

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

МАТЕМАТИЧЕСКИЕ
СТРУКТУРЫ
И
МОДЕЛИРОВАНИЕ

Выпуск 9

УДК 007, 378, 510, 514, 515, 517, 519, 530, 531, 536, 537, 539, 541, 611, 612, 616, 621,
681

Математические структуры и моделирование: Сб. научн. тр.
/ Под ред. А.К. Гуца. – Омск: Омск. гос. ун-т, 2002. – Вып. 9. – 184 с.

ISBN 5 – 8239 – 0071 – 6

Сборник составлен из статей преподавателей и аспирантов математического факультета, отделения компьютерных наук Омского государственного университета и других вузов.

Для научных работников, аспирантов и студентов старших курсов.

Редакционная коллегия

В.А. Топоногов

д-р физ.-мат. наук, проф.

Институт математики СО РАН (г. Новосибирск)

А.К. Гуц

д-р физ.-мат. наук, проф.

Н.Ф. Богаченко

к-т физ.-мат. наук

Д.Н. Лавров

к-т тех. наук

Художественное оформление

В.В. Коробицын

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

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 "Zentralblatt 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

- А.К. Гуц. *Симметричное управление, не выводящее динамическую систему за пределы конуса* 5
- И.И. Еремин, И.М. Соколинская. *Фейеровские итерационные процессы для несобственных задач линейного программирования* 10
- В.А. Кыров. *Двумерные гильбертовы пространства* 27
- А.А. Звягинцев. *Дифференциальные уравнения в Синтетической дифференциальной геометрии Ловера-Кока* 38
- А.А. Звягинцев. *Связность Леви-Чевита в синтетической дифференциальной геометрии Кока-Ловера* 45
- В.А. Маренко. *Использование нечетких отношений для формализации знаний* 48
- Б.К. Нартов. *Об одном методе оптимизации начальных условий в управлении динамическими системами* 51
- Р.К. Романовский, И.Д. Макарова, С.Е. Макаров. *Условия существования стационарного режима в реакторе с кипящим слоем катализатора* 54
- А.В. Пролубников. *Оценка приближенного решения алгоритмом Хватала задачи о покрытии* 58
- А.В. Пролубников, Р.Т. Файзуллин. *Эвристический алгоритм дешифровки шифра двойной перестановки* 62
- Р.Р. Файзуллин. *Об ультраметрических тетраэдрах и треугольниках* 70

Моделирование

- Н.Ф. Богаченко, Р.Т. Файзуллин. *Механические аналогии в задаче Штейнера* 80
- О.Т. Данилова. *Алгоритм расчета технологических параметров нагрева порошковых образцов в высокочастотном электромагнитном поле* 88

продолжение на след. странице



Л.В. Недорезов, Е.В. Волкова. *Об одной модели динамики численности изолированной популяции* 92

Л.В. Недорезов, Ю.В. Утюпин. *Об одной непрерывно-дискретной модели динамики численности двуполой популяции* 101

В.М. Семенюк, И.Н. Путалова, А.В. Артюхов, А.В. Сырцова, А.К. Гуц. *Применение метода конечных элементов в стоматологии (обзор литературных источников)* 113

В.М. Семенюк, А.К. Гуц, Н.И. Панова. *Компьютерные эксперименты с моделью итфитовой конструкции с большим диаметром итфита* 124

Р.И. Нужин, В.А. Шапцев. *Инструмент моделирования процесса дозирования активности компьютера* 129

Теоретическая физика

С.В. Белим. *Зависимость критического поведения однородных сжимаемых систем от размерности параметра порядка* 135

Е.В. Палешева. *Вклад спинорных духов в интерференцию квантовых частиц* 142

Информационные технологии

Т.Б. Бигильдеева, В.В. Таркаев. *О пакете программ недифференцируемой оптимизации* 158

И.А. Земсков. *Сбор информации о доступных ресурсах Интернет* 168

Преподавание математики и физики

М.П. Ланкина. *Задачи с элементами исследования при обучении физике* 179

СИММЕТРИЧНОЕ УПРАВЛЕНИЕ, НЕ ВЫВОДЯЩЕЕ ДИНАМИЧЕСКУЮ СИСТЕМУ ЗА ПРЕДЕЛЫ КОНУСА

А.К. Гуц

We study a control dynamic system in $\mathbb{R}^3$ that is symmetric with respect to Lie group G which acts simple transitively and affinely in $\mathbb{R}^3$ and linear with respect to control parameters. The purpose is discription of all convex cones in which lies the system trajectory.

Введение

Пусть дана задача оптимального управления

$$\int_0^T s(x, u) dt \rightarrow \inf$$

при условии, что

$$\begin{cases} \dot{x} = f(x, u), \\ x(0) = a, \\ x(T) = b \\ x \in \mathbb{R}^n, \quad u \in U \subset \mathbb{R}^r. \end{cases} \quad (1)$$

В случае, когда функции s, f линейны относительно u , т.е. $s(x, u) = S(x) + Au + a$, $f(x, u) = F(x) + Bu + b$, и множество U является выпуклым многогранником (или выпуклым компактом), то эта задача сводится к следующей [1–3]:

$$J[x(\cdot)] = \int_{x(\cdot)} \omega \rightarrow \inf \quad (2)$$

при условии, что

$$\dot{x} \in K_{x(t)}, \quad t \in [0, 1], \quad (3)$$

$$x(0) = a, \quad x(1) = b, \quad (4)$$

где ω – дифференциальная 1-форма на n -мерном гладком многообразии $X \approx \mathbb{R}^n$, K_x выпуклый конус в каждой точке $x \in X$ и лежащий в касательном пространстве $T_x X$.

Таким образом, управление $u \in U$ в виде (1) сводится к управлению в форме задания семейства выпуклых конусов $\mathcal{K} = \{K_x \subset T_x X : x \in X\}$.

© 2002 А.К. Гуц

E-mail: guts@univer.omsk.su

Омский государственный университет

Работа поддержана грантом РФФИ (проект № 01-01-00303).

1. Симметричное управление

Группа Ли G действует на многообразии X , если каждому $g \in G$ соответствует диффеоморфизм $\alpha(g) : X \rightarrow X$ такой, что произведению gh отвечает композиция $\alpha(g) \circ \alpha(h)$ диффеоморфизмов, а единице $e \in G$ – тождественное отображение $id_X : X \rightarrow X$. Иначе говоря, действие G на X – это гомоморфизм $\alpha : G \rightarrow Diff(X)$.

Определение 1. Управление динамической системой (2)–(4) называется *симметричным* относительно действия группы G , если для любого $g \in G$

$$d\alpha(g)_x[K_x] = K_{[\alpha(g)](x)} \quad (5)$$

и

$$\alpha^*(g)[\omega_{[\alpha(g)](x)}] = \omega_x. \quad (6)$$

Здесь $d\alpha(g)_x$ – дифференциал диффеоморфизма $\alpha(g)$ в точке $x \in X$, а $\alpha^*(g)_x : T_{[\alpha(g)](x)}^*X \rightarrow T_x^*X$ соответствующий кодифференциал.

2. Упорядоченное симметричное управление

Пусть X и семейство подмножеств $\mathcal{P} = \{P_x \subset X : x \in X\}$ задает порядок в X , т.е. выполняются условия:

- 1) $x \in P_x$;
- 2) если $y \in P_x$, то $P_y \subset P_x$;
- 3) если $y \neq x$, то $P_y \not\subset P_x$.

Мы будем предполагать далее, что $X = \mathbb{R}^n$ и группа Ли G действует просто транзитивно на X . Зафиксируем точку $a \in X$. Тогда имеем диффеоморфизм

$$\varphi : G \cong \mathbb{R}^n, \quad (7)$$

$$\varphi(g) = [\alpha(g)](a), \quad \varphi(e) = [\alpha(e)](a) = a.$$

Определение 2. Порядок $\mathcal{P}$ инвариантен относительно действия группы G (G -инвариантный порядок), если для любой точки $x \in X$ и любого $g \in G$

$$\alpha(g)[P_x] = P_{[\alpha(g)](x)}. \quad (8)$$

Нетрудно убедиться, что если $\mathcal{P}$ – G -инвариантный порядок на X , то $S = \varphi^{-1}(P_a)$ – подполугруппа группы G .

Касательный объект к S – это множество вида $L(S) = \{\xi \in \mathfrak{g} : \xi = \lim_{n \rightarrow \infty} n\xi_n, \exp \xi_n \in S\}$, где $\mathfrak{g}$ алгебра Ли группы Ли G .

Контингенция множества P_x в точке x – это совокупность векторов, касательных в точке x к гладким кривым, исходящим из точки x и лежащим в P_x . Для контингенции используем обозначение: $cont(\mathcal{P}, x)$. Известно, что $K_x = cont(\mathcal{P}, x)$ – замкнутый выпуклый конус, лежащий в T_xX .

Ясно,

$$d\varphi_e[L(S)] = K_a \quad \text{и} \quad d\varphi_g[d(l_g)_e[L(S)]] = d\alpha(g)_a[K_a] = K_x, \quad (9)$$

$$x = \alpha(g)(a),$$

где $l_g : G \rightarrow G$ левый сдвиг на g .

Предложение 1. Пусть подполугруппа S порождает G . Тогда $L(S) = \{\xi \in \mathfrak{g} : \exp(\mathbb{R}^+) \xi \subset \overline{S}\}$, т.е. $\exp[L(S)] \subset \overline{S}$. ■

Мы можем использовать семейство $\mathcal{K}(\mathcal{P}) = \{K_x : x \in X\}$, $K_x = \text{cont}(\mathcal{P}, x)$ в качестве управления для динамической системы (3)–(4).

Определение 2. Управление $\mathcal{K}(\mathcal{P})$ называется упорядоченным, если $\mathcal{P}$ задает порядок на X .

Предложение 2. Если порядок $\mathcal{P}$ инвариантен относительно группы G , то очевидно, что $\mathcal{K}(\mathcal{P})$ -управляемая система симметрична относительно действия группы G .

Доказательство. Это очевидно, поскольку из (9) вытекает (5). ■

Предложение 3. Если порядок $\mathcal{P}$, $P_a \neq \emptyset$, инвариантен относительно группы G , то $\mathcal{K}(\mathcal{P})$ -управляемая система (3)–(4) не выходит за пределы множества $\overline{P_a}$.

Доказательство. Следует из условия 2) в определении порядка, определения контингенции, формул (9) и предложения 1. ■

3. Упорядоченное аффинное управление

Пусть действие группы Ли G на $X = \mathbb{R}^3$ является аффинным, т.е. $\alpha : G \rightarrow \text{Aff}(\mathbb{R}^3)$. Просто транзитивное аффинное действие α порождает полную левоинвариантную аффинную структуру $\mathcal{A}$ на самой группе Ли G . Действительно, диффеоморфизм

$$\begin{aligned} \phi : G \cong \mathbb{R}^3, \quad \varphi(g) = [\alpha(g)](0, 0, 0) = (x^1, x^2, x^3) = x \in \mathbb{R}^3, \\ \phi(e) = a = (0, 0, 0) \end{aligned} \quad (10)$$

можно использовать для задания глобальной аффинной системы координат на G , в которых левые сдвиги $l_h : G \rightarrow G$, $l_h(g) = hg$ имеет вид

$$[l_h]^k(g) = [l_h]^k(\phi^{-1}(x^1, x^2, x^3)) = \sum_{i=1}^3 L_i^k x^i + L^i \quad (k = 1, 2, 3). \quad (11)$$

Предположим, что порядок $\mathcal{P}$ в $\mathbb{R}^3$ является коническим, т.е. состоит из замкнутых выпуклых конусов. Тогда можно отождествить $P_x = K_x$.

Предложение 4. Если порядок $\mathcal{P}$ инвариантен относительно просто транзитивного аффинного действия группы Ли G , то $\mathcal{K}(\mathcal{P})$ -управляемая система (3)–(4) не выходит за пределы конуса K_a . ■

Поставим задачу описать все конусы K_a в $\mathbb{R}^3$, за которые не выходит $\mathcal{K}(\mathcal{P})$ -управляемая система, эволюционирующая в $\mathbb{R}^3$. Данная задача сводится к задаче классификации и описания всех G -инвариантных конических порядков в

пространстве $\mathbb{R}^3$ относительно разрешимых¹ односвязных 3-мерных групп Ли, действующих аффинно и просто транзитивно на $\mathbb{R}^3$. Такое описание содержится в работе [5], а краткая формулировка результата приводится в следующей теореме 1.

Теорема 1. *Если G_3 не является группой Гейзенберга, тогда G_3 допускает левоинвариантный эллиптический конический порядок относительно полной левоинвариантной аффинной структуры $\mathcal{A}$, содержащей канонические координаты 2-го рода. Следовательно, $X = \mathbb{R}^3$ допускает G_3 -инвариантный эллиптический конический порядок. При этом аффинное просто транзитивное действие α , соответствующее аффинной структуре $\mathcal{A}$, является нормальным² и строится на основе метода Ямагучи [6].* ■

4. Упорядоченное однородное аффинное управление

Если потребовать, чтобы упорядоченное аффинное управление было не только симметричным, но и *однородным*, т.е. либо внутри (*Int*-однородность), либо на границе, за исключением вершины (*∂*-однородность) конуса K_a , либо вне $K_a \cup K_a^-$ (*ext*-однородность)³ действовала транзитивно аффинная группа $Aut(\mathcal{P})_a \subset Aff(\mathbb{R}^3)$, состоящая из порядковых автоморфизмов⁴, то необходимо найти описание всех *однородных аффинных причинных порядков* на 3-мерных разрешимых группах Ли. Эта задача решена в статье [8] следующим образом.

Пусть $\mathcal{A}_\gamma$ – полная левоинвариантная аффинная структура на разрешимой группе Ли G_3 , которая вводится с помощью метода С.П.Гаврилова [7]. Эта структура определяет естественное просто транзитивное аффинное действие α_γ в $\mathbb{R}^3$. В статье [7] С.П.Гаврилов описывает $\alpha_\gamma(G_3)$ -инвариантные лоренцевы метрики в $\mathbb{R}^3$. Пусть $\langle \cdot, \cdot \rangle$ – такая метрика. Конус $P_x \subset \mathbb{R}^3$ назовем *причинным*, если луч из P_x исходит в направлении ξ , лежащем в фиксированной половине касательного конуса $K_x = \{\zeta : \langle \zeta, \zeta \rangle_x \geq 0\}$. Порядок $\mathcal{P}$ в $\mathbb{R}^3$ называется *аффинно причинным*, если он состоит из множества эллиптических причинных конусов $\{P_x : x \in \mathbb{R}^3\}$.

Теорема 2. *Пусть $\mathcal{P}$ аффинно причинный порядок относительно аффинной структуры Гаврилова $\mathcal{A}_\gamma$. Тогда для связных односвязных разрешимых групп Ли G_3 типа I, VI_0 , VII_0 класса I (см. [7]) порядок $\mathcal{P}$ является одновременно *Int*–, *∂*– и *ext*–однородными. Соответствующие лоренцевы метрики плоские. Для остальных типов групп Ли порядок $\mathcal{P}$ не является однородным ни в одном из указанных выше смыслах.*

Доказательство. Дано в статье [8]. ■

¹Разрешимость есть следствие диффеоморфности (7).

²Просто транзитивное аффинное действие α на $\mathbb{R}^3$ и соответствующая аффинная структура на G_3 называются нормальными, если $\alpha(g)$ – параллельный перенос для любого $g \in T$, где T – максимальная абелева нормальная подгруппа группы G_3 .

³Здесь K_a^- – конус центрально симметричный K_a относительно точки a .

⁴Аффинное преобразование $A : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ является порядковым автоморфизмом, если $A(P_x) = P_{A(x)}$ для любой точки $x \in \mathbb{R}^3$.

ЛИТЕРАТУРА

1. Зеликин М.И. *Синтез оптимальных траекторий на пространствах представлений групп Ли* // Мат. сб. 1987. Т.132. С.541–555.
2. Зеликин М.И. *Необходимые условия для оптимальности особых траекторий в линейной задаче управления* // Некоторые вопросы современного анализа / Ред. В.М.Тихомиров. МГУ, мехмат, 1984. С.35–41.
3. Zelikin M.I. *Totally extremal manifolds for optimal control problems* // Semigroups in Algebra, Geometry and Analysis (ed. K.H.Hofmann, J.D.Lawson, E.B.Vinberg). De Gruyter Expositions in Mathematics. Berlin, 1995. 368 pp. P.339–354.
4. Винберг Э.Б. *Инвариантные конусы и упорядочивание в группах Ли* // Функ. анализ и прилож. 1980. Т.14. С.1–13.
5. Абдрахимова Н.Р., Гуц А.К., Грибанова И.А. *Описание аффинных конических порядков на трехмерных разрешимых группах Ли* // Ученый совет мат. фак. ОмГУ. Деп. в ВИНТИ 15.06.94, N 1467–В94. 35 с.
6. Yamaguchi S. *On complete affinely flat structures of some solvable Lie groups.* // Mem. Fac. Sci. Kyuchi Univ. 1979. Ser.A33. P.209–218.
7. Гаврилов С.П. *Левинвариантные метрики на разрешимых односвязных 3-мерных группах Ли* // Теория относительности и гравитация (Казань). 1985. Вып.22. С.31–64.
8. Гуц А.К., Ермакова Е.В. *Однородные аффинные причинные порядки на трехмерных разрешимых группах Ли* // Ученый совет мат. фак. ОмГУ. Деп. в ВИНТИ 05.07.93, N 1841–В93. 42 с.

ФЕЙЕРОВСКИЕ ИТЕРАЦИОННЫЕ ПРОЦЕССЫ ДЛЯ НЕСОБСТВЕННЫХ ЗАДАЧ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ

И.И. Еремин, И.М. Соколинская

In this article, we consider the Fejer's methods of solving the intrinsic (having a solution) and non-intrinsic (having no solution) linear optimization problems of 1-st, 2-nd and 3-rd types. The method is based on the reduction of mentioned problems to consistent and inconsistent systems of linear inequalities. For these systems, we construct the various variants of Fejerian processes which are converging to a solution or quasi-solution of the linear inequalities system. We discuss some aspects concerned to constructing of the iterative Fejerian mappings and the iterative sequences generated by them.

1. Введение

Основным методом решения задач линейного программирования (ЛП) является симплекс-метод. Он показал свою эффективность для задач ЛП средней размерности (несколько сотен и даже тысяч переменных и ограничений), решаемых на k -процессорных вычислительных машинах при небольших k , например при $k \leq 10$, и достаточно мощных процессорах. При больших размерностях задач ЛП (десятки и более тысяч переменных и ограничений) возникает необходимость их решения на многопроцессорных машинах (при значениях для k – сотни и тысячи процессоров) на основе принципа распараллеливания. Однако опыт решения больших задач показывает (это подкрепляется и теоретическим анализом [2]), что эффективность конечных методов, в частности симплекс-метода, падает с ростом числа задействованных процессоров в силу нарастания числа обменных операций. Это обстоятельство и порождает ограничения на использование симплекс-метода для решения больших задач ЛП. Определенный оптимизм можно связывать с использованием итерационных методов, дающих решение как предел итерационной последовательности, генерируемой тем или иным итерационным оператором. Известен широкий класс релаксационных методов фейеровского типа [1, 3], которые являются хорошими кандидатами для использования в указанных целях. Фейеровские методы отличаются простотой итерации, задаваемой формульно, удобны для массового распараллеливания

© 2002 И.И. Еремин, И.М. Соколинская

E-mail: ermii@imm.uran.ru, isok@csu.ru

Институт математики и механики УрО РАН, Челябинский государственный университет

Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (проекты 00–15–96041, 01–01–00563)

вычислительного процесса, дают возможность реализации счета при эволюционном состоянии исходных данных задачи, т.е. ее информационной составляющей. Некоторый экспериментальный материал на этот счет имеется в [4].

В статье рассматриваются фейеровские методы для решения задач ЛП с акцентом на несобственные задачи, применительно к которым метод превращается в аппроксимационно-оптимизационный.

В работе будут использованы следующие обозначения:

$\mathbf{R}^n$ – n -мерное евклидово пространство;

$\mathbf{R}_+^n := \{x \in \mathbf{R}^n \mid x^T = [x_1, \dots, x_n] \geq 0\}$;

$\alpha^+ = \max\{\alpha, 0\}$, $\alpha \in \mathbf{R}$ – поле действительных чисел;

$x^+ = [x_1^+, \dots, x_n^+]^T$ – положительная срезка вектора x ;

$\text{Arg} \dots$ – оптимальное множество выписанной вслед Arg оптимизационной задачи или ее номера; если " $\dots$ " – система неравенств, то $\text{Arg} \dots$ – множество ее решений;

$\text{arg} \dots$ – конкретный элемент из $\text{Arg} \dots$;

$\text{opt} \dots$ – оптимальное значение оптимизационной задачи;

$|x|$ – евклидова норма вектора $x \in \mathbf{R}^n$;

F_M – класс M -фейеровских отображений;

$\bar{F}_M$ – класс непрерывных M -фейеровских отображений;

$\text{rank } A$ – ранг матрицы A ;

$\{\varphi^k(x_0)\}_k$ – процесс (или последовательность), порождаемый отображением $\varphi(x) \in \{\mathbf{R}^n \rightarrow \mathbf{R}^n\}$ согласно рекуррентному соотношению: $x_{k+1} = \varphi(x_k)$, x_0 – начальный элемент процесса.

2. Краткие сведения о фейеровских и слабо фейеровских отображениях ([1], гл. II)

Пусть $\varphi \in \{\mathbf{R}^n \rightarrow \mathbf{R}^n\}$, $M \subset \mathbf{R}$. Отображение φ называется M -фейеровским, если

$$\varphi(y) = y, \quad |\varphi(x) - y| < |x - y|, \quad \forall y \in M, \quad \forall x \notin M.$$

Класс M -фейеровских отображений обозначим через F_M . Отображение $\varphi(\cdot) \in \{\mathbf{R}^n \rightarrow \mathbf{R}^n\}$ называется *слабо M -фейеровским*, если в выше приведенном определении неравенство " $<$ " заменить на " $\leq$ ". Некоторые свойства фейеровских отображений:

1. Если $\varphi(\cdot) \in F_M$ и оператор $\varphi(x)$ непрерывен, т.е. $\varphi(\cdot) \in \bar{F}_M$, то

$$\{\varphi^k(x_0) =: x_k\}_k \rightarrow \bar{x} \in M \quad (1)$$

при произвольном начальном $x_0 \in \mathbf{R}^n$.

2. Если $\varphi_j(\cdot) \in F_{M_j}$, $j = 1, \dots, m$ и $M_0 = \bigcap_{j=1}^m M_j \neq \emptyset$, то

$$\varphi(x) := \sum_{j=1}^m \alpha_j \varphi_j(x) \in F_{M_0}, \quad \alpha_j > 0, \quad \sum_{j=1}^m \alpha_j = 1. \quad (2)$$

Следовательно, если $\forall j: M_j = M$, то $\varphi(x) \in F_M$.

3. Если $\varphi_j(\cdot) \in F_{M_j}$, $j = 1, \dots, m$, то

$$\varphi_{j_1} \varphi_{j_2} \cdots \varphi_{j_m}(x) \in F_{M_0}, \quad M_0 = \bigcap_{j=1}^m M_j \neq \emptyset; \quad (3)$$

$(j_1, \dots, j_m)$ – любая перестановка индексов $\{1, \dots, m\}$.

4. Если $F_M \neq \emptyset$, то M автоматически выпукло и замкнуто.

5. Если $\varphi(\cdot)$ – слабо фейеровское отображение с множеством неподвижности $M \neq \emptyset$, то $\varphi_\alpha(x) := \alpha x + (1 - \alpha)\varphi(x) \in F_M$, $\alpha \in (0, 1)$. В частности, если отображение $\varphi(\cdot)$ является нерасширяющим, т.е. $|\varphi(x) - \varphi(y)| \leq |x - y|$, $\forall x, y \in \mathbf{R}^n$, и M – множество его точек неподвижности, то $\alpha x + (1 - \alpha)\varphi(x) \in \bar{F}_M$. Примером нерасширяющего отображения может служить операция проектирования на выпуклое замкнутое множество $M \subset \mathbf{R}^n$.

Примеры M -фейеровских отображений:

1⁰. Пусть M – выпуклое замкнутое множество из $\mathbf{R}^n$ и $\text{Pr}_M(x)$ – оператор метрического проектирования на M . Тогда $\varphi(x) := \text{Pr}_M(x) \in \bar{F}_M$. Как уже было отмечено, $\varphi(x)$ является нерасширяющим.

2⁰. Пусть $H = \{x \mid l(x) := (a, x) - \alpha = 0\}$, $a \neq 0$ и $\varphi(x) = x - \lambda \frac{l^+(x)}{\|a\|^2} \cdot a$, $\lambda \in (0, 2)$. Тогда $\varphi(\cdot) \in \bar{F}_H$. Оператор $\varphi(x)$ реализует проектирование на H с релаксацией λ .

3⁰. Пусть $P = \{x \geq 0 \mid l(x) := (a, x) - \alpha \leq 0\}$, тогда $\varphi(x) := \left[x - \lambda \frac{l^+(x)}{\|a\|^2} a \right]^+ \in \bar{F}_P$.

4⁰. Пусть $M = \{x \mid l(x) := (a_j, x) - b_j \leq 0, j = 1, \dots, m\}$, $\forall j: a_j \neq 0$; $\varphi(x) = x - (\lambda / \delta) \sum_{j=1}^m l_j^+(x) a_j$, $\delta := \sum_{j=1}^m \|a_j\|^2$, $\lambda \in (0, 2)$. Тогда: если $M \neq \emptyset$, то $\varphi(\cdot) \in \bar{F}_M$;

если $M = \emptyset$, то $\varphi(\cdot) \in \bar{F}_{\widetilde{M}}$, $\widetilde{M} = \{x \mid \nabla d(x) = 0\}$, где $d(x) = \sum_{j=1}^m l_j^{+2}(x)$.

5⁰. Пусть $H := \{x \mid Ax = b\}$ и строчки матрицы A линейно независимы. Тогда

$$\varphi(x) := \text{Pr}_H(x) = x - A^T(AA^T)^{-1}(b - Ax) \quad (4)$$

и $\varphi(\cdot) \in \bar{F}_H$. Формула (4) хорошо известна из линейной алгебры.

3. Базовый фейеровский процесс

Если M – множество решений некоторой системы линейных неравенств, то относительно этой системы формально можно конструировать отображения $\varphi(\cdot) \in F_M$, но сходимость процесса $\{\varphi^k(x_0)\}_k$ гарантируется только при условии $M \neq \emptyset$. Мы хотим выделить из всех способов конструирования $\varphi(\cdot)$ базовый способ, нацеленный на возможность использования его для построения сходящихся

итерационных процессов в ситуации несовместности системы, т.е. когда $M = \emptyset$. Предел же последовательности, порожденной базовым отображением, должен давать *квазирешение* этой системы. Этот же процесс может быть приспособлен к нахождению квазирешений несобственных задач ЛП (посредством сведения задачи ЛП к системе линейных неравенств).

Будем исходить из системы линейных неравенств и уравнений, а также одного включения:

$$\left. \begin{aligned} l_j(x) &:= (a_j, x) - b_j \leq 0, \quad j \in J_{\leq}; \\ l_i(x) &:= (a_i, x) - b_i = 0, \quad i \in J_{=}; \\ x &\in M_0. \end{aligned} \right\} \quad (5)$$

Система может быть как совместной, так и несовместной. Множество M_0 предполагается выпуклым и замкнутым. Положим

$$\varphi(x) = x - (\lambda / \delta) \left[\sum_{j \in J_{\leq}} R_j l_j^+(x) a_j + \sum_{i \in J_{=}} R_i l_i(x) a_i \right], \quad (6)$$

$\lambda \in (0, 2)$, $l_k(x) = (a_k, x) - b_k$, $k \in J := J_{\leq} \cup J_{=}$, $\delta = \sum_{k \in J} R_k |a_k|^2$, $\{R_k > 0\}_{k \in J}$

– параметры. Пусть

$$\psi(x) := \text{Pr}_{M_0} \varphi(x), \quad \psi_{\alpha}(x) := (1 - \alpha)\psi(x) + \alpha x, \quad (7)$$

$\alpha \in (0, 1)$. Под *базовой* конструкцией фейеровского оператора и *базовым* фейеровским процессом будем понимать $\psi_{\alpha}(x)$ и

$$\{\psi_{\alpha}^k(x_0)\}_k, \quad (8)$$

соответственно. Повсюду ниже x_0 – произвольный начальный элемент процесса.

Процесс (8), порождаемый оператором $\psi_{\alpha}(x)$ при произвольном начальном $x_0 \in \mathbf{R}^n$, сходится к некоторой точке $\bar{x}$, являющейся решением системы (5) в случае ее совместности; в случае же несовместности $\bar{x}$ является некоторым *аппроксимационным квазирешением*. Вопросы сходимости процесса (8) и содержательного смысла предельной точки этого процесса будут рассмотрены ниже.

Пусть $d(x) = \sum_{j \in J_{\leq}} R_j l_j^{+2}(x) + \sum_{i \in J_{=}} R_i l_i^2(x)$, $\widetilde{M} := \text{Arg} \min_{x \in M_0} d(x) \neq \emptyset$, $\varphi(x)$ – согласно (6).

Теорема 1. Пусть отображение $\psi_{\alpha}(x)$ задано согласно (7) и $\widetilde{M} \neq \emptyset$. Тогда $\psi_{\alpha}(\cdot) \in \bar{F}_{\widetilde{M}}$. ■

Следствие 1. Так как $\psi_{\alpha}(x)$ непрерывно, то согласно п. 1 раздела 2 процесс (8) сходится к некоторой точке $\bar{x} \in \widetilde{M}$. В частности, если $M = \mathbf{R}^n$ и система (5) совместна, то процесс (8) сходится к одному из ее решений. ■

Доказательству теоремы 1 предпошлем ряд лемм.

Лемма 1. (Она хорошо известна; см., например, [3], стр. 188). Пусть M – выпуклое замкнутое множество из $\mathbf{R}^n$, $p \notin M$ и $\bar{p} := \text{Pr}_M(p)$. Тогда $(p - \bar{p}, x - \bar{p}) \leq 0$, $\forall x \in M$. Верно и обратное утверждение. ■

Лемма 2. Для любых x и y из $\mathbf{R}^n$ справедливо неравенство

$$|\text{Pr}_M(x) - \text{Pr}_M(y)| \leq |x - y|, \quad (9)$$

т.е. $\text{Pr}_M(x)$ является оператором нестяжения (см., например, [1], гл. II). ■

Лемма 3. Пусть $P_j := \{x \mid l_j(x) := (a_j, x) - b_j \leq 0\}$, $a_j \neq 0$, $\varphi_j(x) = x - \lambda_j \frac{l_j^+(x)}{|a_j|^2} \cdot a_j$, $\lambda_j \in (0, 2)$. Тогда $\varphi_j(\cdot) \in F_{P_j}$ (см. п. 3⁰ раздела 2) и

$$|\varphi_j(x) - \varphi_j(y)| \leq |x - y| \quad (10)$$

для любых x и y из $\mathbf{R}^n$. Проверяется непосредственно. ■

Замечание. Если $P_i = \{x \mid l_i(x) = 0\}$ и $\bar{\varphi}_i(x) = x - \lambda \frac{l_i(x)}{|a_i|^2} a_i$, то свойство нестяжения для $\bar{\varphi}_i$ также выполняется.

Пусть M – выпуклое замкнутое множество из $\mathbf{R}^n$.

Лемма 4. Отображение $\psi(x)$, заданное согласно $\psi(x) := \text{Pr}_M \varphi(x)$, $\varphi(x)$ – согласно (6), является оператором нестяжения.

Доказательство. Отображение (6) можно записать в виде:

$$\varphi(x) = \sum_{j \in J_{\leq}} \alpha_j \left(x - \lambda \frac{l_j^+(x)}{|a_j|^2} \cdot a_j \right) + \sum_{i \in J_{=}} \alpha_i \left(x - \lambda \frac{l_i(x)}{|a_i|^2} \cdot a_i \right),$$

где $\alpha_k = \frac{R_k |a_k|^2}{\delta}$, $\delta = \sum_{j \in J_{\leq}} R_j |a_j|^2 + \sum_{i \in J_{=}} R_i |a_i|^2$, $k \in J_{\leq} \cup J_{=}$. Следовательно,

$\sum_{k \in J_{\leq} \cup J_{=}} \alpha_k = 1$. Опираясь последовательно на леммы 2 и 3, будем иметь:

$$\begin{aligned} & |\text{Pr}_M \varphi(x) - \text{Pr}_M \varphi(y)| \leq |\varphi(x) - \varphi(y)| = \\ & = \left| \sum_{j \in J_{\leq}} \alpha_j (\varphi_j(x) - \varphi_j(y)) + \sum_{i \in J_{=}} \alpha_i (\bar{\varphi}_i(x) - \bar{\varphi}_i(y)) \right| \leq \\ & \leq \sum_{j \in J_{\leq}} \alpha_j |\varphi_j(x) - \varphi_j(y)| + \sum_{i \in J_{=}} \alpha_i |\bar{\varphi}_i(x) - \bar{\varphi}_i(y)| \leq \\ & \leq \sum_{j \in J_{\leq}} \alpha_j |x - y| + \sum_{i \in J_{=}} \alpha_i |x - y| = \left(\sum_{j \in J_{\leq}} \alpha_j + \sum_{i \in J_{=}} \alpha_i \right) |x - y| = |x - y|, \end{aligned}$$

что и требовалось. ■

Отображение (7) перепишем в следующей форме

$$\varphi(x) = x - (\lambda / 2\delta) \nabla d(x), \quad (11)$$

где $d(x)$ и δ уже были введены, ∇ – символ градиента.

Лемма 5. Пусть $f(x)$ – выпуклая дифференцируемая функция и $\nabla f(\bar{x}) \neq 0$. Тогда при достаточно малом $\alpha > 0$: $f(\bar{x} - \alpha \nabla f(\bar{x})) < f(\bar{x})$ (известный факт из математического анализа). ■

Лемма 6. Пусть $\bar{x} \in M_0$ и $\widetilde{M} = \text{Arg} \min_{x \in M_0} f(x) \neq \emptyset$. Тогда

$$\bar{x} \in \widetilde{M} \iff \text{Pr}_M(\bar{x} - \gamma \nabla f(\bar{x})) = \bar{x}, \quad \gamma > 0. \quad (12)$$

Доказательство. Докажем " $\Leftarrow$ ", т.е. достаточность правого соотношения в (12). Если $\nabla d(\bar{x}) = 0$, то $\bar{x} \in \text{Arg} \min_{(x)} d(x)$. Но так как $\bar{x} \in M_0$, то $\bar{x} \in$

$\text{Arg} \min_{x \in M_0} d(x)$, т.е. $\bar{x} \in \widetilde{M}$. Пусть теперь $\nabla d(\bar{x}) \neq 0$. Обозначим $p := \bar{x} - \gamma \nabla f(\bar{x})$. По лемме 1: $(p - \text{Pr}_M(p), x - \text{Pr}_M(p)) \leq 0, \forall x \in M_0$, или $(\bar{x} - \gamma \nabla f(\bar{x}) - \bar{x}, x - \bar{x}) = -\gamma(\nabla f(\bar{x}), x - \bar{x}) \leq 0, \forall x \in M_0$. Так как $f(x)$ выпуклая функция, то из полученного неравенства вытекает $0 \leq (\nabla f(\bar{x}), x - \bar{x}) \leq f(x) - f(\bar{x}), \forall x \in M_0$, или $f(x) \geq f(\bar{x}), \forall x \in M_0$, т.е. $\bar{x} \in \text{Arg} \min_{x \in M_0} f(x)$.

Докажем " $\Rightarrow$ ", т.е. необходимость правого соотношения в (12). Если $\nabla d(\bar{x}) = 0$, то доказываемое соотношение выполняется очевидным образом, ибо $\bar{x} \in M_0$. Пусть $\nabla f(\bar{x}) \neq 0$. Тогда $p := \bar{x} - \gamma \nabla f(\bar{x}) \notin M_0$. Если это не так, то при любом $\alpha \in (0, 1)$: $p_\alpha := (1 - \alpha)\bar{x} + \alpha p = \bar{x} - \alpha\gamma \nabla f(\bar{x}) \in M_0$. Но тогда в силу леммы 5 при достаточно малом α : $f(p_\alpha) < f(\bar{x})$. Так как $p_\alpha \in M_0$, то получилось противоречие с тем, что $\bar{x} \in \text{Arg} \min_{x \in M_0} f(x)$. Итак, $p \notin M_0$.

Рассмотрим множество $\mathcal{N} := \{x \mid f(x) \leq f(\bar{x})\}$. Так как $M_0 \cap \mathcal{N}^0 = \emptyset$ ($\mathcal{N}^0 := \{x \mid f(x) < f(\bar{x})\}$), то по теореме отделимости существует гиперплоскость $H := \{x \mid (h, x - \bar{x}) = 0\}$ с $h \neq 0$, разделяющая M и $\mathcal{N}$, при этом можно считать, что h – внешняя нормаль к $\mathcal{N}$ в точке $\bar{x}$, поэтому $(h, x - \bar{x}) \geq 0, \forall x \in M_0$. Так как $f(x)$ – выпуклая гладкая функция, то $h = \gamma_0 \nabla f(\bar{x}), \gamma_0 > 0$. Предыдущее неравенство можно переписать в виде: $((\underbrace{\bar{x} - \gamma_0 \nabla f(\bar{x})}_p) - \bar{x}, x - \bar{x}) \leq 0, \forall x \in M_0$.

По лемме 1: $\bar{x}$ – проекция точки p на M_0 , т.е. $\text{Pr}_{M_0}(\bar{x} - \gamma_0 \nabla f(\bar{x})) = \bar{x}$. Заметим, что выполнимость или невыполнимость этого соотношения не зависит от значения величины $\gamma_0 > 0$.

Лемма 6 доказана полностью. ■

Перейдем к доказательству теоремы 1. По существу, оно вытекает из леммы 6 и п. 5 раздела 2. Действительно, пусть $\bar{x} \in \widetilde{M}$ и $x \notin M$. Согласно лемме 6: $\psi(\bar{x}) = \bar{x}$, а потому и $\psi_\alpha(\bar{x}) = \bar{x}$ (напомним: $\psi(x) = \text{Pr}_M \varphi(x), \varphi(x)$ – согласно (6)).

Далее, $|\psi(x) - \bar{x}| = |\psi(x) - \psi(\bar{x})| \stackrel{\text{лемма 4}}{\leq} |x - \bar{x}|$. Отсюда вытекает (согласно п. 5 раздела 2): $|\psi_\alpha(x) - \bar{x}| < |x - \bar{x}|$.

Теорема 1 доказана.

Замечание. Если $M_0 := \mathbf{R}^n$, то, очевидно, $\psi_\alpha(x) = \varphi_\alpha(x) := (1-\alpha)\varphi(x) + \alpha x$, причем $\psi_\alpha(\cdot) \in F_{\widetilde{M}}$. На самом деле уже само отображение $\varphi(x)$ является $\widetilde{M}$ -фейеровским, т.е. $\varphi(\cdot) \in F_{\widetilde{M}}$. Действительно, положив $\lambda' = \lambda / (1-\alpha)$, будем иметь: $\varphi(x) = (1-\alpha) \underbrace{[x - (\lambda' / \delta) \nabla d(x)]}_{\varphi_0(x)} + \alpha x$, при этом $\lambda' \in (0, 2)$ при $\alpha < 1 - \lambda / 2$.

Так как оператор $\varphi_0(x)$ – нестягивающий, то опять-таки в силу п. 5 раздела 2: $\varphi(x) = (1-\alpha)\varphi_0(\cdot) + \alpha x \in F_{\widetilde{M}}$.

4. Фейеровский процесс для решения задачи линейного программирования

Пусть

$$L : \max \{ (c, x) \mid Ax \leq b, \ x \geq 0 \} \quad (13)$$

– разрешимая задача ЛП;

$$L^* : \min \{ (b, u) \mid A^T u \geq c, \ u \geq 0 \} \quad (14)$$

– двойственная к L .

Задачам L и L^* поставим в соответствие систему линейных неравенств

$$S : \left. \begin{array}{ll} Ax \leq b, \ x \geq 0 & ; \quad (15)_1 \\ A^T u \geq c, \ u \geq 0 & ; \quad (15)_2 \\ (c, x) = (b, u) & , \quad (15)_3 \end{array} \right\} \quad (15)$$

называемую *симметрической*.

В теории ЛП хорошо известен следующий факт:

$$\boxed{\text{Arg } S = \text{Arg } L \times \text{Arg } L^*} \quad (16)$$

В (16): $\text{Arg } S$ – множество решений системы S .

Если

$$\begin{aligned} \varphi_1(x) &\in F_{M_1}, & M_1 &:= \text{Arg } (15)_1; \\ \varphi_2(u) &\in F_{M_2}, & M_2 &:= \text{Arg } (15)_2; \\ \varphi_3(x, u) &:= \text{Pr}_H(x, u), & H &= \text{Arg } (15)_3 \end{aligned}$$

и

$$\varphi(x, u) := \text{Pr}_H(\varphi_1^+(x), \varphi_2^+(u)), \quad (17)$$

то $\varphi(x, u) \in \widetilde{F}_{\widetilde{M}}$, где $\widetilde{M} = \text{Arg } S$, т.е. $\varphi(x, u)$ является непрерывным фейеровским отображением относительно множества $\text{Arg } L \times \text{Arg } L^*$. Следовательно (в силу п. 1 раздела 3):

$$\{\varphi^k(x_0, u_0)\}_k \rightarrow [\bar{x}, \bar{u}] \in \text{Arg } S. \quad (18)$$

Что касается выбора отображений $\varphi_1(x)$ и $\varphi_2(x)$, то это можно делать разными способами, в частности *базовым* способом, изложенными в разделе 3, а именно:

$$\left. \begin{aligned} \varphi_1(x) &= x - (\lambda / \delta_1) \sum_{j=1}^m l_j^+(x) a_j, \\ \varphi_2(u) &= u - (\lambda / \delta_2) \sum_{i=1}^m h_i^+(u) h_i, \\ \varphi_3(x, u) &= [x, u]^T - \frac{(c, x) - (b, u)}{|c|^2 + |b|^2} [c, -b]^T; \end{aligned} \right\} \quad (19)$$

выше $\{h_i\}$ – столбцы матрицы A , $h_i(u) = c_i - (h_i, u)$, $\delta_1 = \sum_{j=1}^m |a_j|^2$, $\delta_2 = \sum_{i=1}^n |h_i|^2$,

$$\lambda \in (0, 2), \quad [x, u]^T = \begin{bmatrix} x \\ u \end{bmatrix}, \quad [c, -b]^T = \begin{bmatrix} c \\ -b \end{bmatrix}.$$

Заметим, что вид отображений $\varphi_1(\cdot)$ и $\varphi_2(\cdot)$ будет меняться в зависимости от того, в какой форме записана исходная задача линейного программирования. Если, например, задача L задана в канонической форме

$$\min \{(c, x) \mid Ax = b, x \geq 0\}, \quad (20)$$

то двойственная к ней будет иметь вид

$$\max \{(b, u) \mid A^T u \geq 0\}, \quad (21)$$

а отображения $\varphi_1(\cdot)$, $\varphi_2(\cdot)$ и $\varphi(x, u)$ соответственно примут вид:

$$\left. \begin{aligned} \varphi_1(x) &= x - (\lambda / \delta_1) \sum_{j=1}^m l_j(x) a_j, \\ \varphi_2(u) &= u - (\lambda / \delta_2) \sum [(h_i, u) - c_i]^+ h_i, \\ \varphi_3(x, u) &= \text{Pr}_H(\varphi_1(x), \varphi_2^+(u)). \end{aligned} \right\} \quad (22)$$

Вид $\varphi_3(x, u)$ формулы проектирования на H сохраняется.

В связи с тем, что основные ограничения в (20) записаны в форме системы линейных уравнений $Ax = b$, пусть с множеством решений H_1 , то $\varphi_1(\cdot)$ можно сконструировать, исходя из операции проектирования на H_1 , аналогично (4):

$$\text{Pr}_{H_1}(x) = x - A^T(AA^T)^{-1}(b - Ax) \quad (23)$$

(в предположении, что $\text{rank } A = n$). Полагая $\varphi_1(x) = \text{Pr}_{H_1}(x)$, заключительный итерационный оператор $\varphi(x, u)$ будет иметь ту же форму, что и в (22).

Случай задания отображения $\varphi(x, u)$ применительно к задаче (20) в форме

$$\varphi(x, u, v) := \text{Pr}_{H_0}(x, u^+, v^+), \quad (24)$$

где H_0 – множество решений системы

$$Ax = b, \quad A^T u + v = c, \quad (c, x) = (b, u), \quad (25)$$

рассматривался в работе [4], посвященной численным экспериментам решения задач ЛП большой размерности на многопроцессорных вычислительных машинах.

5. Базовый аппроксимационно-фейеровский процесс для несобственной задачи ЛП 1-го рода

Если исходная задача ЛП неразрешима, т.е. по принятой терминологии является *несобственной*, то ее соответствующая симметрическая система S является несовместной (верно и обратное). Для этой задачи в этом случае можно ввести понятие *квазирешения* через определение *квазирешения* системы S . Поясним это для случая задания задачи ЛП в форме

$$L : \max \{(c, x) \mid Ax \leq b\}. \quad (26)$$

Двойственной к ней будет

$$L^* : \min \{(b, u) \mid A^T u = c, u \geq 0\}. \quad (27)$$

Симметрической системой S является

$$\begin{aligned} Ax \leq b, \quad A^T u = c, \quad u \geq 0, \\ (c, x) = (b, u). \end{aligned} \quad (28)$$

Если L – несобственная задача ЛП 1-го рода, т.е.

$$M := \{x \mid Ax \leq b\} = \emptyset, \quad M^* := \{u \mid A^T u = c, u \geq 0\} \neq \emptyset,$$

то, введя

$$H = \left\{ \begin{bmatrix} x \\ u \end{bmatrix} \mid A^T u = c, (c, x) = (b, u) \right\}, \quad d(x, u) = |(Ax - b)^+|^2 + |(-u)^+|^2,$$

можно сформировать задачу

$$\min \{d(x, u) \mid u \in H\} \quad (29)$$

– как аппроксимационную для несовместной системы (28).

Если $\begin{bmatrix} \bar{x} \\ \bar{u} \end{bmatrix} \in \text{Arg}(29)$, то $\bar{x}$ и будем называть квазирешением системы (28), а вместе с тем и квазирешением задачи (26). Если последняя разрешима, то $\bar{x}$ – обычное ее решение. Для (29) можно записать фейеровское отображение относительно множества $\text{Arg}(29)$ по стандарту из раздела 3, но применительно к системе (28):

$$\begin{aligned} \varphi(x, u) &= \begin{bmatrix} x \\ u \end{bmatrix} - (\lambda / 2 \delta) \nabla_{x, u} d(x, u), \\ \psi(x, u) &= \text{Pr}_H \varphi(x, u), \end{aligned}$$

$$\psi_\alpha(x, u) = (1 - \alpha) \Psi(x, u) + \alpha \begin{bmatrix} x \\ u \end{bmatrix}, \quad (30)$$

где $\lambda \in (0, 2)$, $\delta = \sum_{j=1}^m |a_j|^2 + m$, $\alpha \in (0, 1)$.

В соответствии с теоремой 1 $\psi_\alpha(x, u)$ является $\widetilde{M}$ -фейеровским непрерывным отображением относительно $\widetilde{M} = \text{Arg} (5.4)$. Следовательно, процесс

$$\{\psi_\alpha^k(x_0, u_0)\}_k$$

сходится к некоторому вектору $\begin{bmatrix} \bar{x} \\ \bar{u} \end{bmatrix}$, при этом по определению

$\bar{x}$ – квазирешение задачи (26),

$\bar{u}$ – квазирешение задачи (27).

Операция $\text{Pr}_H(\cdot)$, фигурирующая в определении $\psi(x, u)$ и входящая, следовательно, в формулу (30), может быть записана согласно соотношению (23) с заменой матрицы A на матрицу

$$\bar{A} := \begin{bmatrix} 0 & A^T \\ c^T & -b^T \end{bmatrix},$$

а вектор b (из (23)) – на $\begin{bmatrix} c \\ 0 \end{bmatrix}$.

6. Аппроксимационно-фейеровский процесс для несобственной задачи ЛП 2-го рода

Рассмотрение вопроса соотнесем к задачам (26)–(28). Ситуация несобственности 2-го рода соответствует тому, что $M \neq \emptyset$, $M^* = \emptyset$, где смысл символов M и M^* тот же, что и в разделе 5. Систему (28) разобьем на две части:

$$A^T u = c, \quad u \geq 0; \quad (31)$$

$$Ax + v = b, \quad v \geq 0, \quad (c, x) = (b, u); \quad (32)$$

выше система $Ax \leq b$ заменена на $Ax + v = b$, $v \geq 0$ – обычная перезапись системы неравенств, если в этом есть необходимость. Положим, $d(x, u, v) = |A^T u - c|^2 + |(-u)^+|^2 + |(-v)^+|^2$, H_0 – множество решений системы (32). В силу предположения о непустоте M имеем: $H_0 \neq \emptyset$. В согласии с базовой конструкцией фейеровского отображения $\psi_\alpha(\cdot)$ из раздела 3, реализованного для НЗ ЛП 1-го рода в разделе 5, соответствующим отображением для рассматриваемого случая НЗ ЛП 2-го рода будет отображение

$$\psi_\alpha(x, u, v) = \text{Pr}_{H_0} \left(\begin{bmatrix} x \\ u \\ v \end{bmatrix} - (\lambda / 2\delta) \nabla_{x,u,v} d(x, u, v) \right); \quad (33)$$

здесь $\lambda \in (0, 2)$, $\delta = \sum_{i=1}^n |h_i|^2 + 2m$.

Процесс, порожденный отображением (33), будет сходиться к некоторому вектору $[\bar{x}, \bar{u}, \bar{v}]^T$, при этом $\bar{x}$ – квазирешение задачи (26), $\bar{u}$ – квазирешение задачи (27).

7. Фейеровский процесс для НЗ ЛП 3-го рода

В предыдущих разделах 5 и 6 мы апеллировали к несовместности симметрической системы S , поставленной в соответствие исходной задаче ЛП, которая предполагалась либо НЗ ЛП 1-го рода, либо 2-го. Аппроксимационная суть, соотношенная системе S , заключалась в формировании ее совместной подсистемы из уравнений и квадратичной функции невязки $d(x)$ для оставшихся ограничений этой системы. Понятно, что эта операция неоднозначна, что является весьма полезным обстоятельством, ибо во всяком конкретном случае, т.е. в случае конкретного формата задачи (например, формата транспортной задачи, блочного формата и т.д.), указанная операция выделения подсистемы может быть увязана с потребностью простой и эффективной численной реализации итерационного шага, порождаемого итерационным оператором. К этому следует еще добавить, что построение итогового фейеровского итерационного оператора будет зависеть от конкретного вида исходной задачи ЛП. Всякий раз при построении нужного фейеровского отображения необходимо все это учитывать.

Для рассмотрения НЗ ЛП 3-го рода возьмем задачу ЛП, ей двойственную, и систему S в формате (13)–(15).

Так как итоговое итерационное отображение $\psi_\alpha(\cdot)$ формируется из фрагментов $\varphi(x)$, $\psi(x) = \text{Pr}_M \varphi(x)$, а они в свою очередь – из M , $d(\cdot)$ и δ , то, формируя аналоги отображений $\psi_\alpha(\cdot)$ применительно к системе (15), мы в рассматриваемых ниже вариантах будем приводить лишь вид множества M , функции невязки $d(\cdot)$ и числа δ .

$$\text{Вариант 1. } M_1 := \left\{ \begin{bmatrix} x \\ u \end{bmatrix} \mid (c, x) = (b, u) \right\}, \quad d_1(x, u) = |(Ax - b)^+|^2 + |(c - A^T u)^+|^2 + |(-x)^+|^2 + |(-u)^+|^2, \quad \delta_1 = \sum_{j=1}^m |a_j|^2 + \sum_{i=1}^n |h_i|^2 + m + n.$$

$$\text{Вариант 2. } M_2 := \left\{ \begin{bmatrix} x \\ u \\ v \end{bmatrix} \mid A^T u - v = c, (c, x) = (b, u) \right\}, \quad d_2(x, u, v) = |(Ax - b)^+|^2 + |(-x)^+|^2 + |(-u)^+|^2 + |(-v)^+|^2, \quad \delta_2 = \sum_{j=1}^m |a_j|^2 + m + 2n.$$

$$\text{Вариант 3. } M_3 := \left\{ \begin{bmatrix} x \\ u \\ v \end{bmatrix} \mid Ax - w = b, (c, x) = (b, u) \right\}, \quad d_3(x, u, w) = |(c - A^T u)^+|^2 + |(-x)^+|^2 + |(-u)^+|^2 + |(-w)^+|^2, \quad \delta_3 = \sum_{i=1}^n |h_i|^2 + 2m + n.$$

Вариант 4. $M_4 := \mathbf{R}^n$, $d_4(x, u) = |(Ax - b)^+|^2 + |(c - A^T u)^+|^2 + |(-x)^+|^2 + |(-u)^+|^2$,

$$\delta_4 = \sum_{j=1}^m |a_j|^2 + \sum_{i=1}^n |h_i|^2 + m + n.$$

Последний вариант годится как для разрешимой задачи, так и для неразрешимой любого рода неразрешимости (1-го, 2-го или 3-го).

8. Приближенный метод решения разрешимой задачи ЛП, основанный на ее редукции к задаче минимизации кусочно-квадратичной выпуклой функции

Нам безразлично, в какой форме представить исходную задачу ЛП, но для простоты последующего формульного текста запишем ее в виде

$$L : \min \{(c, x) \mid Ax \leq b\}. \quad (34)$$

Известен результат по методам штрафных функций [3, теорема 20.2]: *Задача (34) асимптотически по $R = [R_1, \dots, R_m] > 0$ эквивалентна задаче*

$$\min \left\{ (c, x) + \sum_{j=1}^m R_j l_j^{+2}(x) \right\}, \quad (35)$$

т.е. задача (35) приближает (как по аргументу, так и по значению) задачу (34), при этом тем точнее, чем больше число $\min_{(j)} R_j$. На самом деле имеются оценки для соответствующих уклонений, но приводить их не будем. Результат верен и для нелинейной целевой функции в задаче (34).

Пусть γ_0 – некоторая нижняя оценка для $\text{opt}(8.1)$, т.е. $(c, x) \geq \gamma_0$, $\forall x \in M := \{x \mid Ax \leq b\}$. В конкретных задачах такую оценку получить не составляет труда. Именно в силу того, что $(c, x) - \gamma_0 \geq 0$ для всех допустимых x , задачу (34) можно записать в эквивалентном виде:

$$\min \{[(c, x) - \gamma_0]^2 \mid x \in M\}. \quad (36)$$

Применив к ней приведенный выше метод квадратичных штрафных функций, получим задачу

$$\min_{(x)} \{l_0^2(x) + \sum_{j=1}^m R_j l_j^{+2}(x)\}, \quad (37)$$

где $l_0(x) = (c, x) - \gamma_0$. Так что задача (37) приближенно решает исходную задачу (34). Но минимизируемая функция в задаче (37) является частным случаем для функции $d(x)$ из раздела 3, так что для поиска точки из $\text{Arg} \min_{(x)} d(x)$ можно применить фейеровский процесс

$$\{\varphi^k(x_0)\}_k, \quad (38)$$

в котором

$$\varphi(x) = x - (\lambda / \delta) \nabla d(x), \quad \lambda \in (0, 2), \quad \delta = |c|^2 + \sum_{j=1}^m |a_j|^2,$$

$$d(x) = [(c, x) - \gamma_0]^2 + \sum_{j=1}^m R_j l_j^{+2}(x),$$

$$\nabla d(x) = 2 \left\{ [(c, x) - \gamma_0]c + \sum_{j=1}^m R_j l_j^+(x) a_j \right\}.$$

Если $\bar{x}$ – предел сходящейся последовательности, порождаемой процессом (37) при произвольном $x_0 \in \mathbf{R}^n$, то $\bar{x}$ решает приближенно задачу (34), и тем точнее, чем больше числа $R_j > 0$, $j = 1, \dots, m$. Поскольку $\bar{x} \in \operatorname{Arg} \min_{(x)} d(x)$, то $\nabla d(\bar{x}) = 0$, а потому признаком останова вычислительного процесса может служить выполнимость неравенства $|\nabla d(x_k)| \leq \varepsilon$, ε – наперед заданное положительное число.

9. Декомпозиционные методы построения фейеровских отображений применительно к системам линейных неравенств

Как уже отмечалось (см. раздел 4), решение задачи ЛП сводится к решению системы линейных неравенств S с матрицей коэффициентов, имеющей блочную структуру. Если взять в качестве исходной задачи задачу (13), то системой S является (15) с матрицей коэффициентов перед переменными x_i и y_j :

$$\bar{A} := \begin{bmatrix} \boxed{A} & & \\ & \boxed{-A^T} & \\ \boxed{c^T} & & \boxed{-b^T} \end{bmatrix}.$$

Построение отображения (17) было основано на декомпозиционном принципе, учитывающем блочную структуру матрицы $\bar{A}$. Ниже этот вопрос рассмотрим подробнее применительно к канонической системе линейных уравнений и неравенств

$$Ax = b, \quad x \geq 0 \quad (39)$$

при разных вариантах блочности матрицы A (в скобках заметим, что символ A в данном контексте не следует связывать с символом A в предыдущих записях).

Вариант 1. Пусть матрица A разбита на горизонтальные подматрицы A_i :

$$\left[\begin{array}{c} \boxed{A_1} \\ \vdots \\ \boxed{A_n} \end{array} \right], \quad (40)$$

что соответствует разбиению системы (39) на подсистемы

$$A_i x = b^i, \quad x \geq 0, \quad i = 1, \dots, n. \quad (41)$$

Пусть $H_i = \{x \mid A_i x = b^i\}$. Тогда

$$\varphi(x) = \left[\sum_{i=1}^n \alpha \Pr_{H_i}(x) \right]^+ \in F_M,$$

где $M = \{x \geq 0 \mid Ax \leq b\}$ (см. п. 2 и 5⁰ из раздела 2). На самом деле $\varphi(x)$ непрерывно, поэтому $\varphi(\cdot) \in \bar{F}_M$, что дает

$$\{\varphi^k(x_0)\}_k \rightarrow \bar{x} \in M.$$

Целесообразность разбиения системы (39) на подсистемы можно связать с проблемой преодоления трудностей, связанных с обращением матрицы AA^T (см. формулу проектирования (4) на $H = \{x \mid Ax = b\}$). При разбиении системы (39) на подсистемы (41) нужно будет обращаться матрицы $A_i A_i^T$ меньших размеров. К тому же само разбиение, учитывающее структуру матрицы A , может быть нацелено именно на эффективность указанных обращений.

Вариант 2. Пусть

$$A := \left[\begin{array}{ccc} \boxed{A_1} & & \\ & \ddots & \boxed{A_n} \\ \boxed{A_0} & & \end{array} \right],$$

тогда система (39) имеет вид

$$\left. \begin{array}{l} A_i x^i = b^i, \quad i = 1, \dots, n, \\ A_0 x = b^0, \quad x^T = [x^1, \dots, x^n] \geq 0 \end{array} \right\} \quad (42)$$

и $H_i := \{x^i \mid A_i x^i = b^i\} \subset \mathbf{R}^{n_i}$, $H_0 = \{x \mid A_0 x = b^0\}$.

Если подсистемам $A_i x^i = b^i$, $i = 1, \dots, n$, $A_0 x = b^0$ соотнести $\varphi_i(\cdot) \in \bar{F}_{H_i}$ и $\varphi_0(x) \in \bar{F}_{H_0}$, то

$$\varphi(x) := \varphi_0^+(\varphi_1(x^1), \dots, \varphi_n(x^n)) \in \bar{F}_M,$$

причем выбор отображений $\{\varphi_i\}_{i=0}^n$ может осуществляться разными способами, например, $\varphi_i(x^i) = \Pr_{H_i}(x^i)$, $i = 1, \dots, n$; $\varphi_0(x) = \Pr_{H_0}(x)$.

Вариант 3. Рассмотрим матрицу

$$\begin{bmatrix} \boxed{A_1} & & & \\ & \ddots & & \\ & & \boxed{A_n} & \\ & & & \begin{bmatrix} B_1 \\ \vdots \\ B_n \end{bmatrix} \end{bmatrix}$$

и ей соответствующую систему

$$\left. \begin{aligned} A_i x^i + B_i x^0 &= b^i, \quad i = 1, \dots, n, \\ x^T &= [x^1, \dots, x^n, x^0] \geq 0. \end{aligned} \right\} \quad (43)$$

Подсистеме $A_i x^i + B_i x^0 = b^i$ поставим в соответствие $\varphi_i(x^i, x^0) \in \bar{F}_{H_i}(x^i, x^0)$, где $H_i = \{[x^i, x^0] \mid A_i x^i + B_i x^0 = b^i\}$. Обозначим через $\bar{x}^i$ алгебраический след вектора $\varphi_i(x^i, x^0)$ в $\mathbf{R}^{n_i}$ – как подпространство пространства $(\prod_{i=1}^n \mathbf{R}^{n_i}) \times \mathbf{R}^{n_0}$ исходной переменной $x^T = [x^1, \dots, x^n, x^0]$, $(\bar{x}^0)_i$ – аналогичный след в $\mathbf{R}^{n_0}$. Образует отображение

$$\varphi(x) = \left[\bar{x}^1, \dots, \bar{x}^n, \frac{1}{n} \sum_{i=1}^n (\bar{x}^0)_i \right]^+.$$

Оно является непрерывным M -фейеровским [5].

10. Вычислительные аспекты применения фейеровских отображений

Обсудим некоторые вопросы эффективных реализаций фейеровских процессов, соотнесенных задачам поиска решения, либо системы линейных неравенств, либо задачи линейного программирования. Остановимся просто на системе линейных неравенств

$$Ax \leq b \sim l_j(x) := (a_j, x) - b_j \leq 0, \quad j = 1, \dots, m, \quad (44)$$

поскольку задачи ЛП сводятся к таким системам. Будем предполагать $M := \{Ax \leq b\} \neq \emptyset$. В разделе 9 мы уже обсуждали вопрос об учете структуры матрицы A при построении фейеровского итерационного отображения $\varphi(\cdot) \in \bar{F}_M$. Несомненно, эффективность вычислительного процесса, порождаемого отображением $\varphi(\cdot)$, зависит от разумного учета особенностей структуры матрицы A . Но это только одна сторона дела. Само генерирование итерационной последовательности допускает большое число модификаций, основанных на иных соображениях. Пусть выбрано для системы (44) базовое отображение

$$\varphi(x) = x - (\lambda / \delta) \sum_{j=1}^m l_j^+(x) a_j \in \bar{F}_M.$$

Для $k < m$ введем

$$\varphi_k(x) = x - (\lambda / \delta_k) \sum_{j=1}^k l_j^+(x) a_j, \quad \delta_k = \sum_{j=1}^k |a_j|^k$$

– это отображение, соотнесенное подсистеме $l_j(x) \leq 0, \quad j = 1, \dots, k$. При больших значениях m (тысячи и десятки тысяч) целесообразней счет вести согласно соотношению

$$x_{k+1} = \varphi_k^{t_k}(x_k), \quad (45)$$

при этом $\varphi_k(x) = \varphi(x)$ и $t_k = 1$ при $k \geq m$.

В процессе (45) предусмотрены по крайней мере два обстоятельства:

1) на шаге с номером k (при $k < m$) итерация $\varphi_k(x_k)$ выполняется для подсистемы из k неравенств, что сокращает число арифметических и логических операций, необходимых для подсчета $\varphi_k(x_k)$;

2) так вычисляемые $\varphi_k(\cdot)$ можно повторить несколько, пусть t_k , раз, до обеспечения, например, условия $\sum_{j=1}^k l_j^+(x_k) \leq \varepsilon$. После этого срабатывает интегрированная итерация (45).

При такой тактике перехода от x_k к x_{k+1} подсчет δ_{k+1} ведется рекуррентно: $\delta_{k+1} = \delta_k + |a_{k+2}|^2$.

При описанной организации счета факт сходимости итерационной последовательности $\{x_k\}$ к решению системы (44) сохраняется.

Заметим, что рекуррентное соотношение (45) пригодно и для решения бесконечных систем линейных неравенств ([1], стр. 106).

Что касается скорости сходимости рассмотренных фейеровских процессов применительно к совместным системам линейных неравенств, пусть это будет (44), то ее можно охарактеризовать *скоростью* по геометрической прогрессии, понимаемой в смысле

$$|x_{k+1} - M| \leq \theta |x_k - M|, \quad \forall k, \quad \theta \in (0, 1);$$

здесь $|x - M| := \inf_{y \in M} |x - y|$. Оценки для θ были даны в работе одного из авторов ([1], стр. 68).

По фейеровским методам проводились вычислительные эксперименты на многопроцессорной вычислительной машине МВС-100/1000 с применением методов распараллеливания (см. [2–4]). Решались задачи ЛП с числом переменных $n \leq 4000$ и числом ограничений $m \leq 1000$. В основу метода был положен итерационный процесс $z_{k+1} = \text{Pr}_M^+(z_k)$ для системы $Az = b, \quad z \geq 0$, к которой редуцировалась задача ЛП. Результаты экспериментов были удовлетворительными. Эксперименты планируется продолжить.

ЛИТЕРАТУРА

1. Еремин И.И., Мазуров В.Д. *Нестационарные процессы математического программирования*. М.: Наука, 1979. –205 с. (гл. II).
2. Попов Л.Д. *Вопросы реализации методов ЛП в траспьютерных сетях* / Сб.: «Алгоритмы и программные средства параллельных вычислений». Екатеринбург: Изд-во УрО РАН, 1995. N.1. С.148-156.
3. Еремин И.И. *Теория линейной оптимизации*. Екатеринбург: Изд-во УрО РАН, 1999. –312 с.
4. Бердникова Л.Д., Попов Л.Д. *О применении декомпозиции при реализации фейеровских методов решения больших систем линейных неравенств на МВС-100* / Сб.: «Алгоритмы и программные средства параллельных вычислений». Екатеринбург: Изд-во УрО РАН, 2000. N.4. С.51-62.
5. Еремин И.И. *Синтез фейеровских отображений с несовпадающими пространствами их образов* // ДАН. 2001. Т.378, N.1. С.11-13.

ДВУМЕРНЫЕ ГЕЛЬМГОЛЬЦЕВЫ ПРОСТРАНСТВА

В.А. Кыров

The Helmholtz planes are defined and their motion groups are given. The quasi-metric tensor that is analogues of metric tensor of the Riemannian spaces is found.

В данной работе определяются гельмгольцевы плоскости и приводятся группы их движений. Находятся метрические функции f гельмгольцевых двумерных пространств, которые в бесконечно малом имеют структуру гельмгольцевых плоскостей. Определяется квазиметрический тензор, который является в некотором смысле аналогом метрического тензора в римановых пространствах. Доказывается теорема о существовании единственной квазиметрической (согласованной) связности с нулевым кручением, относительно которой при параллельном переносе сохраняется метрическая функция f . Устанавливаются свойства тензора кривизны R гельмгольцевых пространств. Вводится секционная гельмгольцева кривизна K и доказывается, что только для локально плоских пространств она является константой, причем равной нулю.

Приступим к определению гельмгольцевых плоскостей.

Г.Г. Михайличенко в монографии [1] проводит полную классификацию двумерных феноменологически симметричных геометрий, то есть геометрий со следующим свойством. Существует достаточно гладкое многообразие N , в котором можно ввести единую систему координат. Существует плотное подмногообразие N' прямого произведения $N \times N$ некоторого двумерного многообразия N на себя. Существует также достаточно гладкая невырожденная функция $f : N' \rightarrow R$, которую будем называть *метрической функцией*, и гладкая функция шести переменных $\Phi : R^6 \rightarrow R$, что для любого кортежа из четырех произвольных точек $\langle xyzi \rangle$, каждая пара из которого принадлежит множеству N' , имеет место функциональное уравнение: $\Phi(f(xy), f(xz), f(xu), f(yz), f(yu), f(zu)) = 0$.

Этому свойству удовлетворяют некоторые известные, а также и неизвестные геометрии. К известным геометриям с таким свойством принадлежат: плоскость Евклида, псевдоевклидова плоскость, симплектическая плоскость и т.д. Неизвестные геометрии со свойством феноменологической симметрии — это: *плоскость Гельмгольца* Γ^2 , метрическая функция $f(xy)$ которой в специальной единой системе координат имеет следующее представление:

$$f(xy) = [(x^1 - y^1)^2 + (x^2 - y^2)^2] \exp \left(2\gamma \arctg \frac{x^2 - y^2}{x^1 - y^1} \right), \quad (1)$$

где $\gamma > 0$, x^1 и x^2 — координаты точки x , причем функция $arctg$ рассматривается однозначной с областью значений в промежутке $(-\pi/2, \pi/2]$ (этот термин появился из анализа знаменитой работы Гельмгольца «О фактах, лежащих в основании геометрии», где он предлагал изучать геометрию, в которой роль окружности выполняет логарифмическая спираль); *псевдогельмгольцева плоскость* $P\Gamma^2$, метрическая функция f которой в некоторой системе координат имеет вид:

$$f(xy) = [(x^1 - y^1)^2 - (x^2 - y^2)^2] \exp \left(2\beta Ar(c)th \frac{x^2 - y^2}{x^1 - y^1} \right), \quad (2)$$

где $\beta > 0$ и $\beta \neq 1$, причем выбирается функция $Arth$, если аргумент по модулю меньше единицы, и выбирается функция $Arcth$, если аргумент по величине больше единицы; *дуальногельмгольцева плоскость* D^2 с метрической функцией f в некоторой системе координат:

$$f(xy) = (x^1 - y^1)^2 \exp \left(2 \frac{x^2 - y^2}{x^1 - y^1} \right) \quad (3)$$

и, наконец, *симплициальная плоскость* S^2 с метрической функцией $f(ij)$, которая в некоторой системе координат имеет следующее представление:

$$f(xy) = \frac{x^2 - y^2}{x^1 - y^1}. \quad (4)$$

Объединим в одном выражении метрические функции (1)-(3):

$$f(xy) = [(x^1 - y^1)^2 - \varepsilon(x^2 - y^2)^2] \exp 2\alpha\Phi_\varepsilon \left(\frac{x^2 - y^2}{x^1 - y^1} \right), \quad (5)$$

где для плоскости Гельмгольца Γ^2 $\varepsilon = -1$, $\alpha = \gamma$ и $\Phi_{-1}(x) = arctgx$; для псевдогельмгольцевой плоскости $P\Gamma^2$ $\varepsilon = 1$, $\alpha = \beta$ и $\Phi_1(x) = Ar(c)thx$; для дуальногельмгольцевой плоскости D^2 $\varepsilon = 0$, $\alpha = 1$ и $\Phi_0(x) = x$. Ниже под термином плоскость Гельмгольца, если нет опасности путаницы, будет пониматься одна из этих четырех плоскостей, которую обозначим через F^2 .

Теперь введем строгое определение плоскости Гельмгольца F^2 . Итак, будем говорить, что множество точек x двумерного многообразия N принадлежат F^2 , если существует такая точка y из N , что пара $\langle xy \rangle \in N'$.

Будем предполагать, что эта плоскость одновременно является подмногообразием многообразия N .

Перейдем теперь к определению группы движений плоскости F^2 и ее подгруппы вращений.

Пусть G — группа преобразований плоскости F^2 . Преобразование g назовем *движением*, если оно сохраняет метрическую функцию f , то есть оставляет ее инвариантной: $f \circ (g \times g) = f$

Михайличенко Г.Г. в монографии [1] показал, что по метрической функции f находится трехмерная группа движений G , а по этой группе движений восстанавливается метрическая функция f с точностью до «масштабной» гладкой

функции $\varphi : R \rightarrow R$. Решая эту задачу для выше приведенных плоскостей, приходим к группам движений G_{F^2} , которые в выше определенных координатах и специальной системы параметров для плоскостей Γ^2 , $P\Gamma^2$, D^2 задаются следующими уравнениями:

$$x^{1'} = ax^1 + \varepsilon bx^2 + c, \quad x^{2'} = bx^1 + ax^2 + d, \quad (a^2 - \varepsilon b^2) \exp 2\alpha\Phi_\varepsilon \left(\frac{b}{a} \right) = 1; \quad (6)$$

для плоскости S^2 :

$$x^{1'} = ax^1 + c, \quad x^{2'} = ax^2 + d, \quad a \neq 0. \quad (7)$$

Очевидно, что эти группы являются подгруппами группы аффинных преобразований. Выделим в группах G_{F^2} подгруппы $\tilde{G}_{F^2}$ по следующему принципу. В G_{F^2} существует такая нормальная подгруппа T , являющаяся группой сдвигов, что фактор-группа G_{F^2}/T изоморфна подгруппе $\tilde{G}_{F^2}$. Тогда G_{F^2} является полупрямым произведением $\tilde{G}_{F^2}$ и T . Эту группу будем называть *группой вращений плоскости F^2* и обозначать $O(F^2)$. В соответствующей системе координат и параметров эти группы имеют, очевидно, следующие выражения:

$$x^{1'} = ax^1 + \varepsilon bx^2, \quad x^{2'} = bx^1 + ax^2, \quad (a^2 - \varepsilon b^2) \exp 2\alpha\Phi_\varepsilon \left(\frac{b}{a} \right) = 1 \quad (6')$$

и

$$x^{1'} = ax^1, \quad x^{2'} = ax^2, \quad a \neq 0. \quad (7')$$

Можно показать, что у группы $O(F^2)$ имеется три независимых двухточечных гладких инварианта ψ_1, ψ_2, ψ_3 : $O(F^2) \rightarrow (\psi_1, \psi_2, \psi_3)$, от которых удобно перейти к трем инвариантам [2] (необязательно независимым и невырожденным). Эти инварианты в специальной системе координат имеют следующие представления: для плоскостей $F^2 = \Gamma^2, P\Gamma^2, D^2$:

$$(x, y)_{F^2} = (x^1 y^1 - \varepsilon x^2 y^2) \exp \left(\alpha\Phi_\varepsilon \left(\frac{x^2}{x^1} \right) + \alpha\Phi_\varepsilon \left(\frac{y^2}{y^1} \right) \right), \quad (8a)$$

$$< x, y >_{F^2} = (x^1 y^2 - x^2 y^1) \exp \left(\alpha\Phi_\varepsilon \left(\frac{x^2}{x^1} \right) + \alpha\Phi_\varepsilon \left(\frac{y^2}{y^1} \right) \right), \quad (8b)$$

$$(x, x)_{F^2} = ((x^1)^2 - \varepsilon (x^2)^2) \exp 2\alpha\Phi_\varepsilon \left(\frac{x^2}{x^1} \right); \quad (8c)$$

для симплицальной плоскости S^2 :

$$(x, y)_{S^2} = \sqrt{\frac{x^2 y^2}{x^1 y^1}}, \quad (9a)$$

$$< x, y >_{S^2} = \frac{x^1 y^2}{x^2 y^1}, \quad (9b)$$

$$(x, x)_{S^2} = \frac{x^2}{x^1}, \quad (9c)$$

где, напомним, x^1 и x^2 – координаты точки x .

Пусть V^2 – двумерное вещественное линейное пространство, в котором фиксируем базис e_1, e_2 . Относительно этого базиса произвольный вектор ξ имеет координаты (ξ^1, ξ^2) . В V^2 определим гельмгольцево квазискалярное произведение (аналог скалярного произведения в римановых пространствах) между векторами ξ и η , которое относительно базиса e_1, e_2 в координатах задается формулой (9а) или (10) и обозначается $(\xi, \eta)_{F^2}$, где уже $F^2 = \Gamma^2, P\Gamma^2, D^2, S^2$. Вторые формулы определяют в V^2 относительно этого же базиса e_1, e_2 площадь между этими векторами, которая обозначается $< \xi, \eta >_{F^2}$. Заметим, что введенное здесь квазискалярное произведение не удовлетворяет всем свойствам обычного скалярного произведения. Так, оно не является линейным и не определено в нуле.

1. Двумерные гельмгольцевы пространства

С данного момента приступим к определению двумерных гельмгольцевых пространств, которые в бесконечно малой окрестности произвольной точки устроены как гельмгольцевы плоскости. Предварительно введем понятие структурных функций, через которые мы выразим метрические функции этих пространств.

Пусть M – гладкое n -мерное многообразие. Пусть $T(M)$ – касательное расслоение со стандартным слоем V^n и касательным пространством $T_x(M)$ в точке x многообразия M . Пусть $L(M)$ – расслоение линейных реперов u со структурной группой $GL(n, R)$. Слой $L_x(M)$ в $L(M)$ над точкой $x \in M$ состоит из линейных реперов u , отличающихся друг от друга на элементы из $GL(n, R)$, то есть $v = uA$, $A \in GL(n, R)$, $u, v \in L_x(M)$. Очевидно, что в точке $x \in M$ $u = (X_1, \dots, X_n)$. Каждый линейный репер u из $L(M)$ можно рассмотреть как изоморфизм V^n на $T_x(M)$ [3]. Пусть $(e_1, \dots, e_n)$ – фиксированный базис в V^n . Тогда $ue_i = X_i$, где $i = 1, \dots, n$, следовательно, $u\xi = \xi^i X_i \in T_x(M)$, причем $\xi = \xi^i e_i$.

Рассмотрим отображение

$$\omega : L_x(M) \times L_x(M) \times T_x(M) \rightarrow V^n, \quad (10)$$

которое в явном виде задается следующей формулой. Пусть u и v – произвольные реперы из $L_x(M)$. Пусть произвольный вектор $X \in T_x(M)$ в базисе v имеет координаты $X^1, \dots, X^n$. Тогда положим

$$\omega(u, v, X) = a_i^j X^i e_j. \quad (11)$$

Предположим, что $v = (Y_1, \dots, Y_n)$ и $u = (X_1, \dots, X_n)$. Поэтому $u = (X_1^j Y_j, \dots, X_n^j Y_j)$, причем $i, j = 1, \dots, n$. В отношении отображения (10) выдвинем еще требование: $a = X^{-1}$, где X – матрица отличия репера u от v . Предполагается, что соответствие (10) является гладким по своим аргументам. Кроме того, будем полагать, что относительно открытого покрытия U_α многообразия M в силу

локальной тривиальности расслоения $L(M)$ существует семейство изоморфизмов $\kappa_\alpha : L(U_\alpha) \rightarrow U_\alpha \times GL(n, R)$, которые определяют сечения через u и v : $\sigma_\alpha : U_\alpha \rightarrow L(M)$ по формулам $\sigma_{1\alpha}(x) = \kappa_{1\alpha}^{-1}(x, e)$ и $\sigma_{2\alpha}(x) = \kappa_{2\alpha}^{-1}(x, e)$ соответственно. Тогда соответствие $x \mapsto \omega$ является необходимым число раз дифференцируемым в произвольной координатной окрестности $U \subset M$. Поэтому элементы матрицы a являются дифференцируемыми функциями в U , которые будем называть *структурными функциями*.

Естественно рассмотреть следующее сужение отображения (10):

$$\omega : L_x(M) \times \{v\} \times T_x(M) \rightarrow V^n.$$

Его мы обозначим

$$\omega_v : L_x(M) \times T_x(M) \rightarrow V^n. \quad (12)$$

Предположим, что $v = (\partial x^1, \dots, \partial x^n)$ – координатный базис в координатной окрестности U произвольной точки многообразия M . Тогда

$$\omega_v(u, \partial x^i) = e_i^* = a_i^j e_j.$$

Очевидно следующее свойство отображения ω_v :

$$u \circ \omega_v(u, \partial x^i) = \partial x^i$$

и в более общем случае

$$u \circ \omega_v(u, X) = X.$$

Лемма 1. При переходе от системы координат U к системе координат U' структурные функции a преобразуются по закону

$$a_i'^j = a_k^j \frac{\partial x^k}{\partial x'^i}, \quad (13)$$

где a – структурные функции в координатной окрестности U , а a' – структурные функции в координатной окрестности U' , причем $i, j, k = 1, \dots, n$. ■

Из леммы следует, что функции (11) инвариантны относительно произвольной замены координат.

Рассмотрим подгруппу Ли G группы $GL(n, R)$ и редукцию общей линейной группы $GL(n, R)$ к подгруппе G . Редуцированное подрасслоение будем обозначать $Q(M, G)$ или просто Q . Рассмотрим сужение отображения (10) на Q_x относительно первого аргумента.

$$\omega : Q_x(M) \times L_x(M) \times T_x(M) \rightarrow V^n. \quad (14)$$

Функция (14), так же, как и функция (10), является гладкой. Тогда для фиксированного репера $v \in L_x(M)$

$$\omega_v(u, X) = a_i^j X^i e_j, \quad (15)$$

где u – произвольный репер из Q_x . Заметим, что функции (15) в силу выше доказанной леммы инвариантны относительно произвольной замены координат. Поскольку u – произвольный репер из Q_x , а v – некоторый фиксированный репер из $L_x(M)$, то матрица a представима в таком виде:

$$a = bc \quad (16)$$

или в координатах

$$a_j^i = b_k^i c_j^k,$$

где b – произвольный элемент подгруппы G , а c – некоторая фиксированная матрица из $GL(n, R)$. Пусть U_α – некоторое открытое покрытие многообразия M . В силу локальной тривиальности расслоения $L(M)$ и того, что $Q(M)$ – редуцированное подрасслоение расслоения линейных реперов, существует семейство изоморфизмов $\kappa_\alpha : L(U_\alpha) \rightarrow U_\alpha \times GL(n, R)$, которое индуцирует семейство изоморфизмов $\varkappa_\alpha : Q(U_\alpha) \rightarrow U_\alpha \times G$. Рассмотрим отображения $\sigma_\alpha : U_\alpha \rightarrow L(M)$ и $\tau_\alpha : U_\alpha \rightarrow Q(M)$, которые определяют сечения: $\sigma_\alpha(x) = \kappa_\alpha^{-1}(x, e)$ через репер $v \in L_x(M)$, в частности, если v – координатный базис, то в качестве сечения $\sigma_\alpha(x)$ удобно взять сечение, задаваемое реперами $\partial_{x^1}, \dots, \partial_{x^n}$, $\tau_{1\alpha}(x) = \varkappa_{1\alpha}^{-1}(x, e)$ через произвольный репер $u \in Q_x(M)$ и $\tau_{2\alpha}(x) = \varkappa_{1\alpha}^{-1}(x, e)$ через некоторый репер $u' \in Q_x(M)$. Тогда a^{-1} – матрица отличия репера $\tau_{1\alpha}(x)$ от репера $\sigma_\alpha(x)$, а c^{-1} – матрица отличия репера $\tau_{2\alpha}(x)$ от репера $\sigma_\alpha(x)$. Заметим, что c – это некоторая специальным образом подобранная матрица. Из выше приводимых рассуждений следует, что функция (14), так же как и функция (10) в координатной окрестности U , является гладкой с гладким соответствием $x \mapsto \omega$. Очевидно, что разложение (16) справедливо и для структурных функций в координатной окрестности U .

Пусть теперь M – гладкое двумерное многообразие. Пусть G – это группа гельмгольцевых вращений $O(F^2)$. Рассмотрим редуцированное подрасслоение $Q(M) = Q(M, O(F^2))$ расслоения линейных реперов $L(M)$. Этому подрасслоению соответствует отображение

$$\omega_v : Q_x(M) \times T_x(M) \rightarrow V^2.$$

Поскольку для G -пространства справедливо разложение структурных функций (16), то отображение ω_v в координатах принимает такой вид:

$$\omega_v(u, X) = b_k^j c_i^k X^i e_j, \quad (17)$$

где b – произвольная матрица из $O(F^2)$. Заметим, что выражение (17) при произвольном u задает семейство векторов в V^2 , которые определяют одно и то же значение квазискалярного произведения вектора на себя $(\xi, \xi)_{F^2}$ по формуле (8а) или (9с). Так как в (17) X – произвольный вектор из $T_x(M)$, то естественно переносится квазискалярное произведение в касательное пространство $T_x(M)$ по формуле

$$f(X, Y) = (\omega_v(u, X), \omega_v(u, Y))_{F^2}, \quad (18)$$

причем $X, Y \in T_x(M)$. Таким образом, мы получили отображение

$$f : Q_x(M) \times T_x(M) \times T_x(M) \rightarrow R.$$

Это отображение, очевидно, является гладким по своим аргументам и как соответствие $x \mapsto f$. Формула (18) позволяет в $L(M)$ ввести отношение эквивалентности. Классами эквивалентности являются редуцированные подрасслоения $Q(M, O(F^2))$. Функцию (18) можно положить в основу определения подрасслоения $Q(M, O(F^2))$. Эту функцию будем называть метрической функцией *гельмгольца двумерного пространства*.

Найдем явный вид метрической функции (18) в координатной окрестности U . Пусть линейное отображение ω_v переводит векторы $X, Y \in T_x(M)$ в соответственно следующие векторы

$$\omega_v(u, X) = a_i^j X^i e_j, \quad \omega_v(u, Y) = a_i^j Y^i e_j,$$

где под v теперь понимается координатный базис в U . Следовательно, метрическая функция f принимает такой вид: для собственно гельмгольцевых, псевдогельмгольцевых и дуальногельмгольцевых пространств:

$$f(X, Y) = g_{ij}^\varepsilon X^i Y^j \exp \left(\alpha \Phi_\varepsilon \left(\frac{a_k^2 X^k}{a_l^1 X^l} \right) + \alpha \Phi_\varepsilon \left(\frac{a_k^2 Y^k}{a_l^1 Y^l} \right) \right); \quad (19)$$

для симплицальных пространств:

$$f(X, Y) = \sqrt{\frac{a_k^2 X^k a_j^2 Y^j}{a_l^1 X^l a_i^1 Y^i}}, \quad (20)$$

$i, j, k, l = 1, 2$, причем

$$g_{ij}^\varepsilon = a_i^1 a_j^1 - \varepsilon a_i^2 a_j^2, \quad (21)$$

где $\varepsilon = -1, 1, 0$ и под ε не понимается индекс суммирования. Заметим, что здесь под a уже понимается фиксированная матрица. Можно показать, что символы (21) образуют тензор. Легко установить, что метрическая функция (19) или (20) остается инвариантной при переходе к произвольной другой системе координат.

Дифференциалы координат dx^i образуют контравариантный тензор первого ранга, которому однозначно соответствует вектор в касательном пространстве $T_x(M)$. Положим в (19) и (20) $X^i = Y^i = dx^i$. Тогда будем иметь

$$f = g_{ij}^\varepsilon dx^i dx^j \exp \left(2\alpha \Phi_\varepsilon \left(\frac{a_k^2 dx^k}{a_l^1 dx^l} \right) \right); \quad (19')$$

$$f = \frac{a_j^2 dx^j}{a_i^1 dx^i}, \quad (20')$$

где $i, j, k, l = 1, 2$.

Следует заметить, что если вместо $Q(M, O(F^2))$ взять расслоение ортонормированных реперов $O(M)$ и провести аналогичные рассуждения, то получим римановы пространства [3].

2. Квазиметрическая связность

Заметим, что по построению гельмгольцевы пространства являются пространствами с линейной связностью. Приступим к исследованию связностей в этих пространствах. Докажем существование и единственность квазиметрической связности с нулевым кручением.

Связность в расслоении линейных реперов $L(M)$ называется *квазиметрической или согласованной связностью*, если параллельный перенос слоев из $T(M)$ сохраняет метрическую функцию f гельмгольцева пространства.

Из определения квазиметрической связности следует равенство нулю ковариантной производной:

$$\nabla_k f(X, X) = 0.$$

Кручение определяется обычным образом. Несложно убедиться в справедливости следующей теоремы.

Теорема 1. *Гельмгольцево двумерное многообразие в координатной окрестности U произвольной точки допускает единственную квазиметрическую связность с нулевым кручением, символы Кристоффеля которой задаются выражениями*

$$\Gamma_{ik}^{\varepsilon l} = \frac{1}{2} h^{\varepsilon lk} \left(\frac{\partial h_{jk}^{\varepsilon}}{\partial x^i} + \frac{\partial h_{ki}^{\varepsilon}}{\partial x^j} - \frac{\partial h_{ij}^{\varepsilon}}{\partial x^k} \right) - \alpha h^{\varepsilon lk} (\lambda_{jki} + \lambda_{kij} - \lambda_{ijk}), \quad (22)$$

$$\Gamma_{ik}^l = \frac{1}{2} h^{lk} \left(\frac{\partial h_{jk}}{\partial x^i} + \frac{\partial h_{ki}}{\partial x^j} - \frac{\partial h_{ij}}{\partial x^k} \right) - h^{lk} (\lambda_{jki} + \lambda_{kij} - \lambda_{ijk}), \quad (23)$$

где для квазиметрического тензора имеем выражение

$$h_{ij}^{\varepsilon} = a_i^1 a_j^1 - \varepsilon a_i^2 a_j^2 + \alpha (a_i^1 a_j^2 - a_j^1 a_i^2), \quad h_{ij} = a_i^1 a_j^2 - a_j^1 a_i^2, \quad (24)$$

для символов

$$\lambda_{ijk} = a_j^2 \frac{\partial a_i^1}{\partial x^k} - a_j^1 \frac{\partial a_i^2}{\partial x^k} \quad (25)$$

и, наконец, $h^{ij} h_{jk} = \delta_k^i$, $h^{\varepsilon ij} h_{jk}^{\varepsilon} = \delta_k^i$, $\varepsilon = -1, 1, 0$ и $\alpha = \gamma, \beta, 1$, причем под ε не следует понимать индекс суммирования. ■

Из доказательства теоремы 1 и выражения для ковариантной производной тензора следует, что

$$\nabla_k h_{ij}^{\varepsilon} = 2\alpha q_{ijk}, \quad q_{ijk} = \Gamma_{ki}^{\varepsilon l} \tilde{h}_{jl} + \lambda_{ijk}, \quad \nabla_k h_{ij} = 2\alpha q_{ijk}, \quad q_{ijk} = \Gamma_{ki}^l \tilde{h}_{jl} + \lambda_{ijk}.$$

Следует заметить, что для римановых пространств ковариантная производная метрического тензора относительно римановой связности равна также нулю.

Легко показать, что символы h_{ij}^{ε} и h_{ij} преобразуются по тензорному закону, которые будут называться *квазиметрическими тензорами*. Символы λ_{ijk} преобразуются по закону:

$$\lambda'_{ijk} = \lambda_{lmn} \frac{\partial x^l}{\partial x'^i} \frac{\partial x^m}{\partial x'^j} \frac{\partial x^n}{\partial x'^k} + h_{lm} \frac{\partial x^l}{\partial x'^i} \frac{\partial x^m}{\partial x'^k} \frac{\partial x^n}{\partial x'^j},$$

где $i, j, k, l = 1, 2$.

Предположим, что в координатной окрестности U произвольной точки $x \in M$ функции $a_j^i = \text{const}$. Тогда мы приходим локально к гельмгольцевой плоскости F^2 . Осуществим локально обратимую замену координат $x = a_i^1 x^i$, $y = a_i^2 x^i$, где по «немому» индексу i ведется суммирование от 1 до 2. Значит, метрические функции (19'), (20') принимают такой вид:

$$f = [dx^2 - \varepsilon dy^2] \exp \left(2\alpha \Phi_\varepsilon \left(\frac{dy}{dx} \right) \right); \quad (19'')$$

$$f = \frac{dy}{dx}. \quad (20'')$$

Сравнивая с (19'), (20'), имеем $a_1^1 = a_2^2 = 1$, $a_2^1 = a_1^2 = 0$. Поэтому квазиметрический тензор гельмгольцевых плоскостей равен:

$$h_{ij}^\varepsilon = \begin{pmatrix} 1 & \alpha \\ -\alpha & -\varepsilon \end{pmatrix}, \quad h_{ij} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

3. Секционная гельмгольцева кривизна

В этом параграфе введем понятие секционной гельмгольцевой кривизны $O(F^2)$ -пространств, где уже строго под F^2 понимается либо Γ^2 , либо $P\Gamma^2$, либо D^2 . Заметим, что для симплицальных пространств понятие секционной кривизны не определяется. Сначала определим свойства компонент тензора кривизны гельмгольцевых пространств. С данного момента под $O(F^2)$ будем понимать изоморфный образ группы вращений с конечными уравнениями (6') в общей линейной группе $GL(n, R)$. Справедлива следующая

Теорема 2. Алгебра Ли $o(F^2)$ группы Ли $O(F^2)$ состоит из матриц

$$\begin{pmatrix} -\alpha a & \varepsilon a \\ a & -\alpha a \end{pmatrix}, \quad (26)$$

где, напомним, $\alpha = \gamma, \beta, 1$ и $\varepsilon = -1, 1, 0$ соответственно. ■

Напомним, что связность в расслоении линейных реперов $L(M)$ называется согласованной со структурой, если при параллельном переносе слоев в $T(M)$ сохраняется метрическая функция f гельмгольцева пространства. Тогда форма кривизны [3] связности, согласованной со структурой, является элементом алгебры Ли $o(F^2)$, то есть имеет вид (26). Переходя от формы кривизны к тензору кривизны R [3], устанавливаем его свойства:

1. Антисимметрия по нижним индексам: $R_{jkl}^i = -R_{jlk}^i$;
2. При фиксированных индексах k и l в точке x $R_{jkl}^i \in o(F^2)$.

Из этих свойств следуют равенства

$$R_{112}^1 = -\alpha R_{112}^2, \quad R_{212}^1 = \varepsilon R_{112}^1, \quad R_{212}^2 = -\alpha R_{112}^2.$$

Перейдем теперь к выводу выражений для секционной кривизны. Рассмотрим двумерное гельмгольцево многообразие M с касательным пространством

$T_x(M)$. Касательное пространство $T_x(M)$ является также двумерным и поэтому содержит единственную двумерную плоскость, то есть M имеет единственное двумерное направление. Рассмотрим два произвольных линейно независимых вектора $X, Y \in T_x(M)$, которые относительно координатного базиса имеют координаты X^i и Y^i , где $i = 1, 2$. Секционную кривизну K определим следующей формулой:

$$K = \frac{f(R(X, Y)Y, X)}{\langle X, Y \rangle_{F^2}} \varphi(X, Y), \quad (27)$$

где введено сокращающее обозначение

$$\varphi(X, Y) = \exp \left[\alpha \Phi_\varepsilon \left(\frac{a_i^2 X^i}{a_j^1 X^j} \right) + \alpha \Phi_\varepsilon \left(\frac{a_i^2 Y^i}{a_j^1 Y^j} \right) \right],$$

причем a_j^i — структурные функции. Из выражений (8.2) для гельмгольцевой площади следует, что в координатах

$$\langle X, Y \rangle_{F^2} = h_{ij} X^i Y^j \varphi(X, Y).$$

Рассмотрим подмножество $S_x(M)$ прямого произведения $T_x(M) \times T_x(M)$, состоящее из линейно независимых пар векторов $\langle X, Y \rangle$, то есть этому множеству не принадлежит диагональ из прямого произведения. Пару векторов будем называть линейно независимой, если определитель, составленный из их координат, равен нулю. Очевидно, что в $S_x(M)$ можно естественно ввести гладкую структуру. Следует заметить, что для ненулевых векторов $\langle X, Y \rangle_{F^2} = 0$ тогда и только тогда, когда $\langle X, Y \rangle \in T_x(M) \times T_x(M) / S_x(M)$. В таком случае секционную кривизну K можно представить как отображение

$$K : M \times S_x(M) \rightarrow R. \quad (28)$$

Пусть векторы $X, Y \in T_x(M)$, причем $\langle X, Y \rangle \in S_x(M)$, включены в гладкие векторные поля из $T(M)$. Тогда функции $f, \langle, \rangle_{F^2}, \varphi$ будут гладко зависеть от точек из M и пар векторов из $S_x(M)$. Значит, функция (28) является гладкой.

Выражение (27) можно еще записать в такой форме:

$$K = \frac{f(R(X, Y)Y, X)}{h_{ij} X^i Y^j}. \quad (27')$$

Из выше доказанной леммы следует, что при замене координат в точке x секционная кривизна K остается неизменной.

Выясним теперь, как изменяется секционная кривизна K в зависимости от преобразований векторов в касательном пространстве $T_x(M)$.

Пусть H_x — совокупность диффеоморфизмов касательного пространства $T_x(M)$. Очевидно, что H_x является группой Ли, совпадающей с $GL(T_x(M))$. Элементы h_x группы H_x тогда являются невырожденными матрицами. Будем предполагать, что соответствие $x \mapsto h_x$ является гладким. Группа H_x в касательном пространстве $T_x(M)$ действует так: $h_x : X \mapsto h_x(X)$. Определим действие группы H_x в прямом произведении $T_x(M) \times T_x(M)$ по формуле

$h_x \times h'_x : \langle X, Y \rangle \mapsto \langle h_x(X), h'_x(Y) \rangle$. Эту группу мы обозначим $(H \times H)_x$. Очевидно, что пространство $S_x(M)$ является инвариантом группы $(H \times H)_x$. Рассмотрим диагональную подгруппу группы $(H \times H)_x$, которую будем обозначать просто H_0 . Пусть элемент $(h \times h)_x$ группы H_0 пару векторов $\langle X, Y \rangle$ из пространства S_x переводит в пару $\langle h_x(X), h_x(Y) \rangle$. Выясним, как при этом будет изменяться значение секционной кривизны K . Имеет место следующее утверждение. Если h_x является элементом группы гельмгольцевых вращений $O(F^2)$, где $F^2 = \Gamma^2, P\Gamma^2, D^2$, то справедлива формула

$$K(x, X, Y) = K(x, h_x(X), h_x(Y)). \quad (29)$$

Последнее утверждение приводит к тому, что пространство $S_x(M)$ разбивается на классы $S_x(H_0)$. Эти классы состоят из пар векторов $\langle X, Y \rangle$, которые отличаются друг от друга на матрицы группы гельмгольцевых вращений $O(F^2)$. Тогда для произвольных пар векторов класса $S_x(H_0)$ имеет место формула (29).

Заметим, что в случае римановых пространств это утверждение справедливо для произвольных преобразований из группы H_0 .

Если $K = \text{const}$ для всех точек $x \in M$, то пространство называется пространством постоянной кривизны.

От (27') приходим к формуле

$$f(R(X, Y)Y, X) = Kh_{ij}X^iY^j, \quad (30)$$

которая нам ниже понадобится. Справедлива следующая

Теорема 3. Если секционная гельмгольцева кривизна K в некоторой координатной окрестности U некоторой точки гельмгольцева двумерного пространства не зависит от векторов X и Y , то $K = 0$ в этой окрестности.

Доказательство. Предположим, что в некоторой окрестности U некоторой точки K не зависит от X и Y . Тогда в (30) правая часть является билинейной формой, а левая часть не является такой. Это возможно только при $K = 0$. Теорема доказана. ■

Из этой теоремы следует, что гельмгольцево пространство постоянной кривизны может иметь только нулевую кривизну.

ЛИТЕРАТУРА

1. Михайличенко Г.Г. *Полиметрические геометрии*. Новосибирск, 2001.
2. Кыров В.А. *Векторы некоторых двумерных феноменологически симметричных геометрий* // Наука, культура, образование (Горно-Алтайск). 2000. N.6/7. С.111 - 114.
3. Кобаяси Ш., Номидзу К. *Основы дифференциальной геометрии*. М.: Наука, 1981.

ДИФФЕРЕНЦИАЛЬНЫЕ УРАВНЕНИЯ В СИНТЕТИЧЕСКОЙ ДИФФЕРЕНЦИАЛЬНОЙ ГЕОМЕТРИИ ЛОВЕРА-КОКА

А.А. Звягинцев

Theory of differential equations in intuitionistic Synthetic Differential Geometry of Kock-Lawvere is considered. The problem of solution uniqueness is discussed.

Как и в классическом случае, под дифференциальным уравнением первого порядка над многообразием (микролинейным пространством [1]) M будем понимать пару (M, X) , состоящую из самого многообразия M и векторного поля X над ним, т. е. отображения вида $X : M \rightarrow T(M)$, здесь $T(M)$ – обозначение для M^D [2]. В свою очередь, морфизмом векторных полей $(M_1, X_1) \rightarrow (M_2, X_2)$ назовем такое отображение $f : M_1 \rightarrow M_2$, что коммутативен следующий квадрат:

$$\begin{array}{ccc} T(M_1) & \xrightarrow{Tf} & T(M_2) \\ X_1 \uparrow & & \uparrow X_2 \\ M_1 & \xrightarrow{f} & M_2 \end{array}$$

Объединяя теперь эти определения, построим категорию обыкновенных дифференциальных уравнений первого порядка (1ODE), взяв в качестве объектов всевозможные пары вида (M, X) , а в качестве морфизмов – морфизмы векторных полей.

1. Формальная теория дифференциальных уравнений первого и второго порядка [3]

Изучение свойств этой категории и будет составлять основную часть этого параграфа.

Следует заметить, что понятие решения дифференциального уравнения также можно записать внешним образом, т.е. через некоторую диаграмму в категории (1ODE). А именно, если $(M, X) \in Ob(1ODE)$, то его (глобальным) решением называется отображение из $(R, \frac{\partial}{\partial x})$ в (M, X) .

Подобным образом 2ODE – категория дифференциальных уравнений второго порядка – обычно строится как категория векторных полей на $T(M)$, т.е. отображений вида

$$\xi : TM \rightarrow TTM$$

В качестве ее морфизмов берется такое отображение $f : M_1 \rightarrow M_2$, что коммутативен следующий квадрат:

$$\begin{array}{ccc} TT(M_1) & \xrightarrow{TTf} & TT(M_2) \\ \xi_1 \uparrow & & \uparrow \xi_2 \\ T(M_1) & \xrightarrow{Tf} & T(M_2) \end{array}$$

В отличие от решений уравнений первого порядка решения уравнений второго порядка в общем случае не являются гомоморфизмами из R , если, конечно, это уравнение не струя на M . Кроме того, верна следующая теорема.

Теорема 1 [3]. Пусть $\xi : M^D \rightarrow M^{D^2}$ является объектом 2ODE на M , и $\tilde{\xi} : M^D \xrightarrow{\xi} M^{D^2} \xrightarrow{M^+} M^{D \times D}$ соответствует 1ODE на M^D , и пусть задан морфизм $y : R \rightarrow M$. Тогда, если y – решение 2ODE ξ , то $\dot{y} : R \rightarrow M^D$ – решение 1ODE $\tilde{\xi}$. Обратное верно, если M^+ мономорфизм. ■

В контексте всего вышесказанного, необходимо потребовать, чтобы категория $\mathcal{E}$, являлась атомом, т.е. функтор $(-)^D : \mathcal{E} \rightarrow \mathcal{E}$ имел бы не только левый присоединенный $- \times D$, но также и правый присоединенный $(-)^{1/D}$, который обычно называется дробной экспонентой (терминология Ловера). Подробнее о свойствах моделей, удовлетворяющих этому условию, можно прочитать в [1]. Обратимся теперь к категориям 1ODE и 2ODE, но вначале заметим (см. [1]), что дифференциальное уравнение второго порядка эквивалентным образом можно определить как отображение

$$M^D \rightarrow M^{D^2},$$

которое «продолжает» 1-джет на 2-джет.

Определение 1. Предсвязностью на M называется отображение $\xi : M^{D(2)} \rightarrow M^{D \times D}$ такое, что

$$M^{D(2)} \xrightarrow{\xi} M^{D \times D} M^i \rightarrow M^{D(2)} = \text{id}_{M^{D(2)}},$$

где $i : D(2) \rightarrow D \times D$ – каноническое включение.

Теорема 2 [3]. Пусть $\mathcal{E}$ – топос Гротендика, который одновременно является моделью SDG, и пусть $\mathbf{G}$ – одна из следующих категорий:

1. Векторных полей (1ODE) на $\mathcal{E}$;
2. Дифференциальных уравнений второго порядка (2ODE) на $\mathcal{E}$;
3. Предсвязностей на $\mathcal{E}$;
4. Струй на $\mathcal{E}$;
5. Аффинных связностей на $\mathcal{E}$.

Тогда $\mathbf{G}$ – топос Гротендика, и точный функтор $(M, \xi) \mapsto M$ является обратным образом неотъемлемой геометрической сюръекции $\mathcal{E} \rightarrow \mathbf{G}$. ■

2. Примеры и решения дифференциальных уравнений первого и второго порядка

Важной задачей как классической, так и синтетической теории дифференциальных уравнений, базирующейся на Синтетической дифференциальной геометрии Ловера-Кока (СДГ) [2], является задача нахождения одного или нескольких решений некоторого дифференциального уравнения. Однако изложенная выше формальная теория не справляется с ней, поскольку порой поиск векторного поля, удовлетворяющего исходному уравнению, становится трудоемким, а иногда и неразрешимым процессом. Исходя из этого, введем иное определение дифференциального уравнения, основанное на понятии функции, определенной на $M|_U \times M^{D_n}$ со значениями в R . Здесь U подобъект объекта M . Далее будем считать, что $M = R$, а $U = [a, b]$.

Определение 2. Дифференциальным уравнением над кольцом R называется уравнение вида

$$F(t, y(t), \dots, y^{(n)}(t)) = 0 \quad \forall t \in [a, b]. \quad (*)$$

Определение 3. Функция $g : [a, b] \rightarrow R$ называется решением уравнения (*), если

$$F(t, g(t), \dots, g^{(n)}(t)) \equiv 0.$$

Перейдем теперь к интерпретации дифференциальных уравнений над кольцом R . Итак, пусть $y \in R^{[a,b]}$ и $F : R^{n+1} \rightarrow R$ – функция, определенная на многообразии $S = \{(t, y(t), y'(t), \dots, y^{(n)}(t)) \mid t \in [a, b]\}$. Тогда дифференциальное уравнение

$$F(t, y(t), y'(t), \dots, y^{(n)}(t)) = 0 \quad (*)$$

на стадии lA переписывается в следующем виде:

$$F(x, t, y(t), y'(t), \dots, y^{(n)}(t)) = 0 \quad \text{mod } \pi^*(I). \quad (**)$$

Следовательно, решение уравнения (*) — это класс $Y(x, t) \text{ mod } \pi^*(I)$, т.ч.

$$F\left(x, t, Y(x, t), \frac{\partial Y(x, t)}{\partial t}, \dots, \frac{\partial^n Y(x, t)}{\partial t^n}\right) \equiv 0 \quad \text{mod } \pi^*(I).$$

Рассмотрим несколько важных примеров.

Пример 1. Найдем решение задачи Коши

$$\begin{aligned} y'(t) &= y(t) \\ y(0) &= a, \quad a \in R \end{aligned}$$

Решение. На стадии $lA = lC^\infty(\mathbf{R}^n)/I$ она запишется следующим образом:

$$\begin{aligned} \frac{\partial Y(x, t)}{\partial t} &= Y(x, t) \quad \text{mod } \pi^*(I) \\ Y(x, 0) &= A(x) \quad \text{mod } I, \quad A \in lC^\infty(\mathbf{R}^n). \end{aligned}$$

Решим вначале дифференциальное уравнение. Из доказательства истинности аксиомы интегрирования следует, что:

$$Y(x, t) = (Y(x, 0) + \int_0^t Y(x, s) ds) \mod \pi^*(I).$$

Отсюда видно, что решение д.у. = решение классического д.у. в частных производных $+ \mod \pi^*(I)$. Таким образом, достаточно решить дифференциальное уравнение:

$$\frac{\partial Y(x, t)}{\partial t} = Y(x, t).$$

Это классическое квазилинейное дифференциальное уравнение. Запишем уравнения характеристик:

$$\frac{dt}{1} = \frac{dx}{0} = \frac{dy}{y}.$$

Находя характеристики, имеем:

$$\{x = C_1, \ln y - t = C_2\} \Rightarrow \Phi(x, \ln y - t) = 0,$$

$$\ln y - t = C(x) \Rightarrow \ln y = t + C(x) \Rightarrow y = e^{C(x)} e^t,$$

$$(e^{C(x)} = C_1(x)) \Rightarrow Y(x, t) = C_1(x) e^t.$$

Проверяем начальное условие:

$$Y(x, 0) = A(x) \Rightarrow C_1(x) = A(x) \Rightarrow y(x, t) = A(x) \cdot e^t.$$

■

Пример 2. Рассмотрим дифференциальное уравнение $y' = c \cdot y$, где $c \in R$ — необратимо.

Решение. На стадии $lA = lC^\infty(\mathbf{R}^n)/I$ оно запишется следующим образом

$$\frac{\partial H(u, t)}{\partial t} = C(u) \cdot Y(u, t) \mod \pi^*(I).$$

Очевидно, что $Y(u, t) = e^{t \cdot C(u)}$ — решение дифференциального уравнения. Положим, что $F(u, t)$ — решение дифференциального уравнения. Следовательно,

$$\frac{\partial F(u, t)}{\partial t} = C(u) \cdot F(u, t),$$

тогда запишем $F(u, t)$ в виде:

$$F(u, t) = \frac{F(u, 0)}{e^{C(u) \cdot t}} \cdot e^{C(u) \cdot t}.$$

Очевидно, что $G(u, t) \stackrel{df}{=} \frac{F(u, t)}{e^{C(u) \cdot t}}$ — гладкая функция. Подставляем $F(u, t) = G(u, t) \cdot e^{t \cdot C(u)}$ в дифференциальное уравнение:

$$\frac{\partial(G(u, t) \cdot e^{t \cdot C(u)})}{\partial t} = C(u) \cdot G(u, t) \cdot e^{t \cdot C(u)}.$$

Тогда

$$\left(\frac{\partial G(u, t)}{\partial t} \right) \cdot e^{t \cdot C(u)} + C(u) G(u, t) \cdot e^{t \cdot C(u)} = C(u) \cdot G(u, t) \cdot e^{t \cdot C(u)}.$$

Отсюда получаем:

$$e^{t \cdot C(u)} \cdot \frac{\partial G(u, t)}{\partial t} = 0.$$

А следовательно:

$$\frac{\partial G(u, t)}{\partial t} = 0.$$

Это уравнение является классическим квазилинейным дифференциальным уравнением. Запишем его уравнение характеристик:

$$\frac{dt}{1} = \frac{du}{0} = \frac{dG}{0}.$$

Находим характеристики:

$$u = C_1 \quad G = C_2.$$

Значит, $\Phi(u, G) = 0$. Тогда: $G(u, t) = C_1(u) \Rightarrow F(u, t) = C_1(u) \cdot e^{t \cdot C(u)}$.

Совершая обратный переход из модели в теорию, получаем:

$$f = C_1(u) \cdot e^{t \cdot C(u)} \mod \pi^*(I) = c_1 \cdot e^{c \cdot t},$$

здесь c_1 — произвольный элемент кольца R .

В случае, если $c = 0$, получаем, что $f = c_1$. То есть всякое решение уравнения $y' = 0$ есть константа. ■

Пример 3. Рассмотрим дифференциальное уравнение

$$d \cdot y'(t) = y(t), \tag{1}$$

где $d \in D \stackrel{df}{=} \{x \in R \mid x^2 = 0\}$.

Решение. Очевидно, что $y = 0$ является решением уравнения (1).

Теперь умножим на d обе части равенства. Тогда слева будет стоять выражение

$$d \cdot d \cdot y'(t), \tag{2}$$

а справа

$$d \cdot y(t).$$

Так как $d \cdot d$ равно нулю (по определению d), тогда

$$d \cdot y(t) = 0. \quad (3)$$

Взяв производную от выражения (3), получим, что

$$d \cdot y'(t) = 0. \quad (4)$$

Если $g : R \rightarrow R$ решение уравнения (1), то для него выполнено и уравнение (4). Отсюда следует, что $g = 0$. Таким образом, получили, что уравнение (1) имеет единственное решение, равное нулю. ■

Но не для всякой задачи Коши первого порядка можно найти точное решение.

Пример 4. Рассмотрим дифференциальное уравнение

$$d \cdot y'(t) = y^2(t), \quad (5)$$

где $d \in D \stackrel{df}{=} \{x \in R \mid x^2 = 0\}$.

Решение. Очевидно, что $y = 0$ является решением уравнения (5). Теперь умножим на d обе части равенства. Тогда слева будет стоять выражение

$$d \cdot d \cdot y'(t), \quad (6)$$

а справа

$$d \cdot y^2(t).$$

Так как $d \cdot d$ равно нулю (по определению d), тогда

$$d \cdot y^2(t) = 0. \quad (7)$$

Взяв производную от выражения (7), получим, что

$$2 \cdot d \cdot y(t) \cdot y'(t) = 0. \quad (8)$$

Если $g : R \rightarrow R$ решение уравнения (5), то для него выполнено и уравнение (8). Отсюда следует, что $g^3 = 0$. Таким образом, получили, что решением уравнения (5) является некоторая функция $g : R \rightarrow D_2$.

Кроме этого, для уравнения (5) можно указать в явном виде ненулевое решение

$$d_1 \cdot \exp \left(\int_0^t f(s) ds \right) + \frac{d}{2} \cdot f(t),$$

где d_1 – произвольный элемент D , а $f : R \rightarrow R$ – произвольная функция. Таким образом, добавление одного начального условия к уравнению (5) однозначно не определяет решение полученной задачи Коши. Следовательно, теорема о единственности решения для

$$\begin{aligned} d \cdot y'(t) &= y^2(t) \\ y(0) &= a, \quad a \in D_2 \end{aligned}$$

не выполняется. ■

Обобщая последние два примера, получим следующий

Принцип 1. Если $y : R \rightarrow R$ является решением уравнения

$$d \cdot y'(t) = y^n(t),$$

где $d \in D$, то $y^{2^n-1} = 0$.

Заменяя теперь D на D_k , получаем

Принцип 2. Если $y : R \rightarrow R$ является решением уравнения

$$d \cdot y'(t) = y^n(t),$$

где $d \in D_k$, то $y^{k \cdot n + (n-k)} = 0$.

ЛИТЕРАТУРА

1. Moerdijk I., Reyes G.E. *Models for Smooth Infinitesimal Analysis*. Springer-Verlag, 1991.
2. Kock A. *Synthetic Differential Geometry*. Cambridge University Press, 1981.
3. Kock A., Reyes G.E. *Aspects of Fractional Exponent Fctors* // Theory and Applications of Categories. 1999, V.5, No.10.

СВЯЗНОСТЬ ЛЕВИ-ЧЕВИТА В СИНТЕТИЧЕСКОЙ ДИФФЕРЕНЦИАЛЬНОЙ ГЕОМЕТРИИ КОКА-ЛОВЕРА

А.А. Звягинцев

In the article we consider correspondence between connection and the Riemannian metric on a microlinear spaces.

Цель данной работы – восполнить пробел, возникший в литературе по синтетической дифференциальной геометрии Кока-Ловера (СДГ). А именно дадим определение аффинной связности, согласованной с римановой метрикой, заданной на произвольном микролинейном пространстве, и выведем известное в классической римановой геометрии соотношение:

$$\Gamma_{jk}^i = \frac{1}{2} g^{li} \left(\frac{\partial g_{lj}}{\partial x^k} + \frac{\partial g_{lk}}{\partial x^j} - \frac{\partial g_{jk}}{\partial x^l} \right).$$

Вкратце напомним некоторые факты и определения из СДГ, большую часть которых можно найти в [2].

1. Основные определения

Аффинной связностью на микролинейном пространстве M будем называть отображение $\nabla : M^D \times_M M^D \rightarrow M^{D \times D}$ такое, что

$$\nabla(t_1, t_2)(d_1, 0) = t_1(d_1), \quad \nabla(t_1, t_2)(0, d_2) = t_2(d_2), \quad (1)$$

$$\nabla(\alpha \cdot t_1, t_2)(d_1, d_2) = \nabla(t_1, t_2)(\alpha \cdot d_1, d_2), \quad (2)$$

$$\nabla(t_1, \alpha \cdot t_2)(d_1, d_2) = \nabla(t_1, t_2)(d_1, \alpha \cdot d_2). \quad (3)$$

Для всех $(t_1, t_2) \in M^D \times_M M^D$, $d_i \in D$, $\alpha \in R$. Здесь R – коммутативное локальное кольцо (подробнее см. [2]), а $D = \{d \in R | d^2 = 0\}$.

Аффинную связность будем называть симметричной, если

$$\nabla(t_1, t_2)(d_1, d_2) = \nabla(t_2, t_1)(d_2, d_1).$$

В том случае, когда M совпадает с некоторым конечномерным R -модулем V , аффинная связность однозначно описывается функцией

$$\tilde{\nabla}_x : V \times V \rightarrow V,$$

определенной на главных частях касательных векторов пространства M в точке x . При этом

$$\nabla(t_1, t_2)(d_1, d_2) = x + b_1 d_1 + b_2 d_2 + \tilde{\nabla}_x(b_1, b_2)d_1 d_2. \quad (4)$$

Кроме того, если $\{e_1, \dots, e_n\}$ является базисом V , то

$$\tilde{\nabla}_x(e_i, e_j) = -\Gamma_{ij}^k(x)e_k. \quad (5)$$

Коэффициенты Γ_{ij}^k , как и в классическом случае, будем называть символами Кристоффеля второго рода.

Римановой метрикой на микролинейном пространстве M будем называть такое отображение $g : M^D \times_M M^D \rightarrow R$, для которого выполняются следующие условия (см. [1]):

$$t = 0 \Rightarrow g(t, t) = 0, \quad (6)$$

$$t \neq 0 \Rightarrow g(t, t) > 0, \quad (7)$$

$$g(t_1 + t, t_2) = g(t_1, t_2) + g(t, t_2), \quad (8)$$

$$g(t_1, t + t_2) = g(t_1, t) + g(t_1, t_2), \quad (9)$$

$$g(\alpha \cdot t_1, t_2) = g(t_1, \alpha \cdot t_2) = \alpha \cdot g(t_1, t_2) \quad (10)$$

для любых $t, t_1, t_2 \in M^D$ и $\alpha \in R$.

2. Связность Леви-Чевита

Мы непосредственно подошли к цели этой статьи и сейчас дадим следующее:

Определение 1. Если на микролинейном пространстве M задана риманова метрика g , то будем говорить, что аффинная симметричная связность ∇ согласована с метрикой g , если

$$g(\nabla(t, t_1)(d), \nabla(t, t_2)(d)) = g(t_1, t_2) \quad (11)$$

для любых $t, t_1, t_2 \in M^D$ и $d \in D$. Связность ∇ в этом случае будем называть связностью Леви-Чевита.

Теорема 1. Аффинная связность согласована с римановой метрикой на конечномерном R -модуле V тогда и только тогда, когда

$$\frac{\partial g_{ij}}{\partial x^k} - \Gamma_{ik}^l g_{lj} - \Gamma_{jk}^l g_{il} = 0. \quad (12)$$

Здесь g_{ik} — обозначение для $g(x, e_i, e_k)$.

Доказательство. Пусть $\{e_1, \dots, e_n\}$ — базис на V , тогда каждый касательный вектор пространства V в точке x можно представить как пару $(x, \alpha^i e_i)$. Ограничимся рассмотрением пар вида (x, e_i) . Заметим также из (4), что

$$\nabla((x, e_i), (x, e_j))(d) = (x + d \cdot e_i, e_j + d \cdot \nabla_x(e_i, e_j)).$$

Тогда

$$\begin{aligned}
 g(\nabla((x, e_k), (x, e_i))(d), \nabla((x, e_k), (x, e_j))(d)) &= g(x + d \cdot e_k, e_i - d \cdot \Gamma_{ik}^l e_l, e_j - d \cdot \Gamma_{jk}^l e_l) = \\
 &= g(x + d \cdot e_k, e_i, e_j - d \cdot \Gamma_{jk}^l e_l) - d \cdot \Gamma_{ik}^l g(x + d \cdot e_k, e_l, e_j - d \cdot \Gamma_{jk}^l e_l) = \\
 &= g(x + d \cdot e_k, e_i, e_j) - d \cdot \Gamma_{ik}^l g(x + d \cdot e_k, e_l, e_j) - d \cdot \Gamma_{jk}^l g(x + d \cdot e_k, e_i, e_l) = \\
 &= g(x, e_i, e_j) + d \cdot \frac{\partial g}{\partial x^k}(e_i, e_j) - d \cdot \Gamma_{ik}^l g(x, e_l, e_j) - d \cdot \Gamma_{jk}^l g(x, e_i, e_l) = g(x, e_i, e_j). \quad (13)
 \end{aligned}$$

Отсюда для любого $d \in D$ выполняется равенство

$$d \cdot \frac{\partial g}{\partial x^k}(e_i, e_j) - d \cdot \Gamma_{ik}^l g(x, e_l, e_j) - d \cdot \Gamma_{jk}^l g(x, e_i, e_l) = 0. \quad (14)$$

А значит, и

$$\frac{\partial g}{\partial x^i}(e_j, e_k) - \Gamma_{ik}^l g(x, e_j, e_l) - \Gamma_{ij}^l g(x, e_l, e_k) = 0 \quad (15)$$

или

$$\frac{\partial g_{ij}}{\partial x^k} - \Gamma_{ik}^l g_{lj} - \Gamma_{jk}^l g_{il} = 0. \quad (16)$$

■

Предложение 1. Если на конечномерном R -модуле V задана риманова метрика g , тогда на V существует и единственная симметричная аффинная связность, согласованная с этой метрикой. Эта связность в любой системе координат задается формулами

$$\Gamma_{jk}^i = \frac{1}{2} g^{li} \left(\frac{\partial g_{lj}}{\partial x^k} + \frac{\partial g_{lk}}{\partial x^j} - \frac{\partial g_{jk}}{\partial x^l} \right). \quad (17)$$

Доказательство. Доказательство совпадает с классическим. ■

ЛИТЕРАТУРА

1. Guts A.K., Grinkevich E.B. *Toposes in General Theory of Relativity*. Preprint gr-qc/9610073 (1996).
2. Moerdijk I., Reyes G.E. *Models for Smooth Infinitesimal Analysis*. Springer-Verlag, 1991.

ИСПОЛЬЗОВАНИЕ НЕЧЕТКИХ ОТНОШЕНИЙ ДЛЯ ФОРМАЛИЗАЦИИ ЗНАНИЙ

В.А. Маренко

The description of knowledge representation is given in terms of fuzzy sets.

Нечеткое отношение – некоторый критерий, по которому можно отличать одни упорядоченные пары элементов множества от других. Если перечень всех упорядоченных пар задан, то о каждой паре можно сказать, находится она в рассматриваемом отношении или нет. Рассмотрим два подмножества P_1 и P_2 , из множества P ключевых слов; пусть элементами первого подмножества являются ключевые слова p_i , второго – p_j . Множество упорядоченных пар (p_i, p_j) определяет прямое произведение $P_1 \times P_2$. Нечетким подмножеством $\underline{U}$ будет такое подмножество, что

$$\forall (p_i, p_j) \in P_1 \times P_2 : \mu_{\underline{U}(p_i, p_j)} \in M,$$

где M – множество значений функции принадлежности из интервала $[0, 1]$. Каждый элемент множества M – $\mu_{\underline{U}(p_i, p_j)}$ есть значение упорядоченной пары (p_i, p_j) , предложенное экспертом [1].

Подмножество ключевых слов в одном документе информационно-консультационной системы «Электромагнитная совместимость» (КС ЭМС) [2]: динамический диапазон, блокирование, помеха взаимная, а в другом – обстановка электромагнитная, интермодуляция. Сформируем $\underline{U}$. Это нечеткое отношение можно рассматривать как понимание экспертом «сходства» в решаемой задаче. Для каждой пары ключевых слов значения функции принадлежности – субъективная оценка экспертом степени сходства каждого элемента из этой пары. Их функции принадлежности, установленные экспертом, следующие:

$$\underline{U} = \{((p_{11})|0, 5), ((p_{12})|0, 5), ((p_{21})|0, 4), ((p_{22})|1), ((p_{31})|1), ((p_{32})|0, 5)\}.$$

Из этого нечеткого подмножества можно получить обычное подмножество путем упорядочения пар элементов в интервале $\{0, 1\}$:

$$U = \{((p_{11})|1), ((p_{12})|1), ((p_{21})|1), ((p_{32})|1)\}.$$

Пары p_{22} , и p_{31} не принадлежат полученному обычному подмножеству (их функции принадлежности равны нулю). Эксперт также указал на непринадлежность

этих пар ключевых слов (помеха взаимная, обстановка электромагнитная; блокирование, интермодуляция) U , мотивируя тем, что они находятся «как бы на другом уровне» по сравнению с остальными парами, то есть принадлежат другому подмножеству ключевых слов. Приведенное нечеткое отношение используется в КС ЭМС при формировании модели пользователя, его можно назвать одношаговым отношением в том смысле, что оно описывает результаты лишь попарного сравнения ключевых слов друг с другом.

Пусть $\mu_1(p_i, p_j)$ – функция принадлежности нечеткого бинарного отношения «сходство» на заданном наборе ключевых слов. Для каждой пары ключевых слов p_i, p_j значение $\mu_1(p_i, p_j)$ есть субъективная оценка человеком степени их сходства p_i, p_j . Это нечеткое отношение можно рассматривать как «экспериментальные данные», отражающие понимание человеком «сходства» в данной задаче. Следующим этапом использования этих «данных» становится этап классификации всех ключевых слов. Для $\mu_1(p_i, p_j)$ вводится n -шаговое отношение [3]

$$\mu_n(p_i, p_j) = \sup_{p_1, \dots, p_{n-1} \in P_i} \min[\mu_1(p_i, p_1), \dots, \mu_1(p_{n-1}, p_j)].$$

Для любых $p_i, p_j \in P_i$ выполняется цепочка неравенств

$$0 \leq \mu_1(p_i, p_j) \leq \mu_2(p_i, p_j) \leq \dots \leq \mu_n(p_i, p_j),$$

из которой следует, что для любых $p_i, p_j \in P_i$ последовательность $\{\mu_k(p_i, p_j)\}$ имеет предел при $k \rightarrow \infty$. Таким образом, существует предельное отношение сходства, определяемое равенством

$$\mu(p_i, p_j) = \lim_{k \rightarrow \infty} \mu_k(p_i, p_j) \forall p_i, p_j \in P_i.$$

Это предельное отношение является конечным продуктом обработки результатов нечетких измерений $\mu_1(p_i, p_j)$ и используется для классификации ключевых слов.

Отношение предпочтения возникает в результате формализации условий выбора. Это отношение используется в КС ЭМС в связи с выдачей необходимых документов по запросу пользователя. Проблема решается с помощью применения эталонного подхода к выбору наилучшей альтернативы. Эталоны назначаются извне пользователем КС ЭМС. Если необходимо выдать документ с определенным набором ключевых слов, то этот документ принимается за эталон, и поиск осуществляется по определенной схеме. Обозначим множества эталонов – Y , эталонное отношение – S , экспертные оценки – F' , нормированные экспертные оценки – F . Путем экспертной оценки выбирается для каждого x_i максимально близкий к нему эталон y_j при $i = \text{const}$ и оценивается степень их сходства единицей, а остальные оценки назначаются, исходя из этой максимально схожей пары. На основании закона взаимодействия отношений на множестве X строится нечеткое отношение предпочтения R и интерпретируется как степень выраженности представления о том, что элемент x «не хуже» элемента y .

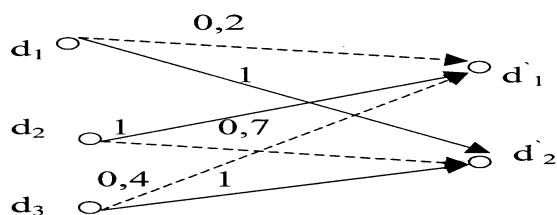


Рис. 1. Нормированное отношение в виде орграфа

Пример 1. Пусть дано подмножество документов $D_1 = \{d_1, d_2, d_3\}$, подмножество эталонов представлено двумя документами $D_2 = \{d'_1, d'_2\}$, на которых определен четкий линейный порядок:

$$S = \{((p_{11})|1), ((p_{12})|1), ((p_{21})|0), ((p_{22})|1)\}.$$

Заданы экспертные оценки степени сходства каждого документа с каждым эталоном:

$$\underline{F'} = \{((p_{11})|0.16), ((p_{12})|0.8), ((p_{21})|0.75), ((p_{22})|0.52)), ((p_{31})|0.36), ((p_{32})|0.9)\}.$$

После нормирования $\underline{F'}$ (рис.1) и в соответствии с законом взаимодействия отношений определим нечеткое отношение предпочтения на D и нечеткое отношение строгого предпочтения на D [4].

Из рис.1 следует, что из данных трех документов документ d_2 наиболее предпочтителен, так как он со степенью 1 схож с наиболее предпочтительным эталоном d'_1 . Этот вывод подтвержден и вычислительной процедурой отношения строгого предпочтения.

Язык нечетких отношений позволяет рассматривать широкий спектр качественных данных об исследуемых объектах. Опираясь на нечеткие отношения, определяются отношения сходства и различия документов в КС ЭМС, строится логический вывод.

ЛИТЕРАТУРА

1. Кофман А. *Введение в теорию нечетких множеств*. М.: Радио и связь, 1982.
2. Маренко В.А. , Маренко В.Ф. *Консультационная экспертная система «Электромагнитная совместимость»* // Сб.материалов международного информационного конгресса «МИК-2001», 2001.
3. Орловский С.А. *Добавление в книгу Заде Л. «Понятие лингвистической переменной и его применение к принятию приближенных решений»*. Мир, 1976. –168 с.
4. Кузьмин В.Б. *Построение групповых решений в пространстве четких и нечетких бинарных отношений*. М.:Наука, 1982. – 168с.

ОБ ОДНОМ МЕТОДЕ ОПТИМИЗАЦИИ НАЧАЛЬНЫХ УСЛОВИЙ В УПРАВЛЕНИИ ДИНАМИЧЕСКИМИ СИСТЕМАМИ

Б.К. Нартов

In the paper the new method's possibilities of the orientation optimization the begining conditions in the dinamical sistems control are presented.

Предложенный нами [1] метод направленной оптимизации начальных условий (метод возврата) предназначался первоначально для оптимизации вектора начальных координат в частной модели конфликта подвижных объектов, характеристики которых ухудшались в результате взаимодействия с объектами противника и старения. Модель связывала характеристики (в терминологии теории динамических систем – состояния) и координаты объектов дифференциальными уравнениями типа уравнений Ланчестера. Далее становилась и решалась конкретная задача оптимального управления движениями группы объектов, противодействующих другой группе объектов с заданными на интервале управления траекториями (по критерию минимизации некоторой функции конечных состояний объектов).

Существенно сложнее опорной оказалась задача построения приемлемого по времени счета и точности алгоритма оптимизации начального вектора управления, то есть начального размещения группы управляемых объектов. Найденный подход оказался весьма общим и, как недавно удалось показать, позволяет направленно оптимизировать начальный вектор управления по меньшей мере в классе управляемых гладких систем с отрицательной правой частью.

В самом общем виде идея метода состоит в том, что для оптимизации, в смысле избранного функционала качества, начальных условий исходной задачи оптимального управления записывается вспомогательная двойственная задача и реализуется итеративный процесс, в шагах которого чередуются исходная и двойственная задачи, а в качестве части начальных условий очередного шага итерации используется часть конечных значений предыдущего шага. При этом двойственная задача отличается от исходной обращением знаков правых частей исходной системы дифференциальных уравнений и знака функционала качества, а также обращением заданных временных процессов (например, заданных движений). Дополнительно используются лишь некоторые элементарные

© 2002 Б.К. Нартов

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

Омский филиал Института математики СО РАН

Работа выполнена при финансовой поддержке РФФИ (проекты 01-01-00303 и 01-07-90003).

правила коррекции части конечных значений четных шагов и ограничений на вектор управления исходной задачи. (Таким образом, если в конкретной задаче управления состояние суть монотонно, точно ухудшающаяся физическая характеристика, схему процесса оптимизации начальных условий можно представить в виде последовательности решений чередующихся задач «исчерпывания» (прямая) и «восстановления» (обратная).)

Ограничимся здесь рассмотрением следующей динамической системы:

$$\dot{P}_i = f_i(\bar{p}(t), \bar{a}(t), \bar{u}_i(t)), \quad i = \overline{1, N}, \quad (1)$$

где f_i – функция, непрерывная на заданном интервале управления $(0, t_f)$; $\bar{p}(t) = (p_1(t), \dots, p_N(t))$;

$\bar{a}(t)$ – заданные на $(0, t_f)$ временные процессы;

$\bar{u}_i(t) = (u_{i1}(t), u_{i2}(t), u_{i3}(t))$ – i -ое управление, координата в R^3 .

Не оговаривая ограничений на управления, запишем для заданных начальных условий задачу оптимального управления

$$J(\bar{p}(t_f)) \rightarrow \inf, \quad (2)$$

где J непрерывна по $(p_1(t), \dots, p_N(t))$. Определив далее

$$\tilde{f}_i = -f_i,$$

$$\tilde{a}(t) = \bar{a}(t_f - t),$$

рассмотрим динамическую систему

$$\dot{\tilde{P}}_i = \tilde{f}_i(\tilde{p}(t), \tilde{a}(t), \tilde{u}_i(t)), \quad i = \overline{1, N} \quad (3)$$

с начальными условиями

$$\tilde{p}(0) = \bar{p}(t_f),$$

$$\tilde{u}(0) = \bar{u}(t_f), \quad i = \overline{1, N}$$

и запишем для (3) задачу оптимального управления, двойственную задаче (2):

$$J(\tilde{p}(t_f)) \rightarrow \sup. \quad (4)$$

Задав теперь ограничения на управления

$$|\dot{u}_{ij}| < c_{ij}, |\dot{\tilde{u}}_{ij}| < c_{ij}, \quad i = \overline{1, N}, \quad j = \overline{1, 3}$$

и рассмотрев последовательность решений задачи (2) – первый шаг – и (4) – второй шаг, – можно заметить, что

$$J(\tilde{p}(t_f)) \geq J(\bar{p}(0)). \quad (5)$$

Очевидно, что на втором шаге достигается по меньшей мере равенство функционалов, для чего достаточно обратить оптимальные управления, найденные на первом шаге:

$$\tilde{u}_i(t) = \bar{u}_i^*(t_f - t), \quad i = \overline{1, N}.$$

Далее нас интересует реализация строгого неравенства

$$J(\tilde{\bar{p}}(t_f)) > J(\bar{p}(0)). \quad (6)$$

Примечательно, что (6) выполняется при весьма общих предположениях. Достаточно потребовать для двойственной обратной задачи (4) существования кривых, смещение по которым произвольных неоптимальных начальных управлений (начальных координат) монотонно улучшает функционал качества – при сохранении начального вектора состояния (в частных случаях это условие удастся существенно ослабить). Дополнительно в доказательстве используется лишь ограниченность функций f и J на интервале управления $(0, t_f)$. Заметим также, что размерность векторов $\bar{U}_i(t)$ вполне произвольна, а трехмерное пространство управления в (1) и (3) назначено лишь из соображений наглядности.

Повторяя теперь приведенные рассуждения и условия реализации (6) для исходной задачи (2), мы получаем итеративный процесс (2), (4), (2), (4), (2)... с монотонным возрастанием $\Delta J = J(0) - J(t_f)$ исходной задачи.

Существование предела ΔJ и построенные нами алгоритмы сохранения и параллельной оптимизации начального вектора состояния исходной задачи требуют отдельного обсуждения, однако возможности прямых эвристических реализаций представленного подхода достаточно очевидны.

ЛИТЕРАТУРА

1. Nartov B.K. *Conflict of Moving Systems*. AMSE Press, France, 1994. 87 p.

УСЛОВИЯ СУЩЕСТВОВАНИЯ СТАЦИОНАРНОГО РЕЖИМА В РЕАКТОРЕ С КИПЯЩИМ СЛОЕМ КАТАЛИЗАТОРА

Р.К. Романовский, И.Д. Макарова, С.Е. Макаров

The article considers the boundary value problem, simulating the process in chemical reactor with the boiling layer of catalyst. The conditions of existence of stationary solutions are established.

При математическом моделировании реакторов идеального вытеснения возникают краевые задачи для гиперболических систем [1–5]. В частности, процесс в реакторе с кипящим слоем катализатора при реакции первого порядка (скорость реакции линейно зависит от концентрации реагирующего вещества) моделируется [2, 3] краевой задачей для почти линейной гиперболической системы на плоскости

$$\left\{ \begin{array}{l} \frac{\partial C}{\partial t} + \frac{\partial C}{\partial x} = \alpha(1 - C), \\ \beta \frac{\partial \theta}{\partial t} + \frac{\partial \theta}{\partial x} = \gamma(1 - C)e^{\theta} - \delta(\theta - \theta_r), \\ \frac{\partial \theta_r}{\partial t} - \frac{\partial \theta_r}{\partial x} = \delta(\theta - \theta_r), \quad (x, t) \in \Pi, \\ (\theta - \theta_r)_{x=0} = 0, \quad C|_{x=0} = 0, \quad \theta_r|_{x=1} = \theta_0, \\ (C, \theta, \theta_r)_{t=0} \quad \text{заданы.} \end{array} \right. \quad (1)$$

Здесь Π – полуполоса $(0, 1) \times (0, \infty)$, C – концентрация реагирующего вещества, θ, θ_r – температуры в реакторе и холодильнике, $\alpha, \beta, \gamma, \delta, \theta_0$ – константы, из них первые четыре положительны. Предполагаются выполненными условия согласования нулевого и первого порядков в точках $(0, 0), (1, 0)$.

Корректная разрешимость краевых задач такого класса в различных функциональных пространствах обсуждалась в ряде работ [5–9]. В данной работе исследуются условия существования стационарных решений задачи (1). Попутно предложена процедура построения стационарных решений. Полученные результаты существенно дополняют результаты из [1] для задачи (1).

© 2002 Р.К. Романовский, И.Д. Макарова, С.Е. Макаров

E-mail: makarov@math.omsu.omskreg.ru

Омский государственный технический университет

Омский государственный университет

Подставляя в (1) $C = u_0(x)$, $\theta = u_1(x)$, $\theta_r = u_2(x)$, получим: $u_0 = 1 - e^{-\alpha x}$,

$$\begin{cases} \frac{du_1}{dx} = \gamma e^{u_1 - \alpha x} - \delta(u_1 - u_2), \\ \frac{du_2}{dx} = -\delta(u_1 - u_2), \quad x \in (0, 1), \\ u_1(0) = u_2(0), \quad u_2(1) = \theta_0. \end{cases} \quad (2)$$

Введем параметр $\Delta = \alpha + \delta/\gamma e^{\theta_0}$.

Теорема 1. Для разрешимости краевой задачи (2) и тем самым для существования хотя бы одного стационарного решения краевой задачи (1) необходимо выполнение неравенства

$$\rho = \frac{\alpha}{\gamma e^{\theta_0}} + e^{-\alpha} > 1 \quad (3)$$

и достаточно выполнение неравенств

$$\alpha > \gamma e^{\theta_0} > 1, \quad \frac{\alpha + \ln \Delta}{\alpha + \delta} > \frac{e^\delta - 1}{\delta}, \quad \rho \leq 1 + \delta \left(1 - \frac{1}{\gamma e^{\theta_0}}\right). \quad (4)$$

Заметим, что из первого неравенства (4) следует, в частности, (3), $\Delta > 1$, а также второе неравенство (4) при достаточно малых $\delta > 0$.

Доказательство. 1. Из (2) легко получить

$$e^{-(u_1 - u_2)} \frac{d(u_1 - u_2)}{dx} = \gamma e^{u_2 - \alpha x} \geq \gamma e^{\theta_0 - \alpha x}.$$

Интегрируя это неравенство по отрезку $[0, 1]$, получим

$$1 - e^{u_2(1) - u_1(1)} \geq \gamma e^{\theta_0} \frac{1 - e^{-\alpha}}{\alpha}.$$

Для выполнения этого неравенства необходимо, чтобы правая часть была меньше 1, что эквивалентно (3).

Отметим, что из (2) следуют соотношения

$$\varepsilon = e^{u_1(0) - \theta_0} > 1, \quad T = u_1(1) - u_2(1) > 0, \quad u_1 - u_2 \in [0, T]. \quad (5)$$

2. Выполняя в (2) замену $(u_1, u_2) \rightarrow (u, v)$ по формулам

$$u = u_1 - u_2, \quad v = \gamma e^{u_1 - \alpha x}, \quad (6)$$

получим

$$\begin{cases} \frac{du}{dx} = v, \quad \frac{dv}{dx} = v(v - \delta u - \alpha), \quad v > 0, \\ u(0) = 0, \quad v(1) = \gamma e^{\theta_0 - \alpha + T}, \end{cases} \quad (7)$$

где T – постоянная (5). Очевидно, краевая задача (7) эквивалентна, с учетом (6), задаче (2). Далее нетрудно убедиться, что (7) эквивалентна краевой задаче на $[0, T]$ с неизвестными $v(u), x(u), T$:

$$\begin{cases} \frac{dv}{du} = v - \delta u - \alpha, & \frac{dx}{du} = \frac{1}{v}, & v > 0, \\ v(T) = \gamma e^{\theta_0 - \alpha + T}, & x(0) = 0, & x(T) = 1. \end{cases} \quad (8)$$

Задача (8) представляет собой подкласс задач со свободной границей. Покажем, что при условиях (4) эта задача разрешима.

Первое неравенство (5) в обозначениях (6) имеет вид

$$\varepsilon = \frac{v(0)}{\gamma e^{\theta_0}} > 1. \quad (9)$$

Нетрудно убедиться, что функция

$$v(u) = \delta u + \alpha + \delta + \gamma e^{\theta_0}(\varepsilon - \Delta)e^u \quad (10)$$

удовлетворяет первому уравнению (8) и начальному условию (9). С учетом (10) первое краевое условие (8) принимает вид

$$\delta T + \alpha + \delta = \gamma e^{\theta_0}(\Delta + e^{-\alpha} - \varepsilon)e^T. \quad (11)$$

Наложим на начальную константу (9) дополнительное требование $\Delta < \varepsilon < \Delta + e^{-\alpha}$. Тогда $v > 0$ и уравнение (11) имеет точно один положительный корень T . Подставляя (10) во второе уравнение (8), интегрируя по отрезку $[0, u]$ и учитывая второе краевое условие (8), получим

$$x(u) = \int_0^u \frac{ds}{v(s)}, \quad u \in [0, T]. \quad (12)$$

Из выполненного построения следует: при каждом $\varepsilon \in (\Delta, \Delta + e^{-\alpha})$ тройка $T, v(u), x(u)$, где $T = T(\varepsilon)$ – положительный корень уравнения (11), $v(u), x(u)$ – функции (10), (12), удовлетворяет всем соотношениям (8), кроме, быть может, третьего краевого условия. Для завершения доказательства теоремы достаточно показать: существует $\varepsilon \in (\Delta, \Delta + e^{-\alpha})$, при котором выполняется и это условие.

С учетом (12) третье краевое условие принимает вид

$$J(\varepsilon) = \int_0^{T(\varepsilon)} \frac{ds}{v(s, \varepsilon)} = 1, \quad (13)$$

где $v(u, \varepsilon)$ – функция (10). Так как функция $J(\varepsilon)$ непрерывна на $[\Delta, \Delta + e^{-\alpha})$, достаточно показать: при условиях (4) имеют место неравенства

$$J(\Delta) > 1, \quad \lim_{\varepsilon \uparrow \Delta + e^{-\alpha}} J(\varepsilon) < 1. \quad (14)$$

Имеем

$$J(\Delta) = \int_0^{T(\Delta)} \frac{ds}{\delta s + \alpha + \delta} = \frac{1}{\delta} \ln \left(1 + \frac{\delta}{\alpha + \delta} T(\Delta) \right). \quad (15)$$

Из второго неравенства (4) следует

$$1 + \frac{\delta}{\alpha + \delta} T_0 > e^\delta, \quad T_0 = \alpha + \ln \Delta. \quad (16)$$

Число $T(\Delta)$ – корень уравнения (11) при $\varepsilon = \Delta$, число T_0 – корень уравнения, получаемое из него отбрасыванием первого слагаемого, откуда следует $T(\Delta) > T_0$. С учетом этого из (15), (16) следует первое неравенство (14). Далее, так как $T(\varepsilon) \rightarrow \infty$ при $\varepsilon \uparrow \Delta + e^{-\alpha}$, то

$$\lim_{\varepsilon \uparrow \Delta + e^{-\alpha}} J(\varepsilon) = \int_0^\infty \frac{ds}{v(s, \Delta + e^{-\alpha})} < \frac{e^\alpha}{\gamma e^{\theta_0}} \int_0^\infty \frac{ds}{e^s + \Delta e^\alpha} = \frac{\alpha + \ln(e^{-\alpha} + \Delta)}{\alpha + \delta}. \quad (17)$$

Из последнего неравенства (4) легко получить: $\ln(e^{-\alpha} + \Delta) \leq \ln(1 + \delta) \leq \delta$. Отсюда и из (17) следует второе неравенство (14). Теорема доказана. ■

Отметим, что изложенное доказательство содержит процедуру построения решения краевой задачи (8). Параметры $\varepsilon, T(\varepsilon)$ могут быть найдены численно из системы уравнений (11), (13), затем функции $v(u), x(u)$ находятся по формулам (10), (12).

ЛИТЕРАТУРА

1. Зеленьяк Т.И. *О стационарных решениях смешанных задач, возникающих при изучении некоторых химических процессов* // Дифференц. уравнения. 1966. Т.2, N.2. С.205–213.
2. Шеплев В.С., Мещеряков В.Д. *Математическое моделирование реакторов с кипящим слоем катализатора* // В кн.: Математическое моделирование химических реакторов. Новосибирск: Наука. Сиб. Отд. 1984. С.44–65.
3. Иванов Е.А. *Управление процессом в реакторе с псевдоожиженным слоем* // Там же. С.116–127.
4. Акрамов Т.А. *Качественный и численный анализ модели реактора с противотоком компонентов* // Математическое моделирование каталитических реакторов. Новосибирск: Наука. 1989. С.195–214.
5. Лаврентьев М.М.(мл.), Люлько Н.А. *Повышение гладкости решений некоторых гиперболических задач* // Сиб. матем. журнал. 1997. Т.38, N.1. С.109–124.
6. Аболиня В.Э., Мышкис А.Д. *Смешанная задача для почти линейной гиперболической системы на плоскости* // Мат. сб. 1960. Т.50, N.4. С.423–442.
7. Годунов С.К. *Уравнения математической физики*. М.: Наука. 1979.
8. Елтышева Н.А. *К вопросу об устойчивости стационарных решений некоторых гиперболических систем* // Докл. АН СССР. 1986. Т.289, N.1. С.30–32.
9. Воробьева Е.В., Романовский Р.К. *Метод характеристик для гиперболических краевых задач на плоскости* // Сиб. матем. журнал. 2000. Т.41, N.3. С.531–540.

ОЦЕНКА ПРИБЛИЖЕННОГО РЕШЕНИЯ АЛГОРИТМОМ ХВАТАЛА ЗАДАЧИ О ПОКРЫТИИ

А.В. Пролубников

In the paper we consider the Set Cover Problem. This problem is NP-hard[2]. So it becomes very important to find approximate solution of this problem in polynomial time. Algorithm developed by V. Chvatal is such algorithm and it is most effective algorithm developed for set cover problem. Followed statement true for this algorithm[1]:

$$\frac{c(Chv)}{c(Opt)} \leq \ln m + 1.$$

We have found another proof for this statement. Our proof is more simple and vivid by our opinion.

В задаче о покрытии даны: конечное множество $U, |U| = m$ и конечный набор множеств $P = \{P_1, \dots, P_n\}$ таких, что

$$\forall j \ P_j \subseteq U, \bigcup_{j=1}^n P_j = U.$$

Будет рассматриваться также набор индексов $I = \{1, 2, \dots, n\}$, соответствующий P .

Набор множеств $P' = \{P_{j_1}, \dots, P_{j_p}\}$ ($P_{j_i} \in P$) называется *покрытием* множества U , если $\bigcup_{i=1}^p P_{j_i} = U$. Набору множеств P' соответствует набор индексов входящих в него множеств $I' = \{j_1, \dots, j_p\}$.

Задан также набор положительных чисел $\{c_1, \dots, c_n\}$ *стоимостей* множеств из P . Иначе говоря, задана аддитивная неотрицательная функция $c : P \rightarrow R_+$. Далее будем обозначать $c_j = c(P_j)$. *Стоимость* набора множеств $P' = \{P_{j_1}, \dots, P_{j_p}\}$ равна

$$c(P') = \sum_{i=1}^p c(P_{j_i}) = \sum_{i=1}^p c_{j_i}.$$

Целью является отыскание покрытия минимальной стоимости.

Задача о покрытии является NP-трудной [2]. Следовательно, важным является поиск приближенных решений задачи за полиномиальное время. Широко

известным алгоритмом отыскания приближенного решения задачи о покрытии является алгоритм Хватала, для которого верна [1] оценка

$$\frac{c(Chv)}{c(Opt)} \leq H(|A|) \leq H(m) \leq \ln m + 1, \quad (1)$$

где Chv – решение, которое дает алгоритм Хватала, Opt – оптимальное решение задачи, m – мощность покрываемого множества U , $|A|$ – максимальная мощность множеств из оптимального покрытия, $H(p) = \sum_{k=1}^p \frac{1}{k}$.

Алгоритм В. Хватала является полиномиальным алгоритмом, который находит приближенное решение задачи о покрытии. В схеме алгоритма Хватала множества в покрытие (набор множеств Chv) выбираются соответственно минимальному отношению $\frac{c(P_j)}{|P_j|}$ – удельной стоимости множества следующим образом:

Алгоритм Хватала

Шаг 0. $Chv := \emptyset$;

Шаг 1. Если Chv – покрытие U , то работу завершить. Иначе – **Шаг 2**.

Шаг 2. Выбираем k так, что

$$\frac{c_k}{|P_k|} = \min_j \left(\frac{c_j}{|P_j|} : P_j \in P, P_j \not\subset \cup(P_l : l \in Chv) \right);$$

$Chv := Chv \cup \{P_k\}, \forall P_j : P_j := P_j \setminus P_k$. Перейти на **Шаг 1**.

Для решений, находимых алгоритмом Хватала, верна оценка:

$$\frac{c(Chv)}{c(Opt)} \leq H(|A|) \leq H(m), \quad (2)$$

где $|A|$ – максимальная мощность множеств из оптимального покрытия. Существует несколько вариантов доказательства этой оценки, например доказательство самого В. Хватала [1]. В доказательстве В. Хватала задача о покрытии рассматривается как задача целочисленного линейного программирования. Нами получено доказательство, где задача о покрытии рассматривается как комбинаторная задача. И хотя можно сказать, что получена более слабая оценка для решений, отыскиваемых алгоритмом Хватала, а именно оценка:

$$\frac{c(Chv)}{c(Opt)} \leq H(m),$$

полученное доказательство более наглядно демонстрирует мажорирование отношения $\frac{c(Chv)}{c(Opt)}$ гармоникой $H(m)$. Доказательство основано на ключевом соотношении (**лемма 2**):

$$\frac{c_i}{|P_i|} \leq \frac{c(Opt_i)}{m_i},$$

где P_i – множество, которое выбирается алгоритмом на i -й итерации; Opt_i – оптимальное покрытие части множества U , непокрытой к i -й итерации работы алгоритма, m_i – мощность этой части множества U .

Лемма 1. Пусть $a, b, c, d > 0$. Тогда $\frac{a}{b} \leq \frac{c}{d} \Rightarrow \frac{a}{b} \leq \frac{a+c}{b+d} \leq \frac{c}{d}$.

Доказательство. $\frac{a}{b} \leq \frac{c}{d} \Leftrightarrow ad \leq bc$;

$$\begin{aligned} \frac{a}{b} &= \frac{a}{b} \cdot \frac{a+c}{a+c} \cdot \frac{b+d}{b+d} = \frac{a+c}{b+d} \cdot \frac{a(b+d)}{b(a+c)} = \frac{a+c}{b+d} \cdot \frac{ab+ad}{ab+bc} \leq \frac{a+c}{b+d} \cdot \frac{ab+bc}{ab+bc} = \\ &= \frac{a+c}{b+d}; \\ \frac{c}{d} &= \frac{c}{d} \cdot \frac{a+c}{a+c} \cdot \frac{b+d}{b+d} = \frac{a+c}{b+d} \cdot \frac{c(b+d)}{d(a+c)} = \frac{a+c}{b+d} \cdot \frac{bc+cd}{ad+cd} \geq \frac{a+c}{b+d} \cdot \frac{ad+cd}{ad+cd} = \\ &= \frac{a+c}{b+d}. \end{aligned}$$

■

Следствие 1. $\frac{a_1}{b_1} \leq \frac{a_2}{b_2} \leq \dots \leq \frac{a_k}{b_k} \Rightarrow \frac{a_1}{b_1} \leq \frac{\sum_{i=1}^k a_i}{\sum_{i=1}^k b_i}$.

■

Пусть $(P_1, P_2, \dots, P_l)$ – покрытие, которое дает алгоритм Хватала, и алгоритм Хватала выбирает множество P_i на i -й итерации, $(A_1, A_2, \dots, A_k)$ – оптимальное покрытие, a_i – стоимость множества A_i .

Лемма 2. $\frac{c_1}{|P_1|} \leq \frac{c(Opt)}{m} = \frac{\sum_{i=1}^k c(A_i)}{m}$.

Доказательство. Положим $A'_1 = A_1$, $A'_2 = A_2 \setminus A_1$, ..., $A'_k = A_k \setminus \left(\bigcup_{i=1}^{k-1} A_i \right) \Rightarrow A'_i$ – попарно непересекающиеся. Упорядочим множества A_i и введем новую индексацию в соответствии с неубыванием отношения $\frac{a_i}{|A'_i|}$:

$$\frac{a_1}{|A'_1|} \leq \frac{a_2}{|A'_2|} \leq \dots \leq \frac{a_k}{|A'_k|} \Rightarrow \frac{a_1}{|A'_1|} \leq \frac{\sum_{i=1}^k a_i}{\sum_{i=1}^k |A'_i|} = \frac{c(Opt)}{m} \text{ (по лемме 1). По}$$

алгоритму $\frac{c_1}{|P_1|} \leq \frac{a_1}{|A_1|} = \frac{a_1}{|A'_1|} \leq \frac{c(Opt)}{m}$.

■

Теорема 1.

$$\frac{c(Chv)}{c(Opt)} \leq H(m).$$

Доказательство. Пусть на i -й итерации алгоритм Хватала выбирает множество P_i^i , где P_i^i – множество P_i к i -й итерации: $P_i^i = P_i \setminus (\bigcup_{k=1}^{i-1} P_k)$; обозначим $p_i = |P_i^i| \Rightarrow \sum_{i=1}^l p_i = m$. Обозначим Opt_i – оптимальное покрытие для i -й задачи (т.е. задачи на i -й итерации). По **лемме 2**

$$\frac{c_1}{p_1} \leq \frac{c(Opt)}{m}, \quad \frac{c_2}{p_2} \leq \frac{c(Opt_2)}{m - p_1}, \quad \dots, \quad \frac{c_l}{p_l} \leq \frac{c(Opt_l)}{m - p_1 - p_2 - \dots - p_{l-1}}.$$

Так как $c(Opt) \geq c(Opt_2) \geq \dots \geq c(Opt_l)$, то

$$\begin{aligned} c(Chv) &= \sum_{i=1}^l c_i \leq \left(\frac{p_1}{m} + \frac{p_2}{m - p_1} + \dots + \frac{p_l}{m - p_1 - p_2 - \dots - p_{l-1}} \right) \cdot c(Opt) \leq \\ &\leq \left[\left(\frac{1}{m} + \frac{1}{m - 1} + \dots + \frac{1}{m - p_1 + 1} \right) + \left(\frac{1}{m - p_1} + \frac{1}{m - p_1 - 1} + \dots \right. \right. \\ &\dots + \left. \frac{1}{m - p_1 - p_2 + 1} \right) + \dots + \left(\frac{1}{m - p_1 - \dots - p_{l-1}} + \frac{1}{m - p_1 - \dots - p_{l-1} - 1} + \right. \\ &\left. \dots + \frac{1}{m - p_1 - p_2 - \dots - p_l + 1} \right) \Big] \cdot c(Opt) = \left[\frac{1}{m} + \frac{1}{m - 1} + \dots + \frac{1}{2} + \right. \\ &\left. + 1 \right] \cdot c(Opt) = H(m) \cdot c(Opt). \quad \blacksquare \end{aligned}$$

ЛИТЕРАТУРА

1. Chvatal V. *A greedy heuristic for the set-covering problem* // Mathematics of operation research. 1979. V.4, N 3. P.233–235.
2. Гэри М. Джонсон Д. *Вычислительные машины и труднорешаемые задачи*. М.: Мир, 1982.

ЭВРИСТИЧЕСКИЙ АЛГОРИТМ ДЕШИФРОВКИ ШИФРА ДВОЙНОЙ ПЕРЕСТАНОВКИ

А.В. Пролубников, Р.Т. Файзуллин

In the paper we consider heuristic algorithm for solving graph isomorphism problem. The algorithm based on a successive splitting of the eigenvalues of the matrices which are modifications (to positive defined) of graphs' adjacency matrices. Modification of the algorithm allows to find a solution for Frobenius problem. Formulation of the Frobenius problem is following one. Given a pair of two matrices with the same number of rows and columns. We must find out whether one of the matrix can be acquired from another by permutation of it's rows and strings or not. Solution of Frobenius problem can give to us efficient way for decrypting of double permutation cyphers problem for high dimension matrices.

1. Изоморфизм графов

В работе представлен эвристический алгоритм решения задачи определения изоморфности графов, основанный на последовательном расщеплении собственных чисел, видоизмененных (до положительно определенных) матриц смежности, и решении систем линейных уравнений, определяющих обратные матрицы. Сравнение норм столбцов обратных матриц позволяет получить решающую перестановку.

Задача проверки изоморфности графов принадлежит к задачам, относительно которых нет ясности: являются ли они полиномиально разрешимыми или нет [1]. В то же время известно, что задача определения изоморфности графа является полиномиально разрешимой для некоторых классов графов, в частности для планарных, регулярных графов и некоторых других построены эффективные алгоритмы, для решения этой задачи [2], [3], [4]. Предлагаемый нами алгоритм является эвристическим алгоритмом для проверки изоморфности графов.

В задаче даны два неориентированных графа $G_A = \langle V_A, E_A \rangle$ и $G_B = \langle V_B, E_B \rangle$, где V_A, V_B – множества вершин графов, E_A, E_B – множества ребер. Предполагается, что $|V_A| = |V_B|$, $|E_A| = |E_B|$. Задача изоморфизма графов формулируется следующим образом: существует ли биективное отображение $\varphi : V_A \rightarrow V_B$, такое, что если $(i, j) \in E_A$, то $(\varphi(i), \varphi(j)) \in E_B$?

© 2002 А.В. Пролубников, Р.Т. Файзуллин

E-mail: e-mail: faizulin@univer.omsu.ru, prolubnikov@math.omsu.omscreg.ru

Омский государственный университет

Предлагаемый алгоритм для решения задачи определения изоморфности графов работает с видоизмененными матрицами смежности графов.

Пусть A_0 – матрица смежности G_A , то есть $A_0 = (a_{ij}^0)$, где

$$a_{ij}^0 = \begin{cases} 1, & \text{если } (i, j) \in E_A, \\ 0, & \text{иначе.} \end{cases}$$

B_0 – матрица смежности графа G_B .

По матрице A_0 строим матрицу D_{A_0} :

$$\begin{pmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_n \end{pmatrix}.$$

D_{A_0} – диагональная матрица со следующими элементами:

$$d_i = \sum_{j=1}^n a_{ij}^0 + 1.$$

Аналогично строится матрица D_{B_0} по матрице смежности графа G_B .

Рассматриваемые далее матрицы

$$A = A_0 + D_{A_0}, \quad B = B_0 + D_{B_0} \quad (1)$$

– матрицы, с которыми будет работать алгоритм, являются симметричными положительно определенными матрицами.

Если графы $G_A = \langle V_A, E_A \rangle$ и $G_B = \langle V_B, E_B \rangle$ изоморфны, то соответствующие им матрицы описанного типа могут быть получены одна из другой перестановкой строк с одновременной перестановкой столбцов с теми же номерами. Таким образом, задача проверки изоморфности двух графов может быть сформулирована как частный случай задачи Фробениуса: может ли быть получена матрица B из матрицы A последовательной перестановкой строк с одновременной перестановкой столбцов?

Для произвольной матрицы перестановка строк с номерами i и j эквивалентна ее умножению на матрицу перестановки P_{ij} . Перестановка столбцов матрицы эквивалентна ее умножению справа на ту же матрицу. При умножении вектора на матрицу P_{ij} как справа, так и слева происходит перестановка компонент вектора с номерами i и j .

Рассмотрим две системы уравнений следующего вида:

$$Ax = e_j, \quad By = e_k, \quad (2)$$

где векторы $e_i = (0, \dots, 0, 1, 0, \dots, 0)$ – базисные векторы в пространстве R^n , матрицы A и B – описанного выше вида. Обе системы уравнений имеют решение, и решение единственно, так как A и B – матрицы с диагональным преобладанием, и, следовательно, их определители не равны нулю. Пусть далее x^j –

решение системы линейных уравнений $Ax = e_j$, y^k – решение системы линейных уравнений $Bu = e_k$.

Отметим, что, решив системы уравнений (1), получим обратные матрицы для A и B . Так, для i -й компоненты вектора x^j верно: $x_i^j = (-1)^{i+j} A_{ij} / \det(A)$, где A_{ij} – алгебраическое дополнение элемента a_{ij} матрицы A . То есть векторы решений x^j , $j = 1, \dots, n$ являются столбцами обратной к A матрицы.

Если $B = P_{jk}AP_{jk}$, то для решения систем уравнений (2) должны выполняться равенства:

$$x_i = y_i, \quad i \neq j, \quad i \neq k;$$

$$x_j = y_k, \quad x_k = y_j.$$

Действительно: $(Ax = e_j) \sim (P_{jk}Ax = P_{jk}e_j) \sim (P_{jk}AxP_{jk} = P_{jk}e_jP_{jk}) \sim (P_{jk}AP_{jk}x = e_j) \sim (P_{jk}AP_{jk}xP_{jk} = e_jP_{jk}) \sim (BxP_{jk} = e_k)$. То есть $xP_{jk} = y$.

Если матрица B получена из матрицы A многократной одновременной перестановкой строк и столбцов, то:

$$B = P_{j_l k_l} \dots P_{j_1 k_1} A P_{j_1 k_1} \dots P_{j_l k_l},$$

и, соответственно, $xP_{j_1 k_1} \dots P_{j_l k_l} = y$.

Таким образом, если мы будем менять вектор e_k в системе уравнений (2) при фиксированном j , то есть индекс k будет пробегать значения $1, \dots, n$, то векторы x^j и y^k – соответствующие друг другу решения полученных систем (2) будут совпадать с точностью до перестановки компонент только в том случае, если строке j матрицы A соответствует строка k матрицы B . То есть элементы строки k матрицы B есть переставленные элементы строки i матрицы A . То же верно и для столбцов матриц.

При нахождении в матрице B строки и столбца, соответствующих строке и столбцу с номером i матрицы A , на каждой i -й итерации алгоритма будем последовательно производить возмущения ее диагональных элементов. Алгоритм работает с симметричными матрицами, которые могут быть приведены к диагональной форме с помощью ортогональных преобразований. То есть

$$\tilde{A} = U_A A U_A^T,$$

$$\tilde{B} = U_B B U_B^T,$$

где $\tilde{A}$, $\tilde{B}$ – диагональные матрицы с собственными значениями на диагонали, а U_A , U_B – матрицы ортогональных преобразований. Диагональные элементы $\tilde{A}$, $\tilde{B}$ – собственные числа матриц A и B .

Спектры матриц смежности изоморфных графов совпадают [5]. То же можно сказать и о спектрах матриц, с которыми работает алгоритм, так как описанная выше процедура замены нулевых диагональных элементов матрицы смежности приводит лишь к сдвигу спектра матрицы. Причем, если матрицы соответствуют изоморфным графам, то спектры матриц будут совпадать и после сдвига.

Очевидно, если спектр каждой матрицы является простым, то есть среди собственных значений матрицы нет кратных, то задача определения изоморфности графов разрешима однозначно путем сопоставления матриц $\tilde{A}$, $\tilde{B}$, U_A и U_B .

Основные трудности возникают при рассмотрении графов, спектры которых содержат кратные собственные значения. Возмущая матрицы в ходе итераций алгоритма, мы будем получать возмущение спектра матриц, при котором происходит расщепление кратных собственных значений, что позволяет установить однозначное соответствие между строками и столбцами матриц.

В итоге, если среди собственных значений матриц A и B есть кратные, то в ходе работы алгоритма они будут расщеплены, и будет возможно получение перестановки, задающей искомое отображение φ , устанавливающее изоморфизм графов $G_A = \langle V_A, E_A \rangle$ и $G_B = \langle V_B, E_B \rangle$. Численные эксперименты показывают, что расщепление спектра матриц, необходимое для определения соответствующих друг другу строк и столбцов матриц, происходит значительно раньше заключительной итерации.

В ходе работы алгоритма на каждой итерации будем работать не с исходными, но с уже возмущенными в ходе предыдущих итераций алгоритма матрицами. Так, получив соответствие между строкой j матрицы A^j и строкой k матрицы B^j на итерации j , а также столбцами с теми же номерами, рассматриваем далее возмущенные матрицы A^{j+1} и B^{j+1} :

$$A^{j+1} = A^j + \varepsilon C^j, \quad B^{j+1} = B^j + \varepsilon C^k.$$

Возмущения производим с помощью диагональных матриц C^k с диагональными элементами

$$c_i = \begin{cases} 1, & \text{если } i = j = k, \\ 0, & \text{иначе.} \end{cases}$$

В результате, если матрица B может быть получена из матрицы A последовательными одновременными перестановками строк и столбцов, то пока j пробегает значения от 1 до n , будет получена, вообще говоря, одна из возможных перестановок строк и столбцов матрицы A :

$$\begin{pmatrix} 1 & 2 & \dots & n \\ k_1 & k_2 & \dots & k_n \end{pmatrix},$$

где k_j – номер строки матрицы B , полученный на j -й итерации работы алгоритма.

Перестановка задает такую перенумерацию вершин графа G_A , то есть отображение $\varphi : V_A \rightarrow V_B$, при которой матрицы A и B , представляющие графы, совпали бы, что возможно только в случае изоморфности графов.

Алгоритм спектрального расщепления проверки изоморфности графов

Шаг 0. $A^0 := A$, $j := 1$.

Шаг 1. Если $j < n$, то $A^j := A^{j-1} + \varepsilon C^j$, иначе работу алгоритма завершить.

Шаг 2. Решение системы уравнений $Ax = e_j$. x^j — полученное решение.

Шаг 3. $k := 1$. Если $k < n$, то — Шаг 3.1, иначе перейти на Шаг 4.

Шаг 3.1. $B^k := B^{k-1} + \varepsilon C^k$.

Шаг 3.2. Решение системы уравнений $B^k y = e_k$. y^k — полученное решение.

Шаг 3.3. $k := k + 1$. Перейти на Шаг 3.

Шаг 4. Сравнение норм x^j и y^k , $k = 1, \dots, n$.

Если $\forall k \ ||x^j|| \neq ||y^k||$, то графы G_A и G_B неизоморфны. Работу алгоритма завершить.

Если $\exists k : ||x^j|| = ||y^k||$ и $x_j^j = y_k^k$, то вершине j графа G_A ставим в соответствие вершину k графа G_B . Иначе графы G_A и G_B неизоморфны. Работу алгоритма завершить.

Шаг 5. $j := j + 1$. Перейти на Шаг 1.

Трудоемкость представленной схемы алгоритма равна $O(n^4)$, где n — число строк в квадратных матрицах A и B , подающихся на вход алгоритма. Следует отметить, что с добавлением процедуры проверки спектра матриц A и B на расщепленность схема алгоритма может быть модифицирована до схемы с трудоемкостью $O(n^{3.5})$.

2. Решение задачи Фробениуса для произвольных квадратных матриц полного ранга

Алгоритм спектрального расщепления проверки изоморфности неориентированных графов может быть применен без каких-либо модификаций для проверки изоморфности взвешенных неориентированных графов. Схема алгоритма остается в точности той же, модифицируются только представляющие графы матрицы, с которыми работает алгоритм.

В задаче даны два неориентированных графа $G_A = \langle V_A, E_A \rangle$ и $G_B = \langle V_B, E_B \rangle$, где V_A, V_B — множества вершин графов, E_A, E_B — множества ребер. Предполагается, что $|V_A| = |V_B|$, $|E_A| = |E_B|$. На множествах ребер графов G_A и G_B определены функции $H_A : E_A \rightarrow R$ и $H_B : E_B \rightarrow R$, задающие веса ребер. Задача определения изоморфизма взвешенных графов формулируется следующим образом: существует ли биективное отображение $\varphi : V_A \rightarrow V_B$, такое, что если $(i, j) \in E_A$, то $(\varphi(i), \varphi(j)) \in E_B$ и $H_A(i, j) = H_B(\varphi(i), \varphi(j))$?

Матрицы смежности взвешенных неориентированных графов также преобразуются до положительно определенных матриц с диагональным преобладанием, но при этом диагональные элементы подбираются так, чтобы числа обусловленности матриц, с которыми работает алгоритм, были ограничены.

Пусть $A_0 = (a_{ij}^0)$ — матрица смежности взвешенного ориентированного графа G_A . D_{A_0} — диагональная матрица с элементами d_i :

$$d_i = d + \sum_{j=1}^n a_{ij}^0,$$

где $d = \max_{1 \leq i \leq n} \sum_{j=1}^n a_{ij}^0$. Таким же образом строится матрица D_{B_0} для взвешенного неориентированного графа G_B по его матрице смежности B_0 .

Для числа обусловленности $\mu(A)$ симметричной матрицы A справедлива следующая оценка [6]:

$$\mu(A) \leq \frac{\eta(A)}{\chi(A)},$$

где $\eta(A) = \max_{1 \leq i \leq n} (a_{ii} + \sum_{j \neq i} |a_{ij}|)$, $\chi(A) = \min_{1 \leq i \leq n} (a_{ii} - \sum_{j \neq i} |a_{ij}|)$.

При нашем выборе d

$$\begin{aligned} \eta(A) &= \max_{1 \leq i \leq n} (a_{ii} + \sum_{j=1}^n |a_{ij}^0|) = a_{i_1 i_1} + \sum_{j=1}^n |a_{i_1 j}^0| = d_{i_1} + \sum_{j=1}^n |a_{i_1 j}^0| = d + 2 \sum_{j=1}^n |a_{i_1 j}^0| = \\ &= 3 \sum_{j=1}^n |a_{i_1 j}^0|, \end{aligned}$$

$$\begin{aligned} \chi(A) &= \min_{1 \leq i \leq n} (a_{ii} - \sum_{j=1}^n |a_{ij}^0|) = a_{i_2 i_2} - \sum_{j=1}^n |a_{i_2 j}^0| = d_{i_2} - \sum_{j=1}^n |a_{i_2 j}^0| = \sum_{j=1}^n |a_{i_2 j}^0| + d - \\ &- \sum_{j=1}^n |a_{i_2 j}^0| = \sum_{j=1}^n |a_{i_2 j}^0| + \sum_{j=1}^n |a_{i_1 j}^0| - \sum_{j=1}^n |a_{i_2 j}^0| = \sum_{j=1}^n |a_{i_1 j}^0|. \end{aligned}$$

Следовательно,

$$\mu(A) = \frac{\eta(A)}{\chi(A)} \leq \frac{3 \sum_{j=1}^n |a_{i_1 j}^0|}{\sum_{j=1}^n |a_{i_1 j}^0|} = 3.$$

В задаче Фробениуса даны две матрицы F_A и F_B с одинаковым числом строк и столбцов. Требуется определить, может ли одна из матриц быть получена из другой при помощи некоторой перестановки ее строк и столбцов.

Будем считать, что матрицы F_A и F_B — квадратные матрицы полного ранга с числом строк, равным n . Матрица A_0 вида

$$\begin{pmatrix} 0 & F_A \\ F_A^T & 0 \end{pmatrix}$$

может рассматриваться как матрица смежности некоторого взвешенного неориентированного графа. Построив для матрицы A_0 матрицу D_{A_0} описанного выше вида, ставим в соответствие исходной матрице F_A матрицу A :

$$A = A_0 + D_{A_0}.$$

То есть

$$A = \begin{pmatrix} D_{A_0}^1 & F_A \\ F_A^T & D_{A_0}^2 \end{pmatrix}$$

и

$$D_{A_0} = \begin{pmatrix} D_{A_0}^1 & 0 \\ 0 & D_{A_0}^2 \end{pmatrix},$$

где

$$D_{A_0}^1 = \begin{pmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_n \end{pmatrix}, \quad D_{A_0}^2 = \begin{pmatrix} d_{n+1} & 0 & \dots & 0 \\ 0 & d_{n+2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_{2n} \end{pmatrix}.$$

Итак, A — симметричная положительно определенная матрица с диагональным преобладанием, соответствующая некоторому двудольному взвешенному неориентированному графу. Матрица F_B — квадратная матрица с тем же числом строк, что и матрица F_A . Аналогично ставим в соответствие матрице F_B матрицу B . Применяя к матрицам A и B описанный выше алгоритм, сформируем решающую перестановку строк и столбцов матриц A и B .

По перестановкам строк и столбцов матриц A и B формируем перестановки строк и столбцов исходных матриц F_A и F_B следующим образом. Пусть алгоритмом получено соответствие строки и столбца номер i матрицы A столбцу и строке номер j матрицы B . Очевидно, что если $1 \leq i \leq n$, $1 \leq j \leq n$, то перестановка пары строк $\{i, j\}$ и столбцов $\{i, j\}$ матрицы A соответствует перестановке строк i и j исходной матрицы F_A . Если $n+1 \leq i \leq 2n$, $n+1 \leq j \leq 2n$, то перестановка пары строк $\{i, j\}$ и столбцов $\{i, j\}$ матрицы A соответствует перестановке столбцов i и j исходной матрицы F_A . Так как матрицы A и B , с которыми работает алгоритм, имеют вид:

$$\begin{pmatrix} * & 0 & \dots & 0 & * & * & \dots & * \\ 0 & * & \dots & 0 & * & * & \dots & * \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & * & * & * & \dots & * \\ * & * & \dots & * & * & 0 & \dots & 0 \\ * & * & \dots & * & 0 & * & \dots & 0 \\ \vdots & \vdots & & * & \vdots & \vdots & \ddots & \vdots \\ * & * & \dots & * & 0 & 0 & \dots & * \end{pmatrix},$$

где $*$ обозначены позиции, допустимые для ненулевых элементов, то ситуация, когда алгоритм поставил бы в соответствие друг другу пару строк и столбцов матриц A и B с номерами $\{i, j\}$ и $1 \leq i \leq n$, $n+1 \leq j \leq 2n$ невозможна, так как после такой перестановки строк и столбцов, если исходные матрицы F_A и F_B не являются диагональными, в левой верхней части матрицы A (в матрице $D_{A_0}^1$) и правой нижней части (в матрице $D_{A_0}^2$) появились бы кроме диагональных дополнительные ненулевые элементы, чего не может быть по построению матриц A и B . Следовательно, получение такого соответствия в ходе работы алгоритма невозможно.

3. Дешифрование шифра двойной перестановки при помощи алгоритма спектрального расщепления

Пусть исходный текст представлен некоторой квадратной матрицей символов. Формируем матрицу шифра, где каждому символу текста поставлен в соответствие его некоторый числовой код. Вторая матрица представляет собой также некоторый кодированный таким же образом текст. Требуется определить, является ли текст, представленный второй матрицей, шифром двойной перестановки текста, представленного первой матрицей. То есть требуется определить, является ли он текстом, полученным из первой текстовой матрицы при помощи перестановок ее строк и столбцов. Такая задача может быть интерпретирована как задача Фробениуса для пар числовых матриц, и, следовательно, для ее решения может быть применен алгоритм спектрального расщепления проверки изоморфности графов.

В ходе вычислительного эксперимента установлено, что для успешного решения задач проверки изоморфности неориентированных невзвешенных и взвешенных графов при решении систем линейных уравнений достаточно 60-ти итераций метода Зейделя для достижения необходимой точности при том, что расщепление спектра происходит за число итераций, не превышающее числа групп кратных собственных значений матриц A и B , подаваемых на вход алгоритма. Для задач проверки изоморфности графов проведены вычислительные эксперименты, в частности, для регулярных графов с числом вершин до 2500. Для задач дешифрования шифра двойной перестановки проведены вычислительные эксперименты с текстами с числом символов до 5000.

ЛИТЕРАТУРА

1. Гэри М., Джонсон Д. *Вычислительные машины и труднорешаемые задачи*. М.: Мир, 1982.
2. Hopcroft, Wong *A linear time algorithm for isomorphism of planar graphs*. Proceedings of the Sixth Annual ACM Symposium on Theory of Computing. 1974. P.172–184.
3. Luks *Isomorphism of graphs of bounded valence can be tested in polynomial time*. Proc. 21st IEEE FOCS Symp., 42, 49, 1980.
4. Hoffmann *Group-Theoretic Algorithms and Graph Isomorphism*. Lecture Notes in Computer Science (Chapter V). 1982. P.127–138.
5. Цветкович Д. и др. *Спектры графов. Теория и применение*. Киев: Наукова думка, 1984.
6. Годунов С.К. и др. *Гарантированная точность решения систем линейных уравнений в евклидовых пространствах*. Новосибирск: Наука, 1988.

ОБ УЛЬТРАМЕТРИЧЕСКИХ ТЕТРАЭДРАХ И ТРЕУГОЛЬНИКАХ

Р.Р. Файзуллин

In this article relationships between ultrametric spaces and spaces of negative curvature are considered.

В данной работе представлено доказательство двух предложений о пространствах отрицательной кривизны.

1. Общие утверждения

Определение 1. Ультраметрическое пространство (УМП) – это метрическое пространство, в котором неравенство треугольника заменяется более сильным неравенством:

$$\rho(A, B) \leq \max(\rho(A, C), \rho(B, C)). \quad (1)$$

Предложение 1. Метрическое пространство U является ультраметрическим тогда и только тогда, когда каждая тройка его точек при вложении в S_k^2 есть подмножество окружности. Когда три точки не подчиняются ультраметрике, то составленный из них треугольник можно вписать в S_k^2 не для всякого k . Пусть даны три точки в S_k^2 и они изометричны подмножеству УМП. Тогда при любом k они лежат на окружности. ■

Предложение 2. Любые четыре точки УМП изометрически вкладываются в S_k^3 , образуя при любом k в S_k^3 вписанный в сферу тетраэдр. И обратно, если метрическое пространство U таково, что каждая четверка его точек изометрически вложена в подмножество сферы в S_k^3 при любом $k \leq 0$, то U есть УМП.

Если же тетраэдр в S_k^3 вписывается в шар, то его вершины вложены из УМП. ■

Предварительно докажем несколько необходимых геометрических фактов о тройках и четверках точек УМП.

Мы будем пользоваться теоремой косинусов и известным свойством углов параллельности в пространстве Лобачевского: $\cos P_k(a) = \text{th}(ka)$ [1].

Найдем соотношения между тройками точек УМП и соответствующие им треугольники. Определим, какие ограничения (1) накладывает на эти треугольники. Рассмотрим четыре типа треугольников, изображенных на рис. 1, а)-г).

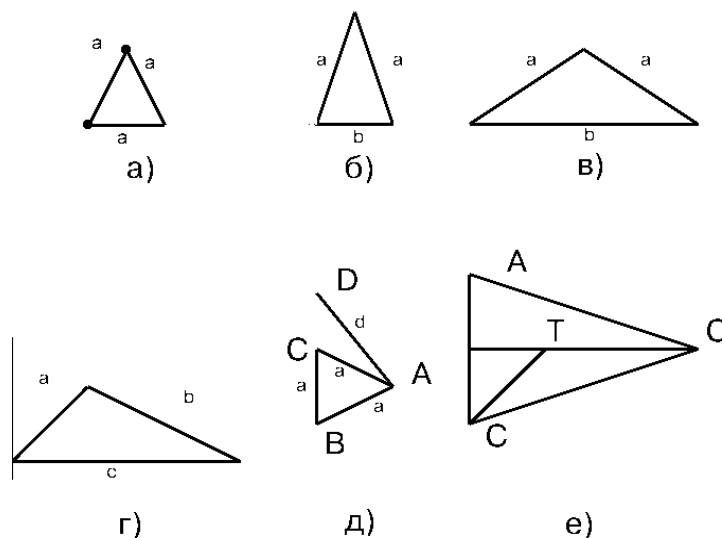


Рис. 1.

Из них ультраметрическому неравенству будут удовлетворять треугольники следующих двух типов:

Тип 1. Равносторонние треугольники: сторона $a = a = a$ (рис.1,а).

Тип 2. Равнобедренные с основанием меньшим, чем боковая сторона: $a > b$ (рис.1,б).

Рассматриваем также два типа треугольников, для которых не выполняется ультраметрическое неравенство:

Тип 3. Равнобедренные с основанием большим, чем боковая сторона: $b > a$ (рис.1,в).

Тип 4. Треугольники с тремя различными сторонами: $c > \max(a, b) = b$ (рис.1, г).

Итак, из четырех выделенных типов треугольников допустимыми мы считаем два типа треугольников – 1-й и 2-й. Точно так же найдем типы тетраэдров, которые можно составить из допустимых треугольников.

Проведем построение по шагам, отсеивая тетраэдры, для вершин которых не выполняется неравенство (1).

На первом шаге выбираем тип треугольника ABC . Вспомнив классификацию треугольников, замечаем, что целесообразно рассмотреть два варианта.

Вариант 1.

Первый шаг. Рассматриваем треугольник ABC типа 1, сторона $AD = d$ (рис.1,д).

Второй шаг. Вариант разобьем на подварианты:

1а) $AD > a$. Следовательно, по вышеизложенной классификации $\triangle ADC$ 2-го типа, $DA = DC = d$, $DB = DC = d$ из рассмотрения $\triangle DBC$.

1б) $AD = a$, $AD = AC = a$.

Удобно рассмотреть три случая:

1б₁) $DC = a$. По (1) $DB \leq a$.

1б₂) $DC < a$, значит, или $DB = BC$ или $DC = BC$ (из подчинения вершин $\triangle DBC$ неравенству (1)).

1б₃) $DC > a$ и $\triangle ADC$ 3-го типа, таким образом, случай 1б₃) противоречит (1).

1в) $AD < a$, и по подчинению вершин $\triangle ADC$ неравенству (1) выполняется $DC = AC$. А по рассмотрению $\triangle DBC$ верно $DB = DC$.

Вариант 2.

Первый шаг. Рассматриваем треугольник ABC типа 2, сторона $a = AB = AC > BC = b$.

Второй шаг. Вариант также разобьем на подварианты:

2а) $AD < a$ из рассмотрения $\triangle ADC$, по (1), $DC = a$. Из рассмотрения $\triangle ADB$, по (1), $DB = AD$. Вершины тетраэдра $ABCD$ не противоречат (1) и AD с BC в любом отношении.

2б) $AD = a$, по (1), $DC \leq a$, $DB \leq a$, $BC \leq a$. Не ограничивая общности: $DB = DC \geq BC$

2в) $AD > a$. Из (1) для $\triangle ADC$ и $\triangle ADB$ $AD = DB = DC \geq AB = AC \geq BC$.

Итак, рассмотрев эти тетраэдры, для вершин которых выполняется неравенство (1), – будем называть ультраметрическими – выделим два типа тетраэдров, полностью описывающие все тетраэдры с вершинами, изометричными точкам УМП.

Тип I: $AD = DC = DB = d \geq a = AB = AC \geq BC = b$.

Тип II: $AD = d \leq DC = DB = AB = AC = a \geq BC = b$. Отношения между AD и BC произвольные.

2. Доказательство предложения 1

Необходимость. Без потери общности будем считать, что $a > b \geq c$. Если окружность для этого треугольника существует, то ее центр – это пересечение срединных перпендикуляров. Мы докажем, что срединные перпендикуляры, проведенные к b и c , не пересекаются. Достаточно будет показать, что они не пересекают биссектрису угла A . Рассмотрим углы параллельности отрезков M_1A , M_2A . Надо сравнить $A/2$ с $P_k(c/2)$. Так как $\text{th}(t)$ монотонно стремится снизу к 1 при $t \rightarrow \infty$, то

$$\cos(P_k(c/2)) = \text{th}(kc/2) \leq \text{th}(kb/2) = \cos(P_k(b/2)). \quad (2)$$

То есть при $k \rightarrow \infty$ $P_k(c/2) \rightarrow +0$.

Следовательно, нужно показать, что $\cos(A_k/2) < \cos(P_k(c/2))$ при больших k . Это будет достаточным, так как из (2) следует, что $P_k(c/2) > P_k(b/2)$. По

теореме косинусов:

$$\cos(A_k) = \frac{\cosh(kb) \cosh(kc) - \cosh(ka)}{\sinh(kb) \sinh(kc)}. \quad (3)$$

Так как A, B, C – точки метрического пространства, расстояния между ними подчиняются неравенству треугольника, то $a < b + c$, $\cos(A_k) \rightarrow 1$, $\lim_{k \rightarrow \infty} A_k = 0$.

Сравним при возрастающем k числа $\cos(A_k/2) = \sqrt{(1 + \cos(A_k))/2}$ и $\text{th}(kc/2)$. Покажем, что

$$\sqrt{0.5 + \frac{\cosh(kb) \cosh(kc) - \cosh(ka)}{2 * \sinh(kb) \sinh(kc)}} < \text{th}\left(\frac{kc}{2}\right). \quad (4)$$

Очевидно, что обе части неравенства (4) положительны (далее используем символы сравнения – $\bigvee, \bigvee$) и

$$\frac{\sinh kb \sinh kc + \cosh kb \cosh kb - \cosh ka}{2 \sinh kb \sinh kc} \bigvee \frac{\exp(2kc) - 2 \exp(kc) + 1}{\exp(2kc) + 2 \exp(kc) + 1}, \quad (5)$$

$$\frac{2 \exp(kb + kc) + 2 \exp(-kb - kc) - 2(\exp(ka) + \exp(-ka))}{2(\exp(kb) - \exp(-kb))(\exp(kc) - \exp(-kc))} \bigvee \left(\frac{\exp(kc) - 1}{\exp(kc) + 1} \right)^2. \quad (6)$$

Упростим эти выражения, введя слагаемые:

$$\begin{aligned} C_1(k) &= \frac{\exp(-ak)(\exp(2ck) + 2 \exp(ck) + 1)}{(\exp(kc) - \exp(-kc))(\exp(kb) - \exp(-kb))(\exp(kc) + 1)^2}, \\ C_2(k) &= \frac{\exp(ak)(2 \exp(ck) + 1)}{(\exp(kc) - \exp(-kc))(\exp(kb) - \exp(-kb))(\exp(kc) + 1)^2}, \\ C_3(k) &= \frac{\exp(bk)(\exp(ck) + \exp(-ck) - 2)}{(\exp(kc) - \exp(-kc))(\exp(kb) - \exp(-kb))(\exp(kc) + 1)^2}, \\ C_4(k) &= \frac{\exp(-bk)(\exp(3ck) - 2 \exp(2ck) + \exp(ck) + 4)}{(\exp(kc) - \exp(-kc))(\exp(kb) - \exp(-kb))(\exp(kc) + 1)^2}, \\ C_5(k) &= \frac{-\exp(ak + 2ck) + 4 \exp(bk + 2ck)}{(\exp(kc) - \exp(-kc))(\exp(kb) - \exp(-kb))(\exp(kc) + 1)^2}. \end{aligned}$$

Рассмотрим сравнение:

$$C_1(k) + C_2(k) + C_3(k) + C_4(k) + C_5(k) \bigvee 0. \quad (7)$$

При $kc > 0$ функция $\exp(kc) - \exp(-kc) > 0$, т.е. знак неравенства (7) при умножении на $(\exp(kc) - \exp(-kc))$ не изменится.

Получаем

$$(\exp(kc) - \exp(-kc))(C_1(k) + C_2(k) + C_3(k) + C_4(k) + C_5(k)) \bigvee 0. \quad (8)$$

Теперь рассмотрим поведение слагаемых в формуле (8) при увеличении k , при этом учтем, что $a > b \geq c$. У нас возникают два набора слагаемых.

14) Слагаемые, стремящиеся к 0 при $k \rightarrow +\infty$. Это $(C_1(k), C_2(k), C_3(k), C_4(k))$:

$$\begin{aligned}\lim_{k \rightarrow +\infty} \frac{\exp(-ak)(\exp(2ck) + 2\exp(ck) + 1)}{(\exp(kb) - \exp(-kb))(\exp(kc) + 1)^2} &= 0, \\ \lim_{k \rightarrow +\infty} \frac{\exp(ak)(2\exp(ck) + 1)}{(\exp(kb) - \exp(-kb))(\exp(kc) + 1)^2} &= 0, \\ \lim_{k \rightarrow +\infty} \frac{\exp(bk)(\exp(ck) + \exp(-ck) - 2)}{(\exp(kb) - \exp(-kb))(\exp(kc) + 1)^2} &= 0, \\ \lim_{k \rightarrow +\infty} \frac{\exp(-bk)(\exp(3ck) - 2\exp(2ck) + \exp(ck) + 4)}{(\exp(kb) - \exp(-kb))(\exp(kc) + 1)^2} &= 0.\end{aligned}$$

Обоснование: $-a+c < -a+2c < b+2c$; $a < a+c < b+2c$; $b-c < b+c < b+2c$; $-b+c < -b+2c < -b+3c < b+2c$;

5) Слагаемое $(\exp(kc) - \exp(-kc))C_5$, которое не стремится к 0 при $k \rightarrow +\infty$. Заметим, что $a+2c > b+2c$. А $(-\exp(ak+2ck) + 4\exp(bk+2ck))/((\exp(kb) - \exp(-kb))(\exp(kc) + 1)^2)$ при возрастании k есть функция, эквивалентная функции $-\exp(ak+2ck-bk-2ck) = -\exp(ak-bk)$ и $\lim_{k \rightarrow +\infty} [-\exp(ak-bk)] = -\infty$.

Это значит, что с некоторого k_0 знак в (7) меньше нуля. Следовательно, когда кривизна пространства менее $-k_0^2$, то срединные перпендикуляры к b и c не могут пересекать идущую между ними прямую – биссектрису угла A . Значит, они тем более не пересекаются между собой. Отсюда следует, что точки A, B, C не лежат на окружности.

Достаточность. Найдем точку M : $|MA| = |MB| = |MC|$. Без ограничения общности считаем, что $|AB| = c \leq |AC| = |BC| = a$ (см. рис.1,е).

Опустим перпендикуляр из C на AB . Точка T произвольна: $T \in CH$. Найдем функцию $f_8(t) = |CT| - |BT|$, где t есть расстояние от T до C . $f_8(t) = t - |BT(t)|$. Используем теорему синусов геометрии Лобачевского для треугольника HCB . Надо показать, что $CH > BH$. Рассмотрим $\triangle ABC$, в нем против большей стороны лежит больший угол.

В треугольнике ABC $\angle ACB$ против основания, а угол $\angle ABC$ против боковой стороны. Вспомним нашу классификацию треугольников. Для 1-го и 2-го типов треугольников это значит, что $\angle ACB \leq \angle ABC$, $\angle HCB = 1/2\angle ACB < \angle ABC$. Видим, что в $\triangle HCB$ сторона CH , лежащая против $\angle ACB$, больше стороны BH , лежащей против $\angle HCB$. Следовательно, $f_8(|CH|) > 0$. В то же время $f_8(t)$, являясь композицией непрерывных функций, принимает на концах промежутка $[0, |CH|]$ разные знаки. По теореме Больцано-Вейерштрасса на $(0, |CH|)$ есть точка t_0 такая, что $f_8(t_0) = 0$. Таким образом, найдена искомая точка M , она лежит на расстоянии t_0 от C . Так как CH срединный перпендикуляр к AB , то расстояния от $\forall T \in CP$ до A и B совпадают. А так как $f_8(t_0) = 0$, то совпадает расстояние от M до C и B . Таким образом, точки A, B, C принадлежат окружности.

3. Доказательство предложения 2

Рассмотрим функции

$$f(u) = \frac{\operatorname{ch}^2(u) - \operatorname{ch}(u)}{\operatorname{sh}^2(u)}$$

и

$$\cos(\gamma) = f_1(u, a) = \frac{\operatorname{ch}^2(u) - \operatorname{ch}(a)}{\operatorname{sh}^2(u)}.$$

Необходимость.

Надо доказать, что:

- 1) Ультраметрические тетраэдры вкладываются в S_k^3 при каждом k .
- 2) Эти тетраэдры – подмножества сфер.

Докажем утверждение 1). Для этого необходимо и достаточно показать, что справедливы следующие утверждения:

- а) Сумма углов при каждой вершине $< 2\pi$.
- б) Для углов при вершине верно неравенство треугольника: $(\alpha \leq \beta + \gamma)$.

Вычислим косинусы углов (γ_0) для треугольников 1-го типа (d), 2-го типа ($d > b, \gamma < \gamma_1$). Рассмотрим

$$\cos(\gamma_0) = f(u) = \frac{\operatorname{ch}^2(u) - \operatorname{ch}(u)}{\operatorname{sh}^2(u)}. \quad (9)$$

Так как $\lim_{u \rightarrow 0} f(u) = 1/2$, а $u = |d\sqrt{K}|$, $f'(u) > 0$ при $u > 0$, $f((0, +\infty)) = (1/2, 1)$, $f(u) = \cos(\gamma_0)$, значит, $\gamma_0 \in (0, \pi/3)$.

Это вычисление было проведено для равностороннего треугольника.

Для треугольника 2-го типа γ – это угол при вершине, $\cos(\gamma) = f_1(u, a)$,

$$f_1(u, a) = \frac{\operatorname{ch}^2(u) - \operatorname{ch}(a)}{\operatorname{sh}^2(u)}, \quad (10)$$

$f_1(u, a) \geq f(u)$ при $\operatorname{ch}(a) \leq \operatorname{ch}(u)$. Для этого неравенства достаточно того, что основание треугольника b меньше боковой стороны d . Теперь рассмотрим полную производную функции $f_1(u, a)$ по кривизне k

$$f'_1(u, a) = \frac{2 \operatorname{ch}(u)(\operatorname{ch}(a) - 1)}{\operatorname{sh}(u) \cdot (\operatorname{ch}(u)^2 - 1)} \cdot d - \frac{\operatorname{sh}(a)}{\operatorname{sh}^2(u)} \cdot b > 0. \quad (11)$$

Функция $f_1(u(k), a(k))$ возрастает по k при $k > 0$. Мы вспомнили также, что u и a есть функции от k .

Очевидно, что $f_1([a, +\infty), a) = [f(a), 1)$ и $\lim_{k \rightarrow +\infty} f_1(u(k), a(k)) = 1$. Следовательно, $\cos(\gamma) \in (f(a), 1)$, $\gamma \in (0, \arccos(f(a)))$, $\arccos(f(a)) \in (0, \pi/3)$.

Используем равенство

$$\cos(\gamma_1) = f_3(u, a) = \frac{\operatorname{ch}(u) \operatorname{ch}(a) - \operatorname{ch}(u)}{\operatorname{sh}(u) \operatorname{sh}(a)}. \quad (12)$$

Сравним $f_3(u, a)$ с $f(u)$:

$$\frac{\operatorname{ch}(u) \operatorname{ch}(a) - \operatorname{ch}(u)}{\operatorname{sh}(u) \operatorname{sh}(a)} \vee \frac{\operatorname{ch}^2(u) - \operatorname{ch}(u)}{\operatorname{sh}^2(u)}. \quad (13)$$

Для решения этой задачи надо исследовать функцию

$$f_4(t) = \frac{\operatorname{ch}(t) - 1}{\operatorname{sh}(t)}. \quad (14)$$

Также, учитывая, что $f'_4(t) = (\operatorname{ch}(t) - 1)/(\operatorname{sh}^2(t)) > 0$ при $t > 0$, по (1) $f_4(a) \leq f_4(u)$. А $f_3(u, a) \leq f(u)$. Угол $\gamma_1 \in (\arccos f_4(a), \arccos f(a)]$. Получены пределы и отношения для углов треугольников 1 и 2 типов: угол $\gamma_0 \in (0, \pi/3]$. Угол $\gamma \in (0, \arccos(f(a)))$, $\gamma_1 \in (\arccos f_4(a), \arccos f(a)]$. Отсюда следует, что $\gamma_i \in (0, \pi/2)$, $\gamma \in (0, \pi/3)$,

$$0 < \gamma_0 < \gamma < \gamma_1 < \pi/2. \quad (15)$$

Докажем пункт (а). Выберем любую вершину, три ее плоских угла α, β, η подчиняются (15), значит, $\alpha + \beta + \eta < 3\pi/2 < 2\pi$.

Доказательство пункта (б) опирается на проведенную выше классификацию тетраэдров по двум типам.

Тип I. Здесь $d = DA = DC = DB \geq a = AB = AC \geq BC = b$ и $u = \sqrt{|K|}$. Проведем доказательство по вершинам.

Вершина A. Тривиальные неравенства $\angle DAB + \angle BAC \geq \angle DAC = \angle DAB$; $\angle DAC + \angle BAC \geq \angle DAB = \angle DAC$. Сравним $\cos(BAC)$ и $\cos(DAB)$ по величине. Имеем

$$\cos(BAC) = \frac{\operatorname{ch}(ua) \operatorname{ch}(ua) - \operatorname{ch}(ub)}{\operatorname{sh}(ua) \operatorname{sh}(ua)}, \quad (16)$$

$$\cos(DAB) = \frac{\operatorname{ch}(ud) \operatorname{ch}(ua) - \operatorname{ch}(ud)}{\operatorname{sh}(ud) \operatorname{sh}(ua)}. \quad (17)$$

Рассмотрим цепочку неравенств

$$\frac{\operatorname{ch}^2(ua) - \operatorname{ch}(ub)}{\operatorname{sh}^2(ua)} \geq \frac{\operatorname{ch}^2(ua) - \operatorname{ch}(ua)}{\operatorname{sh}^2(ua)} \geq \frac{\operatorname{ch}(ud) \operatorname{ch}(ua) - 1}{\operatorname{sh}(ud) \operatorname{sh}(ua)}. \quad (18)$$

Докажем последнее неравенство. После упрощения достаточно исследовать $f_5(t) = \operatorname{ch}(t)/\operatorname{sh}(t)$. При $t > 0$

$$f'_5(t) = \frac{-4 \exp(2t)}{(\exp(2t) - 1)^2} < 0. \quad (19)$$

Следовательно, функция $f_5(t)$ убывает при $t > 0$. И $ud \geq ua$, следовательно, $f_5(ud) \leq f_5(ua)$, $\cos(BAC) \geq f_5(ua) \geq \cos(DAB) > 0$, значит, $\angle BAC < \angle DAB < 2\angle DAB$.

Перейдем к вершине D. Очевидны неравенства: $\angle BDA = \angle CDA < \angle CDA + \angle BDC$. Для другого угла можно записать

$$\cos(BDC) = \frac{\operatorname{ch}(ud) - \operatorname{ch}(ub)}{\operatorname{sh}^2(ud)} > \frac{\operatorname{ch}^2(ud) - \operatorname{ch}(ua)}{\operatorname{sh}^2(ud)} = \cos(BDA). \quad (20)$$

Значит, $\angle BDC < \angle BDA < 2\angle BDA$.

В силу симметрии (углы при B, C одинаковы) осталось рассмотреть B . Докажем, что $\angle CBD > \angle DBA$ и $\angle CBD > \angle ABC$. Для разрешения нижеследующего сравнения (21) используем функцию f_4 :

$$\cos(CBD) = \frac{\operatorname{ch}(ub) \operatorname{ch}(ud) - \operatorname{ch}(ud)}{\operatorname{sh}(ub) \operatorname{sh}(ud)} \vee \cos(DBA) = f_5(ud) f_4(ua). \quad (21)$$

Очевидно, что $f_4(ub) < f_4(ua)$, так как $ua > ub$ и $f'_4(t) > 0$. Следовательно, $\cos(CBD) < \cos(DBA)$, $\cos(ABC) = f_5(ua) f_4(ub)$, $\cos(CBD) = f_5(ud) f_4(ub)$, $f_5(t)$ монотонно убывает при $t > 0$ и $d > a$. Поэтому $\cos(CBD) < \cos(ABC)$. Таким образом, верны следующие неравенства: $\angle DBA + \angle CBD > \angle CBD > \angle ABC$; $\angle ABC + \angle CBD > \angle DBA$. Надо доказать, что

$$\angle ABC + \angle DBA > \angle CBD. \quad (22)$$

Имеем $\angle ABC + \angle DBA \in (0, \pi)$, $\angle CBD \in (0, \pi/2)$. Отсюда получаем

$$\angle ABC + \angle DBA \geq \angle CBD \Leftrightarrow \cos(ABC + DBA) \leq \cos(CBD). \quad (23)$$

Выражение (23) эквивалентно формуле

$$\cos(ABC) \cos(DBA) - \sin(ABC) \sin(DBA) \vee \cos(CBD).$$

Для сокращения записи введем следующие функции:

$$\begin{aligned} T_1(u, b, a, d) &= \sqrt{\operatorname{sh}^2(ud) \operatorname{sh}^2(ub) - \frac{(\operatorname{ch}(ub) - 1)^2 \operatorname{ch}^2(ua) \operatorname{sh}^2(ud)}{\operatorname{sh}^2(ua)}}, \\ T_2(u, b, a, d) &= \sqrt{\operatorname{sh}^2(ud) \operatorname{sh}^2(ub) - \frac{((\operatorname{ch}(ua) - 1))^2 \operatorname{ch}^2(ud) \operatorname{sh}^2(ub)}{\operatorname{sh}^2(ua)}}, \\ T_r(u, b, a, d) &= T_1(u, b, a, d) T_2(u, b, a, d). \end{aligned}$$

Неравенство (23) для углов равносильно неравенству

$$(\operatorname{ch}(ub) - 1) f_5(ua) f_4(ua) \operatorname{ch}(ud) - T_r(u, b, a, d) \leq (\operatorname{ch}(ub) - 1) \operatorname{ch}(ud). \quad (24)$$

После некоторых преобразований получаем

$$\begin{aligned} (\operatorname{ch}(ub) - 1) f_5(ua) f_4(ua) \operatorname{ch}(ud) - T_r(u, b, a, d) &\leq \\ &\leq (\operatorname{ch}(ub) - 1) f_5(ua) f_4(ua) \operatorname{ch}(ud) \end{aligned} \quad (25)$$

или

$$(\operatorname{ch}(ub) - 1) f_5(ua) f_4(ua) \operatorname{ch}(ud) = (\operatorname{ch}(ub) - 1) f(ua) \operatorname{ch}(ud). \quad (26)$$

Так как $ua \in [0, +\infty)$, то $f(ua) \leq 1$ и, значит,

$$\operatorname{ch}(ud)(\operatorname{ch}(ub) - 1) f(ua) \leq \operatorname{ch}(ud)(\operatorname{ch}(ub) - 1).$$

Неравенство (24) выполняется, значит, неравенство (22) доказано. Первый тип тетраэдров исследован.

Докажем пункт б) для II-го типа тетраэдров, $AD = d$; $BC = b$. Рассмотрим вершину A . Нетривиальное неравенство: $\angle CAB \vee 2\angle CAD$, причем косинусы этих углов равны

$$\cos(CAB) = \frac{\text{ch}^2(ua) - \text{ch}(ub)}{\text{sh}^2(ua)}, \quad (27)$$

$$\begin{aligned} \frac{\text{ch}^2(ua) - \text{ch}(ub)}{\text{sh}^2(ua)} &\geq \frac{\text{ch}^2(ua) - \text{ch}(ua)}{\text{sh}^2(ua)} = \\ &= f_4(ua)f_5(ua) \geq f_4(ud)f_5(ua) = \cos(CAD). \end{aligned} \quad (28)$$

Следовательно, $\angle CAB \leq \angle CAD < 2\angle CAD$. В случае рассмотрения вершины D доказательство проводится с точностью до переобозначений. Для вершин B , C по аналогии можно показать, что $\angle DBA < 2\angle CBD$.

2) Докажем вложение четверки точек в сферу. Для доказательства используем результаты о вписываемости ультраметрических треугольников в окружности.

У тетраэдра I-го типа основание (ABC) вписывается в окружность, ее центр H . Отрезок DH – равноудален от A, B, C . На этом отрезке выберем некоторую точку T . По аналогии с доказательством вписываемости ультраметрического треугольника в окружность достаточно доказать, что $DH > HA$. Проведем AH до пересечения с BC , получим M . Сравним наложением медианы к BC у $\triangle ABC$ и $\triangle DBC$, получим $DM \geq AM$, $\angle DAM \geq \angle MDA > \angle MDA/2 = \angle HDA$, следовательно, в $\triangle MAH$ стороны $DH > AH$ – по противолежащим углам. Используя рассуждения из доказательства вписания ультраметрического треугольника в окружность, получим $\exists T_0 \in DH$: $|T_0D| = |T_0A|$. Так как T_0 равноудалена от A, B, C , то T_0 – центр искомой сферы.

Рассмотрим тетраэдр II типа $c = CD < DA = DB = AC = AB = d > AB = b$. Отметим точки M_1 и M_2 – середины AB и CD . Получим $M_1M_2 = 1/kch[(ch(kd))/(ch(kc/2)ch(kb/2))]$. Найдем гладкую функцию для $|DT| - |AT|$ $T \in M_1M_2$. Сравнивая углы в треугольниках $\triangle AM_1D$ $\triangle AM_2D$, получим, что $AM_1 < DM_1$, $AM_2 > DM_2$ и

$$\begin{aligned} |AT| - |DT| &= f_9(t) = \\ &= \frac{1}{k} \left(\text{arcch} \left(\text{ch} \left[\frac{kb}{2} \right] \text{ch}(C_1 - kt) \right) - \text{arcch} \left(\text{ch}(kt) \text{ch} \left(\frac{kc}{2} \right) \right) \right), \end{aligned} \quad (29)$$

где

$$C_1 = \text{arcch} \left[\frac{\text{ch}(kt)}{\text{ch}(kb/2) \text{ch}(kc/2)} \right].$$

Функция $f_9(t)$ гладкая. На M_1 и M_2 она принимает значения, противоположные по знаку. По теореме Больцано-Вейерштрасса на M_1M_2 найдется точка M_0 : $AM_0 = DM_0$. По построению $AM_0 = BM_0$ $DM_0 = CM_0$. Мы получили

центр сферы, описанной около тетраэдра $ABCD$. Утверждения 1) и 2) доказаны. То есть четверки точек, подчиняющиеся ультраметрическому неравенству треугольника (1), вкладываются в подмножества сфер в S_k^3 при любом $k > 0$.

Достаточность. Из четырех вершин выбираем три. Опустим из центра шара перпендикуляр на проходящую через выбранные три точки плоскость Φ . Пересечением перпендикуляра с плоскостью Φ будет точка L . От трех выбранных вершин точка L равноудалена, тем самым они лежат на окружности. Используя предложение 1, получаем: любая тройка вершин из четырех данных вложена в УМП. Отсюда следует, что все четыре точки вложены в УМП.

ЛИТЕРАТУРА

1. Лобачевский Н.И. *Геометрические исследования по теории параллельных линий*. М.-Л.: Изд-во АН СССР, 1945.

МЕХАНИЧЕСКИЕ АНАЛОГИИ В ЗАДАЧЕ ШТЕЙНЕРА

Н.Ф. Богаченко, Р.Т. Файзуллин

The statement of the Shteiner problem and some approximate approaches are covered in this article. It is suggested the heuristic algorithm to the decision of the Shteiner problem which based on the mechanical analogies.

1. Постановка задачи и точные алгоритмы

Рассмотрим задачу Штейнера о построении кратчайшей связывающей сети на плоскости и способы ее решения, в частности возможность использования механических аналогий. Данная проблема имеет многочисленные приложения при решении транспортных задач, в теории графов, при конструировании интегральных электронных схем, при планировании топологии телефонных сетей, трубопроводов, шоссеиных дорог и т.д. [1, 5].

Введем ряд определений. Кратчайшей связывающей сетью (КСС) для точек $a_1, a_2, \dots, a_n$, образующих множество M , называется дерево с вершинами в этих точках, имеющее минимальную длину, то есть наименьшую сумму длин ребер. Далее, если не оговорено противное, будем рассматривать задачу на плоскости. Известны простые алгоритмы построения такого дерева (остова минимального веса), например алгоритм Краскала [4, с.60–61], алгоритм Прима [4, с.61–62]. Но иногда возможно построить более короткое дерево, связывающее исходные точки путем добавления новых вершин, которые в дальнейшем, согласно общепринятой терминологии, будем называть точками Штейнера. В общем случае под задачей Штейнера понимается следующая проблема [5]: Пусть M – фиксированное конечное множество точек плоскости. Необходимо описать все КСС, «затягивающие» множество M , то есть выбрать из всех плоских графов, множество вершин которых содержит M , сети наименьшей длины. Формулируя задачу в терминах евклидовой плоскости, требуется найти кратчайшую сеть прямолинейных отрезков, связывающих между собой заданное множество точек. Таким образом, для решения задачи Штейнера необходимо построить КСС с использованием новых узлов – точек Штейнера. Дерево, имеющее минимальную длину среди деревьев с любым допустимым числом точек Штейнера, называется *минимальным* деревом Штейнера (МДШ). Под топологией дерева понимается его матрица смежностей [4, с.27]. Относительно минимальным деревом Штейнера является дерево, имеющее минимальную длину среди деревьев с той же топологией.

© 2002 Н.Ф. Богаченко, Р.Т. Файзуллин

E-mail: zhihalkina@math.omsu.omskreg.ru, faizulin@univer.omsk.su

Омский государственный университет

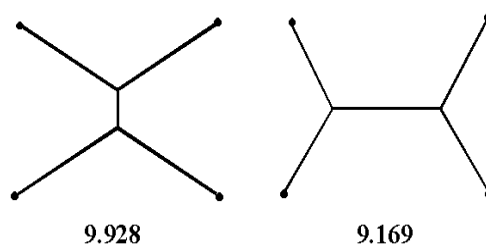


Рис. 1. Относительно минимальные деревья Штейнера

В общей форме задача Штейнера была впервые сформулирована в 1934 году М. Кёсслером и В. Ярником. Однако сама эта проблема не приобрела широкой известности вплоть до 1941 г., когда Р. Курант и Е. Роббинс включили ее в свою книгу [6], связав с именем Я. Штейнера, работавшего над задачей поиска одной точки, сумма расстояний от которой до всех точек заданного множества M была бы минимальной. Имеется точное решение для случая, когда мощность множества M равна 3 [1]. Существующие точные алгоритмы построения кратчайшего дерева Штейнера [2] для произвольного числа исходных вершин и точек Штейнера громоздки и, как следствие, мало эффективны на практике. Кроме того, выявление наиболее приемлемой топологии для заданных вершин и определение числа точек Штейнера зачастую является отдельной задачей.

Относительно минимальное дерево Штейнера, отвечающее заданной топологии, может не быть минимальным (пример двух различных топологий и соответствующие суммы длин ребер указаны на рис. 1).

Действуя методом полного перебора, найти минимальное дерево возможно путем построения всех относительно минимальных деревьев Штейнера (для заданной топологии и заданных вершин существует не более одного относительно минимального дерева Штейнера [2]). Алгоритм Мелзака [1] теоретически решает поставленную задачу, но на практике он нереализуем даже для небольших задач в следствие огромного числа переборов. Дальнейшее развитие этого алгоритма позволило сократить объем вычислений, но время решения задачи определяется геометрией исходной конфигурации и экспоненциально зависит от количества точек. Более того, задача Штейнера относится к классу $\mathcal{NP}$ -полных задач [4, с.62–63], а это дает повод отказаться от поиска эффективного и точного алгоритма и направить усилия на получение приближенных методов.

В теоретическом плане многие свойства относительно минимальных деревьев Штейнера и точек Штейнера хорошо изучены [2]. Далее приведены некоторые из этих свойств.

1. Каждый угол $\geq 120^\circ$, следовательно, каждая точка соединяется с остальными вершинами дерева не более чем тремя ребрами.
2. Число ребер, исходящих из точки Штейнера, равно трем, образуя углы по 120° .
3. Число ребер дерева на единицу меньше суммарного числа исходных вершин и точек Штейнера.
4. Все точки Штейнера лежат в выпуклой оболочке исходных вершин

$a_1, a_2, \dots, a_n$.

5. Число точек Штейнера $s \leq (n - 2)$ (равенство достигается тогда и только тогда, когда каждая исходная вершина a_i имеет ровно одно ребро).

Необходимо отметить, что степень чувствительности к исходным данным слишком высока: незначительное изменение геометрии задачи может коренным образом изменить кратчайшую сеть, являющуюся ее решением.

2. Приближенные алгоритмы

В практических приложениях нередко достаточно построить *условно минимальное* дерево Штейнера, то есть дерево, длина которого меньше КСС, но минимальность которого не гарантирована, либо *локально минимальную* сеть [5], характеризующуюся тем, что любая ее малая деформация увеличивает длину сети.

Используя механические аналогии [2], дерево можно интерпретировать как механическую систему, потенциальная энергия которой равна сумме расстояний между смежными вершинами. Положение устойчивого равновесия такой механической системы соответствует дереву, имеющему относительно минимальную длину. В такой модели ребра дерева представляют собой упругие нити, натяжение которых не зависит от того, насколько они растянуты. Исходные вершины $a_1, a_2, \dots, a_n$ закреплены, а точки Штейнера могут свободно перемещаться. Устойчивое положение этих точек и есть относительно минимальное дерево Штейнера.

Предложенные механические аналогии и теорема Максвелла [5, с.75] позволяют доказать некоторые факты, связанные с МДШ. Например, как отмечалось ранее, в относительно минимальном дереве Штейнера углы между ребрами в точке Штейнера равны 120° , что означает равенство нулю результирующей трех единичных сил, действующих на точку Штейнера.

Курант и Роббинс [1] использовали в качестве модели для создания такой механической системы мыльную пленку, натянутую между двумя параллельными плоскостями, соединенными между собой стержнями в точках $a_1, a_2, \dots, a_n$. Так как мыльная пленка всегда стремится занять такое положение, в котором ее площадь оказалась бы наименьшей по сравнению с площадями всех достаточно близких поверхностей, и может менять «свою топологию», расщепляя вершину на две и тем самым образуя новую точку Штейнера, то положение равновесия соответствует локально минимальной сети [5].

Другой эксперимент также использует механическую аналогию, при этом локально минимальной сети отвечает «равновесное состояние некоторой механической системы точек, для которой потенциал – это сумма некоторых попарных расстояний» [5, с.56]. В этом случае появляются вершины, в которых встречаются по четыре ребра и больше. Но «поскольку система находится в состоянии равновесия, а ее энергия пропорциональна длине сети, полученные сети также минимальны в некотором разумном смысле» [5, с.57].

3. Компьютерный аналог динамической системы материальных точек

Заменяем эти механические модели компьютерным аналогом динамической системы материальных точек, часть из которых жестко закреплена, а часть свободно перемещается. Положение устойчивого равновесия такой системы соответствует максимуму потенциальной энергии. На базе этой модели предлагается приближенный алгоритм построения условно минимального дерева Штейнера.

Далее рассматриваются полные деревья Штейнера, то есть деревья, в которых число точек Штейнера $s = n - 2$ (так как любое дерево Штейнера, которое не является полным, может быть представлено в виде объединения полных деревьев Штейнера [2]).

В основе алгоритма лежат свойства 2), 4), 5). В начальный момент времени внутри выпуклой оболочки исходных или граничных вершин $a_1, a_2, \dots, a_n$, которые являются неподвижными, случайным образом генерируется $s = (n - 2)$ точек Штейнера. Необходимо отметить, что задача построения выпуклой оболочки сама по себе является достаточно трудоемкой. На первом этапе разработки алгоритма представляется возможным ограничить область распределения точек Штейнера прямоугольником, объемлющим исходные вершины. Если предположить, что граничные точки образуют выпуклую оболочку или выделено необходимое подмножество этих вершин, то могут быть приемлемы следующие стратегии.

1. Добавочные вершины генерируются внутри окружности с центром в центре масс системы исходных вершин и радиуса, равного минимальному расстоянию от центра до граничных точек.

2. Выпуклый многогранник разбивается на треугольники с общей вершиной во вновь сгенерированной точке. Сумма площадей треугольников должна быть равна площади этого многогранника.

На каждой итерации изменение координат точек Штейнера происходит по следующему правилу: для текущей вершины определяется три ближайших «соседа», из которых граничными могут быть максимум две точки (при этом каждая граничная вершина выбирается в качестве «ближайшей» не более одного раза). Далее координаты точки Штейнера изменяются согласно формулам:

$$x_{sht} = \frac{m_1 x_1 + m_2 x_2 + m_3 x_3}{m_1 + m_2 + m_3}, \quad y_{sht} = \frac{m_1 y_1 + m_2 y_2 + m_3 y_3}{m_1 + m_2 + m_3}, \quad (1)$$

где m_1, m_2, m_3 – «массы», а $(x_1, y_1), (x_2, y_2), (x_3, y_3)$ – координаты точек, соответствующих вершинам треугольника. Иными словами, точка Штейнера перемещается в центр масс системы, состоящей из трех ближайших вершин. Соединения точки Штейнера с этими вершинами образуют ребра графа, который строится на данной итерации. Критерием останова является стабилизация системы, то есть ее переход в состояние устойчивого равновесия.

Заметим, что в произвольной гравитирующей системе материальных точек частицы смещаются в направлении центра масс либо системы в целом, либо подмножества ближайших соседей. Иными словами, система стремится максимизировать потенциальную энергию. Система частиц, представляющих собой

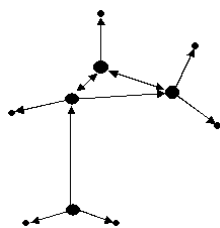


Рис. 2. Ошибочные конфигурации

вершины дерева Штейнера, также обладает максимальной потенциальной энергией, но при условии неподвижности граничных точек.

Как показывают численные эксперименты, алгоритм наиболее эффективен в случае, когда исходные вершины «тяжелее» точек Штейнера. Величина разности «масс» неподвижных и свободно перемещающихся точек, при которой удастся найти условно минимальное дерево Штейнера, является настраиваемым параметром и зависит как от общего числа вершин, так и от начальной конфигурации исходных точек.

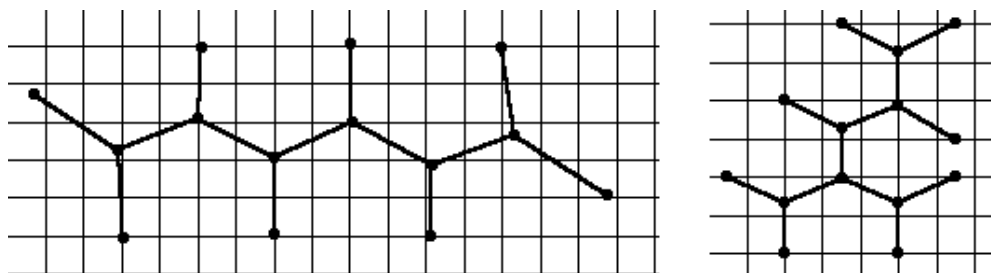
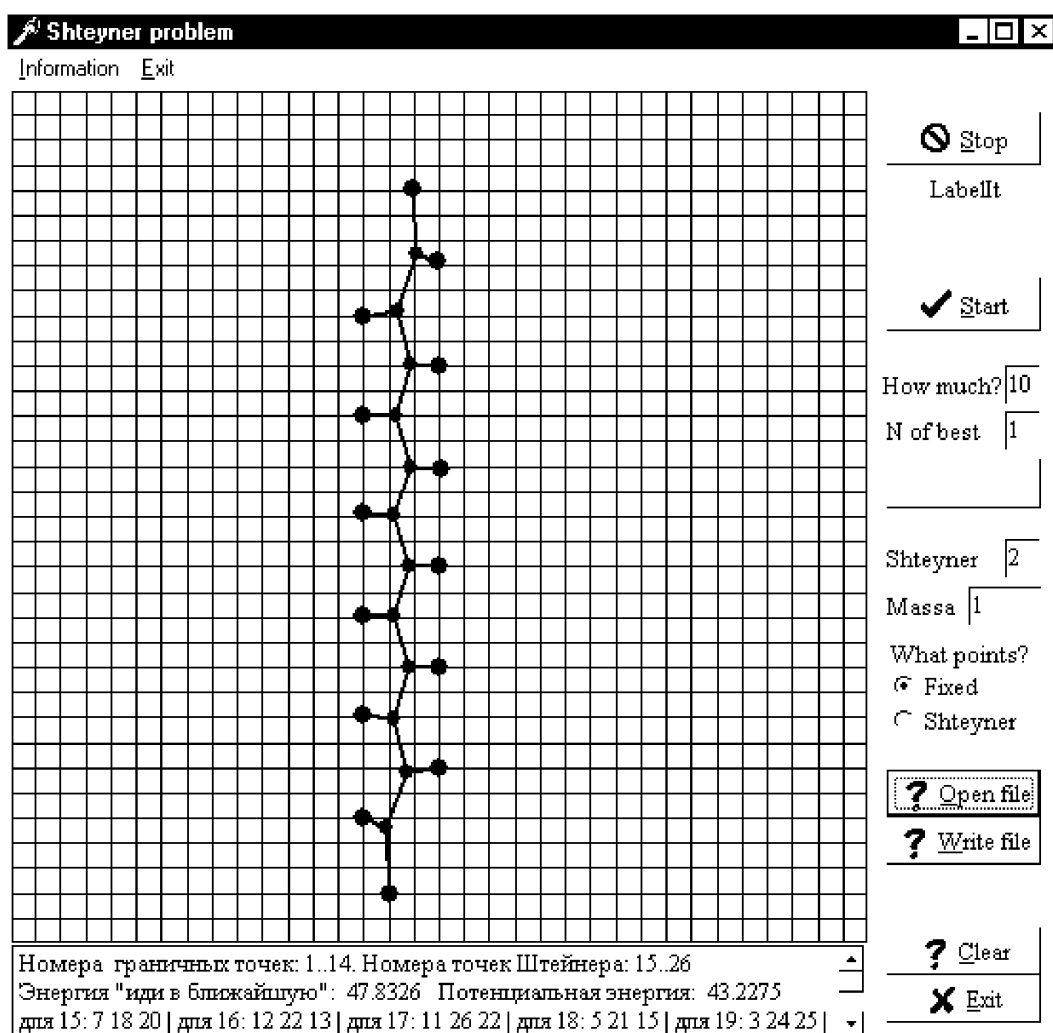
Изложенный вариант алгоритма может привести к потере связности получаемого графа. Одна из причин – «склейка» точек Штейнера. Кроме того, появляются вершины, имеющие более трех ребер. Решить последнюю проблему возможно путем формирования графа не последовательно («от треугольника к треугольнику»), а строя КСС (дерево минимальной длины) на полученных $2(n-1)$ вершинах. Что касается вопроса связности, то один из вариантов решения – это обязательная принадлежность точки Штейнера найденному треугольнику. (В противном случае самая дальняя точка «блокируется» и вновь ищутся три ближайших; если после перебора всех точек допустимый треугольник не найден, то происходит изменение координат текущей точки Штейнера согласно некоторому правилу. Но, как показывает вычислительный эксперимент, наиболее эффективна случайная смена координат.) К сожалению, эта интерпретация алгоритма также не защищена от появления «висячих» вершин или неприемлемых конфигураций (см. рис. 2).

Отслеживание всех недопустимых ситуаций приводит к обычным комбинаторным проблемам.

С другой стороны, на начальном этапе решения задачи, когда требуется определить геометрию вершин, приемлем исходный вариант алгоритма, трудоемкость которого значительно меньше.

Кроме того, предложенный алгоритм легко переносится на случай пространства, размерность которого больше двух, так как ребра искомого дерева по-прежнему являются геодезическими и степень каждой вершины не превосходит трех [5, с.57].

В работе не приводятся оценки трудоемкости и сравнительные характеристики алгоритма, а также не затрагиваются вопросы сходимости. Целью являлось описание одного из возможных подходов к решению поставленной проблемы. Успешное решение отдельных тестовых задач позволяет надеяться на дальнейшее развитие и применение предложенного алгоритма. На рисунках 3, 4

Рис. 3. Минимальные деревья Штейнера для $n = 8$ Рис. 4. Минимальное дерево Штейнера для $n = 14$

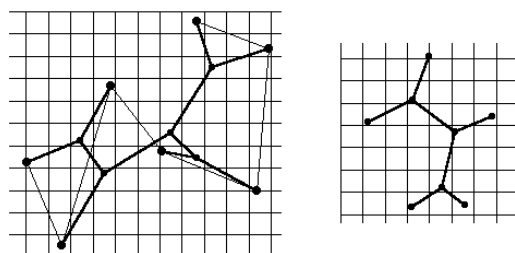
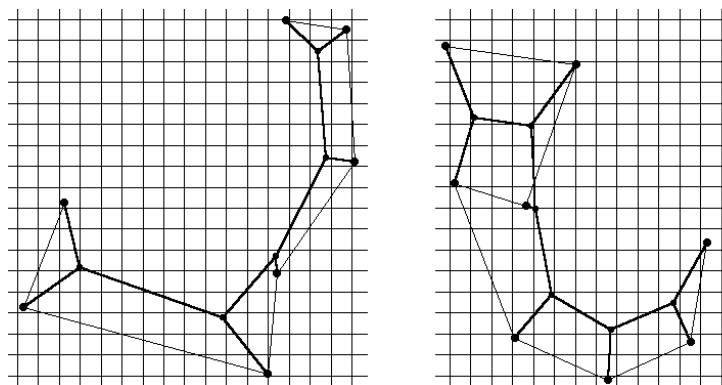
Рис. 5. Условно минимальные деревья Штейнера для $n = 7, 5$ Рис. 6. Условно минимальные деревья Штейнера для $n = 7, 8$

Таблица 1.

Рис.	Исходные вершины		S	S_{sht}
	n	m		
5	7	2,5		
5	5	1,5	14,46	13,97
6	7	2,0	38,01	37,47
6	8	1,2	39,03	38,95

представлены результаты работы алгоритма по поиску топологии, отвечающей МДШ (исходные конфигурации представляют собой множества из n висячих вершин результирующих деревьев, $n = 8$ и 14). Данные графы совпадают с деревьями, построенными при помощи точных алгоритмов [2]. Рисунки 5, 6 демонстрируют работу алгоритма по построению условно минимального дерева (исходные конфигурации представляют собой множества висячих вершин результирующих деревьев). Пусть S – длина КСС, найденной приближенным алгоритмом: первое ребро соединяет две ближайшие вершины, затем на каждой итерации добавляется ближайшая, еще не присоединенная вершина и строится соответствующее ребро, из результирующей замкнутой цепи удаляется ребро максимальной длины; S_{sh} – длина дерева Штейнера, полученного при работе предложенного алгоритма, основанного на механических аналогиях; n – число исходных вершин; m – «масса» исходных вершин («масса» точки Штейнера равна 1).

В заключении еще раз вернемся к вопросу начального распределения точек Штейнера. Возможно, что наиболее «благоприятным» является такое расположение этих точек внутри выпуклой оболочки исходных вершин, что проведенная через них статистическая кривая (например, по методу наименьших квадратов) «достаточно» гладкая [3].

ЛИТЕРАТУРА

1. Берн М. У., Рональд Л. Г. *Поиск кратчайших сетей* // В мире науки. 1989. N 3. С.64–70.
2. Гилберт Э. Н., Поллак Г. О. *Минимальные деревья Штейнера* // Кибернетический сборник, вып. 8. М.: Мир, 1971. С.19–49.
3. Горбунов С. В., Кисель И. В., Конотопская Е. В., Файзуллин Р. Т. *Сравнение методов гарантированной гладкости и эластичной сети для задачи коммивояжера на плоскости* // Сообщения ОИЯИ Р5-97-258. Дубна, 1997. С.1–7.
4. Емеличев В. А., Мельников О. И., Сарванов В. И., Тышкевич Р. И. *Лекции по теории графов*. М.: Наука, 1990. 383 с.
5. Иванов А.О., Тужилин А.А. *Геометрия минимальных сетей и одномерная проблема Плато* // Успехи математических наук. М.: Наука, 1992. Т.47. В.2(284). С.53–115.
6. Courant R., Robbins H. *What is Mathematics?* Oxford University Press, New York, 1941. (Русский перевод: Курант Р., Роббинс Г. *Что такое математика?* М.: Просвещение, 1967. 559 с.)

АЛГОРИТМ РАСЧЕТА ТЕХНОЛОГИЧЕСКИХ ПАРАМЕТРОВ НАГРЕВА ПОРОШКОВЫХ ОБРАЗЦОВ В ВЫСОКОЧАСТОТНОМ ЭЛЕКТРОМАГНИТНОМ ПОЛЕ

О.Т. Данилова

The method for mathematical description of the characteristics of high-frequency heating is developed.

Электромагнитные поля влияют на кинетику химических реакций и термодинамику фазовых превращений, изменяют скорость, выход и даже направление химической реакции, изменяют свойства обрабатываемых материалов. На основании обзора теории и методов электротермического нагрева разработан расчетный алгоритм, позволяющий получить информацию о характере тепловых, электромагнитных и иных параметров в системе при высокочастотном индукционном и диэлектрическом нагреве порошковых материалов. Алгоритм состоит из трех частей (блоков): тепловой, электрической, информационно-логической. В тепловом и электрическом блоках производится расчет соответствующих полей. Информационно-логический блок осуществляет ввод и вывод информации, а также взаимодействие остальных частей программы, определяемое конкретной структурой нагревателя и режимом его работы. Тепловой расчет выполняется конечно-разностным методом, а электрический – по методу численного расчета осесимметрических систем [1].

При организации алгоритма принято во внимание следующее: в отличие от индукционного нагрева металлов при диэлектрическом нагреве диэлектриков поверхностный эффект является вредным, так как приводит к неравномерному распределению температуры, которая не может выровняться из-за низкой теплопроводности диэлектриков. Во избежание заметных проявлений поверхностного эффекта надо выбирать частоту поля такой, чтобы глубина проникновения в 3-4 раза превосходила размеры нагреваемого образца. Так как нагреваемый образец имеет конечные размеры, то из-за отражения электромагнитных волн от границ образца внутри него устанавливаются стоячие волны подобно тому, что происходит в электрических цепях с распределенными параметрами. Это явление в сочетании с поверхностным эффектом может приводить к весьма сложной картине распределения поля по объему тела.

Особенностью индукционного ввода энергии является регулирование пространственного расположения зоны протекания вихревых токов. Во-первых,

© 2002 О.Т. Данилова

E-mail: danilova@univer.omsk.su

Омский государственный университет

вихревые токи протекают в пределах площади, охватываемой индуктором, нагревается только та часть образца, которая находится в магнитной связи с индуктором, независимо от размеров тела. Во-вторых, глубина зоны циркуляции вихревых токов и, следовательно, зоны выделения энергии зависит кроме других факторов от частоты тока индуктора (увеличивается при низких частотах и уменьшается с повышением частоты) [2]. Для расчета параметров процесса индукционного нагрева в зависимости от геометрических размеров образцов определяются: глубина проникновения тока в образец; оптимальная частота нагрева; время нагрева; реактивные и активные сопротивления; число и высота витков; удельная мощность.

Нижний и верхний пределы выбора частоты тока для индукционного нагрева образца диаметром d определяются соотношениями:

$$f_0 > \frac{3 \cdot 10^6 \rho}{(\mu d^2)}, \quad (1)$$

$$f_x < \frac{6 \cdot 10^6 \rho}{(\mu d^2)}, \quad (2)$$

где μ – относительная магнитная проницаемость материала; ρ – его удельное сопротивление.

Предельный электрический КПД индуктора определяется выражением [2]:

$$\eta_u = \frac{1}{1 + \frac{d}{gD} \sqrt{\frac{\rho_u}{\mu\rho}}}, \quad (3)$$

где g – коэффициент заполнения; ρ_u – удельное сопротивление материала индуктора; D – диаметр индуктора.

Внешнее емкостное сопротивление $X_c = 1/\omega C_L$ (где C_L – внешняя, или, точнее, краевая емкость рабочего конденсатора) можно рассчитать, используя некоторые свойства электрического поля конденсатора и магнитного поля индуктора. Если рассмотреть схему замещения индуктора с нагреваемым образцом, основанную на общности потока замыкания, и схему замещения рабочего конденсатора (рис.1), то можно заметить полную аналогию.

Схема замещения индуктора по общему потоку получается из схемы замещения конденсатора путем замены всех его емкостей индуктивностями, и сопротивление r_1 становится сопротивлением провода индуктора. Расчет сопротивления X_L (при диэлектрическом нагреве это внешнее сопротивление, при индукционном-индуктивное сопротивление обратного замыкания) основан на предположении, что внешнее магнитное поле индуктора с нагрузкой подобно полю пустого индуктора. Очевидно, справедливо и аналогичное утверждение: краевое электрическое поле конденсатора с нагрузкой подобно полю пустого конденсатора [3]. Отсюда следует способ расчета:



Рис. 1. Схемы замещения рабочего конденсатора и индуктора с нагреваемым образцом

$$C_L = C - C_{\text{ср}}, \quad (4)$$

где $C_{\text{ср}}$ – емкость средней части конденсатора с нагреваемым материалом; C – емкость пустого конденсатора.

Полная аналогия наблюдается и при расчете индуктора с нагрузкой:

$$R_a = r_1 + Lr_2 \approx Lr_2, \quad (5)$$

$$X_{\text{э}} = L[X_2 + X_s + \frac{r_2^2 + (X_r + X_s)}{X_L}], \quad (6)$$

где L – коэффициент приведенного активного сопротивления загрузки:

$$L = \frac{1}{(\frac{r_2}{X_s})^2 + (\frac{X_2 + X_s}{X_L})^2}. \quad (7)$$

При проведении расчетов следует принимать во внимание, что если обрабатываемые образцы являются полупроводниками различного типа, то можно ожидать, что высокочастотное электромагнитное поле окажется интенсификатором процесса восстановления. Поэтому необходимо учитывать дополнительную энергию, вносимую в реакционную зону путем преобразования электромагнитной энергии в тепловую, которая выделяется в результате перемангничивания и индуктивных потерь. Время нагрева образцов определялось по формуле [4]:

$$t_k = \frac{D^2}{4a} \cdot \frac{S(\alpha, 0) - (T_0/T_c)S(\alpha, 0)}{T_0/T_c - 1}, \quad (8)$$

где a – тепловой коэффициент для заданных температур; T_0, T_c – значения температур на поверхности и на оси образца соответственно; $S(\alpha, 0), S(\alpha, 1)$ –

критерий Фурье. Расчеты показывают, что при индукционном нагреве оксидов железа при частоте 0,44 МГц резкое увеличение температуры как внутри образца, так и снаружи начинается с $230 - 250^0 \text{ C}$. Затем температура во внутренних и наружных слоях возрастает с разной интенсивностью. На этом этапе преобладают уже вихревые токи, которые на поверхности образца максимальны, что является проявлением скин-эффекта, который характеризуется глубиной проникновения ВЧП в слой материала $\Delta = 0,49 - 0,64 \text{ м}$. Пределы изменения Δ определяются температурой процесса $400 - 700^0 \text{ C}$. Если материал восстанавливается до железа, то глубина проникновения резко уменьшается до $0,003 - 0,006 \text{ м}$.

ЛИТЕРАТУРА

1. Глуханов Н.П. *Физические основы высокочастотного нагрева*. Л.: Машиностроение, 1989. 460 с.
2. Нетушил А.В., Жуховицкий Б.Я., Кудин В.Н. *Высокочастотный нагрев в электрическом поле*. М.: Высшая школа, 1961. 146 с.
3. Донской А.В., Рамм Г.С., Вигдорович Ю.Б. *Высокочастотные электротермические установки с ламповыми генераторами*. Л.: Энергия, 1974. 208 с.
4. Александров В.И., Осико В.В., Татаринцев В.М. *Плавление тугоплавких диэлектрических материалов высокочастотным нагревом* // Приборы и техника эксперимента. 1970. N.5. С.222–226.

ОБ ОДНОЙ МОДЕЛИ ДИНАМИКИ ЧИСЛЕННОСТИ ИЗОЛИРОВАННОЙ ПОПУЛЯЦИИ

Л.В. Недорезов, Е.В. Волкова

At present paper there is the analysis of continuous-discrete model of isolated population dynamics. It's assumed that population has nonoverlapping generations and appearance of individuals of new generation correlates with the death of previous. Also it's assumed that during the winter period there isn't the activity of individuals in population (for example, it's typical for insects in boreal forests), and part of individuals eliminates. Productivity of individuals is constant and death rate during the winter period depends on the food conditions (i.e. the respective function in model depends on the values of integral on any time intervals). For model there is the analysis of dynamic regimes and, in particular, determination of conditions of stability of origin and existance of nontrivial equilibrium. For one partial case of model it was shown that the chaotic regimes can be realized under certain values of parameters.

Работа посвящена анализу непрерывно-дискретной модели динамики численности изолированной популяции. Предполагается, что поколения в популяции не перекрываются и появление особей новой генерации сопровождается гибелью предыдущего. Также предполагается, что в течение всего зимнего периода никакой жизнедеятельности особей не происходит (что характерно, например, для насекомых в лесах бореальной зоны) и часть особей погибает. При этом плодовитость особей считается постоянной величиной, а гибель особей в течение зимы зависит от условий питания (т. е. соответствующая функция в модели зависит от значений интеграла на определенных интервалах времени). Для модели изучаются динамические режимы и, в частности, выявляются условия устойчивости начала координат и существования нетривиального равновесия. Для одного частного случая модели с помощью численных методов показано, что при определенных значениях параметров в модели реализуются хаотические режимы.

Введение

Видимо, первой моделью, предназначенной для описания динамики численности изолированной популяции и базирующейся на обыкновенных дифференциальных уравнениях с импульсами, является модель В.А. Костицына [1]. В рамках модели предполагается, что существуют фиксированные моменты времени τ_k , $k = 0, 1, 2, \dots$, когда происходит разрыв траектории, что соответствует

© 2002 Л.В. Недорезов, Е.В. Волкова

E-mail: leo@nsu.ru

Новосибирский государственный университет

моменту появления особей новой генерации. Между этими моментами времени численность популяции монотонно снижается (процесс гибели особей), что описывается обыкновенным дифференциальным уравнением.

Дальнейшее развитие данного подхода в моделировании динамики популяций представлено в работах [2–6], где, в частности, показано, что дискретным аналогом модели Ферхюльста [7] является модель Скллама [8], а не дискретная логистическая модель [8], как это обычно указывается в литературе. Также было показано, что при постоянной плодовитости и смертности, зависящей от текущего значения численности популяции (и удовлетворяющей лишь общепринятым ограничениям), в модели нет осцилляторных режимов и существует не более одной глобальной устойчивой точки равновесия.

Настоящая работа является продолжением этих исследований, и в ней также рассматривается математическая модель динамики численности изолированной популяции. Предполагается, что плодовитость особей постоянна, а доля погибающих особей за зимний период зависит от условий их питания в течение некоторого интервала времени. Считается, что поколения в популяции не перекрываются и появление особей новой генерации сопровождается гибелью предыдущего. Исследуются общие свойства модели, определяются частные случаи модели при различных видах зависимости скорости гибели от численности популяции. Для одного частного случая определяются условия вырождения популяции, стабилизации численности на нетривиальном уровне, а также показывается, что в пространстве параметров существуют области, соответствующие хаотическому изменению численности.

1. Описание модели

Будем предполагать, что динамику численности популяции на каждом отрезке времени $[t_k, t_{k+1})$, $k = 0, 1, 2, \dots$ – между моментами наступления зимних погодных условий, в течение которых часть особей погибает, – можно описать с помощью следующего обыкновенного дифференциального уравнения:

$$\frac{dx}{dt} = -xR(x), \quad (1)$$

где интенсивность гибели особей $R(x)$ удовлетворяет следующим ограничениям, которые считаются общепринятыми [9–12]:

$$R(0) > 0, \quad \frac{dR}{dx} > 0, \quad R(+\infty) = +\infty. \quad (2)$$

Величина $R(0)$ является интенсивностью естественной смертности особей (мальтузианский параметр), а увеличение значений функции $R(x)$ обусловлено усилением действия внутрипопуляционных саморегуляторных механизмов с ростом численности популяции.

Обозначим через $x(\tau_k - 0)$ количество особей, которые выжили к моменту времени τ_k и которые могут дать потомство (будем считать, что τ_k лежит в интервале (t_k, t_{k+1}) , $\tau_k \in (t_k, t_{k+1})$), через $Y = \text{const}$ – коэффициент плодовитости,

т. е. среднее число потомков, которое порождается каждой выжившей особью. Тогда численность особей новой генерации определяется выражением

$$x(\tau_k) = Yx(\tau_k - 0), \quad Y > 1. \quad (3)$$

Для многих природных популяций и, в частности, для насекомых, характерно отсутствие какой-либо жизнедеятельности в течение всего зимнего периода. Каждая особь за этот период под воздействием различных факторов – флуктуаций температуры, глубины снежного покрова и др. – либо выживает, либо гибнет. Поэтому можно считать, что в моменты времени t_k , $t_{k+1} - t_k = h > 0$ происходит скачкообразное уменьшение значений численности популяции:

$$x(t_{k+1}) = Qx(t_{k+1} - 0), \quad (4)$$

где через $x(t_{k+1} - 0)$ обозначена численность особей, доживших до момента времени t_{k+1} , через Q – доля выживших особей за зимний период $(k + 1)$ -го года.

Как уже отмечалось, выживаемость особей за зимний период зависит от многих факторов, но в первую очередь от условий питания особей в течение всего периода жизнедеятельности $[t_k, t_{k+1})$. Поэтому целесообразно считать, что величина $Q = Q(u)$, где

$$u = \frac{1}{h} \int_{t_k}^{t_{k+1}} x(t) dt.$$

Функция $Q(u)$ монотонно убывает, и чем выше средняя численность популяции, тем хуже условия питания особей (в результате усиления внутривидовой конкуренции особей):

$$Q(u) \in [0, 1], \quad \frac{dQ}{du} < 0. \quad (5)$$

Объединяя вместе соотношения (1) – (5), получаем требуемую модель динамики численности изолированной популяции. Необходимо отметить, что если в (3) $Y < 1$, то тогда последовательность $x(t_k)$ в модели (1) – (5) монотонно убывает при любых начальных значениях и, следовательно, стационарное состояние 0 является глобально устойчивым равновесием.

2. Сведение модели к рекуррентному уравнению

Пусть функция

$$\Psi(x) = \int \frac{dx}{xR(x)}.$$

Из ограничений (2) следует, что функция $\Psi(x)$ монотонно возрастает, неотрицательна и «выпукла вверх»: $\Psi'(x) > 0$, $\Psi''(x) < 0$. Таким образом, существует обратная функция Ψ^{-1} , которая монотонно возрастает, неотрицательна и «выпукла вниз».

Из (1) получаем:

$$x(t) = \Psi^{-1}(-t + c_1), \quad (6)$$

где константа интегрирования $c_1 = t_k + \Psi(x_k)$. Таким образом, выражение (6) приобретает вид: $x(t) = \Psi^{-1}(t_k - t + \Psi(x_k)) \quad \forall t \in [t_k, \tau_k]$. В момент τ_k происходит разрыв траектории уравнения (1) (появление особой новой генерации), и, следовательно,

$$x(\tau_k) = Yx(\tau_k - 0) = Y\Psi^{-1}(\Psi(x_k) - h_1), \quad (7)$$

где $h_1 = \tau_k - t_k$.

Решая (1) на интервале $[\tau_k, t_{k+1})$ с начальным условием (7), окончательно получаем:

$$x(t_{k+1}) = Q\left(\frac{1}{h} \int_{t_k}^{t_{k+1}} x(t) dt\right) \Psi^{-1}(c_2 - t_{k+1}), \quad (8)$$

где $c_2 = \tau_k + \Psi(x(\tau_k))$. Обозначим правую часть рекуррентного уравнения (8) через $F(x_k)$.

3. Достаточные условия монотонного роста функции $F(x)$

Если функция F в (8) монотонно возрастает, то тогда в модели реализуются режимы только с монотонным изменением численности популяции. Производная функции $F(x)$ определяется выражением:

$$\begin{aligned} \frac{dF}{dx_k} &= \frac{d}{dx_k} \left(Q\left(\frac{1}{h} \int_{t_k}^{t_{k+1}} x(t) dt\right) \right) \Psi^{-1}(c_2 - t_{k+1}) + \\ &Q\left(\frac{1}{h} \int_{t_k}^{t_{k+1}} x(t) dt\right) \frac{d}{dx_k} (\Psi^{-1}(c_2 - t_{k+1})). \end{aligned} \quad (9)$$

Оценим знак каждого слагаемого в (9) по отдельности. Рассмотрим

$$\frac{d}{dx_k} (\Psi^{-1}(c_2 - t_{k+1})) = (\Psi^{-1})'(c_2) \Psi'.$$

Функция $c_2 = c_2(x_k)$ представима в виде:

$$c_2 = \tau_k + \Psi(Y\Psi^{-1}(\Psi(x_k) - h_1)),$$

и, следовательно, ее производная имеет вид:

$$(c_2)' = \Psi'Y(\Psi^{-1})'\Psi'.$$

С учетом последнего получаем неравенство

$$\frac{d}{dx_k} (\Psi^{-1}(c_2 - t_{k+1})) = Y(\Psi')^2((\Psi^{-1})')^2 > 0. \quad (10)$$

Далее

$$\begin{aligned} \frac{d}{dx_k} \left(Q\left(\frac{1}{h} \int_{t_k}^{t_{k+1}} x(t) dt\right) \right) = \\ Q' \frac{1}{h} \left\{ \int_{t_k}^{\tau_k} \frac{d\Psi^{-1}(c_1 - t)}{dx_k} dt + \int_{\tau_k}^{t_{k+1}} \frac{d\Psi^{-1}(c_2 - t)}{dx_k} dt \right\}. \end{aligned}$$

Очевидно, справедливо неравенство

$$\frac{d}{dx_k}(\Psi^{-1}(c_1 - t_{k+1})) = (\Psi^{-1})'\Psi' > 0.$$

В силу ограничений модели $Q' < 0$. Следовательно, с учетом (10) получаем неравенство

$$\frac{d}{dx_k} \left(Q \left(\frac{1}{h} \int_{t_k}^{t_{k+1}} x(t) dt \right) \right) < 0.$$

Оценим знак выражения $\Psi^{-1}(c_2 - t_{k+1})$. Как уже было отмечено, $\Psi^{-1}(x) \geq 0 \quad \forall x \geq 0$, или, иначе, $\Psi^{-1}(c_2 - t_{k+1}) \geq 0$, если $c_2 - t_{k+1} \geq 0$. При $c_2 - t_{k+1} \geq 0$ имеем $\tau_k + \Psi(x(\tau_k)) \geq t_{k+1}$, $x(\tau_k) \geq \Psi^{-1}(h_2)$, где $h_2 = t_{k+1} - \tau_k$. Преобразуя это выражение, получаем неравенство $x(\tau_k) = Y\Psi^{-1}(\Psi(x_k) - h_1) \geq \Psi^{-1}(h_2)$ - и окончательно

$$x_k \geq \Psi^{-1} \left(\Psi \left(\frac{\Psi^{-1}(h_2)}{Y} \right) + h_1 \right) = M.$$

Очевидно, константа M положительна.

Таким образом, $\Psi^{-1}(c_2 - t_{k+1}) \geq 0 \quad \forall x \geq M$ и $\Psi^{-1}(c_2 - t_{k+1}) \leq 0$ при $x < M$. Следовательно, справедливо следующее утверждение:

Утверждение 1. 1. На интервале $0 \leq x_k < \Psi^{-1}(\Psi(\frac{\Psi^{-1}(h_2)}{Y}) + h_1)$ функция $F(x)$ монотонно возрастает.

2. Если $x_k \geq \Psi^{-1}(\Psi(\frac{\Psi^{-1}(h_2)}{Y}) + h_1)$, то для монотонного роста функции $F(x)$ дополнительно требуется выполнение неравенства

$$Q \left(\frac{1}{h} \int_{t_k}^{t_{k+1}} x(t) dt \right) \frac{d}{dx_k}(\Psi^{-1}(c_2 - t_{k+1})) \geq \\ - \frac{d}{dx_k} \left(Q \left(\frac{1}{h} \int_{t_k}^{t_{k+1}} x(t) dt \right) \right) \Psi^{-1}(c_2 - t_{k+1}).$$

■

4. Частные случаи

В математической экологии при построении моделей динамики численности популяции используются различные типы зависимостей популяционных параметров от показателей условий питания особей [1–10]. Наиболее часто встречаются экспоненциальная и гиперболическая зависимости. Поэтому представляет определенный интерес выявление возможных частных случаев модели (1) – (5) при различных видах функций R и Q .

Утверждение 2. При $R(x) = b$ и $Q(x) = e^{-x}$ отображение (8) принимает вид:

$$x_{k+1} = \alpha x_k e^{-\beta x_k}, \quad (11)$$

где α, β - некоторые положительные константы. ■

Заметим, что уравнение (11) – известная модель Морана-Риккера, анализу которой посвящено достаточно большое число работ (см., например, [13–16]). Таким образом, модель Морана-Риккера «выводима» в рамках модели (1) – (5), однако в условиях, когда в популяции не действуют саморегуляторные механизмы ($R = \text{const}$), плодовитость особей постоянна и единственным регулятором численности выступают зимние погодные условия. Здесь необходимо отметить, что, по мнению некоторых авторов [10,17,18], именно зимние условия являются главными регуляторами отдельных массовых видов лесных насекомых.

Утверждение 3. При $R(x) = b$, $Q(x) = \frac{1}{x+1}$ отображение $x_{k+1} = F(x_k)$ является моделью Скеллама [8]

$$x_{k+1} = \frac{\alpha x_k}{\beta x_k + \gamma}. \quad (12)$$

■

Таким образом, модель Скеллама (12) также «выводима» в рамках модели (1) – (5), как и модель Морана-Риккера (11). В отличие от последней, модель Скеллама «выводима» при условии, что доля выживающих за зимний период особей подчиняется гиперболическому закону. Заметим, что данная модель «выводима» и при некоторых других законах динамики популяций, и, в частности, когда саморегуляторные механизмы в популяции подчиняются закону Ферхюльста [7]. Как и предыдущая модель, модель Скеллама (12) широко используется для описания динамики численности различных популяций [2,8,10,11].

Утверждение 4. При $R(x) = ax+b$, $Q(x) = e^{-xp}$, $p = \text{const} \geq 0$ отображение $x_{k+1} = F(x_k)$ принимает вид:

$$x_{k+1} = \frac{K_1 x_k}{K_2 x_k + K_3} \left(\frac{K_4 x_k + K_5}{K_6 x_k^2 + K_7 x_k + K_5} \right)^{\frac{p}{ha}}, \quad K_i \geq 0 \quad i = 1, \dots, 7. \quad (13)$$

■

В отличие от моделей (11) и (12) уравнение (13) получено в предположении, что действие саморегуляторных внутривидовых механизмов подчиняется закону Ферхюльста [7] (R - линейная функция), что является более реалистичным. Заметим, что модель (13) в литературе не встречается и ее построение на основе общебиологических предположений о функционировании подобных систем представляется весьма проблематичным.

Утверждение 5. При $R(x) = ax + b$, $Q(x) = \frac{1}{x+1}$, $Y = \text{const}$ отображение принимает вид

$$x_{k+1} = \frac{K_1 x_k}{K_2 x_k + K_3} \left(\frac{ha}{ha + \ln(K_6 x_k^2 + K_7 x_k + K_5) - \ln(K_4 x_k + K_5)} \right), \quad (14)$$

причем K_i , $i = 1, \dots, 7$ определяются так же, как и в утверждении 4. ■

Как и предыдущая, модель (14) также не встречается в литературе и ее построение в том виде, в котором она была получена в рамках модели (1) – (5), проблематично.

Замечание 1. Если рассмотреть случаи функций $R(x) = a \ln(x+1) + b$ (модель Гоматама [19]) и $R(x) = ax^\gamma + b$ (модель Розенцвейга [20,21]), то отображение $F(x)$ не имеет представления в элементарных функциях.

5. Анализ модели (13)

Рассмотрим отображение (13). Стационарные точки находятся из уравнения $F(x) = x$:

$$x = \frac{K_1 x}{K_2 x + K_3} \left(\frac{K_4 x + K_5}{K_6 x^2 + K_7 x + K_5} \right)^{\frac{p}{ha}}, \quad (15)$$

и, очевидно, точка $x_0 = 0$ удовлетворяет этому уравнению. Для поиска нетривиального равновесия представим (15) в следующем виде:

$$\frac{K_2 x + K_3}{K_1} = \left(\frac{K_4 x + K_5}{K_6 x^2 + K_7 x + K_5} \right)^{\frac{p}{ha}}. \quad (16)$$

Заметим, что у функции

$$\Phi(x) = \left(\frac{K_4 x + K_5}{K_6 x^2 + K_7 x + K_5} \right)^{\frac{p}{ha}}$$

все особые точки расположены на отрицательной полуоси OX ; $\Phi(x)$ пересекает ось OY в единице, ось OX при $x = -\frac{K_5}{K_4} < 0$ и при $x \geq 0$ $\Phi(x)$ положительна и монотонно убывает. Следовательно, учитывая (16), существование нетривиальной стационарной точки возможно лишь при условии $\frac{K_3}{K_1} < 1$. При этом соотношение $\frac{K_3}{K_1} = 1$ является уравнением бифуркационной поверхности. Таким образом, справедливо следующее утверждение:

Утверждение 6.

1. Если $\frac{K_3}{K_1} > 1$, т. е. $e^{b(h_1+h_2)} > Y$, то начало координат является единственной глобально устойчивой стационарной точкой отображения (13).
2. Если $e^{b(h_1+h_2)} \leq Y$, то помимо нулевой стационарной точки существует нетривиальное равновесие x' , удовлетворяющее соотношению (16). ■

Анализ поведения бифуркационной диаграммы показывает, что в рамках модели (13) кроме режимов стабилизации реализуются циклические режимы с циклами различной длины (пример бифуркационной диаграммы представлен на рис. 1; соответствующее изменение показателя Ляпунова λ [22] показано на

рис. 2). Как видно из представленных рис. 1 и рис. 2 результатов, при определенных значениях параметров модели показатель Ляпунова достаточно высок (наблюдается экспоненциальное разбегание траекторий). Численная проверка выполнения условий теоремы Даймонда [23] показывает, что в пространстве параметров имеются области, соответствующие хаотическому поведению траекторий модели (сечение пространства параметров при $Y = 224$ представлено на рис. 3).

Заключение

При построении моделей с дискретным временем исследователи исходят из предположения, что коэффициент размножения (отношение численностей двух смежных генераций) является монотонно убывающей функцией от численности популяции. При этом, естественно, пытаются выбрать наиболее простую функцию, параметры которой достаточно просто интерпретируются с «биологических позиций». Именно так построены модель Скллама (коэффициент размножения – гиперболическая функция), модель Морана-Риккера (коэффициент размножения представлен в виде монотонно убывающей экспоненциальной функции) и многие другие.

Однако при биологической интерпретации параметров модели нередко встречаются серьезные проблемы. В частности, в модели Морана-Риккера (11) параметр β обычно интерпретируется как «показатель действия саморегуляторных внутривидовых механизмов». Увеличение значения этого показателя соответствует усилению действия саморегуляторных механизмов – при той же численности популяции коэффициент размножения меньше. Но при этом возникает естественный вопрос: благодаря какому именно внутривидовому механизму произошло увеличение значения этого параметра? Это может произойти за счет снижения плодовитости особей, снижения выживаемости особей в течение зимнего периода или усиления действия саморегуляторных механизмов. В рамках дискретных моделей дать ответ на этот вопрос принципиально нельзя.

В то же время использование непрерывно-дискретных моделей типа (1)–(5) позволяет дифференцировать все регуляторные механизмы и дать однозначный ответ на поставленный вопрос. Более того, в рамках модели (1)–(5), каждый параметр которой имеет вполне определенную биологическую интерпретацию, можно получить такие дискретные модели, параметры которых практически не допускают соответствующей интерпретации (и включают в себя показатели самых разных механизмов) и построение которых непосредственно в виде рекуррентных уравнений (без использования модели (1)–(5)) представляется маловероятным.

ЛИТЕРАТУРА

1. Kostitzin V.A. *La Biologie Mathematique*. Paris: A.Colin, 1937. 236 p.
2. Недорезов Л.В. *Моделирование массовых размножений лесных насекомых*. Новосибирск: Наука, 1986. 125 с.

3. Недорезов Л.В., Недорезова Б.Н. *Модификация моделей Морана-Риккера динамики численности изолированной популяции* // Журнал общей биологии. 1984. N.4-5. С.514–517.
4. Недорезов Л.В., Неклюдова В.Л. *Непрерывно-дискретная модель динамики численности двухвозрастной популяции* // Сибирский экологический журнал. 1999. N.4. С.371–375.
5. Nedorezov L.V., Nazarov I.N. *About some models of population dynamics with nonoverlapping generations* // Advances in modeling and analysis. 1997. V.32, N.1,2. P.16–24.
6. Nedorezov L.V., Nedorezova B.N. *Correlation between models of population dynamics in continuous and discrete time* // Ecological Modelling. 1995. V.82. P.93–97.
7. Verhulst P.F. *Notice sur la loi que la population suit dans son accroissement* // Corresp. Math. et Phys. 1838. V.10. P.113–121.
8. Шаров А.А. *Моделирование динамики популяций насекомых*. ВИНТИ, Энтомология. 1986. Т.6. С.1–115.
9. Вольтерра В. *Математическая теория борьбы за существование*. М.: Наука, 1976. 288 с.
10. *Динамика численности лесных насекомых* / Исаев А.С., Хлебопрос Р.Г., Недорезов Л.В. и др. Новосибирск: Наука, 1984. 224 с.
11. Berryman A.A. *Population systems: a general introduction*. New York, Plenum Press, 1981. 222 p.
12. Смит Дж.М. *Модели в экологии*. М.: Мир, 1976. 184 с.
13. Фрисман Е.Я., Шапиро А.П. *Избранные математические модели дивергентной эволюции популяций*. М.: Наука, 1977. 150 с.
14. Шапиро А.П. *Математические модели конкуренции* // Управление и информация. Владивосток. 1974. Вып.10. С.5–75.
15. Шапиро А.П., Луппов С.П. *Рекуррентные уравнения в теории популяционной биологии*. М.: Наука, 1983. 133 с.
16. Якобсон М.В. *О свойствах однопараметрического семейства динамических систем $x \rightarrow Ax \exp(-x)$* // УМН. 1976. Т.31, N.2. С.239–240.
17. Воронцов А.И. *Патология леса*. М.: Лесная промышленность, 1978. 272 с.
18. Голубев А.В., Инсаров Г.Э., Страхов В.В. *Математические методы в лесозащите*. М.: Лесная промышленность, 1980. 104 с.
19. Gomati J. *A new model for interacting populations: I. Two species systems* // Bull. Math. Biol. 1974. V.36 P.347–353.
20. Rosenzweig M.L. *Why the prey curve has a hump* // Amer. Natur. 1969. V.103. P.81–87.
21. Rosenzweig M.L., MacArthur R.H. *Graphical representation and stability conditions of predator-prey interactions* // Amer. Natur. 1963. V.97. P.209–223.
22. Шустер Г. *Детерминированный хаос*. М.: Мир, 1988. 240 с.
23. Пых Ю.А. *Равновесие и устойчивость в моделях популяционной динамики*. М.: Наука, 1983. 182 с.

ОБ ОДНОЙ НЕПРЕРЫВНО-ДИСКРЕТНОЙ МОДЕЛИ ДИНАМИКИ ЧИСЛЕННОСТИ ДВУПОЛОЙ ПОПУЛЯЦИИ

Л.В. Недорезов, Ю.В. Утюпин

In present paper there is the consideration of parametric model of isolated population dynamics with sex structure. It's assumed that appearance of individuals of new generations is realized at any fixed time moments t_k and death rate has a continuous nature. The dynamic regimes of model are analyzed. In particular, it was obtained that cyclic and chaotic regimes can be realized in model under certain values of it's parameters.

Введение

Анализ динамики численности двуполой популяции представляется крайне важной задачей не только с теоретических позиций, но и с практических. Различные методы управления численностью вредных видов насекомых (метод выпуска стерильных самцов, феромонные ловушки и др. [1–12]) ориентированы именно на создание определенного дисбаланса в половой структуре популяции, что способствует снижению скорости ее размножения и нередко приводит к вырождению. Разработка моделей данного типа представляется также весьма актуальной и для решения отдельных задач эпидемиологии.

Модели динамики численности двуполой популяции разрабатывались в основном для случаев непрерывного размножения [1,3,4,11]. Однако модели данного типа, построенные преимущественно как системы обыкновенных дифференциальных уравнений, представляются малопригодными для описания динамики большинства видов, для которых характерна сезонная приуроченность (дискретность процессов размножения). Для таких видов более приемлемыми являются математические модели, построенные на основе систем обыкновенных дифференциальных уравнений с импульсами [5,13–15]. Рассматриваемая в настоящей работе модель построена именно в таком виде. В рамках модели предполагается, что поколения не перекрываются и появление особей новой генерации сопровождается гибелью предыдущих.

© 2002 Л.В. Недорезов, Ю.В. Утюпин

E-mail: leo@nsu.ru, adm.cnigri@alrosa-mir.ru

Институт молекулярной биологии и биофизики СО РАН

Мирнинский политехнический институт Якутского государственного университета

1. Описание модели

Пусть t_k , $k = 0, 1, 2, \dots$, моменты времени появления особей нового поколения, $t_{k+1} - t_k \equiv \text{const} = h > 0$. Будем предполагать, что между этими моментами времени (на интервалах $[t_k, t_{k+1})$) происходит монотонное снижение численностей особей обоих полов (в результате естественной смертности и действия внутри-популяционных саморегуляторных механизмов), что описывается следующей системой уравнений:

$$\begin{aligned}\frac{dx}{dt} &= -\alpha_1 x - \beta_1 x(x + \gamma y), \\ \frac{dy}{dt} &= -\alpha_2 y - \beta_2 y(x + \gamma y),\end{aligned}\tag{1}$$

где $x(t)$ – численность мужских, $y(t)$ – численность женских особей в популяции в момент времени t , α_j – коэффициенты естественной гибели особей, β_j – коэффициенты саморегуляции. Коэффициент γ отражает неравнозначность «вклада» особей различных полов в процесс саморегуляции численностей, $\alpha_j, \beta_j, \gamma > 0$. Заметим, что (1) – это аналог известной модели Ферхюльста, учитывающий наличие в популяции соответствующей структуры.

Обозначим через $x(t_k - 0)$, $y(t_k - 0)$ – численности особей соответствующих полов, выживших к моменту размножения t_k , через f – численность оплодотворенных самок. В наиболее простом случае величина f будет определяться соотношением:

$$f = \min \{y(t_k - 0), \varepsilon x(t_k - 0)\},$$

где ε – «коэффициент активности» самцов, который отражает не только их потенциальные возможности, но и характер взаимодействия особей различных полов. В частности, если все особи строго разбиваются на пары, то $\varepsilon = 1$.

Пусть m_1 , m_2 – среднее число потомков мужского и женского полов соответственно, порождаемых одной оплодотворенной самкой, $m_1, m_2 = \text{const} > 0$. Тогда в моменты времени t_k появления особей новых генераций выполняются соотношения:

$$\begin{aligned}x(t_k) &= m_1 f, \\ y(t_k) &= m_2 f.\end{aligned}\tag{2}$$

Будем считать, что в начальный момент времени численности обоих полов в популяции положительны, $x(t_0) = x_0 > 0$, $y(t_0) = y_0 > 0$. Очевидно, если $x_0 = 0$ или $y_0 = 0$, за время h траектории модели (1)–(2) «падают» в стационарное состояние $(0, 0)$ (популяция вырождается). Также заметим, что, не уменьшая общности, можно считать $\varepsilon = 1$ и $h = 1$.

2. Свойства модели (1)–(2)

1. Учитывая, что

$$\left. \frac{dx}{dt} \right|_{x=0} = 0, \quad \left. \frac{dy}{dt} \right|_{y=0} = 0$$

и $m_1, m_2 > 0$, решения системы (1)–(2) неотрицательны.

2. Заметим, что решения уравнения

$$\frac{dx}{dt} = -\alpha_1 x - \beta_1 x^2 \quad (3)$$

со следующими условиями в точках разрыва траекторий

$$x(t_k) = m_1 x(t_k - 0) \quad (4)$$

и начальными данными $x(t_0) = x_0 > 0$ ограничивают решения первого уравнения системы (1) – (2). Для решения уравнения (3) – (4) существует притягивающее множество

$$\Delta_1 = \left[0, \frac{\alpha_1(m_1 - e^{\alpha_1})}{\beta_1(e^{\alpha_1} - 1)} \right],$$

и, следовательно, переменная x системы (1) – (2) также будет «притягиваться» полосой Δ_1 . Аналогично можно показать, что переменная y системы (1) – (2) будет «притягиваться» полосой Δ_2 :

$$\Delta_2 = \left[0, \frac{\alpha_2(m_2 - e^{\alpha_2})}{\beta_2 \gamma (e^{\alpha_2} - 1)} \right].$$

Таким образом, все решения системы (1) – (2) при любых конечных начальных значениях переменных асимптотически «входят» в прямоугольник

$$\Delta = \Delta_1 \times \Delta_2 = \left[0, \frac{\alpha_1(m_1 - e^{\alpha_1})}{\beta_1(e^{\alpha_1} - 1)} \right] \times \left[0, \frac{\alpha_2(m_2 - e^{\alpha_2})}{\beta_2 \gamma (e^{\alpha_2} - 1)} \right]$$

и никакое решение с начальными данными, лежащими в Δ , не может выйти за его границы.

Заметим, что прямоугольник Δ существует только тогда, когда выполняются условия

$$m_1 > e^{\alpha_1}, m_2 > e^{\alpha_2}. \quad (5)$$

Если хотя бы одно из условий (5) не выполняется, то $x \rightarrow 0, y \rightarrow 0$ при $t \rightarrow \infty$ (популяция вырождается при любых начальных значениях численностей).

3. Начало координат является стационарной точкой системы (1)–(2). Если условия (5) выполняются, то она неустойчива, иначе – асимптотически устойчива.

4. Из условий (2) следует, что в моменты времени t_k выполняется соотношение

$$x(t_k) = \frac{m_1}{m_2} y(t_k). \quad (6)$$

Кроме этого, для системы (1) существует первый интеграл

$$\frac{x^{\beta_2}}{y^{\beta_1}} e^{(\alpha_1 \beta_2 - \alpha_2 \beta_1)t} = C, \quad (7)$$

где C – константа интегрирования. Подставляя в (7) начальные данные, получаем следующее соотношение:

$$\frac{x^{\beta_2}}{y^{\beta_1}} = e^{(\alpha_2\beta_1 - \alpha_1\beta_2)(t-t_k)} \frac{x^{\beta_2}(t_k)}{y^{\beta_1}(t_k)}. \quad (8)$$

С учетом соотношения (6) получаем, что на отрезках $[t_k, t_{k+1})$ выполняется равенство

$$x = y^{\frac{\beta_1}{\beta_2}} e^{\frac{(\alpha_2\beta_1 - \alpha_1\beta_2)}{\beta_2}(t-t_k)} \frac{m_1}{m_2} y^{\frac{\beta_2 - \beta_1}{\beta_2}}(t_k). \quad (9)$$

С помощью соотношений (6) и (9) из системы (1)–(2) можно исключить переменную x :

$$\frac{dy}{dt} = -\alpha_2 y - \beta_2 y \left(y^{\frac{\beta_1}{\beta_2}} \frac{m_1}{m_2} e^{\frac{(\alpha_2\beta_1 - \alpha_1\beta_2)}{\beta_2}(t-t_k)} y^{\frac{\beta_2 - \beta_1}{\beta_2}}(t_k) + \gamma y \right), \quad (10)$$

$$y(t_{k+1}) = m_2 \min \left\{ y(t_{k+1} - 0), y^{\frac{\beta_1}{\beta_2}}(t_{k+1} - 0) e^{\frac{(\alpha_2\beta_1 - \alpha_1\beta_2)}{\beta_2}(t_{k+1} - t_k)} \frac{m_1}{m_2} y^{\frac{\beta_2 - \beta_1}{\beta_2}}(t_k) \right\}. \quad (11)$$

Далее будем предполагать, что $\frac{\beta_1}{\beta_2} < 1$. Если выполняется обратное неравенство, то, замечая, что x и y входят в систему (1)–(2) симметрично, можно привести ее к виду (10)–(11) относительно переменной x . Пусть

$$B = \frac{\beta_1}{\beta_2}, \quad D = \frac{\alpha_2\beta_1 - \alpha_1\beta_2}{\beta_2}, \quad M = \frac{m_1}{m_2}.$$

С учетом сделанных предположений система (10)–(11) примет вид:

$$\frac{dy}{dt} = -\alpha_2 y - \beta_2 y (y^B M e^{D(t-t_k)} y^{1-B}(t_k) + \gamma y), \quad (12)$$

$$y(t_{k+1}) = m_2 \min \{ y(t_{k+1} - 0), y^B(t_{k+1} - 0) e^D M y^{1-B}(t_k) \}. \quad (13)$$

Система (12)–(13) представляет собой одно неавтономное уравнение, траектории которого имеют периодические разрывы.

Рассмотрим отдельно уравнение (12). На каждом временном интервале $[t_k, t_{k+1})$ решается следующая задача Коши:

$$\frac{dy}{dt} = -\alpha_2 y - \beta_2 y (y^B M e^{D(t-t_k)} \mu^{1-B} + \gamma y), \quad y(t_k) = \mu. \quad (14)$$

Сделав в (14) замену переменных $y = \mu u$, получаем задачу

$$\frac{du}{dt} = -\alpha_2 u - \beta_2 \mu (u^{1+B} M e^{D(t-t_k)} + \gamma u^2), \quad u(t_k) = 1. \quad (15)$$

В соответствии со сделанными предположениями $0 \leq t - t_k \leq 1$ величину $M e^{D(t-t_k)}$ можно ограничить снизу и сверху некоторыми положительными константами N_1 и N_2 . Соответственно, рассмотрим две задачи Коши:

$$1) \quad \frac{dv}{dt} = -\alpha_2 v - \beta_2 \mu N_1 v^{1+B}, \quad v(t_k) = 1, \quad (16)$$

$$2) \quad \frac{dw}{dt} = -\alpha_2 w - \beta_2 \mu (N_2 + \gamma) w^{1+B}, \quad w(t_k) = 1. \quad (17)$$

Очевидно, решение задачи Коши (15) лежит между решениями задач Коши (16) и (17), т. е. $\forall t \in (t_k, t_{k+1}) \quad w < u < v$. Решая задачи (16) и (17), получаем

$$v(t) = \left(\frac{\alpha_2}{\alpha_2 e^{\alpha_2 B t} + (e^{\alpha_2 B t} - 1) \beta_2 \mu N_1} \right)^{\frac{1}{B}}, \quad (18)$$

$$w(t) = \left(\frac{\alpha_2}{\alpha_2 e^{\alpha_2 B t} + (e^{\alpha_2 B t} - 1) \beta_2 \mu (N_2 + \gamma)} \right)^{\frac{1}{B}}. \quad (19)$$

Таким образом, из (18) и (19) следует, что для модели (1)–(2) выполняются неравенства:

$$y(t_k) \left(\frac{a}{C_1 y(t_k) + 1} \right)^{\frac{1}{B}} < y(t_{k+1} - 0) < y(t_k) \left(\frac{a}{C_2 y(t_k) + 1} \right)^{\frac{1}{B}}, \quad (20)$$

где параметры имеют вид:

$$a = e^{-B\alpha_2}, \quad C_1 = \frac{\beta_2}{\alpha_2} (N_2 + \gamma) (1 - e^{-B\alpha_2}), \quad C_2 = \frac{\beta_2}{\alpha_2} N_1 (1 - e^{-B\alpha_2}).$$

Рассмотрим сначала соотношения (12)–(13) при выполнении условия

$$e^D M \geq 1. \quad (21)$$

При выполнении неравенства (21) выражение (13) преобразуется к виду

$$y(t_{k+1}) = m_2 y(t_{k+1} - 0). \quad (22)$$

Пусть $y_k = y(t_k)$. Уравнения (12), (22) определяют рекурсивную зависимость $y_{k+1} = g(y_k)$. Пользуясь теоремой о непрерывной зависимости от начальных данных и непрерывностью функции, определяющей уравнения разрыва, можно показать непрерывность функции $g(y)$.

Неравенства (20) в данном случае принимают вид:

$$m_2 y_k \left(\frac{a}{C_1 y_k + 1} \right)^{\frac{1}{B}} < g(y_k) < m_2 y_k \left(\frac{a}{C_2 y_k + 1} \right)^{\frac{1}{B}}. \quad (23)$$

Уравнения

$$y_{k+1} = \frac{m_2 a^{\frac{1}{B}} y_k}{(C_i y_k + 1)^{\frac{1}{B}}} \quad (24)$$

представляют собой известные дискретные уравнения Хасселла, используемые для моделирования динамики численности изолированных популяций [2,3,13,16]. Приведем некоторые свойства уравнения (24).

i). При выполнении условия

$$m_2 a^{\frac{1}{B}} \leq 1 \quad (25)$$

уравнение (24) имеет в неотрицательной части прямой единственное глобально устойчивое равновесие $y = 0$.

ii). Если выполняется условие

$$m_2 a^{\frac{1}{B}} > 1,$$

то $y = 0$ – неустойчивая стационарная точка; кроме нее появляется нетривиальное состояние равновесия уравнения (24)

$$\bar{y}_i = \frac{am_2^B - 1}{C_i}.$$

iii). При выполнении неравенства

$$\frac{1}{B} \leq 2 \quad (26)$$

$\bar{y}_i$ – глобально устойчивое состояние равновесия.

iv). Если справедливо неравенство

$$\frac{1}{B} > 2, \quad (27)$$

то с увеличением значения параметра m_2 в модели (24) возникает бесконечная серия бифуркаций удвоения периода с рождением хаотических режимов.

Заметим, что условие (25) совпадает с условием вырождения популяции, когда $m_2 \leq e^{\alpha_2}$. Покажем, что для уравнения (12), (22) эти свойства также будут верны.

Предложение 1. Если выполняются условия (5), то (12), (22) имеет единственное нетривиальное состояние равновесия.

Доказательство. Пусть $g(y)$, как и прежде, определяет рекурсивную зависимость $y_{k+1} = g(y_k)$, определяемую уравнением (12), (22). Тогда из неравенств (23) следует, что для любого y справедливы соотношения

$$m_2 y \left(\frac{a}{C_1 y + 1} \right)^{\frac{1}{B}} - y < g(y) - y < m_2 y \left(\frac{a}{C_2 y + 1} \right)^{\frac{1}{B}} - y,$$

откуда следует, что $g(\bar{y}_1) - \bar{y}_1 > 0$ и $g(\bar{y}_2) - \bar{y}_2 < 0$, где $\bar{y}_i$ – неподвижные точки уравнений (24) ($i = 1, 2$). Отсюда следует, что между $\bar{y}_1$ и $\bar{y}_2$ существует $\bar{y}$ такой, что $g(\bar{y}) - \bar{y} = 0$. Последнее означает, что $\bar{y}$ – состояние равновесия уравнения (12), (22).

Для доказательства единственности этого состояния равновесия воспользуемся тем, что решение уравнения (15) является непрерывной, монотонно убывающей функцией от μ . Отсюда следует, что $\frac{g(y)}{y}$ также монотонно убывающая функция. Поэтому уравнение $\frac{g(y)}{y} = 1$ может также иметь одно решение, что и требовалось доказать. ■

Предложение 2. Если выполняются условия (5), (21), (26), то $\bar{y}$ – глобально устойчивый аттрактор уравнения (12), (22).

Доказательство. Сначала докажем, что при выполнении условий утверждения будет верно неравенство

$$-1 < g'(\bar{y}) < 1.$$

Для этого в (12), (22) сделаем замену переменных $y = e^p$. Тогда данные выражения примут вид:

$$\frac{dp}{dt} = -\alpha_2 - \beta_2 (e^{Bp} M e^{D(t-t_k)} e^{(1-B)p(t_k)} + \gamma e^p), \quad (28)$$

$$p(t_{k+1}) = \ln m_2 + p(t_{k+1} - 0). \quad (29)$$

При этом стационарные точки и периодические траектории уравнения (12), (22) при такой замене переменных переходят в стационарные точки и периодические траектории уравнения (28), (29) соответственно.

Соотношения (28), (29) определяют рекурсивную функцию $p_{k+1} = q(p_k)$, где $p_k = p(t_k)$, $q(p)$ – непрерывная функция. Очевидно, $q'(e^p) = g'(y)$. Необходимо показать, что если $\bar{p}$ – стационарная точка отображения (28), (29), то тогда $-1 < q'(\bar{p}) < 1$. Для доказательства этого рассмотрим следующую задачу Коши:

$$\frac{dp}{dt} = -\alpha_2 - \beta_2 (e^{Bp} M e^{Dt} e^{(1-B)\lambda} + \gamma e^p), \quad (30)$$

$$p(0) = \lambda. \quad (31)$$

Пусть $\rho(\lambda) = p(1)$. Тогда $q(p) = \ln m_2 + \rho(p)$, $q'(p) = \rho'(p)$. В (30), (31) сделаем еще одну замену переменных $p = \lambda + r$:

$$\frac{dr}{dt} = -\alpha_2 - \beta_2 e^\lambda (e^{Br} M e^{Dt} + \gamma e^r), \quad (32)$$

$$r(0) = 0. \quad (33)$$

Для $\frac{\partial r}{\partial \lambda}$, очевидно, выполняются соотношения:

$$\dot{\frac{\partial r}{\partial \lambda}} = -\beta_2 e^\lambda (e^{Br} M e^{Dt} + \gamma e^r) - \beta_2 e^\lambda (B e^{Br} M e^{Dt} + \gamma e^r) \frac{\partial r}{\partial \lambda}, \quad (34)$$

$$\frac{\partial r}{\partial \lambda}(0) = 0. \quad (35)$$

Из вида выражений (34), (35) получаем, что

$$-\frac{1}{B} < \left. \frac{\partial r}{\partial \lambda} \right|_{t=1} \leq 0. \quad (36)$$

Кроме этого, при $\lambda \rightarrow -\infty$ имеем

$$\left. \frac{\partial r}{\partial \lambda} \right|_{t=1} \rightarrow 0.$$

Рассмотрим систему

$$\begin{aligned} \dot{r}_m &= -\beta_2 e^\lambda e^{Br_m} Me^{Dt}, \\ \frac{\partial \dot{r}_m}{\partial \lambda} &= -\beta_2 e^\lambda e^{Br_m} Me^{Dt} \left(1 + B \frac{\partial r_m}{\partial \lambda} \right). \end{aligned}$$

Интегрируя последнюю систему, получаем, что

$$\frac{\partial r_m}{\partial \lambda} \Big|_{t=1} \rightarrow -\frac{1}{B}$$

при $\lambda \rightarrow \infty$. Очевидно, что

$$\frac{\partial r_m}{\partial \lambda} \Big|_{t=1} > \frac{\partial r}{\partial \lambda} \Big|_{t=1}$$

для любого λ . Следовательно, учитывая неравенство (36), получаем, что

$$\frac{\partial r}{\partial \lambda} \Big|_{t=1} \rightarrow -\frac{1}{B}.$$

Поскольку

$$\rho'(\lambda) = 1 + \frac{\partial r}{\partial \lambda} \Big|_{t=1},$$

величина $\rho'(\lambda)$ меняется от 1 до $1 - \frac{1}{B}$, когда λ меняется от $-\infty$ до ∞ . По условию утверждения $\frac{1}{B} \leq 2$, поэтому $q'(p) = \rho'(p) \in (-1, 1)$. Отсюда следует, что $-1 < g'(\bar{y}) < 1$. Следовательно, $\bar{y}$ – глобально устойчивое состояние равновесия. ■

Замечание 1. Рассматривая уравнение в вариациях для $\frac{\partial^2 r}{\partial \lambda^2}$, можно показать, что

$$\frac{\partial^2 r}{\partial \lambda^2} \Big|_{t=1} < 0,$$

а, следовательно, для $\forall p$ $q''(p) < 0$.

Предложение 3. При выполнении условий (5), (21) и $\frac{1}{B} > 2$ с увеличением значений параметра m_2 состояние равновесия $\bar{y}$ теряет устойчивость и происходит бифуркация рождения устойчивого цикла длины два.

Доказательство. Из доказательства утверждения 2 при $m_2 \rightarrow \infty$ $g'(\bar{y}) \rightarrow 1 - \frac{1}{B}$, следовательно, при некотором m_2 производная в стационарной точке становится меньше -1, что и требовалось доказать. ■

Покажем, что если условие (21) не выполняется, то тем не менее все утверждения остаются в силе. Для этого снова рассмотрим задачу Коши (14) и обозначим через $f(\mu)$ решение этой задачи в точке $t = 1$. Тогда функция

$$g(y) = m_2 \min \{ f(y), f(y)^B y^{1-B} Me^D \} \quad (37)$$

определяет рекурсивную зависимость, определяемую (12) – (13). Для функции $g(y)$ имеем неравенство, аналогичное (23):

$$\min \left\{ m_2 y \left(\frac{a}{C_1 y + 1} \right)^{\frac{1}{B}}, \frac{m_2 M a e^D y}{C_1 y + 1} \right\} < g(y) < \\ < \min \left\{ m_2 y \left(\frac{a}{C_2 y + 1} \right)^{\frac{1}{B}}, \frac{m_2 M a e^D y}{C_2 y + 1} \right\}. \quad (38)$$

Поскольку выполняется соотношение $m_2 M a e^D = m_1 e^{-\alpha_1}$, то при выполнении условий (5) ограничивающие $g(y)$ отображения (38) имеют нетривиальные неподвижные точки. Следовательно, $g(y)$ также имеет нетривиальную неподвижную точку (это можно показать аналогично доказательству утверждения 1).

Для доказательства утверждений 2 и 3 необходимо сделать в (14) замену $y = \mu i$ и рассмотреть задачу Коши (15). Пусть $\phi(\mu) = u|_{t=1}$, тогда

$$g(y) = m_2 y \min \{ \phi(y), \phi^B(y) M e^D \}.$$

Поскольку $B < 1$, а $\phi(y)$ монотонно убывает и $\phi(y) \rightarrow 0$ при $y \rightarrow \infty$, то при $e^D M < 1$ $\phi(y) > \phi^B(y) M e^D$ только на некотором интервале $y \in (0, y^*)$. Отсюда следует справедливость утверждения 3 (без выполнения условия (21)).

Для доказательства утверждения 2 (также без выполнения условия (21)) можно воспользоваться тем, что доказано в утверждении 2. Поскольку показано, что для $g(y)$ выполняются неравенства $-1 < g'(\bar{y}) < 1$, то для любого $\bar{y}$, такого, что $\phi(\bar{y}) = \frac{1}{m_2}$, справедливо соотношение:

$$\bar{y} \phi'(\bar{y}) + \phi(\bar{y}) > -\frac{1}{m_2}. \quad (39)$$

Требуется доказать, что для любого $\bar{y}$, такого, что

$$\phi(\bar{y}) = \left(\frac{1}{m_2 M e^D} \right)^{\frac{1}{B}},$$

выполняется неравенство $(m_2 y \phi^B(y) M e^D)' > -1$. Последнее неравенство следует из $B < 1$ и $\phi'(y) < 0$:

$$\begin{aligned} (m_2 y \phi^B(y) M e^D)' &= m_2 M e^D \phi^{B-1}(y) (\phi(y) + y B \phi'(y)) > \\ &> m_2 M e^D \phi^{B-1}(y) (\phi(y) + y \phi'(y)) > -1, \end{aligned}$$

откуда и вытекает справедливость утверждения 2.

Таким образом, доказана

Теорема 1.

1. При выполнении условий $m_1 > e^{\alpha_1}, m_2 > e^{\alpha_2}$ задача (10), (11) имеет единственное нетривиальное состояние равновесия $\bar{y}$.

2. Если $1 \leq \frac{1}{B} \leq 2$, то $\bar{y}$ – глобально устойчивое состояние равновесия. Никаких циклов в этом случае не существует.

3. Если $\frac{1}{B} > 2$, то с увеличением параметра m_2 (при неизменном значении M) $\bar{y}$ теряет устойчивость. ■

Для доказательства существования хаотических режимов при условии $\frac{1}{B} > 2$ воспользуемся теоремой П. Дيامонда [17]:

Теорема (Диамамонд). Пусть I – множество в R^N и $f : I \rightarrow R^N$ – непрерывное отображение. Предположим, что существует непустое компактное множество $X \subset I$, удовлетворяющее условиям:

а) $X \cup f(X) \subset f^2(X) \subset I$,

б) $X \cap f(X) = \emptyset$.

Тогда

I. Для каждого $k = 1, 2, \dots$ в I существует k – периодическое множество.

II. Существует несчетное множество $S \subset I$, которое не содержит периодических точек и для которого имеют место следующие соотношения:

1) $f(S) \subset S$,

2) для любых двух различных точек $p \in S$ и $q \in S$

$$\limsup_{k \rightarrow \infty} |f^k(p) - f^k(q)| > 0,$$

3) для любой точки $p \in S$ и любой периодической точки $q \in I$

$$\limsup_{k \rightarrow \infty} |f^k(p) - f^k(q)| > 0.$$

■

Таким образом, при выполнении условий а) и б) теоремы Дيامонда в системе возникают хаотические режимы.

Как было показано выше, после замены переменных $y = e^p$ в системе (12), (13) получается некоторая рекурсивная зависимость

$$p_{k+1} = F(p_k) + A, \quad (40)$$

где A – параметр, увеличение которого соответствует увеличению параметра m_2 в (12), (13). Для функции $F(p)$ показано, что при изменении p от $-\infty$ до ∞ производная $F'(p)$ монотонно меняется от 1 до $1 - \frac{1}{B}$. Следовательно, эта функция имеет единственную точку максимума.

Пусть максимум функции $F(p)$ достигается в некоторой точке p_1 и пусть параметр A достаточно большой для того, чтобы точка p_1 лежала левее точки покоя отображения (40). Обозначим за p_0 точку, которая лежит левее точки p_1 и которую отображение (40) переводит в точку p_1 . Пусть p_0, p_1, p_2, p_3 – точки траектории уравнения (40) с начальной точкой p_0 .

Учитывая то, что $\lim_{p \rightarrow -\infty} F'(p) = 1$, а $\lim_{p \rightarrow \infty} F'(p) < -1$, получаем, что при достаточно большом параметре A точка p_3 будет лежать левее точки p_0 . Заметим, что отрезок $[p_0, p_1 - \varepsilon_1]$ отображается в отрезок $[p_1, p_2 - \varepsilon_2]$, причем эти отрезки не пересекаются. Но отрезок $[p_1, p_2 - \varepsilon_2]$ отображается в отрезок $[p_2, p_3 - \varepsilon_3]$, который при достаточно малом ε_1 содержит в себе оба предыдущих отрезка. Условия теоремы Дيامонда выполняются. Следовательно, при $\frac{1}{B} > 2$ с увеличением параметра m_2 в задаче (10), (11) появляются хаотические режимы.

Теорема 2. Если $\frac{1}{B} > 2$, то с увеличением параметра m_2 (при неизменном значении M) в задаче (10), (11) появляются хаотические режимы. ■

Обобщая теоремы 1 и 2 на систему (1), (2), можно сформулировать теорему:

Теорема 3.

1. При выполнении условий $m_1 > e^{\alpha_1}, m_2 > e^{\alpha_2}$ система (1), (2) имеет единственное нетривиальное состояние равновесия $(\bar{x}, \bar{y})$.

2. Если $\frac{1}{2} \leq \frac{1}{B} \leq 2$, то $(\bar{x}, \bar{y})$ – глобально устойчивое состояние равновесия. Никаких циклов в этом случае не существует.

3. Если $\frac{1}{B} > 2$ или $\frac{1}{B} < \frac{1}{2}$, то с увеличением параметра m_2 (при неизменном соотношении $\frac{m_1}{m_2}$) точка $(\bar{x}, \bar{y})$ теряет устойчивость, и с дальнейшим увеличением в системе возникают бифуркации удвоения периода с появлением хаотических режимов. ■

3. Заключение

Анализ модели динамики численности изолированной популяции с половой структурой показывает, что даже в наиболее простых случаях, когда смертность особей в популяции подчиняется закону Ферхюльста, а плодовитость особей постоянна, существуют определенные значения параметров модели, при которых в системе возникают хаотические режимы. Важно отметить, что циклические и хаотические динамические режимы могут возникать только в тех случаях, когда различия в воздействии саморегуляторных механизмов на особей разных полов достигают определенного критического значения. Если же подобных различий нет, то в системе наблюдается единственное глобально устойчивое равновесие.

ЛИТЕРАТУРА

1. Базыкин А.Д. *Математическая биофизика взаимодействующих популяций*. М.: Наука, 1985. 181 с.
2. *Динамика численности лесных насекомых* / А.С. Исаев, Р.Г. Хлебопрос, Л.В. Недорезов и др., Новосибирск: Наука, 1984. 224 с.
3. *Популяционная динамика лесных насекомых* / А.С. Исаев, Р.Г. Хлебопрос, Л.В. Недорезов и др., М.: Наука, 2001. 374 с.
4. *Динамическая теория биологических популяций* / Гимельфарб и др., М.: Наука, 1974. 456 с.

5. Недорезов Л.В. *Моделирование массовых размножений лесных насекомых*. Новосибирск: Наука, 1986. 125 с.
6. Алексеев В.И., Гинзбург Р.Л. *Регулирование численности популяции биологическими методами* // Журнал общей биологии. 1969. Т.30, N.5. С.616–620.
7. Базыкин А.Д. *О сравнительной эффективности некоторых способов регуляции плотности популяции* // Журнал общей биологии. 1967. Т.26, N.4. С.463–466.
8. Брежнев А.И., Гинзбург Л.Р. *К оценке норм выпуска стерильных насекомых* // Журнал общей биологии. 1974. Т.35, N.6. С.911–916.
9. Доутт Р.Л., Де Бах П. *Некоторые теоретические положения и вопросы биологической борьбы* // Биологическая борьба с вредными видами и сорняками. М.: Колос, 1968. С.96–113.
10. Де Бах П. *Успехи, тенденции и перспективы* // Биологическая борьба с вредными видами и сорняками. М.: Колос, 1968. С.507–536.
11. Гинзбург Л.Р., Юзефович Г.И. *О динамике численности полов в двуполой популяции* // Генетика. 1968. Т.4, N.12. С.116–119.
12. Недорезов Л.В. *Влияние выпуска стерильных особей на динамику популяции* // Известия СО АН СССР, сер. биол. наук. 1983. N.10, вып. 2, С.119–122.
13. Nedorezov L.V., Nedorezova B.N. *Correlation between models of population dynamics in continuous and discrete time* // Ecological Modelling. 1995. V.82. P.93–97.
14. Aagaard-Hansen H., Yeo G.F. *A stochastic discrete generation birth, continuous death population growth model and its approximate solution* // J. of Math. Biology. 1984. V.20. P.69–90.
15. Kostitzin V.A. *La Biologie Mathématique*. Paris: A.Colin, 1937. 236 p.
16. Варли Д.К., Градуэлл Д.Р., Хасселл М.П. *Экология популяций насекомых*. М.: Колос, 1978. 222 с.
17. Diamond P. *Chaotic behaviour of systems of difference* // Int. J. Syst. Sci. 1976. V.7, N.8. P.953–956.

ПРИМЕНЕНИЕ МЕТОДА КОНЕЧНЫХ ЭЛЕМЕНТОВ В СТОМАТОЛОГИИ (ОБЗОР ЛИТЕРАТУРНЫХ ИСТОЧНИКОВ)

**В.М. Семенюк, И.Н. Путалова, А.В. Артюхов,
А.В. Сырцова, А.К. Гуц**

The world scientific press has paid more attention to publications on studies based on FEAM. This article is devoted to medical articles published by researchers from Russia, Japan, USA and other countries, who use FEAM in dentistry. FEAM applications are not exhausted. Compilers of the abstract consider that it is interesting and necessary to apply modern methods of mathematical analysis in their own scientific studies.

Самым современным методом при биомеханических исследованиях признано математическое моделирование [1,2,4,5,8]. В настоящее время большую область применения имеет метод конечных элементов (МКЭ). Наиболее важными преимуществами МКЭ, благодаря которым он широко используется, являются следующие:

1. Свойства материалов смежных элементов не должны быть обязательно одинаковыми. Это позволяет применять метод к телам, составленным из нескольких материалов.

2. Криволинейная область может быть аппроксимирована с помощью прямолинейных элементов или точно описана с помощью криволинейных элементов. Таким образом, методом можно пользоваться не только для областей с «хорошей» формой границы.

3. Размеры элементов могут быть переменными. Это позволяет укрупнить или измельчить сеть разбиения области на элементы, если в этом есть необходимость.

4. С помощью МКЭ не представляет труда рассмотрение граничных условий с разрывной поверхностной нагрузкой, а также смешанных граничных условий.

Все эти преимущества могут быть применены для точного описания биомеханических процессов живых систем, в частности зубочелюстного аппарата человека. В последнее десятилетие в мировой научной прессе все чаще встречаются публикации об исследованиях, проведенных с помощью МКЭ. Медицинские издания наиболее развитых в научно-техническом плане государств

© 2002 В.М. Семенюк, И.Н. Путалова, А.В. Артюхов,
А.В. Сырцова, А.К. Гуц

E-mail: guts@univer.omsk.su

Омская государственная медицинская академия, Омский государственный университет

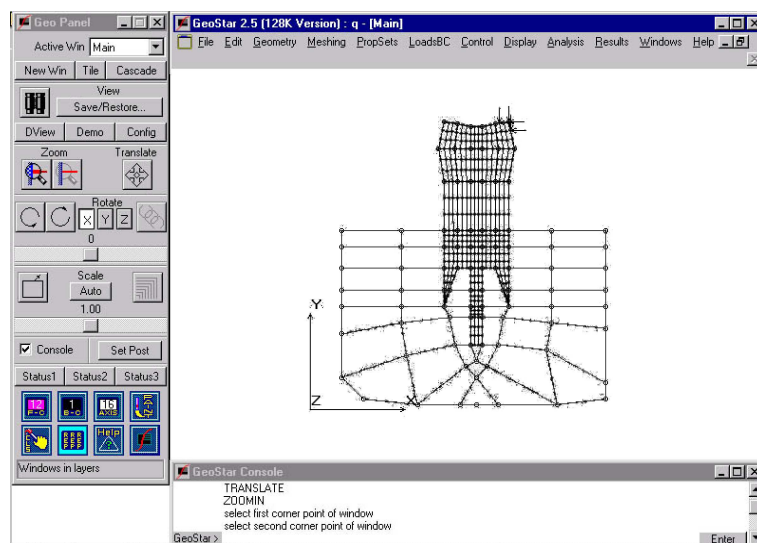


Рис. 1. Окно ППП COSMOS/M 2.5

(Япония, США, Германия и др.) не являются исключением. Большой интерес к МКЭ проявляют и врачи [2,3,5,6,8–17,19–24], производящие изыскания в области материаловедения, имплантологии, экспериментальной и практической стоматологии.

Особенно перспективным является использование компьютерного моделирования, позволяющего проводить вычислительные эксперименты на основе МКЭ, проверяя различные предположения о характере поведения зубов, имплантантов, штифтовых и других искусственных конструкций до того, как они могут быть применены при лечении больных [3,6,7,8,9 и др.].

Одной из новейших компьютерных программ на основе метода конечных элементов является пакет прикладных программ (ППП) COSMOS/M (Structural Research and Analysis Co., Santa Monica, USA.) и COSMOS Work. Пакет программ COSMOS/M позволяет автоматически осуществить расчет напряжений и деформаций любой модели, построенной пользователем. Для выполнения этих расчетов в базу данных программы необходимо ввести свойства материалов, из которых составлен моделируемый объект (это модуль Юнга, плотность материала и коэффициент Пуассона). Метрическими единицами пакета COSMOS/M являются фунты (lb) и дюймы (in). Модуль эластичности (модуль Юнга) и напряжение выражаются в lb/in^2 , плотность в $\text{lb}\cdot\text{s}^2/\text{in}^4$, силы в lb. В пакете COSMOS Work используется система СИ.

Пользователь с помощью команд меню и «мыши» задает геометрию исследуемого объекта, т.е. определяется плоскость, на которой строятся точки, линии и поверхности, задаются граничные условия. После того, как закончено построение модели, ППП COSMOS/M, используя введенные данные, обеспечивает построение системы уравнений (на основе МКЭ) и ее решение, т.е. вычисляются вектор узловых значений деформации и напряжения и визуализация результатов.

При решении любой задачи можно выделить несколько этапов:

1. Задание имени базы данных для рассматриваемой задачи. База состоит из набора файлов и располагается в той же директории, откуда был вызван GEOSTAR.
2. Задается плоскость, на которую будут наноситься точки, линии и т.д., задается рабочая сетка на плоскости (размер шага сетки, смещение, цвет).
3. Задание группы элементов, свойств материалов моделируемого объекта и констант.
4. С помощью команд меню GEOMETRY на плоскости строятся точки, линии и поверхности.
5. Активируются требуемая группа элементов, свойства материалов и наборы действительных констант.
6. Создается узловое соединение, состоящее, например, для плоских задач из PLANE2D – двумерных плоских элементов.
7. С помощью команд меню MESHING производится объединение полученных узловых соединений.
8. Ограничиваются степени свободы моделируемого объекта.
9. Прикладываются силы (нагрузки).
10. Выполняется требуемый анализ выбором соответствующей команды.
11. Проверка и исследование результатов.

Одними из первых работ в России, насколько это известно авторам данного обзора, по применению МКЭ с использованием пакета COSMOS/M в стоматологии являются исследования по надежности штифтовых конструкций, опубликованные в [2,3,8,9,10,17].

В статье [2] было доказано, что разрушенные корни фронтальных зубов ниже уровня десны на $1/4$ длины корня могут быть использованы в клинической стоматологии. Рекомендуемые штифтовые конструкции были усовершенствованы за счет особого «воротничка», что повысило их надежность [3].

В работе [17] описываются эксперименты с моделью штифтовой конструкции для фронтальных зубов с диаметром штифта, равным 2 мм. Такой штифт требует расширения канала зуба, и это наводит на мысль, что такая штифтовая конструкция будет способствовать ослаблению прочностных характеристик зуба. Это должно повлечь его разрушение. Насколько справедливо это предположение, и проверялось в ходе проделанных исследований.

В статье [18] «проведено компьютерное исследование 99 типов математических моделей зубов, армированных широко распространенными в стоматологической практике культевыми штифтовыми вкладками (анкерные штифты – первая серия эксперимента и литые культевые штифтовые вкладки, предложенные авторами, – вторая серия). Модели зубов различались формой, размерами и свойствами конструкционного материала. Доказано, что форма и параметры культевой штифтовой вкладки важнее свойств конструкционного материала. Установлено, что выбор штифтовой конструкции зависит от клинической ситуации, анатомо-топографических особенностей разрушения коронки зуба и патологии зубочелюстно-лицевой системы».

Анализ «возможности повышения эффективности ортопедического лечения путем использования внутрикостных имплантатов известных и перспективных

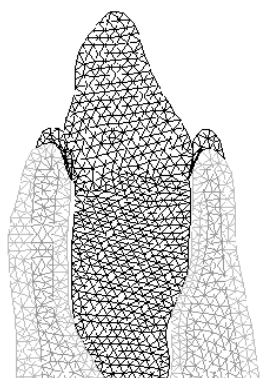


Рис. 2. Плоская конечно элементной модели зубочелюстного сегмента (деформированное состояние при действии горизонтальной силы) [13]

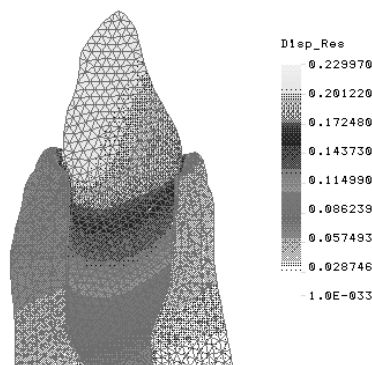


Рис. 3. Поле суммарных перемещений при действии вертикальной силы [13]

конструкций» приводится в статье [19]. «Излагается методика математического моделирования процессов взаимодействия имплантантов с костной тканью. Предлагаются численная модель и прикладная программа, разработанная на базе метода конечных элементов, позволяющие проанализировать напряженно-деформированное состояние кости и определить предельно допустимые нагрузки на имплантант. Показано, что наиболее нагруженной зоной является слой компактной кости, непосредственно примыкающий к имплантанту в области его шейки, что согласуется с результатами клинических исследований, согласно которым именно в этой зоне наиболее высок процент осложнений. Обсуждаются способы дальнейшей оптимизации имплантантов и протезных конструкций» [19].

В сферу интересов современной стоматологии входит повышение эффективности ортопедического лечения больных на основе перспективных конструкций имплантантов, в частности внутрикостных. Эта задача решается группой отечественных исследователей на основе многофакторной оценки состояния зубочелюстной системы с использованием математического моделирования МКЭ, результаты изложены в журнале «Стоматология» [20].

Модель «зуб-имплантант-пародонт» представляет собой сложную в геометрическом и физическом отношении систему, расчет которой возможен только численным методом. При этом МКЭ наиболее предпочтителен, т.к. позволяет учитывать реальные особенности структуры биомеханического объекта и легко адаптируется к изменению формы и внешних нагрузок. Для моделирования напряженно-деформированных состояний челюстной кости при различных вариантах дентальных имплантантных систем группа российских авторов предлагает подход, сочетающий пространственную теорию упругости, МКЭ, цифровую обработку изображений, полученных на компьютерном томографе, представление результатов анализа с помощью средств машинной графики. Результаты представлены в виде атласа с цветными иллюстрациями [21].

Исследования на Украине в области стоматологии, опирающиеся на моделирование с применением МКЭ, представлены в статьях [11–15]. В работе [13] изучаются физиологическая и патологическая подвижности зубов (рис.1,2).

Примером описания использования ППП COSMOS/М в стоматологии является статья японских исследователей [22]. Они сравнивали напряжения в костной ткани вокруг опоры дентальных имплантантов под действием статической вертикальной нагрузки. Объектами моделирования были мостовидные протезы, укрепленные на имплантантах в боковом отделе нижней челюсти. На основе МКЭ конструировались двумерные модели протезов разной конструкции. В случае, когда опоры мостовидного протеза располагались по прямой линии, снижение напряжений в костной ткани вокруг имплантантов наблюдалось по мере уменьшения расстояния между крайними опорами и увеличения числа опорных элементов.

В случае моделирования консольных протезов (разновидность несъемных протезов, имеющих одностороннюю опору) наблюдалось увеличение значений напряжения по сравнению с традиционными мостовидными протезами, и значительное увеличение напряжения в костной ткани вокруг опорных элементов происходило наряду с уменьшением дистанции между крайними опорами и с удлинением консольного плеча. Авторы доказывают незначительную эффективность увеличения общего числа имплантантных опор для уменьшения напряжений в костной ткани.

В статье "A facial growth analysis based on FEM employing three dimensional surface measurement by a rapid laser device" [25] описывается применение МКЭ и ППП COSMOS/М для изучения развития черепа при долгосрочном наблюдении. Целью исследования являлось трехмерное измерение (с помощью лазера), моделирование и анализ характера изменения направления напряжений в лицевом скелете на каждой стадии развития человека.

Обследовали черепа людей разного возраста (0, 3, 6, 9, 12, 18 лет) в трехмерном пространстве с помощью лазерной техники. Из 70000 точек на лицевом скелете выбирали 16 ориентиров для формирования условий применения МКЭ. Применяли ППП COSMOS/М для анализа методом конечных элементов на персональном компьютере. Изучая распределение напряжений, авторы пришли к двум выводам:

1. Уровень напряжений в возрасте 0-3 и 3-6 лет постепенно возрастает от мышечковой области до подбородочной, в то же время максимальный уровень напряженности по сравнению со всеми другими областями наблюдается в теле челюсти.

2. В общем, уровень напряженности всех областей, исключая подбородочную, назальную и тело, является равномерным. Результаты указывают на значительные изменения нижней челюсти, особенно в ранней стадии развития, направление развития: от подбородочной области к мышечковой.

Для сравнительной оценки различных стоматологических конструкций помимо лабораторных и клинических испытаний применимо и математическое моделирование, в частности МКЭ. Немецкие ученые в статье "All-ceramic resin-bonded bridges. A 3-dimensional finite-element analysis study" [26] определили

устойчивость к нагрузкам несъемных мостовидных протезов различной конструкции.

По мнению авторов, протезы с металлическим каркасом имеют ряд недостатков: серое мерцание металла сквозь облицовку, высокий уровень коррозии, значительный аллергенный потенциал, особенно в случае применения недрагоценных сплавов. Современные технологии позволяют облицовывать керамическими массами полимерные каркасы. Цель исследования: выяснить влияние направления нагрузки (под углом в 45 , 60 градусов) и формы промежуточной части на распределение напряжений в мостовидных протезах.

Чтобы смоделировать анатомическую ситуацию, синтезировали трехмерную модель, основываясь на методе конечных элементов. Ударная нагрузка имела значение в 250 Н и была ориентирована в орально-щечном направлении. Точка приложения нагрузки была смещена на 1,5 мм в небную сторону от режущего края. Обнаружено, что напряжение в конструкциях значительно больше при возрастании угла приложения нагрузки до 60 градусов. Тонкие перемычки (3 мм в высоту) и/или глубокая разделка в протезах обуславливают напряжения вплоть до 455 МПа (при угле в 45 градусов) или 534 МПа (при угле в 60 градусов). Отсутствие разделки, закругление краев и увеличение перемычек (4 мм) снижают напряжения до 122 МПа (45 градусов) и 143 МПа (60 градусов). Так как средняя прочность на разрыв для керамики составляет 340-400 МПа, цельнокерамические мостовидные протезы на полимерном каркасе могут быть рекомендованы только при высоте перемычек не менее 4 мм. Закругление краев и минимальная разделка способствуют уменьшению напряжений в ортопедических конструкциях.

Трехмерный анализ МКЭ проводился, чтобы оценить уровень напряженности в кости, имплантанте и его головке при применении золотых сплавов, керамических или полимерных (пластмассовых и композитных) несъемных протезов [27]. Увеличивающаяся нагрузка была приложена по оси и в щечно-язычном направлении по центру мостовидного протеза. Для золотых сплавов и керамики максимальные эквивалентные напряжения были обнаружены во всех частях модели. Почти во всех случаях в моделях с пластмассовыми протезами напряжения были подобными или выше по сравнению с двумя предыдущими материалами. Наибольший уровень напряженности был обнаружен в головках имплантантов при осевом нагружении пластмасс. Защитная роль пластмассы для системы «имплантант-кость» не была подтверждена при условиях данного исследования.

Трехмерный анализ МКЭ был применен, чтобы оценить напряжение в кости вокруг имплантанта из титана при использовании разных конструкции для протезирования частичного отсутствия зубов нижней челюсти [28]. Использовались осевое, щечно-язычное или мезиодистальное направление нагружающей силы. Для каждого из данных направлений наивысшее напряжение было характерно для консольных протезов, поддерживаемых двумя имплантантами (М2). Меньшие напряжения были рассчитаны для модели традиционного мостовидного протеза на двух опорах (М3). Наименьшее напряжение обнаружено у модели шины из трех искусственных коронок на трех имплантантах (М1).

При щечно-язычном направлении силы, приложенной к МЗ, напряжение в кортикальной кости выше по сравнению с рассчитанным для М2 при тех же условиях. При осевом или мезио-дистальном направлении силы, приложенной к модели МЗ, напряжение в кортикальной кости ниже по сравнению с моделью М1, нагруженной аналогично.

Американские ученые [24] исследовали влияние формы и размера на распределение напряженности в имплантатах из высокопрочного нитрида кремния под действием вертикальной и наклонной сил, используя трехмерный анализ МКЭ. Математические модели были построены на основе изучения серии разрезов нижней челюсти. Используя программное обеспечение "Auto-CAD", моделировали расположение имплантатов в области нижних левых жевательных зубов.

Результаты расчетов показали, что на распределение напряжений в окружающей кости влияют, в первую очередь, форма имплантата и направление действующих сил. Менее важным фактором является размер имплантата. Наибольшую площадь опоры в костной ткани имеют по сравнению с конусовидными и цилиндрическими опорами имплантаты зубчатой формы. Следовательно, вокруг зубчатых имплантатов наблюдают низкое сжимающее напряжение.

При увеличении диаметра и длины имплантатов сжимающее напряжение уменьшается. Распределение напряжения вокруг зубчатых имплантатов равномернее по сравнению с цилиндрическими и конусовидными. Результаты исследований при двух направлениях сил (вертикальном и наклонном) показали, что сжимающее напряжение в кортикальной кости вокруг шейки имплантата выше, чем в ячеистой кости на протяжении имплантата. Наиболее высокими являются сжимающие напряжения при наклонном приложении силы. Данное исследование предоставляет первую информацию о взаимосвязях между формой имплантатов из нитрида кремния и напряжениях в окружающей кости.

Помимо сообщений о моделировании искусственных конструкций и имплантатов, в иностранной литературе есть сведения о применении МКЭ для расчетов биомеханических показателей естественных зубов. Примером может служить статья под названием: "Analysis of a central maxillary incisor by using a three-dimensional finite element method" [23]. Это попытка определить напряжения в верхнем первом резце, основываясь на МКЭ. Зуб моделировался изотропным, однородным, эластичным и несимметричным. Сила в 450 Н прилагалась под углом 26 градусов по отношению к длинной оси на режущем угле зуба. Выявлено распределение сжимающих, растягивающих и сдвигающих напряжений для всей структуры резца. При данных параметрах нагрузки возможны трещины и переломы твердых тканей зуба.

В работе немецких ученых [29] МКЭ применялся для анализа напряжений, возникающих в нижней челюсти. Нагрузка задавалась в трех узлах переднего зуба (рис.3,4).

Метод конечных элементов является основой исследовательского проекта "Effects of thermal and mechanical loading on restored human teeth", разрабатываемого на факультете инженерной механики Королевского колледжа (г.Лондон) [30]. Авторам проекта принадлежит серия публикаций по данной теме [31–36].

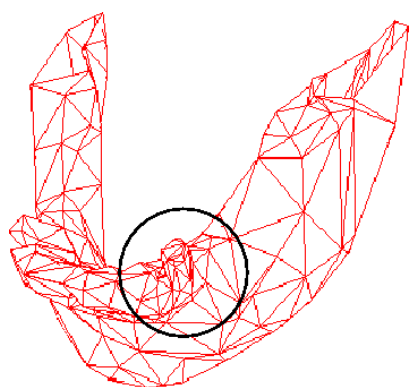


Рис. 4. Объемная конечно элементная модель нижней челюсти с одним передним зубом (225 треугольников) [29]

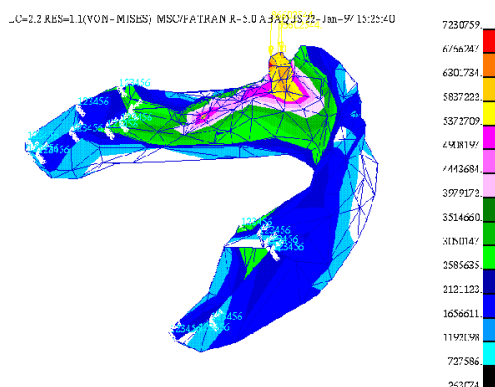


Рис. 5. Поле напряжений при действии вертикальной силы на передний зуб [29]

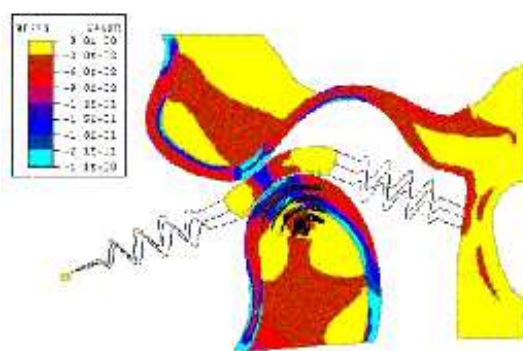


Рис. 6. Распределение напряжений при открывании рта [37]

Двумерные и трехмерные модели, основанные на МКЭ, были построены и использованы для анализа и компьютерных экспериментов по изучению биомеханики височно-нижнечелюстного сустава (ВНЧС) в работе американских исследователей [37] (рис. 5).

Таким образом, математическое моделирование является современным методом при биомеханических исследованиях. Возможности применения метода конечных элементов в стоматологических изысканиях очень широки. Об этом свидетельствует приведенный выше анализ отечественных и зарубежных литературных источников. Однако область приложения МКЭ и основанных на данном методе компьютерных программ далеко не исчерпана. Составители обзора считают интересным и необходимым применение современных методов математического анализа и в своих научных исследованиях.

ЛИТЕРАТУРА

1. Воробьев В.А. *Выбор конструкции зубных протезов и имплантантных систем на основе программного математического моделирования при лечении больных с различными дефектами зубных рядов*. Дис. д-ра мед. наук. Иркутск, 1996.
2. 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 ВИНТИ 21.06.95. N. 1790 - И95. 22 с.
3. 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 Омского университета. 1996. N.2. С.17–19.
4. Маркин В.А. *Прогнозирование осложнений при использовании металлокерамических протезов с помощью математического моделирования*. Дис. канд. мед. наук. М., 1999.
5. Олесова В.Н., Клепилин Е.С., Балгурина О.С. и др. *Биомеханическое и клиническое обоснование штифтовых конструкций на основе стекловолокна* // Панорама ортопедической стоматологии. 2001. N.2. С.4–5.
6. Олесова В.Н., Клепилин Е.С., Балгурина О.С. и др. *Сравнение биомеханики штифтовых конструкций со стекловолоконным и титановым штифтами* // Панорама ортопедической стоматологии. 2001. N.3. С.22–23.
7. Погожев И.Б. *Применение математических моделей заболеваний в клинической практике*. В кн. Вычислительные процессы системы. Вып.3 / Под ред. Г.И. Марчука. М., 1985. С.108–165
8. Семенюк В.М., Гуц А.К., Капотина Т.Н., Вагнер В.Д., Яковлев К.К. *Отдаленные результаты использования корней фронтальных зубов, разрушенных ниже уровня десны, в качестве опоры для штифтовых конструкций* // Пути развития стоматологии: итоги и перспективы. Материалы конференции 27-27 апреля 1995. Екатеринбург, 1995. С.204–206.
9. Семенюк В.М., Гуц А.К., Капотина Т.Н., Вагнер В.Д. *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. Екатеринбург: Уральская гос. мед. академия, 1997. С.211–213.
10. Капотина Т.Н., Семенюк В.М., Буланова Е.Л., Гуц А.К., Ахметянов А.Ш. *Методические подходы при использовании корней фронтальных зубов, разрушенных ниже уровня десны, для протезирования (хирургические и терапевтические мероприятия, особенности формирования культи корня зуба и получения оттисков, технология штифтовых конструкций)*. Рекомендации для врачей-стоматологов, интернов и студентов. Омск: Омская гос.мед.академия, 1996. 20 с.
11. Чуйко А.Н. *О возможностях конечно-элементного моделирования в ортопедической стоматологии* // Стоматолог. 2000. N.3. С.37–38.
12. Чуйко А.Н., Бережная Е.О., Бахуринский Н.Ю. *О современных возможностях биомеханического анализа в стоматологии* // Стоматолог. 2001. N.1–2. С.36–41.
– http://www.100matolog.com/archive/2001/1/art_37.htm

13. Чуйко А.Н., Бережная Е.О. *Подвижность и податливость зуба. Биомеханический анализ* // Стоматолог. 2001. N.4. С.15–19.
– http://www.100matolog.com/archive/2001/4/art_15.htm
14. Чуйко А.Н. *Конечно элементный анализ в стоматологии*.
– <http://fem.bi.com.ua/index.shtml>
15. Чуйко А.Н. *О роли и современных возможностях биомеханического анализа в стоматологии*. – <http://fem.bi.com.ua/article/templ.shtml>
16. Семенюк В.М., Артюхов А.В., Сырцова А.В., Гуц А.К. *Математические модели интактного моляра и моляра после гемисекции* // Математические структуры и моделирование. 2001. Вып.8. С.52–55.
17. Семенюк В.М., Гуц А.К., Панова Н.И. *Компьютерные эксперименты с моделью штифтовой конструкции с большим диаметром штифта* // Математические структуры и моделирование. 2002. Вып.9 (данный сборник).
18. Арутюнов С.Д., Чумаченко Е.Н., Лебеденко И.Ю., Арутюнов А.С. *Сравнительный анализ результатов математического моделирования напряженно-деформированного состояния различных конструкций штифтовых зубных протезов* // Стоматология. 2001. N.2. С.41–46.
– <http://www.mediasphera.ru/stomatol/2001/2/r2-01con.htm>
19. Матвеева А.И., Иванов А.Г., Гветадзе Р.Ш., Гаврюшин С.С., Карасев А.В. *Повышение эффективности ортопедического лечения больных на основе математического моделирования перспективных конструкций имплантантов* // Стоматология. 1997. N.5. С.44–48. – <http://www.mediasphera.ru/stomatol/97/5/r5-97ref.htm>
20. Матвеева А.И., Канатов В.А., Гаврюшин С.С. *Математическое моделирование ортопедических конструкций с опорой на внутрикостные имплантанты* // Стоматология. 1991. N.4. С.62–65.
21. *Атлас напряженных состояний дентальных имплантантных систем* / В.А. Воробьев, Т.М. Ратинер, А.С. Рубинов, В.В. Дудкин; под ред. М.З. Миргазизова. Ангарск, 1997. 58 с.
22. Arataki T., Adachi Y., Kishi M. *Two-dimensional finite element analysis of the influence of bridge design on stress distribution in bone tissues surrounding fixtures of osseointegrated implants in the lower molar region* // Bull. Tokyo Dent. Coll. 1998. V.39, N.3. P.199–209.
23. Darendeliler S, Darendeliler H, Kinoglu T. *Analysis of a central maxillary incisor by using a three-dimensional finite element method* // J. Oral Rehabil. 1992. V.19, N.4. P.371–383.
24. Lin S, Shi S, LeGeros RZ, LeGeros JP. *Three-dimensional finite element analyses of four designs of a high-strength silicon nitride implant* // Implant Dent. 2000. V.9, N.1. P.53–60.
25. Motegi N, Tsutsumi S, Wakatsuki E. *A facial growth analysis based on FEM employing three dimensional surface measurement by a rapid laser device* // Okajimas Folia Anat Jpn. 1996. V.72, N.6. P.323–328.
26. Pospiech P, Rammelsberg P, Goldhofer G, Gernet W. *All-ceramic resin-bonded bridges. A 3-dimensional finite-element analysis study* // Eur J. Oral Sci. 1996. V.104, N.4 (Pt.1). P.390–405.
27. Stegaroiu R, Kusakari H, Nishiyama S, Miyakawa O. *Influence of prosthesis material on stress distribution in bone and implant: a 3-dimensional finite element analysis* // Int J. Oral Maxillofac Implants. 1998. V.13, N.6. P.781–790.

28. Stegaroiu R, Sato T, Kusakari H, Miyakawa O. *Influence of restoration type on stress distribution in bone around implants: a three-dimensional finite element analysis* // Int J. Oral Maxillofac Implants. 1998. V.13, N.1. P.82–90.
29. Fütterling S., Klein R., Kremer J. *Modeling and finite element analysis of the human mandible*. – <http://www.gris.uni-tuebingen.de/gris/jb96/node20.html>.
30. Fenner D.N. *Research Project: Effects of thermal and mechanical loading on restored human teeth*. – <http://www.eee.kcl.ac.uk/mecheng/dnf/teeth.html>
31. Fenner D.N., Cheung P.W-Y., Robinson P.B. *Finite element analysis of transient heat conduction in a restored human premolar* // Proc.2nd Int.Symposium on Computer Methods in Biomechanics & Biomedical Engineering, 1994, Swansea.
32. Cheung P.W-Y., Fenner D.N., Robinson P.B. *Thermal shock in a restored premolar: a three-dimensional finite element analysis* // J.Dental Research. 1995. V.74, N.3. P.882.
33. Fenner D.N., Cheung P.W-Y., Robinson P.B. *Thermal shock in a restored premolar: a three-dimensional finite element analysis* // Proc. Annual Conference of British Society for Dental Research, April 1995, Manchester.
34. Fenner D.N., Cheung P.W-Y., Robinson P.B. *Finite element analysis of transient heat conduction in a restored human premolar*. – Computer Methods in Biomechanics & Biomedical Engineering / J.Middleton et al. Gordon & Breech, 1996. P.451-459.
35. Fenner D.N., Cheung P.W-Y., Robinson P.B. *Finite element analysis of stresses in dental tissues due to transient heat conduction* // Proc. Annual Conference of British Society for Dental Research, Bristol, April 1996.
36. Cheung P.W-Y., Fenner D.N., Robinson P.B. *Finite element analysis of dental stresses due to transient heat conduction* // J.Dental Research. 1996. V.75, N.5, P.1166–1167.
37. DeVocht J.W., Goel V.K., Zeitler D.L., Lew D., Hoffman E.A. *Development of a Finite Element Model to Simulate and Study the Biomechanics of the Temporomandibular Joint*.
38. Chen J., Xu L. *A finite element model of the human temporomandibular joint* // J.Biomechanic Engr. 1994. V.116. P.401–407.

КОМПЬЮТЕРНЫЕ ЭКСПЕРИМЕНТЫ С МОДЕЛЬЮ ШТИФТОВОЙ КОНСТРУКЦИИ С БОЛЬШИМ ДИАМЕТРОМ ШТИФТА

В.М. Семенюк, А.К. Гуц, Н.И. Панова

Mathematical model of the tooth with a cast and core with diameter 2 mm was built with the help of a special software package for simulating. The model was tested with simulating loads of various values and directions. The model shows that core with diameter 2 mm can not be strong and can not be used.

Введение

Компьютерные модели, разработанные с целью проведения вычислительных экспериментов для проверки пригодности той или иной штифтовой конструкции в стоматологической практике, имеют то неоспоримое преимущество, что эти конструкции могут быть испытаны до того, как предложены для лечения пациентов. Более того, можно «опробовать» конструкции, которые заранее считаются рискованными или просто непригодными. Компьютерная модель позволяет провести проверку интуиции стоматолога и дает возможность увидеть последствия применения интуитивно неприемлемого варианта лечения.

В этой статье описываются эксперименты с моделью штифтовой конструкции для фронтальных зубов с диаметром штифта, равным 2 мм. Такой штифт требует расширения канала зуба, и это наводит на мысль, что такая штифтовая конструкция будет способствовать ослаблению прочностных характеристик зуба, должных повлечь его разрушение. Насколько справедливы эти предположения, и проверялось в ходе проделанных исследований.

Исследования содержали следующие этапы:

1. Строилась математическая модель зуба с литым штифтом, монолитно соединенным с культевой вкладкой на уровне $h = H/4$ ниже уровня десны, где H – длина корня зуба вместе с окружающей его челюстной костью. Периодонтальная щель и слой цемента между штифтом и дентином в расчет не принималась (их толщина $< 0,15\text{--}0,2$ мм). Для упрощения расчетов на данном этапе рассматривалась плоская модель, т.е., по существу, мы пытались увидеть то, что происходит внутри зуба с штифтовой конструкцией в мысленно выделенном плоском сечении, проходящем через геометрическую ось зуба. Это

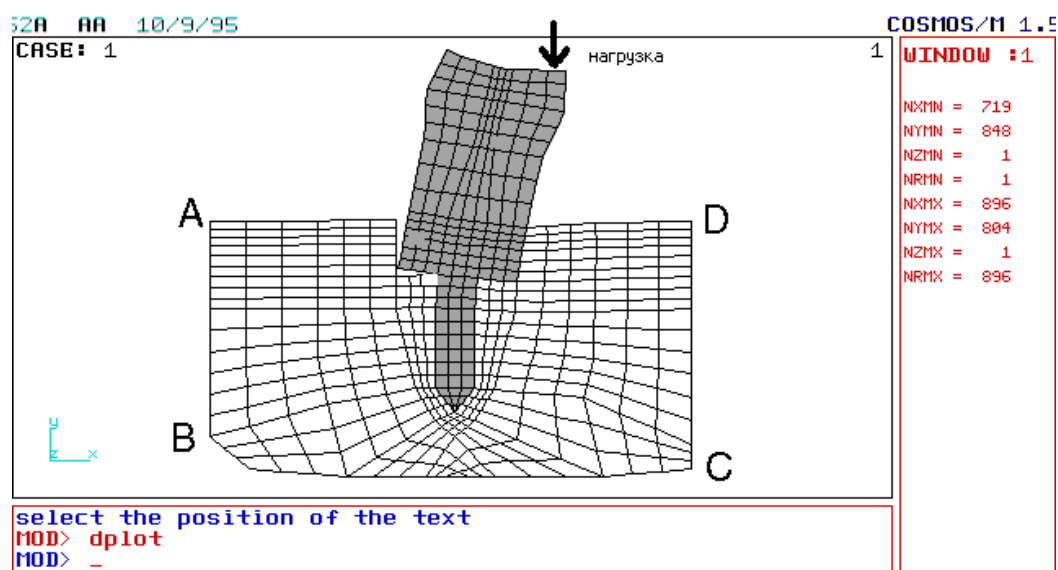


Рис. 1. Конечно-элементная модель штифтовой конструкции (под несимметричной вертикальной нагрузкой)

достаточно стандартная плоская модель упругой среды, составленная из различных материалов (рис.1), для расчета деформаций и напряжений которой под воздействием внешней статической нагрузки использовался метод конечных элементов. Появление обширных (по величине площади рассматриваемого плоского сечения сохранившейся части корня зуба) зон в корне зуба с напряжением, превосходящим предел прочности дентина (главным образом) на разрыв, трактовалось как ситуация, ведущая к разрушению остатка корня, особенно при циклическом повторе данной нагрузки (пережевывание пищи).

2. Модель реализовывалась с помощью специального компьютерного пакета прикладных программ «Космос/М» для ПЭВМ РС IBM 386/387, ¹ который позволял наглядно (на экране дисплея) увидеть, как происходит деформирование зуба и окружающей челюстной кости под влиянием заданной нагрузки (вплоть до мультипликации), а также давал возможность увидеть полную картины распределения (в виде цветных зон одинакового напряжения) напряженных состояний зуба с вкладкой.

Накладывалось условие нулевого граничного перемещения (жесткого закрепления) по линии $AB+BC+CD$. С левой стороны между костью и вкладкой узлы брались двойными – соединение между ними отсутствовало, следовательно, воздействие через образованную «узловую щель» не передавалось. Это делалось для того, чтобы при рассмотрении вектора нагрузки, принадлежащего 4-му квадранту, моделировалась ситуация, при которой считается, что стальная вкладка не срывается с живой тканью. При увеличении нагрузки щель между вкладкой и тканью увеличивается.

Нагрузка бралась точечная, сосредоточенная в трех узлах (имеющая одну и ту же величину в каждом узле).

¹Исследования проводились в 1995 году.

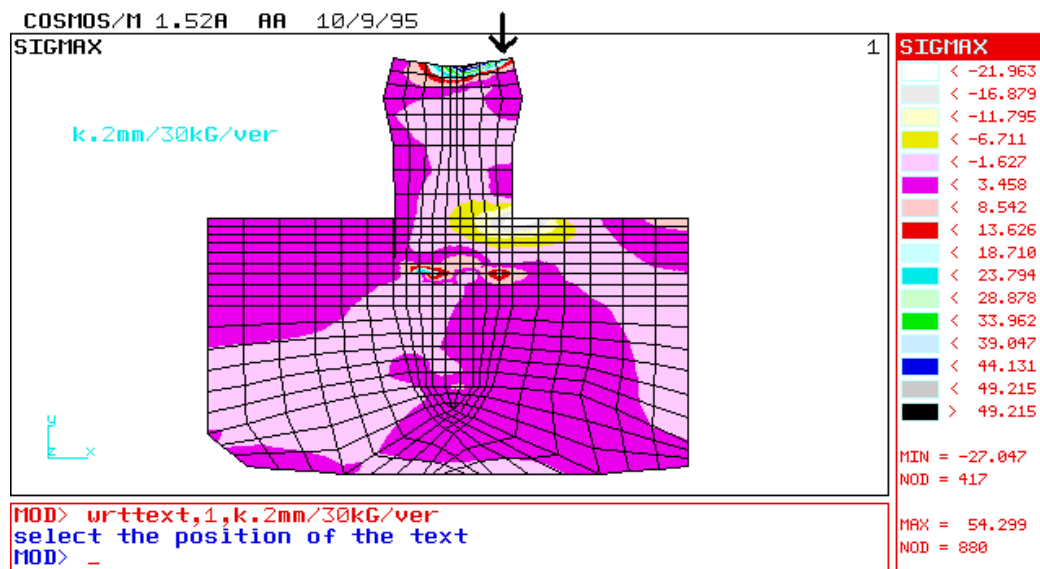


Рис. 2. Картина напряжений σ_x при вертикальной несимметричной нагрузке в 30 кГ

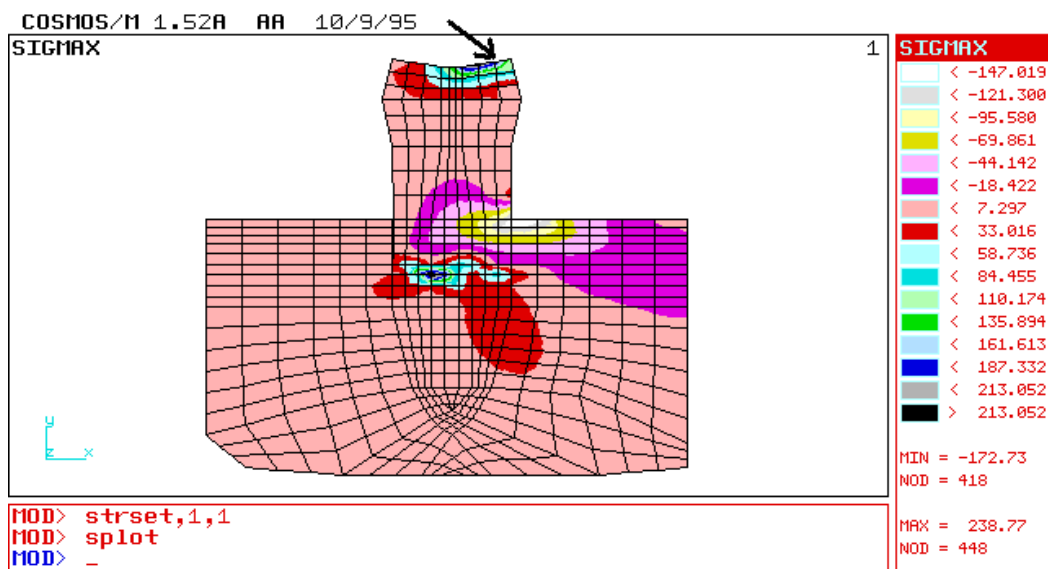
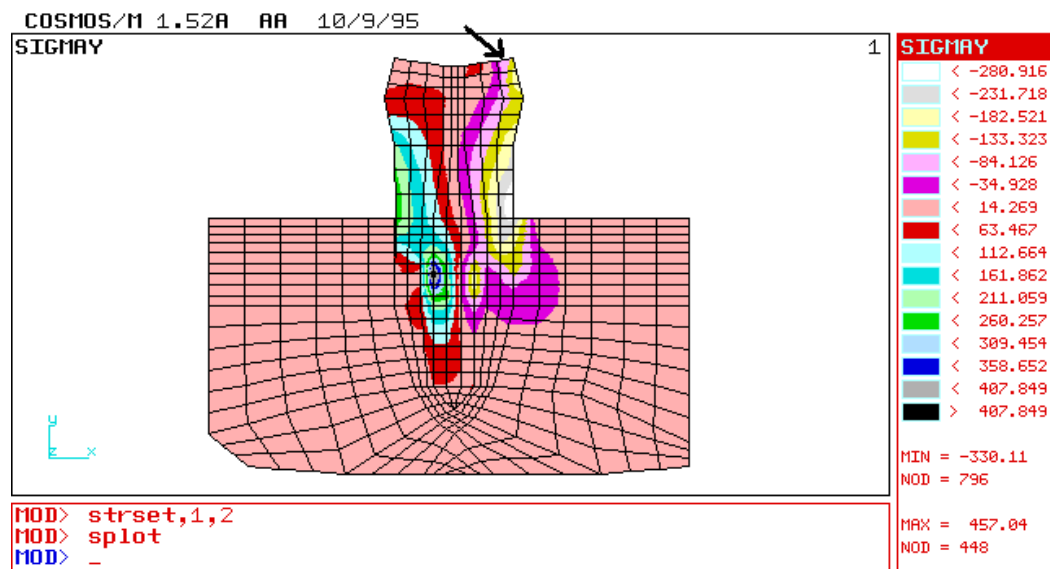
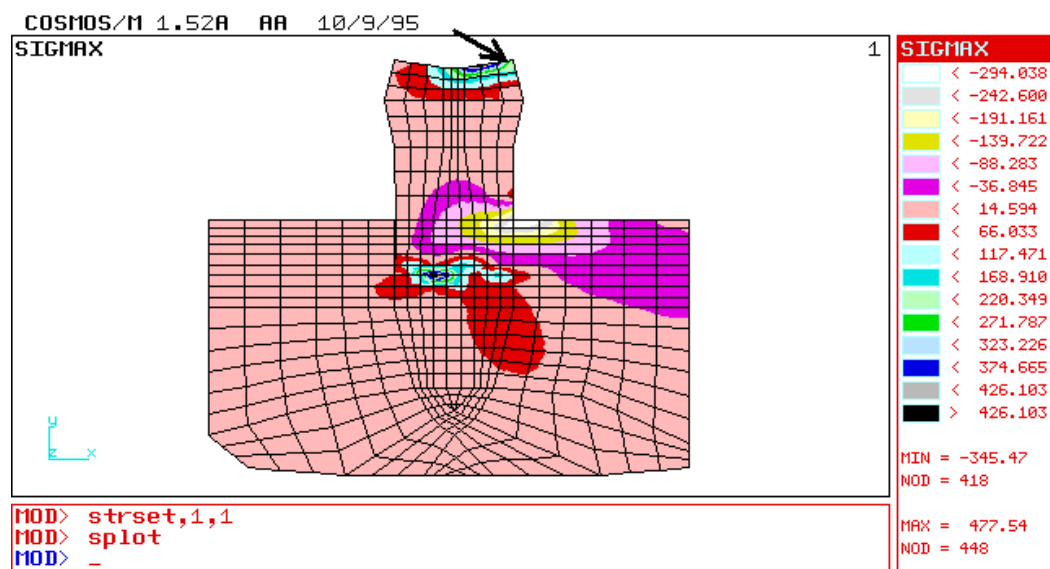


Рис. 3. Картина напряжений σ_x при боковой нагрузке в 45° в 30 кГ

Рис. 4. Картина напряжений σ_y при боковой нагрузке в 45° в 30 кГРис. 5. Картина напряжений σ_x при боковой нагрузке в 45° в 60 кГ

Была проведена серия компьютерных экспериментов при проверке надежности предложенной штифтовой конструкции.

Эксперименты показали, что для исследуемой штифтовой конструкции для $h = H/4$ при вертикальной несимметричной нагрузке до 30 кГ в районе соединения штифта с культовой вкладкой напряжения по горизонтали σ_x не попадают в критическую зону (для разрыва – это положительные напряжения от 20 до 42 н/мм² – см. рис.2).

Однако при боковой нагрузке до 30 кГ под углом в 45° (см. рис.3) в районе соединения штифта с культовой вкладкой возникают напряжения σ_x , попадающие в критическую зону. По существу, это означает реальность разрушений корня зуба. Напряжения по вертикали σ_y значительно больше – рис.4.

Как показали исследования, опубликованные в [1, 2], в случае диаметра штифта в 1 мм напряжения по горизонтали σ_x значительно меньше и не попадают в критическую зону.

Можно проводить исследование и на разрушение при сжатии (отрицательные напряжения – см. рис.2,3). Но, как правило, материал более прочен при сжатии; большие отрицательные напряжения достигались в корне лишь при нагрузке около 60 кГ. Например, при боковой нагрузке в 60 кГ под углом в 45° (см. рис.5) в районе соединения штифта с культовой вкладкой резко нарастают не только напряжения σ_x на растяжения, но и на сжатия (напряжения отрицательные), далеко выходящие за пределы критической зоны для нагрузок на растяжение. Точная оценка ситуации для напряжений на сжатие затруднялась отсутствием у авторов данных о пределе прочности дентина на сжатие. Напомним, однако, что более важно было проследить значения положительных напряжений.

Вывод: не следует использовать штифты с диаметром более 1 мм.

ЛИТЕРАТУРА

1. 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*. Ученый совет мат. фак. ОмГУ. Деп.в ВИНТИ 21.06.95. N. 1790 – В95. 22 с.
2. Семенюк В.М., Капотина Т.Н., Яковлев К.К., Гуц А.К., Панова Н.И., *Математическое обоснование к использованию культовой штифтовой вкладки с «воротничком» при разрушении корней зубов ниже уровня десны* // Вестник Омского университета. 1996. N.2. С.17–19.

ИНСТРУМЕНТ МОДЕЛИРОВАНИЯ ПРОЦЕССА ДОЗИРОВАНИЯ АКТИВНОСТИ КОМПЬЮТЕРА

Р.И. Нужин, В.А. Шапцев

The computer activity dozing algorithm and its general architecture are announced. Characteristics of a simulation system are described as a some engineering tool for a creation of effective information source dozing activity algorithms.

1. Введение

На запрос «информационная экология» поисковый сервер Rambler выдал более 500 ссылок на серверы и более 2000 ссылок на документы. Просмотр первых трех десятков документов позволяет примерно 30 процентов из них признать адекватными по содержанию действительно проблемам нового научного направления – информационной экологии. Этот факт позволяет говорить о растущей актуальности проблем, сформулированных в [1]. Одним из подходов к их решению в [2] заявлено дозирование активности источников информации. Ниже в этом контексте рассматривается область образования.

В связи с развивающейся компьютеризацией образовательного процесса, интенсивным расширением сферы действия технологии открытого образования обучающийся всех уровней обучения вынужден все больше времени проводить за компьютером. Становится актуальной проблема принудительного дозирования активности компьютера из-за неспособности человека контролировать момент перехода в состояние глубокой усталости [1], [2]. Решение проблемы требует включения в операционные среды персональных компьютеров программ контроля и предсказания состояний усталости и принудительного выключения операционной среды, но так, чтобы не нанести пользователю психологическую травму. Для разработки адекватных конкретным образовательным и проблемно-производственным условиям алгоритмов дозирования активности компьютера (АДАК) целесообразно создать имитационную программно-аппаратную систему. Ориентировочный состав ее: компьютер, цифровая видеокамера, программы тестирования пользователя и дозирования активности, программа статистической обработки результатов апробации создаваемых алгоритмов, средства дистанционной диагностики состояния человека [2]. Настоящая работа посвящена формированию проекта системы имитации дозирования компьютера (СИДАК). При этом учтены рекомендации, приведенные в литературе [3–5].

© 2002 Р.И. Нужин, В.А. Шапцев

E-mail: driiz@surguttel.ru, vash@surguttel.ru, surgpi@surguttel.ru

Сургутский государственный педагогический институт

2. Основные принципы имитации дозирования активности ПК

Ниже формулируются основные характеристики, точнее, требования к организации процесса дозирования активности компьютера.

Процесс дозирования активности компьютера должен проходить в стандартной операционной среде компьютеров, принятой в конкретной организации. Запуск программы должен проходить вместе со стартом операционной системы и сохранять все настройки ОС для конкретного пользователя. В стадии имитации АДАК разработчик должен иметь возможность отказаться от некоторых настроек, в том числе и в имитируемом алгоритме. Для этого имитационная система будет иметь ряд ответвлений и процедур отказа от части сервисных функций при возникновении серьезных проблем и неудобств и для наиболее плодотворного контакта разработчика с имитационной системой.

АДАК должен проявлять активность в отношении пользователя (интерактивность): осуществлять предупреждения его о предстоящем изменении состояния ОС и аппаратной среды компьютера, проводить тестирование состояния пользователя, вести сбор данных и их анализ. При этом имитационная система (СИДАК), отображая реальный процесс работы за компьютером, должна в реальном времени позволять разработчику алгоритма дозирования активности вводить в имитационную систему комментарии, изменения параметров, формировать описание процесса (создавать вербально-графическую модель) дозирования. С этой целью, в частности, в СИДАК создается специальное окно для оперативного введения комментариев. Они мгновенно заносятся в текстовый файл и в дальнейшем могут быть просмотрены коллективом разработчиков. То есть СИДАК должна иметь совершенный механизм интерактивности.

АДАК должен иметь дружественный (по возможности адаптирующийся) к разработчику интерфейс. Для этого в АДАКе запланировано проведение тестирования пользователя в начале каждого сеанса работы, в частности, посредством теста Люшера. В этом тесте посредством выбора пользователем цветового фрагмента определяется его эмоциональное состояние, т.е. характерная степень работоспособности. По результатам такого тестирования изменяются те или иные настройки рабочего поля экрана и длительность сеанса активности компьютера в АДАКе. Обязательным компонентом АДАКа должны быть музыкальные фрагменты стимулирования работоспособности и предотвращения стрессовых состояний. СИДАК в свою очередь должна позволять осуществлять выбор наиболее эффективных средств тестирования и стимулирования производительности труда пользователя (обучающегося в частности).

Для формирования адаптивных свойств АДАК в нем предусматривается компонент, который будет фиксировать значения перестраиваемых параметров алгоритма в каждом сеансе работы. Далее эти данные могут автоматически классифицироваться по различным типам пользователей, обрабатываться некоторым образом с целью оптимизации параметров процесса дозирования. На основе анализа этих данных можно строить обучающие варианты АДАКа, т.е. осуществлять установку значений своих параметров в соответствии с прогнозом

состояния пользователя. В СИДАК также планируется создание средств получения информации в процессе многочисленных имитационных экспериментов, ее структуризации (классификации) и использования для дальнейшего совершенствования структуры и принципов, закладываемых в АДАК.

Здесь речь идет о создании в АДАКе подсистемы отображения личности и состояния обучающегося (формализации модели пользователя). При этом тестирование пользователя с этой целью целесообразно проводить в режиме реального времени без заметного для него отвлечения от работы.

СИДАК в этом контексте служит инструментом построения эффективных средств автоматического формирования модели пользователя. То есть в составе СИДАК создаются подсистемы отображения личности и состояния обучающегося. По-видимому, целесообразно создание некоторой базы таких моделей с целью формирования в последующем некоторой экспертной системы для автоматического распознавания состояния пользователя (обучающегося) посредством сопоставления его текущей модели с определенным образом классифицированными накапливаемыми образами пользователей из предшествующих экспериментов.

В самом общем виде структуру АДАК можно представить ниже приведенной блок-схемой 1.

Прокомментируем блок-схему 1.

Производится принудительный запуск АДАКа при загрузке операционной системы. Здесь можно выбрать автоматическую установку длительностей временных интервалов k , m , l (см. блоки 2-й и 3-й) с адаптацией (по данным последующего анализа предшествующих сеансов активности) и без адаптации (фиксированные параметры, ориентированные на СНИП и внутриобъектный режим).

Здесь также возможно включение модуля тестирования пользователя по психологическим тестам и соответствующая автоматическая установка адекватных интервалов k , m , l .

Общий подход к построению предупреждений – интервалы времени активности компьютера и форма предупреждений не должны отрицательно влиять на процесс труда или обучения. Интервалы выбираются в зависимости от результатов тестовых сведений о характере, темпераменте, работоспособности и психолого-эмоциональном состоянии пользователя в некоторые фиксированные или адаптивно выбираемые моменты времени.

По результатам анализа этих данных определяется количество предупреждений, подбирается их форма.

Осуществляется блокировка рабочей станции (компьютера) и ее «минутное» включение, если пользователь действительно нуждается в завершении некоторой операции. Цель этого блока – «заставить» пользователя перейти к физически активной форме отдыха. При этом на экране терминала могут появиться сведения о благоприятном и отрицательном влиянии той или иной формы отдыха, некоторые (ненавязчивые) сведения из медицины. Рассматривается возможность демонстрации на экране различных рекомендаций эффективного времяпрепровождения. Блокировка – это временное отключение периферийных

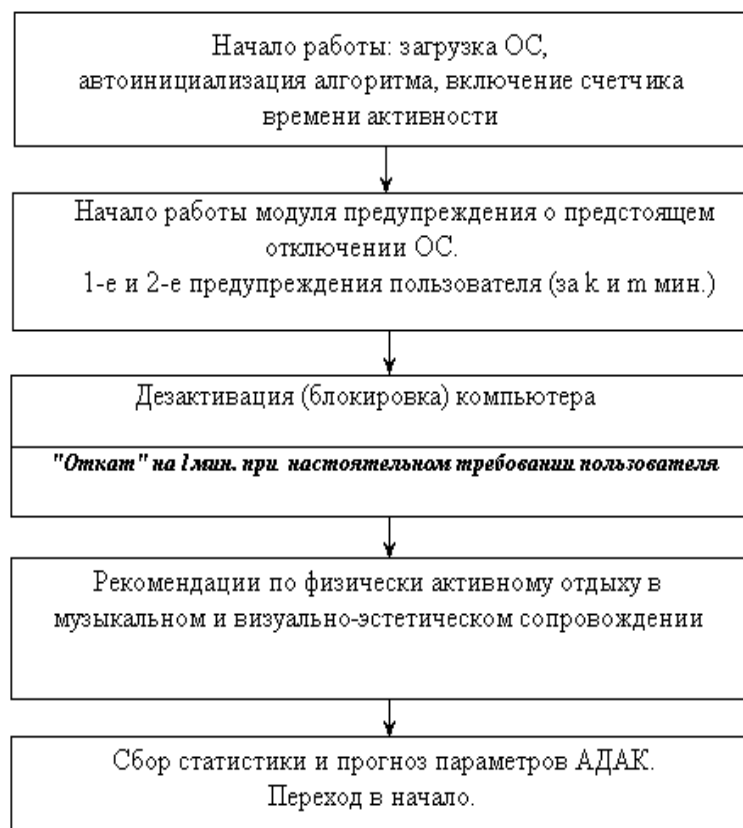


Рис. 1. Процесс дозирования активности компьютера в общем виде

устройств на программном уровне. Попытки пользователя продолжить работу не дадут успеха.

Однако, когда пользователь, например, трехкратным нажатием определенной клавиши требует восстановить активность компьютера, то включается модуль отката (модуль для забывчивых). Он позволяет вернуться к работе на время до, например, 3-х минут для сохранения или каких-либо срочных манипуляций.

На экране компьютера принудительно демонстрируется «клип», музыкальные фрагменты или другая информация, стимулирующая физически активный отдых. Если заглянуть в перспективу, то возможно создание контура управления «компьютер - массажное кресло».

Формируются данные для совершенствования параметров изменения АДАК в режиме адаптивного функционирования. Целесообразно иметь блок сбора отзывов и предложений пользователя о свойствах АДАКа с целью последующего их анализа специалистами, сопровождающими эксплуатации компьютерных станций. В сетевом варианте эта информация автоматически может накапливаться на специализированном сервере, осуществляющем мониторинг активности и работоспособности компьютерных учебных или рабочих мест.

3. Структура системы имитации алгоритмов дозирования активности компьютера

Исходя из общих принципов построения имитационных систем и учитывая комментарии, изложенные выше, структуру СИДАК можно представить в виде, показанном на 2.

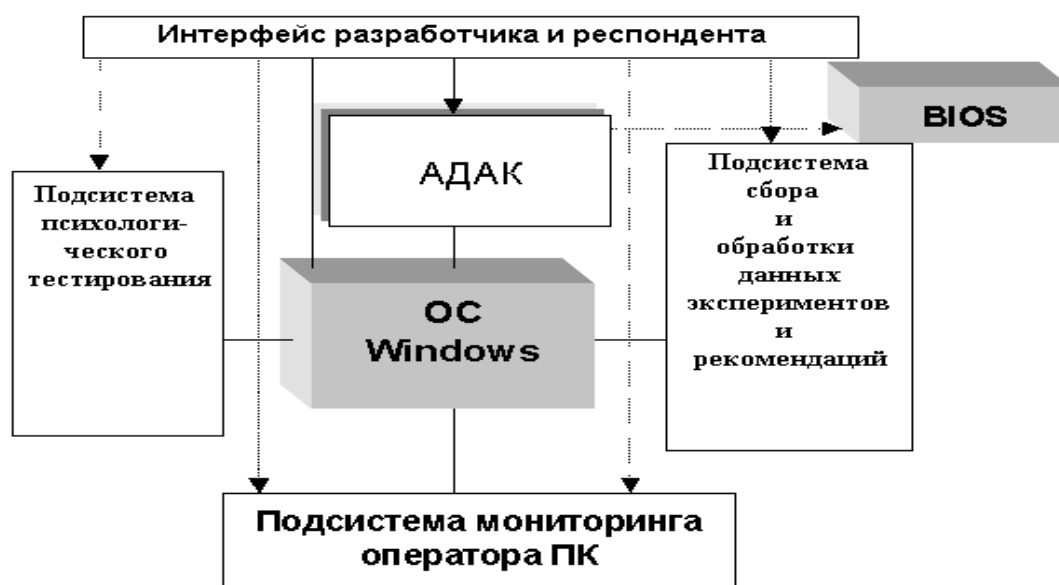


Рис. 2. Обобщенная структура СИДАКа

СИДАК скорее всего будет создаваться в наиболее распространенной среде Windows с целью снятия проблемы адаптации создаваемых АДАКов к различным платформам, по меньшей мере, на начальном этапе развития этого направления.

- [Предполагается создание двойственного интерфейса,] ориентированного как на разработчика алгоритмов, так и на испытуемых (респондентов), участвующих в апробации АДАКов, в имитационных экспериментах с ними.
- Подсистема психологического тестирования, с одной стороны, содержит развиваемую базу психологических тестов, блок архивирования данных тестирования, блок анализа этих данных и представления результатов этого анализа, блок передачи данных анализа в АДАК по санкции интерфейса. Возможно наличие в этой подсистеме интеллектуального компонента выбора теста по ситуации, определяемой подсистемой мониторинга.
- Подсистема сбора и обработки данных экспериментов и рекомендаций очевидна по своему назначению и содержанию. Возможна тесная взаимосвязь этой подсистемы с Excel или пакетом типа Statistics.
- Подсистема мониторинга оператора ПК [2] должна иметь, по меньшей мере, средства снятия частоты пульса, давления и температуры тела. Важ-

ным компонентом этой подсистемы является цифровая видеокамера.

Разумеется, приведенное описание СИДАКа – поверхностное и имеет целью, в первую очередь, анонсирование идеи. Возможно уточнение и появление новых компонентов при реализации этого предложения при использовании конкретного программного инструментария.

4. Заключение

Выше сделан набросок структуры и оговорены основные блоки алгоритма дозирования активности компьютера, основные функции и компоненты имитационной системы для разработки таких алгоритмов. На основе изложенного предстоит осуществить выбор инструмента реализации имитационной системы и осуществить его разработку и реализацию. Разумеется, в первой версии СИДАКа будет иметь место ряд пустых (нулевых) ответвлений (заглушек) для его усовершенствования в процессе проектирования, реализации, отладки и использования (эксплуатации).

Представляется, что развиваемый в статье подход ограничения свободы общения с компьютером посредством принудительного, адаптивного дозирования интервалов общения принесет ощутимые плоды как в отношении сохранения здоровья людей, учащихся в первую очередь, повышения эффективности интеллектуального труда, а также в борьбе с надвигающейся глобальной эпидемией компьютеро-интернетомании.

ЛИТЕРАТУРА

1. Шапцев В.А. *Человекоцентрический подход в информатизации* // Тез. докл. VI Междун. форум по информатизации МФИ-97, VI-й Конгресс «Общественное развитие и общественная информация». Москва, 21–24 нояб. 1997. С.67–70.
2. Shaptsev V., Laptev I. *Intellectual Adaptive Informational Filter. Constructional Idea* // Computer Science and Information Technologies: Proc.of the 2nd Intern. Workshop CSIT'2000. Ufa, Sept. 18–23. 2000. Ufa: USATU Publisher, 2000. P.240–246.
3. Мунипов В.М., Зинченко В.П. *Эргономика: человекоориентированное проектирование техники, программных средств и среды*. М.: Логос, 2001. 356 с.
4. Концевой М.П. *Здоровье и компьютер* // Информатика и образование. 2000. N.1. С.88–91.
5. Розенблюм Ю.З., Фрейгин А.А., Корнюшина Т.А. *Компьютерный зрительный синдром. Компьютер и зрение* // По материалам Международного конгресса «Человек в большом городе XXI века». Москва, июнь 1998 г. С.36–38.

ЗАВИСИМОСТЬ КРИТИЧЕСКОГО ПОВЕДЕНИЯ ОДНОРОДНЫХ СЖИМАЕМЫХ СИСТЕМ ОТ РАЗМЕРНОСТИ ПАРАМЕТРА ПОРЯДКА

С.В. Белим

The renormalization-group method is applied to analysis of phase transitions in systems where the order parameter is coupled to nonordering additional elastic variable. A variety of critical and tricritical behavior is found as function of the physical variables and possible macroscopic constraints imposed on the system. The tricritical exponents were calculated in two-loop order with using Padé-Borel summation technique.

В сжимаемых системах связь параметра порядка с упругими деформациями играет важную роль. Как впервые было показано в [1], для упруго-изотропного тела критическое поведение сжимаемых систем с квадратичной стрикцией неустойчиво относительно связи параметра порядка с акустическими модами и реализуется фазовый переход первого рода, близкий ко второму. Однако выводы работы [1] справедливы только в области низких давлений и, как показано в [2], в области высоких давлений, начиная с некоторого трикритического значения P_t , деформационные эффекты, индуцируемые внешним давлением, оказывают на систему более радикальное влияние, приводя к смене знака эффективной константы взаимодействия флуктуаций параметра порядка и, как следствие, рода фазового перехода. При этом в [2] для однородных сжимаемых систем предсказывается два типа трикритического поведения и существование критической точки четвертого порядка, в которой пересекаются две трикритические кривые. Расчеты, проведенные в рамках двухпетлевого приближения [3], подтвердили наличие двух типов трикритического поведения для изинговских систем и позволили получить значения трикритических индексов.

Согласно критерию, полученному в [1], стрикционные эффекты, рассматриваемые как дополнительные термодинамические параметры, приводят к смене режима критического поведения только в системах с сингулярным поведением теплоемкости в отсутствии деформаций. Индекс теплоемкости α ($C \sim |T - T_c|^{-\alpha}$) положителен лишь для изинговских магнетиков. Для ХУ-модели и модели Гейзенберга индекс теплоемкости жестких систем положителен, и, следовательно, упругие деформации не должны сказываться на критическом поведении. Отсюда вытекает, что критическое значение размерности параметра порядка $n_c < 2$.

© 2002 С.В. Белим

E-mail: belim@univer.omsk.su

Омский государственный университет

В настоящей работе осуществлено развитие модели фазовых превращений в однородных сжимаемых системах, характеризующихся различной размерностью флуктуирующего параметра порядка [4, 5], рассматриваемых методами ренорм-группы в двухпетлевом приближении непосредственно в трехмерном пространстве. Исследуются также условия реализации трикритического поведения за счет эффектов дальнего действия взаимодействия флуктуаций параметра порядка, обусловленного длинноволновыми акустическими модами. В связи с тем, что в критической области основной вклад в стрикционные эффекты дает зависимость обменного интеграла от расстояния, рассматриваются лишь упруго-изотропные системы.

Гамильтониан однородной изингоподобной модели с учетом упругих деформаций может быть записан в виде:

$$H_0 = \int d^D x \left[\frac{1}{2} (\tau_1 + \nabla^2) \vec{S}(x)^2 + \frac{u_0}{4!} (\vec{S}(x)^2)^2 \right] + \int d^D x \left[a_1 \left(\sum_{\alpha=1}^3 u_{\alpha\alpha}(x) \right)^2 + a_2 \sum_{\alpha,\beta=1}^3 u_{\alpha\beta}^2 \right] + \frac{1}{2} a_3 \int d^D x \vec{S}(x)^2 \left(\sum_{\alpha=1}^3 u_{\alpha\alpha}(x) \right), \quad (1)$$

где $\vec{S}(x)$ – n -мерный параметр порядка, u_0 – положительная константа, $\tau_0 \sim |T - T_c|/T_c$, T_c – температура фазового перехода, $u_{\alpha\beta}$ – тензор деформаций, a_1, a_2 – упругие постоянные кристалла, a_3 – параметр квадратичной стрикции. Переходя в (1) к фурье-образам переменных и интегрируя по слагаемым, зависящим от нефлуктуирующих переменных, не взаимодействующих с параметром порядка $S(x)$, и вводя для удобства новую переменную $y(x) = \sum_{\alpha=1}^3 u_{\alpha\alpha}(x)$, получим гамильтониан системы в следующем виде:

$$H_0 = \frac{1}{2} \int d^D q (\tau_0 + q^2) \vec{S}_q \vec{S}_{-q} + \frac{u_0}{4!} \int d^D q_i (\vec{S}_{q_1} \vec{S}_{q_2}) (\vec{S}_{q_3} \vec{S}_{-q_1-q_2-q_3}) + \quad (2)$$

$$+ a_3 \int d^D q y_{q_1} \vec{S}_{q_2} \vec{S}_{-q_1-q_2} + \frac{a_3^{(0)}}{\Omega} y_0 \int d^D q \vec{S}_q \vec{S}_{-q} + \frac{1}{2} a_1 \int d^D q y_q y_{-q} + \frac{1}{2} \frac{a_1^{(0)}}{\Omega} y_0^2.$$

В (2) выделены слагаемые y_0 , описывающие однородные деформации. Как показано в работе [1], такое разделение необходимо, так как неоднородные деформации y_q отвечают за обмен акустическими фононами и приводят к эффектам дальнего действия, которые отсутствуют при однородных деформациях.

Определим эффективный гамильтониан системы, зависящий только от сильно флуктуирующего параметра порядка S , следующим образом:

$$\exp\{-H[\vec{S}]\} = B \int \exp\{-H_R[\vec{S}, y]\} \prod dy_q. \quad (3)$$

Если эксперимент осуществляется при постоянном объеме, то y_0 является константой, интегрирование в (3) проводится только по неоднородным деформациям и однородные деформации вклада в эффективный гамильтониан не вносят.

При постоянном давлении в гамильтониан добавляется слагаемое $P\Omega$, объем представляется в терминах компонент тензора деформации в виде

$$\Omega = \Omega_0[1 + \sum_{\alpha=1} u_{\alpha\alpha} + \sum_{\alpha \neq \beta} u_{\alpha\alpha} u_{\beta\beta} + O(u^3)], \quad (4)$$

и интегрирование в (3) осуществляется также и по однородным деформациям. Как отмечено в [6], учет в (4) квадратичных слагаемых может оказаться важным в случае высоких давлений и кристаллов с большими стрикционными эффектами. Пренебрежение в [1] данными квадратичными слагаемыми ограничивает применение результатов работы Ларкина и Пикина только к случаю низких давлений. В результате:

$$\begin{aligned} H &= \frac{1}{2} \int d^D q (\tau_0 + q^2) \vec{S}_q \vec{S}_{-q} + \left(\frac{u_0}{4!} - \frac{z_0}{2} \right) \int d^D \{q_i\} \vec{S}_{q_1} \vec{S}_{q_2} \vec{S}_{q_3} \vec{S}_{-q_1-q_2-q_3} \\ &+ \frac{1}{2\Omega} (z_0 - w_0) \int d^D \{q_i\} \vec{S}_{q_1} \vec{S}_{-q_1} \vec{S}_{q_2} \vec{S}_{-q_2}, \\ z_0 &= a_1^2/(4a_3), \quad w_0 = a_1^{(0)2}/(4a_3^{(0)}). \end{aligned} \quad (5)$$

Возникающий в гамильтониане эффективный параметр взаимодействия $v_0 = u_0 - 12z_0$ за счет влияния стрикционных эффектов, определяемых параметром z_0 , может принимать не только положительные, но и отрицательные значения. В результате данный гамильтониан описывает как фазовые переходы первого, так и второго рода. При $v_0 = 0$ в системе реализуется трикритическое поведение. В свою очередь, эффективное взаимодействие в (5), определяемое разностью параметров $z_0 - w_0$, при $z_0 - w_0 > 0$ может вызывать в системе фазовый переход второго рода, а при $z_0 - w_0 < 0$ – фазовый переход первого рода. Из данного вида эффективного гамильтониана следует возможность осуществления критической точки более высокого порядка, в которой пересекаются трикритические кривые, при одновременном выполнении условий $v_0 = 0$, $z_0 = w_0$ [2]. Следует отметить, что при трикритическом условии $z_0 = w_0$ гамильтониан модели (5) изоморфен гамильтониану однородной жесткой системы.

В рамках теоретико-полевого подхода [7] асимптотическое критическое поведение и структура фазовых диаграмм во флуктуационной области определяется ренорм-групповым уравнением Каллана-Симанчика для вершинных частей неприводимых функций Грина. Для вычисления β - и γ -функций как функций, входящих в уравнение Каллана-Симанчика перенормированных вершин взаимодействия $u, a_1, a_1^{(0)}$ или более удобных для определения критического и трикритического поведения модели комплексных вершин $z = a_1^2/(4a_3)$, $w = a_1^{(0)2}/(4a_3^{(0)})$, $v = u - 12z$, был применен стандартный метод, основанный на диаграммной технике Фейнмана и процедуре перенормировки [8]. В результате в рамках двухпетлевого приближения были получены следующие выражения для

β -функций:

$$\begin{aligned}\beta_v &= -v \left(1 - \frac{n+8}{6}v + \frac{41n+190}{243}v^2 \right), \\ \beta_z &= -z \left(1 - \frac{n+2}{3}v - 2nz + \frac{23(n+2)}{243}v^2 \right), \\ \beta_w &= -w \left(1 - \frac{n+2}{3}v - 4nz + 2nw + \frac{23(n+2)}{243}v^2 \right).\end{aligned}\quad (6)$$

Известно, что ряды теории возмущений являются асимптотическими, а вершины взаимодействия флуктуаций параметров порядка во флуктуационной области достаточно велики, чтобы можно было непосредственно применять выражения (6). Поэтому с целью извлечения из полученных выражений нужной физической информации был применен обобщенный на трехпараметрический случай метод Паде-Бореля. При этом прямое и обратное преобразования Бореля имеют вид

$$\begin{aligned}f(v, z, w) &= \sum_{i_1, i_2, i_3} c_{i_1, i_2, i_3} v^{i_1} z^{i_2} w^{i_3} = \int_0^\infty e^{-t} F(vt, zt, wt) dt, \\ F(v, z, w) &= \sum_{i_1, i_2, i_3} \frac{c_{i_1, i_2, i_3}}{(i_1 + i_2 + i_3)!} v^{i_1} z^{i_2} w^{i_3}.\end{aligned}\quad (7)$$

Для аналитического продолжения борелевского образа функции вводится ряд по вспомогательной переменной θ

$$\tilde{F}(v, z, w, \theta) = \sum_{k=0}^{\infty} \theta^k \sum_{i_1, i_2, i_3} \frac{c_{i_1, i_2, i_3}}{k!} v^{i_1} z^{i_2} w^{i_3} \delta_{i_1+i_2+i_3, k}, \quad (8)$$

к которому применяется аппроксимация Паде [L/M] в точке $\theta = 1$. Данная методика была предложена и апробирована в работах [9] для описания критического поведения ряда систем, характеризующихся несколькими вершинами взаимодействия флуктуаций параметра порядка. Выявленное в [9] свойство сохранения симметрии системы в процессе применения Паде-аппроксимант по переменной θ становится существенным при описании многовершинных моделей.

В двухпетлевом приближении для вычисления β -функций был использован аппроксимант [2/1]. Природа критического поведения определяется существованием устойчивой фиксированной точки, удовлетворяющей системе уравнений

$$\beta_i(v^*, z^*, w^*) = 0 \quad (i = 1, 2, 3). \quad (9)$$

Требование устойчивости фиксированной точки сводится к условию, чтобы собственные значения b_i матрицы

$$B_{i,j} = \frac{\partial \beta_i(u_1^*, u_2^*, u_3^*)}{\partial u_j} \quad (u_i, u_j \equiv v, z, w) \quad (10)$$

Таблица 1. Значения фиксированных точек и собственных значений матрицы устойчивости

N	v^*	z^*	w^*	b_1	b_2	b_3
n=1						
I0	0	0	0	-1	-1	-1
I1	1,064472	0	0	0,6536	-0,1692	-0,1692
I2	1,064472	0,089187	0	0,6536	0,1702	0,1710
I3	1,064472	0,089187	0,089187	0,6536	0,1702	-0,1710
I4	0	0,5	0	-1	1	-0,16923
I5	0	0,5	0,5	-1	1	-1
n=2						
X0	0	0	0	-1	-1	-1
X1	0,934982	0	0	0,6673	-0,0017	0,1053
X2	0,934982	0,000439	0	0,6673	0,0017	0,1087
X3	0,934982	0,000439	0,000439	0,6673	0,0017	-0,1053
X4	0	0,25	0	-1	1	1
X5	0	0,25	0,25	-1	1	-1
n=3						
G0	0	0	0	-1	-1	-1
G1	0,829620	0	0	0,6813	0,1315	0,2173
G2	0,829620	0,022909	0	0,6813	-0,1311	-0,0518
G3	0,829620	0,022909	0,022909	0,6813	-0,1311	-0,2170
G4	0	1/6	0	-1	1	1
G5	0	1/6	1/6	-1	1	-1

лежали в правой комплексной полуплоскости. Фиксированная точка с $v^* = 0$, соответствующая трикритическому поведению, является седловой точкой и должна быть устойчивой в направлениях, задаваемых переменными z, w , и неустойчивой в направлении, определяемом переменной v . Стабилизация трикритической фиксированной точки в направлении, задаваемом переменной v , осуществляется в результате учета в эффективном гамильтониане модели членов шестого порядка по флуктуациям параметра порядка. Фиксированная точка с $z^* = w^*$, соответствующая трикритическому поведению второго типа, является также седловой точкой и должна быть устойчивой в направлениях, задаваемых переменными v, z , и неустойчивой в направлении, определяемом переменной w . Ее стабилизация может осуществляться за счет влияния ангармонических эффектов. Полученная система просуммированных β -функций содержит широкое разнообразие фиксированных точек. В таблице 1 приведены наиболее интересные для описания критического и трикритического поведения фиксированные точки для модели Изинга ($n = 1$), XY-модели ($n = 2$) и модели Гейзенберга ($n = 3$), лежащие в физической области значений вершин с $v, z, w \geq 0$. В таблице приведены также собственные значения матрицы устойчивости для соответствующих фиксированных точек.

Анализ значений фиксированных точек и их устойчивости позволяет сделать ряд выводов. Гауссовы фиксированные точки I_0 , X_0 , G_0 являются трикритическими и неустойчивы относительно влияния упругих деформаций. Критическое поведение несжимаемых систем относительно деформационных степеней свободы неустойчиво для модели Изинга (I_1) и устойчиво для модели Гейзенберга (G_1). Для ХУ-модели (X_1) собственное значение $b_2 < 0$, но по порядку величины сравнимо с точностью вычислений, вследствие чего нельзя сделать однозначного вывода об устойчивости данной фиксированной точки. Повидимому, сложности в описании ХУ-модели связаны с близостью критической размерности параметра порядка n_c к двум. Согласно критерию, полученному в работе [1], $n_c < 2$, тогда как двухпетлевое приближение дает $n_c = 2,011$. Для изинговских систем оказывается устойчивой фиксированная точка при постоянной деформации (I_2), для гейзенберговских систем соответствующая точка неустойчива (G_2), для ХУ-модели нельзя дать однозначный ответ в силу все той же близости критической размерности к двум. Фиксированные точки I_3 , X_3 , G_3 описывают первый тип трикритического поведения сжимаемых систем, наблюдаемый при постоянном давлении. Фиксированные точки I_4 , X_4 , G_4 являются трикритическими для систем, исследуемых при постоянном объеме. Точки I_5 , X_5 , G_5 являются критическими точками четвертого порядка, в них пересекаются две трикритические линии.

Полученные в двухпетлевом приближении значения вершин в фиксированных точках, соответствующих критическому и трикритическому поведению сжимаемой модели Изинга, позволяют вычислить критические индексы для данных систем на основе просуммированных методом Паде-Бореля выражений для индексов ν и η :

$$\begin{aligned}\nu &= \frac{1}{2} \left(1 + \frac{n+2}{12} v^* + n z^* - n w^* - \frac{11(n+2)}{3888} v^{*2} \right), \\ \eta &= \frac{2(n+2)}{243} v^{*2}.\end{aligned}\quad (11)$$

Значения остальных критических индексов могут быть получены из скейлинговых соотношений, связывающих их с индексами ν , η .

Критическое поведение сжимаемых изинговских систем при постоянном давлении (I_2) характеризуется перенормированными индексами согласно теории Фишера о влиянии дополнительных термодинамических переменных [10]:

$$\nu^{(I)} = 0,632; \eta^{(I)} = 0,028; \alpha^{(I)} = 0,103; \beta^{(I)} = 0,325; \gamma^{(I)} = 1,247.$$

Для трикритического поведения первого типа (I_3 , X_3 , G_3) гамильтониан (5) изоморфен гамильтониану однородной несжимаемой модели, поэтому и критические индексы совпадают с индексами несжимаемой модели:

$$\begin{aligned}\nu^{(I)} &= 0,708; \eta^{(I)} = 0,028; \alpha^{(I)} = -0,125; \beta^{(I)} = 0,364; \gamma^{(I)} = 1,397; \\ \nu^{(XY)} &= 1; \eta^{(XY)} = 0; \alpha^{(XY)} = -1; \beta^{(XY)} = 0,5; \gamma^{(XY)} = 2; \\ \nu^{(G)} &= 1; \eta^{(G)} = 0; \alpha^{(G)} = -1; \beta^{(G)} = 0,5; \gamma^{(G)} = 2.\end{aligned}$$

Трикритическое поведение второго типа (I4, X4, G4) соответствует критическому поведению сферической модели и определяется соответствующими индексами:

$$\begin{aligned}\nu^{(I)} &= 1; \eta^{(I)} = 0; \alpha^{(I)} = -1; \beta^{(I)} = 0,5; \gamma^{(I)} = 2; \\ \nu^{(XY)} &= 1; \eta^{(XY)} = 0; \alpha^{(XY)} = -1; \beta^{(XY)} = 0,5; \gamma^{(XY)} = 2; \\ \nu^{(G)} &= 1; \eta^{(G)} = 0; \alpha^{(G)} = -1; \beta^{(G)} = 0,5; \gamma^{(G)} = 2.\end{aligned}$$

Фиксированные точки четвертого порядка (I4, X4, G4) характеризуются среднеполевыми значениями критических индексов:

$$\begin{aligned}\nu^{(I)} &= 0,5; \eta^{(I)} = 0; \alpha^{(I)} = 0,5; \beta^{(I)} = 0,25; \gamma^{(I)} = 1; \\ \nu^{(XY)} &= 0,5; \eta^{(XY)} = 0; \alpha^{(XY)} = 0,5; \beta^{(XY)} = 0,25; \gamma^{(XY)} = 1; \\ \nu^{(G)} &= 0,5; \eta^{(G)} = 0; \alpha^{(G)} = 0,5; \beta^{(G)} = 0,25; \gamma^{(G)} = 1.\end{aligned}$$

Проведенные исследования показали существенность влияния упругих деформаций на критическое поведение сжимаемых систем, проявляющееся как в изменении значений критических индексов для изинговских систем, так и появлении мультикритических точек на фазовых диаграммах всех трех моделей. Мы надеемся, что выявленные эффекты и определенные значения индексов найдут подтверждение в экспериментальных исследованиях.

ЛИТЕРАТУРА

1. Ларкин А.И., Пикин С.А. // ЖЭТФ. 1969. Т.56. С.1664.
2. Imry Y. // Phys. Rev. Lett. 1974. V.33. P.1304.
3. Белим С.В., Прудников В.В. // ФТТ. 2001. Т.45. С.1299.
4. Laptev V.M., Skryabin Yu.N. // Phys. Stat. Sol. 1979. **B91**, K143.
5. Skryabin Y.N., Shchanov A.V. // Phys. Lett. 1997. **A234**, 1. P.147.
6. Bergman D.J., Halperin B.I. // Phys. Rev. 1976. **B13**, 4. P.2145.
7. Amit D. *Field theory the renormalization group and critical phenomena*. New York: McGraw-Hill, 1976.
8. Zinn-Justin J. *Quantum field theory and critical phenomena*. Oxford: Clarendon Press, 1989.
9. Sokolov A.I., Varnashev K.B. // Phys.Rev. 1999. **B59**, 13. P.8363.
10. Fisher M.E. // Phys.Rev. 1976. **176**, 1. P.257.

ВКЛАД СПИНОРНЫХ ДУХОВ В ИНТЕРФЕРЕНЦИЮ КВАНТОВЫХ ЧАСТИЦ

Е.В. Палешева

In this article a question of the ghost spinors's influence to the quantum particles's interference is investigated. The interaction between spinors and ghost spinors are considered. Furthermore the conditions of zero stress-energy tensor in private cases are found. Also we consider a question of experimental test of existance of Deutsch's shadow pacticles.

Введение

Как известно, одним из подтверждений квантовой механики являются опыты, связанные с интерференцией элементарных частиц. В основе объяснения получающихся экспериментальных данных лежит представление о волновой природе квантовых частиц, волновая функция которых сама по себе не несет никакого физического смысла, а квадрат модуля амплитуды вероятности интерпретируется как распределение квантового потока. При этом само явление интерференции нескольких волн, являющееся одним из фундаментальных понятий в современной квантовой теории поля, объясняется наличием у потока частиц заданной интенсивности распределения в пространстве. Но на вопрос: почему в опыте с двумя щелями электрон не может попасть в определенные области экрана – сложно получить удовлетворительный ответ. Что мешает электрону удариться о край щели так, чтобы попасть в недоступную область экрана? В лучшем случае можно было услышать о вероятности попадания частицы в заданные точки пространства. А почему электрон имеет именно такое распределение вероятности? На этот вопрос ответил Дэвид Дойч [2]. Теневой электрон – вот что отталкивает реальный электрон и мешает ему попасть в любую точку экрана. В [6, 7] вводится предположение, что теневой электрон и электронный дух представляют одну и ту же частицу. Именно то, что спинорные духи обладают нулевой энергией и ненулевой дираковской плотностью тока, делает данное предположение достаточно привлекательным. В представленной работе описываются результаты, полученные в ходе дальнейших исследований поставленной задачи. Во-первых, необходимо показать, что известные опыты с интерференцией не опровергают возможность влияния духов квантовых частиц на распределение волнового потока реальных частиц. Во-вторых, следует выяснить характер взаимодействия между спинорами и их духами. Именно эти вопросы частично раскрываются в последующих разделах.

© 2002 Е.В. Палешева

E-mail: m82palesheva@math.omsu.omskreg.ru

Омский государственный университет

В [7] были приведены некоторые рассуждения, в основе которых лежало представление о реальности как о мультиверсе [2], при этом несложно увидеть неточность, основанную на том, что импульс пропорционален току. В действительности дираковский ток не пропорционален импульсу, такой вывод можно сделать, опираясь на известное разложение Гордона [5, с.45].

Недавно были получены новые результаты, позволяющие более корректно говорить о мультиверсе Дойча, в работе [3, 4] А.К.Гуц представил математическую модель такого мультиверса. Формализованный подход к представлению реальности как совокупности параллельных вселенных, в основе которого лежит синтетическая дифференциальная геометрия Кока-Ловера, позволяет использовать понятие мультиверса, подразумевая под этим конкретную математическую теорию.

1. Условия зануления тензора энергии-импульса

Уравнение Дирака для свободной частицы в пространстве-времени Минковского, как известно, имеет вид

$$i\hbar\gamma^{(k)}\frac{\partial\psi}{\partial x^k} - mc\psi = 0, \quad (1)$$

где $\gamma^{(k)}$ – матрицы Дирака в стандартном представлении:

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

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

В этом случае тензор энергии-импульса спинорной материи определяется выражением:

$$T_{ik} = \frac{i\hbar c}{4} \left\{ \psi^* \gamma^{(0)} \gamma_i \frac{\partial\psi}{\partial x^k} - \frac{\partial\psi^*}{\partial x^k} \gamma^{(0)} \gamma_i \psi + \psi^* \gamma^{(0)} \gamma_k \frac{\partial\psi}{\partial x^i} - \frac{\partial\psi^*}{\partial x^i} \gamma^{(0)} \gamma_k \psi \right\}, \quad (2)$$

где символ $*$ означает эрмитово сопряжение. При этом

$$\gamma_i = g_{ik} \gamma^{(k)}.$$

В [6, 7] были найдены спинорные духи в плоском пространстве-времени, метрика которого отлична от метрики специальной теории относительности. Спинорные духи – это частицы с нулевым тензором энергии-импульса и ненулевой плотностью тока, вычисляемой в пространстве Минковского по формуле

$$j^{(k)} = \psi^* \gamma^{(0)} \gamma^{(k)} \psi. \quad (3)$$

При этом $j^{(0)}$ определяет квадрат модуля амплитуды вероятности волны ψ , характеризующий вероятность появления данной частицы в пространстве, причем в рамках специальной теории относительности $j^{(0)} = \psi^* \psi$, а $j^{(\alpha)}$ определяет скорость изменения потока плотности вероятности ¹.

¹Греческие индексы пробегают значения 1,2,3.

Приведем следующие результаты относительно спинорных духов. При этом в дальнейшем будем работать с метрикой специальной теории относительности.

Теорема 1. Пусть $\psi = u \cdot G(x)$ – решение уравнения Дирака, при этом полагаем $\psi^* \psi \neq 0$,

$$G(x) = f(x) + i \cdot g(x),$$

где $f(x)$ и $g(x)$ – гладкие вещественные функции, а

$$u = \begin{bmatrix} u_0 \\ u_1 \\ u_2 \\ u_3 \end{bmatrix},$$

где $\forall i \quad u_i \in \mathbb{C}$. В рассмотренных условиях ψ является спинорным духом $\iff g(x) = a \cdot f(x)$, где $a = \text{const} \in \mathbb{R}$.

Доказательство. По определению решение уравнения Дирака называется спинорным духом $\iff T_{ik} \equiv 0$ и дираковский ток $j^{(k)} \neq 0$. Из условия $\psi^* \psi \neq 0$ следует выполнение того, что $j^{(k)} \neq 0$.

Теперь осталось показать, что

$$T_{ik} \equiv 0 \iff g(x) = a \cdot f(x).$$

Для этого заметим, что ²

$$\psi^* = u^* \bar{G} = u^* (f(x) - i \cdot g(x)).$$

Далее разобьем доказательство теоремы на несколько этапов.

а) $T_{00} \equiv 0 \iff$

$$\begin{aligned} 0 &= \psi^* \gamma^{(0)} \gamma^{(0)} \frac{\partial \psi}{\partial x^0} - \frac{\partial \psi^*}{\partial x^0} \gamma^{(0)} \gamma^{(0)} \psi = \psi^* \frac{\partial \psi}{\partial x^0} - \frac{\partial \psi^*}{\partial x^0} \psi = u^* \bar{G} u \frac{\partial G}{\partial x^0} - \\ &- u^* \frac{\partial \bar{G}}{\partial x^0} u G = u^* u \left(\bar{G} \frac{\partial G}{\partial x^0} - \frac{\partial \bar{G}}{\partial x^0} G \right). \end{aligned}$$

Из того, что $\psi^* \psi \neq 0$, следует выполнение неравенства $u^* u \neq 0$. Поэтому

$$T_{00} \equiv 0 \iff \bar{G} \frac{\partial G}{\partial x^0} - \frac{\partial \bar{G}}{\partial x^0} G = 0,$$

а это в свою очередь верно $\iff$

$$f \frac{\partial g}{\partial x^0} = g \frac{\partial f}{\partial x^0}.$$

б) $T_{01} \equiv 0 \iff$

$$0 = \psi^* \gamma^{(0)} \gamma^{(0)} \frac{\partial \psi}{\partial x^1} - \frac{\partial \psi^*}{\partial x^1} \gamma^{(0)} \gamma^{(0)} \psi - \psi^* \gamma^{(0)} \gamma^{(1)} \frac{\partial \psi}{\partial x^1} + \frac{\partial \psi^*}{\partial x^1} \gamma^{(0)} \gamma^{(1)} \psi =$$

²Здесь $\bar{G}$ – комплексно-сопряженная функция.

$$\begin{aligned}
&= \psi^* \frac{\partial \psi}{\partial x^1} - \frac{\partial \psi^*}{\partial x^1} \psi - u^* \bar{G} \gamma^{(0)} \gamma^{(1)} u \frac{\partial G}{\partial x^0} + u^* \frac{\partial \bar{G}}{\partial x^0} \gamma^{(0)} \gamma^{(1)} u G = \\
&= u^* u \left(\bar{G} \frac{\partial G}{\partial x^1} - \frac{\partial \bar{G}}{\partial x^1} G \right) - u^* \gamma^{(0)} \gamma^{(1)} u \left(\bar{G} \frac{\partial G}{\partial x^0} - \frac{\partial \bar{G}}{\partial x^0} G \right).
\end{aligned}$$

Используя результаты пункта а), получаем зануление второго слагаемого. Поэтому зануление компоненты T_{01} происходит только лишь в случае выполнения равенства

$$f \frac{\partial g}{\partial x^1} = g \frac{\partial f}{\partial x^1}.$$

в) Последовательно выполняя аналогичные процедуры для компонент T_{02} и T_{03} , находим, что $T_{0k} \equiv 0 \iff$ выполнена следующая система линейных дифференциальных уравнений первого порядка в частных производных:

$$f \frac{\partial g}{\partial x^k} = g \frac{\partial f}{\partial x^k}.$$

Полученная система имеет решение

$$g(x) = a f(x), \quad (4)$$

где $a = \text{const} \in \mathbb{R}$.

Теперь покажем, что из выполнения условия (4) следует, что

$$\forall \alpha, \beta \quad T_{\alpha\beta} \equiv 0.$$

Имеем

$$\begin{aligned}
T_{\alpha\beta} \equiv 0 \iff 0 &= \psi^* \gamma^{(0)} \gamma^{(\alpha)} \frac{\partial \psi}{\partial x^\beta} - \frac{\partial \psi^*}{\partial x^\beta} \gamma^{(0)} \gamma^{(\alpha)} \psi + \psi^* \gamma^{(0)} \gamma^{(\beta)} \frac{\partial \psi}{\partial x^\alpha} - \\
&- \frac{\partial \psi^*}{\partial x^\alpha} \gamma^{(0)} \gamma^{(\beta)} \psi = u^* \gamma^{(0)} \gamma^{(\alpha)} u \left(\bar{G} \frac{\partial G}{\partial x^\beta} - \frac{\partial \bar{G}}{\partial x^\beta} G \right) + u^* \gamma^{(0)} \gamma^{(\beta)} u \left(\bar{G} \frac{\partial G}{\partial x^\alpha} - \frac{\partial \bar{G}}{\partial x^\alpha} G \right).
\end{aligned}$$

Из условия, наложенного на функцию $G(x)$, получаем выполнение равенства. Теорема 1 доказана. ■

Следствие 1. Если в условиях теоремы положить

$$G(x) = e^{\alpha(x) + i\beta(x)},$$

где $\alpha(x)$ и $\beta(x)$ – гладкие вещественные функции, то ψ является спинорным духом $\iff \beta(x) = \text{const} \in \mathbb{R}$.

Доказательство.

$$G(x) = e^{\alpha(x) + i\beta(x)} = e^{\alpha(x)} \cos[\beta(x)] + i e^{\alpha(x)} \sin[\beta(x)].$$

По теореме 1 ψ является спинорным духом $\iff \text{ctg}[\beta(x)] = \text{const}$. ■

Теорема 2. Пусть

$$\psi = \begin{bmatrix} G_0(x) \\ G_1(x) \\ G_2(x) \\ G_3(x) \end{bmatrix}$$

– решение уравнения Дирака, при этом $\psi^* \psi \neq 0$ и

$$\forall \quad k \quad G_k(x) = f_k(x) + i g_k(x),$$

где $f_k(x)$ и $g_k(x)$ – гладкие вещественные функции. Если при этом

$$\forall \quad i, k \quad f_i(x) = c_{ik} \cdot g_k(x),$$

здесь $c_{ik} = \text{const} \in \mathbb{R}$, то ψ является спинорным духом.

Доказательство. Простой подстановкой биспинора проверяется следующее равенство:

$$\psi^* \gamma^{(0)} \gamma_i \frac{\partial \psi}{\partial x^k} - \frac{\partial \psi^*}{\partial x^k} \gamma^{(0)} \gamma_i \psi = 0.$$

Поэтому $T_{ik} \equiv 0$. Вследствие того, что $\psi^* \psi \neq 0$, дираковский ток отличен от нуля. Теорема доказана. ■

Следствие 2. Если решение уравнения Дирака представляется в виде

$$\psi = u f(x), \tag{5}$$

где $f(x)$ – гладкая вещественная функция, а компоненты биспинора u – комплексные числа, удовлетворяющие условию $\psi^* \psi \neq 0$, то ψ является спинорным духом.

Доказательство. Очевидно. ■

2. Интерференция и взаимодействие реальной волны и спинорного духа

Теневые электроны Дойча, взаимодействуя с реальными электронами в опытах по интерференции, приводят к наблюдаемой экспериментаторами свето-теневой картине. И сразу же возникает вопрос о характере взаимодействия между этими частицами. Дэвид Дойч утверждает, что реальный фотон взаимодействует только с собственным теневым фотоном: «...каждая дробноатомная частица имеет двойников в других вселенных, и только эти двойники им мешают. Любые другие частицы этих вселенных не оказывают на нее непосредственного воздействия» [2]. Здесь приводится результат, говорящий о том, что при столкновении волны с некоторым спинорным духом возникает интерференционная картина. Кроме этого в ряде случаев столкновения соответствующих частиц результирующая волна оказывается спинорным духом.

2.1. Флуктуации материи во вселенной

Пусть волновая функция Дирака имеет вид:

$$\psi = u(f(x) + ig(x)),$$

$f(x)$ и $g(x)$ – гладкие вещественные функции, удовлетворяющие условию

$$f(x) \neq \text{const} \cdot g(x), \quad [g(x) + f(x)]^2 \neq 0.$$

При этом биспинор

$$u = \begin{bmatrix} u_0 \\ u_1 \\ u_2 \\ u_3 \end{bmatrix}$$

такой, что $\forall i \quad u_i \in \mathbb{R}$ и $u^*u \neq 0$. По теореме 1 такая спинорная волна не является духом.

Введем новую частицу, находящуюся в состоянии $\theta = u[g(x) - f(x)]$. Тогда, используя результаты теоремы 1, несложно убедиться, что частица $\theta(x)$ находится в состоянии духа. Так как функции $g(x)$ и $f(x)$ достаточно произвольны, то можно предположить, что в пространстве существует точка, в которой частицы сталкиваются. Тогда результирующая волна представляется биспинором $\psi + \theta = u(1 + i)g(x)$. Снова применяя теорему 1, можно увидеть, что $\psi + \theta$ является спинорным духом.

Таким образом, мы показали наличие взаимодействия у двух частиц, одна из которых реальная волна, а другая – теневая, т.е. находящаяся в состоянии духа. В общем случае из этого результата можно заключить, что реальный электрон может перейти в совершенно иное состояние без столкновения с другой реальной частицей. Если при этом вспомнить, что теневой электрон определяется Дойчем как электрон в параллельной вселенной, то получаем возможность «исчезновения» частицы в одном мире и «появления» в другом. При этом описанная флуктуация материи не вызвана взаимодействием только лишь частиц одной вселенной. В итоге мы получаем, что флуктуации пространства-времени могут быть вызваны именно взаимодействием различных вселенных единого мультиверса, модель которого предложена в [3,4]. Реальность «кипит» – внутри одной вселенной происходит рождение частиц из «ничего» и «уход в никуда».

2.2. Интерференция между спинорным духом и реальной частицей

Будем предполагать, что решение уравнения Дирака (1) имеет вид

$$\psi = \begin{bmatrix} u_0 \\ u_1 \\ u_2 \\ u_3 \end{bmatrix}, \quad (6)$$

тогда, подставляя (6) в (1), получаем следующую систему линейных дифференциальных уравнений в частных производных первого порядка:

$$\begin{cases} u_{0,0} + u_{3,1} - iu_{3,2} + u_{2,3} = -i\frac{mc}{\hbar}u_0 \\ u_{1,0} + u_{2,1} + iu_{2,2} - u_{3,3} = -i\frac{mc}{\hbar}u_1 \\ -u_{2,0} - u_{1,1} + iu_{1,2} - u_{0,3} = -i\frac{mc}{\hbar}u_2 \\ -u_{3,0} - u_{0,1} - iu_{0,2} + u_{1,3} = -i\frac{mc}{\hbar}u_3 \end{cases},$$

здесь $u_{i,k}$ означает дифференцирование по k -й координате. Будем искать решение, удовлетворяющее условию $u_0 = u_1 = -u_2 = u_3 = u$. Кроме этого, предполагая, что

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

получаем следующее ограничение на биспинор ψ :

$$\frac{\partial u}{\partial x^0} = \frac{\partial u}{\partial x^3}, \quad \frac{\partial u}{\partial x^1} = 0.$$

Тогда биспинор

$$\psi = \begin{bmatrix} 1 \\ 1 \\ -1 \\ 1 \end{bmatrix} e^{\frac{mc}{\hbar}x^2 + f(x^0+x^3) + ig(x^0+x^3)} \quad (7)$$

является решением уравнения Дирака в отсутствии гравитационного поля. Здесь $g(x^0 + x^3)$ и $f(x^0 + x^3)$ – гладкие вещественные функции.

Опираясь на результаты теоремы 1, получаем, что (7) описывает спинорного духа только лишь в том случае, когда $g(x^0 + x^3) = \text{const} \in \mathbb{R}$.

Возьмем решение для реальной волны в виде:

$$\psi = \begin{bmatrix} 1 \\ 1 \\ -1 \\ 1 \end{bmatrix} e^{\frac{mc}{\hbar}x^2 + i(x^0+x^3)}, \quad (8)$$

а для спинорного духа положим

$$\theta = \begin{bmatrix} 1 \\ 1 \\ -1 \\ 1 \end{bmatrix} e^{\frac{mc}{\hbar}x^2}. \quad (9)$$

Как (8) так и (9) имеют следующий 4-вектор дираковского тока:

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

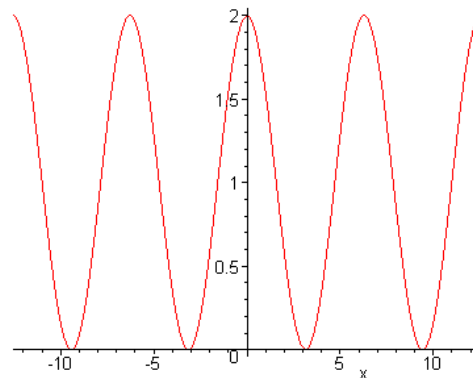


Рис. 1. Интенсивность распределения квадрата модуля амплитуды вероятности при интерференции спинорного духа и реальной частицы в точке $8e^{2\frac{mc}{\hbar}x^2} = 1$, $x^0 = 0$

Так как оба решения имеют одинаковое распределение вероятности и одинаковое направление тока, то мы можем посчитать их результирующую волну при столкновении этих частиц. Найдем квадрат модуля амплитуды вероятности результирующей волны $\psi + \theta$. Имеем

$$|\psi + \theta|^2 = (\psi + \theta)^*(\psi + \theta) = 8e^{2\frac{mc}{\hbar}x^2}(1 + \cos(x^0 + x^3)). \quad (10)$$

При фиксированных x^0 и x^2 получаем следующую интерференционную картину (см. рис.1). Таким образом, мы наблюдаем эффект интерференции теневой и реальной квантовой частицы. Еще один вариант взаимодействия параллельных вселенных.

2.3. Интерференция на двух щелях и спинорные духи

Только что было показано, что спинорные духи и истинные спинорные волны могут по-разному взаимодействовать друг с другом. Причем проявления таких взаимодействий реально должны быть наблюдаемы. И все-таки хотелось бы подробнее исследовать вопрос о влиянии спинорных духов на интерференционную картину в эксперименте с двумя щелями (см. рис.2).

В начале эксперимента частица находится в точке s , на экране B в точке x установлен детектор, фиксирующий попадание электрона в данную область экрана. На экране A расположены две щели a_1 и a_2 , симметричные относительно оси S , вдоль которой происходит распространение волны. Нас интересует распределение частиц на экране B . Для того, чтобы исключить влияние на результат эксперимента столкновений между частицами, будем испускать электроны по-одному и с достаточно большим интервалом между двумя излучениями. Как известно, в таких случаях интерференционная картина такая же, как и при излучении потока электронов.

Пусть ψ_1 и ψ_2 – амплитуды вероятности соответственно для электрона, прошедшего через щель a_1 и a_2 . В этом случае распределение результирующей волны определяется выражением $|\psi_1 + \psi_2|^2$. Будем предполагать, что мы наблюдаем интерференцию, т.е. расстояние между щелями a_1 и a_2 и экранами A

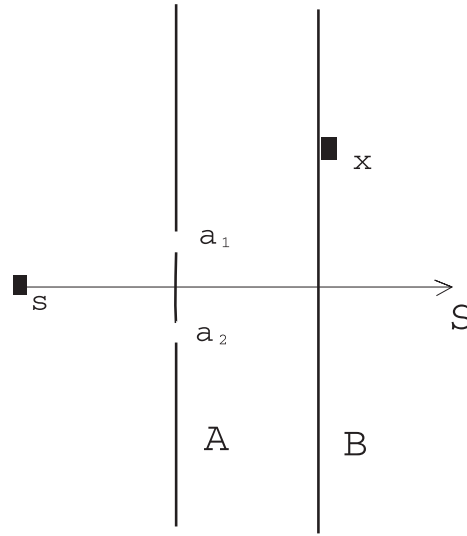


Рис. 2. Опыт по интерференции на двух щелях

и B , а также частота излучения таковы, что на экране B наблюдается чередование максимумов и минимумов.

Пусть теперь θ_1 – теневой электрон, проходящий через щель a_1 , а θ_2 – теневой электрон, проходящий через щель a_2 . Мы можем использовать для определения теневых электронов решения уравнения Дирака в силу того, что было сделано предположение тождественности теневых спинорных полей и соответствующих спинорных духов, являющихся решениями уравнения Дирака с отличной от нуля вероятностью появления в пространстве-времени [6, 7]. Будем использовать известные обозначения квантовой механики. Пусть в начальном состоянии электрон находится в точке y , тогда положим, что электрон в состоянии $|y\rangle$. Соответственно, если в конечном состоянии электрон находится в точке y , то положим, что электрон в состоянии $\langle y|$. При этом также будем полагать, что символ $\langle | \rangle$ означает результирующее состояние системы в ходе эксперимента, а символы $\langle | \rangle_\psi$ и $\langle | \rangle_\theta$ – соответственно состояния для реального спинорного поля ψ и спинорного духа θ . В рассмотренных обозначениях

$$\psi_1 = \langle x|a_1\rangle_{\psi_1} \langle a_1|s\rangle_{\psi_1}, \quad (11)$$

$$\psi_2 = \langle x|a_2\rangle_{\psi_2} \langle a_2|s\rangle_{\psi_2}, \quad (12)$$

$$\theta_1 = \langle x|a_1\rangle_{\theta_1} \langle a_1|s\rangle_{\theta_1}, \quad (13)$$

$$\theta_2 = \langle x|a_2\rangle_{\theta_2} \langle a_2|s\rangle_{\theta_2}. \quad (14)$$

Тогда состояние

$$\langle x|s\rangle_1 = \langle x|a_1\rangle_{\psi_1} \langle a_1|s\rangle_{\psi_1} + \langle x|a_2\rangle_{\psi_2} \langle a_2|s\rangle_{\psi_2} \quad (15)$$

определяет распределение вероятности $|\psi_1 + \psi_2|^2$ того, что реальная частица попадет из точки s в точку x , пройдя либо через щель a_1 , либо через щель a_2 . При этом неизвестно через какую именно щель прошла частица и также

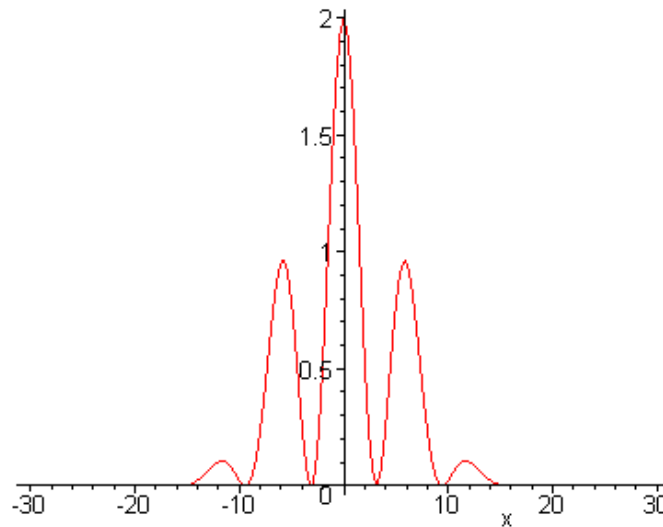


Рис. 3. Интерференция на двух щелях с темным центром

не учитывается влияние теневых частиц на вероятность появления в некоторой точке пространства реальной частицы.

Вычислим состояние $\langle x|s\rangle_2$, определяющее интенсивность попадания реального электрона на экран B из излучающего устройства s и учитывающее взаимодействие реальных и теневых частиц в случае, когда мы не можем знать, через какую щель прошли электрон и теневой электрон. Заметим, что существует четыре возможных варианта в ходе эксперимента в силу того, что либо теневой электрон и реальный электрон проходят через одну и ту же щель, либо через разные щели. Итак, учитывая также, что состояния $\langle a_1|s\rangle_{\theta_2}$, $\langle a_2|s\rangle_{\theta_1}$, $\langle a_1|s\rangle_{\psi_2}$ и $\langle a_2|s\rangle_{\psi_1}$ невозможны в силу определения волновых функций ψ_i и θ_i ³, получаем:

$$\begin{aligned} \langle x|s\rangle_2 = & \langle x|a_1\rangle_{\psi_1} \langle a_1|s\rangle_{\psi_1} \langle x|a_1\rangle_{\theta_1} \langle a_1|s\rangle_{\theta_1} + \langle x|a_2\rangle_{\psi_2} \langle a_2|s\rangle_{\psi_2} \langle x|a_2\rangle_{\theta_2} \langle a_2|s\rangle_{\theta_2} + \\ & + \langle x|a_1\rangle_{\psi_1} \langle a_1|s\rangle_{\psi_1} \langle x|a_2\rangle_{\theta_2} \langle a_2|s\rangle_{\theta_2} + \langle x|a_2\rangle_{\psi_2} \langle a_2|s\rangle_{\psi_2} \langle x|a_1\rangle_{\theta_1} \langle a_1|s\rangle_{\theta_1}. \end{aligned} \quad (16)$$

В итоге, используя (11) – (14), находим:

$$|\langle x|s\rangle_2|^2 = |\psi_1\theta_1 + \psi_1\theta_2 + \psi_2\theta_1 + \psi_2\theta_2|^2 = |\psi_1 + \psi_2|^2 \cdot |\theta_1 + \theta_2|^2. \quad (17)$$

Необходимым условием того, что теневые электроны действительно влияют на распределение частиц на экране B , является наблюдение интерференционной картины в случае учета этого влияния, т.е. мы должны показать наличие чередования минимумов и максимумов функции, определяемой выражением (17), как функции, зависящей от точки x экрана B .

Так как мы полагаем, что интерференционная картина наблюдается, то для функции $|\psi_1 + \psi_2|^2$, являющейся квадратом модуля выражения (15) и определяющей интенсивность попадания электрона на экран B , происходит чередование максимумов и минимумов. Точнее говоря, соответствующий график будет

³В данном случае $i = 1, 2$.

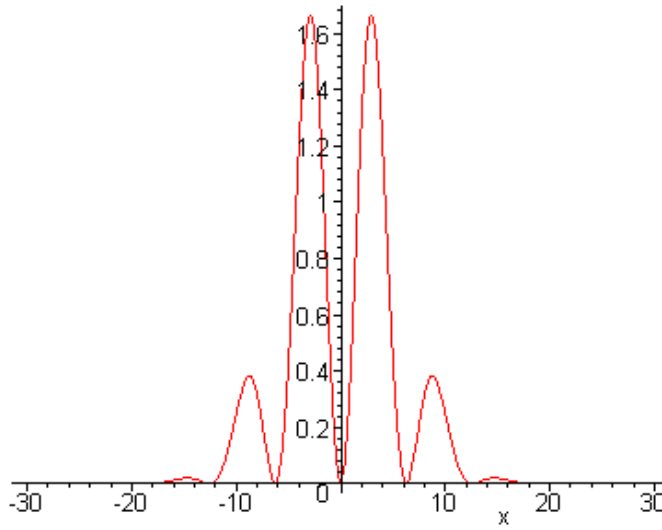


Рис. 4. Интерференция на двух щелях со светлым центром

качественно подобен одному из графиков, представленных на (рис. 3,4). Как известно, в случае, когда мы пропускаем поток электронов через одну щель, то вероятность попадания электрона на экран B определяется гауссовым нормальным распределением. Поэтому соответствующие волновые функции ψ_1 и ψ_2 должны иметь вид ⁴:

$$\psi_1 = u \cdot e^{-A(x+d)^2 + i\alpha(x)}, \quad \psi_2 = u \cdot e^{-A(x-d)^2 + i\beta(x)}, \quad (18)$$

при этом $u^*u \neq 0$, u – биспинор, компонентами которого являются комплексные числа, $A \neq 0$ и не зависит от x , а $\alpha(x)$, $\beta(x)$ такие, что квадрат модуля суммы функций (18) соответствует наблюдению интерференционной картины (рис. 3,4). Кроме этого, d является половиной расстояния между щелями.

Заметим также, что так как теневые частицы, влияющие на интерференционную картину, определяются Дойчем как частицы в параллельных вселенных, практически идентичных нашей, то θ_1 должна иметь такое же распределение, как и ψ_1 , а θ_2 соответственно такое же, как ψ_2 . Тогда, опираясь на результаты теоремы 1, мы с необходимостью должны положить, что

$$\theta_1 = u \cdot e^{-A(x+d)^2 + i \cdot c_1}, \quad \theta_2 = u \cdot e^{-A(x-d)^2 + i \cdot c_2}, \quad (19)$$

$c_i \in \mathbb{R}$. Это означает, что если d достаточно мало, график соответствующей функции $|\theta_1 + \theta_2|^2$ соответствует отсутствию интерференции (рис.5).

⁴Заметим, что, несмотря на то, что мы рассматриваем случай, соответствующий (17), мы должны предполагать и наличие чередования максимумов и минимумов для функции $|\psi_1 + \psi_2|^2$, которая соответствует варианту (15), не учитывающему влияние теневых частиц на интерференционную картину. Поскольку если гипотеза теневых частиц верна, то в приближении, т.е. без учета теневых частиц, должно выполняться существующее ранее объяснение интерференции квантовых частиц. А это означает, что $|\psi_1 + \psi_2|^2$ имеет описанный выше вид. Поэтому определение (18) амплитуд ψ_1 и ψ_2 в случае существования теневых частиц корректно.

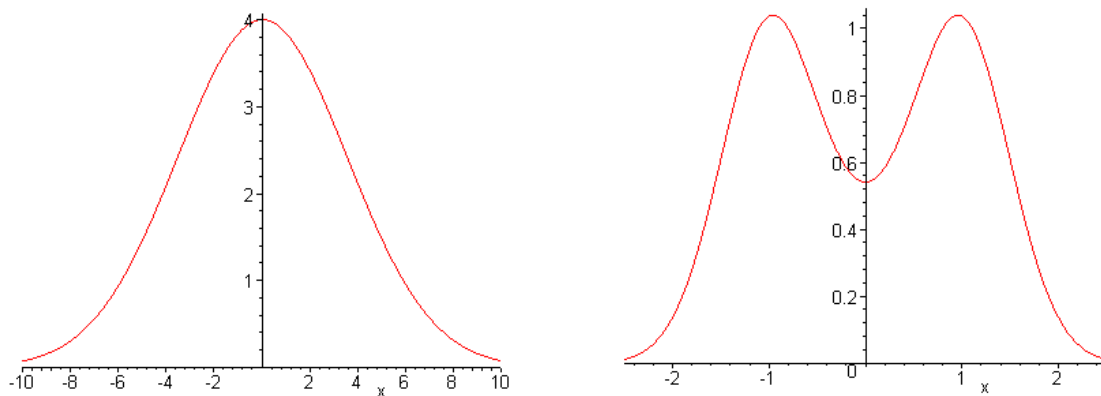


Рис. 5. Амплитуда вероятности $\theta_1 + \theta_2$ в случае достаточно малого расстояния $2d$ между щелями для наблюдения интерференции имеет одно из представленных распределений

Поэтому получаем, что у графика функции (17) происходит чередование максимумов и минимумов. Более того, соответствующая функция и функция $|\psi_1 + \psi_2|^2$ имеют минимумы в одних и тех же точках, так как в силу (19) функция $|\theta_1 + \theta_2|^2$ нигде не равна нулю. В результате мы имеем, что в случае объяснения интерференционной картины столкновением реальных и теневых электронов [2] интерференция действительно наблюдается, если мы отождествляем теневые электроны с электронными духами.

Таким образом, мы показали, что в случае теории теневых частиц Дойча в опыте с двумя щелями интерференция по-прежнему наблюдается, если мы отождествим соответствующие частицы со спинорными духами. При этом точки минимумов интерференционной картины с учетом теневых частиц и интерференционной картины без учета этого взаимодействия совпадают. Но соответствующие графики имеют лишь качественное равенство! Точнее говоря, максимумы функций $|\psi_1 + \psi_2|^2$ и $|\psi_1 + \psi_2|^2 \cdot |\theta_1 + \theta_2|^2$ различны.

2.4. Возможность экспериментальной проверки

Выше нами был рассмотрен опыт по интерференции на двух щелях (рис.2), при этом мы учитывали возможность существования частиц Дойча. Но в соответствующем эксперименте (рис.2) мы предполагали, что за экран A попадает только один теневого электрон. Но за экран A может попасть и большее количество теневых электронов. Как тогда будет выглядеть интерференционная картина? Заметим, что в действительности мы не можем определить количество теневых электронов, прошедших через экран A , в силу их свойств.

Пусть теперь кроме реального электрона, определяемого через волновые функции ψ_1 и ψ_2 , через щели a_1 и a_2 соответственно прошло n теневых электронов, характеризуемых амплитудами $\{\theta_1^{(m)}\}$ и $\{\theta_2^{(m)}\}$, где $m = \overline{1, n}$. При этом ψ_1 и ψ_2 имеют вид (18), а $\{\theta_1^{(m)}\}$ и $\{\theta_2^{(m)}\}$ определяются выражением (19), в котором $\forall m = \overline{1, n} \quad c_i = c_i^m$. Тогда для квадрата модуля амплитуды вероятности попадания электрона в точку x экрана B аналогично формуле (16) получим

выражение:

$$\begin{aligned} \langle x|s \rangle_2 = & \left(\psi_1 \theta_1^{(1)} \cdot \dots \cdot \theta_1^{(n)} + \psi_1 \theta_1^{(1)} \cdot \dots \cdot \theta_1^{(n-1)} \theta_2^{(n)} + \psi_1 \theta_1^{(1)} \cdot \dots \cdot \theta_1^{(n-2)} \theta_2^{(n-1)} \theta_1^{(n)} + \right. \\ & \left. + \psi_1 \theta_1^{(1)} \cdot \dots \cdot \theta_1^{(n-2)} \theta_2^{(n-1)} \theta_2^{(n)} + \dots + \psi_2 \theta_1^{(1)} \cdot \dots \cdot \theta_1^{(n)} + \dots \right). \end{aligned} \quad (20)$$

Сгруппируем члены в (20) следующим образом:

$$\begin{aligned} \langle x|s \rangle_2 = & \left(\psi_1 \theta_1^{(1)} \cdot \dots \cdot \theta_1^{(n-1)} + \psi_1 \theta_1^{(1)} \cdot \dots \cdot \theta_2^{(n-1)} + \dots + \psi_2 \theta_1^{(1)} \cdot \dots \cdot \theta_1^{(n-1)} + \dots \right) \theta_1^{(n)} + \\ & + \left(\psi_1 \theta_1^{(1)} \cdot \dots \cdot \theta_1^{(n-1)} + \psi_1 \theta_1^{(1)} \cdot \dots \cdot \theta_2^{(n-1)} + \dots + \psi_2 \theta_1^{(1)} \cdot \dots \cdot \theta_1^{(n-1)} + \dots \right) \theta_2^{(n)}. \end{aligned}$$

Откуда получаем, что

$$\begin{aligned} \langle x|s \rangle_2 = & \left(\psi_1 \theta_1^{(1)} \cdot \dots \cdot \theta_1^{(n-1)} + \psi_1 \theta_1^{(1)} \cdot \dots \cdot \theta_2^{(n-1)} + \dots + \psi_2 \theta_1^{(1)} \cdot \dots \cdot \theta_1^{(n-1)} + \dots \right. \\ & \left. \dots + \psi_2 \theta_2^{(1)} \cdot \dots \cdot \theta_2^{(n-1)} \right) \left(\theta_1^{(n)} + \theta_2^{(n)} \right). \end{aligned} \quad (21)$$

В результате, последовательно выполняя аналогичную процедуру для выражения (21), получим следующее выражение, определяющее интенсивность попадания электрона на экран B :

$$|\langle x|s \rangle_2|^2 = |\psi_1 + \psi_2|^2 \cdot \prod_{i=1}^n |\theta_1^{(i)} + \theta_2^{(i)}|^2, \quad (22)$$

где n – количество теневых частиц, прошедших через щели за экран A .

Теперь, учитывая вид функций $\theta_1^{(i)}$ и $\theta_2^{(i)}$, указанный выше, мы получаем, что для любого i функции $|\theta_1^{(i)} + \theta_2^{(i)}|$ определяются графиками, качественно подобными графикам, представленным на (рис.5). А это соответствует тому, что (22) описывает некоторую интерференционную картину, т.е. если мы предположим, что за экран A проходит более одной теневой частицы, то на экране B по-прежнему будем фиксировать чередование максимумов и минимумов, соответствующее наблюдению интерференции. Более того, независимо от количества теневых частиц, оказавших влияние на результат эксперимента, минимумы соответствующих функций остаются в одних и тех же точках.

Посмотрим на полученные результаты с точки зрения имеющихся экспериментальных данных. Во всех опытах, связанных с интерференцией квантовых частиц, предполагалось, что полученная экспериментально интенсивность распределения электронов на экране B соответствует функции $|\psi_1 + \psi_2|^2$, умноженной на коэффициент подобия. Это вызвано тем, что функция $|\psi_1 + \psi_2|^2$ должна являться функцией распределения. Поэтому А. К. Гуц предложил связать умножение на произведение

$$\prod_{i=1}^n |\theta_1^{(i)} + \theta_2^{(i)}|^2 \quad (23)$$

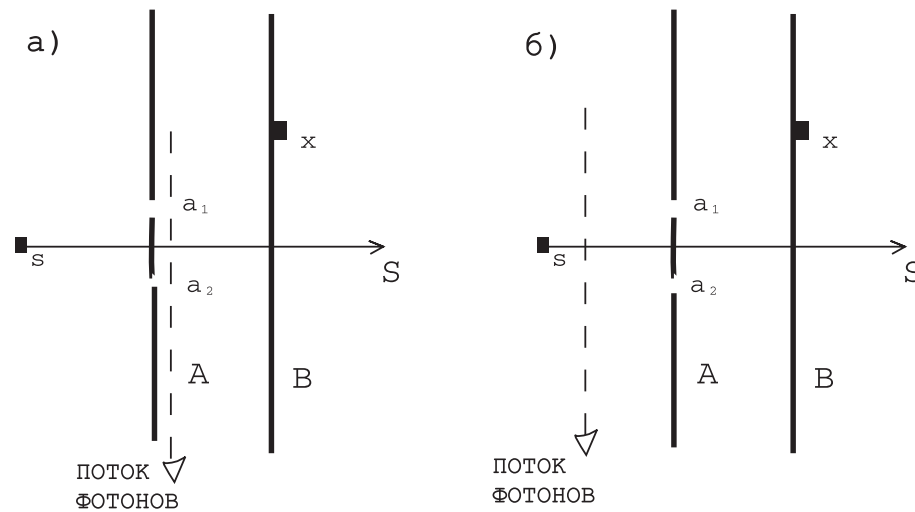


Рис. 6. Представлены а) классический эксперимент, б) предлагаемый эксперимент

с умножением на коэффициент подобия. Каждый новый теневой электрон будет добавлять свой коэффициент, который больше нуля. Если в (23) встретятся взаимнообратные коэффициенты, то их влияние аннулируется. Поэтому когда мы рассмотрим бесконечно много теневых частиц, т.е. положим $n = \infty$, то множитель (23) окажется равным единице и соответственно получим равенство интенсивностей распределения электрона с учетом и без учета теневых частиц. Только дело в том, что щели имеют конечный размер и электроны также должны обладать некоторым конечным размером. Поэтому на результат эксперимента может повлиять только конечное число теневых электронов, т.е. $n \neq \infty$. Что мы и наблюдаем в эксперименте: коэффициент подобия имеет разные значения.

Можем ли мы из полученных данных каким-либо образом проверить существование теневых частиц? Такая вероятность существует. Для этого достаточно научиться контролировать количество теневых электронов, влияющих на результат эксперимента. Проблема состоит в том, что мы не можем никакими приборами зафиксировать частицу Дойча. Но это можно обойти следующим образом.

В действительности нам необязательно знать, сколько теневых частиц прошло через щели a_1 и a_2 . Вспомним про один из экспериментов, определяющих, через какую из двух щелей прошел электрон (рис.6а). Для этого между экранами A и B , как можно ближе к экрану A , пускали поток фотонов; возле которой из щелей обнаруживается рассеивание света – там и прошел электрон. В результате подобного эксперимента интерференционная картина не наблюдалась. Посмотрим на этот эксперимент с точки зрения существования теневых частиц. Реальный электрон взаимодействует с потоком фотонов, но и теневой электрон так же взаимодействует с потоком теневых фотонов, в результате теневой электрон отклонится от своего первоначального движения так же, как и реальный. Столкновение реальной и теневой частиц, возможное в эксперименте без фотонов, может не произойти при испускании потока фотонов. При этом чем ближе

поток фотонов к экрану A , тем меньше теневых частиц, вероятно, влияет на движение реального электрона. Из мысленного эксперимента мы заключаем, что, вероятно, с помощью потока фотонов можно каким-то образом контролировать количество теневых частиц, влияющих на интерференцию.

Попробуем теперь поток фотонов поместить перед экраном A (рис.6б). Возможно, что некоторый сдвиг вправо рассматриваемого потока вызовет уменьшение количества частиц, прошедших сквозь щели, что, в свою очередь, скажется на уменьшении разницы значений двух соседних максимумов. Если в предлагаемом эксперименте будет зафиксирован описанный эффект, то можно говорить о существовании теневых частиц. Заметим, что в тот момент, когда поток фотонов будет находиться достаточно близко от экрана A для того, чтобы можно было бы сказать, через какую из двух щелей прошел реальный электрон, влияние теневых частиц максимально уменьшится и интерференция не будет наблюдаться. При этом, так как в этом случае мы получим дополнительную информацию о пути реального электрона, амплитуда вероятности уже будет удовлетворять не соотношению (22), а следующему выражению:

$$|\langle x|s\rangle_2|^2 = (|\psi_1|^2 + |\psi_2|^2) \cdot \prod_{i=1}^n |\theta_1^{(i)} + \theta_2^{(i)}|^2. \quad (24)$$

Итак, суть предлагаемого эксперимента состоит в изучении интерференционной картины при движении точки испускания потока фотонов вдоль оси S (рис.6б).

Кроме представленного выше эксперимента, можно привести еще один эксперимент, предложенный М. С. Шаповаловой. Предлагается влиять на количество теневых частиц, взаимодействующих в ходе эксперимента с реальными частицами, при помощи изменения размеров щелей: чем меньше щели, тем меньше количество частиц Дойча повлияют на наблюдаемое явление. Правда, в этом случае функция $|\psi_1 + \psi_2|^2$ также будет изменяться. Но, на наш взгляд, из подобного эксперимента также можно получить некоторые сведения о взаимодействии реальных и теневых частиц.

Заключение

Когда мы не в состоянии объяснить наблюдаемое физическое явление на основе существующих теорий, необходимо появление новых идей, иногда значительно отличающихся от старых. Так в начале прошлого века была создана теория относительности, разрешившая ряд противоречий, возникших в физике. Существующая сейчас квантовая теория не в состоянии до конца объяснить некоторые явления – одно из них явление интерференции. Следовательно, нужна теория, наиболее полно объясняющая появления максимумов и минимумов в наблюдаемой интерференционной картине. Теневые частицы Дойча дают нам возможность точнее понять эффекты, связанные с интерференцией квантовых частиц. В данной работе и в [3,4,6,7] был сделан шаг к формальному определению частиц Дойча, существование которых не противоречит экспериментальным данным в области интерференции. Кроме этого в [3,4] определены теневые

фотоны. Заметим, что существовавшее ранее объяснение природы интерференции, основанное на влиянии измерения на результат эксперимента, в действительности неявно учитывает теневые частицы. Поскольку, проводя некоторое измерение, мы влияем на движение теневых частиц «теневыми измерениями», что, в свою очередь, и отражается на получаемых нами результатах. Какое из объяснений наиболее точно, может указать лишь эксперимент. Существование частиц Дойча опирается на иное восприятие физической реальности. Не одна вселенная, а мультиверс Гуца-Дойча [3, 4] дает возможность использования теневых частиц. Признание множественности различных вселенных требует существенного изменения сложившихся взглядов, «сейчас мы обладаем несколькими чрезвычайно глубокими теориями о структуре реальности, главным образом благодаря ряду экстраординарных научных открытий. Если мы хотим понять мир не поверхностно, а более глубоко, нам помогут эти теории и разум, а не наши предрассудки, приобретенные мнения и даже не здравый смысл» [2]. Найденные спинорные духи обладают в некоторой мере экзотическим свойством зануления тензора энергии-импульса, что ранее означало отсутствие каких-либо полей. Но они существуют в силу ненулевой вероятности появления в пространстве. Именно это и послужило поводом к появлению гипотезы тождественности спинорных духов и теневых спинорных полей [6, 7]. В этой работе мы постарались придать этой гипотезе некоторое обоснование. В свете полученных результатов, по крайней мере, можно говорить о том, что пока не видно оснований для опровержения выдвинутого предположения.

ЛИТЕРАТУРА

1. Ландау Л.Д., Лифшиц Е.М. *Теория поля*. М.: Наука, 1988.
2. Дойч Д. *Структура реальности*. Ижевск: НИЦ «Регулярная и хаотическая динамика», 2001.
3. Гуц А.К. *Теоретико-топосная модель мультиверса Дойча* // Математические структуры и моделирование / Под ред. А.К. Гуца. Омск: ОмГУ. 2001. Вып.8. С.76–90.
4. Guts A.K. *Topos-theoretic Model of the Deutsch Multiverse*. Los Alamos E-paper: physics/0203071 (2002).
5. Бьеркен Дж., Дрелл С.Д. *Релятивистская квантовая теория. Т.1. Релятивистская квантовая механика*. М.: Наука, 1978.
6. Palesheva E.V. *Ghost spinors, shadow electrons and Deutsch's Multivers*. Los Alamos E-print gr-qc/0108017 (2001)
7. Палешева Е. В. *Спинорные духи, теневые электроны и Мультиверс Дойча* // Математические структуры и моделирование / Под ред. А.К. Гуца. Омск: ОмГУ. 2001. Вып.8. С.66–75.

О ПАКЕТЕ ПРОГРАММ НЕДИФФЕРЕНЦИРУЕМОЙ ОПТИМИЗАЦИИ

Т.Б. Бигильдеева, В.В. Таркаев

In the article the complex of the nondifferential optimization programs including special means of maintenance of input-output of the data and visualization of optimization process is considered. These features of the software package make its convenient as by development both research of new algorithms and methods of optimization, and with use it in educational process.

Введение

В Челябинском государственном университете с начала девяностых годов под руководством профессора В.Д.Батухтина разрабатываются методы решения негладких и разрывных задач оптимизации, базирующиеся на понятии аппроксимационного градиента [1]. В этой связи возникла потребность в программировании разрабатываемых алгоритмов и проведении численных экспериментов. Однако почти сразу программистская задача была поставлена существенно шире - наряду с программированием численных методов была начата разработка набора средств, позволяющих использовать имеющиеся наработки сразу в нескольких тесно связанных областях научной и преподавательской деятельности.

А именно:

- 1) использование при разработке и тестировании новых алгоритмов;
- 2) использование при решении конкретных прикладных задач;
- 3) иллюстрирование выступлений, посвященных изложению полученных результатов;
- 4) разработка иллюстративно-обучающих программ по курсу «методы оптимизации»;
- 5) использование студентами в ходе подготовки курсовых и дипломных работ по родственной проблематике.

В результате был разработан описываемый программный комплекс : «Пакет программ недифференцируемой оптимизации». Неоднократно он был использован для иллюстрирования выступлений на международных семинарах по недифференцируемой оптимизации. В течение нескольких лет он установлен в локальной сети математического факультета ЧелГУ и был использован уже при подготовке более чем двадцати курсовых и дипломных работ. Но, разумеется, наиболее важное его использование - это разработка и тестирование новых вычислительных методов.

© 2002 Т.Б. Бигильдеева, В.В. Таркаев

E-mail: tbig@cgu.chel.su

Челябинский государственный университет

1. Цели разработки и общая характеристика пакета

При разработке описываемого программного комплекса преследовались следующие цели:

1. Максимальное упрощение для исследователя самого процесса программирования алгоритмов и тестовых задач благодаря использованию специализированных библиотек.
2. Наличие удобных средств ввода/вывода данных.
3. Наличие средств визуализации хода вычислительного процесса.
4. Возможность сохранения информации о ходе и результатах вычислительного процесса.
5. Наличие библиотек задач оптимизации и методов оптимизации с максимально упрощенным доступом к ним.
6. Поддержка коллективной работы.
7. Наличие средств, облегчающих реализацию иллюстративных программ, представляющих собой заранее подготовленный набор вычислительных экспериментов, с возможностью интерактивного изменения ряда параметров и отображением поясняющих текстов.
8. Наличие нетрудоемкой технологии экспорта/импорта алгоритмов и задач оптимизации на уровне исходного текста.

Разработчики описываемого программного комплекса прежде всего ориентировались на потребности вынужденного программировать математика-исследователя и на родственные им потребности студента, программирующего в ходе выполнения курсовых и дипломных работ. В соответствии с этим при разработке структуры пакета и технологии его использования определяющим было стремление, во-первых, максимально упростить стадию программирования алгоритма, обеспечив, однако, во-вторых, максимум удобств на стадии тестирования программы и апробации алгоритма. Ясно, что одновременное достижение двух этих целей возможно лишь при условии достаточно жесткой стандартизации поддерживаемых типов задач. При этом обязательно должна быть оставлена возможность выхода за рамки этого стандарта, платой за это является повышение сложности программирования.

В случае пакета программ недифференцируемой оптимизации таким стандартом является вычислительный процесс, реализующий решение задачи безусловной оптимизации.

Причем он должен, во-первых, быть итерационным и, во-вторых, допускать деление на три блока: оптимизируемая функция, главный метод и вспомогательный метод (может отсутствовать). Как правило, вспомогательный метод - это одномерный метод оптимизации, осуществляющий выбор величины шага по определенному главному методом направлению.

Деление на логически независимые стандартные блоки обеспечивает возможность замены каждого из них на любой другой блок, относящийся к тому же типу, и, кроме того, упрощает их независимую разработку. Тем более, что каждый из таких блоков реализуется как отдельный ехе-файл.

Нематематическая часть программы, реализующей каждый из блоков ал-

горитма, удовлетворяющего двум выше указанным требованиям, в пакете недифференцируемой оптимизации сведена, по сути дела, к нетрудоемкому заполнению своеобразной «анкеты» метода или задачи. Требования же, предъявляемые к реализации математической части, сродни стилистическим. Состоят они в необходимости написания определенных функций с определенными типами и интерпретацией параметров и возвращаемых величин. Например, для задачи оптимизации это отдельная реализация функций, вычисляющих сам функционал и его градиент (если таковой определен) с предопределенными типами возвращаемых величин и параметров (см. пример в пункте 2). В реализации метода обязательно должна быть функция, выполняющая очередную итерацию. Кроме того, в каждом модуле должна быть реализована функция его инициализации.

Реализация методов, предназначенных для запуска в качестве главных и в качестве вспомогательных, почти ничем не различается. Во всяком случае, методы, написанные как вспомогательные, могут запускаться и как главные, что, в частности, упрощает их тестирование.

Простота интеграции в пакет блоков, стандартных с точки зрения описанной схемы, достигается, естественно, за счет стандартизации их интерфейсной части, то есть набора функций, реализованных в некотором блоке и предназначенных для использования другими блоками. Для задач безусловной оптимизации такая стандартизация является достаточно естественной и необременительной для программирующего математика. Но разработка чего-то подобного для более широких классов вычислительных процессов, включающих, например, решение задач условной оптимизации, оказалась практически невозможной, так как пришлось бы учесть слишком много вариантов, что свело бы удобства к минимуму. Поэтому было принято решение пойти по пути реализации простых в использовании средств конструирования интерфейсов самим пользователем пакета. Практически любая функция из одного блока может быть «объявлена» глобальной, что обеспечивает возможность ее вызова из другого блока. В этом случае возможность взаимной замены блоков обеспечивается уже не средствами пакета, а на уровне дисциплины программирования. Подчеркнем, что использование этой технологии не усложняет доступ к сервисным возможностям пакета, а лишь снимает ограничения на структуру реализуемого процесса.

2. Примеры программирования

Приведенные ниже примеры содержат полные тексты программ, реализующих простейшие метод и задачу оптимизации. Подчеркнем, что тексты эти являются именно полными, то есть ехе-файлы, полученные в результате их компиляции и линковки, вполне могут быть запущены в составе пакета для проведения экспериментов.

2.1. Простейшая задача оптимизации (минимизация квадратичной функции)

Математическая часть программы включает функции, вычисляющие значение и градиент оптимизируемого функционала (запись $x*x$ означает скалярное произведение векторов, а запись $2*x$ означает умножение вектора x на 2).

Нематематическая часть состоит из обязательной функции инициализации. В этой функции размерность задачи и координаты начальной точки считываются из файла данных. После этого заполняется «анкета» задачи. В данном примере в модуле содержится только одна задача, и поэтому значение параметра `nproblem` игнорируется.

```
/*Математическая часть */
double func ( vector & x ) { return ( x * x ); }
vector grad ( vector & x ) { return ( 2 * x ); }
/* Нематематическая часть */
int problem_init ( int nproblem )
{
    int n;
    n=input («размерность задачи»);
    vector start_point(n);
    start_point=input («начальная точка»);
    set_problem_name («КВАДРАТИЧНАЯ ФУНКЦИЯ»);
    set_problem_start_point (start_point);
    set_function (func);
    set_gradient (grad);
    return (1);
}
```

2.2. Простейший градиентный метод оптимизации

Метод состоит в запуске одномерного подметода выбора шага по направлению антиградиента. Функция, вычисляющая градиент (`grad`), и сам подметод выбираются на этапе запуска метода. Процесс прекращается, если значение нормы градиента меньше заданного минимального значения градиента (`min_grad`).

Нематематическая часть состоит из функции инициализации и функции, осуществляющей графический вывод данных на каждой итерации. В данном примере выводится только текущая точка (`gr_current_point`) и норма градиента (`gr_norma_grad`).

В функции инициализации заполняется «анкета» метода (имя метода, имя функции, осуществляющей очередную итерацию, и функции, осуществляющей вывод). Инициализируются переменные метода, запрашиваются задача и вспомогательный метод, инициализируются переменные, обеспечивающие графический вывод.

```
/* Переменные метода */
method_type submethod; double min_grad; vector x;
array gr_norma_grad, gr_current_point;
```

```

/* Математическая часть */
int next_iteration (void)
{
    submethod.run(x,-grad(x)); x=submethod.current_point();
    if(norma(grad(x)) < min_grad) return(IR_STOP);
    else return(IR_CONTINUE);
}
/* Нематематическая часть */
void func_output(void)
{
    gr_norma_grad < < norma(grad(x));
    gr_current_point < < x;
}
int method_init (int nmethod)
{
    set_method_name(«Простейший градиентный метод»);
    set_method_next_iteration(next_iteration);
    set_method_output(func_output);
    min_grad=input(«Минимальное значение градиента»);
    submethod=load_submethod();
    load_problem();
    x=problem_start_point();
    gr_norma_grad = output_array ( «grad» );
    int dim = problem_dimension ();
    gr_current_point = output_array («Траектория», dim);
    return(1);
}

```

3. Проведение вычислительного эксперимента

Что же дает включение некоторого модуля в пакет? Или, иначе говоря, какие возможности предоставляются математику на этапе проведения вычислительного эксперимента при условии, что составляющие вычислительный процесс блоки реализованы в соответствии с предъявляемыми пакетом требованиями?

1. Исходные данные вычислительного процесса считываются из текстового файла. В обязательной его части содержится указание на задачу, запускаемый метод и подметод. Подметод может отсутствовать. В остальном набор исходных данных, их имена и типы определяются при реализации конкретных модулей. Например, если где-то в тексте программы содержалась строка

```
start_point = input ( «начальная точка» );
```

где переменная `start_point` имеет тип `vector`, то в файле будет осуществляться поиск строки вида

```
начальная точка = {1.0, 2.0, 3.0 },
```

а если таковой найдено не будет, то на экране появится соответствующий запрос.

Такой способ задания исходной информации ориентирован на этап численного экспериментирования с разрабатываемым методом, когда нередко возникает потребность в смене состава параметров. Для «устоявшихся» модулей имеется другой способ. Он предполагает перечисление в специальном формате имен входных параметров, их рекомендуемых значений и кратких пояснений к ним. После чего все необходимое будет переноситься в файл данных средствами пакета.

2. Каждый из трех блоков – оптимизируемая функция, главный метод и вспомогательный метод – может быть заменен на любой другой ему подобный. Достигается это изменением соответствующего параметра в файле, содержащем исходные данные.

3. Выбор и/или изменение файла данных может осуществляться непосредственно перед запуском вычислительного процесса. Все это сводит к минимуму трудоемкость проведения серий вычислительных экспериментов вроде испытания одного метода на различных задачах. Такие серии могут готовиться и заранее для исполнения без непосредственного участия экспериментатора.

4. По ходу вычислительного процесса и по его окончании на экране могут отображаться последовательности, элементами которых являются либо вещественные числа, либо двумерные векторы, либо векторы произвольной размерности. Номер элемента последовательности есть номер итерации главного метода. Ломанные на плоскости, как правило, являющиеся траекториями метода, могут изображаться на линиях уровня оптимизируемого функционала. Реализованы и стандартные манипуляции с графиками, такие как сдвиги и изменение масштаба. Графические возможности пакета этим не исчерпываются, но для обеспечения выше описанного минимума в тексте программ достаточно иметь по две строки на каждый график: в первой он должен быть определен с заданием отображаемого на экране имени, например

```
array current_point=output_array(«текущая точка»,2);
```

(второй параметр - размерность элемента последовательности). Во второй строке осуществляется сам вывод очередного элемента:

```
current_point << x;
```

Об изображении линий уровня при программировании задачи вообще можно не заботиться. Если она была реализована в соответствии с описанной технологией, то все будет сделано средствами пакета.

5. По окончании расчета или после его прерывания данные (если они вводились описанным выше способом) могут быть сохранены в текстовом файле. Имеется также и возможность бинарного сохранения данных, что делает возможным последующий просмотр их с использованием выше перечисленных графических возможностей.

6. При необходимости могут быть включены счетчики вычислений оптимизируемого функционала и его градиента, причем возможен отдельный подсчет вычислений, осуществляемых из главного и вспомогательного методов. Если задача оптимизации была реализована в соответствии с выше описанными требованиями, то использование счетчиков не требует ничего от программы.

4. Структура и составные части пакета

4.1. Структура пакета

Структура пакета продиктована стремлением одновременно решить две сформулированные выше задачи: во-первых, максимально упростить программирование определенных классов оптимизационных задач, обеспечив при этом, во-вторых, максимум сервиса, специфичного для этого класса, на этапе проведения численных экспериментов.

И если для решения первой задачи достаточно разработать специализированные библиотеки, то более качественное решение второй может быть обеспечено благодаря использованию специализированных программ.

Именно этими соображениями и было продиктовано решение реализовать описываемый программный комплекс не как одну многофункциональную программу, а как набор специализированных программ. Решение это было в значительной степени предопределено выбором WINDOWS в качестве основной операционной системы.

Итак, пакет можно разделить на три части.

1. Служебная часть пакета, состоящая из нескольких .dll и .exe файлов. Они обеспечивают подготовку и запуск вычислительных процессов, управление ими, хранение, просмотр и сохранение информации, текстовый ввод/вывод данных.

2. Библиотеки функций и классов в смысле языка C++, используемые при разработке программ, предназначенных для работы в составе пакета.

3. Реализованные на базе пакета модули, относящиеся к одному из четырех следующих типов :

- А) задачи оптимизации,
- Б) методы оптимизации,
- В) иллюстративно-обучающие программы,
- Г) модули без предопределенного набора интерфейсных функций.

Пользователи пакета занимаются разработкой приложений, относящихся к третьей категории.

Реализация пользовательских модулей как самостоятельных WINDOWS-приложений позволяет:

- во-первых, избежать существенной части проблем, которые возникли бы при работе нескольких человек над одной программой;

- во-вторых, пользуясь многозадачностью WINDOWS, формировать вычислительный процесс из нескольких независимо разработанных модулей (часть из которых может быть объединена в библиотеки);

- в-третьих, обеспечивает определенную устойчивость пакета в целом к ошибкам, случившимся в одном из модулей.

4.2. Вспомогательные библиотеки

В состав пакета включены несколько файлов .lib. Объединенные в них функции и классы можно разделить на несколько групп :

1. Функции, позволяющие реализовать некоторый модуль как модуль определенного типа (метод, задача, иллюстративно-обучающая программа). Они, в

частности, позволяют начинающему пользователю разрабатывать WINDOWS-приложения, совершенно не зная особенностей программирования и тонкостей под WINDOWS.

2. Функции и классы, обеспечивающие стандартное взаимодействие стандартных блоков (методов, подметодов, задач), составляющих вычислительный процесс.

3. Функции и классы, обеспечивающие текстовый ввод/вывод данных различных типов.

4. Функции и классы, обеспечивающие доступ к графической подсистеме пакета.

5. Функции и классы, обеспечивающие программное управление подсистемой сохранения результатов вычислительного процесса.

6. Функции, используемые при разработке иллюстративно-обучающих программ, реализуемых на базе пакета.

7. Вспомогательные типы данных: вектор, матрица, динамический массив.

Возможности, предоставляемые пакетом при программировании, можно условно разделить на два уровня по сложности освоения. Возможности первого уровня ориентированы на пользователей, либо обладающих лишь начальными навыками программирования на C++ (например, студентов непрограммистских специализаций), либо на пользователей, не желающих тратить время и силы на освоение дополнительных возможностей. Таким пользователям предоставляется возможность реализовать задачи и методы оптимизации, почти не тратя сил на нематематическую часть программ. При этом практически все сервисные возможности пакета остаются доступными при проведении вычислительных экспериментов с таким образом написанными модулями.

С другой стороны, имеется возможность при разработке приложений, выходящих за пределы схемы «главный метод, вспомогательный метод, задача оптимизации», либо приложений, нуждающихся в пользовательском интерфейсе, учитывающем специфику конкретной задачи (прежде всего иллюстративно-обучающих программ), программировать на более низком уровне, что, разумеется, требует более высокой программистской квалификации и несколько больших затрат на освоение соответствующих средств.

Доступна также возможность программирования с использованием непосредственно программного интерфейса WINDOWS.

4.3. Библиотека задач оптимизации

Библиотека задач оптимизации в пакете недифференцируемой оптимизации представляет собой набор ехе-файлов, реализующих задачи оптимизации и запрограммированных для работы в составе пакета.

Понятие задачи включает в себя две обязательные компоненты: оптимизируемый функционал и начальную точку процесса. Могут быть заданы еще и несколько необязательных: градиент оптимизируемого функционала, точка минимума, название задачи.

Библиотека задач оптимизации делится на три части:

1) задачи безусловной оптимизации дифференцируемых функций;

2) задачи безусловной оптимизации недифференцируемых, но непрерывных функций;

3) задачи безусловной оптимизации разрывных функций.

Библиотека задач оптимизации дифференцируемых функций включает в себя задачи оптимизации известных функций, таких, как расширенная функция Розенброка, функция Бокса, обобщенная функция Пауэлла, функция Вуда, функций Кларка и ряд других функций [3, 4, 6], каждая из этих задач обладает некоторой «изюминкой», создающей проблемы в процессе оптимизации, например плохой обусловленностью, вырожденностью точки минимума, многоэкстремальностью и так далее. Апробация численных методов и алгоритмов на этих задачах позволяет получить достаточно полную картину о возможностях методов, их достоинствах и недостатках.

Задачи оптимизации недифференцируемых функций состоят из задач выпуклого программирования (это задача аппроксимации полиномами в L_1 , минимизации негладкой функции Шора и другие) [4], а также из задач оптимизации невыпуклых функций.

Задачи оптимизации разрывных функций сконструированы из известных задач оптимизации гладких функций путем «введения» разрывов (к таким задачам, в частности, относится задача минимизации функции Розенброка, дополненной разрывом по гиперплоскости, проходящей через точку минимума вдоль оврага), а также путем сведения задач оптимизации с ограничениями к задачам безусловной минимизации разрывных функций (это возможно благодаря введению точных штрафов или специального продолжения минимизируемой функции вне области задания). При построении этих задач моделировались различные ситуации, усложняющие решение задачи, например прохождение разрыва через точку минимума, медленное изменение функции вдоль разрыва, что затрудняет движение к минимуму вдоль разрыва и так далее. В этом классе задач присутствуют задачи минимизации различных кусочно-линейных разрывных функций, а также задачи, возникающие на практике, например задача оптимизации распределения потока воды в ГРЭС [8], и другие.

Библиотека задач оптимизации постоянно пополняется, что расширяет возможность проведения численных экспериментов и позволяет глубже исследовать достоинства и недостатки создаваемых методов и алгоритмов.

4.4. Библиотека методов оптимизации

Библиотека методов оптимизации в пакете недифференцируемой оптимизации представляет собой набор ехе-файлов, реализующих методы оптимизации и запрограммированных для работы в составе пакета. Этот набор, в частности, включает в себя методы одномерной оптимизации, предназначенные для использования в качестве вспомогательных для выбора величины шага по выбранному главным методом направлению.

Библиотека методов оптимизации делится условно на две части:

- 1) методы оптимизации дифференцируемых функций;
- 2) методы недифференцируемой оптимизации.

Методы оптимизации дифференцируемых функций включают в себя известные методы безусловной оптимизации: градиентные методы, квазиньютоновские методы, метод Ньютона и его модификации [4, 5, 7]. Эти методы могут работать с различными процедурами выбора шага по направлению (выбором шага путем одномерной минимизации методом золотого сечения, выбором шага по правилу Армихо, по правилу Голдстейна) [5, 7].

Методы оптимизации недифференцируемых функций включают в себя методы прямого поиска (метод покоординатного спуска, метод многогранника Нелдера-Мида), субградиентные методы (метод Шора, метод Поляка, субградиентный метод последовательной релаксации) [4, 5], а также метод Конна-Монго минимизации кусочно-линейных разрывных функций, основанный на декомпозиции функции [2].

Основное место в библиотеке методов занимают методы, созданные на базе понятия аппроксимационного градиента. К этим методам относятся метод аппроксимационного градиента и его модификации, аппроксимационный аналог метода Ньютона, адаптивные алгоритмы [1, 9]. Данные методы предназначены для решения широкого класса задач безусловной оптимизации разрывных функций. Эти методы постоянно совершенствуются, создаются их новые модификации.

В настоящее время пакет программ недифференцируемой оптимизации успешно используется как при выполнении научно-исследовательских работ, так и в учебном процессе. В частности, на базе пакета создано компьютерное пособие «Соревнование методов оптимизации» [10]. Лекции с использованием этого пособия вызывают большой интерес у студенческой аудитории.

ЛИТЕРАТУРА

1. Батухтин В.Д., Майборода Л.А. *Разрывные экстремальные задачи*. С.-П.: Гиппократ, 1995. С.358.
2. Conn A.R., Mongeau M. *Discontinuous piecewise linear optimization*. // Mathematical Programming. 1998. V.80. P.315–380.
3. Дэннис Дж., Шнабель Р. *Численные методы безусловной оптимизации и решения нелинейных уравнений*. М.: Мир, 1988.
4. Поляк Б.Т. *Введение в оптимизацию* М.: Наука, 1983. С.384.
5. Гилл Ф., Мюррей У., Райт М. *Практическая оптимизация*. М.: Мир, 1985. С.509.
6. Кларк Ф. *Оптимизация и негладкий анализ*. М.: Наука, 1988. С.279.
7. Бертсекас Д. *Условная оптимизация и методы множителей Лагранжа*. М.: Радио и связь, 1987. С.400.
8. Klein E.M. and Sim S.H. *Discharge allocation for hydro - electric generating stations* // European Journal of Operational Research. 1994. V.73. P.132–138.
9. Батухтин В.Д., Бигильдеев С.И., Бигильдеева Т.Б. *Численные методы решения разрывных экстремальных задач* // Изв.РАН. Теория и системы управления. 1997. N.3. С.113–120.
10. Бигильдеева Т.Б., Таркаева О.В. *Компьютерное пособие «Соревнование методов оптимизации»* // Тез.докл. ВВМШ «Понтрягинские чтения» 3-9 мая 2001. Воронеж, ВГУ. С.23.

СБОР ИНФОРМАЦИИ О ДОСТУПНЫХ РЕСУРСАХ ИНТЕРНЕТ

И.А. Земсков

We first study the generic search engine architecture, then review all problems of collecting information about Web's resources and finally describe three building conception of search engine's module for collecting information about Web's resources.

Введение

Согласно последним исследованиям, в Интернете уже опубликовано более двух миллиардов страниц, и их число экспоненциально увеличивается [22, 25]. Таким образом, с каждым годом, месяцем, днем становится труднее находить «дорогу» к нужной информации. По этой причине становятся все более актуальными научные исследования в области информационного поиска в Интернет. Результаты этих исследований [2, 4] нацелены прежде всего на создание новых поисковых систем [26, 27] или на усовершенствование алгоритмов работы уже существующих средств информационного поиска.

Целью данной статьи является рассмотрение такого аспекта работы информационно-поисковых систем в Интернет, как технологии сбора информации о содержимом Веб-ресурсов. Для того чтобы обсуждать технологии, предложенные различными исследователями, нужно удостовериться в том, что мы правильно понимаем общие принципы функционирования поисковых систем в общем и процесса сбора информации о Веб в частности. Поэтому вначале мы рассмотрим некоторые общие понятия и получим краткое описание процессов, протекающих в поисковых системах. Далее мы опишем основные проблемы, с которыми сталкивается процесс сбора информации, и будем опираться на них при рассмотрении различных концепций поведения, направленных на получение «представления» о Веб-ресурсах.

1. Основные компоненты поисковых средств

Множество работ по проблемам поиска в Интернете описывают схему организации поисковых средств [17, 19, 21, 23], но для целей данной статьи мы приведем эти схемы к общему виду. Результат можно видеть на рисунке 1.

Основные компоненты поисковых средств Интернет (1): модуль сбора информации о доступных ресурсах, репозиторий собранной информации, модуль

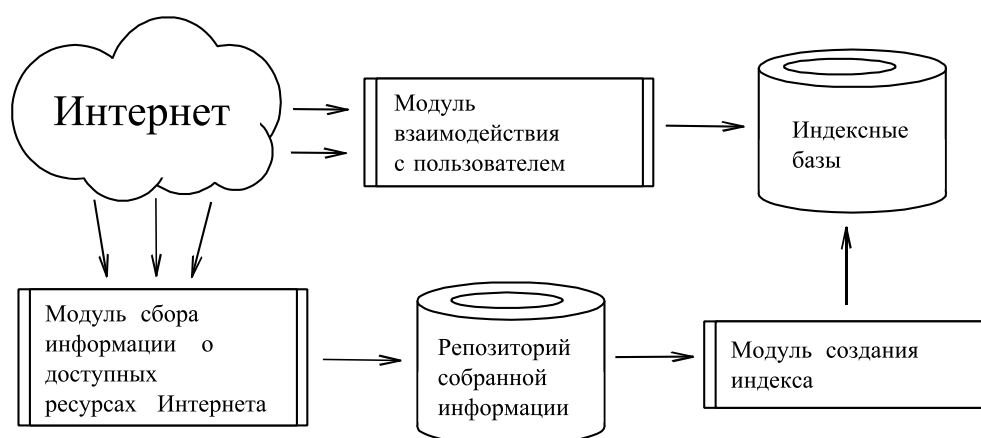


Рис. 1. Обобщенная схема организации поисковых систем Интернет

создания индексов, сами индексы (индексные базы) и модуль взаимодействия с пользователем.

Постараемся описать каждый элемент схемы, не выходя при этом за рамки обобщения.

Интернет. Следует отметить, что в Интернет реализовано и функционирует несколько технологий организации информационных ресурсов (например протоколы: NNTP, WAIS, Gopher, HTTP, FTP). Самый популярный источник информационных ресурсов – это Веб (WWW реализовано с использованием протокола HTTP). В пользу этого утверждения говорят факты увеличения как пользователей ресурсов этого типа, так и рост количества самих ресурсов. Однако многие поисковые системы позволяют своим пользователям производить поиск не только по источникам Веб. Например, стало популярным производить поиск по сообщениям новостных групп (реализация новостных групп основана на использовании протокола NNTP).

В дальнейшем изложении мы сознательно ограничим себя информационными источниками из Веб. Сузим тем самым область охвата поисковых систем. Однако это сужение позволит нам досконально разобраться с одной средой (под средой понимается множество информационных источников, реализованных в рамках одной технологии), и в дальнейшем попытаться перенести свои наработки на другие информационные среды.

Ограничившись средой Веб, необходимо разобраться с тем, что же мы будем понимать под словами «информационный ресурс» в данном контексте. Для начала надо вспомнить, что в этой среде есть несколько типов ресурсов, имеющих свою вполне определенную «природу». Основными типами ресурсов являются следующие: форматированный текст (для форматирования используется HTML код), графические изображения (форматы: JPEG, GIF, PNG, SWF), аудиофайлы (форматы: WAV, MIDI, MP3, RA), видео. Таким образом, помня о том, что ресурс любого из этих типов может стать объектом поиска, мы будем понимать под словами «информационный ресурс» некий файл, имеющий внутреннюю структуру согласно спецификации одного из объявленных типов и

находящийся на специальном сервере (информационный источник) в Интернет. Сервер должен обеспечивать доступ к этому файлу по протоколу HTTP.

Модуль сбора информации о доступных ресурсах. На входе этот модуль имеет все множество информационных ресурсов Веб (помним о нашем ограничении на Интернет). Функция модуля заключается в том, чтобы, используя некий внутренний алгоритм, собрать в одном месте информацию обо всем многообразии доступных информационных ресурсов. Таким образом, на выходе этого модуля получаем репозиторий собранной информации, т.е. некое представление о Веб, готовое к удобному использованию.

Репозиторий собранной информации. Как было указано ранее, репозиторий – это некое представление о Веб. Далее мы поясним данное утверждение.

Учитывая огромные размеры хранимой в Вебе информации разного типа, становится понятным желание избежать простого складирования на своих дисках копий тех данных, которые есть в Веб. Тем самым мы приходим к решению о целесообразности создания «некоего представления» о Веб, описывающего существующие ресурсы Веб без потребности хранить их у себя в полном объеме.

Из всего выше сказанного вытекает, что репозиторий – это пассивный элемент схемы, т.к. сам он никаких действий по отношению к другим элементам схемы не выполняет. Его основная функция заключается в накоплении в нужном виде того, что ему передаёт на хранение модуль сбора информации. Другой его функцией является предоставление накопленных сведений как модулю сбора информации, так и модулю создания индексов.

Модуль создания индексов. Выполнение поиска на данных, накопленных в репозитории, является неэффективным. На этом факте основано решение об использовании индексных структур. Таким образом, на вход рассматриваемого модуля поступает информация, накопленная в репозитории. Он использует её для построения индексных структур. Количество и состав индексов зависит от цели поставленной перед разработчиками поисковой системы. То есть на выходе этого модуля мы получаем набор из индексных баз (минимум одну базу).

Индексы. Индексы сами по себе также являются пассивным элементом схемы, т.к. не производят действий по отношению к другим элементам схемы. Основная функция индексов заключается в специальной организации данных, с помощью которых можно с приемлемой эффективностью производить поиск нужной информации среди всего массива накопленной информации о Веб.

Модуль взаимодействия с пользователем. На вход данного модуля поступают запросы пользователей. Причем реализация языка запросов пользователей преимущественно стремится к естественному языку. Таким образом, основная функция данного модуля заключается в переводе запросов пользователя с языка, понятного пользователю, на язык, понятный машине, работающей с индексом. На выходе этого модуля получается список ссылок на информационные ресурсы Веб.

2. Проблемы

Перед разработчиками поисковых систем встает большое количество проблем, их рассмотрению посвящено множество работ [11, 13, 15, 17, 21]. В данном параграфе делается попытка обобщения проблем, которые стоят перед всеми исследователями и непосредственно влияют на архитектурные решения при разработке модуля сбора информации о доступных ресурсах. В дальнейшем такое обобщение позволит нам при рассмотрении вопросов разработки рассматриваемого модуля глубже понимать суть применяемых стратегий и концепций.

Возникающие перед исследователями проблемы можно разделить на три класса. В первый класс мы отнесём проблемы, связанные с «природой» информационных ресурсов. В этот класс попали следующие проблемы [15, 17, 21]:

- **Большой объем информационных ресурсов:** размер накопленных человечеством в рамках Веб данных поистине огромен. С этим фактом трудно спорить, т.к. исследователи в данной области давно оперируют сотнями гигабайт в своих работах [2]. Также нельзя выпускать из виду факт экспоненциального роста объемов накапливаемой в рамках Веб информации.
- **Разное время существования информационных ресурсов:** документы или файлы могут быть легко добавлены и так же легко удалены в Веб. Для большинства других членов сети Интернет эти манипуляции с файлами могут остаться незамеченными, но в целях обеспечения эффективного поиска видится необходимым четкое отслеживание производимых изменений в содержимом информационных ресурсов. Тем самым, например, можно заметно уменьшить шанс столкновения в результатах поиска со «сломанной» ссылкой, т.е. ссылкой на уже несуществующий ресурс.
- **Разнородность информационных ресурсов:** данные информационных ресурсов в Веб разнотипны. Они создаются в различных форматах, имеют различную медиаприроду (текст, звук, изображение), а также различаются по применяемым естественным языкам.
- **Динамичность изменения содержимого информационных ресурсов:** для пояснения этой проблемы достаточно вспомнить страницы сайтов с расположенным на них разделом «Новости» или «Объявления». В зависимости от активности владельца ресурса информация в этих разделах может меняться от «очень часто» (раз в 10 минут или еще чаще) до «очень редко» (раз в год или еще реже). Независимо от скорости изменения содержимого страницы её адрес для посетителей сайта остается прежним, благодаря чему становится большой вероятностью из поисковой системы посетителю попасть на страницу, которая уже не содержит нужную информацию.
- **Различное качество и уровень полезности информационных ресурсов:** здесь под словом «качество» понимается широкий круг проблем, начиная с того, что в процесс создания многих ресурсов не привлекаются профессиональные программисты, дизайнеры, редакторы, и заканчивая тем, что в сети очень много дублирующих друг друга ресурсов. Отдель-

ным блоком можно выделить правовые и морально-этические вопросы, касающиеся качества публикуемой информации. Еще один блок вопросов ставит нас перед философской дилеммой целесообразности владения информацией о некотором доступном ресурсе. Своими корнями дилемма уходит к проблеме большого объема доступных в Веб ресурсов.

- **«Скрытность» информационных ресурсов:** в последнее время появились работы, в которых рассматривается проблема организации поиска по так называемому скрытому Веб [4]. Под словами «скрытый Веб» здесь понимаются информационные ресурсы для получения доступа, к которым нужно пройти сложную процедуру регистрации или сформировать с помощью предлагаемой разработчиками ресурса формы некий запрос. Однако на этапе регистрации и при составлении запросов возникают большие сложности. Например, они начинаются уже с того, что заранее не известны ни предлагаемые формы, ни смысл их полей ввода, и заканчиваются тем, что нужно научиться понимать ответы систем на посланные нами запросы. А этап регистрации вдобавок может состоять из нескольких стадий, что еще более усложняет задачу. Одним словом, разработка системы автоматического сбора информации о таких ресурсах сродни разработке системы искусственного интеллекта.
- **Доступность информационных ресурсов:** суть этой проблемы заключается в различном качестве коммуникационных связей между распределенными по всему миру узлам и сегментам сети Интернет. Бывают моменты, когда узел по причинам плохого уровня связи с Интернет становится труднодоступен остальным членам сети. Однако для нас является нежелательным отказ от учета информации, предоставленной этим узлом (источником).

Во второй класс отнесём проблемы, касающиеся нагрузки на различные элементы, участвующие во взаимодействии:

- **Минимизация нагрузки на информационный источник:** независимо от выбираемой стратегии сбора информации о доступных информационных ресурсах информационный источник (место хранения информационных ресурсов) будет нести потери вычислительной мощности и неэффективной нагрузки на его аппаратные ресурсы (например создание нагрузки на дисковые накопители). Это, в свою очередь, может вызвать вполне законное негодование со стороны владельцев информационного источника.
- **Минимизация нагрузки на каналы связи:** здесь следует вспомнить о том, что сбор информации о доступных ресурсах подразумевает перекачивание определенных объемов информации с источника информации на узел сбора по существующим коммуникационным линиям. Если принять во внимание упоминавшиеся большие объемы хранимой информации, то станет понятно, что перекачка со всех источников создаст некоторую (довольно заметную) нагрузку на каналы Интернет. Что вполне закономерно отражается на плате за использование этих самых каналов связи. Еще нужно помнить о том, что каналы связи не имеют способности по первому же

запросу увеличивать беспрестанно свою «ширину», т.е. они имеют вполне определенную максимальную пропускную способность.

- **Оптимизация нагрузки на модули сбора и накопления информации:** описывая данную проблему, следует еще раз напомнить тот факт, что модуль сбора информации должен будет обрабатывать большие объемы данных, приводя их в надлежащий вид, пригодный для, возможно, своей последующей работы и работы модуля создания индексов. Т.е. без должной оптимизации всех нюансов (алгоритмов, структур данных и т.п.), протекающих в модуле процессов, мы невольно обрекаем себя на бесполезную трату дорогостоящих ресурсов.

К третьему классу относятся все проблемы, касающиеся внедрения технологий, положенных в основу модуля сбора информации о доступных ресурсах. Весь список проблем мы приводить не будем, однако отметим, что большинство из них вызвано большой разобщенностью владельцев информационных ресурсов и разработчиков поисковых систем.

3. Концепции

В прошлом параграфе мы кратко рассмотрели проблемы, стоящие перед разработчиками модуля сбора информации о доступных ресурсах. Вполне резонно предположить, что различные группы разработчиков по-разному ставят приоритеты при решении этих проблем. Однако при всем кажущемся многообразии подходов к решению этих проблем все существующие стратегии реализации модуля сбора укладываются в три конкурирующие концепции. В рамках каждой из концепций общие проблемы, описанные в предыдущем параграфе, принимают новый вид, новое осмысление. В проблемах появляется некоторая конкретика, касающаяся направления их решения. Как следствие выбора одной из концепций становится постановка нового круга проблем, на решение которых исследователи бросают все свои усилия. Так происходит во всех отраслях знаний, а область исследования проблем поиска информации в Интернет не стала исключением. Но есть косвенные факты, которые заставляют задуматься о правильности выбираемых направлений приложения усилий. Например, сейчас модными стали публикации об исследовании объемов «охвата» Веб поисковыми системами [25]. Эти исследования показывают, что даже самые «большие» поисковые системы охватывают своим поиском лишь малый процент доступных ресурсов (конкретные цифры объемов на текущий момент можно посмотреть в приведенной ссылке). С учетом непрерывного роста объемов Веб еще более актуальным становится вопрос о правильности выбранной разработчиками «больших» поисковых систем концепции сбора информации о доступных ресурсах Веб. Другим косвенным фактом может служить само существование нескольких концепций.

После всего выше сказанного видится актуальным применение широко известных научных методов исследования, например имитационного моделирования, для исследования эффективности и перспективности существующих, а также вновь разрабатываемых в рамках каждой концепции технологий сбора

информации о доступных ресурсах Веб.

Далее мы кратко опишем все три концепции. Рассмотрение начнем с самой популярной и наиболее проработанной концепции:

Концепция роботов. Согласно этой концепции весь модуль сбора информации располагается и работает на аппаратном обеспечении разработчика поисковой системы. Другими словами, разрабатывается некий программный комплекс, который реализует модуль сбора информации для поисковой системы и оставляет без изменения программные технологии, лежащие в основе средств создания информационных источников (т.е. Веб-сервера). Основу реализаций программного комплекса составляет некоторый программный код, именуемый в литературе сетевым роботом, пауком, краулером и т.п. Т.к. единого мнения по этому вопросу нет, то мы договоримся далее в статье называть его роботом. Алгоритм работы робота заключается в рекурсивном «обходе» ресурсов Веб и извлечении из «обойденных» ресурсов ссылок (URL) на новые ресурсы. Его работа начинается с некоторого набора ссылок на ресурсы Веб и заканчивается при выполнении некоторого условия. Под словом «обход» здесь понимается скачивание ресурса к себе для последующей обработки. Таким образом, «обойденные» ресурсы - это ресурсы Веб, которые уже скачаны роботом к себе и о содержании которых составлено некоторое представление.

Первым результатом данной концепции становится отодвигание на задний план или практически полное снятие с рассмотрения одной из проблем, описанной в прошлом параграфе, а именно проблемы разобщенности разработчиков поисковых систем и владельцев информационных ресурсов. Фактически проблема остается, но теперь разработчики поисковых средств остаются одни перед лицом других, не менее серьезных, проблем и пытаются преодолеть их только своими силами. А круг оставшихся проблем настолько широк и разнообразен, что вызывает у них логичное желание сконцентрировать свои усилия не на всех, а только на некоторых проблемах. Это выражается, например, в желании разрабатывать специализированных роботов, т.е. роботов, которые имеют свою стратегию обхода Веб и свои условия остановки работы, продиктованные его специализацией. Такие роботы имеют различный объем охвата доступных ресурсов Веб. Например, робот для создания представления о доступных музыкальных ресурсах Веб может с известной долей рвения отвергать ссылки на графические ресурсы и прекращать составление представления об информационном источнике (Веб-сервере) при достижении определенного уровня проникновения в его «глубины». Еще одним интересным примером специализированного робота может быть робот, направленный на сканирование скрытого Веб [4].

Создание специализированных роботов многим исследователям кажется некоторой панацеей. Зная некоторые особенности организации и существования ресурсов в какой-либо информационной области интересов, исследователи могут далеко продвинуться в решении проблемы максимального охвата при сборе информации о ресурсах в рамках выбранной тематики. Однако собранная информация будет касаться только одной области интересов! А как же быть с другими интересными тематиками? Создавать новых роботов?

В рамках данной концепции проблема охвата максимально возможного объ-

ема доступных ресурсов неразрывно связана с проблемами, попавшими во второй класс прошлого параграфа, а именно минимизации и оптимизации нагрузок на участвующие во взаимодействии аппаратные ресурсы. С ростом объемов Веб эти проблемы только обостряются. Даже самые простые подсчеты передаваемых от информационных источников к поисковым системам по каналам связи объемов данных показывают большие объемы финансовых расходов [25]. Но сами по себе большие финансовые расходы еще не повод для беспокойства. Беспокойство появляется после обнаружения некоторых подробностей алгоритмов работы модулей поисковой системы. А именно того факта, что в скачиваемой информации содержится большой объем html-кода, который выступает некоторой оболочкой для форматированного текста и который после стадии извлечения «полезной» информации попросту удаляется.

Дополнительные сведения о проблемах и предлагаемых решениях, связанных с данной концепцией, можно почерпнуть из специальных обзоров [15, 21].

Нам остается отметить лишь тот факт, что при построении модели данной концепции нужно будет очень аккуратно подходить к её формализации, т.к. накоплен большой багаж наработок и предложений по улучшению определенных моментов в работе роботов [1, 2, 5–9, 11, 18, 20] и отказ от их учета в модели может повлечь за собой построение неадекватной модели.

Концепция сенсоров. Основной причиной возникновения данной концепции стала попытка найти более дешевый метод изменения представления о содержимом доступных информационных ресурсов вместе с изменениями самих ресурсов. Суть концепции сенсоров кроется в специальной доработке программного обеспечения, находящегося на стороне информационного источника, т.е. Веб-сервера. Доработка Веб-сервера заключается в создании некоего модуля, который доступными ему средствами пытается обнаружить новые ресурсы на данном сервере или пытается обнаружить изменения в уже найденных ресурсах. После обнаружения каких-либо изменений в состоянии информационного источника модуль сообщает об этих изменениях некоему «головному» серверу.

Одна из попыток реализаций данной концепции была осуществлена в [3]. Результатом этой работы стал модуль, который автоматически с определенной периодичностью исследовал содержимое каталогов сервера в поисках изменений с момента его последнего запуска. Текущее состояние сравнивалось с состоянием, сохраненным в специальных файлах. При обнаружении изменений создавался файл с так называемой мета-информацией, которая описывала суть произошедших изменений. Затем файл архивировался и отсылался модулем на специальный сервер. После этого роботы специального сервера на основе принятой мета-информации принимали решение о скачивании ресурса для последующей обработки.

Заметным недостатком данной реализации можно назвать то, что модуль работает только с серверами, на которых информация хранилась в виде файлов. А этот подход к созданию информационных Веб-ресурсов начинает уступать место т.н. динамическим сайтам, в которых содержимое страницы берется из базы данных и соединяется с шаблоном дизайна в момент запроса странички посетителем сайта.

В статье [24] предлагается еще один вариант сенсора. Но теперь сенсор не является активным по отношению к поиску изменений содержимого, т.е. предложенный вариант модуля можно назвать пассивный сенсор (прошлomu варианту больше подходило название активный сенсор). Другими словами, в программное обеспечение Веб-сервера предлагается встроить модуль, который будет «следить» за поступающими запросами информационных ресурсов и «слушать» ответы сервера. Каждому запросу ставится в соответствие ответ сервера, и далее эта пара ищется в уже накопленной базе мета-описаний ресурсов. При обнаружении изменений предлагается рассмотреть два варианта поведения: в первом случае предлагается послать основному серверу некое мета-описание найденных изменений, а во втором предлагается сразу провести предварительную обработку найденного ресурса по очищению от «мусора» (это в большей степени относится к страницам в формате HTML). Такая очистка (например от лишних конструкций языка разметки) может дать уменьшение нагрузки на каналы связи. Но даже в случае передачи одного мета-описания мы можем получить выигрыш, т.к. нет необходимости постоянных повторных обходов, как это наблюдается в концепции роботов.

Большим недостатком данного предложения является то, что оно существует пока только на бумаге, т.е. еще не существует программной реализации, способной показать практические результаты от её применения. Однако уже сейчас можно предложить еще одно направление исследований касательно этого варианта сенсора. Оно заключается в том, чтобы изучить поведение сенсора на объектах самой сложной природы, а именно на сайтах, подпадающих под определение скрытого Веб. Например, видится интересным получение ответа на вопрос: сколько времени потребуется пассивному сенсору для помощи поисковой системе в составлении представления о самых полезных данных, хранимых «за формой» на одном из ресурсов (подразумевается то, что вопросы через форму будут задавать посетители сайта).

В заключение описания данной концепции хочется указать одну общую проблему для её исследователей. Проблема заключается в том, что концепция подразумевает вмешательство в программное обеспечение Веб-серверов. А это является огромным сдерживающим фактором на пути внедрения технологий, основанных на этой концепции.

Концепция мобильных роботов. Эта концепция по своей сути является гибридом двух рассмотренных ранее концепций. Её разработчики осуществили попытку объединить весь накопленный положительный опыт в рамках концепции роботов и заманчивую идею снижения нагрузки на каналы связи за счет снижения объема передаваемых по ним данных, описывающих содержимое ресурсов. В результате этой «попытки» появилась технология, которая предлагает модернизировать программное обеспечение Веб-сервера таким образом, чтобы оно приобрело способность принимать от специального сервера поисковой системы к себе некоторый код, описывающий поведение робота. В дальнейшем этот код должен будет отработать на принявшем его сервере. Результатом работы робота-кода становится представление о найденных информационных ресурсах на данном Веб-сервере. Это представление уже не содержит в себе такого

большого объема мусора, т.к. его основная чистка в найденных ресурсах теперь происходит не на стороне поисковой системы с роботом, а на самом источнике информации.

Однако и в этой бочке мёда есть большой черпак дегтя. Это можно почувствовать еще на стадии поверхностного ознакомления с данной концепцией. Для этого достаточно задать несколько «безобидных» вопросов. Например, как будет себя вести технология данной концепции по отношению к информационным ресурсам, имеющим способность очень часто «незначительно» менять свое содержимое? К сожалению, в этом вопросе данная концепция солидарна с концепцией роботов, а именно предлагается осуществлять периодический обход всех известных Веб-серверов. После такого ответа вопрос о нагрузке на принимающие к себе код робота Веб-сервера возникает сам собой. А он может усложнить и так тяжелый вопрос, связанный с возможностью внедрения «в массы» данной концепции.

За более подробной информацией можно обратиться к [14, 16].

Заключение

Рассмотрена общая структура поисковых систем для сети Интернет. Предпринята попытка сформулировать полный список общих проблем, с которыми сталкиваются разработчики модуля сбора информации о доступных ресурсах Веб. Опираясь на полученный список, были кратко рассмотрены основные концепции при разработке этого модуля. Попутно были обозначены возможные направления дальнейших исследований.

ЛИТЕРАТУРА

1. Najork M., Wiener J.L. *Breadth-First Search Crawling Yields High-Quality Pages*. – <http://www10.org/cdrom/papers/pdf/p208.pdf>
2. Melnik S., Raghavan S., Yang B., Garcia-Molina H. *Building a Distributed Full-Text Index for the Web*. – <http://www-db.stanford.edu/rsram/pubs/www10/www10paper.pdf>
3. Brandman O., Cho J., Garcia-Molina H., Shivakumar N. *Crawler-Friendly Web Servers*. – <http://rose.cs.ucla.edu/cho/papers/cho-server.pdf>
4. Raghavan S., Garcia-Molina H. *Crawling the Hidden Web*. – <http://dbpubs.stanford.edu/pub/2000-36>
5. Shkapenyuk V., Suel T. *Design and Implementation of a High-Performance Distributed Web Crawler*. – <http://cis.poly.edu/tr/tr-cis-2001-03.pdf>
6. Cho J., Garcia-Molina H., Page L. *Efficient Crawling Through URL Ordering*. – <http://rose.cs.ucla.edu/cho/papers/cho-order.pdf>
7. Menczer F., Pant G., Srinivasan P., Ruiz M.E. *Evaluating Topic-Driven Web Crawlers*. – <http://dollar.biz.uiowa.edu/fil/Papers/sigir-01.pdf>
8. Diligenti M., Coetzee F.M., Lawrence S., Giles C.L., Gori M. *Focused Crawling Using Context Graphs*. – <http://www.neci.nec.com/homepages/coetzee/focusCrawler.pdf>

9. Buyukkokten O., Garcia-Molina H., Paepcke A. *Focused Web Searching with PDAs*.
– <http://www-db.stanford.edu/orkut/papers/pb2.pdf>
10. Fox A., Brewer E.A. *Harvest, Yield, and Scalable Tolerant Systems*.
– <http://www.cs.ucsb.edu/tve/cs290i-sp01/papers/fox99harvest.pdf>
11. Najork M., Heydon A. *High-Performance Web Crawling*.
– <http://citeseer.nj.nec.com/najork01highperformance.html>
12. Green J.W. *HYPERDOG - Up To Date Web Monitoring Through Metacomputers*.
– <http://www.cnds.jhu.edu/pub/papers/hyperdog.pdf>
13. Brewington B.E., Cybenko G. *Keeping up with the changing Web*.
– <http://www.ece.eng.wayne.edu/cz xu/ece7995/reading/keep-up-change.pdf>
14. Fiedler J., Hammer J. *Mobile Web crawling*.
– <http://www.cise.ufl.edu/tech-reports/tech-reports/tr98-abstracts.shtml>
15. Arasu A., Cho J., Garcia-Molina H., Paepcke A., Raghavan S. *Searching the Web*.
– <http://rose.cs.ucla.edu/cho/papers/cho-toit01.pdf>
16. Bowman C. M., Danzig P.B., Hardy D.R., Manber U., Schwartz M.F. *The Harvest Information Discovery and Access System*.
– <http://citeseer.nj.nec.com/bowman95harvest.html>
17. Lam S. *The Overview of Web Search Engines*.
– <http://citeseer.nj.nec.com/lam01overview.html>
18. Rennie J., McCallum A.K. *Using Reinforcement Learning to Spider the Web Efficiently*. – <http://www.ai.mit.edu/jrennie/papers/icml99-text.pdf>
19. Brin S., Page L. *The Anatomy of a Search Engine*.
– <http://www7.scu.edu.au/programme/fullpapers/1921/com1921.htm>
20. Koch T., Ardo A., Brumer B., Lundbr S. *The building and maintenance of robot based internet search services – A review of current indexing and data collection methods*.
– <http://www.ub.lu.se/desire/radar/reports/D3.11v0.3/tot.html>
21. Некрестьянов И.С., Пантелеева Н. *Системы текстового поиска для Веб*.
– <http://meta.math.spbu.ru/nadejda/papers/web-ir/web-ir.html>
22. Некрестьянов И.С. *Тематико-ориентированные методы информационного поиска* // Канд. дис., Санкт-Петербург, 2000.
– <http://meta.math.spbu.ru/igor/thesis/thesis.html>
23. Браславский П.И. *Методы повышения эффективности поиска научной информации (на материале Internet)* // Канд. дис., Екатеринбург, 2000.
24. Земсков И.А. *О концепции индексации информационных ресурсов сети Интернет*. – <ftp://cmm.univer.omsk.su/pub/sbornik8/zemskov.zip>
25. *Сайт обзоров поисковых систем*. – <http://www.searchenginewatch.com/>
26. *Поисковая система Google*. – <http://www.google.com/>
27. *Поисковая система Teoma*. – <http://www.teoma.com/>

ЗАДАЧИ С ЭЛЕМЕНТАМИ ИССЛЕДОВАНИЯ ПРИ ОБУЧЕНИИ ФИЗИКЕ

М.П. Ланкина

In this article is presented one example of estimation task on physics solution as the means of projective and cognitive skills formation on the way of physicist training.

Среди особенностей деятельности преподавателей высшей школы выделяют проектировочные, гностические, организаторские и коммуникативные умения [1]. К проектировочным умениям относят целеполагание в информационном поиске, умения строить умозаключения по аналогии, опираться на интуицию в научном поиске, открывать новые связи элементов проблемы, формулировать гипотезы исследования, находить способы проверки гипотез, планировать научный поиск, сравнивать несходные явления, мысленно представить результат, построить его модель. К гностическим относят умения воспринимать возникновение проблемной ситуации, формулировать проблему, решать ее известным способом, находить новые способы решения проблемы путем выдвижения гипотез, обосновать гипотезу, проверить найденное решение.

Рассмотрев все виды учебных физических задач, можно заметить, что процесс решения оценочных задач наиболее полно включает в себя комплекс перечисленных умений. Кроме того, каждая оценочная задача позволяет диагностировать уровень сформированности целого ряда понятий, представлений о явлениях, усвоения различных физических законов, умений выделять существенные факторы при построении моделей.

Для того чтобы студент-физик мог успешно решать учебные, а на старших курсах — научные задачи, формирование перечисленных умений должно начинаться еще в школе. В частности, формирование умений 1) найти в окружающем мире физическую проблему (в ходе наблюдения или эксперимента); 2) сформулировать условие и требование задачи текстом (качественно); 3) построить модель явления; 4) перевести условие и требование задачи в количественную форму; 5) решить задачу любым доступным обучаемому способом; 6) объяснить результаты, обсудить условия применимости выбранной модели.

В качестве примера рассмотрим задачу-оценку, придерживаясь указанной последовательности действий.

1. Гудит водопроводная труба. Почему? — физическая проблема найдена в ходе наблюдений.

2. Сформулируем задачу — оценить частоту излучаемой звуковой волны.

3. Построим модель явления. Предложим один из возможных механизмов возникновения звука в водопроводных трубах.

При увеличении скорости потока в местах сужений в трубах может возникнуть турбулентность, которая приводит к кавитации — образованию в жидкости разрывов сплошности с появлением полостей (кавитационных пузырьков), заполненных газом, паром или их смесью, в результате местного понижения давления [2]. Кавитация возникает в тех участках потока, где давление понижается до некоторого критического значения $P_{кр}$. Минимумы давления возникают на криволинейных твердых телах, а при наличии сильной завихренности — и во внутренних областях жидкости. При этом присутствующие в жидкости пузырьки газа или пара, двигаясь с потоком жидкости и попадая в области давления $P < P_{кр}$, приобретают способность к росту. После перехода в область, где $P > P_{кр}$, рост пузырька прекращается, и он начинает сокращаться. Сокращение кавитационного пузырька происходит с большой скоростью и сопровождается звуковым импульсом тем более сильным, чем меньше газа содержит пузырек.

Если степень развития кавитации такова, что возникает и схлопывается множество пузырьков, то явление сопровождается сильным шумом со сплошным спектром от нескольких сотен Гц до сотен кГц. Спектр расширяется в область низких частот по мере увеличения максимального радиуса пузырька [2]. Колебания пузырьков усиливаются трубами, а также стенами, полами, потолками, к которым трубы прикреплены [3].

Этот механизм возникновения звука позволяет рассмотреть явление кавитации и целый ряд связанных с ним понятий и законов: ламинарное, турбулентное течения, закон Бернулли, гидравлический удар, звуковая волна, акустический резонанс в трубе, стоячая и бегущая волны и многие другие.

Задача о схлопывании кавитационного пузырька в жидкости впервые была рассмотрена Рэлеем в 1917 году. Он использовал гидродинамическую модель, в которой жидкость считалась идеальной, несжимаемой и безграничной; предполагалось, что внутри пузырька вакуум и пузырек сохраняет сферическую форму в течение всего процесса схлопывания, а давление в жидкости вдали от пузырька (в бесконечности) постоянно.

4, 5. Количественная формулировка задачи и решение. Время схлопывания пузырька тем больше, чем больше его первоначальный радиус a , плотность жидкости ρ и чем меньше давление в жидкости P_0 вдали от пузырька. Пусть в первом приближении от других параметров это время не зависит.

Школьники могут использовать метод размерностей.

$$\tau = a^x \rho^y P_0^z \Rightarrow M^{y+z} L^{x-3y-z} T^{-2z} = T \Rightarrow \begin{cases} y + z = 0 \\ x - 3y - z = 0 \\ -2z = 1 \end{cases} \Rightarrow \begin{cases} z = -\frac{1}{2}; \\ y = \frac{1}{2}; \\ x = 1 \end{cases}$$

$$\tau = a \sqrt{\frac{\rho}{P_0}}, \quad (1)$$

тогда частоту звуковой волны, которая при этом возникает, можно оценить $\nu \sim \tau^{-1}$. Взяв для оценки $P_0 \sim 10^5$ Па, $\rho \sim 10^3$ кг/м³, $a \sim 10^{-2}$ м, получим $\nu \sim 10^3$ Гц, тогда длина звуковой волны $\lambda = c/\nu \approx 1,4$ м, где $c = 1400$ м/с — скорость звука в воде.

Для студентов возможен более сложный вариант математической постановки задачи. Студент может повторить решение задачи Рэлея. Он записывает уравнение движения единицы объема жидкости. Для идеальной жидкости это уравнение Эйлера [4]:

$$\frac{\partial \vec{v}}{\partial t} + (\vec{v} \nabla) \vec{v} = -\frac{1}{\rho} \nabla P. \quad (2)$$

Из несжимаемой жидкости, заполняющей все пространство, внезапно удаляется сферический объем радиуса a . Определить время, в течение которого образовавшаяся полость заполняется жидкостью.

Движение жидкости после образования полости будет центрально-симметричным со скоростями, направленными в каждой точке по радиусу к центру. Уравнение Эйлера в сферических координатах для радиальной составляющей скорости $v_r \equiv v < 0$:

$$\frac{\partial v}{\partial t} + v \frac{\partial v}{\partial r} = -\frac{1}{\rho} \frac{\partial P}{\partial r}. \quad (3)$$

Уравнение непрерывности дает:

$$r^2 v = F(t), \quad (4)$$

где $F(t)$ — произвольная функция времени; (4) выражает тот факт, что в силу несжимаемости жидкости объем, протекающий через сферу любого радиуса, не зависит от этого радиуса. Из (4) следует: $v = F(t)/r^2$, подставим это в (3):

$$\frac{\dot{F}}{r^2} + v \frac{\partial v}{\partial r} = -\frac{1}{\rho} \frac{\partial P}{\partial r}. \quad (5)$$

Проинтегрируем это уравнение по r от ∞ до $R(t)$ при $a \geq R(t)$ — радиуса заполняющейся полости. Интегрирование дает:

$$-\frac{\dot{F}}{R(t)} + \frac{V^2}{2} = \frac{P_0}{\rho}, \quad (6)$$

где $V = \dot{R}$ — скорость изменения радиуса полости. Граничные условия: $V(\infty) = 0$, $P(R) = 0$, $P(\infty) = P_0$. На поверхности полости: $F(t) = R^2(t)V(t)$, подставим в (6):

$$-\frac{R}{2} \frac{d(V^2)}{dR} = \frac{P_0}{\rho} + \frac{3V^2}{2} \quad (7)$$

— уравнение с разделяющимися переменными. Проинтегрируем его при начальном условии $V = 0$ при $R = a$ (в начальный момент жидкость покоилась).

Интегрирование и потенцирование дает

$$\frac{2}{3} \frac{P_0}{\rho} \left(\frac{a}{R} \right)^3 = V^2 + \frac{2}{3} \frac{P_0}{\rho} \implies V = \dot{R} = \sqrt{\frac{2}{3} \frac{P_0}{\rho} \left[\left(\frac{a}{R} \right)^3 - 1 \right]}.$$

Время заполнения полости

$$\tau = \int_0^a \frac{dR}{V} = \sqrt{\frac{3\rho}{2P_0}} \int_0^a dR \left[\left(\frac{a}{R} \right)^3 - 1 \right]^{-\frac{1}{2}}.$$

Интегрирование дает

$$\tau = \sqrt{\frac{3a^2\rho\pi}{2P_0}} \frac{\Gamma(\frac{5}{6})}{\Gamma(\frac{1}{3})} = 0,915a\sqrt{\frac{\rho}{P_0}}. \quad (8)$$

При прежних значениях P_0, ρ и a время $\tau \sim 9,15 \cdot 10^{-4}c$, частота $\nu = \tau^{-1} \sim 1,1 \cdot 10^3 \text{Гц}$, $\lambda = c/\nu = 1,28\text{м}$.

6. Задача Рэлея, конечно, дает заниженную оценку времени схлопывания пузырька. Реальная жидкость имеет вязкость, теплопроводность, сжимаемость; внутри кавитационного пузырька имеется некоторое количество пара и газов; изменение давления, обуславливающее сокращение пузырька, происходит не скачком, а с некоторой конечной скоростью. Эти отклонения от модели Рэлея должны увеличивать время сокращения пузырька.

Усложним нашу модель. Учтем противодействие пара, содержащегося в пузырьке. Жидкость будем считать по-прежнему идеальной, т.е. не учитываем процессы теплопроводности и вязкости. Иначе говоря, движение идеальной жидкости считаем адиабатическим. Значит, и процесс сжатия пара будем считать адиабатическим.

В правой части уравнения (5) произойдет изменение

$$\frac{\dot{F}}{r^2} + v \frac{\partial v}{\partial r} = -\frac{1}{\rho} \frac{\partial(P - P_{\text{пара}})}{\partial r}. \quad (9)$$

Уравнение Пуассона для адиабатического процесса:

$$P_{\text{пара}} = P_{\text{пара}}^0 \left(\frac{a}{r} \right)^{3\gamma},$$

где $P_{\text{пара}}^0$ — давление пара в начальном состоянии. Интегрирование (9) по r с учетом $V = 0$ при $R = a$ дает

$$-\frac{\dot{F}}{R} + \frac{V^2}{2} = \frac{P_0}{\rho} + \frac{P_{\text{пара}}^0}{\rho} \left[\left(\frac{a}{R} \right)^{3\gamma} - 1 \right].$$

Подставим $F(t)$, получим

$$R \frac{d(V^2)}{dR} = -3V^2 - \frac{2P_{\text{пара}}^0}{\rho} \left(\frac{a}{R} \right)^{3\gamma} - 2 \frac{P_0 - P_{\text{пара}}^0}{\rho}.$$

Обозначим $y \equiv V^2; x \equiv R \implies$

$$\frac{dy}{dx} + \frac{3}{x}y = -\frac{2P_{\text{пара}}^0 a^{3\gamma}}{\rho} \frac{1}{x^{3\gamma+1}} - \frac{2(P_0 - P_{\text{пара}}^0)}{\rho x}. \quad (10)$$

Решение этого уравнения:

$$y = \frac{2P_{\text{пара}}^0}{3\rho(\gamma-1)} \left(\frac{a}{x}\right)^{3\gamma} + \left[\frac{2(P_0 - P_{\text{пара}}^0)}{\rho} - \frac{2P_{\text{пара}}^0}{3\rho(\gamma-1)} \right] \left(\frac{a}{x}\right)^3 - \frac{2(P_0 - P_{\text{пара}}^0)}{\rho}. \quad (11)$$

Время схлопывания

$$\tau = \int \frac{dR}{V} = \int_0^a \frac{dx}{\sqrt{y}}. \quad (12)$$

Введем обозначение: $A = 3(\gamma-1)(P_0 - P_{\text{пара}}^0)/P_{\text{пара}}^0$, тогда (12) дает:

$$\tau = \sqrt{\frac{3\rho(\gamma-1)}{2P_{\text{пара}}^0}} a \sum_{n=0}^{\infty} \frac{(-1)^n (2n-1)!!}{(2n)!!} \sum_{k=0}^n \frac{C_n^k (A-1)^{n-k} A^k}{3\gamma/2 + 1 + 3n(\gamma-1) + 3k}. \quad (13)$$

Для оценок возьмем $\rho \sim 10^3 \text{ кг/м}^3$, $P_0 \sim 10^5 \text{ Па}$, $\gamma = 1,32$ для H_2O , $P_{\text{пара}}^0 \sim 10^3 \text{ Па}$, тогда $\tau \sim 1,1 \cdot 10^{-3} \text{ с}$; $\nu \approx 910 \text{ Гц}$; $\lambda \approx 1,54 \text{ м}$.

Следующим шагом усложнения модели может быть учет конечности диаметра трубы. Предполагаем, что центр пузырька находится на оси трубы. В этом случае пузырек можно считать по-прежнему сферически симметричным. Если скорость схлопывания полости V значительно меньше скорости звука в воде, то сжимаемость воды можно не учитывать.

На следующем этапе усложнения модели можно показать несущественность растяжимости труб и влияния поля тяготения на процесс распространения звуковой волны в воде, учесть распределение пузырьков по размерам и т.д.

Модель явления можно усложнять сколь угодно долго, учитывая все новые и новые факторы. Любая оценочная задача дает в этом отношении неисчерпаемые возможности. Поэтому задачи-оценки включаются в процесс подготовки физиков как на пропедевтическом этапе обучения в физико-математических классах, так и на формирующем этапе в университете.

ЛИТЕРАТУРА

1. Есарева З.Ф. *Особенности деятельности преподавателя высшей школы*. Л.: Изд-во ЛГУ, 1974. 112 с.
2. *Физическая энциклопедия. Т. II.* /Гл. ред. А.М. Прохоров. М.: Советская энциклопедия, 1990. 703 с.
3. Уокер Дж. *Физический фейерверк*. М.: Мир, 1979. 288 с.
4. Ландау Л.Д., Лифшиц Е.М. *Теоретическая физика: Учеб. пособ. В 10 т. Т. VI. Гидродинамика*. М.: Наука, 1988. 736 с.

Математические структуры и моделирование

Вып. 9

Сборник научных трудов

Редактор Е.В. Брусницына

Подготовлена к печати ООО «Издательство Наследие. Диалог-Сибирь»

Лицензия ЛР № 071680 от 04.06.98.

Подписано в печать 25.05.2002.

ОП. Формат 60 × 84 1/8. Печ.л. 23,5. Уч.-изд.л. 21,0.

Тираж 140 экз.

Отпечатано в Полиграфическом центре КАН
644050, г. Омск, пр. Мира 32, ком.11, тел. (381-2) 65-47-31
Лицензия ПЛД № 58-47 от 21.04.97 г.