ 100%	Д/Б $\cdot$ 100%
Заведено дел оперативной проверки	В	Г/В $\cdot$ 100%	Д/В $\cdot$ 100%
Возбуждено уголовных дел	Г	100%	100%
Окончено уголовных дел	Д		100%

ности органов налоговой полиции. При этом следует понимать, что закладывается результат в начале технологической цепочки в оперативном подразделении, а оформляется (в конечном виде, когда можно оценить его качество) только на последней стадии этого процесса в следственном подразделении. Не следует преуменьшать значимость каждого из подразделений, принимающих участие в достижении конечного результата. Целесообразным будет поддержание разумного баланса между «тянущей» и «толкающей» технологиями.

Вместе с тем данная особенность – наличие растянутой по времени «технологической» процедуры – позволяет проводить анализ, направленный на прогнозирование результатов, и предлагать варианты управленческих решений при неудовлетворительных итогах прогноза.

Сделаем некоторые предположения. Во-первых, будем полагать, что результаты, получаемые одними и теми же подразделениями в том или ином секторе экономики, не подвержены серьезным изменениям и не сильно отклоняются от своих среднестатистических значений. Т.е. будем считать, что в первом приближении ряд величин (вероятность проведения результативной проверки и возбуждения уголовного дела на основании первичной информации, а также количественные показатели этих мероприятий и время на их проведение) можно считать постоянными. Во-вторых, будем полагать, что результаты, получаемые одними и теми же подразделениями по тому или иному направлению работы, не подвержены серьезным изменениям и не сильно отклоняются от своих среднестатистических значений.

В итоге на основании полученных конечных результатов можно изучить основные показатели, характеризующие деятельность органов налоговой полиции в разрезе секторов экономики, видов нарушений и т.п. и рассчитать вероятность получения желаемых конечных результатов. Повышение общей результативности может быть достигнуто посредством активизации работы на наиболее перспективных направлениях (обучением или дисциплинарными методами воздействия на сотрудников) либо путем перераспределения сил и средств. В таблице 1 приведены числовые данные для одного из секторов экономики, раскрывающие сущность этого подхода. Таблица 2 показывает пример применения

Таблица 2. Вероятности получения конечных результатов оперативно-служебной деятельности с разбивкой по секторам экономики

	Топливно-энергетический комплекс		Лесная промышленность		Кредитно-финансовая сфера		Сфера услуг	
	Возб УД	Окон УД	Возб УД	Окон УД	Возб УД	Окон УД	Возб УД	Окон УД
Получено первичных информаций	21%	10%	34%	19%	93%	80%	48%	41%
Заведено подборок материалов	36%	18%	51%	28%	116%	100%	39%	33%
Заведено дел оперативной проверки	57%	28%	150%	83%	350%	300%	137%	115%
Возбуждено уголовных дел	100%	50%	100%	86%	100%	85%	100%	84%
Окончено уголовных дел		100%		100%		100%		100%
	Производство и оборот подакциз. товаров		Строительство		Машиностроение и металлообработ.		Металлургия	
	Возб УД	Окон УД	Возб УД	Окон УД	Возб УД	Окон УД	Возб УД	Окон УД
Получено первичных информаций	14%	5%	45%	51%	33%	33%	28%	19%
Заведено подборок материалов	36%	13%	58%	66%	57%	57%	32%	21%
Заведено дел оперативной проверки	135%	50%	103%	117%	109%	109%	75%	50%
Возбуждено уголовных дел	100%	37%	100%	113%	100%	100%	100%	67%
Окончено уголовных дел		100%		100%		100%		100%

	Транспорт		Пищевая промышленность		С/х-во и животноводство		Иное	
	Возб УД	Окон УД	Возб УД	Окон УД	Возб УД	Окон УД	Возб УД	Окон УД
Получено первичных информаций	27%	13%	93%	64%	45%	13%	47%	21%
Заведено подборок материалов	30%	15%	118%	82%	106%	31%	56%	25%
Заведено дел оперативной проверки	100%	50%	650%	450%	81%	24%	300%	133%
Возбуждено уголовных дел	100%	50%	100%	69%	100%	29%	100%	44%
Окончено уголовных дел		100%		100%		100%		100%

данного подхода к анализу результатов в разрезе секторов экономики [1].

Представленные данные (табл. 2) показывают вероятность получения отдельных конечных результатов (возбужденных и оконченных уголовных дел) в каждом из указанных секторов экономики. При этом важным является не только вероятность доведения до логического завершения полученных первичных информаций, но и иных мероприятий. Если начальная вероятность достаточно велика и она резко нарастает по мере работы с нарушением, заниматься работой в таком секторе экономики выгодно, и наоборот.

В то же время для принятия решения о переориентации сил и средств на выделенные в ходе такого анализа сектора экономики приведенных данных недостаточно. Во-первых, имеются социально значимые сектора экономики, работать в которых необходимо, несмотря на ее невыгодность. Во-вторых, переориентация сил и средств на выделенный сектор экономики возможна, только если он составляет значительную долю в экономике обслуживаемой территории. В-третьих, результаты должны иметь плавную динамику во времени. В противном случае предсказать их значение в будущем будет затруднительно.

Таким образом, для принятия управленческих решений необходимо учитывать:

[1] В связи с ограничением на доступ к сведениям о результатах деятельности органов налоговой полиции в таблице 1 приведены условные данные, демонстрирующие применение данного подхода.

1. Абсолютные величины результатов.
2. Вероятность доведения работы по делу до логического завершения.
3. Плавность динамики результатов во времени.

Кратко смысл этих показателей можно определить как размер выявляемых правонарушений, вероятность их пресечения и стабильность получения результатов на этом направлении. Проведем эту работу и сведем ее результаты в единую таблицу (табл. 3). Для сравнимости между собой по каждому показателю припишем изучаемым секторам экономики места от первого для наилучшего значения, до двенадцатого – для наихудшего. Затем суммируем их и укажем итоговый рейтинг, смыслом которого является комплексная оценка целесообразности работы органов налоговой полиции в данном секторе экономики (табл. 3).

Таким образом, комплексное изучение целесообразности проведения мероприятий в разрезе секторов экономики (табл. 3) показывает, что наиболее согласуется с формально определенной выше целью (1) работа в строительной отрасли, машиностроении и металлообработке, лесной промышленности и сфере услуг [2].

Аналогично тому, как это сделано для секторов экономики (табл. 3), можно провести подобный формальный анализ по видам нарушений и другим основаниям. В итоге можно будет выделить не только приоритетные для работы секторы экономики, но и наиболее перспективные виды нарушений налогового законодательства, активизация работы на которых повысит результативность работы органа налоговой полиции.

Акцентируем внимание на формальной стороне данного метода. В его основу положены данные о результатах работы органа налоговой полиции. При этом анализ форм и методов совершения налоговых правонарушений, характеристики внешней среды, тенденции развития секторов экономики и т.п. в расчет не принимались. Т.е. делать заключения об изменении тактики работы по итогам такого исследования нужно крайне осторожно. Правильнее относиться к данному анализу как к вспомогательному средству, указывающему направления дальнейшего более подробного изучения особенностей складывающейся обстановки. Выше уже упоминалось о наличии социально значимых секторов экономики, некоторые виды нарушений трудно выявляются, но их высокая социальная опасность обусловливает активную работу с ними органов налоговой полиции. Так, в соответствии с проведенным анализом (табл.3) деятельность в сфере производства и оборота подакцизных товаров оказалась почти бесперспективной. В то же время следует учитывать высокую криминогенность этой сферы, ее важность для стабилизации общей социально-экономической и бюджетно-налоговой обстановки в стране. Кроме того, невысокий итоговый рейтинг, видимо, обусловлен трудностями при выявлении и пресечении нарушений в данной сфере.

[2] В данном случае (табл. 3) не учтена представленность секторов экономики на рассматриваемой территории. Более корректно будет дополнить данный анализ, например, суммами налогов, собираемых по каждому из секторов экономики, либо численностью налогоплательщиков в них и т.п.

Таблица 3. Комплексная оценка результативности мероприятий, проводимых в различных секторах экономики

	Топливно-энергетический комплекс	Лесная промышленность	Кредитно-финансовая сфера	Сфера услуг
Итоговый рейтинг	5	3	12	4
1Р. Место по абсолютной величине результата результата	6	4	10	2
2Р. Место по достижимости конечного результата	10	4	7	3
3Р. Место по прогнозируемости получения конечного результата	1	4	12	9

	Производство и оборот подакциз. товаров	Строительство	Машиностроение и металлообработ.	Металлургия
Итоговый рейтинг	7	1	2	6
1Р. Место по абсолютной величине результата результата	3	1	4	10
2Р. Место по достижимости конечного результата	12	1	5	5
3Р. Место по прогнозируемости получения конечного результата	6	3	2	4

	Транспорт	Пищевая промышленность	С/х-во и животноводство	Иное
Итоговый рейтинг	9	8	10	11
1Р. Место по абсолютной величине результата результата	7	9	7	12
2Р. Место по достижимости конечного результата	8	2	11	8
3Р. Место по прогнозируемости получения конечного результата	9	11	7	7

Рассмотрим некоторые из управленческих решений в плане возможности или невозможности их принятия по результатам анализа данного вида. Во-первых, можно сравнить экономическую структуру обслуживаемой территории со структурой, на которую ориентирует данный анализ. В случае ее несовпадения могут быть внесены коррективы в расстановку сил и средств, акценты в работе.

Во-вторых, данный анализ может явиться индикатором того, что появился новый вид налогового правонарушения и/или какое-то подразделение научилось с ним бороться. Признаком этого может явиться рост первичных информаций или иных результатов на каком-либо выделенном направлении работы. По этим результатам может быть принято решение о дополнительном изучении причин такого роста и далее – о систематизации и распространении положительного опыта.

В-третьих, такой вид анализа фактически раскрывает структуру накопленного сотрудниками органа налоговой полиции опыта работы. Как следствие, это должно быть учтено при разработке планов обучения сотрудников, командировок по обмену опытом и т.п.

Могут обсуждаться и другие управленческие решения. В любом случае перед их принятием должны обсуждаться все аспекты того или иного сектора экономики или вида деятельности. В частности нет смысла полностью отказываться от одних направлений работы и полностью ориентироваться на иные. Данный анализ призван помочь найти оптимальное сочетание между основными и «страховочными» направлениями работы. Например, при расстановке акцентов в части выявления налоговых и неналоговых правонарушений и т.п. В каждой конкретной ситуации они могут существенно различаться, и предложить универсальные рекомендации здесь вряд ли возможно. Данный подход позволяет критически взглянуть на структуру деятельности органа налоговой полиции и его результаты, перекомпоновать их наиболее оптимальным образом.

Можно указать еще на одно важное применение рассмотренного подхода – индикативное.

Опыт развития органов налоговой полиции уже научил, что если налоговое нарушение не обусловлено существенными пробелами в действующем законодательстве, то его использование проходит ряд стадий. На первом этапе происходит придумывание новой схемы ухода от налогообложения. На втором этапе она распространяется среди налогоплательщиков. После того как схема становится известной органам налоговой полиции и налоговым органам, начинается третий этап, в ходе которого происходит наработка опыта противодействия данной схеме, поиск форм и методов наиболее эффективного выявления и пресечения налогового преступления. Четвертый этап заключается в активном и массовом противодействии органов налоговой полиции данному нарушению, которое приводит к существенному снижению количества таких нарушений и причиненного ущерба. Это приводит к тому, что на пятом этапе данное нарушение прекращает свое существование либо (при наличии к тому объективных предпосылок) устанавливается равновесие в противоборстве незаконопослушных налогоплательщиков и государственных ведомств.

Работа органов налоговой полиции должна быть согласована с данным циклом развития налоговых преступлений. Например, посредством структурного анализа результатов органа налоговой полиции могут быть своевременно выявлены новые схемы уклонения от налогообложения. По полученным конечным результатам можно установить наличие или отсутствие опыта их пресечения. Таким образом, предложенный выше подход может рассматриваться, во-первых, как индикатор начала третьего, четвертого и пятого этапов в «жизненном цикле» налогового правонарушения. Во-вторых, такой подход является эффективным средством обмена опытом, который позволяет сократить продолжительность этих этапов, принимать меры превентивного характера, затрудняющие их территориальное распространение.

В итоге, с одной стороны, данный подход позволяет формальными методами повысить результаты органа налоговой полиции, не очень задумываясь при этом об особенностях сложившейся обстановки. С другой стороны, он позволяет осмысленно выявлять эти особенности и наиболее эффективно противодействовать негативным тенденциям в ее развитии.

В результате применения данного подхода могут быть сделаны предложения, касающиеся не только выбора линий, форм и методов работы, объектов для проведения проверок. На основе этого подхода могут быть сформированы предложения о внесении иных изменений в структуру и организацию деятельности органа налоговой полиции. Рассмотренный подход показывает, что деятельность органов налоговой полиции поддается моделированию. При его наличии структура их деятельности может рассматриваться как система, самоадаптирующаяся к изменениям оперативной обстановки, развитию налоговой преступности.

Такая самоадаптирующаяся структура должна включать аналитическую, организационную и информационную составляющие. Эти элементы предназначены для повышения гибкости системы сбора информации и управления подчиненными подразделениями, возможности для модификации данной системы адекватно вновь появляющимся условиям. Данная работа призвана сделать один из первых шагов в создании аналитической составляющей данной системы управления.

Литература

1. Девятов А.Н. *Проблемы исследования «теневого» сектора экономики России* / В кн.: Экономическая безопасность. Производство-Финансы-Банки / Под ред. В.К.Сенчагова. М.: ЗАО «Финстатинформ», 1998. С.503-507.
2. *О федеральных органах налоговой полиции.* Закон РФ от 24.06.93г. № 5238-1.
3. *Налоговая полиция* / Учебно-практическое пособие. М., 1994. 110с.

Математические структуры и моделирование
2001, вып. 8, с. 156–166

УДК 51:378

АНАЛИЗ ТИПИЧНЫХ ОШИБОК, ДОПУСКАЕМЫХ СТУДЕНТАМИ ПРИ ИЗУЧЕНИИ ПРЕДЕЛА ФУНКЦИИ И ДИФФЕРЕНЦИАЛЬНОГО ИСЧИСЛЕНИЯ

В.А. Далингер

The article analyses the subject and containing aspects of typical mistakes made by students in the process of the function limit studying and differential counting of the function of a variable.

В данной статье будут проанализированы предметно-содержательные аспекты **типичных ошибок, допускаемых студентами в процессе изучения предела функции** и дифференциального исчисления функции одной переменной.

Прежде чем переходить к разбору типичных ошибок, допускаемых студентами при вычислении пределов функций, сделаем два существенных замечания.

Замечание 1. В отношении понятия «бесконечность» в математическом анализе существует две точки зрения. Довольно значительная часть авторов (А.Ф.Бермант, В.И.Смирнов, Г.П.Толстов, Н.А.Фролов и др.) употребляют три символа бесконечности: $\infty, +\infty, -\infty$, понимая под ними следующее: $x_n \to \infty$, если $|x_n| > M$ при $n > N$; $x_n \to +\infty$, если $x_n > M$ при $n > N$; $x_n \to -\infty$, если $x_n < -M$ при $n > N$.

Другие же авторы (М.К.Гребенча, С.И.Новоселов, А.Я.Хинчин, В.Немецкий и др.) пользуются только двумя символами: $+\infty$ и $-\infty$, причем знак плюс иногда опускается, так что между символами $+\infty$ и ∞ нет никакого различия, как нет различия между числами $+5$ и 5.

Наконец, есть авторы, вставшие на промежуточную точку зрения. Так, в сборнике задач Н.А.Давыдова и др. «Сборник задач по математическому анализу: М.: Гостехизадт, 1953 г.» для предела на бесконечности пишут $x \to \infty$, $(x \to +\infty, x \to -\infty)$, а для бесконечного предела функции лишь $+\infty$ и $-\infty$.

Принятие первой точки зрения в преподавании обычно приводит к тому, что студенты под записью $x \to \infty$ подразумевают все-таки $x \to +\infty$, тем самым лишь укореняя вредную привычку иметь в виду лишь положительные значения. В этом грешат также сами авторы учебников и задачников. Так, рассматривая предел последовательности, пишут $n \to \infty$, тогда как при последовательном

E-mail: dalinger@omgpu.omsk.edu
Омский государственный педагогический университет

проведении своей точки зрения следовало бы писать $n \to +\infty$. Рассматривая бесконечный интервал, пишут $(a; \infty)$ вместо $(a; +\infty)$. Таким образом, символ ∞ употребляют двойственно: и как бесконечность неопределенного знака, и как положительную бесконечность, создавая тем самым ненужную путаницу и неопределенность.

В курсе математического анализа можно найти много примеров, когда требуется знание бесконечности определенного знака и, напротив, всегда можно обойтись без употребления бесконечности неопределенного знака. Так, желая показать, что при $x \to +\infty$ и $x \to -\infty$ функция имеет обыкновенный предел, можно писать $\lim_{x \to \pm\infty} f(x) = A$ вместо $x \to \infty$.

Можно надеяться, что вторая точка зрения, будучи более определенной, в конечном счете вытеснит первую, как не имеющую никаких преимуществ.

Замечание 2. Пусть требуется вычислить предел функции $f(x)$ в точке $a : \lim_{x \to a} f(x) = A$.

Для существования предела не требуется, чтобы функция была определена при $x = a$. Что же касается определения функции в некоторой окрестности этой точки, то в отношении этого существует две различные точки зрения. Одни требуют, чтобы функция была определена во всех точках некоторой окрестности точки а (А.Ф.Бермант, Н.А.Фролов, Н.А.Давыдов и др.). Другие же встают на более общую точку зрения, требуя лишь того, чтобы точка a была предельной для множества значений аргумента (Г.М.Фихтенгольц, Г.П.Толстов, М.К.Гребенча, С.И.Новоселов и др.).

В этом отношении интересен следующий пример. Задача. Существует ли предел функции $f(x) = \frac{\sin \frac{1}{x}}{\sin \frac{1}{x}}$ в точке $x = 0$?

Решение

Принимая вторую точку зрения, ответить на поставленный вопрос следует положительно: в любой окрестности нуля существуют точки, в которых функция определена, причем во всех таких точках ее значение равно 1. Следовательно, предел существует и равен 1.

Если стать на первую точку зрения, то ответ на поставленный вопрос отрицательный: предел не существует, так как в любой окрестности нуля имеются точки, в которых функция не определена.

Заметим, что в теории функций принята вторая точка зрения. Было бы нецелесообразно при изучении математического анализа принять первую точку зрения и дать на поставленный вопрос отрицательный ответ, а спустя некоторое время, когда студенты приступят к изучению курса «Теория функций», встать на более общую точку зрения и на тот же вопрос давать уже положительный ответ. Лучше сразу встать на общую точку зрения (хотя мы должны здесь заметить, что если придерживаться первой точки зрения, то ответ – «предела не существует» будет верным).

Перейдем к анализу типичных ошибок, допускаемых студентами при решении задач на вычисление пределов функций.

Задача 1. Найти $\lim\limits_{x \to +\infty} \frac{1-x}{\sqrt{1+x^2}}$.

Решение

Для вычисления предела надо раскрыть неопределенность вида $\frac{\infty}{\infty}$, для чего разделим числитель и знаменатель дроби на наивысшую степень переменной, то есть на x. Будем иметь

$$\lim_{x\to+\infty}\frac{1-x}{\sqrt{1+x^2}}=\lim_{x\to+\infty}\frac{\frac{1}{x}}{\sqrt{\frac{1}{x^2}+1}}=-1$$

($\frac{1}{x}$ и $\frac{1}{x^2}$ при $x\to+\infty$ являются бесконечно малыми, а поэтому их предел равен нулю).

Теперь поступим иначе. Внесем выражение $(1-x)$ под знак квадратного корня. Будем иметь

$$\lim_{x\to+\infty}\frac{1-x}{\sqrt{1+x^2}}=\lim_{x\to+\infty}\frac{\sqrt{(1-x)^2}}{\sqrt{1+x^2}}=\lim_{x\to+\infty}\frac{\sqrt{1+x^2-2x}}{\sqrt{1+x^2}}=$$

$$=\lim_{x\to+\infty}\sqrt{1-\frac{2x}{1+x^2}}=\lim_{x\to+\infty}\sqrt{1-\frac{\frac{2}{x}}{\frac{1}{x^2}+1}}=\sqrt{1-\frac{0}{1}}=1.$$

Легко увидеть, что функция $f(x)=\frac{1-x}{\sqrt{1+x^2}}$, стоящая под знаком предела, при всех значениях $x>1$ отрицательная и, следовательно, при $x\to+\infty$ не может иметь положительного предела.

Значит, второй способ привел к неверному результату.

Задача 2. Найти $\lim\limits_{x\to-\infty}\frac{\sqrt{1+x+x^2}}{2-x}$.

Решение

Раскроем неопределенность $\frac{\infty}{\infty}$, разделив числитель и знаменатель дроби на наивысшую степень переменной, то есть на x. Будем иметь:

$$\lim_{x\to-\infty}\frac{\sqrt{1+x+x^2}}{2-x}=\lim_{x\to-\infty}\frac{\sqrt{\frac{1}{x^2}+\frac{1}{x}+1}}{\frac{2}{x}-1}=-1.$$

Поступая иначе, а именно внеся знаменатель дроби под знак квадратного корня, приходим к следующему:

$$\lim_{x\to-\infty}\frac{\sqrt{1+x+x^2}}{2-x}=\lim_{x\to-\infty}\sqrt{\frac{1+x+x^2}{4-4x+x^2}}=\lim_{x\to-\infty}\sqrt{\frac{\frac{1}{x^2}+\frac{1}{x}+1}{\frac{2}{x^2}-\frac{4}{x}+1}}=\sqrt{\frac{0+0+1}{0-0+1}}=1.$$

На этот раз, как нетрудно видеть, неверный результат получен первым способом.

Поясним суть ошибок, допущенных при решении этих двух задач.

Решая задачу 1 первым способом, мы делили числитель на x, а подкоренное выражение, стоящее в знаменателе, на x^2, то есть мы применяли первую строчку формулы

$$|x| = \begin{cases} \sqrt{x^2}, & x \geq 0; \\ -\sqrt{x^2}, & x < 0, \end{cases}$$

считая $x > 0$.

Так как в этой задаче $x \to +\infty$, то мы получили правильный результат. Та же формула, применяемая к решению задачи 2, приводит к ошибке, ибо теперь $x \to -\infty$, и, следовательно, мы должны брать вторую строчку указанной формулы.

Аналогичное замечание можно сделать и относительно второго способа. Действительно, достаточно учесть, что в задаче 1 имеем: $(1-x) < 0$, а в задаче 2: $(2-x) > 0$.

Таким образом, при решении этих двух задач тем или иным способом мы получим правильный результат, если воспользуемся той строчкой формулы

$$|x| = \begin{cases} \sqrt{x^2}, & x \geq 0; \\ -\sqrt{x^2}, & x < 0, \end{cases}$$

которая соответствует знаку выражения, вносимого под знак квадратного корня.

Задача 3. Найти предел $\lim\limits_{x\to 0} \frac{x}{\sqrt{1-\cos x}}$.

Решение

Покажем вначале ошибочное решение, приводящее, конечно же, к неверному ответу, которое мы можем найти, например, в учебнике К.А.Поссе, И.И.Привалова «Курс дифференциального исчисления», 1938.

$$\lim_{x\to 0} \frac{x}{\sqrt{1-\cos x}} = \frac{1}{\sqrt{2}} \lim_{x\to 0} \frac{x}{\sin\frac{x}{2}} = \frac{1}{\sqrt{2}} \lim_{x\to 0} \frac{\frac{x}{2}}{\frac{\sin\frac{x}{2}}{2}} = \frac{2}{\sqrt{2}} \lim_{x\to 0} \frac{\frac{x}{2}}{\sin\frac{x}{2}} = \frac{2}{\sqrt{2}} = \frac{2\sqrt{2}}{2} = \sqrt{2}.$$

(При вычислении этого предела использован первый замечательный предел: $\lim\limits_{x\to 0} \frac{\sin x}{x} = 1$ или $\lim\limits_{x\to 0} \frac{x}{\sin x} = 1$).

Ошибка допущена из-за того, что использовано равенство $\sqrt{x^2} = x$ вместо $\sqrt{x^2} = |x|$.

Приведем верное решение.

Так как $\sqrt{1-\cos x} = \sqrt{2}|\sin\frac{x}{2}|$, то имеем

$$\lim_{x\to 0} \frac{x}{\sqrt{1-\cos x}} = \frac{1}{\sqrt{2}} \lim_{x\to 0} \frac{x}{|\sin\frac{x}{2}|}.$$

Полученное выражение $\frac{x}{|\sin\frac{x}{2}|}$ при $x \to 0$ предела не имеет, но существуют односторонние пределы.

При $x \to 0+$ имеем

$$\lim_{x\to 0+} \frac{x}{\sqrt{1-\cos x}} = \frac{1}{\sqrt{2}} \lim_{x\to 0+} \frac{x}{|\sin\frac{x}{2}|} = \frac{1}{\sqrt{2}} \lim_{x\to 0+} \frac{x}{\sin\frac{x}{2}} =$$

$$= \frac{1}{\sqrt{2}} \lim_{x\to 0+} \frac{\frac{x}{2}}{\frac{\sin\frac{x}{2}}{2}} = \frac{2}{\sqrt{2}} \lim_{x\to 0+} \frac{\frac{x}{2}}{\sin\frac{x}{2}} = \frac{2}{\sqrt{2}} = \frac{2\sqrt{2}}{2} = \sqrt{2}.$$

(При вычислении этого предела использован первый замечательный предел: $\lim_{x\to 0} \frac{\sin x}{x} = 1$ или $\lim_{x\to 0} \frac{x}{\sin x} = 1$).

При $x \to 0-$ будем иметь

$$\lim_{x\to 0-} \frac{x}{\sqrt{1-\cos x}} = \frac{1}{\sqrt{2}} \lim_{x\to 0-} \frac{x}{|\sin\frac{x}{2}|} = -\frac{1}{\sqrt{2}} \lim_{x\to 0+} \frac{x}{\sin\frac{x}{2}} = -\frac{2}{\sqrt{2}} = -\frac{2\sqrt{2}}{2} = -\sqrt{2}.$$

Задача 4. Вычислить предел $\lim\limits_{x\to a} \frac{\sqrt{ax}-x}{x-a}$.

В нашей практике все студенты, выполнявшие данное задание, получали ответ, что при $a > 0$ предел равен $(-\frac{1}{2})$. Их решение было таким.

Предполагая, что $x \neq a$ и $a > 0$, они преобразовывали функцию, стоящую под знаком предела:

$$\frac{\sqrt{ax}-x}{x-a} = \frac{(\sqrt{ax}-x)(\sqrt{ax}+x)}{(x-a)(\sqrt{ax}+x)} = \frac{ax-x^2}{(x-a)(\sqrt{ax}+x)} =$$

$$= \frac{x(a-x)}{(x-a)(\sqrt{ax}+x)} = -\frac{x}{\sqrt{ax}+x}.$$

Отсюда они получили:

$$\lim_{x\to a} \frac{\sqrt{ax}-x}{x-a} = -\lim_{x\to a} \frac{x}{\sqrt{ax}+x} = -\frac{1}{2}.$$

Но при $a = 0$ имеем

$$\lim_{x\to 0} \frac{\sqrt{ax}-x}{x-a} = \lim_{x\to 0} \left(-\frac{x}{\sqrt{ax}+x}\right) = \lim_{x\to 0} \left(-\frac{x}{x}\right) = -1.$$

Заметим, что при $a < 0$ существуют лишь односторонние бесконечные пределы: если $x \to a+0$, то предел равен $(+\infty)$, если $x \to a-0$, то предел равен $(-\infty)$.

Задача 5. Вычислить предел $\lim\limits_{x\to\infty} (\frac{x+1}{2x-1})^x$.

Решение

Ошибочное решение состоит в том, что не рассматриваются два разных случая: $x \to +\infty$ и $x \to -\infty$.

Студенты вычисляли предел следующим образом:

$$\lim_{x\to\infty} \left(\frac{x+1}{2x-1}\right)^x = \lim_{x\to\infty} \left(\frac{1}{2} + \frac{3}{2}\cdot\frac{1}{2x-1}\right)^x = \lim_{x\to\infty} \left(\frac{1}{2} + \frac{3}{2}\cdot 0\right)^2 = 0.$$

Верное решение состоит в том, чтобы учесть, что при достаточно больших $|x|$ имеет место неравенство $|\frac{x+1}{2x-1}| < 1$ и при $x \to +\infty$ $\lim\limits_{x\to+\infty}(\frac{x+1}{2x-1})^x = 0$, а при $x \to -\infty$ будем иметь $\lim\limits_{x\to-\infty}(\frac{x+1}{2x-1})^x = +\infty$.

Задача 6. Вычислить предел $\lim\limits_{x\to\infty}(\sqrt{x^2-2x-1} - \sqrt{x^2-7x+3})$

Решение

Ошибочное решение выглядит следующим образом:

$$\lim_{x\to\infty}(\sqrt{x^2-2x-1} - \sqrt{x^2-7x+3}) =$$

$$\lim_{x\to\infty}\frac{(\sqrt{x^2-2x-1} - \sqrt{x^2-7x+3})(\sqrt{x^2-2x-1} + \sqrt{x^2-7x+3})}{\sqrt{x^2-2x-1} + \sqrt{x^2-7x+3}} =$$

$$\lim_{x\to\infty}\frac{(x^2-2x-1)-(x^2-7x+3)}{\sqrt{x^2-2x-1} + \sqrt{x^2-7x+3}} = \lim_{x\to\infty}\frac{5x-4}{\sqrt{x^2-2x-1} + \sqrt{x^2-7x+3}} =$$

$$\lim_{x\to\infty}\frac{5-\frac{4}{x}}{\sqrt{1-\frac{2}{x}-\frac{1}{x^2}} + \sqrt{1-\frac{7}{x}+\frac{3}{x^2}}} = \frac{5-0}{\sqrt{1-0-0}+\sqrt{1-0+0}} = \frac{5}{1+1} = \frac{5}{2}.$$

Неверное решение состоит в том, что не рассмотрены отдельно два случая: $x \to -\infty$ и $x \to +\infty$.

Приведенное решение справедливо лишь при $x \to +\infty$. Если же $x \to -\infty$, то предел будет равен $(-\frac{5}{2})$.

Покажем верное решение этой задачи.

$$\lim_{x\to\pm\infty}(\sqrt{x^2-2x-1} - \sqrt{x^2-7x+3}) = \lim_{x\to\pm\infty}\frac{(x^2-2x-1)-(x^2-7x+3)}{\sqrt{x^2-2x-1} + \sqrt{x^2-7x+3}} =$$

$$\lim_{x\to\pm\infty}\frac{5x-4}{|x|(\sqrt{1-\frac{2}{x}-\frac{1}{x^2}} + \sqrt{1-\frac{7}{x}+\frac{3}{x^2}})} = \pm\frac{5}{2},$$

где знак плюс соответствует условию $x \to +\infty$, а знак минус — условию $x \to -\infty$.

Подобного рода ошибки мы наблюдаем и при решении такой задачи.

Задача 7. Вычислить $\lim_{x\to\infty} x(\sqrt{x^2+1} - x)$

Решение

Решение будет ошибочным, если мы его проведем следующим образом

$$\lim_{x\to\infty} x(\sqrt{x^2+1} - x) = \lim_{x\to\infty}\frac{x(\sqrt{x^2+1}-x)(\sqrt{x^2+1}+x)}{\sqrt{x^2+1}+x} =$$

$$= \lim_{x\to\infty}\frac{x(x^2+1-x^2)}{\sqrt{x^2+1}+x} = \lim_{x\to\infty}\frac{x}{\sqrt{x^2+1}+x} = \frac{1}{2}.$$

На самом же деле этот ответ может быть получен лишь тогда, когда мы x устремим к $(+\infty)$, если же $x \to -\infty$, то будем иметь

$$\lim_{x\to\infty} x(\sqrt{x^2+1}-x) = -\infty.$$

Сделаем вначале некоторые пояснения общего характера, прежде чем переходить к разбору **типичных ошибок, допускаемых при решении задач на нахождение производной функции.**

Пусть на некотором множестве задана функция и требуется найти для нее производную. Мы должны найти одно или несколько выражений, которые в совокупности давали бы результат, справедливый на той области определения данной функции, где она имеет производную (в большинстве случаев, встречающихся на практике, это будет вся область определения, за исключением, быть может, отдельных точек).

В качестве подтверждения сказанному приведем несколько примеров.

1)

$$(|x|)' = \begin{cases} 1, & x > 0; \\ -1, & x < 0. \end{cases}$$

Данная функция определена в интервале $(-\infty; +\infty)$, в то время как производная существует лишь для значений $x \neq 0$.

2) $(\ln x)' = \frac{1}{x}, x > 0$

Здесь необходимо указать ограничение, так как естественная область определения правой части шире области определения данной функции.

3) $(\mathrm{arctg}\frac{1+x}{1-x})' = \frac{1}{1+x^2}, x \neq 1.$

В этой задаче особенно важно указать ограничение. Если его не учитывать, то, зная, что $(\mathrm{arctg}\frac{1+x}{1-x})' = \frac{1}{1+x^2}$, мы бы сделали вывод: $\mathrm{arctg}\frac{1+x}{1-x} = \mathrm{arctg} x + C$, который оказался бы неверным.

Дело в том, что данная функция и ее производная существуют и непрерывны в двух промежутках: $(-\infty; 1)$ и $(1; +\infty)$. Для каждого из этих промежутков мы получили свое равенство:

$$\left(\mathrm{arctg}\,\frac{1+x}{1-x}\right)' = \begin{cases} \mathrm{arctg}\, x + C_1, & x < 1; \\ \mathrm{arctg}\, x + C_2, & x > 1. \end{cases}$$

При $x = 0$ имеем $\mathrm{arctg}\ 1 = \mathrm{arctg}\, 0 + C_1$, откуда $C_1 = \frac{\pi}{4}$. При $x = \sqrt{3}$ получим для левой части $\mathrm{arctg}\,\frac{1+\sqrt{3}}{1-\sqrt{3}} = \mathrm{arctg}1 + \mathrm{arctg}\sqrt{3} - \pi = -\frac{5}{12}\pi$, а для правой части $\mathrm{arctg}\sqrt{3} + C_2 = \frac{\pi}{3} + C_2$, откуда $C_2 = -\frac{3}{4}\pi$.

Данный пример показывает, что указание ограничения $x \neq 1$ дало возможность избежать ошибки.

В русле рассмотренных трех примеров небезынтересно рассмотреть вывод с помощью дифференцирования ряда тождеств связывающих между собой аркфункции.

4) Из равенства $(\arcsin x + \arccos x)' = \frac{1}{\sqrt{1-x^2}} - \frac{1}{\sqrt{1-x^2}} = 0, |x| < 1$, следует, что $\arcsin x + \arccos x = C$.

Подставив $x = 0$, получим $C = \frac{\pi}{2}$. Этот же результат получим и на концах сегмента [-1; 1]. Действительно

$$\arcsin(-1) + \arccos(-1) = -\frac{\pi}{2} + \pi = \frac{\pi}{2},$$

$$\arcsin 1 + \arccos 1 = \frac{\pi}{2} + 0 = \frac{\pi}{2}.$$

Следовательно, выведено следующее тождество:

$$\arcsin x + \arccos x = \frac{\pi}{2}, \quad |x| < 1.$$

5) Функция $f(x) = 2\mathrm{arctg}\sqrt{\frac{1-x}{1+x}}$ определена на множестве $-1 < x \le 1$, а ее производная $f'(x) = -\frac{1}{\sqrt{1-x^2}}$ на множестве $|x| < 1$. Следовательно, на множестве $|x| < 1$ справедливо тождество $2\mathrm{arctg}\sqrt{\frac{1-x}{1+x}} = \arccos x + C$, а так как при $x = 0$ $2\,\mathrm{arctg}\,1 = \frac{\pi}{2}$ и $\arccos 0 = \frac{\pi}{2}$, то $C = 0$. При $x = 1$ мы также имеем равенство $2\,\mathrm{arctg}\,0 = \arccos 1 = 0$.

Окончательно $2\mathrm{arctg}\sqrt{\frac{1-x}{1+x}} = \arccos x, -1 < x \le 1$.

Особенностью приведенных примеров является то, что производные существуют не на всей области определения данных функций. Поэтому для установления тождества необходима особая проверка его справедливости для тех точек, где производная не существует.

Перейдем к разбору типичных ошибок по теме «Дифференциальное исчисление».

Задача 8. Найти производную функции $y =\mathrm{arcsec} x$ при $|x| \le 1$.

Решение

Ошибочное решение выглядит следующим образом.

Воспользовавшись тождеством $\mathrm{arcsec} x = \arccos\frac{1}{x}, |x| \ge 1$. Тогда

$$(\mathrm{arcsec}\, x)' = \left(\arccos\frac{1}{x}\right)' = -\frac{1}{\sqrt{1-\frac{1}{x^2}}} \cdot (-\frac{1}{x^2}) = \frac{1}{x^2\sqrt{\frac{x^2-1}{x^2}}} =$$

$$= \frac{1}{\frac{x^2}{x}\sqrt{x^2-1}} = \frac{1}{x\sqrt{x^2-1}}$$

при $|x| > 1$.

Здесь, как и во многих других случаях, неверно использовано тождество $\sqrt{x^2} = |x|$. Следовало бы отдельно рассмотреть два случая $x > 1$ и $x < -1$, тогда:

а) при $-\infty < x < -1$ будем иметь

$$(\mathrm{arcsec}\, x)' = \frac{|x|}{x^2\sqrt{x^2-1}} = \frac{-x}{x^2\sqrt{x^2-1}} = \frac{-1}{x\sqrt{x^2-1}}.$$

б) при $1 < x < +\infty$ будем иметь

$$(\operatorname{arcsec} x)' = \frac{|x|}{x^2\sqrt{x^2-1}} = \frac{x}{x^2\sqrt{x^2-1}} = \frac{1}{x\sqrt{x^2-1}}.$$

Итак,

$$(\operatorname{arcsec} x)' = \begin{cases} \frac{1}{x\sqrt{x^2-1}}, & x > 1; \\ -\frac{1}{x\sqrt{x^2-1}}, & x < -1. \end{cases}$$

Покажем несколько иной способ нахождения производной указанной функции y=arcsec$x$ при $|x| \geq 1$.

Обратная функция $x = \sec y$ возрастает на каждом из промежутков $[0; \frac{\pi}{2})$ и $(0; \frac{\pi}{2}]$ и имеет конечную производную $(\sec y)' = \sec y \cdot \mathrm{tg} y$ в каждом из указанных промежутков.

Но так как при $y = 0$ и $y = \pi$ $(\sec y)' = 0$, то в этих точках не будет конечной производной для y=arcsec x.

Таким образом, производную y=arcsec$x$ следует искать в отдельности при $-\infty < x < -1$ и при $1 < x < +\infty$.

а) $1 < x < +\infty, 0 < y < \frac{\pi}{2}$.

$$(\operatorname{arcsec} x)'_x = \frac{1}{(\sec y)'_y} = \frac{1}{\sec y \cdot tgy} = \frac{1}{\sec y\sqrt{\sec^2 y - 1}} = \frac{1}{x\sqrt{x^2-1}}.$$

б) $-\infty < x < -1, \frac{\pi}{2} < y < \pi$.

$$(\operatorname{arcsec} x)'_x = \frac{1}{(\sec y)'_y} = \frac{1}{\sec y \cdot tgy} = \frac{1}{\sec y(-\sqrt{\sec^2 y - 1})} = -\frac{1}{x\sqrt{x^2-1}}.$$

В случае а) перед квадратным корнем взят знак плюс, так как $\mathrm{tg} y > 0$ при $0 < y < \frac{\pi}{2}$, а в случае б) взят знак минус, так как $\mathrm{tg} y < 0$ при $\frac{\pi}{2} < y < \pi$.

Использование тождества $\sqrt{x^2} = |x|$ требует и дифференцирование функций $y = \arcsin f(x)$ и $y = \arccos f(x)$. Покажем это.

Пусть нужно найти, например, производную функции $y = \arcsin \frac{a}{x}$.

Решение

$$y' = \frac{1}{\sqrt{1 - \frac{a^2}{x^2}}} \cdot (-\frac{a}{x^2}) = \frac{-a|x|}{x^2\sqrt{x^2-a^2}} = -\frac{a}{|x|\sqrt{x^2-a^2}}.$$

Задача 9. Найти производную функции

$$y = \arccos x + \arccos \frac{x + \sqrt{1-x^2}}{\sqrt{2}}.$$

Решение

Ошибочное решение состоит в том, что находятся производные каждого слагаемого и затем не проводится тождественное преобразование полученного выражения, а именно оно позволяет сделать верный вывод. Итак,

$$y' = \left(\arccos x + \arccos\frac{x+\sqrt{1-x^2}}{\sqrt{2}}\right)' = -\frac{1}{\sqrt{1-x^2}} - \frac{\sqrt{1-x^2}-x}{\overline{\overline{\sqrt{1-x^2}\cdot\sqrt{1-2x\sqrt{1-x^2}}}}}.$$

Произведем тождественное преобразование полученного выражения, для чего введем обозначение $\sqrt{1-2x\sqrt{1-x^2}} = N$. Это выражение преобразуем по формуле «сложного» квадратного корня:

$$\sqrt{A\pm\sqrt{B}} = \sqrt{\frac{A+\sqrt{A^2-B}}{2}} \pm \sqrt{\frac{A-\sqrt{A^2-B}}{2}}.$$

При этом мы должны рассматривать четыре случая:
а) $\frac{1}{\sqrt{2}} \le x \le 1$;

$$N = \sqrt{1-\sqrt{4x^2-4x^4}} = \sqrt{\frac{1+|1-2x^2|}{2}} - \sqrt{\frac{1-|1-2x^2|}{2}} = x - \sqrt{1-x^2}.$$

б) $0 \le x \le \frac{1}{\sqrt{2}}$;

$$N = \sqrt{1-\sqrt{4x^2-4x^4}} = \sqrt{\frac{1+|1-2x^2|}{2}} - \sqrt{\frac{1-|1-2x^2|}{2}} = \sqrt{1-x^2} - x.$$

в) $-\frac{1}{\sqrt{2}} \le x \le 0$;

$$N = \sqrt{1+\sqrt{4x^2-4x^4}} = \sqrt{\frac{1+|1-2x^2|}{2}} + \sqrt{\frac{1-|1-2x^2|}{2}} = \sqrt{1-x^2} - x.$$

г) $-1 \le x \le -\frac{1}{\sqrt{2}}$;

$$N = \sqrt{1+\sqrt{4x^2-4x^4}} = \sqrt{\frac{1+|1-2x^2|}{2}} + \sqrt{\frac{1-|1-2x^2|}{2}} = \sqrt{1-x^2} - x.$$

Таким образом:

$$N = \begin{cases} x-\sqrt{1-x^2}, & \frac{1}{\sqrt{2}} \le x \le 1; \\ \sqrt{1-x^2}-x, & -1 \le x \le \frac{1}{\sqrt{2}}. \end{cases}$$

Для производной заданной функции получаем:

$$y' = \begin{cases} -\frac{2}{\sqrt{1-x^2}}, & -1 < x < \frac{1}{\sqrt{2}}; \\ 0, & \frac{1}{\sqrt{2}} < x < 1. \end{cases}$$

Заметим, что при $x = \frac{1}{\sqrt{2}}$ имеем угловую точку, в которой производная не существует.

Задача 10. Найти кратчайшее расстояние от точки $A(a;0)$ до окружности $x^2+y^2=r^2$.

Решение

Выразим квадрат расстояния от точки A до искомой точки $M(x;y)$, лежащей на заданной окружности:

$$d^2 = (x-a)^2 + y^2 = x^2 + y^2 - 2ax + a^2.$$

Так как точка M принадлежит окружности, то вместо $x^2 + y^2$ можно подставить r^2. Получаем $d^2 = r^2 - 2ax + a^2 = f(x)$. Искомая точка должна находиться среди корней уравнения $f'(x) = 0$, однако производная данной функции $f'(x) = -2a$ ни при каком значении x в нуль не обращается, если $a \neq 0$, и всегда равна нулю, если $a = 0$.

При $a = 0$, то есть при условии, что точка A есть центр заданной окружности, имеем $d = r$, причем это есть расстояние до любой точки окружности.

При $a \neq 0$ из условия $f'(x) = -2a \neq 0$ следует, что кратчайшее расстояние от точки до окружности не существует, а это противоречит фактам, известным из элементарной геометрии.

Для объяснения этой ошибки отметим, что рассматриваемая функция $f(x) = r^2 - 2ax + a^2$, выражающая квадрат расстояния от данной точки $A(a;0)$ до произвольной точки окружности, определена на сегменте $[-r;r]$. Так как эта функция непрерывна на данном сегменте, то согласно известной теореме Вейерштрасса она достигает на нем своего наибольшего и наименьшего значения. Если $a \neq 0$, то $f'(x)$ также отличается от нуля, то есть функция монотонна на всем сегменте, и, следовательно, наибольшее и наименьшее значения достигаются на концах сегмента. При $a > 0$ наименьшее значение будет при $x = r$:

$$d_{\min} = \sqrt{f(r)} = \sqrt{r^2 - 2ar + a^2} = |r-a|,$$

при $a < 0$ наименьшее значение будет при $x \to -r$:

$$d_{\min} = \sqrt{f(-r)} = \sqrt{r^2 + 2ar + a^2} = |r+a|.$$

Этот пример показывает, что при решении задач на отыскание наибольшего и наименьшего значения необходимо всегда указывать область определения рассматриваемой функции.

Мы рассмотрели наиболее характерные ошибки студентов, хотя у каждого преподавателя есть свой набор таких и им подобных ошибок.

Математические структуры и моделирование
2001, вып. 8, с. 167–173

УДК 51:372.851

ФОРМИРОВАНИЕ КУЛЬТУРЫ МАТЕМАТИЧЕСКОГО МЫШЛЕНИЯ СРЕДСТВАМИ ИСТОРИИ МАТЕМАТИКИ

Т.Н. Алешкова

In this article the concept of mathematical thinking's culture is determined, and methods of its forming by means of history of mathematics are offered.

Введение

В настоящее время в образовании отдается приоритет воспитательным целям, относящимся к интеллектуальной деятельности и формированию характера личности. Применительно к математике эти цели сводятся к воспитанию логического мышления (способности рассуждать, критиковать, мыслить дедуктивно, систематизировать, обобщать) и работе над его качествами: рационально мыслить, четко, лаконично выражать свои мысли; воспитывать способность сосредоточиться, работать упорядоченно и настойчиво. Поэтому перед учителем на первый план выдвигается задача воспитания мышления и культуры математического мышления.

1. Культура математического мышления

Понятие «культура математического мышления» встречается в педагогических исследованиях и не определяется в работах психологов.

Вопросам развития математической культуры посвящено диссертационное исследование Д.Икрамова. Под математической культурой он понимает «систему математических знаний, умений и навыков, органически входящих в фонд общей культуры учащихся, и свободное оперирование ими в практической деятельности» [1, с. 107]. Согласно его исследованию школьник, обладающий культурой мышления:

а) аргументировано рассуждает, «умеет определять, сопоставлять, классифицировать, анализировать, систематизировать, обобщать, устанавливать причинно-следственные связи и т.п.» [1, с. 107];

б) точно, ясно, лаконично, выразительно выражает свои мысли, владеет математическим языком.

Таким образом, Д.Икрамов связывает культуру мышления с развитием логического мышления при обучении математике.

E-mail: aleshkova@math.omsu.omskreg.ru
Омский государственный университет

Вопросам воспитания культуры мышления уделял достаточное внимание А.Я.Хинчин. В культуре мысли он выделяет *«правильность мышления» и «стиль мышления»*.

Считая основным признаком *правильности мышления* полноценность аргументации, А.Я.Хинчин говорит: «В математике нет и не может быть «наполовину доказанных» и «почти доказанных» утверждений: либо полноценность аргументации такова, что никакие споры о правильности доказываемого утверждения более невозможны, либо аргументация полностью отсутствует». Отмечая важность воспитания школьников логической полноценности аргументации, А.Я.Хинчин говорит о том, что учащимся следует разъяснять, что обобщения, сделанные на основе неполной индукции, или аналогии, приводят только к гипотезам и требуют дальнейшего математического обоснования или опровержения.

Полноценность аргументации определяет и *стиль мышления*. А.Я.Хинчин выделяет 4 его признака:

- доведенное до предела доминирование логической схемы рассуждения. Эта черта в максимальной степени позволяет следить за правильностью течения мысли;
- лаконизм мышления: предельная скупость, суровая строгость мысли и ее изложения;
- четкая расчлененность хода рассуждения: «если, например, при доказательстве какого-либо предложения рассматривается несколько случаев, а последние распадаются еще на подслучаи, то в каждый момент рассуждения математик должен отчетливо помнить, в каком случае и подслучае его мысль обретается и какие случаи и подслучаи ему еще остается рассмотреть»;
- скрупулезная точность символики [3, с. 141-144].

Таким образом, культура мысли по А.Я.Хинчину является формулированием свойств мышления применительно к математическому мышлению.

Вопросам математического мышления и воспитания культуры математического мышления уделяет достаточное внимание Л.М.Фридман [2].

Он определяет математическое мышления как «предельно абстрактное теоретическое мышление, объекты которого лишены всякой вещественности и могут интерпретироваться самым произвольным образом, лишь бы при этом сохранялись заданные между ними отношения» [2, с. 165] и считает его составной частью общей культуры мышления.

Согласно Л.М.Фридману культура мышления обладает рядом признаков, таких как *«разумность, логичность, дисциплинированность»* [2, с. 167].

Под *разумностью* Л.М.Фридман понимает высшую ступень мышления, следующую за рассудком. Сравнивая их, он пишет: «Если рассудочное мышление осуществляется без изменения наличной ситуации — объекта мышления, то разумное мышление — это «способность находить причины и сущность явлений, рассматривать их всесторонне, вскрывать единство противоположностей».

Логичность культуры мышления, согласно Л.М.Фридману, предусматривает рассуждения и умозаключения в соответствии с законами логики.

Наконец, культурное мышление предполагает использование разных мыслительных операций в определенной системе, в полном соответствии с характером решаемой задачи. Между тем для учащихся, у которых не воспитана дисциплина мышления, характерна крайняя беспорядочность, хаотичность мышления. *Дисциплина мышления* предполагает, во-первых, анализ объекта мысли, во-вторых, планирование на основе этого анализа своей мыслительной деятельности и, в-третьих, пошаговый самоконтроль и самооценку выполненной деятельности с целью установления соответствия намеченному плану и его корректировки при необходимости [2, с. 168].

Так как перечисленные признаки культуры мышления включают в себя глубину мысли, гибкость и самостоятельность, критичность мышления, то можно сделать вывод, что Л.М.Фридман отождествляет свойства мышления с культурой мышления.

Вопросы исследования компонентов культуры мышления мы находим в работе методиста Т.А.Ивановой [1].

Обобщая выше изложенные точки зрения по этой проблеме, она выделяет компоненты культуры мышления и связывает их с некоторыми «составляющими» математической деятельности. Поэтому, на наш взгляд, корректнее их назвать компонентами культуры математического мышления.

Итак, по Т.А.Ивановой *культура математического мышления* включает в себя такие компоненты, как:

1. Осознание предмета математики, ее метода, ее ведущих понятий и осмысленное оперирование ими как при изучении математики, так и в ее приложениях и в практической деятельности;
2. Владение логической составляющей математической деятельности:
   - понимание логической структуры понятия (род, видовые понятия, их конъюнктивная или дизъюнктивная связь, наличие и смысл кванторов, умение формулировать отрицание понятия);
   - умение оперировать определением понятия: подводить под понятия, выводить следствия;
   - умение сравнивать объекты по указанному признаку, выделять существенные основания для их сравнения;
   - умение проводить классификацию понятий по заданному и самостоятельно найденному основанию;
   - понимание логической структуры теоремы, умение формулировать обратное, противоположное, противоположное обратному утверждения и понимание логической связи между этими четырьмя предложениями;
   - понимание сущности доказательства, полноценности аргументации;
   - владение дедуктивными методами доказательств и опровержений: синтетическим, аналитическим, от противного, методом исчерпывающих проб, полной индукции, контрапозиции, методом математической индукции;
3. Владение эвристической составляющей математической деятельности:
   - умение выявлять закономерности и устанавливать аналогии;

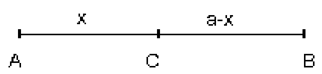


Рис. 1.

- умение выдвигать гипотезы на основе аналогии, неполной индукции, обобщения, конкретизации, пространственного воображения, интуиции как для постановки проблем, так и для их решения;

4. Умение отличать достоверные выводы от правдоподобных, вероятностных;
5. Владение алгоритмической составляющей математической деятельности:
   - понимание сущности алгоритма;
   - умение пользоваться готовыми алгоритмами;
   - умение создавать алгоритм какого-либо действия;
6. Владение математическим языком (математической терминологией, символикой), умение четко, последовательно, лаконично, логично выражать свои мысли как устно, так и письменно [1].

Таким образом, Т.А.Иванова глубже, чем Л.М.Фридман и А.Я.Хинчин, конкретизирует свойства мышления, учитывая специфику математической деятельности, формулирует их для математического мышления.

2. Роль истории математики в формировании культуры математического мышления

Педагогические пути формирования культуры математического мышления предполагают различную методическую работу. Покажем, как история математики влияет на воспитание потребности полно аргументировать свои высказывания и влияет на осознанное оперирование понятиями. Предложим содержание урока по теме «Экстремумы функций» с использованием исторического материала [4].

После знакомства учащихся с определением точек максимума и минимума дальнейшее изучение необходимого и достаточного признаков экстремума можно мотивировать, сообщая историю возникновения данной проблемы. В XIV в. епископ и профессор Пражского университета Никола Оресм (ок. 1323-1382), изображая графически зависимость интенсивности физических явлений от времени, заметил, что изменение вблизи точек экстремума самое медленное. Но это наблюдение не получило развития и тем более применения. Еще через два с половиной столетия Кеплер, решая задачи на экстремум, указал, что «по обе стороны от места наибольшего значения убывание вначале нечувствительно». Но и он не использовал замеченного им свойства и продолжал идти путем, проторенным древними греками.

Новый метод отыскания экстремумов первым разработал Ферма. Он не публиковал научных результатов, а, как правило, сообщал их в письмах своим

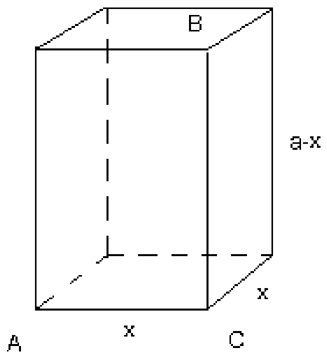


Рис. 2.

корреспондентам. В 1638 г. он написал Декарту, что решил задачу нахождения максимумов и минимумов величин; под величинами подразумевались многочлены. Разберем метод Ферма на конкретной задаче: «Разбить отрезок AB (рис.1) точкой C так, чтобы тело, построенное на квадрате AC и на отрезке CB, было наибольшим». Другими словами, разбить отрезок AB точкой C так, чтобы прямоугольный параллелепипед, основанием которого служит квадрат со стороной AC, а высотой — отрезок CB (рис. 2), имел максимальный объем. Найдем точку максимума, следуя Ферма, но применяя привычные обозначения.

Пусть C — искомая точка (рис.1). Обозначим $AB = a$, $AC = x$, тогда $CB = a - x$. Объем параллелепипеда $V(x) = x^2(a - x)$. Дадим приращение h точке C; объем нового параллелепипеда $V(x + h) = (x + h)^2(a - x - h)$. Поскольку в точке x объем достигает наибольшего значения, выражения $V(x)$ и $V(x + h)$ приближенно равны. К этой идее должен был прийти еще раньше Кеплер. Ведь у него возникла мысль о том, что в малой окрестности точки экстремума изменение функции незаметно. Но он не развил свою мысль. Ферма же пишет $V(x) = V(x+h)$ и называет это равенство вымышленным. Приведя в нем подобные члены и поделив его на h, получаем $2x(a-x)-x^2+h(a-3x-h) = 0$. Отбросив теперь члены, содержащие h, найдем «истинное» равенство (по терминологии Ферма). Откуда $x = \frac{2a}{3}$. Задача решена.

Посмотрим теперь, как будут выглядеть рассуждения Ферма применительно к произвольному многочлену $P(x)$. Дадим аргументу x малое приращение h и найдем $P(x+h)$. Если x — точка экстремума, то $P(x) = P(x+h)$ - «вымышленное» равенство, или $P(x)-P(x+h) = 0$. Приведем подобные члены и, поделив на h, придем к равенству $\frac{P(x+h)-P(x)}{h} = 0$. Отбросив в нем члены, содержащие h, получим уравнение, из которого определяется точка экстремума. Для многочлена последняя операция равносильна предельному переходу $\lim\limits_{h\to 0} \frac{P(x+h)-P(x)}{h} = 0$, т.е. полученное в результате уравнение для точки экстремума имеет вид $P'(x) = 0$.

Сейчас аналогичное утверждение для любой функции, дифференцируемой в точке x, называют теоремой Ферма. Ферма, конечно, к пределу не переходил. У него h было хоть и сколько угодно малым, но постоянным. Доказывал он свою теорему алгебраическим способом (в другом письме от 1643 г.).

Если рассмотреть произвольную функцию $y = f(x)$, то левая часть равен-

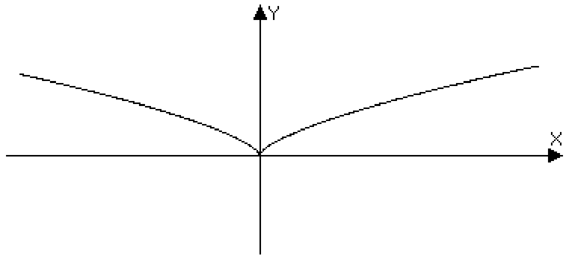


Рис. 3.

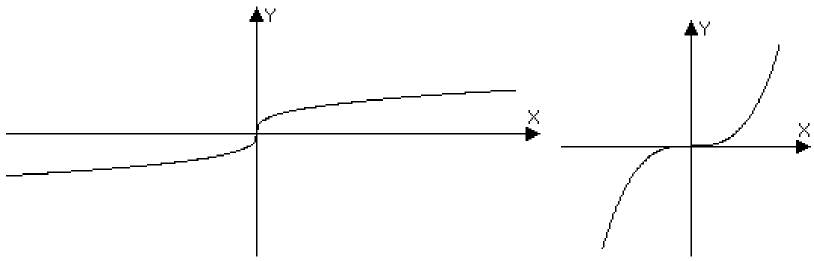


Рис. 4.

ства $\frac{f(x+h)-f(x)}{h} = 0$ является многочленом от h. Например, для $f(x) = \sqrt{x}$ имеем $\frac{f(x+h)-f(x)}{h} = \frac{\sqrt{x+h}-\sqrt{x}}{h} = \frac{1}{\sqrt{x+h}+\sqrt{x}} = 0$ при $h \neq 0$. Здесь выделить слагаемые, не содержащие h, не удается. Но если положить $h = 0$, тоже придем к условию Ферма $\frac{1}{2\sqrt{x}} = 0$. Так и поступали последующие аналитики. Теорема Ферма позволяет с помощью производной найти точки, «подозрительные на экстремум». А если производной нет?

Для многочленов такой вопрос излишен: все многочлены имеют производную в каждой точке числовой оси. Но у других функций вполне могут быть точки, в которых производной нет, а экстремум есть. Возьмем функцию $f(x) = \sqrt[3]{x^2}$. Она имеет минимум в начале координат (рис.3), в то время как ее производная $f(x) = \frac{2}{3\sqrt[3]{x}}$ в нуле не определена. Случаями, когда производная равна нулю или не существует, исчерпываются все возможные точки, «подозрительные на экстремум», — их также называют критическими.

Далее, обобщая выше сказанное, даем строгое определение критической точки и обсуждаем со школьниками, почему данный признак является необходимым. В этих точках функция не всегда имеет экстремум. Например, функция $f(x) = x^3$ имеет в начале координат производную, равную нулю, а функция $f(x) = \sqrt[3]{x}$ не имеет производной в той же точке. Тем не менее ни у той, ни у другой функции в точке $x = 0$ экстремума нет (рис.4).

Таким образом, условие: производная функции в точке x равна нулю или не существует — является необходимым условием наличия в точке x=0 экстремума, но недостаточным. Затем формулируем достаточный признак экстремума.

Такой подход в изучении данной темы не только делает содержание интересным, но и приводит к осмысленному оперированию понятием экстремума функции. Отметим, что в школьном учебнике нет доказательства теоремы, а показывается наглядный геометрический смысл. Подобное же изложение представляет собой попытку дать «строгое» обоснование материала, и в этом его преимущество в формировании потребности полной аргументации математических утверждений.

Таким образом, обучение математике с помощью элементов историзма ориентировано на формирование культуры математического мышления.

Литература

1. Иванова Т.А. *Гуманитаризация общего математического образования.* – Н.Новгород. 1998. 206с.
2. Фридман Л.М. *Теоретические основы методики обучения математике.* – М.: Московский психолого-социальный институт: Флинта, 1998. 224с.
3. Хинчин А.Я. *Педагогические статьи.* М., 1963.
4. Шибасов Л.П., Шибасова З.Ф. *За страницами учебника математики.* М.: Просвещение, 1997. 269с.

Математические структуры и моделирование
2001, вып. 8, с. 174–177

УДК 51:378

СВЯЗЬ ДОПОЛНИТЕЛЬНОГО И ОСНОВНОГО КУРСОВ МАТЕМАТИКИ

Н.Г. Русанова

In the article additional education and standart course of mathematics as basic making common education, their connection and succession, are considered. Moreover the problem of break between preparation of learning high schools and the university requirements is studied.

1. Место дополнительного образования в системе общего образования

Для понимания места дополнительного образования детей в современной системе общего образования необходимо выяснить, какие проблемы общего образования могут быть решены за счет использования возможностей дополнительного образования.

Общее образование — это совокупность знаний, умений и навыков, способов творческой деятельности, ценностных ориентиров, необходимых каждому человеку независимо от его профессии.

Можно выделить три основные функции общего образования. Первой из них является функция социализации личности, под которой понимается приобщение индивида к культуре, ее ценностям и нормам, что обеспечивает возможность выполнения человеком своих социальных функций — гражданина, работника, родителя и так далее. Вторая функция – подготовка к продолжению образования, к получению профессионального образования. Третья функция – индивидуализация личности, то есть выявление специфики их интересов, интеллекта, потребностей и способностей. Реализация первой функции выражается в готовности и способности личности к решению социально значимых проблем, второй функции – возможности получить профессиональное образование, третьей функции — в развитии способности к самопознанию, самооценке и самоопределению.

Система дополнительного образования обладает большими возможностями для совершенствования общего образования, его гуманитаризации. Все виды детских объединений, независимо от их профиля, способствуют развитию у детей способности к самопознанию и самоопределению. Они дают опыт общения со специалистами в различных видах практической деятельности. В условиях

E-mail: rusanova@math.omsu.omskreg.ru
Омский государственный университет

учреждения дополнительного образования педагоги имеют возможность быть менее «функциональными» по сравнению со школьными учителями.

Дополнительное образование позволяет полнее использовать потенциал школьного образования за счет углубления и расширения школьных знаний. Оно компенсирует неизбежную ограниченность школьного образования путем реализации досуговых и индивидуальных образовательных программ, дает возможность каждому ребенку удовлетворить свои индивидуальные познавательные и творческие запросы.

Дополнительное образование не только существенно расширяет знания о творческих возможностях и творческом потенциале учащихся; оно обеспечивает возможность успеха в избранной сфере деятельности и тем самым способствует развитию таких качеств личности, которые важны для успеха в любой сфере деятельности; оно создает возможность формирования круга общения на основе общих интересов.

2. Преемственность дополнительного и основного курсов математики

Школьное и дополнительное образование являются составными частями общего образования, и поэтому их взаимодействие – одно из условий решений общих задач образования.

Основная задача обучения математике в школе – «обеспечить прочное и сознательное овладение учащимися системой математических знаний и умений, необходимых в повседневной жизни и трудовой деятельности каждому человеку современного общества, достаточных для продолжения образования» [3] – проявляется в ориентации на достижение всеми учащимися определенного уровня математической подготовки, зафиксированной в программе по математике. Но задачи обучения математике не исчерпываются достижением только этого результата. Как указывается в программе дополнительного образования, «нельзя ограничивать обучение всех учащихся уровнем обязательных требований: важно стремиться к возможно более полному раскрытию их математических способностей и дарований. В этом смысле уровень обязательной математической подготовки определяет нижнюю ее границу, на базе которой должно осуществляться дальнейшее математическое развитие школьников» [4].

Проблему повышения математической подготовки школьников, в особенности проявляющих склонности к математике, школа решает не всегда удовлетворительно. А тенденции к сокращению и упрощению математического курса, которые наблюдаются в последние годы, еще более усугубляют данную проблему. Результатом чего является снижение уровня подготовки выпускников школ, в то время как уровень, предъявляемый высшими учебными заведениями к будущим студентам, не меняется. То есть возникает разрыв между реальной подготовкой учащихся средних школ и требованиями вузов. Эта действительность становится в том числе и психологической проблемой уже для всей семьи. Не секрет, что родители поступающего принимают близко к сердцу удачи и неудачи своих детей, нервничает и сам ребенок, что тоже в определенной мере ухудшает результаты экзаменов.

Решить данную проблему помогает распространение различных форм повышения математической подготовки: школ и классов с углубленным изучением математики; факультативных и кружковых занятий по математике; подготовительных курсов при вузах; школ юных математиков; частного репетиторства и так далее.

Особое место среди этих форм занимают очно-заочные школы при университетах (в Москве, Санкт-Петербурге, Новосибирске, Омске, Челябинске и многих других), имеющие большой опыт и давние традиции в работе со школьниками по математике. Многие преподаватели университетов вносят свой вклад в эту работу: читают лекции для школьников, руководят кружками, организуют олимпиады, пишут книги для школьников.

Одной из причин, побуждающих преподавателей вузов работать со школьниками, является соображение о том, что для подготовки математиков существенны не только «количество и качество» работы с ними, но и время начала работы, возраст, в котором человек впервые знакомится и заинтересовывается своей будущей профессией.

Существование оптимального (и довольно раннего) возраста для подготовки к целому ряду специальностей общепризнано: подготовка спортсменов, музыкантов и так далее. Так же и в математике, если школьник только в 10 классе заинтересовался математикой, то ему уже крайне трудно догнать тех сверстников, которые находили время для углубленного изучения математики в 7, 8 и 9 классах. Опыт работы со школьниками, данные психологии и педагогики, относящиеся к развитию интересов и возрастных особенностей школьников, позволяют сделать вывод, что склонности, требующиеся для успешных систематических занятий по углубленному изучению математики, определяются у школьников на уровне 7-9 классов. Другие особенности этого возраста: определенный уровень развития мышления, устойчивость внимания, способность к систематическим занятиям – и все это вместе со значительным запасом знаний дает возможность развития этих склонностей.

Более 25 лет при Омском государственном университете работает Профориентационная школа математического факультета (ПШМФ), она пользуется большой популярностью и имеет многочисленные филиалы в Омской области. Методические материалы, разработанные для разных классов (с 5 по 11 кл.), содержат единую тематику. Основные темы повторяются от класса к классу, однако повышается уровень сложности рассматриваемых по данной теме задач.

Для одиннадцатиклассников темы, задачи, уровень сложности задач ориентированы на вступительные экзамены по математике в ОмГУ, к ним же приближены материалы десятого класса.

Основная цель ПШМФ (и других подобных школ при университетах) — не просто подготовка к вступительным экзаменам, а подготовка к продолжению образования, к дальнейшему изучению математики в школе и в вузе, к самообразованию, к широкому и свободному применению математики в дальнейшей профессиональной деятельности. Такие школы способствуют повышению научного уровня знаний школьников по математике, развитию интереса школьников к изучаемой науке, развитию их способностей, а также осуществлению преемственности с основным курсом математики.

Для полноценной подготовки к продолжению образования должны быть решены следующие задачи:

- обобщение и систематизация основного материала общеобразовательного курса математики, на котором строится дальнейшее обучение;
- достижение достаточного уровня умений и технических навыков;
- ознакомление с различными подходами к понятиям и идеям школьного курса математики, получение представлений и развитие понятий и методов, изучаемых в основном курсе;
- дополнительное раскрытие прикладного содержания некоторых тем обязательного курса, подготовка к восприятию курса математики при дальнейшем обучении в вузе, к условиям обучения в вузе;
- развитие устойчивого интереса к математике, творческого мышления.

Мы должны двигаться навстречу интересам наших учащихся, среди которых, как правило, нет случайных людей. Наших детей надо учить мыслить широко, учить совершать открытия, а не замыкать в рамки школьной программы.

Дополнительное образование является необходимым компонентом современного общего образования. Его отсутствие вряд ли может быть компенсировано использованием других источников общего образования. Есть основания полагать, что для некоторых учащихся дополнительное образование на некоторых этапах процесса общего образования является не только необходимым, но и ведущим компонентом.

Литература

1. *Внешкольные учреждения:* Пособие для работников внешкольных учреждений / Сост. Л.И. Филатова, В.С. Муратова; под. ред. Л.К. Балясной. – М.: Просвещение, 1978.
2. *После школьного звонка: О работе внешкольных учреждений:* Сборник. / Сост. Г. Чубарова. М.: Молодая гвардия, 1978.
3. *Программа по математике для средней общеобразовательной школы (5-10 кл.)* // Математика в школе. 1985. №6.
4. *Программы факультативных занятий по математике* // Математика в школе. 1980. №4.

Математические структуры и моделирование

Вып. 8

Сборник научных трудов

Редактор Е.В. Брусницина

Подготовлена к печати ООО «Издательство Наследие. Диалог-Сибирь»
Лицензия ЛР № 071680 от 04.06.98.
Подписано в печать 04.12.2001.
ОП. Формат 60×84 1/8. Печ.л. 19,5. Уч.-изд.л. 13,3.
Тираж 140 экз.

Полиграфический центр КАН
644050, г. Омск, пр. Мира 32, ком.11, тел. (381-2) 65-47-31
Лицензия ПЛД № 58-47 от 21.04.97 г.