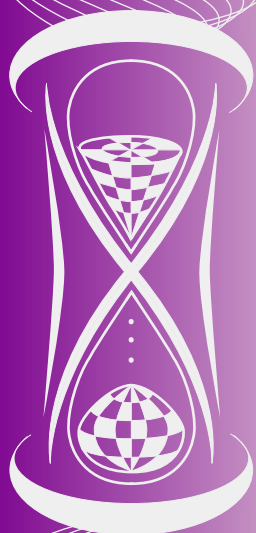


ISSN 2222-8772

МАТЕМАТИЧЕСКИЕ
СТРУКТУРЫ
И
МОДЕЛИРОВАНИЕ

№ 2(34)
2015



**МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ**

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

**МАТЕМАТИЧЕСКИЕ
СТРУКТУРЫ
И
МОДЕЛИРОВАНИЕ**

№ 2(34)

Омск
2015

Математические структуры и моделирование. — Омск : Омский государственный университет, 2015. — № 2(34). — 112 с.

ISSN 2222-8772 (print)

ISSN 2222-8799 (online)

Редакционная коллегия

Н. Ф. Богаченко	технический редактор, канд. физ.-мат. наук, доцент, Омский государственный университет им. Ф.М. Достоевского.
В. Я. Волков	доктор техн. наук, профессор, зав. кафедрой начертательной геометрии, инженерной и машинной графики, Сибирская государственная автомобильно-дорожная академия (СибАДИ).
В. П. Голубятников	доктор физ.-мат. наук, профессор Новосибирского государственного университета, главный научный сотрудник Института математики СО РАН, г. Новосибирск.
С. И. Горлов	доктор физ.-мат. наук, профессор, ректор Нижневартковского государственного университета.
А. Г. Гринь	доктор физ.-мат. наук, профессор, кафедра кибернетики, Омский государственный университет им. Ф.М. Достоевского.
А. К. Гуц	главный редактор, председатель редакционной коллегии, доктор физ.-мат. наук, профессор, зав. кафедрой кибернетики, Омский государственный университет им. Ф.М. Достоевского.
В. А. Еровенко	доктор физ.-мат. наук, профессор, зав. кафедрой общей математики и информатики Белорусского государственного университета, г. Минск, Республика Беларусь.
А. Н. Кабанов	канд. физ.-мат. наук, кафедра кибернетики, Омский государственный университет им. Ф.М. Достоевского.
П. А. Корчагин	доктор техн. наук, профессор, Сибирская государственная автомобильно-дорожная академия (СибАДИ).
Д. Н. Лавров	выпускающий редактор, канд. техн. наук, доцент, зав. каф. компьютерных технологий и сетей, Омский государственный университет им. Ф.М. Достоевского.
A. A. Fedorenko	Ph.D., Researcher (CR1) at the French National Centre of Scientific Research (CNRS) Laboratoire de Physique de l'ENS-Lyon, France.
V. Kreinovich	Ph.D., Professor, Computer Science Department, University of Texas at El Paso, Texas, USA.

Учредитель

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

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

Адрес научной редакции

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

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

МАТЕМАТИЧЕСКИЕ СТРУКТУРЫ И МОДЕЛИРОВАНИЕ

Журнал основан в 1998 году.

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

Все статьи журнала проходят обязательное рецензирование.

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

Электронная версия журнала представлена в сети Интернет по адресам:

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

<http://msm.omsu.ru>

Подписной индекс на второе полугодие по каталогу «Пресса России»: 94082

Электронная почта главного редактора:

guts@omsu.ru

Электронная почта выпускающего редактора:

lavrov@omsu.ru

СОДЕРЖАНИЕ

Фундаментальная математика и физика

А.К. Гуц, Г.Б. Гольдина. *Контингенция порядков в однородных аффинных многообразиях* 5

С.Н. Чуканов. *Определение устойчивости состояния равновесия системы управления динамическим объектом на основе построения оператора гомотопии* 16

О. Kosheleva, V. Kreinovich. *Why We Need Extra Physical Dimensions: A Simple Geometric Explanation* 24

V. Kreinovich, O. Kosheleva. *Towards A Physics-Motivated Small-Velocities Approximation to General Relativity* 29

Прикладная математика и моделирование

О.В. Адмаев, И.В. Ильинский, А.В. Казаков, С.В. Ковтун, Е.О. Смольников, А.М. Ткаченко. *Создание экспертной системы транспортного комплекса на урбанизированных территориях* 39

М.И. Зимин, О.А. Кумукова, С.А. Зими́на. *Оперативный прогноз селевой опасности в неучтенном районе на основе математического моделирования* 45

Ю.Ф. Стругов, А.М. Семенов, С.М. Добровольский, И.А. Батырев. *Стохастическое моделирование каналов с аддитивными и мультипликативными помехами. Схема реализации* . . 48

В.А. Маренко, О.Н. Лучко, В.А. Ляпин, С.Ю. Гуца, Л.В. Алексеенко. *Анализ когнитивных моделей «Артериальная гипертензия»* 64

В.А. Шовин, В.В. Гольяпин. *Методы вращения факторных структур* 75

Продолжение на следующей странице

Наши публикации



V. Kreinovich, O. Kosheleva. *Wiener-Process-Type Evasive Aircraft Actions are Indeed Optimal Against Anti-Aircraft Guns: Wiener's Data Revisited*85

O. Kosheleva, V. Kreinovich. *When an Idea Comes, Write It Down Right Away: Mathematical Justification of Vladimir Smirnov's Advice*90

Компьютерные науки

Д.В. Постарнак. *Метод выбора направления для ускорения классификации с помощью вероятностной нейронной сети адаптивного резонанса*94

Д.В. Постарнак, В.А. Шапцев. *К реализации модифицированной вероятностной нейронной сети адаптивного резонанса*105

КОНТИНГЕНЦИЯ ПОРЯДКОВ В ОДНОРОДНЫХ АФФИННЫХ МНОГООБРАЗИЯХ

А.К. Гуц

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

Г.Б. Гольдина

студентка, e-mail: zoloto13@list.ru

Омский государственный университет им. Ф.М. Достоевского

Аннотация. Доказан слабый вариант теоремы Александра о контингенции для однородных аффинных многообразий.

Ключевые слова: Порядки, аффинная структура, однородное аффинное многообразие, контингенция.

Введение

Исследования связно упорядоченного аффинного пространства A^n , $n \geq 2$, в случае порядка $\mathcal{P}$, инвариантного относительно группы параллельных переносов, привели, благодаря работе А.Д. Александра [1], к полному описанию группы порядковых автоморфизмов $Aut(\mathcal{P})$.

Аналогичный результат был достигнут при изучении порядка, инвариантного относительно основной аффинной группы Ли [2, 3].

Основным инструментом этих исследований была теорема о контингенции, с помощью которой изучение порядковых автоморфизмов сводилось к изучению автоморфизмов порядков, задаваемых конусами.

В данной статье дается доказательство «слабой» теоремы о контингенции для упорядоченных связных односвязных разрешимых групп Ли, снабжённых полной левоинвариантной аффинной структурой.

1. Предпорядок, порядок и порядковые автоморфизмы

Определение 1. Предпорядок на множестве $\mathcal{M}$ — это семейство подмножеств $\mathcal{P} = \{P_x : x \in \mathcal{M}\}$, удовлетворяющее следующим условиям:

- О1) каждой точке $x \in \mathcal{M}$ сопоставлено подмножество $P_x \subset \mathcal{M}$;
- О2) $x \in P_x$ для любой точки $x \in \mathcal{M}$;
- О3) если $y \in P_x$, то $P_y \subset P_x$.

Если $P_x \neq \{x\}$, то предпорядок называется *нетривиальным*. В противном случае — *тривиальным*.

Вводим отношение предпорядка

$$x \preceq y \Leftrightarrow y \in P_x.$$

Полагаем

$$P_x^- = \{y \in \mathcal{M} : y \preceq x\}.$$

Определение 2. Порядок на множестве $\mathcal{M}$ — это семейство подмножеств $\mathcal{P} = \{P_x : x \in \mathcal{M}\}$, удовлетворяющее условиям О1) - О3) определения 1 и дополнительному условию:

О4) если $y \neq x$, то $P_x \neq P_y$.

Определение 3. Биективное отображение $f : \mathcal{M} \rightarrow \mathcal{M}$, удовлетворяющее условию $f(P_x) = P_{f(x)}$ для любой точки $x \in \mathcal{M}$, где $\mathcal{P}$ — предпорядок в $\mathcal{M}$, будем называть *порядковым автоморфизмом*, или *$\mathcal{P}$ -автоморфизмом*.

Из определения следует, что обратное отображение f^{-1} также является $\mathcal{P}$ -автоморфизмом, т.е. $f^{-1}(P_x) = P_{f^{-1}(x)}$ для любой точки $x \in \mathcal{M}$.

Обозначим через $Aut(\mathcal{P})$ множество всех $\mathcal{P}$ -автоморфизмов, а через $Aut_c(\mathcal{P})$ множество всех непрерывных $\mathcal{P}$ -автоморфизмов.

Ясно, $Aut(\mathcal{P})$ является группой относительно операции композиции.

Пусть дан предпорядок $\mathcal{P}$ в топологическом пространстве $\mathcal{M}$.

Говорим, что предпорядок $\mathcal{P}$ — *замкнутый*, если все P_x — замкнутые множества; предпорядок $\mathcal{P}$ — *открытый*, если все множества $P_x \setminus \{x\}$ открыты.

2. Инвариантные порядки в A^n

Пусть на аффинном пространстве A^n действует группа G преобразований этого пространства.

Определение 4. Порядок $\mathcal{P} = \{P_x : x \in A^n\}$, заданный в A^n , называется *G -инвариантным* (или инвариантный относительно группы G), если для любого $x \in A^n$ и любого $g \in G$ $g(P_x) = P_{g(x)}$.

Говорим, что группа G преобразований действует на A^n *транзитивно*, если для любых $x, y \in A^n$ найдётся $g \in G$ такое, что $g(x) = y$. Если для любых $x, y \in A^n$ найдётся единственный элемент $g \in G$ такой, что $g(x) = y$, то группа G действует *эффективно* или *просто транзитивно* на A^n .

В случае порядка G -инвариантного относительно просто транзитивной группы G свойства множеств P_x точно такие же, как у конкретного множества P_e , взятого в выделенной точке $e \in A^n$. Поэтому часто можно говорить о свойствах порядка, говоря только о множестве P_e . В силу того, что $P_x = g(P_e)$, где $x = g(e)$, $g \in G$, видна определяющая роль множества P_e при задании порядка $\mathcal{P}$. Будем в таком случае говорить, что *множество P_e задаёт порядок $\mathcal{P}$* .

Будем также писать P вместо P_e .

Определение 5. Порядок $\mathcal{P} = \{P_x : x \in A^n\}$, заданный в A^n , называется *релятивистским*, если он инвариантен относительно группы параллельных переносов.

2.1. Конусы

Под *конусом* в аффинном пространстве A^n ($n \geq 2$) понимается множество точек K , состоящее из лучей — *образующих конуса* — с началом в точке e , которую называем *вершиной* конуса.

В случае, когда P_x — конус и $P_x \setminus \{x\}$ — открытое множество, говорим об *открытом конусе*.

2.2. Теоремы о контингенции для релятивистских порядков в A^n

Определение 6. Пусть множество $M \subset A^n$. *Контингенцией* $\text{cont}(M, a)$ множества M в точке a называется конус, составленный из всевозможных пределов лучей $l^+(a, x)$, где $x \in M$ при стремлении x к a . Если a не является предельной точкой множества M , то такого конуса нет. Но тогда можно считать, что $\text{cont}(M, a) = \{a\}$ — «нулевой конус».

Нам нужна

Аксиома A_2 . Для любых точек $x, y \in M$ таких, что $y \in P_x$, множество $P_x \cap P_y^-$ ограничено.

Пусть $\mathcal{P} = \{P_x : x \in A^n, n \geq 2\}$ задаёт порядок в аффинном пространстве A^n . Назовём *направленной кривой*, исходящей из точки x , образ полуоси $[0, \infty)$ при непрерывном и монотонном (не убывающем по отношению к порядку на $[0, \infty)$ и порядку $\mathcal{P}$ в A^n) отображении $f : [0, \infty) \rightarrow A^n$, при котором число 0 отображается в x . Очевидно, всякая направленная кривая, исходящая из x , содержится в P_x .

Следовательно,

$$\forall t_1, t_2 \in [0, \infty) \ (t_1 \leq t_2 \Rightarrow f(t_1) \preceq f(t_2) \text{ или } P_{f(t_2)} \subset P_{f(t_1)}).$$

Следующая теорема А.Д. Александрова [1] даёт информацию о «внешнем виде» контингенции:

Теорема 1. Пусть $\mathcal{P}$ задаёт предпорядок в A^n ($n \geq 2$) и $C = \text{cont}(P, e)$. Тогда

- 1) $C \subset \bar{P}$ и C — замкнутый выпуклый конус.
- 2) Если P — замкнутое множество, удовлетворяющее аксиоме A_2 , то C — конус с острой вершиной, совпадающий с объединением S всех направленных кривых, исходящих из точки e .

И, наконец, ответ на вопрос, чем же так хорошо множество $\text{cont}(P_x, x)$, связанное с порядком $\mathcal{P} = \{P_x : x \in A^n\}$, даёт следующая [1]

Теорема 2. Пусть $f : A^n \rightarrow A^n$, $n \geq 2$, — непрерывный $\mathcal{P}$ -автоморфизм, $\mathcal{P}$ — замкнутый предпорядок, удовлетворяющий аксиоме A_2 .

Тогда $f(C_x) = C_{f(x)}$, где $C = \text{cont}(P, e)$, для любой точки $x \in A^n$.

3. Аффинные структуры и аффинные многообразия

Пусть V^n n -мерное дифференцируемое многообразие, где $n \geq 1$.

Определение 7. Аффинный атлас $\mathcal{A}$ на n -мерном многообразии V^n есть совокупность покрывающих V^n локальных карт таких, что любая функция перехода между картами из $\mathcal{A}$ продолжается до аффинного преобразования пространства $\mathbb{R}^n$. Максимальный аффинный атлас есть *аффинная структура* на V^n . Многообразие V^n , оснащённое аффинной структурой, называется n -мерным *аффинным многообразием*.

Каждая локальная карта аффинной структуры определяет *аффинные координаты*.

Отображение $f : V^n \rightarrow V^k$ аффинных многообразий V^n, V^k называется *аффинным*, если оно, будучи выраженным в аффинных координатах, является ограничением аффинного преобразования из $\mathbb{R}^n$ в $\mathbb{R}^k$.

Множество всех аффинных отображений аффинного многообразия V^n обозначим через $\text{Aff}(V^n)$.

На аффинном многообразии V^n имеется естественная линейная связность ∇ с нулевой кривизной и кручением: в аффинных координатах она является стандартной связностью на $\mathbb{R}^n$. Ковариантное дифференцирование относительно ∇ в V^n отвечает обычному дифференцированию в $\mathbb{R}^n$.

Аффинное многообразие *полное*, если оно геодезически полное (относительно связности ∇).

Аффинное многообразие V^n полное тогда и только тогда, когда оно представимо в виде $\mathbb{R}^n/\Gamma$, где Γ подгруппа аффинных преобразований, действующая свободно и вполне разрывно на $\mathbb{R}^n$ [4, с.61].

Геодезические аффинного многообразия $\langle V^n, \mathcal{A} \rangle$ относительно связности ∇ будем называть *прямыми*. Полугеодезическую, исходящую из точки x , именуем *лучом* с началом x .

Очевидно, образ прямой при аффинной биекции будет всегда прямой. Кроме того, на аффинном многообразии через каждую точку в любом направлении можно провести единственную прямую.

3.1. Однородные аффинные многообразия

Аффинное многообразие $\langle V^n, \mathcal{A} \rangle$ будем называть *однородным*, если на нем действует транзитивно аффинная подгруппа $T \subset \text{Aff}(V^n)$. Причём, если T связная группа Ли, действующая просто транзитивно на V^n , то T диффеоморфна V^n . Следовательно, T оснащается аффинной структурой, в которой левые сдвиги являются аффинными биекциями. Другими словами, группа T приобретает левоинвариантную аффинную структуру.

Если G вещественная связная односвязная группа Ли, оснащённая левоинвариантной аффинной структурой, тогда G допускает *аффинное представление*, т.е. существует гомоморфизм

$$\alpha : G \rightarrow \text{Aff}(\mathbb{R}^n).$$

Причём $\alpha(G)$ сохраняет область $D(G) \subset \mathbb{R}^n$ и действует транзитивно на ней [5].

Обратно, если $\dim G = n$ и $\alpha : G \rightarrow \text{Aff}(\mathbb{R}^n)$ аффинное представление, имеющее открытую орбиту, тогда можно оснастить G единственной левоинвариантной аффинной структурой [5].

При изучении вещественных связных односвязных групп Ли G с левоинвариантными аффинными структурами, которые являются полными, полезно иметь в виду, что группа $\alpha(G)$ действует просто транзитивно на $\mathbb{R}^n$. Известно, что при этом G должна быть разрешимой, и можно изучать вместо аффинной геометрии на G фигуры аффинной геометрии на $\mathbb{R}^n$, но инвариантные относительно группы $\alpha(G)$.

Предложение 1. Пусть группа Ли G оснащена полной левоинвариантной структурой. Каждая прямая есть однопараметрическая подгруппа группы Ли G тогда и только тогда, когда каждая однопараметрическая подгруппа есть прямая.

Доказательство. Пусть каждая прямая есть однопараметрическая подгруппа группы Ли G является прямой. Докажем, что каждая однопараметрическая подгруппа есть прямая.

Предположим, что это не верно. Возьмём однопараметрическую подгруппу $\omega_\xi(t)$ с касательным вектором ξ в e . В направлении ξ проведём прямую L . По условию прямая L является однопараметрической подгруппой. Поскольку в одном направлении двух однопараметрических подгрупп быть не может, то получаем, что $\omega_\xi(t) = L$. Получили противоречие с нашим предположением.

Обратно. Пусть каждая однопараметрическая подгруппа есть прямая, но не каждая прямая является однопараметрической подгруппой. Например таковой является прямая L . Пусть эта прямая проходит через e в направлении ξ . Выпустим в направлении ξ однопараметрическую подгруппу $\omega_\xi(t)$. По условию она является прямой. Поскольку в одном направлении двух прямых не бывает, то $L = \omega_\xi(t)$. Получили противоречие с нашим предположением.

Таким образом, предложение 1 доказано. ■

4. Теорема о контингенции

Рассматриваем связную односвязную разрешимую группу Ли G , оснащённую полной левоинвариантной аффинной структурой. Через e обозначаем единицу группы.

Пусть $\mathcal{P} = \{P_x : x \in G\}$ — левоинвариантный порядок на G , т.е.

$$g(P_x) = P_{g \cdot x}.$$

Определение 8. Контингенцией $C = \text{cont}(P_e, e)$ множества P_e в e называется конус, составленный из всевозможных пределов лучей $l^+(e, x)$, исходящих из e и проходящих через x , где $x \in P_e$, при стремлении x к e . Если e не является предельной точкой множества P_e , то такого конуса нет. Но тогда можно считать, что $\text{cont}(P_e, e) = \{e\}$ — «нулевой конус».

Пусть $\mathcal{P} = \{P_x : x \in G, \dim G \geq 2\}$ задаёт предпорядок в группе Ли G . Назовём *направленной кривой*, исходящей из точки e , образ полуоси $[0, \infty)$ при непрерывном и монотонном (не убывающем по отношению к порядку на $[0, \infty)$ и порядку $\mathcal{P}$ в G) отображении $\omega : [0, \infty) \rightarrow G_n$, при котором число 0 отображается в e . Очевидно, всякая направленная кривая, исходящая из e , содержится в P_e .

Следовательно,

$$t_1 \leq t_2 \Rightarrow \omega(t_1) \preceq \omega(t_2) \text{ или } P_{\omega(t_2)} \subset P_{\omega(t_1)}.$$

Вещественная конечномерная группа Ли G , для которой экспоненциальное отображение $\exp : \mathfrak{g} \rightarrow G$, где $\mathfrak{g}$ — алгебра Ли группы G является диффеоморфизмом, называется *экспоненциальной*.

Любая экспоненциальная группа Ли разрешима и односвязна. Всякая вполне разрешимая группа Ли (в частности, нильпотентная группа Ли) экспоненциальна, если она односвязна¹.

Следующая теорема даёт информацию о «внешнем виде» контингенции:

Теорема 3. Пусть G , $\dim G \geq 2$ связная односвязная экспоненциальная разрешимая группа Ли, оснащённая полной левоинвариантной аффинной структурой и каждая прямая которой является однопараметрической подгруппой. Пусть $\mathcal{P}$ задаёт замкнутый предпорядок в G , удовлетворяющий условию (А): существует окрестность U точки e такая, что $P_e \cap P_e^- \cap U = \{e\}$. Тогда

(1) $C = \text{cont}(P_e, e) \subset P_e$ и C — замкнутый выпуклый конус с острой вершиной e .

(2) C содержится в объединении S всех направленных кривых, исходящих из точки e .

Доказательство. (1) Луч контингенции C — это луч l , исходящий из e и содержащийся в C .

Пусть l — луч контингенции. Он является пределом лучей $l_N = l(e, x_N)$, $x_N \in P_e$.

По предложению 1 каждый $l_N = l(e, x_N)$ луч является однопараметрической подполугруппой $\omega_N(t)$, $t \geq 0$, и $x_N = \omega_N(t_0)$. Имеем

$$e \preceq x_N \Rightarrow x_N \preceq x_N \cdot x_N = \omega_N(t_0)\omega_N(t_0) = \omega_N(2t_0) = x_N^2 \subset l_N \cap P_e.$$

Обозначим, $k \cdot x_N \equiv \omega_N(kt_0)$, $k = 1, 2, \dots$. Тогда получаем, что все точки вида $k \cdot x_N$ будут принадлежать лучу $l_N \cap P_e$.

При $x_N \rightarrow e$ точки $k \cdot x_N \in P_e$ сгущаются на лучах l_N , и их пределы дают точки луча l , т.е. каждая точка $x_0 \in l$ луча l есть предел точек множества P_e . Значит, $x_0 \in \overline{P_e}$ и, следовательно, $l \subset P_e$, т.е. $C \subset P_e$.

Докажем это. Берём $x_0 \in l$, и пусть $\omega(t)$ параметризация луча l .

¹См. «Математическую энциклопедию».

Имеем

$$l = \overline{\left\{ \omega \left(\frac{m}{n} \right) : \forall m, n \geq 0 - \text{целые} \right\}}.$$

Пусть U_{x_0} – произвольная окрестность точки x_0 и $x_0 = \omega(t_0)$. Выберем m, n так, что $\omega(m/n) \in U_{x_0}$.

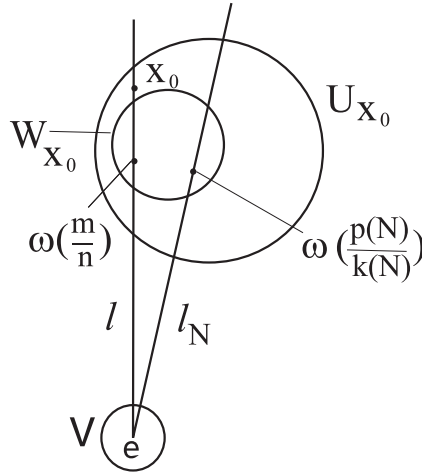


Рис. 1.

Полагаем

$$\omega \left(\frac{m}{n} \right) = \underbrace{\omega \left(\frac{1}{n} \right) \cdot \dots \cdot \omega \left(\frac{1}{n} \right)}_m \equiv m \cdot \omega \left(\frac{1}{n} \right).$$

Так как групповая операция непрерывна, то по окрестности U_{x_0} можно найти окрестность единицы V такую, что

$$m \cdot \omega \left(\frac{1}{n} \right) V \subset U_{x_0}.$$

Поясним это. Пусть

$$a = m \cdot \omega \left(\frac{1}{n} \right).$$

Если дан левый сдвиг $L_a : x \rightarrow a \cdot x$, то $L_a : e \rightarrow a$. Но L_a — это гомеоморфизм на G и, значит, для любой окрестности O_a существует окрестность единицы V такая, что $L_a(V) \subset O_a$. В нашем случае, $a \in U_{x_0}$. Берём $O_a \subset U_{x_0}$. Но тогда

$$L_a(V) \equiv a \cdot V \equiv m \cdot \omega \left(\frac{1}{n} \right) O_a \subset U_{x_0}.$$

Пусть теперь

$$\omega_N(t) - \text{параметризация луча } l_N.$$

Имеем

$$l_N = \overline{\left\{ \omega_N \left(\frac{p}{k} \right) : \forall p, k \geq 0 - \text{целые} \right\}}.$$

Берём окрестность $W_{x_0} \subset U_{x_0}$. Существуют числа $p(N), k(N)$ такие, что

$$\omega_N \left(\frac{p(N)}{k(N)} \right) \in W_{x_0}$$

или

$$p(N) \cdot \omega_N \left(\frac{1}{k(N)} \right) \in W_{x_0} \subset U_{x_0}.$$

Без ограничения общности считаем, что

$$x_N = \omega_N \left(\frac{1}{k(N)} \right).$$

Тогда

$$p(N) \cdot x_N \in W_{x_0} \subset U_{x_0},$$

что мы и доказывали.

Тем самым мы пояснили, что значит «точки $k \cdot x_N$ » сгущаются на лучах l_N , и тем самым доказали часть первого утверждения теоремы 3.

Выпуклость контингенции следует из работы Э.Б. Винберга [6].

Покажем, что конус C имеет острую вершину e . Предположим, что это не верно. Тогда имеется прямая $L \subset C$ и $e \in C$. По доказанному выше, $L \subset P_e$. По предложению 3 прямая L является однопараметрической подгруппой, т.е. $L = \omega(t), t \in \mathbb{R}$. Рассмотрим луч $l = \omega(t), t \geq 0$ (или $t \leq 0$). Это однопараметрическая подполугруппа; для любого $t \geq 0$ $\omega(t) \succeq e$. Следовательно, $\omega(-t)\omega(t) \succeq \omega(-t)$ и $\omega(-t) \preceq e$, т.е. $\omega(-t) \in P_e^-$.

Поскольку знак t при взятии луча не играет роли, то получаем, что $\forall t \omega(t) \in P_e^-$ и, таким образом, $L \subset P_e^-$. Следовательно, $L \subset P_e \cap P_e^-$. Но это противоречит условию (А).

Итак, утверждение (1) теоремы доказано.

(2) Докажем второе утверждение теоремы.

Пусть l – луч контингенции. Покажем, что он является направленной кривой. Надо показать, что если $x, y \in l$, то либо $x \preceq y$, либо $y \preceq x$, либо $x = y$.

Пусть $\omega(t)$, $t \geq 0$ параметризация луча l и $x = \omega(t_1)$, $y = \omega(t_2)$, $t_1 \leq t_2$.

Тогда, раз $l \subset C \subset P_e$, то $x, y \in P_e$

$$x \cdot l = \{\omega(t_1) \cdot \omega(t) : t \geq 0\} = \{\omega(t_1 + t) : t \geq 0\} \subset \{\omega(t) : t \geq 0\} = l.$$

Так как $t_2 = t_1 + \tau$ и $l \subset P_e$, то

$$y \in x \cdot l \subset x \cdot P_e = P_{x \cdot e} = P_x,$$

т.е. $P_y \subset P_x$, или $x \preceq y$. Второе утверждение теоремы 3 доказано.

Теорема 3 доказана. ■

5. Группы Ли, каждая прямая в которых является однопараметрической подгруппой

Приведём примеры экспоненциальных групп Ли, оснащённых аффинной структурой, в которых каждая прямая является однопараметрической подгруппой.

Алгебра Ли $\mathfrak{g}$ является *нильпотентной шага* $k > 0$, где k — целое число, если

$$\mathfrak{b}^{k-1}\mathfrak{g} \neq \{0\}, \quad \mathfrak{b}^k\mathfrak{g} = \{0\},$$

где операция $\mathfrak{b}^k$ определяется посредством индукции:

$$\mathfrak{b}^0\mathfrak{g} = \mathfrak{g}, \quad \mathfrak{b}^k\mathfrak{g} = [\mathfrak{g}, \mathfrak{b}^{k-1}\mathfrak{g}].$$

1. Нильпотентные группы Ли шага 2.

Обозначим $\exp^{-1}(x) = X$, $x \in G$, $X \in \mathfrak{g}$. Если $X = \sum_i x^i X_i$, где X_i — базис в $\mathfrak{g}$, то элементу x ставим в соответствие *координаты 1-го рода*:

$$G \ni x \rightarrow (x^1, \dots, x^n) \in \mathbb{R}^n.$$

Для нильпотентных алгебр Ли шага 2

$$\exp^{-1}(xy) = X * Y = X + Y + \frac{1}{2}[X, Y].$$

Имеем

$$\begin{aligned} X * Y &= \sum_i x^i X_i + \sum_i y^i X_i + \frac{1}{2} \sum_{i,j} x^i y^j [X_i, X_j] = \\ &= \sum_k \left(x^k + y^k + \frac{1}{2} \sum_{i,j} c_{ij}^k x^i y^j \right) X_k. \end{aligned}$$

Поэтому

$$(xy)^k = x^k + y^k + \frac{1}{2} \sum_{i,j} c_{ij}^k x^i y^j.$$

Левый сдвиг $L_a(x) = ax$ записывается в таком случае в виде

$$[L_a(x)]^k = a^k + x^k + \frac{1}{2} \sum_{i,j} c_{ij}^k a^i x^j \quad (1)$$

$$a \rightarrow (a^1, \dots, a^n), \quad x \rightarrow (x^1, \dots, x^n).$$

Видим, что левые сдвиги задаются аффинными преобразованиями (1). Следовательно, имеем аффинную структуру на нильпотентной группе Ли шага 2.

Прямая $\omega(t)$, проходящая через единицу e , в канонических координатах 1-го рода задаётся как

$$[\omega(t)]^k = ta^k.$$

Пусть $Z = \sum_i a^i X_i$ — направляющий вектор данной прямой. Имеем

$$\omega(t_1)\omega(t_2) = (t_1 Z) * (t_2 Z) = t_1 Z + t_2 Z + \frac{1}{2}[t_1 Z, t_2 Z] = (t_1 + t_2)Z = \omega(t_1 + t_2).$$

Это говорит о том, что прямая является однопараметрической подгруппой. По предложению 1 верно и обратное, т.е. однопараметрические подгруппы являются прямыми.

2. Группа Гейзенберга $G_3\text{II}$.

Алгебра Ли $\mathfrak{g}_3\text{II}$ группы Гейзенберга задаётся коммутационными соотношениями:

$$[X_1, X_2] = 0, \quad [X_2, X_3] = X_1, \quad [X_3, X_1] = 0.$$

Имеем

$$b^0 \mathfrak{g} = \mathfrak{g}_3\text{II}, \quad b^1 \mathfrak{g} = \{X_1\}, \\ b^2 \mathfrak{g} = [\mathfrak{g}_3\text{II}, b^1 \mathfrak{g}] = \{[X_1, X_1], [X_2, X_1], [X_3, X_1]\} = \{0\}.$$

Следовательно, алгебра Гейзенберга является нильпотентной шага 2. Поскольку в канонических координатах 1-го рода

$$\begin{aligned} [L_x(y)]^1 &= (xy)^1 = x^1 + y^1 + \frac{1}{2}(x^2 y^3 - x^3 y^2), \\ [L_x(y)]^2 &= (xy)^2 = x^2 + y^2, \\ [L_x(y)]^3 &= (xy)^3 = x^3 + y^3, \end{aligned} \tag{2}$$

то имеем на $G_3\text{II}$ левоинвариантную аффинную структуру.

ЛИТЕРАТУРА

1. Александров А.Д. Отображения упорядоченных пространств // Тр. Математ. ин-та АН СССР. 1972. Т. 128. С. 3–21.
2. Гуц А.К. Отображения упорядоченного пространства Лобачевского // Докл. АН СССР. 1974. Т. 215, № 1. С. 35–37.
3. Гуц А.К. Отображения упорядоченного пространства Лобачевского // Сиб. мат. ж. 1986. Т. 27, № 3. С. 51–67.
4. Вольф Дж. Пространства постоянной кривизны. М. : Наука, 1982.
5. Fried D., Goldman M., Hirsch M. Affine manifolds with nilpotent holonomy // Comment. Math. Helv. 1981. V. 56, № 4. P. 487–523.
6. Винберг Э.Б. Инвариантные выпуклые конусы и упорядочения в группах Ли // Функ. анализ и его прил. 1980. Т. 14, Вып. 1. С. 1–13.

CONTINGENCE OF ORDERS IN HOMOGENEOUS AFFINE MANIFOLDS

A.K. Guts

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

G.B. Goldina

Student, e-mail: zoloto13@list.ru

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

Abstract. The weak version of the Alexandrov's theorem for the homogeneous affine manifolds is proven.

Keywords: Orders, affine structures, homogeneous affine manifolds, contingency.

ОПРЕДЕЛЕНИЕ УСТОЙЧИВОСТИ СОСТОЯНИЯ РАВНОВЕСИЯ СИСТЕМЫ УПРАВЛЕНИЯ ДИНАМИЧЕСКИМ ОБЪЕКТОМ НА ОСНОВЕ ПОСТРОЕНИЯ ОПЕРАТОРА ГОМОТОПИИ

С.Н. Чуканов

профессор, д-р техн. наук, ведущий научный сотрудник, e-mail: ch_sn@mail.ru

Институт математики им. С.Л. Соболева СО РАН

Аннотация. В работе предложены геометрические методы определения устойчивости состояния равновесия динамических систем в критических точках. Метод распространён на системы управления динамическими объектами, задаваемые системой дифференциальных уравнений первого порядка. Формирование управляющих воздействия системы управления приводит к изменению геометрических характеристик системы, что позволяет переводить систему из состояния неустойчивого равновесия в состояние устойчивого равновесия. Состояние устойчивости равновесия определяется методом определения индексов невырожденной критической точки функции Морса, в качестве которой используется потенциальная функция динамической системы.

Ключевые слова: система управления динамическим объектом, устойчивость состояния равновесия динамической системы, функция Морса, оператор гомотопии, теорема Якоби.

1. Введение

Для определения устойчивости состояния равновесия динамических систем в критических точках используются геометрические методы исследования [1, 3, 6]. В работе приводится метод формирования такого метрического тензора Якоби для консервативной динамической системы, что уравнения движения динамической системы совпадают с уравнениями для геодезических траекторий в римановом пространстве с этим тензором.

Метод распространён на системы управления динамическими объектами, задаваемые системой дифференциальных уравнений первого порядка. При этом формирование управляющих воздействия приводит к изменению геометрических характеристик системы, что позволяет переводить систему из состояния неустойчивого равновесия в состояние устойчивого равновесия.

Состояние неустойчивого (устойчивого) равновесия определяется методом определения индексов невырожденной критической точки функции Морса, в

качестве которой используется потенциальная функция динамической системы [4]. В случае системы управления динамическим объектом показано, что индексы функции Морса могут быть приведены к значениям, соответствующим состоянию устойчивого равновесия динамической системы.

Потенциальная функция динамической системы с учётом сигналов системы управления формируется методом построения оператора гомотопии [2, 7, 8].

2. Исследование состояния равновесия консервативной динамической системы

Известно, что уравнения движения динамической системы могут быть получены с использованием вариационного исчисления [3]. Согласно принципу наименьшего действия Гамильтона движения системы определяются траекториями в фазовом пространстве, удовлетворяющими условию:

$$\delta \int_{t_0}^{t_1} L(\mathbf{q}, \dot{\mathbf{q}}) dt = 0, \quad (1)$$

где $\mathbf{q}$ — координаты системы в конфигурационном пространстве Q : $\mathbf{q} \in Q \in \mathbb{R}^n$. Рассмотрим лагранжиан в форме:

$$L = \frac{1}{2} a_{ij}(\mathbf{q}) \dot{q}^i \dot{q}^j - V(\mathbf{q}); i, j = 1, \dots, n, \quad (2)$$

где $a_{ij}(\mathbf{q})$ — тензор, определяющий «кинетическую» энергию $\frac{1}{2} a_{ij}(\mathbf{q}) \dot{q}^i \dot{q}^j$, $V(\mathbf{q})$ — потенциальная функция лагранжиана.

Возможно ли такое конформное преобразование тензора $a_{ij}(\mathbf{q})$, что новый тензор учитывал бы потенциал $V(\mathbf{q})$? Ответ на этот вопрос даёт теорема Якоби [6].

Теорема 1 (Якоби). Для консервативной динамической системы с полной энергией E и лагранжианом (2) можно найти конформное отображение метрического тензора:

$$g_{ij}(\mathbf{q}) = \exp(\phi(\mathbf{q})) \cdot a_{ij}, \quad (3)$$

где функция $\phi(\mathbf{q}) = \ln[2(E - V(\mathbf{q}))]$ такая, что геодезические в римановом пространстве с метрическим тензором g_{ij} — есть траектории динамической системы.

В приложении 1 показано, что можно сформировать метрический тензор:

$$\mathbf{g} = (g_{ij}); g_{ij}(\mathbf{q}) = 2[E - V(\mathbf{q})] a_{ij}(\mathbf{q}) \quad (4)$$

при наличии функциональной зависимости тензора $a_{ij} = a_{ij}(\mathbf{q})$. Таким образом, учёт потенциальной энергии может быть реализован конформным преобразованием метрического тензора. При этом, используя принцип наименьшего

действия Гамильтона, получим уравнения движения динамической системы в виде:

$$\ddot{q}^i + a^{im} \left(\frac{\partial a_{km}}{\partial q^j} - \frac{1}{2} \frac{\partial a_{jk}}{\partial q^m} \right) \dot{q}^j \dot{q}^k + a^{ik} \frac{\partial V}{\partial q^k} = 0. \quad (5)$$

Поставим задачу нахождения такой функции Лагранжа для консервативной динамической системы, определяемой системой дифференциальных уравнений первого порядка: $\dot{\mathbf{q}} = \mathbf{f}(\mathbf{q})$; $\mathbf{q} \in \mathbb{R}^n$; $\mathbf{f}(\cdot) \in \mathbb{R}^n$, что эта система дифференциальных уравнений приводится к динамическим уравнениям в форме (5). Сформируем систему дифференциальных уравнений второго порядка дифференцированием $\dot{\mathbf{q}}$ по времени:

$$\ddot{\mathbf{q}} = \left[\partial \mathbf{f}(\mathbf{q}) / \partial \mathbf{q} \right] \mathbf{f}(\mathbf{q}) = \mathbf{F}(\mathbf{q}) = \partial V(\mathbf{q}) / \partial \mathbf{q}, \quad (6)$$

так как система консервативная, в правой части (10) присутствует только потенциальный вектор [5]. Найдём скалярную потенциальную функцию $V(\mathbf{q})$ формированием оператора гомотопии [7–9]:

$$V(\mathbf{q}) = - \int_0^1 \mathbf{q}^T \mathbf{F}(\lambda \mathbf{q}) \cdot d\lambda. \quad (7)$$

Пусть задана система с лагранжианом (2); точка $\mathbf{q}_0 \in Q$ — есть точка равновесия лагранжиана, если [3]: $\frac{\partial V}{\partial \mathbf{q}}(\mathbf{q}_0) = 0$. Точка $\mathbf{q}_0$ является точкой устойчивого равновесия, если $\mathbf{q}_0$ — локальный максимум потенциальной функции V и индекс невырожденной критической точки функции V , рассматриваемой в качестве функции Морса, $\text{ind}(H_V(\mathbf{q}_0)) = 0$; если $\text{ind}(H_V(\mathbf{q}_0)) > 0$, то $\mathbf{q}_0$ является точкой неустойчивого равновесия (см. Приложение 2).

Пример 1: Рассмотрим динамическую систему:

$$\begin{cases} \dot{q}_1 = \omega q_2; \\ \dot{q}_2 = \omega q_1. \end{cases} \quad (8)$$

Применяя дифференцирование по времени, получим:

$$\begin{cases} \ddot{q}_1 = \omega^2 q_1 = -\partial V / \partial q_1; \\ \ddot{q}_2 = \omega^2 q_2 = -\partial V / \partial q_2. \end{cases}$$

Найдём скалярную потенциальную функцию формированием оператора гомотопии [7–9]: $V(q_1, q_2) = \omega^2 (-q_1^2 - q_2^2)$. Метрический тензор Якоби имеет вид:

$$\begin{aligned} g_{ij}(\mathbf{q}) &= 2 [E + \omega^2 (q_1^2 + q_2^2)] \delta_{ij}; \\ ds^2 &= 2 [E + \omega^2 (q_1^2 + q_2^2)] (dq_1^2 + dq_2^2). \end{aligned} \quad (9)$$

Уравнения движения (8) соответствуют соотношениям (5). Потенциальная функция $V(q_1, q_2)$, рассматриваемая в качестве функции Морса, имеет индекс

невырожденной критической точки $\text{ind}(H_V(\mathbf{q}_0)) = -2$, то есть система находится в неустойчивом состоянии равновесия. $\square$

Покажем, что формирование управляющих воздействий на динамическую систему приводит к конформному отображению метрического тензора $\tilde{g}_{ij}(\mathbf{q}) = e^{\phi(\mathbf{q})} g_{ij}(\mathbf{q})$.

3. Исследование состояния равновесия системы управления нелинейным динамическим объектом

Рассмотрим систему управления гладкой нелинейной динамической системой с обратной связью по состоянию, заданную каноническими соотношениями: $\dot{\mathbf{q}} = \mathbf{f}(\mathbf{q}, \mathbf{u})$, где $\mathbf{q} \in \mathbb{R}^n$ — вектор состояния системы; $\mathbf{u} \in \mathbb{R}^m$ — вектор управляющих сигналов системы; матричная функция $\mathbf{f}(\cdot, \cdot): \{\mathbf{f}(\cdot, \cdot) | \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n\}$. Дифференцируя канонические соотношения, получим:

$$\ddot{\mathbf{q}} = \left[\partial \mathbf{f}(\mathbf{q}) / \partial \mathbf{q} \right] \dot{\mathbf{q}} + \left[\partial \mathbf{f}(\mathbf{q}) / \partial \mathbf{u} \right] \dot{\mathbf{u}} = \partial V_u(\mathbf{q}) / \partial \mathbf{q}. \quad (10)$$

Если существует функциональная зависимость $\mathbf{u} = \mathbf{u}(\mathbf{q})$, то:

$$\ddot{\mathbf{q}} = \left[\partial \mathbf{f}(\mathbf{q}) / \partial \mathbf{q} \right] \dot{\mathbf{q}} + \left[\partial \mathbf{f}(\mathbf{q}) / \partial \mathbf{u} \right] \left(\partial \mathbf{u}(\mathbf{q}) / \partial \mathbf{q} \right) \dot{\mathbf{q}} = \mathbf{F}_u(\mathbf{q}) = \partial V_u(\mathbf{q}) / \partial \mathbf{q}, \quad (11)$$

где потенциальную функцию $V_u(\mathbf{q})$ можно получить методом формирования оператора гомотопии [7–9]: $V_u(\mathbf{q}) = - \int_0^1 \mathbf{q}^T \mathbf{F}_u(\lambda \mathbf{q}) \cdot d\lambda$.

Пример 2: Рассмотрим управление динамической системой:

$$\begin{cases} \dot{q}_1 = \omega q_2 + u_1(\mathbf{q}); \\ \dot{q}_2 = \omega q_1 + u_2(\mathbf{q}). \end{cases} \quad (12)$$

При управляющих сигналах: $u_1 = +kq_2$; $u_2 = -kq_1$, из (12) получим:

$$\begin{cases} \ddot{q}_1 = (\omega^2 - k^2) q_1 = -\partial V / \partial q_1; \\ \ddot{q}_2 = (\omega^2 - k^2) q_2 = -\partial V / \partial q_2. \end{cases} \quad (13)$$

При отсутствии управления ($k = 0$) метрический тензор Якоби имеет вид (9); при формировании управляющих сигналов u_1, u_2 метрический тензор Якоби имеет вид:

$$\begin{aligned} \tilde{g}_{ij}(\mathbf{q}) &= 2 [E + (\omega^2 - k^2) (q_1^2 + q_2^2)] \delta_{ij}; \\ ds^2 &= 2 [E + (\omega^2 - k^2) (q_1^2 + q_2^2)] (dq_1^2 + dq_2^2), \end{aligned} \quad (14)$$

то есть метрический тензор $\tilde{g}_{ij}(\mathbf{q})$ может быть получен из метрического тензора $g_{ij}(\mathbf{q})$ конформным отображением $\tilde{g}_{ij}(\mathbf{q}) = e^{\phi(\mathbf{q})} g_{ij}(\mathbf{q})$, где $\phi(\mathbf{q}) = \ln [-2k^2 (q_1^2 + q_2^2)]$.

Уравнения движения (13) соответствуют соотношению (5) при потенциальной функции:

$$V(q_1, q_2) = (\omega^2 - k^2) (-q_1^2 - q_2^2). \quad (15)$$

Потенциальная функция (15), рассматриваемая в качестве функции Морса, имеет индекс невырожденной критической точки:

$$\text{ind}(H_V(\mathbf{q}_0)) = \begin{cases} 0; & \text{if } k^2 > \omega^2; \\ -2; & \text{if } k^2 < \omega^2; \end{cases}$$

и для обеспечения устойчивости состояния равновесия необходимо выполнить требование: $k > \omega > 0$. $\square$

4. Формирование состояния устойчивого равновесия линейной системы управления

Рассмотрим линейную систему управления с обратной связью по состоянию, заданную каноническими соотношениями:

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}; \mathbf{u} = -\mathbf{K}\mathbf{x}, \quad (16)$$

где $\mathbf{x} \in \mathbb{R}^n$ — вектор состояния системы; $\mathbf{u} \in \mathbb{R}^n$ — вектор управляющих сигналов системы; $\mathbf{A} \in \mathbb{R}^{n \times n}$; $\mathbf{B} \in \mathbb{R}^{n \times m}$; $\mathbf{K} \in \mathbb{R}^{m \times n}$. Дифференцируя эти соотношения, получим:

$$\ddot{\mathbf{x}} = (\mathbf{A} - \mathbf{BK})^2 \mathbf{x}. \quad (17)$$

Для этой системы методом формирования оператора гомотопии [7–9] можно найти такую потенциальную функцию:

$$V(\mathbf{x}) = - \int_0^1 \mathbf{x}^T (\mathbf{A} - \mathbf{BK})^2 \lambda \mathbf{x} \cdot d\lambda = -\frac{1}{2} \mathbf{x}^T (\mathbf{A} - \mathbf{BK})^2 \mathbf{x}, \quad (18)$$

что соотношение (13) может быть переписано в виде: $\ddot{\mathbf{x}} = -\partial V / \partial \mathbf{x}$.

Динамические соотношения $\ddot{\mathbf{x}} = -\partial V / \partial \mathbf{x}$ могут быть получены из уравнения (5) применением метрического тензора Якоби:

$$g_{ij}(\mathbf{x}) = [2E + \mathbf{x}^T (\mathbf{A} - \mathbf{BK})^2 \mathbf{x}] \delta_{ij}. \quad (19)$$

Использование потенциальной функции $V(\mathbf{x}) = -\frac{1}{2} \mathbf{x}^T (\mathbf{A} - \mathbf{BK})^2 \mathbf{x}$ в качестве функции Морса позволяет определить достаточные условия устойчивости состояния равновесия линейной системы управления: $\text{ind}(\mathbf{x}^T (\mathbf{A} - \mathbf{BK})^2 \mathbf{x}) = 0$.

Пример 3: Рассмотрим линейную систему управления с матрицами: $\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$; $\mathbf{B} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$; $\mathbf{K} = \begin{pmatrix} 0 & k \\ -k & 0 \end{pmatrix}$. Тогда при $k < 1$ система находится в неустойчивом состоянии равновесия; например, при $k = 0$ в критической точке $\mathbf{x}_0 = 0$: $2V(\mathbf{x}) = -x_1^2 - x_2^2$; $\text{ind}(H_V(\mathbf{x}_0)) = -2$. Тогда при $k > 0$ система находится в устойчивом состоянии равновесия; например, при $k = 2$ в критической точке $\mathbf{x}_0 = 0$: $2V(\mathbf{x}) = 3x_1^2 + 3x_2^2$; $\text{ind}(H_V(\mathbf{x}_0)) = 0$. При изменении параметра k от 0 до 2 метрический тензор Якоби изменяется от $g_{ii}(\mathbf{x}) = [2E + x_1^2 + x_2^2]$, до $g_{ii}(\mathbf{x}) = [2E - 3x_1^2 - 3x_2^2]$; $i = 1, 2$, и индекс критической точки $\text{ind}(H_V(\mathbf{x}_0))$ от -2 до 0. $\square$

5. Заключение

Рассмотрен метод определения устойчивости состояния равновесия динамических систем в критических точках на основе использования геометрических методов исследования. Приводится метод формирования такого метрического тензора Якоби для консервативной динамической системы, что уравнения движения динамической системы совпадают с уравнениями для геодезических траекторий в римановом пространстве с этим тензором. Метод распространён на системы управления динамическими объектами, задаваемые системой дифференциальных уравнений первого порядка. Формирование управляющих воздействия приводит к изменению геометрических характеристик системы, что позволяет переводить систему из состояния неустойчивого равновесия в состояние устойчивого равновесия.

Состояние устойчивого равновесия динамической системы определяется методом определения индексов невырожденной критической точки функции Морса, в качестве которой используется потенциальная функция динамической системы. В случае системы управления динамическим объектом индексы функции Морса могут быть приведены к значениям, соответствующим состоянию устойчивого равновесия динамической системы.

Приложение 1. Вывод уравнения движения динамической системы

Покажем, что можно сформировать метрический тензор $\mathbf{g} = (g_{ij})$ такой, что учёт потенциальной энергии может быть реализован конформным преобразованием метрического тензора при наличии функциональной зависимости тензора $a_{ij} = a_{ij}(\mathbf{q})$. Используя принцип наименьшего действия Гамильтона, получим уравнения движения динамической системы, исходя из знания метрического тензора $\mathbf{g}$. Рассмотрим консервативную динамическую систему с лагранжианом: $L = \frac{1}{2}a_{ij}(\mathbf{q})\dot{q}^i\dot{q}^j - V(\mathbf{q}) = T(\mathbf{q}, \dot{\mathbf{q}}) - V(\mathbf{q})$ и интегралом движения — полной энергией: $E = T + V$. В этом случае гамильтонов вариационный принцип сводится к принципу наименьшего действия Мопертюи:

$\delta \int_{t_1}^{t_2} 2T dt = 0$, так как $L = 2T - E$. Репараметризуем время $t = t(\tau)$ так, чтобы [6]: $a_{ik}(\mathbf{q}) \left(\frac{\partial q^i}{\partial \tau} \right) \left(\frac{\partial q^k}{\partial \tau} \right) = 1$, откуда:

$d\tau = \sqrt{2W(\mathbf{q})}dt$, и: $\delta \int_{t_1}^{t_2} (2T) dt = \delta \int_{t_0}^{t_1} a_{ik}(\mathbf{q}) \dot{q}^i \dot{q}^k dt = \delta \int_{\tau_0}^{\tau_1} \sqrt{2W(\mathbf{q})} d\tau = 0$, где величина $W(\mathbf{q}) = E - V(\mathbf{q})$ принимает те же значения, что и кинетическая энергия T , но не содержит скоростей в своём выражении. Подставляя $d\tau = \sqrt{2W(\mathbf{q})}dt = \sqrt{a_{ik}dq^i dq^k}$ в вариацию $\delta \int_{\tau_0}^{\tau_1} \sqrt{2W(\mathbf{q})} d\tau$, получим

выражение, которое не зависит от времени параметризации траекторий: $\delta \int_{\mathbf{q}_0}^{\mathbf{q}_1} \sqrt{2W(\mathbf{q})} a_{ik} dq^i dq^k = 0$, где интегрирование проводится по кривой, соеди-

няющей две фиксированные конечные точки: $\mathbf{q}_0$ и $\mathbf{q}_1$. Таким образом, движения, полученные из принципа Гамильтона, удовлетворяют вариационному условию $0 = \delta \int (2T) dt = \delta \int \sqrt{g_{ij} \dot{q}^i \dot{q}^j} dt = \delta \int ds$. Рассмотрим систему с метрическим тензором Якоби: $g_{ij}(\mathbf{q}) = 2W(\mathbf{q}) a_{ij}(\mathbf{q})$, который является конформным отображением метрического тензора $a_{ij}(\mathbf{q})$ и элементом длины дуги $ds^2 = g_{ij} dq^i dq^j = 2W a_{ij} \dot{q}^i \dot{q}^j dt^2 = 4W^2 dt^2$, следовательно, $ds = 2W dt$.

Кривая $\gamma_0 : \mathbb{R} \rightarrow M$ называется геодезической, если векторное поле, определяемое касательным вектором $\dot{\gamma}_0$, параллельно самой кривой, то есть [1, 3]: $\nabla_{\dot{\gamma}_0} \dot{\gamma}_0 = 0$, где $\nabla_{\dot{\gamma}_0}$ ковариантная производная ассоциируемая с g_{ij} . В локальных координатах уравнение геодезической кривой [1, 3]: $d^2 q^i / ds^2 + \Gamma_{jk}^i dq^j / ds dq^k / ds = 0$, где Γ_{jk}^i — коэффициенты Кристоффеля: $\Gamma_{jk}^i = \frac{1}{2} g^{im} (\partial g_{mk} / \partial q^j + \partial g_{mj} / \partial q^k - \partial g_{jk} / \partial q^m)$; $(g^{ij}) = (g_{ij})^{-1}$. С учётом метрического тензора Якоби $g_{ik}(\mathbf{q}) = 2W(\mathbf{q}) a_{ik}(\mathbf{q})$, и: $\frac{\partial g_{ik}}{\partial q^j} = 2W \frac{\partial a_{ik}}{\partial q^j} - 2a_{ik} \frac{\partial W}{\partial q^j}$, получим:

$$\frac{d^2 q^i}{ds^2} = \frac{d}{ds} \left(\frac{dq^i}{ds} \right) = \frac{1}{4W^3} \left(\ddot{q}^i W + \frac{\partial V}{\partial q^j} \dot{q}^i \dot{q}^j \right);$$

$$\Gamma_{jk}^i \frac{dq^j}{ds} \frac{dq^k}{ds} = \frac{1}{4W^2} a^{im} \left(\frac{\partial a_{km}}{\partial q^j} - \frac{1}{2} \frac{\partial a_{jk}}{\partial q^m} \right) \dot{q}^j \dot{q}^k - \frac{1}{4W^3} \frac{\partial V}{\partial q^j} \dot{q}^i \dot{q}^j + \frac{1}{4W^2} a^{ik} \frac{\partial V}{\partial q^k},$$

и динамические уравнения могут быть представлены в виде (см. (5)):

$$\ddot{q}^i + a^{im} \left(\frac{\partial a_{km}}{\partial q^j} - \frac{1}{2} \frac{\partial a_{jk}}{\partial q^m} \right) \dot{q}^j \dot{q}^k + a^{ik} \frac{\partial V}{\partial q^k} = 0.$$

Приложение 2. Функция Морса [4]

Пусть M — m -многообразие и $f : M \rightarrow \mathbb{R}$ — гладкая функция, определённая на M . Точка $\mathbf{q}_0$ является критической точкой f , если $\partial f(\mathbf{q}_0) / \partial q_i = 0; \forall i = 1, \dots, n$, по отношению к локальной системе координат (СК — $q_1, \dots, q_n$) в окрестности $\mathbf{q}_0$. Матрица $H_f(\mathbf{q}_0) = \left(\partial^2 f / \partial q_i \partial q_j (\mathbf{q}_0) \right) \in \mathbb{R}^{n \times n}; \forall i, j = 1, \dots, n$ — гессиан функции f в критической точке $\mathbf{q}_0$. Критическая точка является невырожденной, если $\det [H_f(\mathbf{q}_0)] \neq 0$; функция $f : M \rightarrow \mathbb{R}$ является функцией Морса, если любая критическая точка невырождена. Можно выбрать такую локальную СК в окрестности невырожденной $\mathbf{q}_0$, что представление f в этой СК будет иметь вид: $f = -q_1^2 - \dots - q_\lambda^2 + q_{\lambda+1}^2 + \dots + q_m^2 + f(\mathbf{q}_0)$, где $\mathbf{q}_0$ соответствует началу координат. Индекс невырожденной критической точки $\text{ind}(H_f(\mathbf{q}_0)) = \lambda$ равен числу отрицательных диагональных элементов гессиана $H_f(\mathbf{q}_0)$ после диагонализации. Векторное поле X на координатной окрестности U многообразия M описывается формулой $X = \xi_1 \partial / \partial q_1 + \dots + \xi_n \partial / \partial q_n$, где $\xi_1, \dots, \xi_n$ — функции, определённые в СК U . При выборе $\xi_i = \partial f / \partial q_i; i = 1, \dots, n$ получим градиентное векторное поле функции f : $X_f = \left(\partial f / \partial q_1 \right) \partial / \partial q_1 + \dots + \left(\partial f / \partial q_n \right) \partial / \partial q_n$. Градиентное векторное поле функции Морса в точке $\mathbf{q} \neq \mathbf{q}_0$ имеет вид:

$X_f = -2q_1 \frac{\partial}{\partial q_1} - \dots - 2q_\lambda \frac{\partial}{\partial q_\lambda} + 2q_{\lambda+1} \frac{\partial}{\partial q_{\lambda+1}} + \dots + 2q_n \frac{\partial}{\partial q_n}$, где $\mathbf{q}_0$ соответствует началу координат.

ЛИТЕРАТУРА

1. Dubrovin B.A., Fomenko A.T., Novikov S.P. Modern Geometry – Methods and Applications Part I. The Geometry of Surfaces, Transformation Groups, and Fields. New York : Springer-Verlag, 1992. 468 p.
2. Edelen D.G.B. Applied Exterior Calculus. Courier Corporation, 2005. 472 p.
3. Lewis A.D. Lagrangian mechanics, dynamics, and control. Math 439 course notes. Queen's University, 2002. 260 p.
4. Matsumoto Yu. An introduction to Morse theory. Tokyo : Iwanami Shoten, Publishers, 1997. 219 p.
5. Merkin D.R. Introduction to the Theory of Stability. New York : Springer-Verlag, 1997. 319 p.
6. Pettini M. Geometry and Topology in Hamiltonian Dynamics and Statistical Mechanics. Springer Science+Business Media, LLC. 2007. 452 p.
7. Ульянов Д.В., Чуканов С.Н. Декомпозиция векторного поля динамической системы на основе построения оператора гомотопии // Проблемы управления. 2012. № 6. С. 2–6.
8. Ульянов Д.В., Чуканов С.Н. Формирование инвариантов при визуализации векторных полей на основе построения оператора гомотопии // Компьютерная оптика. 2012. Т. 36, № 4. С. 623–627.
9. Чуканов С.Н. Определение потенциальной компоненты векторного поля системы управления на основе построения оператора гомотопии // Прикладная физика и математика. 2014. № 8. С. 40–46.

DETERMINATION OF EQUILIBRIUM STABILITY FOR DYNAMIC OBJECT CONTROL SYSTEM BY CONSTRUCTING A HOMOTOPY OPERATOR

S.N. Chukanov

Professor, Doctor of Technical Sciences, Leading Researcher, e-mail: ch_sn@mail.ru

Sobolev Institute of Mathematics of the Siberian Branch of the Russian Academy of Sciences

Abstract. The paper introduces geometric methods for determining stability of the equilibrium state of dynamical systems at the critical points. The methods are extended on dynamic object control systems described by a system of differential equations of the first order. Conditioning of control signals for the control system changes the geometric characteristics of the system, and it allows to transfer the system from a state of unstable equilibrium to a state of stable equilibrium. Stability of the equilibrium is determined by finding the indices of a non-degenerate critical point of the Morse function that is, in our case, the potential function of the dynamic system.

Keywords: control system of dynamic object, stability of equilibrium state of dynamic system, Morse function, homotopy operator, Jacobi theorem.

WHY WE NEED EXTRA PHYSICAL DIMENSIONS: A SIMPLE GEOMETRIC EXPLANATION

O. Kosheleva

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

V. Kreinovich

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

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

Abstract. It is known that a consistent description of point-wise particles requires that we add extra physical dimensions to the usual four dimensions of space-time. The need for such dimensions is based on not-very-intuitive complex mathematics. It is therefore desirable to try to come up with a simpler geometric explanation for this phenomenon. In this paper, we provide a simple geometric explanation of why extra physical dimensions are needed.

Keywords: extra physical dimensions, discreteness of physical space-time, geometric explanation, Meyer's Theorem on quadratic forms.

1. Need for Extra Physical Dimensions: Reminder

Problems with the usual 4-dimensional space-time models. In relativistic physics, elementary particles are points in space; see, e.g., [1]. Point-wise character of elementary particles makes many physical quantities infinite. For example, the energy density $\rho(x)$ of the electric field $\vec{E}(x)$ is known to be proportional to $|\vec{E}(x)|^2$, and the electric field of a point-wise particle decreases with the distance r to the particle according to the Coulomb law $|\vec{E}(x)| \sim \frac{1}{r^2}$. Thus, the energy density $\rho(x)$ is proportional to $|\vec{E}(x)|^2 \sim \frac{1}{r^4}$.

The overall energy is equal to the integral $\int \rho(x) dx$ and is, thus, proportional to the integral $\int \frac{1}{r^4} dx$. In polar coordinates, after integrating over angular coordinates, we get

$$I = \int_0^\infty \frac{2\pi \cdot r^2}{r^4} dr = 2\pi \cdot \int_0^\infty \frac{1}{r^2} dr.$$

This integral is equal to

$$I = -2\pi \cdot \frac{1}{r} \Big|_0^\infty = \infty.$$

Similar physically meaningless infinities appear when we compute other quantities related to a point particle [1].

Comment. The above computations use a non-quantum approximation, but similar infinities appear when we take into account quantum effects as well.

Current solution. It turns out that infinities can be avoided if we assume that the space-time has extra dimensions beyond the four usual ones. For example, string theory shows that we can get a consistent physical theory if we assume that the space-time is 10-dimensional; see, e.g., [4].

Remaining challenge. A problem with this solution is that it is heavily mathematical, there is no simple intuitive geometric explanation of why extra dimensions are needed.

Comment. It should be mentioned that:

- while there is no clear geometric explanation of *why extra dimensions are needed*,
- there are simple geometric explanations of *why namely 10* is a good dimension; see, e.g., [5].

What we do in this paper. In this paper, we provide a possible geometric explanation of why extra space-time dimensions are needed.

2. Analysis of the Problem and the Resulting Explanation of Extra Physical Dimension(s)

Natural idea: discrete space-time. The infinities are caused by integration to $r = 0$. Thus, one possible way to avoid infinities is to assume that spatial coordinates – and other quantities – are discrete. This idea is ubiquitous in physics [1]:

- an electric charge cannot take any possible value, it must be proportional to some constant;
- quantum physics started with Planck's hypothesis that energy of light of a given wavelength cannot take any possible value, it must be proportional to some constant (dependent on this frequency), etc.

Resulting description of space-time. Let us apply the discreteness idea to variables that describe space-time geometry, namely,

- to the space-time coordinates $x_1, \dots, x_n$, and
- to the components g_{ij} of the metric tensor that describes the proper time $s(x, x')$ between two points $x = (x_1, \dots, x_n)$ and $x' = (x'_1, \dots, x'_n)$ as follows:

$$s^2(x, x') = \sum_{i=1}^n \sum_{j=1}^n g_{ij} \cdot (x_i - x'_i) \cdot (x_j - x'_j). \quad (1)$$

For space-time coordinates, discreteness means that all the coordinates must be integer multiples of some fixed quantum q_x , i.e., that for every point x and for each coordinate i , we must have $x_i = X_i \cdot q_x$ for some integer X_i . Similarly, for the components of the metric tensor g_{ij} , discreteness means that there exists some fixed quantum q_g for which, for each component g_{ij} , we have $g_{ij} = G_{ij} \cdot q_g$ for some integer G_{ij} .

Under these two discreteness assumptions, the formula (1) that describes the square $s^2(x, x')$ of the proper time between the points $x = (X_1 \cdot q_x, \dots, X_n \cdot q_x)$ and $x' = (X'_1 \cdot q_x, \dots, X'_n \cdot q_x)$ takes the form

$$s^2(x, x') = S^2(X, X') \cdot q_x^2 \cdot q_g, \quad (2)$$

where we denoted

$$S^2(X, X') \stackrel{\text{def}}{=} \sum_{i=1}^n \sum_{j=1}^n G_{ij} \cdot (X_i - X'_i) \cdot (X_j - X'_j). \quad (3)$$

Empirical fact: there are light-like particles. It is a known physical fact that:

- in addition to usual particles like electrons and protons that travel with speeds smaller than the speed of light, and for which, therefore, $s^2(x, x') > 0$ for every two points $x \neq x'$ on the particle's trajectory,
- there also exist “light-like” particles like photons that always travel with the speed of light and for which $s^2(x, x') = 0$ for every two points $x \neq x'$ on the particle's trajectory.

In the continuous space-time, the possibility of light-like particles is mathematically trivial. In the continuous space-time, when each coordinate x_i can take any real value, it is always possible to find pairs of points $x \neq x'$ for which $s^2(x, x') = 0$ – provided, of course, that the matrix g_{ij} is not positive or negative definite, i.e., provided that:

- there exist pairs (x, x') with $s^2(x, x') > 0$, and
- there exist pairs (x, x') with $s^2(x, x') < 0$.

In discrete space-time, the existence of light-like particles is automatically guaranteed only if we have extra physical dimensions. In the discrete space-time model (2)-(3), however, it is not always true that if a quadratic form (3) with integer coefficients G_{ij} attains both positive and negative values, there exist integer values $X_i - X'_i$ for which this form is equal to 0.

Such a general statement is true if and only if we have at least five variables, i.e., if and only if $n \geq 5$. This result was proven by A. Meyer in 1884 [6] and is known as Meyer's Theorem; see, e.g., [1, 7, 8].

Resulting explanation. Thus, to make sure that a discrete space-time is always consistent with the existence of light-like particles, we must assume that the dimension of space-time is at least five.

This explain the need for at least one extra physical dimension – in addition to the usual four dimensions of space-time.

Acknowledgments.

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

REFERENCES

1. A. Borel, "Valued of indefinite quadratic forms at integral points and flows on spaces of lattices", *Bulletin of the American Mathematical Society*, 1995, Vol. 32, No. 2, pp. 184–204.
2. J. W. S. Cassels, *Rational Quadratic Forms*, London Mathematical Society, London, 1978.
3. R. Feynman, R. Leighton, and M. Sands, *The Feynman Lectures on Physics*, Addison Wesley, Boston, Massachusetts, 2005.
4. M. Green, J. H. Schwarz, and E. Witten, *Superstring Theory*, Vols. 1 and 2, Cambridge University Press. Cambridge, Massachusetts, 1987.
5. A. B. Korlyukov and V. Kreinovich, "Equations of physics become consistent if we take measurement uncertainty into consideration", *Reliable Computing*, 1995, Supplement (Extended Abstracts of APIC'95: International Workshop on Applications of Interval Computations, El Paso, TX, Febr. 23–25, 1995), pp. 111–112.
6. A. Meyer, "Mathematische Mittheilungen", *Vierteljahrschrift der Naturforschenden Gesellschaft in Zürich*, 1884, Vol. 29, pp. 209–222.
7. J. Milnor and D. Husemoller, *Symmetric Bilinear Forms*, Springer Verlag, Berlin, Hidelberg, 1973.
8. J.-P. Serre, *A Course in Arithmetic*, Springer Verlag, Berlin, Hidelberg, 1973.

ПОЧЕМУ НЕОБХОДИМЫ ДОПОЛНИТЕЛЬНЫЕ ФИЗИЧЕСКИЕ РАЗМЕРНОСТИ: ПРОСТОЕ ГЕОМЕТРИЧЕСКОЕ ОБЪЯСНЕНИЕ

О. Кошелева

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

В. Крейнович

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

Техасский университет в Эль Пасо, США

Аннотация. Известно, что последовательное описание точечных частиц требует, чтобы мы добавили дополнительные физические размерности к обычным четырём измерениям пространства-времени. Необходимость добавления таких размерностей основана на не очень интуитивной сложной математике. Поэтому желательно придумать более простые геометрические объяснения этому феномену. В данной работе мы предоставляем простое геометрическое объяснение того, почему дополнительные физические размерности необходимы.

Ключевые слова: дополнительные физические размерности, дискретность физического пространства-времени, геометрическое объяснение, теорема Мейера о квадратичных формах.

TOWARDS A PHYSICS-MOTIVATED SMALL-VELOCITIES APPROXIMATION TO GENERAL RELATIVITY

V. Kreinovich

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

O. Kosheleva

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

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

Abstract. In the general case, complex non-linear partial differential equations of General Relativity are very hard to solve. Thus, to solve the corresponding physical problems, usually appropriate approximations are used. The first approximation to General Relativity is, of course, Newton's theory of gravitation. Newton's theory is applicable when the gravitational field is weak and when all velocities are much smaller than the speed of light. Most existing approximations allow higher velocities, but still limit us to weak gravitational fields. In this paper, we consider the possibility of a different approximation, in which strong fields are allowed but velocities are required to be small. We derive the corresponding equations and speculate on their possible physical consequences.

Keywords: General Relativity, small-velocities case, approximation.

1. Formulation of the Problem

Need for approximations to General Relativity. Since its discovery in 1915 and its first experimental confirmation in 1919, General Relativity has been consistently confirmed by more and more accurate experiments [1, 6, 10, 11]. However, its non-linear differential equations are very difficult to solve in the general case. As a result, most observable physical consequences – starting from Einstein's own – were obtained by using appropriate approximations to General Relativity.

In many cases, when a new problem appeared which could not be easily solved by using the existing approximation, a solution was found when a new approximation was developed. Since General Relativity still has many problems which have not yet been fully analyzed, a reasonable idea is to come up with new approximations, with the hope that these approximations will help us to analyze these problems.

It is desirable to come up with physics-motivated approximations. Sometimes, purely mathematical and/or computational approximation ideas are very helpful. However, the experience of physicists shows that, in general, physics-motivated approximations are usually much more successful [1]. From this view-

point, it is desirable to come up with new physics-motivated approximation to General Relativity.

Which approximations are known. General Relativity is, in its essence, the theory of gravitational interactions. From this viewpoint, it is an extension of the original Newton's theory of gravitation. In this sense, the Newton's theory is the first approximation to General Relativity.

Newton's theory adequately describes gravitational effects when the following two conditions are satisfied:

- the gravitational field φ is weak and
- all the velocities v are much smaller than the speed of light c : $v \ll c$.

In the vicinity of the Solar system, the gravitational field is weak, but there are objects – such a light – which travel with the speed which is equal to the speed of light. As a result, while the Newtonian theory of gravitation provides a very accurate description of the movement of celestial bodies, it is not so adequate in describing the gravitational effects on the objects whose velocities are comparable with the speed of light. For example, the Newtonian estimate of the Sun's gravitational effect on the trajectory of light is twice smaller than what General Relativity predicts, and the observational confirmation of this effect in 1919 was the first of many successes of this theory.

From this viewpoint, it has been reasonable to look for a next approximation in which we take into account that the velocities may be comparable with the speed of light. Such a *post-Newtonian* approximation was indeed developed. This effort was started by Einstein himself. This approximation was later supplemented by taking into account not only terms which are *linear* in terms of the (weak) gravity field (as Newton's theory does), but also terms which are *quadratic* in terms of this field [6, 10, 11].

Possibility of an alternative approximation: what we do in this paper. Newton's theory holds under the two above assumptions: that the field is weak and that the velocities are small. Current approximations largely describe what happens when we do not make the second assumption, while keeping the first assumption largely intact.

From the physical viewpoint, it is also natural to try to analyze what will happen if we do not make the first assumption (about weak fields), while keeping the second one (that velocities are small) intact. In this paper, we show what approximation we get when we try to come up with such a small-velocity approximation.

Comment. Some of the results presented in this paper first appeared in the technical report [8].

2. Analysis of the Problem

To come up with a physics-motivated approximation, let us recall the physics behind these equations. To come with a physics-motivated approximation, let us

start not directly with the equations of General Relativity, but first with the physics behind these equations.

Einstein himself came up with the equations of General Relativity from physical ideas like equivalence principle, but he used these ideas more as a heuristic, he did not show how these ideas lead exactly to his equations and not to any other equations; such derivations only appeared much later (see, e.g., [2, 3]).

Historically the first physics-motivated derivation of Einstein's equation was obtained by Gupta [4, 5] (see also [7]) based on a slightly different idea. Specifically, Gupta used another idea behind General Relativity: that the source of the gravitational field g_{ij} is the *total* energy density, including both:

- the energy of the other fields, *and*
- the energy of the gravitational field itself.

The fact that a gravitational field generated by other bodies and fields itself becomes a source of gravity explains the highly non-linear character of Einstein's equations. This idea is what we will use to come up with our small-velocities approximation.

Comment. It should be mentioned that while this idea can be traced back to Einstein himself, it was only in 1954, after the first Gupta's paper, that it was proven that this idea explains the exact equations of General Relativity.

How to describe a small-velocities approximation: a general reminder. In General Relativity, gravitation is described by the metric tensor g_{ij} , so that the proper times ds between the space-time points x^i and $x^i + dx^i$ has the form $ds^2 = \sum_i \sum_j g_{ij} \cdot dx^i \cdot dx^j$. Here, $dx^0 = dt$ is the difference in time, and $dx^a = v_a \cdot dt$ for $a = 1, 2, 3$, where v_a is the a -th component of the velocity v . Thus,

$$ds^2 = dt^2 \cdot \left(g_{00} + 2 \sum_a g_{0a} \cdot v_a + \sum_a \sum_b g_{ab} \cdot v_a \cdot v_b \right).$$

In the small-velocities approximation, when $v_a \ll c$, we can ignore terms proportional to these small velocities, and thus, consider only the g_{00} component of the gravitational field [1, 10].

So, in this approximation, gravity at each space-time point is described by a single scalar g_{00} . In other words, gravity is described by a single scalar field. For simplicity, let us denote this field by φ .

A general description of physical fields in a small-velocity approximation is well known in classical field theory [1, 9]. Such theories are usually formulated in terms of the corresponding minimum action principle $S = \int L dx \rightarrow \min$, where the corresponding Lagrangian has the form $L = L(\varphi, \varphi_{,a} \cdot \varphi_{,a})$, where, as usual, $\varphi_{,a}$ denotes the partial derivative $\frac{\partial \varphi}{\partial x_a}$, and it is implicitly assumed that we sum over

repeated indices, so that, e.g., $\varphi_{,a} \cdot \varphi_{,a}$ means $\sum_{a=1}^3 \varphi_{,a} \cdot \varphi_{,a}$.

In particular, the original Newton's theory corresponds to the Lagrangian

$$L = k \cdot \rho \cdot \varphi - \frac{1}{2} \cdot \varphi_{,a} \cdot \varphi_{,a},$$

where ρ is the overall energy density of all the other fields, and k is the parameter in the differential equations that describe Newton's theory:

$$\nabla^2 \varphi = -k \cdot \rho. \quad (1)$$

Comment. This small-velocity approximation is what is usually called a *non-relativistic* approximation in field theory, but we will try to avoid this term, since its use of the word "relativistic" (meaning here related to Special Relativity) may be confusing in this text, where relativity means General Relativity.

How to derive equations and energy density from the Lagrangian: a brief reminder. Minimization principle means that the (variational) derivative off the Lagrangian L with respect to the field φ should be equal to 0: $\frac{\delta L}{\delta \varphi} = 0$. It is known [1,9] that this leads to the following partial differential equation:

$$\frac{\partial L}{\partial \varphi} - \frac{\partial}{\partial x_a} \left(\frac{\partial L}{\partial \varphi_{,a}} \right) = 0. \quad (2)$$

The energy density of a field with Lagrangian L is described by the formula

$$\rho_L = -L + \varphi_{,a} \frac{\partial L}{\partial \varphi_{,a}}. \quad (3)$$

Now, we are ready to formulate our main idea in precise terms. We want to describe a field whose source includes both the outside energy density ρ and the energy density ρ_L of the field itself, i.e., a field for which

$$\nabla^2 \varphi = -k \cdot \rho - k \cdot \rho_L. \quad (4)$$

Substituting the expression (3) into the formula (4), we get

$$\nabla^2 \varphi = -k \cdot \rho - k \cdot \left[-L + \varphi_{,a} \frac{\partial L}{\partial \varphi_{,a}} \right]. \quad (5)$$

So, we are looking for a Lagrangian L for which the field equations (2) are equivalent to the equation (5).

3. Deriving the Approximate Lagrangian

Let us fix a reasonable class of Lagrangians. To solve our problem, let us consider the following natural generalization of the above Newtonian Lagrangian, namely, a Lagrangian of the type

$$L = a \cdot \rho \cdot a(\varphi) - \frac{1}{2} \cdot b(\varphi) \cdot \varphi_{,a} \cdot \varphi_{,a} \quad (6)$$

for appropriate functions $a(\varphi)$ and $b(\varphi)$.

What our idea means for a Lagrangian of this type. Substituting the expression (6) into the field equations (2), we conclude that

$$k \cdot a' \cdot \rho - \frac{1}{2} \cdot b' \cdot \varphi_{,a} \cdot \varphi_{,a} + \partial_a(b \cdot \varphi_{,a}) = 0, \quad (6)$$

where a' and b' , as usual, denote the derivatives, and ∂_a denotes partial derivative relative to x_a .

Differentiating the product $b \cdot \varphi_{,a}$, we get

$$k \cdot a' \cdot \rho - \frac{1}{2} \cdot b' \cdot \varphi_{,a} \cdot \varphi_{,a} + b' \cdot \varphi_{,a} \cdot \varphi_{,a} + b \cdot \nabla^2 \varphi = 0, \quad (7)$$

i.e., equivalently, that

$$\nabla^2 \varphi = -k \cdot \frac{a'}{b} \cdot \rho + \frac{1}{2} \cdot \frac{b'}{b} \cdot \varphi_{,a} \cdot \varphi_{,a}. \quad (8)$$

On the other hand, substituting the expression (6) into the formula (5), we get

$$\nabla^2 \varphi = -k \cdot \rho - k^2 \cdot a \cdot \rho - \frac{1}{2} \cdot k \cdot b \cdot \varphi_{,a} \cdot \varphi_{,a} + k \cdot b \cdot \varphi_{,a} \cdot \varphi_{,a},$$

i.e.,

$$\nabla^2 \varphi = -k \cdot (1 - k \cdot a) \cdot \rho + \frac{1}{2} \cdot k \cdot b \cdot \varphi_{,a} \cdot \varphi_{,a}. \quad (9)$$

The equations (8) and (9) should be equivalent. Thus, comparing the equations (8) and (9), we conclude that

$$\frac{b'}{b} = k \cdot b \quad (10)$$

and

$$\frac{a'}{b} = 1 - k \cdot a. \quad (11)$$

Finding the dependence $b(\varphi)$. The equation (10) has the form

$$\frac{db}{d\varphi} = k \cdot b.$$

Moving all the terms related to b to one side and all terms related to φ to the other side, we get

$$\frac{db}{b^2} = k \cdot d\varphi.$$

Integrating both sides, we get

$$-\frac{1}{b} = k \cdot \varphi + C,$$

so

$$b(\varphi) = \frac{1}{-C - k \cdot \varphi}.$$

When $k \rightarrow 0$, we should get the Newtonian Lagrangian, with $b = 1$. Thus, $C = -1$, and the above formula takes the form

$$b(\varphi) = \frac{1}{1 - k \cdot \varphi}. \quad (12)$$

Finding the dependence $a(\varphi)$. Substituting this expression into the formula (11), we conclude that

$$\frac{da}{d\varphi} \cdot (1 - k \cdot \varphi) = 1 - k \cdot a.$$

Moving terms containing a and da to one side and terms containing φ and $d\varphi$ to the other side, we conclude that

$$\frac{da}{1 - k \cdot a} = \frac{d\varphi}{1 - k \cdot \varphi}.$$

Multiplying both sides by $-k$, we get

$$\frac{d(1 - k \cdot a)}{1 - k \cdot a} = \frac{d(1 - k \cdot \varphi)}{1 - k \cdot \varphi},$$

so integration leads to

$$\ln(1 - k \cdot a) = \ln(1 - k \cdot \varphi) + c,$$

i.e., to

$$1 - k \cdot a = C \cdot (1 - k \cdot \varphi).$$

When $k \rightarrow 0$, we should get the Newtonian term $a(\varphi) = \varphi$. Thus, we conclude that $C = 1$. So, $1 - k \cdot a = 1 - k \cdot \varphi$, and thus,

$$a(\varphi) = \varphi. \quad (13)$$

The resulting Lagrangian. Substituting the expressions (12) and (13) into the formula (6), we get the following Lagrangian

$$L = k \cdot \rho \cdot \varphi - \frac{1}{2} \cdot \frac{\varphi_{,a} \cdot \varphi_{,a}}{1 - k \cdot \varphi}. \quad (14)$$

This is the Lagrangian that we propose as the desired small-velocities approximation to General Relativity.

4. Analyzing the Approximate Lagrangian: First Attempts

What we do in this section. Our main objective is to provide the desired approximation, so that other researchers will analyze this approximation and be able to use it to solve physical problems.

In this section, we start this analysis by considering the simplest possible case of an empty-space solution.

Taking into account that most of the Universe is practically empty. It is well known that, just like most of the Solar system is practically empty – most of the mass is concentrated in a few practically point-wise celestial bodies – same way the Universe is mostly practically empty.

In the areas where $\rho = 0$, the above Lagrangian has a simplified form

$$L = -\frac{1}{2} \cdot \frac{\varphi_{,a} \cdot \varphi_{,a}}{1 - k \cdot \varphi}. \quad (15)$$

Deriving the resulting solution. The Lagrangian (15) can be simplified if we find a function $\Phi(\varphi)$ for which

$$\Phi'(\varphi) = \frac{1}{\sqrt{1 - k \cdot \varphi}}. \quad (16)$$

For this function, the Lagrangian (15) takes the Newtonian form

$$L = -\frac{1}{2} \cdot \Phi_{,a} \cdot \Phi_{,a}. \quad (17)$$

Integrating the formula (16), we conclude that

$$\Phi(\varphi) = -\frac{2}{k} \cdot \sqrt{1 - k \cdot \varphi} + c. \quad (18)$$

When $k \rightarrow 0$, we should have $\Phi \approx \varphi$. In this limit, we have $\sqrt{1 - k \cdot \varphi} \approx 1 - \frac{k}{2} \cdot \varphi$, so (18) takes the form

$$\Phi \approx -\frac{2}{k} \cdot \left(1 - \frac{k}{2} \cdot \varphi\right) + c = -\frac{2}{k} + \varphi + c,$$

and the requirement that Φ tend to φ implies that $c = \frac{2}{k}$. Thus, that the formula (18) take the form

$$\Phi = \frac{2}{k} \cdot \left(1 - \sqrt{1 - k \cdot \varphi}\right). \quad (19)$$

For the Lagrangian (17), the variational equations $\frac{\delta L}{\delta \varphi} = 0$ take the usual Newtonian form

$$\nabla^2 \Phi = 0.$$

From Newton's theory, we know the general solution to this equation:

$$\Phi = k \cdot \sum_j \frac{m_j}{r_j}, \quad (20)$$

where m_j is the mass of the j -th body and r_j is the distance to this body.

From (19), we conclude that

$$1 - \sqrt{1 - k \cdot \varphi} = \frac{k}{2} \cdot \Phi,$$

hence

$$\begin{aligned} \sqrt{1 - k \cdot \varphi} &= 1 - \frac{k}{2} \cdot \Phi, \\ 1 - k \cdot \varphi &= \left(1 - \frac{k}{2} \cdot \Phi\right)^2 = 1 - k \cdot \Phi + \frac{k^2}{4} \cdot \Phi^2, \end{aligned}$$

and thus,

$$\varphi = \Phi - \frac{k}{4} \cdot \Phi^2. \quad (21)$$

Conclusion: the resulting solution. The resulting solution has the form (21), where Φ has the form (20), i.e., the form

$$\varphi = k \cdot \sum_j \frac{m_j}{r_j} - \frac{k^2}{4} \cdot \left(\sum_j \frac{m_j}{r_j}\right)^2. \quad (22)$$

An unexpected consequence: repulsion replaces attraction when the gravitational field becomes very strong. For a gravitational field generated by a single body of mass m , this formula takes the form

$$\varphi = k \cdot \frac{m}{r} - \frac{k^2}{4} \cdot \frac{m^2}{r^2}. \quad (23)$$

The resulting force is then equal to

$$F = \frac{d\varphi}{dr} = -k \cdot \frac{m}{r^2} + \frac{k^2}{2} \cdot \frac{m^2}{r^3}. \quad (24)$$

At a distance $r = \frac{k \cdot m}{2}$, when the Newtonian field attains a very large value $\frac{1}{2}$, the force (24) turns into 0, and for smaller r (i.e., for stronger fields), the force changes sign, i.e., gravitation becomes a repulsion force instead of the usual attraction force.

In physical terms, such a strong field corresponds to a very small-size object of a dis-proportionally huge mass – an object that, according to General Relativity, is most probably a black hole [10].

Acknowledgments

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

REFERENCES

1. R. Feynman, R. Leighton, and M. Sands, *The Feynman Lectures on Physics*, Addison Wesley, Boston, Massachusetts, 2005.
2. A. M. Finkelstein and V. Kreinovich, "Derivation of Einstein's, Brans-Dicke and other equations from group considerations," In: Y. Choque-Bruhat and T. M. Karade (eds), *On Relativity Theory. Proceedings of the Sir Arthur Eddington Centenary Symposium, Nagpur India 1984*, World Scientific, Singapore, 1985, Vol. 2, pp. 138–146.
3. A. M. Finkelstein, V. Kreinovich, and R. R. Zapatrin, "Fundamental physical equations uniquely determined by their symmetry groups," *Springer Lecture Notes in Mathematics*, 1986, Vol. 1214, pp. 159–170.
4. S. N. Gupta, "Gravitation and electromagnetism", *Physical Review*, 1954, Vol. 96, pp. 1683–1685.
5. S. N. Gupta, "Einstein's and other theories of gravitation", *Reviews of Modern Physics*, 1957, Vol. 29, pp. 334–336.
6. S. Klioner, P. K. Seidelmann, and M. H. Soffel (eds.), *Relativity in Fundamental Astronomy*, Proceedings of IAU Symposium No. 261, Cambridge University Press, Cambridge, UK, 2009
7. V. Kreinovich. "Gupta's derivation of Einstein equations," *Soviet Physics Doklady*, 1975, Vol. 20, No. 5, pp. 341–342.
8. V. Kreinovich, *Newtonian limit of General Relativity and black holes*, Technical Report, Informatika, St. Petersburg, Russia, 1989 (in Russian).
9. L. D. Landau and E. M. Lifschitz, *The Classical Theory of Fields*, Butterworth-Heinemann, Oxford, UK, 1987.
10. C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation*, W. H. Freeman and Company, San Francisco, California, 1973.
11. C. M. Will, *Theory and Experiment in Gravitational Physics*, Cambridge University Press, Cambridge, Massachusetts, 1993.

НА ПУТИ ОТ ФИЗИЧЕСКИ-МОТИВИРОВАННЫХ НИЗКОСКОРОСТНЫХ АППРОКСИМАЦИЙ К ОБЩЕЙ ТЕОРИИ ОТНОСИТЕЛЬНОСТИ

В. Крейнович

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

О. Кошелева

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

Техасский университет в Эль Пасо, США

Аннотация. В общем случае, получить решение сложных нелинейных дифференциальных уравнений общей теории относительности очень трудно. Поэтому для решения соответствующих физических проблем обычно используются подходящие приближения. Первым приближении общей теории относительности является, конечно, теория тяготения Ньютона. Теория Ньютона применима, когда гравитационное поле является слабым и когда все скорости много меньше скорости света. Большинство существующих приближений позволяет работать на более высоких скоростях, но по-прежнему ограничивают нас в слабых гравитационных полях. В данной работе рассматривается возможность получения различных аппроксимаций, которые разрешены в сильных полях, но скорости должны быть небольшие. Мы выведем соответствующие уравнения и обсудим их возможные физические интерпретации.

Ключевые слова: общая теория относительности, случай небольшой скорости, приближение.

СОЗДАНИЕ ЭКСПЕРТНОЙ СИСТЕМЫ ТРАНСПОРТНОГО КОМПЛЕКСА НА УРБАНИЗИРОВАННЫХ ТЕРРИТОРИЯХ

О.В. Адмаев

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

И.В. Ильинский

аспирант, e-mail: aspirantura@krsk.irkups.ru

А.В. Казаков

аспирант, e-mail: akzk83@mail.ru

С.В. Ковтун

соискатель, e-mail: svkovtun@mail.ru

Е.О. Смольников

аспирант, e-mail: 1244-z@mail.ru

А.М. Ткаченко

аспирант, e-mail: ale42053557@yandex.ru

Красноярский институт железнодорожного транспорта — филиал ИРГУПС,
г. Красноярск, Россия

Аннотация. Актуальность решения проблемы снижения уровня загрязнения атмосферы промышленного города определяется, во-первых, назревшей необходимостью решать рассматриваемую задачу – в Красноярске уже возник термин «режим чёрного неба» как характеристика режима неблагоприятных условий, во-вторых, её недостаточной решённостью в настоящее время, в-третьих, потенциальной возможностью её решения. Объектом исследования является атмосфера города Красноярска, предмет исследования – улично-дорожная сеть, промышленные предприятия, объекты инфраструктуры железнодорожного транспорта. Цели работы – выработка рекомендаций для снижения уровня загрязнения атмосферы города Красноярска и создание механизма их реализации. В связи с поставленной целью возникают задачи: оценка уровня выбросов транспорта на улично-дорожной сети, промышленных предприятий. К методам решения поставленных задач относятся: создание экспертной системы, в которой определены такие базы знаний, как химические вещества и различные соединения с описанием их свойств, характеристика городской застройки, улично-дорожной сети, транспортных потоков, других факторов; постановка математической модели, описывающей эмиссию выбросов и последующее её решение. Авторами получен один из результатов – выяснение причин резкого увеличения значений выбросов.

Ключевые слова: Экспертные системы, улично-дорожная сеть, эмиссия транспортного потока, режим черного неба, математическая модель массопереноса.

Экспертные системы (ЭС) применяются в различных областях человеческой деятельности. ЭС могут быть разработаны с расчётом на процесс обучения, т.к.

они уже содержат необходимые знания и способны объяснить процесс своего рассуждения. Кроме того, должны быть включены знания о методах обучения и возможном поведении пользователя. Реальные ЭС содержат большое количество дополнительных блоков, специфичных для каждой предметной области. Главным структурным отличием ЭС от всех других типов программ является наличие базы знаний и, как следствие, способность к обучению и самообучению. Она предназначена для хранения исходных и промежуточных данных решаемой в текущий момент задачи. Функциональная схема ЭС показана на рис. 1

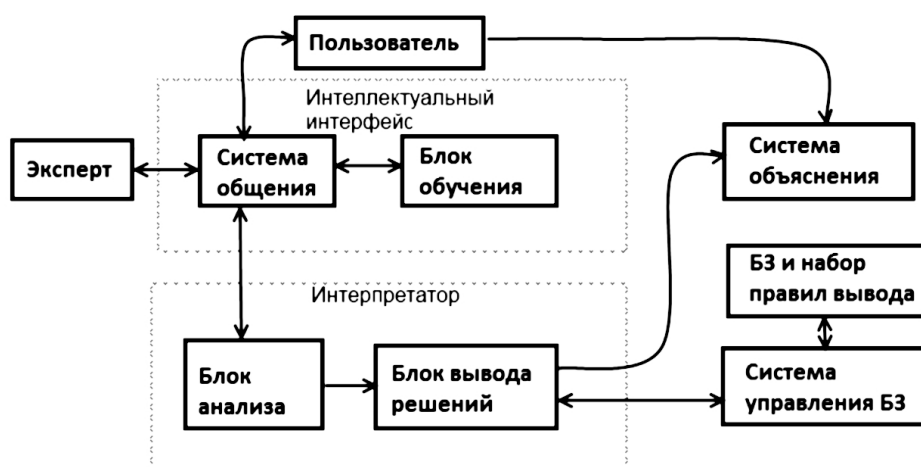


Рис. 1. Функциональная схема экспертной системы

В режиме приобретения знаний общение с ЭС осуществляет эксперт через посредничество инженера знаний. Эксперт описывает предметную область в виде совокупности данных и правил. Данные определяют объекты, их характеристики и значения, существующие в области экспертизы. Правила определяют способы манипулирования данными, характерные для данной предметной области. Рассмотрим элементы базы знаний экспертной системы экологического мониторинга воздушной среды г. Красноярска. Проект «Живой Красноярск» позволяет наблюдать обстановку на городских дорогах как в режиме реального времени, так и в записи (рис.2). На сайте ведётся круглосуточная трансляция с видеокamer, установленных на ключевых магистралях и перекрёстках города. Это позволяет оперативно определять улицы с повышенной загруженностью, количественные оценки транспортного потока во всем городе и, соответственно, оценивать степень экологического неблагополучия его районов (<http://live.krsn.ru>).

Распространение вредных веществ в городской атмосфере представляет собой сложную задачу, так как зависит от очень большого числа параметров: метеорологического режима, рельефа, наличия водных объектов и открытых пространств, плотности и характера застройки и т.д. При расчёте рассеивания



Рис. 2. Пример транспортного потока

выбросов от автотранспорта и определения концентрации токсичных веществ на различном удалении от проезжей части используется модель нормального распределения примесей в атмосфере. Для автомобильных дорог, представляющих собой линейный источник выбросов, распределение рассматривают только в направлении перпендикулярном оси дороги. Если ветер дует со скоростью V под углом ϕ к оси трассы, то выражение для определения концентрации i -го вещества в воздухе определяется по формуле

$$C_i = \frac{2M_i^{\Pi}}{\sqrt{2\pi}\sigma V \sin \phi} + F,$$

где $2M_i^{\Pi}$ — мощность эмиссии, г/(м.с); σ — стандартное отклонение Гауссова рассеивания, зависящее от расстояния от оси дороги, м; V — скорость ветра, м/с; ϕ — угол между направлением ветра и осью трассы (при $\phi < 30^\circ$, $\sin \phi = 0,5$); F — фоновая концентрация загрязнения воздуха, г/м³. Данные по направлению и скорости ветра берутся с метеостанции, расположенной в Центральном районе Красноярска. Необходимая метеорологическая информация поступает в режиме реального времени и доступна на сайте Красноярского отделения Росгидромета <http://meteo.krasnoyarsk.ru>. На рис. 3. представлены метеорологические параметры, зафиксированные во время проведения наблюдений за транспортным потоком и взятия проб воздуха на посту.

Экспертная часть исходных данных может быть получена в результате соответствующей обработки статистических данных, в частности, по методике, изложенной ниже. В качестве элемента базы знаний рассмотрена вероятностная матрица перехода состояний уровня концентрации вредных выбросов в окружающую среду. В г. Красноярске контроль состояния атмосферного воздуха осуществляется территориальным Центром по мониторингу загрязнения

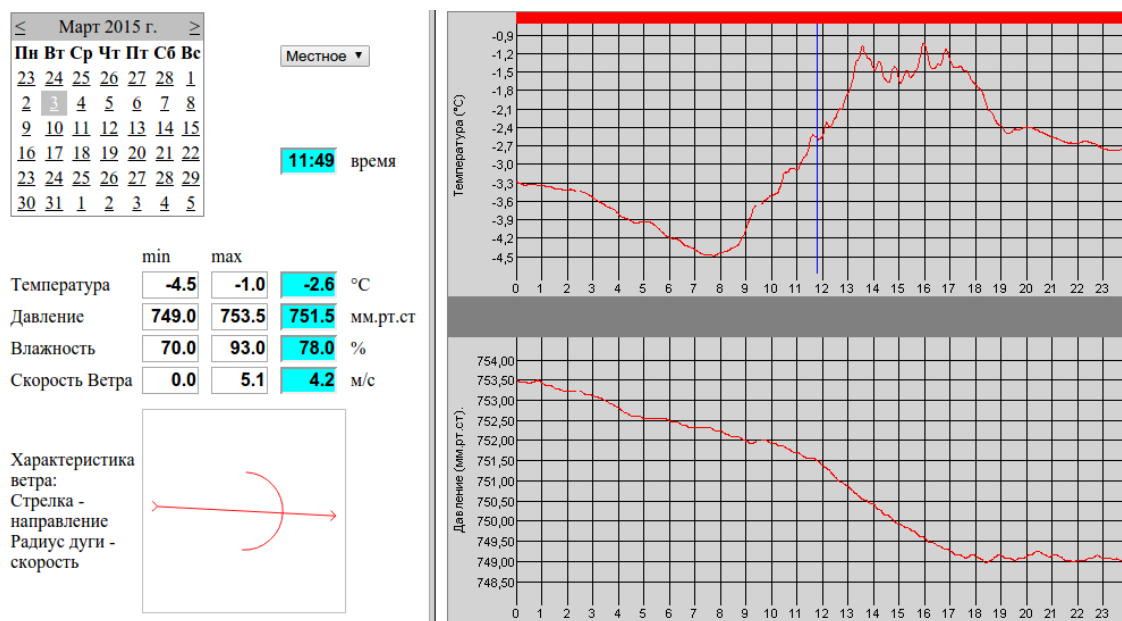


Рис. 3. Данные метеорологической станции Центрального района г. Красноярск

окружающей среды. Наблюдения ведутся в стационарных пунктах, расположенных в разных административных районах города. На рис. 4 приведены еженедельные значения ПДК по диоксиду азота NO_2 . Он был выбран для анализа экологической безопасности воздушного пространства города, так как является токсичным газом. На организм человека он действует как острый раздражитель. Диоксид азота относится ко второму классу опасности и его максимальная разовая величина ПДК составляет $0,085 \text{ мг/м}^3$ («Предельно допустимые концентрации (ПДК) загрязняющих веществ в атмосферном воздухе населённых мест: ГН 2.1.6.695-98. — М.: Российский регистр потенциально опасных химических и биологических веществ Минздрава России, 2000»).

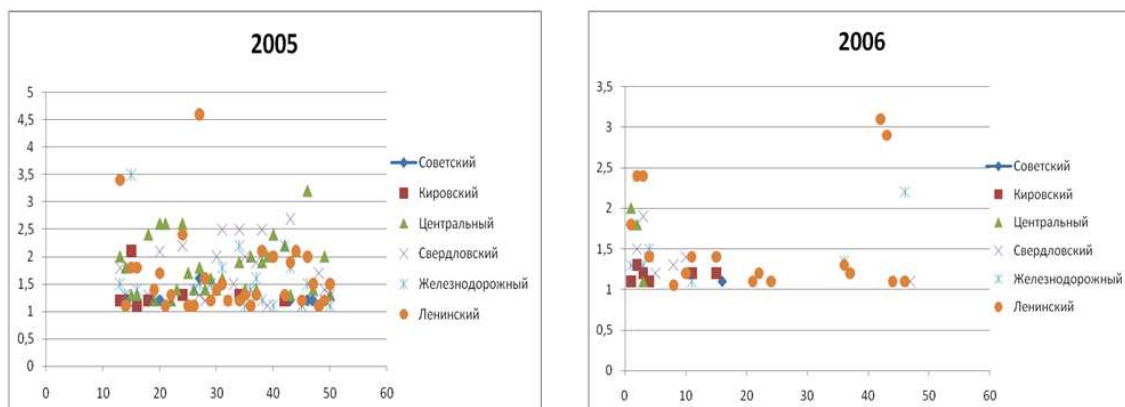


Рис. 4. Данные метеорологической станции Центрального района г. Красноярск

Выводы

Степень достоверности полученных результатов оценивается по общепризнанным критериям — совпадению теоретических и экспериментальных результатов, в данном случае найденное численно рассеивание вредных веществ хорошо коррелирует с распределением эмиссии от транспортных потоков в городе. Результаты исследования сравнивались с авторскими работами других учёных, посвящёнными проблемам загрязнения в схожих геологических и климатических условиях, а также проверенными практикой научными исследованиями для новых строящихся объектов внутри города. Новизна полученных результатов заключается в сочетании учёта транспортных потоков, основанного на рассмотрении случайных процессов, и решении дифференциальных уравнений переноса веществ на территории города. Новое знание (научность результатов) определяется в результате использования экспертной системы в качестве инструмента для выработки рекомендаций по улучшению качества окружающей среды мегаполиса. В работах других авторов, работающих с данной проблемой, в том числе и в Красноярске, данное сочетание почти не рассматривается, соответственно, авторы не имеют возможности моделировать отдельные критические неблагоприятные явления, тем самым, не изучают способы их предотвращения. Практическая значимость работы определяется возможностью увеличения периода времени для возникновения критических неблагоприятных условий с одной стороны и, с другой стороны, возможностью уменьшения продолжительности временного периода существования режима неблагоприятных условий. Указанные эффекты обеспечиваются разработанным программным обеспечением, соответствующими базами знаний, математическими моделями и их решениями при различных определяющих параметрах. Рекомендации по изменению режима работы предприятий, движения и интенсивности транспортных потоков позволят получить желаемый практический эффект, а также связана с использованием в экспертной системе решений нестационарной математической модели при различных параметрах.

CREATION OF EXPERT SYSTEM OF THE TRANSPORT COMPLEX IN THE URBANIZED TERRITORIES

O.V. Admaev

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

I.V. Ilyinskiy

Postgraduate student, e-mail: aspirantura@krsk.irknps.ru

A.V. Kazakov

Postgraduate student, e-mail: akzk83@mail.ru

S.V. Kovtun

Ph.D. doctoral candidate, e-mail: svkovtun@mail.ru

E.O. Smolnikov

Postgraduate student, e-mail: 1244-z@mail.ru

A.M. Tkachenko

Postgraduate student, e-mail: ale42053557@yandex.ru

Krasnoyarsk Institute of Railway Engineering Branch of Irkutsk State Transport
University

Abstract. The problem of reducing the level of atmosphere pollution of an industrial city should be solved nowadays as over the last years the problem has significantly worsened and even spawned such terms as "mode of the black sky" while opportunities for solving it do exist. The object of current research is the atmosphere of the city of Krasnoyarsk. The subjects of the research are the road network, the industrial enterprises and the objects of railway transport infrastructure. The paper purpose is to develop recommendations for reducing pollution of Krasnoyarsk atmosphere and to create a mechanism of their implementation. To fulfill the purpose one should assess the level of emissions generated by transport on the road network and by industrial enterprises. The assessment can be fulfilled via creation of an expert system in which such knowledge bases as chemicals and their compounds with the description of their properties, the characteristics of urban development, the road network, traffic streams and other factors are defined as well as via defining and solving the mathematical model of exhaust emissions.

Keywords: expert systems, road network, issue of a transport stream, mode of the black sky, mathematical model of mass transfer.

ОПЕРАТИВНЫЙ ПРОГНОЗ СЕЛЕВОЙ ОПАСНОСТИ В НЕИЗУЧЕННОМ РАЙОНЕ НА ОСНОВЕ МАТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ

М.И. Зимин²

к.т.н., доктор РАЕН, профессор РАЕН, e-mail: zimin7@yandex.ru

О.А. Кумукова¹

зав. сектором Высокогорного геофизического института, e-mail: kumukova@rambler.ru

С.А. Зимина²

доктор РАЕН, e-mail: zimin7@yandex.ru

¹Высокогорный геофизический институт, Нальчик, Россия

²Торонто, Канада

Аннотация. Рассматривается экспертная диагностическая система для оперативного прогноза селевой опасности в неизученном районе. Используется математическое моделирование, позволяющее оценивать возможность схода селя, основываясь на опорной модели и правилах перехода на новую модель. Приводятся примеры расчёта.

Ключевые слова: селевой поток, предсказание, математическое моделирование, безопасность, мониторинг.

Безопасность эксплуатации объектов является неотъемлемым требованием, предъявляемым к современной инфраструктуре. Она, в частности, обеспечивается проектированием мероприятий и конструкций по их инженерной защите от воздействия опасных природных и природно-техногенных процессов. Одним из таких процессов является сель, представляющий собой грязевой или грязекаменный поток, состоящий из смеси воды и обломков горных пород, внезапно возникающий в бассейнах небольших горных рек [3]. Однако на данный момент отсутствуют вычислительные технологии, учитывающие взаимовлияние и комплексный характер различных факторов, приводящих к их возникновению и развитию, что не позволяет прогнозировать многие ситуации, вызываемые неблагоприятным сочетанием различных компонентов, серьёзно затрудняя инженерные изыскания для проектирования, строительства, реконструкции транспортных сооружений и обоснования размещения объектов транспортной инфраструктуры с учётом требований технической и социальной безопасности. Без адекватных математических моделей проблематично и осуществление мониторинга транспортных природно-технических систем во время их использования в горной местности, а также оценка возможных изменений обстановки в будущем. В то же время, согласно, например, [4] изменчивость метеорологических и сейсмических данных необходимо учитывать с помощью

соответствующих расчётов, и поэтому эти модели весьма востребованы, поскольку явления окружающей среды нередко отличаются многоплановостью и взаимодействием различных параметров, информация о которых может быть весьма неопределённой.

В работе [1] описано применение двухуровневых прогностических систем, имитирующих работу группы экспертов, для оперативного предсказания селей. Используются данные о погребённых льдах, наличии рыхлого материала, осадках, температуре воздуха, биопредвестниках. Приводятся также соответствующие зависимости, позволяющие оценивать селевую опасность в бассейне реки Герхожан-Су.

К недостаткам этого подхода следует отнести необходимость построения отдельных прогностических зависимостей для каждого селевого очага, что достаточно затратно, поскольку это связано с решением систем трансцендентных уравнений. Кроме того, нельзя исключить ситуацию, когда исходных данных для построения математических моделей зарождения и движения селей недостаточно. Это особенно актуально при анализе влияния возможных климатических или экологических изменений на селевую опасность.

В этом случае можно использовать приведённые выше прогностические зависимости в качестве опорных, вводя поправочные параметры. Подобный подход описан в [2]. Ранее он был успешно применён для моделирования средств связи.

Корректируется влияние угла склона, объёма рыхлого материала и объёма погребённого льда. Это было выполнено с помощью математического и компьютерного моделирования.

Как показывает это моделирование, сход селя при угле склона меньше либо равном $3,6^\circ$ нереален. Следовательно, для подобных участков даже при высоком таянии погребённых льдов и сильных осадках селевая опасность отсутствует. В противном случае влияние угла склона учитывается также, как при оценке селевой опасности в бассейне Герхожан-Су.

Объём селя пересчитывается по формуле

$$V_{sc} = V_{sg} \left(\frac{V_{Lrc}}{V_{Lrg}} + \frac{V_{ic}}{V_{ig}} \right),$$

где V_{sc} — объём селя, ожидаемого в заданном очаге, V_{sg} — объём селя, ожидаемого в очаге реки Герхожан-Су при аналогичных условиях, V_{Lrc} — объём рыхлого материала, имеющего размер частиц не более 20 мм, в заданном очаге, V_{Lrg} — объём рыхлого материала, имеющего размер частиц не более 20 мм, в селевом очаге реки Герхожан-Су; V_{ic} — объём погребённого льда в заданном очаге, V_{ig} — объём погребённого льда в селевом очаге реки Герхожан-Су. Если отсутствуют погребённые льды, то сумма среднесуточных температур с момента перехода среднесуточной температурой через 0°C и средняя температура за последние 10 дней не принимаются во внимание. В этом случае не учитывается также влияние сейсмической нагрузки.

В качестве примера можно привести оценку селевой опасности по данным на 17.07.2000 при объёме рыхлого материала в очаге 5000000 м^3 . В этом слу-

чае по прогнозу ожидается сход селей среднего объёма (10–100 тыс. м³), что выглядит правдоподобным, поскольку при аналогичных условиях имел место сход селя в районе г. Тырныауз объёмом несколько сотен тысяч кубометров, но поскольку объём рыхлой породы в рассматриваемом очаге на порядок меньше, то и объём возникшего в нём селя должен быть существенно меньше.

Другим примером является расчёт по предыдущим исходным данным, но при объёме рыхлого материала в очаге 50000 м³. В этом случае ожидается только сель малого объёма, что также логично.

Описанный подход позволяет осуществлять оперативное прогнозирование селей в условиях сильной неопределённости. Это особенно важно в тех случаях, когда сели в каком-либо районе сходят редко, оценка их объёма затруднена, но они всё-таки теоретически возможны.

ЛИТЕРАТУРА

1. Зимин М.И., Зими́на С.А. Прогнозирование воздействий на транспортные сооружения в горной местности на основе бионического подхода // Вестник СибАДИ. 2013. В. 3(31). С. 76–82.
2. Полляк Ю.Г., Филимонов В.А. Статистическое машинное моделирование средств связи. М. : Радио и связь, 1988. 176 с.
3. Сергеев В.С. Безопасность жизнедеятельности, модуль 2. Защита населения и территорий от опасностей в чрезвычайных ситуациях. М. : Российский новый университет, 2010. 89 с.
4. Ontario's 2006 Building Code. V. 1. Toronto : Ministry of Municipal Affairs and Housing, 2008.

OPERATIVE PROGNOSIS OF MUD FLOW RISK IN UNKNOWN AREA WITH THE HELP OF MATHEMATICAL MODELING

M.I. Zimin²

Candidate of Engineering Science, Doctor and Professor of Russian Academy of Natural Sciences, e-mail: zimin7@yandex.ru

O.A. Kumukova¹

Head of the Department of Alpine Geophysical Institute, e-mail: kumukova@rambler.ru

S.A. Zimina²

Doctor of Russian Academy of Natural Sciences, e-mail: zimin7@yandex.ru

¹Alpine Geophysical Institute, Nalchik, Russia

²Toronto, Canada

Abstract. The expert diagnostic system for operative prognosis of mud flow in unknown area is considered. Mathematical modeling allowing to estimate the probability of mud flow utilizing the base model and rules for converting to a new model is used. Examples of analysis are given.

Keywords: mud flow, forecast, mathematical modeling, safety, monitoring.

СТОХАСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ КАНАЛОВ С АДДИТИВНЫМИ И МУЛЬТИПЛИКАТИВНЫМИ ПОМЕХАМИ. СХЕМА РЕАЛИЗАЦИИ

Ю.Ф. Стругов¹

профессор, д.ф.-м.н., зав. каф. «Математический анализ», ИМИТ,
e-mail: strugov77@mail.ru

А.М. Семенов¹

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

С.М. Добровольский¹

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

И.А. Батырев²

ОНИИП, e-mail: b.ivan007@mail.ru

¹Омский государственный университет им. Ф.М. Достоевского

²Омский научно-исследовательский институт приборостроения

Аннотация. В работе представлен краткий обзор и анализ подходов к математическому моделированию многолучевой среды распространения радиоволн УКВ диапазона, рассматривается схема реализации случайных процессов, моделирующих аддитивные и мультипликативные радиопомехи.

Ключевые слова: математические модели радиоканалов, электромагнитные волны, сигнал с многолучевым распространением, доплеровский сдвиг частоты, аддитивные и мультипликативные помехи.

Введение

На практике важной областью моделирования является полунатурное моделирование средств передачи информации (СПИ) на специализированных комплексах или стендах. Для создания подобных комплексов необходимы математические модели радиоканалов и их реализация в имитаторах каналов связи, и именно они в значительной степени определяют эффективность и достоверность результатов полунатурного моделирования.

Основное требование к выбору и обоснованию математической модели — это величина априорной погрешности моделирования, возникающая вследствие неадекватности реализуемой математической модели реальному каналу связи.

В настоящее время существуют различные методы моделирования радиоканалов (например, физическое моделирование, метод записанного канала, метод функционального подобия и другие) [1–9].

Математическим моделированием радиоканалов занимались многие известные учёные: Б.А. Введенский, Д.Д. Кловский, В.В. Марков, А.Г. Самойлов,

Ю.С. Шинаков, Р. Bello, R. Smele, I. Hanzo, В. Sklar и многие другие.

В соответствии с положениями теории распространения радиоволн, электромагнитные волны на входе приёмника представляют собой суперпозицию множества волн, приходящих от антенны передатчика по кратчайшему пути и прибывающих со всех других направлений из-за отражений, дифракций, рассеяний, возникающих при взаимодействии с различными объектами в окружающем пространстве. При этом в искажённом (принятом) сигнале появляются индивидуально различные импульсы. Это эффект многолучевого распространения. Именно поэтому распространение сигнала приближённо считают многолучевым и моделируют принимаемый сигнал как сумму принимаемых сигналов.

Таким образом, сигнал в приёмной антенне содержит сумму волн с различными задержками, амплитудами и фазами. Суперпозиция этих волн приводит к изменению амплитуды и фазы принимаемого сигнала.

Помимо этого, различные компоненты сигнала могут испытывать доплеровский сдвиг частоты различной величины, что обусловлено движением подвижной станции или отражающих объектов. Даже небольшие перемещения на расстояния, соизмеримые с длиной волны передаваемого сигнала, могут вызывать существенные изменения параметров принимаемого сигнала. Различные режимы передачи позволяют получить компромисс между уровнем восприимчивости к межсимвольной интерференции и доплеровскому сдвигу частоты.

Совокупность всех многочисленных обстоятельств приводит к тому, что изменения параметров приходящих радиоволн случайны и интерференционная картина постоянно меняется, то есть является нестационарной. На практике, для упрощения математических расчётов, на определённых временных интервалах процессы замираний (искажений) сигнала полагают локально стационарными.

При моделировании изменчивости напряжённости поля сигнала в радиоканале учитывают потери в тракте, которые характеризуются как зависимость падения средней мощности сигнала от расстояния между передатчиком и приёмником и выделяют следующие виды замираний (искажений):

— быстрые (интерференционные) замирания – это замирания, скорость которых меняется на расстоянии порядка длины волны, что обусловлено в первую очередь изменениями фаз различных компонентов сигнала (и других параметров сигнала во времени), вызванных движением абонента и многолучевым распространением радиоволн;

— медленные замирания – это искажения сигнала, которые происходят на расстояниях большой протяжённости, прежде всего, за счёт изменения уровня потерь из-за затенения окружающими объектами (эти изменения вызываются объектами, такими как здания, холмы, деревья и т.п., оказывающимися на пути сигнала и ограничивающими прямую видимость между передатчиком и приёмником), медленные замирания сигнала можно интерпретировать как изменения медианы (среднего уровня) случайного процесса быстрых замираний, обусловленные изменениями свойств каналов распространения при длительных сеансах связи;

— селективно-частотные замирания [4];

— селективно-временные замирания [4].

Для моделирования быстрых замираний используют различные случайные процессы, параметры которых, как указано выше, выбирают [14-20] в зависимости от характера трассы между источником радиосигнала и приёмником, а также от взаимного расположения передатчика и приёмника. Так функции распределения случайного процесса выбираются соответствующим образом для различных внешних условий. Например, условия могут быть такими:

— передатчик и приёмник находятся в прямой радиовидимости, и дополнительных отражённых сигналов нет;

— передатчик и приёмник находятся в прямой радиовидимости, и есть отражённые сигналы;

— приёмник находится вне зоны прямой радиовидимости от передатчика.

Помехи (шумы естественного и искусственного происхождения), которые накладываются на передаваемый радиосигнал, принято подразделять на аддитивные и мультипликативные.

В настоящее время на практике за основу математических моделей различных замираний берут многопараметрические модели, в которых аддитивные и мультипликативные помехи формируются как случайные процессы с заданными статистическими характеристиками. Часто используются (см. [17]) известные функции распределения Релея, Райса (обобщённое релеевское распределение), бимодальное, усечённо-нормальное, усечённое одностороннее нормальное распределение, распределение Накагами (m -распределение), логарифмически нормальное распределение, комбинация релеевского и логарифмически нормального распределений замираний сигналов (Marvin K.Simon, Mohamed-Slim Alouini, [11]).

Известна модель медленных замираний (Галкин Л.П., Лапин А.Н., Самойлов А.Г., [12]), в которой плотность распределения медианы случайного процесса имеет вид

$$\omega(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - m)^2}{2\sigma^2}},$$

где σ, m — математическое ожидание и дисперсия $\ln \xi$.

С моделированием быстрых замираний можно ознакомиться в работе [1] Волкова А.Н., где распространение сигнала приближённо считается многолучевым, и принимаемый сигнал моделируется как сумма принимаемых сигналов и помех:

$$s(t) = \sum_{i=1}^N A_i(t) u(t-t_i) \cos[(\omega - \Omega_i)(t-t_i) + \varphi_i(t) + \theta(t-t_i)] + n(t) + \sum_{i=1}^{k(t)} \xi_i(t) + p(t).$$

Здесь обозначено:

N — число лучей распространения;

$u(t), \theta(t)$ — параметры передаваемого сигнала;

$A_i(t)$ — модуль коэффициента передачи i -го луча;

t_i — задержка i -го луча (будем считать её постоянной);

Ω_i – доплеровский сдвиг i -го луча;

ω – частота несущих колебаний;

$\varphi_i(t)$ – сдвиг по фазе i -го луча;

ξ_i – импульсные помехи;

$k(t)$ – число импульсных помех ξ_i ;

$p(t)$ – стационарные помехи;

$n(t)$ – белый шум.

В работе [3] проф. Хазана В.Л. для моделирования канала описан алгоритм формирования двумерного, в общем случае, семипараметрического (матожидания двух квадратурных процессов, их дисперсии, коэффициенты автокорреляции и коэффициент взаимной корреляции этих процессов) нормального марковского случайного процесса. Этот алгоритм может быть использован для получения широкого круга законов замираний сигнала, в том числе, часто физически реализующихся в природе и широко используемых исследователями релейского, райсовского и одностороннего усечённого нормального.

Рассмотренный алгоритм формирования двумерного марковского нормального процесса может быть положен в основу программ формирования различных законов замираний сигнала и стационарных помех.

Все феноменологические математические модели дискретного канала связи (то есть реализованные на представлении канала как «черного ящика» с неизвестной внутренней структурой и известными статистическими свойствами выходных данных) совершенно не отражают физики процесса, протекающего в среде распространения, и имитируют результат на выходе КС, основываясь лишь на статистических характеристиках физических процессов. Этим, по мнению Хазана В.Л., объясняется недостаточная адекватность такого рода моделей реальным каналам связи во многих случаях, когда ситуация отличается от штатной. Например, КВ КС отличается многообразием условий распространения сигнала и наличием различного рода аддитивных помех. Классифицировать все возможные ситуации, которые возникают на практике при проведении вычислительных экспериментов, с помощью феноменологических моделей не представляется возможным. Поэтому Хазан В.Л. предложил свою имитационно-аналитическую модель многолучевого дискретного КС с замираниями и стационарными помехами (см. [3]).

Представляет интерес подход к моделированию различных видов замираний реализованный в диссертационной работе Ву Ван Шона [4] (под руководством А.Г. Самойлова). В работе Ву Ван Шона выбрана модель для имитации радиоканала в виде многоотводной линии задержки сигнала с управлением коэффициентами передачи в отводах по законам замираний сигнала и последующим суммированием моделируемых так лучей распространения. Канал моделируется с имитацией быстрых, медленных и частотно-селективных замираний. Предложена структурная схема для построения широкополосного имитатора быстрых и медленных замираний сигнала (с учётом селективно-частотных замираний). Предложена функция распределения как суперпозиция законов распределения быстрых и медленных замираний.

В упомянутых выше работах [1–4] мы видим разные подходы к модели-

рованию каналов связи. Подробную библиографию и существующие подходы к моделированию радиоканалов и разработке их имитаторов можно найти в работах [1–20]. Основной проблемой при выборе математической модели и её практической реализации в имитаторах радиоканалов, является оперативное управление параметрами формируемых в них случайных процессов, необходимых для имитации всех видов замираний. Полное решение её затруднено тем, что при моделировании каналов требуется управление параметрами мультипликативных помех для каждого из моделируемых лучей, а также — параметрами аддитивных помех.

Известные модели каналов связи не являются универсальными и не охватывают всех возможных видов воздействий на сигнал, ибо каждый конкретный канал радиосвязи обладает рядом свойственных ему особенностей.

Математическая модель принимаемого сигнала

Если вещественная форма передаваемого сигнала (амплитудно-фазовая модуляция) имеет вид

$$s(t) = u(t) \cos(\omega t + \theta(t)),$$

то комплексная форма записи сигнала определяется формулой

$$w(t) = u(t)e^{j(\omega t + \theta(t))} = z(t)e^{j\omega t}.$$

Функция $z(t)$ называется комплексной огибающей сигнала.

В данной работе в качестве базовой математической модели для принимаемого сигнала $w(t)$ (в комплексной форме) мы выбрали, близкую к модели в работе [1], следующую модель:

$$w(t) = \sqrt{EK_L} \sum_{i=1}^N \sqrt{p_i w_i} \left[\frac{K_i(t)}{\sqrt{R_i + 1}} + \sqrt{\frac{R_i}{R_i + 1}} D_i(t) \right] z(t - t_i) + n(t) + \lambda(t) + p(t).$$

Здесь E — средняя мощность передаваемого сигнала;

$K_L(f, d, g)$ — коэффициент потери мощности сигнала в тракте (зависит от несущей частоты, расстояния между приёмником и передатчиком, географической сцены);

$p_1, p_2, \dots, p_N$ — реализации значений случайной величины распределённой по лог-нормальному закону (моделируют медленные замирания лучей в тракте);

$w_1, w_2, \dots, w_N$ — веса лучей (моделируют распределение энергии принимаемого сигнала по лучам);

N — количество лучей в канале (предлагаемый диапазон изменения от 1 до 20);

t_i — временные задержки в лучах;

$K_i(t)$ — случайные процессы единичной средней мощности с заданной спектральной плотностью (моделируют доплеровский сдвиг частоты в отражённых лучах);

$D_i(t) = e^{(2f_i\pi t + \nu_i)j}$ — детерминированные сигналы единичной мгновенной мощности (моделируют доплеровский сдвиг в прямых лучах), здесь f_i — доплеровский сдвиг и ν_i — начальная фаза;

R_i — коэффициенты Райса (определяют распределение мощности между прямой и отражённой частями луча);

$n(t)$ — нормированный белый шум (нормирование задаётся параметром модели канала SNR — отношение сигнал/шум);

$\lambda(t)$ — пуассоновский поток реализаций на промежутках времени (своём для каждой реализации) стационарного случайного процесса с заданной средней энергией (моделирует импульсные помехи);

$p(t)$ — стационарный процесс с заданной средней энергией и автокорреляционной функцией (моделирует стационарные помехи).

Эта модель, на наш взгляд, с одной стороны является достаточно адекватной реальности, а с другой стороны, вычислительные проблемы для нее меньше чем у интегральных моделей.

Рассмотрим, прежде всего, каким образом в предложенной модели реализуются случайные процессы, моделирующие белый шум, импульсные и стационарные радиопомехи. Для этого необходимы датчики случайных чисел (ДСЧ).

1. Выбор датчика случайных чисел

В качестве датчика случайных чисел предлагается следующая процедура: выбирая натуральные взаимно простые числа n и m , последовательные «равномерно распределённые» псевдослучайные числа x_k из отрезка $[a, b]$ получаем по формуле

$$x_k = a + \frac{(b-a)((m_0 + km) \bmod n)}{n-1}, \quad k = 1, 2, \dots,$$

где m_0 — начальное состояние датчика. Отметим, что любой моделирующий равномерное на отрезке распределение ДСЧ — это детерминированная последовательность, и его «случайность» заключается в том, что эта последовательность «теоретически» равномерно распределена на этом отрезке: частота появления членов последовательности в произвольно выбранном интервале стремится к длине интервала, делённой на длину исходного отрезка. Реально добиться этого по понятным причинам невозможно, поэтому равномерное распределение заменяют на близкое к нему дискретное распределение, как в предлагаемом ДСЧ. Выбор параметров n и m диктуется, в частности, прогнозируемым объёмом выборки, так как генерируемая последовательность периодична, и необходимо иметь достаточно большой период для гарантии близости распределения полученной последовательности к моделируемому равномерному. Естественно, имеются и другие алгоритмы получения равномерно распределённых последовательностей, но любое усложнение требует хоть какого-то разумного обоснования спецификой задачи. Датчики с различными начальными состояниями можно считать некоррелированными. Более подробная информация по ДСЧ содержится, например, в (Д. Э. Кнут: Искусство программирования, [13]).

2. Реализация белого шума (БШ)

Пусть Δ — фиксированный промежуток времени, c — некоторая константа,

$$x_k = -c + \frac{2c((m_0 + km) \bmod n)}{n-1}, k = 1, 2, \dots,$$

— ДСЧ. Тогда последовательность

$$w_p = \sqrt{\Delta} \sum_{k=1}^p x_k$$

имитирует белый шум с нулевым средним и дисперсией $\sigma_p^2 = \frac{1}{3}c^2p\Delta$. Отметим, что указанные среднее и дисперсия являются асимптотическими характеристиками при неограниченном возрастании объема выборки.

3. Реализация аддитивной импульсной составляющей радиопомехи

В соответствии с выбранной моделью для вещественного сигнала $s(t)$ импульсная составляющая аддитивной радиопомехи представляет собой пуассоновский поток реализаций на промежутках времени (своем для каждой реализации) длины $d = l\Delta$ стационарного случайного процесса со средней энергией E . В первом приближении предлагается для имитации таких случайных процессов использовать отрезки длины l реализации белого шума с соответствующей средней энергией. Для имитации такого случайного процесса рассмотрим два некоррелированных ДСЧ из отрезка $[-1, 1]$: $x_k^{(1)}$ и $x_k^{(2)}$. Фиксируем константу $r < n$ и обозначим $\lambda = \frac{r}{n}$. Число λ будем интерпретировать как вероятность того, что импульс возникнет на следующем за текущим промежутке времени длины Δ .

Пусть датчик $x_k^{(1)}$ порождает белый шум

$$w_p = \sqrt{3E\Delta} \sum_{k=1}^p x_k^{(1)},$$

а датчик $x_k^{(2)}$ порождает последовательность θ_k , где $\theta_k = 1$, если $|x_k^{(2)}| \leq \lambda$ и $\theta_k = 0$, если $|x_k^{(2)}| > \lambda$. Тогда последовательность импульсных помех формируется следующим алгоритмом: если $\theta_k = 1$, то в последовательность импульсных помех вставляется отрезок ряда чисел $w_k, \dots, w_{k+l-1}$. В противном случае полагаем, что соответствующий член в последовательности импульсных помех равен 0.

Что касается составляющей $n(t)$ модели сигнала, то используется модель с соответствующими характеристиками из пункта 2.

4. Реализация стационарного процесса с заданной спектральной плотностью

Рассмотрим стационарный случайный процесс $\xi(t)$, допускающий следующее спектральное представление

$$\xi(t) = \int_{-\infty}^{\infty} e^{j\lambda t} \varphi(\lambda) d\eta(\lambda),$$

где $\eta(\lambda)$ — стандартный белый шум.

Обозначим $S(\lambda) = |\varphi(\lambda)|^2$.

Тогда корреляционная функция процесса

$$R(t) = \int_{-\infty}^{\infty} e^{j\lambda t} S(\lambda) d\lambda$$

и, таким образом, функция $S(\lambda)$ является спектральной плотностью процесса $\xi(t)$. Указанное спектральное представление позволяет реализовать цифровой доплеровский фильтр. Действительно, задавая ступенчатую аппроксимацию спектральной плотности и полагая её равной 0 вне некоторой полосы частот, мы получим представление процесса $\xi(t)$ в виде конечной линейной комбинации комплексных экспонент со случайными коэффициентами, представляющими собой некоррелированные случайные величины с нулевым средним и дисперсиями, равными значениям спектральной плотности на некоторой сетке частот (для случая гауссового шума эти случайные величины нормально распределены). Таким образом, достаточно реализовать набор приращений стандартного белого шума, после чего вычисление значения процесса сводится к некоторому (зависящему от точности аппроксимации спектральной плотности процесса) числу операций сложения и умножения комплексных чисел.

5. Реализация аддитивной стационарной радиопомехи

Стационарные радиопомехи $p(t)$ следует моделировать как стационарный процесс с заданной автокорреляционной функцией $R(t)$. Если аппроксимировать эту функцию тригонометрическим многочленом

$$Re \left(\sum_{k=1}^s f_k e^{j\omega_k t} \right),$$

то соответствующий процесс можно приближённо реализовать как стационарный процесс с дискретным спектром:

$$\sum_{k=1}^s \Phi_k e^{j\omega_k t},$$

где $\Phi_k, k = 1, \dots, s$, — некоррелированные случайные величины с нулевым средним и дисперсиями $D\Phi_k = f_k, k = 1, \dots, s$.

В соответствии с этими замечаниями, осуществив предварительно выбор частот $\omega_1, \omega_2, \dots, \omega_s$ из содержательных соображений (например, известной информации о числе и частотах передачи «мешающих» станций) далее для моделирования случайных коэффициентов используем датчик из п. 1. Возможно также предварительно осуществлять случайный выбор частот из заданного диапазона с помощью ДСЧ, не коррелированного с используемым для коэффициентов.

6. Реализация мультипликативных радиопомех

6.1. Моделирование случайных процессов $K_i(t)$. Доплеровское смещение частоты

Дискретные реализации случайных процессов $K_i(t)$ формируются с помощью доплеровского фильтра из дискретных реализаций стандартного комплексного белого шума. С помощью фильтра у случайных процессов $K_i(t)$ формируют нужную спектральную плотность $S(f)$ (модуль преобразования Фурье автокорреляционной функции). В рекомендациях встречаются 4 основных типа спектральной плотности: спектральная плотность Джексона

$$S(f) = \frac{1}{\pi f_d \sqrt{1 - \left(\frac{f}{f_d}\right)^2}}, |f| < f_d,$$

здесь f_d — максимальное доплеровское смещение. Равномерная

$$S(f) = \frac{1}{2f_d}, |f| \leq f_d,$$

гауссова

$$S(f) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{f^2}{2\sigma^2}},$$

бигауссова

$$S(f) = \frac{1}{C_1 + C_2} \left\{ \frac{C_1}{\sigma_1\sqrt{2\pi}} e^{-\frac{(f-f_1)^2}{2\sigma_1^2}} + \frac{C_2}{\sigma_2\sqrt{2\pi}} e^{-\frac{(f-f_2)^2}{2\sigma_2^2}} \right\}.$$

Профиль задержки мощности и доплеровский спектр можно получить путём зондирования широкополосного канала.

6.2. Моделирование потерь мощности радиоволн при распространении

Для расчёта бюджета канала связи необходимо знать потери мощности при распространении сигнала. Формула потерь при распространении в свободном пространстве с коэффициентами усиления приёмной и передающей антенны G_R и G_T соответственно, расположенных на расстоянии r метров друг от друга, описывается по формуле Фрииса [21]:

$$\frac{P_R}{P_T} = G_T G_R \left(\frac{\lambda}{4\pi r} \right)^2,$$

где r — расстояние в метрах между передающей и принимающей антенной,

P_T — мощность передающей антенны на расстоянии r , в dBm,

P_R — мощность, принимаемая антенной в dBm,

G_T — коэффициент усиления передающей антенны,

G_R — коэффициент усиления принимающей антенны,

λ — длина волны в метрах.

Из данной формулы можно получить выражение для потерь при распространении в dB:

$$L = 10 \cdot \lg \frac{P_R}{P_T} = 10 \cdot \lg G_T + 10 \cdot \lg G_R + 10 \cdot \lg \left(\frac{\lambda}{4\pi r} \right)^2 = 10 \cdot \lg G_T + 10 \cdot \lg G_R + 20 \cdot \lg \frac{c}{4\pi r f},$$

зная, что $20 \cdot \lg \frac{c}{4\pi} = 147,56 \text{ dB}$,

$$L = 10 \cdot \lg G_T + 10 \cdot \lg G_R - 20 \cdot \lg f - 20 \cdot \lg r + 147,56 \text{ dB}.$$

Данная модель является эталонной при анализе распространения радиоволн на различных трассах. В рамках этой модели энергия сигнала зависит только от расстояния между передатчиком и приёмником и убывает обратно пропорционально квадрату расстояния. При наличии препятствий трасса разделяется на расстояние прямой видимости d_0 — расстояние до первого препятствия и $d = r - d_0$ — расстояние не прямой видимости, при этом формула потерь примет вид:

$$L = 10 \cdot \lg G_T + 10 \cdot \lg G_R - 20 \cdot \lg f - 20 \cdot \lg d_0 + 10 \cdot \lg (d_0/d)^n + 147,56 \text{ dB},$$

где n — показатель степени, $3,5 \leq n \leq 5$, [22], зависящий от условий распространения.

Потери мощности в dB подробно описаны в [18–19]. Например, в [18] даётся формула для основных потерь в децибелах, зависящая от физических свойств трассы.

В [19] потери представлены в виде таблиц. Для использования требуется перевод таблиц в электронный вид и интерполяция по расстоянию и частоте.

6.3. Моделирование медленных замираний

Медленные замирания моделируются с помощью нормировки случайных процессов. Нормировка определяется заданием средней энергии лучей. Средняя энергия лучей определяется средней энергией передаваемого сигнала E , коэффициентом потери в тракте K_L , распределением энергии сигнала по лучам $w_1, w_2, \dots, w_N$ (веса лучей: $w_1 + w_2 + \dots + w_N = 1$, $w_i > 0$) и реализациями значений $(\xi_1, \xi_2, \dots, \xi_N)$ случайной величины ξ , распределённой по логнормальному закону.

7. Моделирование временных задержек сигнала в лучах (интерполяция)

Общий подход заключается в применении специальных фильтров к дискретному сигналу, таких как фильтры Котельникова, Фарроу и других. Фильтр Котельникова даёт наилучшие результаты по точности восстановления временных сдвигов сигнала, но достаточно затратен по времени вычислений. Альтернативой фильтру Котельникова являются линейный или фильтр Фарроу (полиномиальный 3-го порядка). Данные фильтры дают меньшую точность восстановления сигнала на частотах близких к частоте Найквиста, но дают значительное преимущество по скорости вычислений. Для сравнения точности и скорости вычислений был проведён численный эксперимент.

Исходными данными для проведения эксперимента являлись: шаг дискретизации сигнала T , количество отсчётов сигнала N , параметр окна в фильтре Котельникова L ($2L + 1$ — ширина окна для свёртки) и соответствующие отсчёты $x(kT)$, $k = 1, 2, \dots, N$ сигналов вида $x(t) = \sin(2\pi ft + \varphi)$, $f = f_1, f_2, \dots, f_m$. Частоты f_k задавались количеством отсчётов n_k на один период, т.е.

$$f_k = \left[\frac{1}{T n_k} \right].$$

Для временного сдвига dt , не равного целому числу шагов дискретизации, вычислены значения сигналов в моменты времени $mT - dt$, $m = 1, 2, \dots, N$. Вычисления проводились с помощью следующих фильтров: Котельникова, линейного и Фарроу. Для каждого фильтра и каждой частоты вычислялись среднее абсолютное отклонение от точных значений и время вычисления (см. рис. 1). Для проведения эксперимента программно реализованы и оптимизированы по скорости выполнения указанные выше фильтры.

Заданы следующие параметры эксперимента:

$$T = 5 \cdot 10^{-5}; N = 128; dt = 1,6T; L = 32.$$

Фильтр Котельникова

Пусть $x_n = x(nT)$. По теореме Котельникова

$$x(mT - dt) = \sum_{n=1}^N x_n \operatorname{sinc} \left(m - n - \frac{dt}{T} \right) \approx \sum_{n=p}^s x_n \operatorname{sinc} \left(m - n - \frac{dt}{T} \right),$$

где

$$p = \max(m - L, 1), s = \min(m + L, N),$$

и

$$\operatorname{sinc}(t) = \frac{\sin \pi t}{\pi t}, \quad t \neq 0.$$

Для ускорения работы разработан алгоритм, суть которого в том, что значения функции sinc вычисляются не $N(2L + 1)$, а только $2L + 1$ раз. Но и это занимает $\frac{4}{5}$ времени работы алгоритма.

Линейный фильтр

$$x(mT - dt) = (1 - r)x_{m-n} + rx_{m-n-1},$$

где n и r — целая и дробная части числа $\frac{dt}{T}$.

Фильтр Фарроу 3-го порядка

$$\begin{aligned} x(mT - dt) = & \left(\frac{r^3}{6} - \frac{r}{6} \right) x_{m-n-2} + \left(-\frac{r^3}{2} + \frac{r^2}{2} + r \right) x_{m-n-1} + \\ & + \left(\frac{r^3}{2} - r^2 + \frac{r}{2} + 1 \right) x_{m-n} + \left(-\frac{r^3}{6} + \frac{r^2}{2} - \frac{r}{3} \right) x_{m-n+1}, \end{aligned}$$

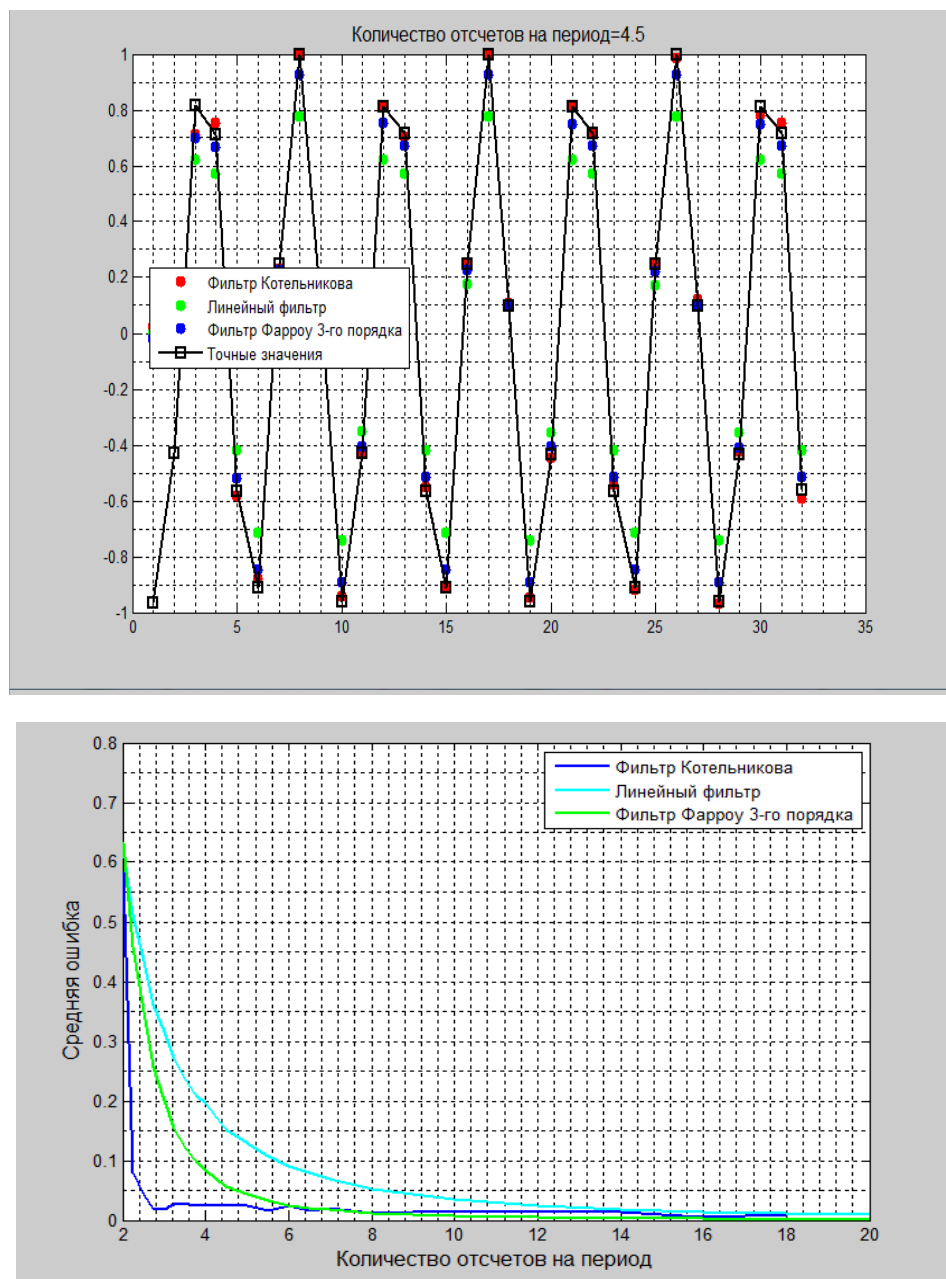


Рис. 1. Среднее абсолютное отклонение от точных значений.

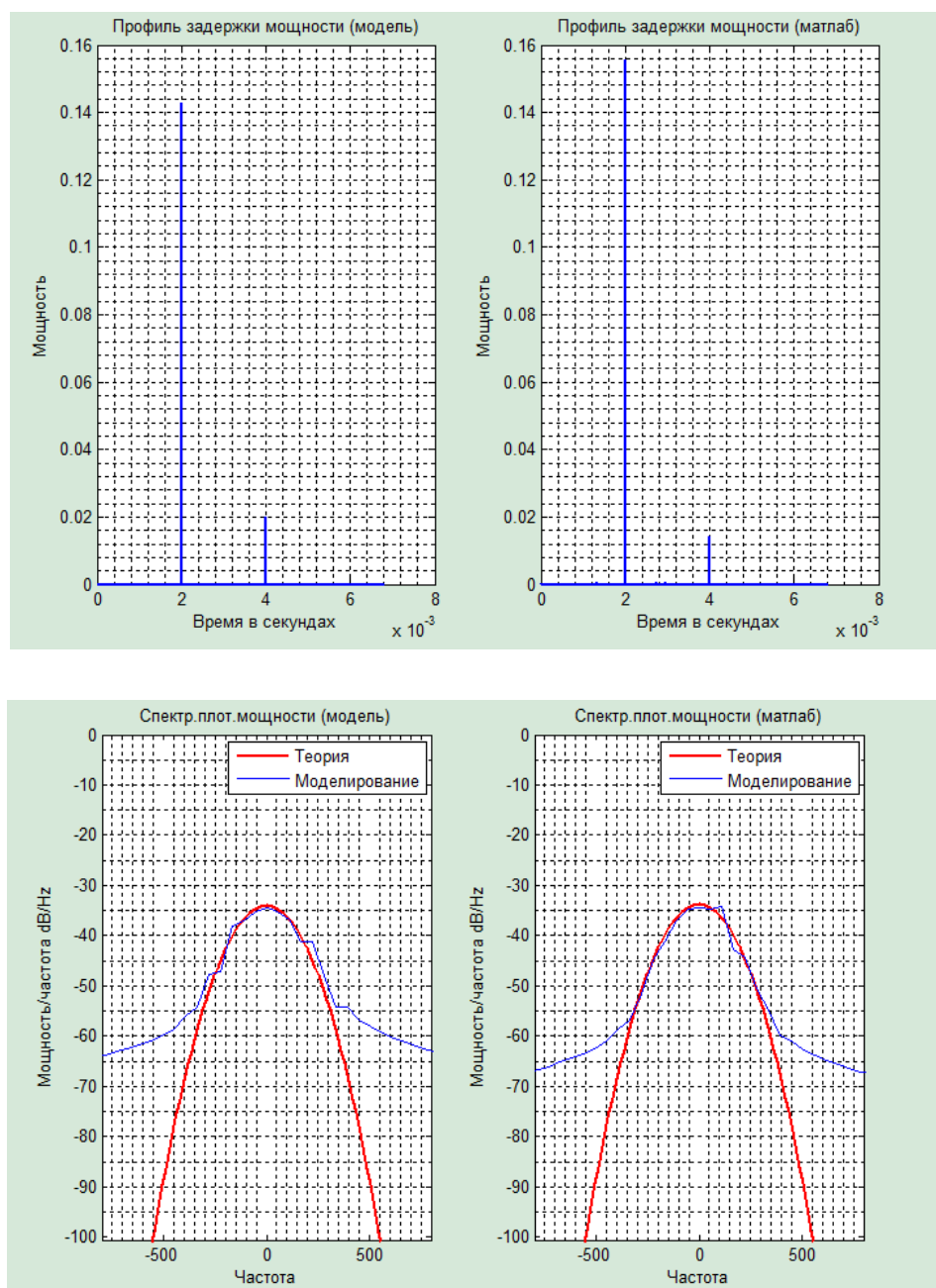


Рис. 2. Сравнение профиля задержки и спектральной плотности мощности

где n и r — целая и дробная части числа $\frac{dt}{T}$.

Быстрый алгоритм вычисления можно найти здесь: <http://www.dsplib.ru/content/farrow/farrow.html>.

По результатам эксперимента можно сделать следующие выводы.

Время интерполяции с помощью фильтра Котельникова в 14 раз больше чем время интерполяции с помощью фильтра Фарроу и в 27 раз больше чем время интерполяции с помощью линейного фильтра (оценка для $N=128$).

В тоже время для достижения достаточной точности восстановления сдвига сигнала по времени частота отсчётов сигнала для линейного фильтра должна быть в 7,5 раз больше, а для фильтра Фарроу в 3,5 раза больше чем для фильтра Котельникова. Оптимальным, по нашему мнению, является фильтр Фарроу.

В Communications System Toolbox в Matlab реализованы многие модели каналов в виде системных объектов без предоставления кодов.

Примеры стандартных моделей каналов в Communications Systems Toolbox из ITU-R HF и из ITU-R 3G можно найти в [14–15].

Проведённый анализ позволяет выделить следующие основные программные модули (функции):

1. Главный модуль. Задание параметров. Спецификация профиля тракта, входного сигнала и сеанса моделирования. Формирование выходного сигнала. Графическое представление преобразования комплексной огибающей сигнала в канале во временной и частотной областях.

2. Генераторы случайных величин. Формирование значений равномерно распределённой случайной величины. Формирование комплексного белого шума. Формирование значений логнормальной случайной величины. Формирование аддитивных помех.

3. Фильтры. Доплеровский фильтр.

4. Модуль интерполяции и децимации. Моделирование временных сдвигов дискретного сигнала. Изменение частоты отсчётов (понадобится для ускорения работы модели, чтобы не формировать очень много отсчётов белого шума).

5. Модуль формирования коэффициента потерь в тракте.

6. Модуль тестирования канала.

Заключение. Для проверки правильности основных алгоритмов программно реализован прототип модели канала. Проведено сравнение работы наших алгоритмов с алгоритмами, реализованными в MATLAB. Установлена адекватность реализации доплеровского фильтра и временных сдвигов дискретного сигнала. На рис. 2 приведены сравнительные графики.

ЛИТЕРАТУРА

1. Волков А.Н. Разработка и исследование методов и средств моделирования декаметровых радиоканалов в реальном масштабе времени: автореферат канд.дисс. Санкт-Петербург : 2002. 184 с.
2. Самойлов А.Г. Имитаторы многолучевых радиоканалов // Проектирование и технология электронных средств. 2003. № 4. С. 32–36.

3. Хазан В.Л. Методы и средства проектирования каналов декаметровый радиосвязи: автореферат докторской дисс. Омск : ОмГТУ, 2007. 385 с.
4. Ву Ван Шон. Разработка имитатора радиоканалов мобильной связи: автореферат диссертации на соискание учёной степени кандидата технических наук. Владимир, 2009. 20 с.
5. Герасимов А.Б. Имитация радиосигналов, рассеянных сложными радиофизическими сценами в реальном масштабе времени: автореферат канд. дисс. Ярославль, 2011. 166 с.
6. Тверской Г.Н., Терентьев Г.К., Харченко И.П. Имитаторы эхо-сигналов судовых радиолокационных станций. JL : Судостроение, 1973. 225 с.
7. Карпов А.А. Модель городского многолучевого радиоканала с предварительной обработкой данных о городской застройке // Журнал Радиоэлектроники Электронный ресурс. 2008. № 8.
8. Марпл-мл. С.Л. Цифровой спектральный анализ и его приложения. М. : Мир, 1990. 584 с.
9. Кловский Д.Д., Конторович В.Я., Широков С.М. Модели непрерывных каналов связи на основе стохастических дифференциальных уравнений / Под ред. Д.Д. Кловского. М. : Радио и связь, 1984. 248 с.
10. Волков Л.Н., Немировский М.С., Шинаков Ю.С. Системы цифровой радиосвязи. М. : Эко Трендз, 2005. 392 с.
11. Simon M.K., Alouini M.-S. Digital Communication over Fading channels. A John Wiley Sons, 2005. 900 с.
12. Галкин Л.П., Лапин А.Н., Самойлов А.Г. Моделирование каналов систем связи. М. : Связь, 1979. 96 с.
13. Кнут Д.Э. Искусство Программирования. Том 2, гл. 3. М. : Вильямс, 2011. 832 с.
14. ITU-R F.1487 (2000). Testing of HF modems with bandwidths of up to about 12 kHz using ionospheric channel simulators (ITU-R HF channel models).
15. ITU-R M.1225 (1997-2). Guidelines for evaluations of radio transmission technologies (ITU-R 3G channel models).
16. РЕКОМЕНДАЦИЯ МСЭ-R P.1407-2. Многолучевое распространение и параметризация его характеристик.
17. РЕКОМЕНДАЦИЯ МСЭ-R P.1057-2. Распределения вероятностей, касающихся моделирования распространения радиоволн.
18. РЕКОМЕНДАЦИЯ МСЭ-R P.1812-2. Метод прогнозирования распространения сигнала на конкретной трассе для наземных служб «из пункта в зону» в диапазонах УВЧ и ОВЧ.
19. РЕКОМЕНДАЦИЯ МСЭ-R P.1546-2. Метод прогнозирования для трасс «точка-зона» для наземных служб в диапазоне частот от 30 МГц до 3000 МГц.
20. Jeruchim M., Balaban P., Shanmugan K.S. Simulation of Communication Systems. Second Edition. New York : Kluwer Academic/Plenum, 2000.
21. Saunders S.R. Antennas and propagation for wireless communication systems. England : John Wiley, Sons Ltd, 2007.
22. Феер К. Беспроводная цифровая связь. Методы модуляции и расширения спектра: Пер. с англ. / Под ред. В.И. Журавлева. М. : Радио и связь, 2000. 520 с.

STOCHASTIC SIMULATION OF THE RADIO CHANNEL WITH ADDITIVE AND MULTIPLICATIVE LINE NOISES. IMPLEMENTATION SCHEME

Yu.F. Strugov¹

Professor, Dr.Sc. (Phys.-Math.), Omsk State University n.a. F.M. Dostoevskiy,
e-mail: strugov77@mail.ru

A.M. Semenov¹

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

S.M. Dobrovolskiy¹

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

I.A. Batirev²

ONIIP, e-mail: batirev007@mail.ru

¹Omsk State University n.a. F.M. Dostoevskiy

²PISC "ONIIP"

Abstract. The article deals gives an overview and analysis of existing approaches to mathematical simulation of the multipath environment of radio wave propagation at VHF band. In particular, implementation of the random processes describing additive and multiplicative line noises is discussed in more detail.

Keywords: simulation of the radio channel, radio wave, multipath environment of the radio wave propagation, Doppler shift, additive and multiplicative line noises.

АНАЛИЗ КОГНИТИВНЫХ МОДЕЛЕЙ «АРТЕРИАЛЬНАЯ ГИПЕРТЕНЗИЯ»

В.А. Маренко¹

к.т.н., старший научный сотрудник, e-mail: marenko@ofim.oscsbras.ru

О.Н. Лучко²

к.п.н., профессор, заведующий кафедрой прикладной информатики и математики,
e-mail: o_luchko@rambler.ru

В.А. Ляпин³

д-р.мед.н., профессор, зав. кафедрой анатомии, физиологии, спортивной медицины и
гигиены, e-mail: v.a.liapin@mail.ru

С.Ю. Гуша³

директор спортивного колледжа Сибирского государственного университета
физкультуры и спорта, e-mail: colledge@sibgufk.ru

Л.В. Алексеенко⁴

начальник отдела защиты прав граждан в системе ОМС ТФОМС Омской области,
e-mail: alexeenko_lv@tfoms.omsk.ru

¹Институт математики им. С.Л. Соболева СО РАН

²Омский государственный институт сервиса

³Сибирский государственный университет физической культуры и спорта

⁴Территориальный фонд обязательного медицинского страхования Омской области

Аннотация. Артериальная гипертензия является большой проблемой общества. Анализ научной информации по проблеме показал, что основным инструментом исследований является математическая статистика. Авторы предлагают использовать когнитивное моделирование. Цель работы — применение экспертных знаний для построения и анализа когнитивных моделей в виде взвешенных ориентированных графов как субъективных образов проблемы. Решены следующие задачи. Построены субъективные пути размышлений специалистов о возможных причинно-следственных связях между объектами проблемной области в виде когнитивных карт или взвешенных ориентированных графов. Проведён компьютерный эксперимент для верификации авторских когнитивных моделей путём наблюдения за распространением возмущений, вводимых в различные вершины графа. Анализируются функциональные состояния студентов-спортсменов и неспортсменов для уточнения значений фактора «физическая активность».

Ключевые слова: гипертензия, физическая активность, когнитивная модель, экспертиза, компьютерный эксперимент, взвешенный ориентированный граф.

Введение

Основная идея построения когнитивных моделей состоит в том, чтобы формализовать наше субъективное представление о реальности путём построения новых субъективных образов. Для этого необходимо выделить проблему, схематически представить информацию о ней, а затем провести её когнитивный анализ на основе опыта, знаний и интуиции исследователя.

Цель работы — применение экспертных знаний для построения и анализа когнитивных моделей, иллюстрирующих влияние существенных факторов среды жизнедеятельности на развитие ситуаций, связанных с повышением артериального давления (АД) у человека.

Биологические объекты исследовать сложно в связи с тем, что они характеризуются большим количеством показателей, описывающих физико-химические свойства внутренней среды организма человека и физические свойства, проявляющиеся при взаимодействии с внешней средой. Оптимальное функциональное состояние организма связано с устойчивыми нормативными показателями внутренней среды. Это состояние может нарушаться, а соответствующие показатели характеризовать отклонение от нормы в сторону увеличения или уменьшения. Под влиянием медикаментов или самопроизвольно, в связи с наличием свойства гомеостаза у человека, показатели могут возвращаться к нормативным значениям. Артериальное давление, например, изменяется при действии следующих основных факторов: стресс, гиподинамия, антропометрические характеристики (избыточный вес) и др.

Связь значений артериальной гипертензии с другими показателями организма описана многими как зарубежными, так и российскими исследователями. Регуляция артериального давления при стрессовых ситуациях приведена в работах М.С. Гавриловой и В.А. Шовина [1, 2]. Влияние вредных привычек (табакокурения, употребления алкоголя) на изменение артериального давления описано в статьях А.И. Счастлиненко и Н.А. Агаджаняна [3, 4]. Построение факторных моделей и анализ корреляционной зависимости биохимических показателей при артериальной гипертензии отражён в работах В.В. Гольяпина и А.И. Лобачева [5, 6].

В статьях зарубежных исследователей Bum Ju Lee, Jong Yeol Kim, Sakurai M., Miura K., Takamura T., Ota T., Ishizaki M. и др. установлены взаимосвязи между артериальной гипертензией и антропометрическими, гендерными и другими показателями с применением математической статистики [7, 8, 9]. Учёные Sim J.J., Shi J., Kovesdy C.P., Kalantar-Zadeh K., Jacobsen S.J. провели статистические исследования и построили регрессионные модели со стратификацией пациентов-гипертоников с наличием заболевания сахарный диабет и без него. А также — в различных возрастных группах [10]. Авторы Cai G., Zhang B. и др. исследовали связь генетических отклонений у пациентов с заболеваемостью гипертонией в различных этнических популяциях [11, 12]. В следующих статьях показаны связи между вегетарианским рационом питания, сахарным диабетом и распространённостью артериальной гипертензии [13, 14]. В работе Yang F. и др. авторов исследуются факторы риска заболеваемости

гипертонией среди китайского населения, такие как курение, низкий уровень фолиевой кислоты и др. [15].

В статье Toyama M., Watanabe S. приведены результаты исследований зависимости между заболеваемостью гипертонией и массой тела. Выявлено, что утренняя гипертония у тучных пациентов выше, чем у не тучных [16]. Установлено, что урбанизация способствует увеличению заболеваемости гипертонией. Анализ данных свидетельствует о том, что неадекватная инфраструктура здравоохранения может выступать в качестве барьера для оптимального управления заболеваемостью гипертонией [17]. Исследовалась зависимость заболеваемости гипертонией от времени года и возраста [18]. Авторы аргументируют необходимость построения базы данных и всестороннего анализа данных о больных гипертонией, так как это заболевание стало проблемой общества [19].

Обзор статей по проблеме показал, что основным инструментом исследователей — математическая статистика. Авторами применена когнитивная методология. Первый этап исследований — формирование проблемного поля, которое состоит из совокупности взаимовлияющих факторов и связей между ними. Когнитивное отображение проблемного поля осуществляется в виде когнитивной карты или взвешенного ориентированного графа $G = \langle V, E \rangle$, где V — множество вершин (базисных факторов), $V_i \in V$, $i = 1, 2, \dots, k$; E — множество дуг, веса которых показывают силу взаимовлияния базисных факторов, $e_{ij} \in E$, $i, j = 1, 2, \dots, n$. Следующий шаг исследования — построение когнитивной модели для проведения имитационного эксперимента. Когнитивная модель $\Phi = (G, X, F)$, где $G = \langle V, E \rangle$ — ориентированный граф; X — множество параметров вершин V , $X = \{X^{v_i}\}$, $i = 1, 2, \dots, k$; $X^{(v_i)} = x_g^{(i)}$, $g = 1, 2, \dots, n$; $x_g^{(i)}$ — параметр вершины V_i , если $g = 1$, то $x_g^{(i)} = x_i$; $X : V \rightarrow R$, R — множество вещественных чисел; $F = F(X, E) = F(x_i, x_j, e_{ij})$ — функционал преобразования дуг, ставящий в соответствие каждой дуге знак, весовой коэффициент ω_{ij} или функцию $f(x_i, x_j, e_{ij}) = f_{ij}$. Параметры когнитивной модели $x_i(t)$, $t = 1, \dots, n$ зависят от времени. Если в момент времени $t - 1$ в вершину поступал импульс $p_j \in P$, то переход модели из состояния $t - 1$ в t осуществлялся по правилу:

$$x_i(t) = x_{i-1}(t) + \sum_{j=1}^{k-1} f(x_i, x_j, e_{ij})p_j(t-1),$$

при известных начальных значениях [20].

1. Построение когнитивных карт

В наших исследованиях базисные факторы проблемного поля разделены на целевой фактор — «артериальная гипертензия» и управляющие факторы (см. табл. 1).

Далее проводится экспертиза, которая призвана ответить на вопрос, какие изменения в значениях управляющих факторов необходимо произвести, чтобы артериальное давление у человека снижалось до нормативных величин. В процессе экспертизы выявлены следующие суждения специалистов и величины

Таблица 1. Базисные факторы (показатели)

Обозначение	Название	Вид	Единицы измерения
А	Артериальная гипертензия	целевой	мм.рт.ст.
Б	Вредные привычки (табакокурение, употребление алкоголя)	управляющий	балл
В	Эмоциональное состояние		
Г	Физическая активность		
Д	Хронические заболевания		
Е	Стресс		

экспертных оценок. Обострения «хронических заболеваний» могут уменьшать «физическую активность» человека (-0,5). Отрицательное «эмоциональное состояние» человека провоцирует «вредные привычки» (-0,8). При снижении «артериального давления» (до нормативных величин) можно увеличивать дозированную физическую нагрузку (-0,2). Уменьшение стрессовых нагрузок приводит к снижению «артериального давления» (до нормативных величин) (+0,5). Увеличение дозированной «физической активности» способствует нормализации (снижению) «артериального давления» (-0,7) и улучшению «эмоционального состояния» человека (+0,5). Положительный эмоциональный фон приводит к снижению «вредных привычек» (-0,8). «Физическая активность» снижает силу и частоту рецидивов «хронических заболеваний» (-0,5) и др. В наших исследованиях когнитивные карты как математические структуры построены двумя независимыми экспертами (см. рис. 1, 2).

Когнитивные карты имеют следующие различия. На когнитивной карте (к-2) отсутствует дуга ДА («хронические заболевания» — «артериальное давление»). Поменялось направление дуги АЕ («артериальное давление» — «стресс»). Дуга ДГ («хронические заболевания» — «физическая активность») стала двухсторонней. Появилась дуга ЕВ («стресс» — «эмоциональное состояние»).

Далее проводился компьютерный эксперимент, результаты которого подтверждали или опровергали интуицию экспертов, использованную при построении когнитивных карт как субъективных моделей проблемы. Взаимосвязи между вершинами графа представляются когнитивной матрицей, в ячейки которой вводятся возмущения. Распространение возмущений на нескольких шагах вычислений наблюдалось визуально с помощью программного средства Excel.

2. Результаты компьютерного эксперимента

Экспериментальные кривые, полученные с применением когнитивной карты к-1, имеют тенденцию к постоянному увеличению значений (рис. 3). В институте проблем управления РАН разработано теоретическое обоснование этого

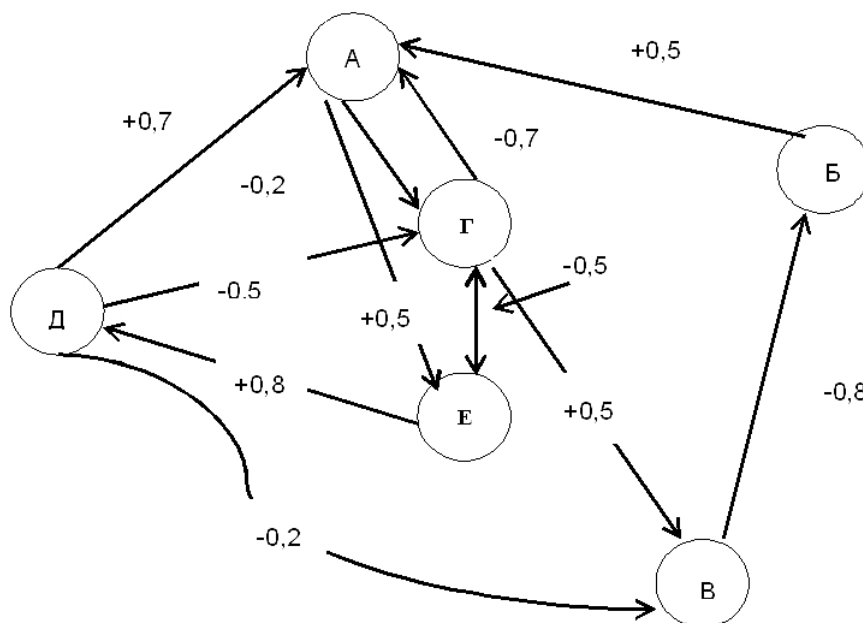


Рис. 1. Когнитивная карта первого эксперта (к-1)

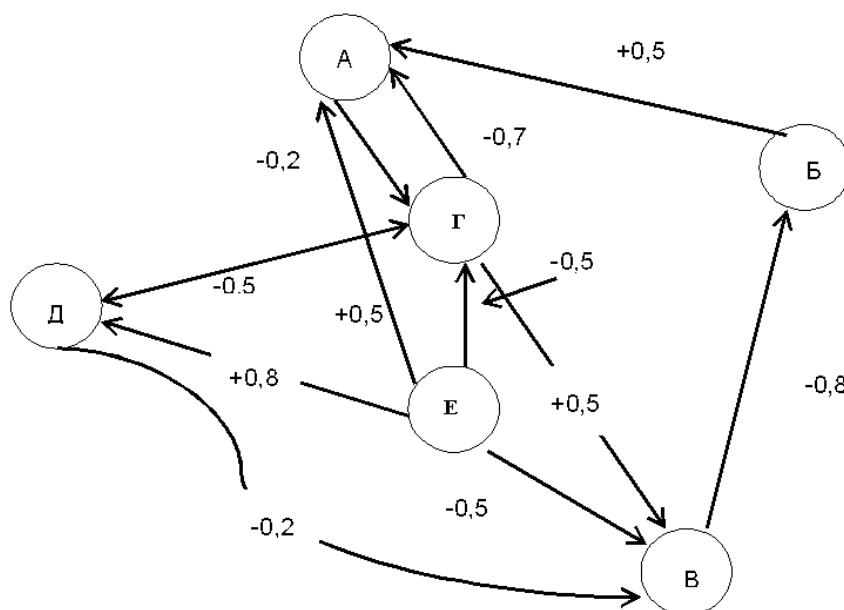


Рис. 2. Когнитивная карта второго эксперта (к-2)

явления, называемого импульсной неустойчивостью [21]. Для того чтобы от него избавиться, необходимо структурное преобразование когнитивной модели проблемы.

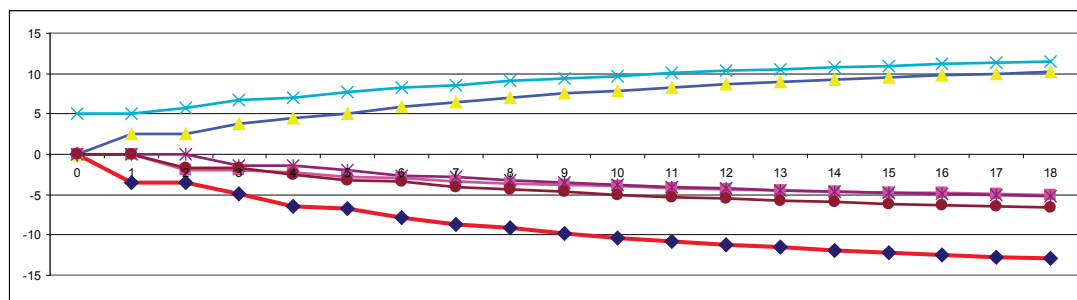


Рис. 3. Результаты компьютерного эксперимента с использованием когнитивной карты к-1 (ось абсцисс – шаги вычислений, ось ординат – значения базисных факторов)

Правомерность внесенных изменений вторым экспертом в структуру когнитивной карты (к-1) подтвердилась. Экспериментальные кривые, полученные с применением когнитивной карты к-2, показывают изменения значений факторов на начальных шагах вычислений, а затем стабилизацию значений на определённых уровнях (рис. 4, 5) [22]. На рисунке 4 приведены результаты компьютерного эксперимента, проведённого с использованием когнитивной карты к-2, при условиях увеличения факторов «физическая активность» (крест), «стресс» (круг) и «вредные привычки» (квадрат) на пять условных единиц. Целевой фактор «артериальное давление» обозначен ромбом. Импульсная неустойчивость отсутствует.

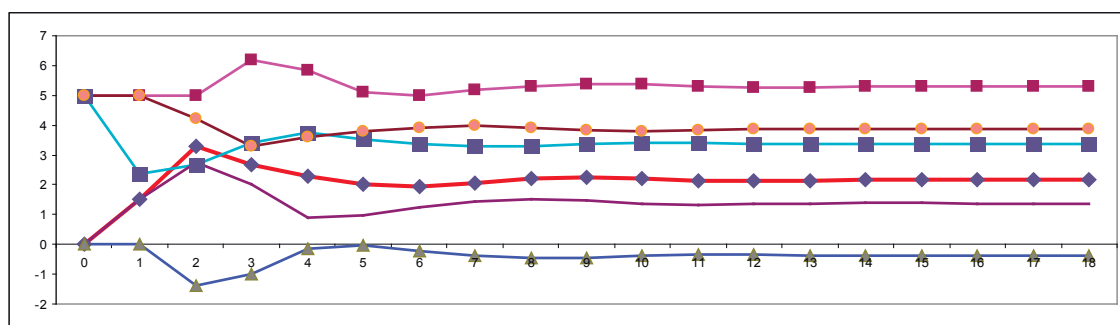


Рис. 4. Результаты компьютерного эксперимента при увеличении факторов «физическая нагрузка» (крест), «стресс» (круг) и «вредные привычки» (квадрат)

Если при предыдущих условиях увеличить ещё и фактор «эмоциональное состояние» (настроение субъекта улучшилось), то на графике видим уменьшение значений целевого фактора «артериальное давление» (ромб) (рис. 5).

Таким образом, одним из существенных управляющих факторов является фактор «эмоциональное состояние» человека.

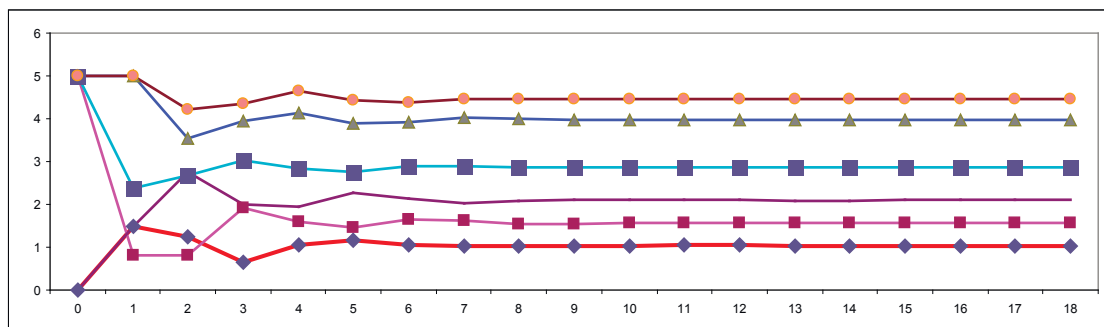


Рис. 5. Результаты компьютерного эксперимента при увеличении факторов «физическая активность» (крест), «стресс» (круг), «вредные привычки» (квадрат) и «эмоциональное состояние» (треугольник)

3. Управляющий фактор «физическая активность»

Проведены эксперименты по оценке функционального состояния студентов с разной степенью повседневной физической активности. Первая группа — студенты, профессионально занимающиеся ИТ-технологиями, вторая группа — студенты Сибирского государственного университета физической культуры и спорта, систематически получающие физические нагрузки.

Эксперимент проводился с применением теста Амосова. Академик Н.М. Амосов в качестве теста предлагал оценивать изменение частоты сердечных сокращений (ЧСС) и общего самочувствия при обычном подъеме пешком на 4-й этаж здания. Функциональное состояние в этом случае оценивается следующим образом: хорошее, если ЧСС не превышает 100-120 ударов в минуту, отсутствуют неприятные ощущения; удовлетворительное, если наблюдается легкая одышка; неудовлетворительное, если уже на 3-м этаже фиксируется выраженная одышка, ЧСС более 140 ударов в минуту, отмечается слабость.

По данным теста построены нечёткие множества «степень функциональности»: (1) для студентов-спортсменов и (2) студентов-неспортсменов. На рис. 6 ось абсцисс — разность ЧСС после и до подъёма испытуемых на 4 этаж здания. Ось ординат — степень выраженности исследуемого свойства.

Сравнение графиков показывает, что степень функциональности у спортсменов выше, чем у неспортсменов, т.к. показатели функциональности после физической нагрузки у них изменяются меньше, чем у неспортсменов. На рис. 7 а, б представлены фазовые портреты показателей разности артериального давления (ось ординат) и разности ЧСС (ось абсцисс) после и до физических нагрузок у студентов-неспортсменов (а) и у студентов-спортсменов (б).

Только десять процентов данных, принадлежащих множеству M_2 , находятся в том же интервале значений, что и данные, принадлежащие множеству M_1 . Таким образом, ослабление повседневной физической активности влияет на функциональные показатели человека и способствует уменьшению выносливости организма в целом.

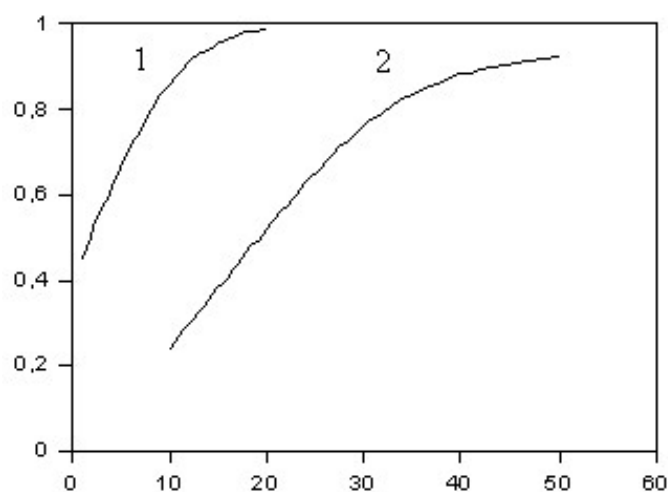


Рис. 6. Нечеткие множества «степень функциональности»
1 — спортсменов, 2 — неспортсменов

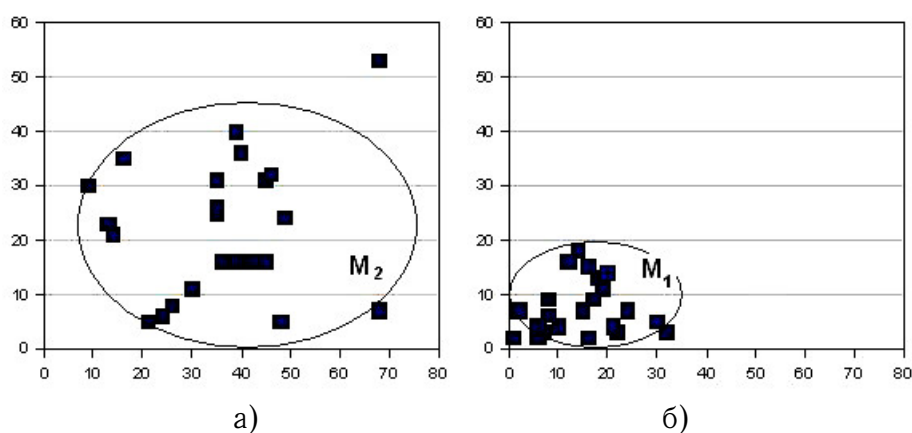


Рис. 7. Фазовые портреты показателей функциональности
а) студентов-неспортсменов, б) студентов-спортсменов

Заключение

Применение когнитивного моделирования позволяет структурировать знания о проблеме с субъективных позиций, систематизировать и преобразовывать их в форму, удобную для анализа, осуществляемого на основе индивидуального опыта, знаний и интуиции специалиста.

ЛИТЕРАТУРА

1. Гаврилова М.С. Математическая модель динамики систолического давления в моменты стрессовых ситуаций у здорового человека // Молодой ученый. Чита. 2009. № 8. С. 6–8.
2. Шовин В.А. Конфирматорная факторная модель артериальной гипертензии // Компьютерное моделирование. 2012. Т. 4, № 4. С. 885–894.

3. Счастливленко А.И., Подпалов В.П., Деев А.Д. Моделирование риска развития артериальной гипертензии // Вестник Витебского государственного медицинского университета. 2005. Т. 1, № 1. С. 38–42.
4. Агаджанян Н.А., Толмачева Н.В., Маслова Ж.В., Капланова А.Ш. Физиологическое обоснование причинно-следственных связей артериальной гипертензии с эколого-биохимическими факторами // Фундаментальные исследования. 2010. № 11. С. 17–21.
5. Гольцяпин В.В., Шовин В.А. Косоугольная факторная модель артериальной гипертензии первой стадии // Вестник Омского университета. 2010. № 4. С. 114–122.
6. Гольцяпин В.В., Лобачев А.И. Факторные и латентные модели в диагностике артериальной гипертензии // Современные проблемы науки и образования. 2013. № 3. URL: www.science-education.ru/109-9519 (дата обращения: 20.08.2014).
7. Lee B.J., Kim J.Y. A Comparison of the Predictive Power of Anthropometric Indices for Hypertension and Hypotension Risk. PLoS ONE 9(1). 2014. P. 1–11.
8. Sakurai M., Miura K., Takamura T., Ota T., Ishizaki M., et al. Gender differences in the association between anthropometric indices of obesity and blood pressure in Japanese. J Hypertens Res 29. 2006. P. 75–80.
9. Barbosa R.A., Munaretti B.D., Coqueiro S.R., Borgatto F.A. Anthropometric indexes of obesity and hypertension in elderly from Cuba and Barbados. J Nutr Health Aging 15. 2011. P. 17–21.
10. Sim J.J., Shi J., Kovesdy C.P., Kalantar-Zadeh K., Jacobsen S.J. Impact of achieved blood pressures on mortality risk and end-stage renal disease among a large, diverse hypertension population // Journal of the American College of Cardiology. 2014. Vol. 64, Issue 6. P. 588–597.
11. Cai G., Zhang B., Weng W., Shi G., Xue S., Song Y., Ma C. E-selectin gene polymorphisms and essential hypertension in Asian population: An updated meta-analysis // PLoS ONE. 2014. Vol. 9, Issue 7. Article number e102058.
12. De Jesus Perez V.A., Yuan K., Lyuksyutova M.A., Dewey F., Orcholski M.E., Shuffle E.M., Mathur M., Yancy Jr.L., Rojas V., Li C.G., Cao A., Alastalo T.-P., Khazeni N., Cimprich K.A., Butte A.J., Ashley E., Zamanian R.T. Whole-exome sequencing reveals TopBP1 as a novel gene in idiopathic pulmonary arterial hypertension // American Journal of Respiratory and Critical Care Medicine. 2014. Vol. 189, Issue 10. P. 1260–1272.
13. Orlich M.J., Fraser G.E. Vegetarian diets in the Adventist Health Study 2: A review of initial published findings // American Journal of Clinical Nutrition. Vol. 100, Issue SUPPL. 1. P. 353S–358S.
14. Salas-Salvado J., Guasch-Ferre M., Bullo M., Sabate J. Nuts in the prevention and treatment of metabolic syndrome // American Journal of Clinical Nutrition. 2014. Vol. 100, Issue SUPPL. 1. P. 399S–407S.
15. Yang F., Liu Q.-Q., Wang L.-J., Guo W.-Y., Yao Y. Risk factors of vascular cognitive impairment among Chinese population: Meta-analysis // Journal of Jilin University Medicine Edition. 2014. Vol. 40, Issue 3. P. 626–632.
16. Toyama M., Watanabe S., Miyauchi T., Kuroda Y., Ojima E., Sato A., Seo Y., Aonuma K. Diabetes and obesity are significant risk factors for morning hypertension: From Ibaraki Hypertension Assessment Trial (I-HAT) // Life Sciences. 2014. Vol. 104, Issue 1–2. P. 32–37.
17. Iwelunmor J., Airhihenbuwa C.O., Cooper R., Tayo B., Plange-Rhule J., Adanu R.,

- Ogedegbe G. Prevalence, determinants and systems-thinking approaches to optimal hypertension control in West Africa // *Globalization and Health*. 2014. Vol. 10, Issue 1. Articlenumber 42.
18. Salwa P., Gorczyca-Michta I., Kluk M., Dziubek K., Wozakowska-Kaplon B. Variability of circadian blood pressure profile during 24-hour ambulatory blood pressure monitoring in hypertensive patients // *Kardiologia Polska*. 2014. Vol. 72, Issue 5. P. 432–437.
19. Kiss I., Kekes E. Hungarian hypertension registry // *Orvosi Hetilap*. 2014. Vol. 155, Issue 19. P. 764–768.
20. Горелова Г.В., Радченко С.А. Когнитивные технологии поддержки управленческих решений в социально-экономических системах // *Известия Южного федерального университета. Технические науки*. 2003. Т. 34, № 5. С. 95–104.
21. Кононов Д.А., Косяченко С.А., Кульба В.В. Формирование и анализ сценариев развития социально-экономических систем с использованием аппарата операторных графов // *Автоматика и телемеханика*. 2007. № 1. С. 121–136.
22. Робертс Ф.С. Дискретные математические модели с приложениями к социальным, биологическим и экологическим задачам. Пер.с англ. А.М. Раппопорта, С.И. Травкина. Под ред. А.И. Теймана. М. : Наука. Гл.ред.физ.-мат.лит., 1986. 496 с.

ANALYSIS OF COGNITIVE MODELS "ARTERIAL HYPERTENSION"

V.A. Marenko¹

Ph.D. (Eng.), Associate Professor, e-mail: marenko@ofim.oscsbras.ru

O.N. Luchko²

Ph.D. (Ped.), Professor, e-mail: o_luchko@rambler.ru

V.A. Liapin³

Doctor of Medical Sciences, Professor, e-mail: v.a.liapin@mail.ru

S.Y. Gushcha³

Director of Sports College, e-mail: collodge@sibgufk.ru

L.V. Alekseenko⁴

The Chief of Department of Protection of the Rights of Citizens in the System of Compulsory Health Insurance, e-mail: alexeenko_lv@tfoms.omsk.ru

¹Sobolev Institute of Mathematics Siberian Branch of the Russian Academy Sciences

²Omsk State Institute of Service

³Siberian State University of Physical Culture and Sports

⁴Territorial'nyy fond obyazatel'nogo meditsinskogo strakhovaniya Omskoy oblasti

Abstract. Arterial hypertension is a big problem of society. The analysis of scientific knowledge on the problem shows that the main instrument of the researches is mathematical statistics. The authors, in turn, suggest to use cognitive modeling. The paper purpose is to apply expert knowledge for construction and analysis of cognitive models described as weighed directed graphs that represent subjective images of the problem. The following tasks are solved. Ways of experts reflections about possible cause-and-effect relations between objects of the problem domain are modelled in the form of cognitive maps or weighed directed graphs. A computer experiment for verification of the author's cognitive models is performed via monitoring distribution

of perturbations introduced into various vertices of the graph. Functional states of students, both athletes and non-athletes, are analyzed to specify the values of the factor "physical activity".

Keywords: hypertension, physical activity, cognitive model, expertise, computer experiment, weighed directed graph.

МЕТОДЫ ВРАЩЕНИЯ ФАКТОРНЫХ СТРУКТУР

В.А. Шовин

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

В.В. Гольяпин

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

Институт математики им. С.Л. Соболева СО РАН

Аннотация. Представлена реализация различных методов факторного вращения на базе численных методов нелинейной оптимизации с условиями. Приведено сравнение ортогональных и косоугольных методов вращений. Доказано превосходство косоугольных методов вращения над ортогональными.

Ключевые слова: факторный анализ, квартимакс, варимакс, облимакс, квартимин, облимин, бинормамин.

Введение

Одной из проблем факторного анализа является проблема вращения и интерпретации. Такие методы получения первичного факторного решения как центроидный метод, метод главных факторов, метод минимальных остатков не позволяют производить интерпретацию факторной структуры. Например, центроидный метод и метод главных факторов имеют высоконагруженный исходными показателями первый фактор, включающий в себя максимум разброса значений переменных, вычисляемых в проекции на первый фактор. Метод минимальных остатков имеет непредсказуемый характер, зависящий от начального приближения, что объясняется его критерием минимизации невязок восстанавливаемой и исходной корреляционной матрицей. Проблема факторного вращения связана с неоднозначностью факторных решений. В рамках критерия минимальных остатков удовлетворительное решение может иметь различный вид.

Для выделения однозначного решения могут использоваться различные критерии вращения. Получаемое в результате вращения факторное решение должно обладать хорошей интерпретируемостью, суть которого заключается в однозначном отнесении каждой исходной переменной лишь к одному из факторов. Результирующие факторы могут быть как ортогональны друг другу, так и косоугольны. Косоугольное вращение приводит, как правило, к лучшему разнесению переменных на факторы в связи с большей возможностью выбора направления факторных осей и получения простой факторной структуры [1], когда исходные переменные максимально прижаты к факторным осям.

Факторное вращение осуществляется посредством умножения первичного факторного отображения на матрицу преобразования. Матрица, полученная как произведение транспонированной матрицы преобразования на саму матрицу преобразования, соответствует матрице корреляций между факторами. Поэтому на матрицу преобразования накладывается как минимум одно ограничение, заключающееся в том, что диагональные элементы матрицы произведения, соответствующие корреляциям между одними и теми же факторами, должны равняться 1, тогда как внедиагональные элементы должны быть по модулю не больше 1. Если получают ортогональное решение, то корреляции между различными факторами должны равняться 0. Матрица преобразования определяется в соответствии с определённым критерием от элементов конечной факторной структуры. Минимизация или максимизация этого критерия позволяет найти оптимальную матрицу преобразования, доставляющую минимум или максимум целевой функции критерия.

В данной работе оптимизация критерия вращения осуществляется различными методами нелинейной оптимизации: метод конфигураций, метод деформируемого многогранника, метод Розенброка и метод случайного поиска. Ограничения, накладываемые на матрицу преобразования, учитываются с помощью метода штрафных функций. В работе приводится сравнение различных методов вращения: кватримакс, варимакс, облимакс, кватримин, облимин, бинормамин. Исследуется оригинальный критерий интерпретируемости, естественным образом учитывающий интерпретационные свойства целевой факторной структуры. Все критерии вычисляются в двух вариантах: с ограничениями на ортогональность матрицы преобразования и без такого ограничения. Тем самым такие известные ортогональные методы как кватримакс и варимакс вычисляются впервые как косоугольные методы без дополнительных ограничений на ортогональность. Доказывается превосходство косоугольных методов вращения над ортогональными.

1. Математическая постановка задачи

Матрица $A \leftrightarrow_{m \times g} a_{ij}$ — матрица первичного факторного отображения размерности $m \times g$ весовых коэффициентов. Где m — число изучаемых параметров, g — число общих факторов.

Вращение заключается в следующей матричной операции:

$$V = A\Lambda,$$

$V \leftrightarrow_{m \times g} v_{ij}$ — косоугольная факторная структура;

$\Lambda \leftrightarrow_{g \times g} \lambda_{ij}$ — матрица вращения.

Матрица корреляций $C \leftrightarrow_{g \times g} c_{ij}$ размерности $g \times g$ между конечными факторами, когда исходные данные стандартизированы (дисперсии переменных равны 1, а средние 0) вычисляется по формуле:

$$C = \Lambda^T \Lambda.$$

Для матрицы вращения Λ должны выполняться соотношения:

$$c_{ii} = \sum_{j=1}^g \lambda_{ij}^2 = 1 \quad (j=1, \dots, g).$$

А также ограничения типа неравенств:

$$|c_{ij}| = \left| \sum_{k=1}^g \lambda_{ki} \lambda_{kj} \right| \leq 1.$$

Если результирующее факторное решение должно быть ортогональным, то ограничения типа неравенств, заменяют ограничениями $c_{ij} = \sum_{k=1}^g \lambda_{ki} \lambda_{kj} = 0$.

Задача факторного вращения соответствует максимизации или минимизации определённого критерия K , как функции от элементов матрицы результирующей факторной структуры:

$K = f(\{v_{ij}\}) = f(\{a_{ij}\}, \{\lambda_{ij}\})$. Поскольку элементы $\{a_{ij}\}$ заданы, то задача сводится к нахождению экстремума функции $K = f(\{\lambda_{ij}\})$ от независимых переменных $\{\lambda_{ij}\}$ с ограничениями:

$$\sum_{j=1}^g \lambda_{ij}^2 = 1 \text{ и } |\sum_{k=1}^g \lambda_{ki} \lambda_{kj}| \leq 1 \text{ или } \sum_{k=1}^g \lambda_{ki} \lambda_{kj} = 0.$$

Также предлагается использовать следующие ограничения на вид результирующей факторной структуры. Общности переменных конечной факторной структуры должны быть не больше общностей переменных исходной факторной структуры, а также не меньше определенного порога значимости:

$$h_i^v = \sqrt{\sum_{k=1}^g v_{ik}^2} \leq \sqrt{\sum_{k=1}^g a_{ik}^2} = h_i^a \text{ и } h_i^v \geq p.$$

2. Критерии вращения

Квартимакс

$$K = \sum_{i=1}^m \sum_{p=1}^g v_{ip}^4.$$

Максимизация данного критерия приводит теоретически к минимальной сложности каждого исходного параметра равной 1, когда исходный параметр выражается только через один фактор [2].

Варимакс

$$K = n \sum_{p=1}^g \sum_{i=1}^m \left(\frac{v_{ip}}{h_i} \right)^2 - \sum_{p=1}^g \left(\sum_{i=1}^m \frac{v_{ip}^2}{h_i^2} \right)^2.$$

Максимизация варимакс критерия соответствует максимизации дисперсий квадратов нагрузок факторов. Тем самым теоретическая сложность фактора уменьшается, нагрузки фактора близки к 0 или 1, и фактор можно наилучшим образом проинтерпретировать. Нормализация факторных нагрузок в данном

критерии устраняет различие между вкладами отдельных параметров пропорциональное их общностям.

Облимакс

$$K = \frac{\sum_{i=1}^m \sum_{p=1}^g v_{ip}^4}{\left(\sum_{i=1}^m \sum_{p=1}^g v_{ip}^2 \right)^2}.$$

Максимизация данного критерия соответствует максимизации эксцесса случайной величины ξ , представленной выборкой v_{ij} и $-v_{ij}$. В результате максимизируется доля больших и маленьких (близких к нулю) элементов факторной структуры.

Квартимин

$$K = \sum_{p < q=1}^g \sum_{i=1}^m v_{ip}^2 v_{iq}^2.$$

Минимизация данного критерия соответствует идее простой факторной структуры, когда для фиксированной пары факторов переменные максимально приближаются к одному из факторов.

Облимин

$$K = \sum_{p < q=1}^g \left[n \sum_{i=1}^m \left(\frac{v_{ip}}{h_i} \right)^2 \left(\frac{v_{iq}}{h_i} \right)^2 - \gamma \left(\sum_{i=1}^m \frac{v_{ip}^2}{h_i^2} \right) \left(\sum_{i=1}^m \frac{v_{iq}^2}{h_i^2} \right) \right],$$

При $\gamma = 0$ — это нормализованный квартимин критерий,

при $\gamma = 1$ критерий называется коваримин,

при $\gamma = 0.5$ критерий называется биквартимин.

Минимизация коваримин-критерия соответствует минимизации ковариации между квадратами элементов различных пар факторов конечной факторной структуры. Коваримин-критерий теоретически даёт ортогональное решение.

Бинормамин

$$K = \sum_{p < q=1}^g \left[\sum_{i=1}^m \left(\frac{v_{ip}}{h_i} \right)^2 \left(\frac{v_{iq}}{h_i} \right)^2 \right] / \left(\sum_{i=1}^m \frac{v_{ip}^2}{h_i^2} \right) \left(\sum_{i=1}^m \frac{v_{iq}^2}{h_i^2} \right).$$

Минимизация данного критерия близка к результатам биквартимин решения.

Критерий интерпретируемости

Получение интерпретируемого факторного решения связано с получением минимальной сложности исходных параметров, когда только одна факторная нагрузка переменной близка к 1, тогда как остальные близки к 0. Поэтому предлагается следующий критерий, непосредственно учитывающий это свойство.

$$K = \sum_{i=1}^m \sum_{p=1}^g |v_i^{\max}| - |v_{ip}|,$$

где $v_i^{\max}$ — максимальная по модулю факторная нагрузка i -ой переменной факторной структуры.

Максимизация данного критерия приводит к тому, что максимальная факторная нагрузка переменной приближается к 1, тогда как остальные к 0.

3. Методы оптимизации

Оптимизацию критерия вращения, как функций от независимых переменных матрицы вращения с ограничениями, предлагается осуществлять методом штрафных функций [3]. В качестве методов безусловной оптимизации метода штрафных функций выбирались следующие альтернативные методы:

- метод конфигураций,
- метод деформируемого многогранника,
- метод Розенброка [4],
- метод случайного поиска [5].

4. Численный эксперимент

Метод штрафных функций с выбором метода безусловной оптимизации, а также адаптация задачи факторного вращения и выбор различных критериев вращения были реализованы в виде отдельной программы RFA (рис. 1, 2). Поскольку использовались неградиентные методы оптимизации, все критерии максимизации были приведены к критериям минимизации с помощью обращения критерия $1/K$. Тем самым нелинейность преобразования критерия существенно не изменяла работу алгоритмов поиска экстремума, и значения критериев всегда оставались положительными.

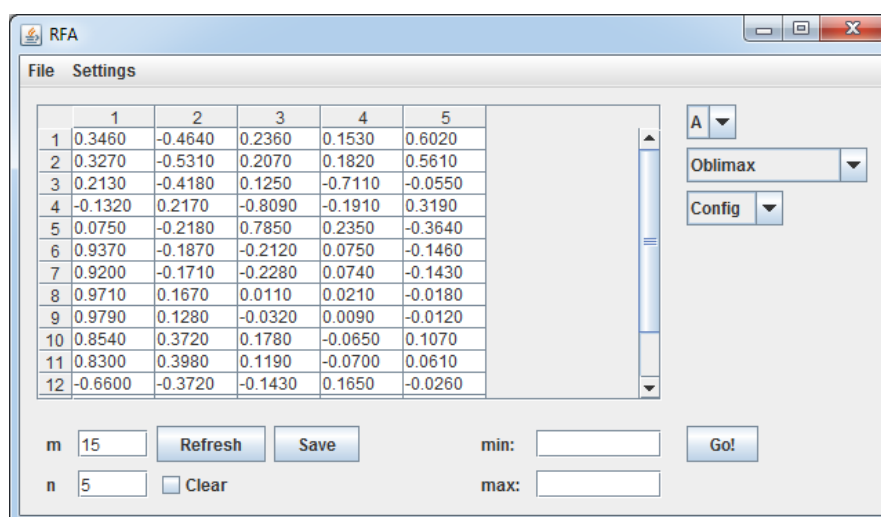


Рис. 1. Интерфейс программы RFA.

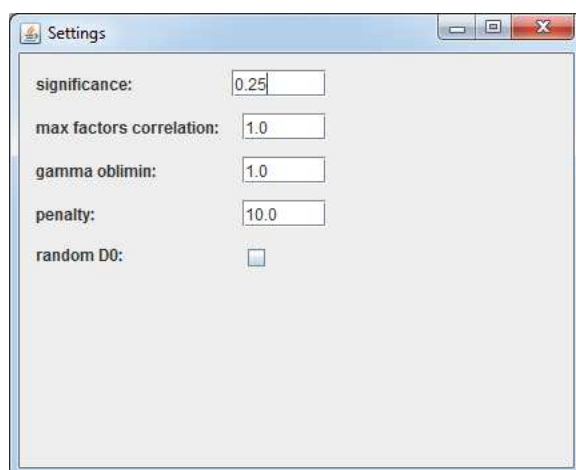


Рис. 2. Интерфейс программы RFA.

Достоверность результатов, получаемых программой RFA, подтверждена на классических примерах, таких как 8 морфологических параметров, 24 психологических параметра, 5 социально-экономических параметров [2], 12 переменных кровяного давления [1].

Сравнение результатов факторных структур для 15 переменных артериальной гипертензии начальной стадии представлены в следующих таблицах (таблице 1–4). В качестве исходных параметров были взяты 15 биофизических показателей для 131 лица с артериальной гипертензией начальной стадии:

- 1) *вес,*
- 2) *индекс массы тела (ИМТ),*
- 3) *частота дыхания (ЧД),*
- 4) *сегментоядерные нейтрофилы (С),*
- 5) *лимфоциты (Л),*
- 6) *конечно-систолический размер левого желудочка (КСР),*
- 7) *конечно-систолический объем левого желудочка (КСО),*
- 8) *конечно-диастолический размер левого желудочка (КДР),*
- 9) *конечно-диастолический объем левого желудочка (КДО),*
- 10) *ударный объем (УО),*
- 11) *минутный объем сердца (МОС),*
- 12) *общее периферическое сосудистое сопротивление (ОПСС),*
- 13) *индекс Хильдебрандта (ИХ),*
- 14) *фракция выброса левого желудочка (ФВ),*
- 15) *фракция укорочения левого желудочка (ФУ).*

Критерий варимакс без ограничений на ортогональность равен 147.1935. Для ортогонального случая его значение 139.8764. Критерий облимакс для косоугольного случая равен 0.0627.

Критерий интерпретируемости для косоугольного случая равен 40.0769.

Критерий облимакс для исходного факторного решения равен 0.0436. Для

Таблица 1. Исходное факторное решение (метод главных факторов)

	<i>F1</i>	<i>F2</i>	<i>F3</i>	<i>F4</i>	<i>F5</i>
Вес	0,346	-0,464	0,236	0,153	0,602
ИМТ	0,327	-0,531	0,207	0,182	0,561
ЧД	0,213	-0,418	0,125	-0,711	-0,055
С	-0,132	0,217	-0,809	-0,191	0,319
Л	0,075	-0,218	0,785	0,235	-0,364
КСР	0,937	-0,187	-0,212	0,075	-0,146
КСО	0,92	-0,171	-0,228	0,074	-0,143
КДР	0,971	0,167	0,011	0,021	-0,018
КДО	0,979	0,128	-0,032	0,009	-0,012
УО	0,854	0,372	0,178	-0,065	0,107
МОС	0,83	0,398	0,119	-0,07	0,061
ОПСС	-0,66	-0,372	-0,143	0,165	-0,026
ИХ	-0,107	0,412	-0,117	0,736	0,068
ФВ	-0,351	0,557	0,408	-0,244	0,261
ФУ	-0,304	0,467	0,419	-0,149	0,239

Таблица 2. Факторная структура варимакс (ортогональный случай)

	<i>F1</i>	<i>F2</i>	<i>F3</i>	<i>F4</i>	<i>F5</i>
Вес	0,1349	-0,0739	0,0789	-0,0699	0,8613
ИМТ	0,0842	-0,1472	0,0952	-0,0701	0,8584
ЧД	0,086	-0,0774	0,0433	-0,8493	0,0873
С	-0,07	-0,052	-0,9135	0,0598	-0,1075
Л	0,0117	0,0282	0,9229	-0,0162	0,0647
КСР	0,7393	-0,6377	0,0145	-0,0933	0,151
КСО	0,7295	-0,6287	-0,0047	-0,0835	0,1379
КДР	0,943	-0,2564	0,0485	-0,0129	0,1191
КДО	0,9316	-0,2985	0,0165	-0,035	0,1329
УО	0,9465	0,0697	0,0524	-0,0015	0,1075
МОС	0,9306	0,0459	0,017	0,0105	0,0393
ОПСС	-0,7759	-0,1167	-0,0223	0,0736	0,0267
ИХ	0,0057	0,0381	-0,028	0,8585	-0,0441
ФВ	-0,0245	0,8448	0,0047	0,024	-0,1124
ФУ	-0,0275	0,7436	0,0715	0,063	-0,0517

Таблица 3. Факторная структура облимакс (косоугольный случай)

	<i>F1</i>	<i>F2</i>	<i>F3</i>	<i>F4</i>	<i>F5</i>
Вес	0,0558	0,1716	-0,0287	-0,0133	0,8187
ИМТ	-0,0133	0,0944	-0,0474	-0,0116	0,8078
ЧД	-0,0391	0,0199	-0,0214	-0,8344	0,0221
С	-0,0082	0,0312	0,9022	0,0241	-0,0219
Л	-0,0476	-0,0758	-0,9164	0,0165	-0,0184
КСР	0,5147	-0,4831	0,0028	-0,0341	0,014
КСО	0,5108	-0,4774	0,0211	-0,0265	0,0051
КДР	0,8197	-0,122	-0,0235	0,0352	0,0217
КДО	0,7965	-0,1533	0,0087	0,0143	0,0322
УО	0,9124	0,182	-0,0231	0,0321	0,0532
МОС	0,8977	0,1432	0,0077	0,0399	-0,0122
ОПСС	-0,7628	-0,1828	0,0022	0,0552	0,0635
ИХ	0,114	-0,039	0,0099	0,8508	0,0087
ФВ	0,2118	0,7631	0,0016	-0,0191	0,003
ФУ	0,1787	0,6716	-0,0643	0,0295	0,0455

Таблица 4. Факторная структура по критерию интерпретируемости (косоугольный случай)

	<i>F1</i>	<i>F2</i>	<i>F3</i>	<i>F4</i>	<i>F5</i>
Вес	0,1949	0,0000	-0,021	-0,0062	0,7783
ИМТ	0,16	-0,0815	0,0000	0,0000	0,7679
ЧД	0,2165	-0,0211	0,01	-0,8305	0,0261
С	-0,0598	-0,0382	-0,888	0,0225	0,1504
Л	0,0000	0,0000	0,9059	0,0205	-0,19
КСР	0,8631	-0,4849	0,008	-0,0288	0,0149
КСО	0,8502	-0,4783	-0,0095	-0,0214	0,0097
КДР	0,9721	-0,0817	0,0251	0,0187	-0,0001
КДО	0,9734	-0,1221	-0,0066	0,0000	0,0178
УО	0,9085	0,2377	0,0152	0,0023	0,0202
МОС	0,8934	0,2084	-0,0105	0,0107	-0,0357
ОПСС	-0,7374	-0,2506	0,0000	0,0829	0,0839
ИХ	-0,1191	0,0000	0,0003	0,8462	0,0000
ФВ	-0,2005	0,8173	-0,0213	-0,0535	-0,0217
ФУ	-0,1863	0,7167	0,0441	0,0000	0,0094

решения облимакс, полученного с помощью специального алгоритма на базе аналитического исследования поведения функции критерия, критерий равнялся 0.0448 [6]. Решение, полученное на базе метода штрафных функций, оказалось более эффективным.

Область допустимых решений косоугольного случая включает в себя область решений ортогонального, поэтому критерий вращения для косоугольного случая будет не хуже, чем для ортогонального. Это доказывает, что косоугольные факторы могут дать меньшую погрешность и большую близость факторов к исходным параметрам. Наличие корреляций между факторами означает, что между ними есть зависимость: изменение одного фактора означает изменение и другого. Поэтому невозможно выявить влияние каждого фактора на изучаемый процесс. Именно для выявления такого влияния и вводится требование ортогональности факторов.

5. Заключение

Предложен общий метод оптимизации критериев вращения на базе метода штрафных функций и методов безусловной оптимизации. Предложен критерий интерпретируемости естественным образом учитывающий свойства интерпретируемого факторного решения. Все алгоритмы реализованы в виде отдельной программы с графическим интерфейсом для пользователя. Показано преимущество облимакс решения на базе метода штрафных функций над решением, получаемого с помощью специального алгоритма на базе аналитического исследования функции критерия. Доказано преимущество косоугольных методов вращения над ортогональными.

ЛИТЕРАТУРА

1. Иберла К. Факторный анализ. Пер. с нем. В.М. Ивановой. Предисл. А.М. Дуброва. М. : Статистика, 1980.
2. Харман Г. Современный факторный анализ. Пер. с англ. В.Я. Лумельского. Научное редактирование и вступительная статья Э.М. Бравермана. М. : Статистика, 1972.
3. Банди Б. Методы Оптимизации. Вводный курс. М. : Радио и связь, 1988.
4. Кокуев А.Г. Оптимальное управление. Поиск экстремумов многомерных функций. Астрахань : АГТУ, 2011. 34 с.
5. Шовин В.А. Конфирматорная факторная модель артериальной гипертензии // Компьютерные исследования и моделирование. 2012. Т. 4, № 4. С. 885–894.
6. Гольцяпин В.В., Шовин В.А. Косоугольная факторная модель артериальной гипертензии первой стадии // Вестник Омского университета. 2010. № 4. С. 120–128.

FACTOR STRUCTURES ROTATION METHODS

V.A. Shovin

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

V.V. Goltyapin

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

Sobolev Institute of Mathematics Siberian Branch of the Russian Academy Sciences

Abstract. Different factors rotational methods were built on the base of numerical methods of nonlinear optimization with conditions. A comparison of orthogonal and oblique rotational methods presented. Advantages of the oblique rotational methods is proven.

Keywords: factor analysis, quartimax, varimax, oblimax, quartimin, oblimin, binormamin.

WIENER-PROCESS-TYPE EVASIVE AIRCRAFT ACTIONS ARE INDEED OPTIMAL AGAINST ANTI-AIRCRAFT GUNS: WIENER'S DATA REVISITED

V. Kreinovich

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

O. Kosheleva

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

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

Abstract. In his 1940s empirical study of evasive aircraft actions, N. Wiener, the father of cybernetics, found out that the pilot's actions follow a Wiener-type-process. In this paper, we explain this empirical result by showing that such evasive actions are indeed optimal against the 1940s anti-aircraft guns.

Keywords: Wiener process, N. Wiener's experiments, evasive aircraft maneuvers, optimization.

1. Introduction

Wiener's empirical data. Many techniques that form the basis of modern communication, signal processing, and control were developed in the 1940s by MIT's Norbert Wiener, as part of his *cybernetics*. Cybernetics is where the now ubiquitous abbreviation "cyber" — ranging from cyberinfrastructure to cyberbullying — comes from. Wiener's work was boosted during the Second World War, when he worked on automatic control devices for anti-aircraft guns; see, e.g., [1, 4].

Wiener used statistical optimization techniques to develop a firing strategy that would, on average, be most efficient against the pilot's random evasive maneuvers. To formulate the corresponding optimization problem, it is necessary to know the probabilities of different evasive trajectories $\vec{x}(t)$. To find these probabilities, N. Wiener, with his collaborator Julian Bigelow, a pilot by training, set up a flight simulator and recorded the corresponding evasive trajectories.

As a result, for the simplified situation with no restriction of airplane maneuvering, the pilot's evasive trajectories followed the Browning motion (what is now known in Mathematics as a Wiener process) [3], when the change

$$\vec{x}(t + \Delta t) - \vec{x}(t)$$

can be in any spatial direction with equal probability, and the current change does not depend on the previous changes. In more realistic situations, when they took into account that the airplane's velocity $\vec{v}(t)$ cannot be changed abruptly, the change in velocities $\vec{v}(t + \Delta t) - \vec{v}(t)$ followed the Wiener process.

Comment. Based on this information, N. Wiener and J. Bigelow developed an optimal controller. For this particular application, the resulting improvement in efficiency was very low, around a few percents, so this controller was not implemented. However, in many other applications, similar controllers were spectacularly successful.

Question. An interesting question is: why did the pilots use Wiener-process-type evasive actions? Are such evasive actions optimal — or are more efficient evasive maneuvers possible?

In this paper, we use simple game theory to show that the Wiener-process-type evasive actions are indeed optimal.

2. Formulation of the Problem

Only 2-D coordinates are important. First, let us recall that from the viewpoint of the anti-aircraft gun, what is important is a 2-D location of an airplane: if the airplane travels in the 3-rd dimension, along the line of fire, it does not help it evade the shells. Because of this, we will only consider 2-D locations and velocities.

An idealized situation when an aircraft can arbitrarily change its velocity: towards the exact formulation of the problem. Let us first consider the simplified setting, when it is assumed that an aircraft can arbitrarily change its speed $\vec{v}(t)$, as long as this speed does not exceed the limit v_0 imposed by its engine. In this case, if at the moment t , the aircraft was at location $\vec{x}(t)$, by the next moment of time $t + \Delta t$ it can travel any distance not exceeding $v_0 \cdot \Delta t$. Thus, at the moment $t + \Delta t$, the aircraft can be anywhere in the disk D of radius $v_0 \cdot \Delta t$ centered at the point $\vec{x}(t)$. Selecting an evasive maneuver means selecting a probability distribution $\rho_p(\vec{x})$ on this disk, a distribution that determines with what probability the plane will be at a given location.

The adversary observes the position $\vec{x}(t)$ and the type of the plane, so the adversary knows the plane's maximum velocity v_0 and thus, knows the disk D of possible locations of the plane at the next moment of time. Once the disk is known, the adversary selects his own probability distribution $\rho_s(\vec{x})$, distribution that describes with what probability the shell is aimed towards a future location $\vec{x}$.

The goal of the pilot is to evade the shell, i.e., to minimize the probability of being hit, while the goal of the gunner is to hit the plane, i.e., to maximize this probability. A shell hits the plane if it is sufficiently close to the plane, i.e., if the position $\vec{x}_s$ of the shell is within a certain small distance $\varepsilon > 0$ from the position $\vec{x}_p$ of the plane. For each position $\vec{x}_s$ of the shell, the plane is hit if this plane is within a circle C of radius ε with a center in $\vec{x}_s$. The probability for a plane to be in this circle is equal to $\int_C \rho_p(\vec{y}) d\vec{y}$. Since the circle is small, the value $\rho_p(\vec{y})$ is practically constant within this circle, so this integral can be approximated as $A_\varepsilon \cdot \rho_p(\vec{x}_s)$, where $A_\varepsilon = \pi \cdot \varepsilon^2$ is the area of this circle.

For each location $\vec{x}_s$ of the shell, the probability of a plane being hit is thus equal to $A_\varepsilon \cdot \rho_p(\vec{x}_s)$. The probability of a shell being in this location is proportional

to $\rho_s(\vec{x}_s)$. Thus, by using the formula of complete probability, we can compute the probability of being hit as $A_\varepsilon \cdot \int_D \rho_s(\vec{x}) \cdot \rho_p(\vec{x}) d\vec{x}$.

This is a zero-sum game: a win for the plane — successful evasion of the shell — is a loss for the adversary. So, according to game theory (see, e.g., [2]), the optimal strategy for a pilot is a *minimax* strategy, i.e., a strategy that minimize the worst-case loss. For this strategy, the worst-case value

$$J(\rho_p) = A_\varepsilon \cdot \max_{\rho_s(\vec{x})} \int_D \rho_s(\vec{x}) \cdot \rho_p(\vec{x}) d\vec{x}$$

is the smallest possible. Let us show how to solve this optimization problem.

Solving the resulting problem. The above integral is the expected (mean) value of the probability density function $\rho_p(\vec{x})$ over the distribution $\rho_s(\vec{x})$. The expected value of any function is always smaller than or equal to the maximum of this function, so $\int_D \rho_s(\vec{x}) \cdot \rho_p(\vec{x}) d\vec{x} \leq \max_{\vec{x} \in D} \rho_p(\vec{x})$. Thus,

$$\max_{\rho_s(\vec{x})} \int_D \rho_s(\vec{x}) \cdot \rho_p(\vec{x}) d\vec{x} \leq \max_{\vec{x} \in D} \rho_p(\vec{x}).$$

On the other hand, if we take a distribution $\rho_s(\vec{x})$ which is located, with probability 1, at a point $\vec{x}$ where the function $\rho_p(\vec{x})$ attains its maximum, then we will get $\int_D \rho_s(\vec{x}) \cdot \rho_p(\vec{x}) d\vec{x} = \max_{\vec{x} \in D} \rho_p(\vec{x})$. Thus,

$$\max_{\rho_s(\vec{x})} \int_D \rho_s(\vec{x}) \cdot \rho_p(\vec{x}) d\vec{x} = \max_{\vec{x} \in D} \rho_p(\vec{x}),$$

and therefore the value $J(\rho_p)$ is equal to $A_\varepsilon \cdot \max_{\vec{x} \in D} \rho_p(\vec{x})$:

$$J(\rho_p) = A_\varepsilon \cdot \max_{\vec{x} \in D} \rho_p(\vec{x}).$$

Minimizing $J(\rho_p)$ is hence equivalent to minimizing the value $\max_{\vec{x} \in D} \rho_p(\vec{x})$. We know that $\int_D \rho_p(\vec{x}) d\vec{x} = 1$. Here, for every $\vec{x}$, we have $\rho_p(\vec{x}) \leq m \stackrel{\text{def}}{=} \max_{\vec{x} \in D} \rho_p(\vec{x})$, thus, $1 = \int_D \rho_p(\vec{x}) d\vec{x} \leq \int_D m d\vec{x} \leq m \cdot A(D)$, where $A(D)$ is the area of the region D . From $1 \leq m \cdot A(D)$, we conclude that $m \geq \frac{1}{A(D)}$. The equality is possible only when there is equality for all $\vec{x}$ in the inequality $\rho_p(\vec{x}) \leq m$, i.e., when $\rho_p(\vec{x}) = m$ for all $\vec{x}$. This is exactly a uniform distribution on the set D — and of course, this distribution should be independent on what was done in the past.

Thus, in the simplified case, we indeed conclude that the Wiener process is an optimal way to perform evasive actions.

A more realistic formulation of the problem. A more realistic description of evasive actions must take into account that the velocity $\vec{v}(t + \Delta t)$ at the next moment of time $t + \Delta t$ cannot be too much different from the velocity $\vec{v}(t)$ at the previous moment of time, there is a limit on acceleration $|\vec{a}(t)| \leq a_0$. In general,

once we know the initial location $\vec{x}(t)$, the initial velocity $\vec{v}(t)$, and the acceleration $\vec{a}(t)$, we can determine the position $\vec{x}(t + \Delta t)$ at the next moment of time as

$$\vec{x}(t + \Delta t) = \vec{x}(t) + \vec{v}(t) \cdot \Delta t + \frac{1}{2} \cdot \vec{a}(t) \cdot (\Delta t)^2.$$

Here, the initial location $\vec{x}(t)$ and the initial velocity $\vec{v}(t)$ are fixed, and the acceleration $\vec{a}(t)$ can take any value for which $|\vec{a}(t)| \leq a_0$.

Thus, the set of locations $\vec{x}(t + \Delta t)$ is a disk D centered at the point $\vec{x}(t) + \vec{v}(t) \cdot \Delta t$ with radius $\frac{1}{2} \cdot a_0 \cdot (\Delta t)^2$. Similarly to the simplified case, we can describe possible evasive actions by a probability density $\rho_p(\vec{x})$ located on this disk, and, similarly to the simplified case, we can conclude that the optimal evasive action corresponds to the uniform distribution on this disk. In this optimal solution, the change in velocity $\vec{v}(t + \Delta t) - \vec{v}(t) = \vec{a}(t) \cdot \Delta t$ is uniformly distributed on the disk of radius $a_0 \cdot \Delta t$ – and is independent on the previous trajectory of the plane.

Thus, in this realistic case, we indeed conclude that the Wiener process for velocities is indeed an optimal way to perform evasive actions. So, Wiener's empirical data indeed corresponds to optimal evasive action.

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

The authors are thankful to all the participants of the International Conference “Norbert Wiener in the 21 Century” (Boston, Massachusetts, June 24–26, 2014) for valuable discussions.

REFERENCES

1. Masini P.R. Norbert Wiener 1894–1964. Basel : Birkhäuser Verlag, 1990.
2. Myerson R.B. Game Theory: Analysis of Conflict. Harvard, Massachusetts : Harvard University Press, 1997.
3. Stark H., Woods J. Probability, Statistics, and Random Processes for Engineers. Upper Saddle River, New Jersey : Prentice Hall, 2011.
4. Wiener N. Cybernetics, or the Control and Communication in the Animal and the Machine. Cambridge, Massachusetts : MIT Press, 1965.

**УКЛОНЕНИЕ САМОЛЁТА ОТ ЗЕНИТНЫХ ОРУДИЙ ПО ТИПУ
ВИНЕРОВСКОГО ПРОЦЕССА ОПТИМАЛЬНО: НОВОЕ ОБЪЯСНЕНИЕ
ДАННЫХ ВИНЕРА**

В. Крейнович

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

О. Кошелева

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

Техасский университет в Эль Пасо, США

Аннотация. В 1940 году в эмпирическом исследовании уклонения авиации от зенитных орудий, Н. Винер, отец кибернетики, доказывает, что действия пилота должны выглядеть как винеровский случайный процесс. В этой статье мы объясним этот эмпирический результат, показав, что такие действия по уклонению действительно оптимальны против зенитных орудий 1940 года.

Ключевые слова: винеровский процесс, эксперимент Винера, манёвры уклонения, оптимизация.

WHEN AN IDEA COMES, WRITE IT DOWN RIGHT AWAY: MATHEMATICAL JUSTIFICATION OF VLADIMIR SMIRNOV'S ADVICE

O. Kosheleva

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

V. Kreinovich

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

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

Abstract. Among several advices to students, Vladimir Smirnov, a renowned Russian mathematician, suggested that when an idea comes, it is better to write it down right away. In this paper, we provide a quantitative justification for this advice.

Keywords: idea, write down, forgetting curve, recall.

1. Formulation of the Problem

Advice of Vladimir Smirnov. When one of us (VK) became a student at the Mathematics Department of St. Petersburg University, the department had a special poster for incoming students with advice from different professors. One of these advices was from Professor Vladimir Smirnov, the author of a widely used course in higher mathematics [2]: when an idea comes, write it down right away, do not delay.

Qualitative explanation. If one does not write down his/her ideas right away, he/she will forget them, and it will require an additional time to recall it. From this viewpoint, to avoid wasting time, it is better to write down the idea right away.

What we plan to do. The objective of this paper is to provide a quantitative explanation for Smirnov's advice.

2. Analysis of the Problem

An idea comes to mind: two possible reactions. Suppose that an idea comes to mind when a person is in the middle of some activity. Then, the person has two options:

- the first option is to interrupt the current activity, write down the idea, and then resume the current activity;

- the second option is to wait until the end of the current activity, and then write down the idea.

Let us analyze when the first reaction is better, and when the second reaction is better.

First option. If we select the first option, then we need extra time to write down the idea, and we also need some additional time to interrupt (and later resume) the current activity. Let:

- t_w denote the time that is needed to write down the idea right away; and
- t_i denote the additional time needed to interrupt the current activity and to resume it again later.

In these terms, in the first option, we spend an additional time $t_w + t_i$.

Second option: analysis of the problem. In the second option, we do not interrupt the current activity. Instead, we wait until the end of this activity. In this case, we do not spend time on the interruption, but we do need to spend time trying to recover the idea.

Let us estimate this recovery time. Human forgetting is well described by the so-called *Ebbinghaus forgetting curve* [1], according to which the amount $a(t)$ of material that we remember decreases with time as

$$\frac{da}{dt} = -k_f \cdot a,$$

where k_f is a parameter describing the forgetting.

This equation makes perfect sense: in general, we can write that $\frac{da}{dt} = f(a)$ for some function $f(a)$. In the first approximation, we can approximate the function $f(a)$ with the first two terms in its Taylor expansion: $\frac{da}{dt} = c_0 + c_1 \cdot a$. When we have no knowledge, i.e., when originally $a(0) = 0$, then of course there is nothing to forget, thus $\frac{da}{dt} = 0$ as well. The condition that $f(0) = c_0 + c_1 \cdot 0 = 0$ implies that $c_0 = 0$ and thus, $\frac{da}{dt} = c_1 \cdot a$. Forgetting means that the amount of remembered material decreases with time, so $c_1 < 0$, and thus, $c_1 = -k_f$ for some $k_f > 0$. This is exactly the Ebbinghaus law.

Because of this equation, the amount of material remembers after time t is equal to $a(0) \cdot \exp(-k_f \cdot t)$.

Let t_e denote the time needed to finish the current activity. Then, by the time t_e , instead of the original amount of information $a(0)$ about our idea, we remember only the amount $a(0) \cdot \exp(-k_f \cdot t_e)$.

Before we write down the idea, we need to recall it. How can we describe a recall? In general, the amount $a(t)$ recalled by time t can also be described by a differential equation $\frac{da}{dt} = g(a)$ for some function $g(a)$. In the first approximation,

we can approximate the function $g(a)$ with the first two terms in its Taylor expansion: $\frac{da}{dt} = c'_0 + c'_1 \cdot a$. Let us consider an ideal situation when eventually, i.e., when $t \rightarrow \infty$, we can recall everything, i.e., we have $a(t) \rightarrow a(0)$. For $t \rightarrow \infty$, we get $\frac{da}{dt} = 0$ and $a = a(0)$. Thus, $c'_0 + c'_1 \cdot a(0) = 0$ and therefore, $c'_0 = -c'_1 \cdot a(0)$. So, the recall equation $\frac{da}{dt} = c'_0 + c'_1 \cdot a$ can be described as $\frac{da}{dt} = c'_1 \cdot (a - a(0))$. When $a < a(0)$, the amount of recalled material increases with time, so $\frac{da}{dt} > 0$ and thus, $c'_1 < 0$. Thus, we can write that $c'_1 = -k_r$ for some parameter k_r that describes a person's recall rate. In terms of this parameter, the recall equation takes the form

$$\frac{da}{dt} = -k_r \cdot (a - a(0)).$$

This equation can be rewritten as $\frac{d(a(0) - a(t))}{dt} = k_r \cdot (a(0) - a(t))$. At moment t_e , when we start the recall process, we have $a(t_e) = a(0) \cdot \exp(-k_f \cdot t_e)$ and thus, $a(0) - a(t_e) = a(0) \cdot (1 - \exp(-k_f \cdot t_e))$. The corresponding solution to the recall differential equation has the form

$$\begin{aligned} a(0) - a(t) &= (a(0) - a(t_e)) \cdot \exp(-k_r \cdot (t - t_e)) = \\ &= a(0) \cdot (1 - \exp(-k_f \cdot t_e)) \cdot \exp(-k_r \cdot (t - t_e)). \end{aligned}$$

Ideally, we should stop recalling at the moment t_s at which we have recovered everything, i.e., at which $a(t_s) = a(0)$ and $a(0) - a(t_s) = 0$. However, the above expression never reaches 0, so we stop when we have recovered the overwhelming part of the original idea, i.e., when $a(t_s) = a(0) \cdot (1 - \varepsilon)$ for some small value $\varepsilon > 0$. In this case, $a(0) - a(t_s) = a(0) \cdot \varepsilon$. By equating

$$a(0) - a(t_s) = a(0) \cdot (1 - \exp(-k_f \cdot t_e)) \cdot \exp(-k_r \cdot (t_s - t_e))$$

with $a(0) \cdot \varepsilon$, we can deduce the time $t_s - t_e$ needed for this recall: namely, by dividing both sides of the equality by $a(0) \cdot (1 - \exp(-k_f \cdot t_e))$, we conclude that

$$\exp(-k_r \cdot (t_s - t_e)) = \frac{\varepsilon}{1 - \exp(-k_f \cdot t_e)}.$$

By taking logarithms of both sides and changing signs, we get

$$k_r \cdot (t_s - t_e) = \ln(1 - \exp(-k_f \cdot t_e)) - \ln(\varepsilon),$$

and thus,

$$t_s - t_e = \frac{1}{k_r} \cdot \ln(1 - \exp(-k_f \cdot t_e)) - \frac{1}{k_r} \cdot \ln(\varepsilon).$$

In this option, the overall additional time needed to record the idea is equal to $t_w + (t - t_e)$.

Conclusion: when is it better to write down the idea right away. It is beneficial to write down the idea right away if the first alternative leads to smaller amount of additional time, i.e., when $t_i + t_w < t_w + (t_s - t_e)$. This inequality is equivalent to $t_s - t_e > t_i$. In view of the above formula for $t_s - t_e$, Thus, it is better to write down the idea if

$$\frac{1}{k_r} \cdot \ln(1 - \exp(-k_f \cdot t_e)) - \frac{1}{k_r} \cdot \ln(\varepsilon) > t_i,$$

i.e., equivalently, when

$$\ln(1 - \exp(-k_f \cdot t_e)) - \ln(\varepsilon) > k_r \cdot t_i,$$

where:

- k_f is the rate width which the person forgets,
- ε is a portion of the original idea that we are willing to ignore,
- k_r is the rate with which a person recalls a forgotten information, and
- t_i is the time needed to interrupt and then resume the current activity.

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

REFERENCES

1. Schacter D.L. Psychology. New York : Worth Publishers, 2009.
2. Smirnov V.I. A Course in Higher Mathematics. Oxford, London, Edinburgh, New York : Pergamon Press, 1964.

КОГДА ИДЕЯ ПРИХОДИТ, ЗАПИШИТЕ ЕЁ НЕМЕДЛЕННО: МАТЕМАТИЧЕСКОЕ ОБОСНОВАНИЕ СОВЕТА ВЛАДИМИРА СМИРНОВА

О. Кошелева

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

В. Крейнович

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

Техасский университет в Эль Пасо, США

Аннотация. Среди прочих советов для студентов, Владимир Смирнов, известный русский математик, предлагал следующее: «Когда приходит идея, лучше запишите её сразу». В этой статье мы даём количественное обоснование этому совету.

Ключевые слова: идея, запись, кривая забывания, воспоминание.

МЕТОД ВЫБОРА НАПРАВЛЕНИЯ ДЛЯ УСКОРЕНИЯ КЛАССИФИКАЦИИ С ПОМОЩЬЮ ВЕРОЯТНОСТНОЙ НЕЙРОННОЙ СЕТИ АДАПТИВНОГО РЕЗОНАНСА

Д.В. Постарнак

аспирант, e-mail: dmi1852@yandex.ru

Тюменский государственный университет

Аннотация. В статье представлен новый метод, обеспечивающий ускорение классификации посредством класса нейронных сетей, базирующегося на теории адаптивного резонанса. Реализация метода возможна в рамках новой модели адаптивного резонанса, имеющей название «Вероятностная модель АРТ». Метод полезен в задачах, требующих максимального быстрого действия в поддержке принятия решений, в задачах распознавания химических соединений движущихся жидкостей, например.

Ключевые слова: нейрон, теория адаптивного резонанса, слой распознавания, синаптический вес.

1. Введение

Существуют задачи распознавания образов, которые требуют максимально-го быстрого действия в поддержке принятия решений, например, задачи распознавания химических соединений движущихся жидкостей. Статья содержит описание нового метода, обеспечивающего ускорение распознавания посредством класса нейронных сетей, базирующегося на теории адаптивного резонанса. Реализация метода возможна в рамках новой модели адаптивного резонанса, имеющей название «Вероятностная модель АРТ».

2. Описание вероятностной модели АРТ

В ходе исследования различных видов нейронных сетей адаптивного резонанса для реализации метода выбора направления разработана вероятностная модель АРТ [1-5] (АРТ-модель). Название класса нейронных сетей на основе теории адаптивного резонанса АРТ происходит от английской аббревиатуры ART — adaptive resonance theory. Вероятностная АРТ-модель разработана автором на основе модели АРТ-1, подробно описанной в работе [6]. АРТ-1 имеет два слоя: слой сравнения и слой распознавания (рис. 1).

Остальные блоки модели АРТ-1 отражают технические аспекты её реализации, поэтому в данной работе они не рассматриваются. Слой сравнения состоит

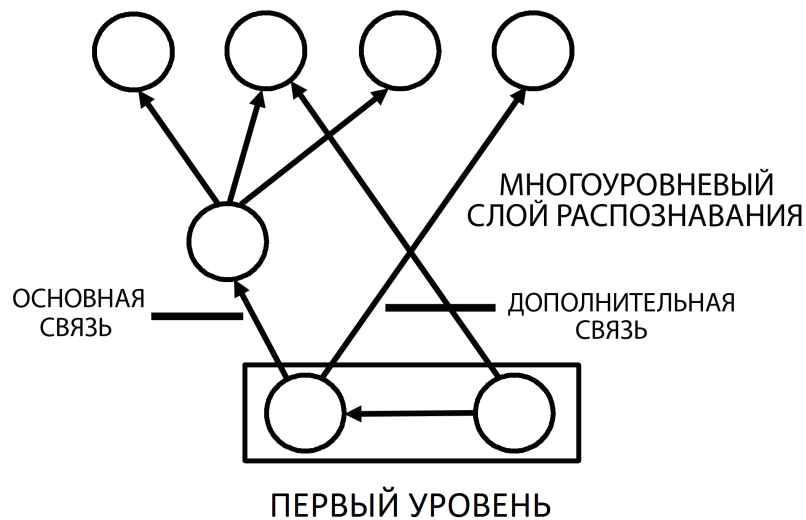


Рис. 2. Многоуровневый слой распознавания

За каждым нейроном слоя распознавания в этой модели закреплён не весь образ, а его конкретный участок (*локальное рецептивное поле*). Этот граф также содержит набор вершин, составляющих так называемый *первый уровень* (исток графа). С рассмотрения нейронов первого уровня всегда начинается классификация.

Преимущество вероятностной АРТ-модели перед АРТ-1 – в возможности использовать данные о вероятностях появления определённых фрагментов в соответствующих местах *глобального рецептивного поля* (области фигурирования образа).

3. Алгоритм метода выбора направления

Под направлением понимается порядок рассмотрения нейронов слоя распознавания. Алгоритм представлен следующим образом:

1. Вычисляются выходные сигналы нейронов первого уровня. Конкурировать по уровню сигнала могут только нейроны с общим *локальным рецептивным полем*:

$$F = \sum_l b_l s_l, s \in \{1, 0\}, \quad (1)$$

где F – выходной сигнал нейрона; l – порядковый номер входа, связанного с нейроном слоя распознавания; s – сигнал входа; b – вес связи между входом и нейроном слоя распознавания.

2. Для *нейрона-победителя* необходимо проверить неравенство, определяю-

шее степень соответствия образа, закреплённого за нейроном, входному образу:

$$\sum_l t_l / \sum_l s_l > \rho, \quad \rho < 1, \quad (2)$$

где l — номер входа, связанного с текущим нейроном слоя распознавания; s — сигнал входа ($s \in \{0, 1\}$); t — элемент образа, закреплённого за нейроном слоя распознавания (отражение того, что в идеальном случае должно быть на месте соответствующего входа s). Величина ρ является переменным параметром нейронной сети, $\rho \in (-\infty, 1)$. Высокий параметр ρ приведёт к повышению числа нейронов слоя распознавания, а низкий параметр ρ приведёт к высокой степени обобщения образов в нейронах.

3. Если степень соответствия образа нейрона входному образу удовлетворительна (неравенство (2) выполняется), это означает, что нейрон сработал, то есть стал частью ответного результата нейронной сети. При этом необходимо уточнить для *сработавшего нейрона* вектор весов связей по формуле, аналогичной той, которая применяется для модели АРТ-1 [6]:

$$b < \frac{W}{W - 1 + \sum_l s_l}, \quad W > 1, \quad (3)$$

где $W \in (1, \infty)$ — параметр, определяющий диапазон между максимальным и минимальным весом (имеет смысл при технической реализации); l — номер входа, связанного с нейроном слоя распознавания; s — сигнал входа. Следует отметить, что образ, закреплённый за *сработавшим нейроном* (элементы t на рис. 3), модифицируется путём присвоения элементам t значений соответствующих элементов s при условии $t \neq 0$ (принцип логического «и»).

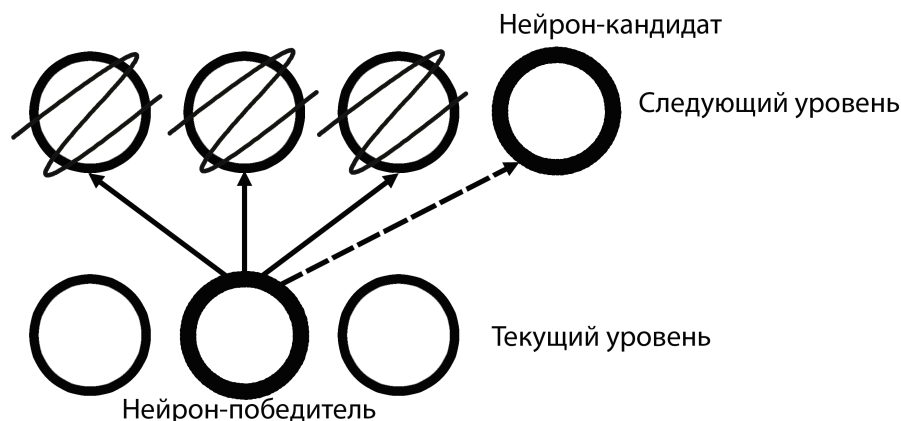


Рис. 3. Элементы t

4. Если степень соответствия образа нейрона входному образу неудовлетворительна (неравенство (2) не выполняется), тогда повторно вычисляются сигналы нейронов слоя распознавания, но без учёта предыдущего нейрона (возврат к шагу 1).

5. Если степень соответствия образа нейрона входному образу неудовлетворительна ни для одного нейрона, тогда в текущем уровне необходимо выделить *нейрон-кандидат* (нейрон, претендующий на то, чтобы быть добавленным в слой распознавания) (рис. 4).

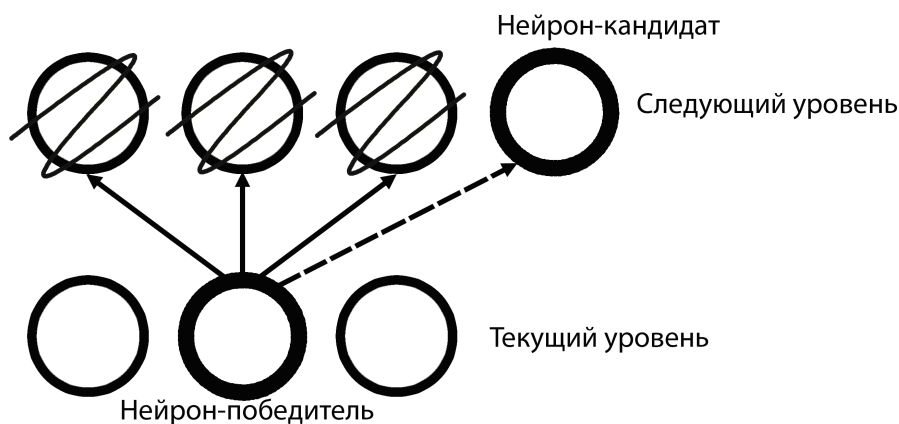


Рис. 4. Выделение нейрона-кандидата

Точный образ *нейрона-кандидата* (элементы t) равен фрагменту входного образа для текущего нейрона. Нейрон-кандидат должен быть связан с предыдущим сработавшим нейроном. Шаги 6 и 7 не являются обязательными.

6. После того как нейрон-кандидат сформирован, необходимо проверить его на возможность замены другим, уже имеющимся, нейроном. Для этого необходимо вычислить характеристики подобия образов K между имеющимся нейроном (любым) и нейроном-кандидатом:

$$K = \sum_l t_l \& s_l, \quad (4)$$

где t — элемент образа текущего сравниваемого нейрона; s — элемент образа нейрона-кандидата, $\&$ — логическое «и».

7. Если все характеристики $K < \varphi$, тогда *нейрон-кандидат* может быть добавлен в систему (рис. 5).

В противном случае вместо *нейрона-кандидата* будет использоваться текущий сравниваемый нейрон (рис. 6).

Величина $\varphi > 0$ является переменным параметром нейронной сети и зависит от желаемой степени сходства образов. При добавлении *нейрона-кандидата* в состав нейронов слоя распознавания классификация не останавливается, поскольку результат может быть завершён с помощью словарного уточнения образа без учёта нового нейрона (завершение классификации по неполному набору сработавших нейронов, с уточнением оставшихся нейронов по словарю).

8. Для *сработавшего нейрона* (или для нового нейрона) вычисляется величина E^* , которая является *скалярным видом образа*. *Скалярный вид образа*

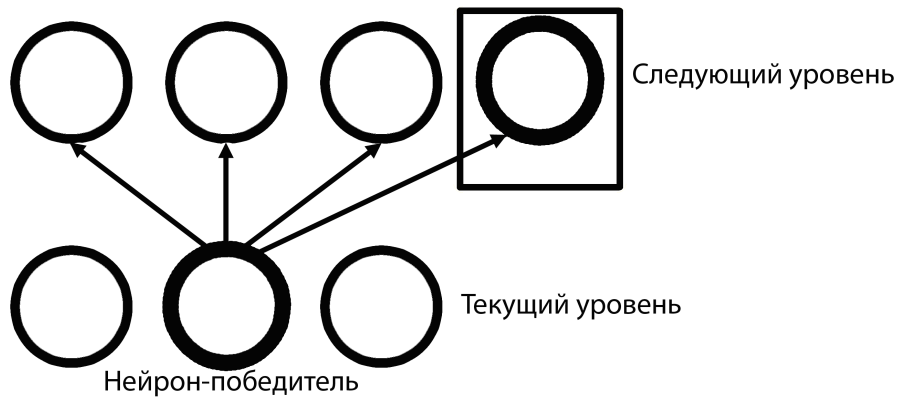


Рис. 5. Добавление нейрона-кандидата в систему

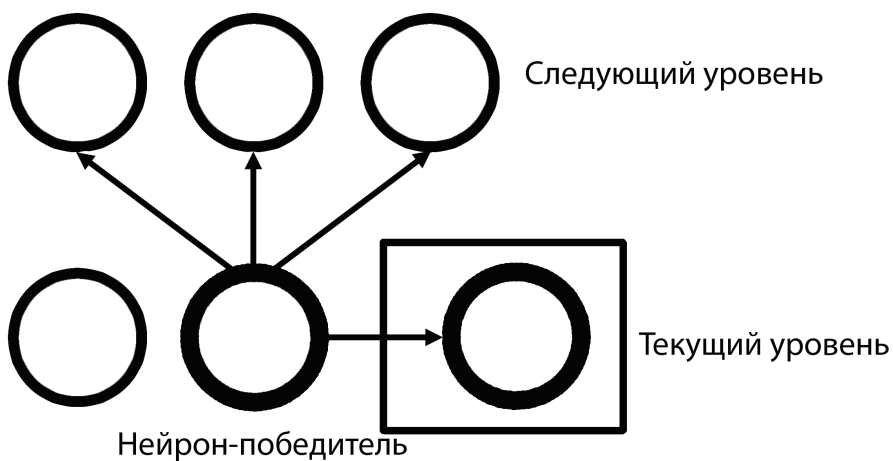


Рис. 6. Отмена нейрона-кандидата

— это целое число, однозначно определяющее весь образ:

$$E^* = \sum_l s_l \times 2^{l-1}, \quad (5)$$

где E^* — скалярный вид образа;

l — порядковый номер входа, связанного с нейроном слоя распознавания;

s — сигнал входа.

9. Осуществляется построение *определяющих векторов* H для каждого нейрона *следующего уровня* (т.е. для нейрона, связанного с последним *сработавшим нейроном*). Каждый определяющий вектор должен соответствовать равенству:

$$E^* = E + \sum_j h_j, h_j \in \{2^u; 0\}, j \in [1 : V], u \in [0 : \infty), \quad (6)$$

где E^* — скалярный вид образа, вычисленный на предыдущем шаге;
 E — скалярный вид образа, вычисленный для *сработавшего нейрона* на предыдущем шаге, но на более ранних этапах работы нейронной сети и закреплённый (сохранённый) за связанным нейроном, который на ранних этапах был следующим *сработавшим*. При первом использовании алгоритма данный шаг пропускается, поскольку все E равны нулю;

h — элемент *определяющего вектора*;

j — порядковый номер элемента *определяющего вектора*;

V — число входов *локального рецептивного поля сработавшего нейрона*.

Необходимо отметить, что при добавлении *нейрона-кандидата* в состав слоя распознавания, величина E этого нейрона должна быть равна *скалярному виду образа* предыдущего *сработавшего нейрона*.

10. Сигналы нейронов *следующего уровня* (нейроны, связанные с последним *сработавшим* связью от него) в случае, когда для этих нейронов формула (6) даёт **определяющий вектор** с минимумом элементов больше нуля вычисляются в первую очередь. Это осуществляется при возврате к шагу 1. Можно ориентироваться не на минимум, а на диапазон, охватывающий близкие к минимуму варианты значений. При срабатывании очередного нейрона его хранимая величина E получает значение *скалярного вида образа* предыдущего *сработавшего нейрона*. Необходимо отметить, что использование именно *скалярного вида образа* вместо вектора обусловлено более высокой потенциальной производительностью данного алгоритма, так как скалярные величины при компьютерной реализации обрабатываются быстрее, чем векторы.

Алгоритм завершается при получении набора *сработавших нейронов*, достаточного для классификации или для её завершения с помощью словарного уточнения.

4. Структура слоя распознавания

Каждый нейрон слоя распознавания вероятностной АРТ-модели представлен совокупностью компонентов Q :

$$Q = \langle G, I, J, Z \rangle. \quad (7)$$

Здесь G — вектор, содержащий набор порядковых номеров нейронов слоя распознавания, связанных с текущим нейроном (связь направлена от текущего нейрона):

$$G = (x_1, x_2, \dots, x_n), \quad E > 0, \quad (8)$$

где x_i ($i = 1, 2, \dots, n$) — порядковый номер нейрона слоя распознавания, связанного с текущим.

Вектор I хранит информацию о том, от каких нейронов существует связь к текущему (в отличие от компонентов вектора G здесь направление связей к текущему нейрону, а не от него), причём каждый порядковый номер связанного нейрона сопровождается скалярным видом образа E_i ($i = 1, 2, \dots, n$), сохранённым на предыдущих этапах работы нейронной сети. Скалярный вид

образа, как было обозначено выше, определяет скалярную величину, которая однозначно выражает образ, закреплённый за нейроном слоя распознавания.

$$I = (E_1, y_1, E_2, y_2, \dots, E_n, y_n), \quad E \geq 0, \quad (9)$$

где y_i ($i = 1, 2, \dots, n$) — порядковый номер нейрона слоя распознавания, связанного с текущим нейроном; E_i ($i = 1, 2, \dots, n$) — хранимый скалярный вид образа.

Вектор J в формуле (7) определяет набор входов, весовые коэффициенты связи нейрона слоя распознавания с входом и элементы образа, на который настроен нейрон (идеального образа) по отношению к соответствующим входам:

$$J = (l_1 \ b_1 \ t_1, \ l_2 \ b_2 \ t_2, \ \dots, \ l_n \ b_n \ t_n), \quad n = V, \quad (10)$$

где l_i ($i = 1, 2, \dots, n$) — порядковый номер входа; b_i ($i = 1, 2, \dots, n$) — весовой коэффициент связи нейрона слоя распознавания с входом; t_i ($i = 1, 2, \dots, n$) — элемент образа, на который настроен нейрон; V — число входов локального рецептивного поля (эти поля могут отличаться по размеру при условии, что они не перекрываются, и полностью охватываются соответствующими нейронами, конкурирующими между собой).

Компонент Z — это скаляр, который определяет принадлежность нейрона к первому уровню (если равен 1):

$$Z \in \{1, 0\}. \quad (11)$$

Если модель не содержит ни одного нейрона в слое распознавания, то необходимо для каждого целого образа определить набор нейронов слоя распознавания, *локальные рецептивные* поля которых не будут пересекаться. Связь между нейронами должна полностью охватывать распознаваемый образ. При этом всё пространство входных нейронов должно быть использовано. В качестве нейронов первого уровня необходимо использовать те, которые отражают отличия образов. Остальные характеристики должны быть следующие:

- вектор I нулевой;
- элементы b вектора J задаются по формуле, аналогичной той, которая применяется для присвоения начальных значений нейронам в модели АРТ-1 [6]:

$$b < \frac{W}{W - 1 + V}, \quad W > 1, \quad (12)$$

где $W \in (1, \infty)$ — параметр, определяющий диапазон между максимальным и минимальным весом (имеет смысл при технической реализации); V — число входов *локального рецептивного поля*, закреплённых за текущим нейроном слоя распознавания;

- все элементы t вектора J равны 1.

5. Обоснование метода выбора направления

Для лучшего восприятия метода рассмотрим задачу распознавания рукописных римских цифр «I», «V» и «X». Для каждого распознаваемого символа

сформирована совокупность из 81 входного нейрона (9×9). За каждой группой конкурирующих нейронов слоя распознавания (с общим *локальным рецептивным полем*) закреплено поле размером 3×3 . На рис. 7 эти поля разделены жирной линией.

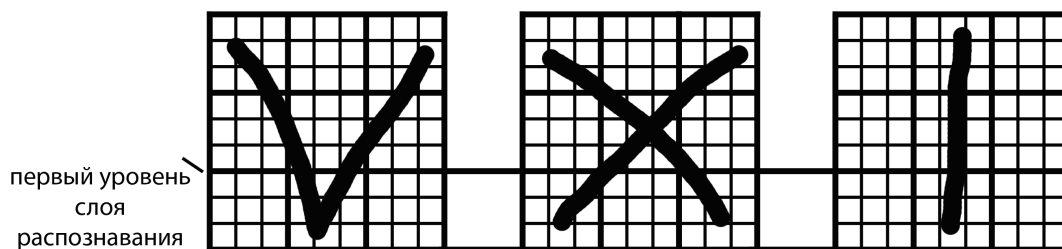


Рис. 7. Примеры локальных рецептивных полей

Процесс распознавания начинается с *первого уровня*. Очевидно, что, если активизируется только крайний правый нейрон среди нейронов *первого уровня*, как показано на рис. 8, то вероятнее всего, что распознаваемый символ является символом «X», поскольку символы «V» и «I» в среднестатистическом случае будут активировать средний нейрон *первого уровня*.

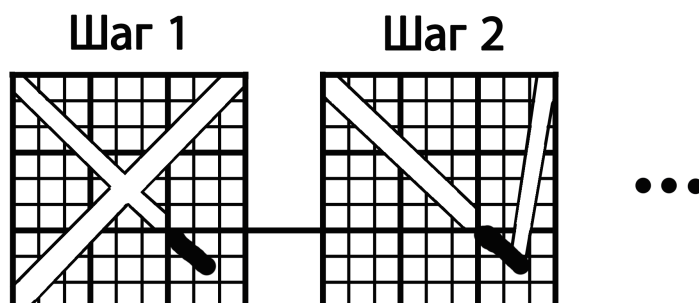


Рис. 8. Вероятность появления образа

Следовательно, дальнейшее рассмотрение нейронов для символа «X» следует производить в тех областях глобального рецептивного поля, в которых этот символ чаще всего представлен. Если вероятность не оправдывается, тогда переходим к следующему шагу и будем рассматривать области, которые соответствуют другому образу, имеющему менее распространённый вариант отображения, например символ «V», поскольку очевидно, что для текущего результата распознавания в правом нейроне первого уровня символ «I» является наименее вероятным.

Таким образом, применение метода выбора направления целесообразно, а конкретная степень увеличения скорости распознавания (по сравнению с моде-

лью ART-1) зависит от распознаваемых образов, то есть от предметной области, в которой данная нейронная сеть может функционировать. Метод, в целом, основан на анализе распределения вероятностей.

Заключение

В работе представлен метод выбора направления, обеспечивающий ускорение классификации. Разработан алгоритм, реализующий этот метод. В результате найден вариант ускорения классификации бинарных образов в нейронных сетях класса ART.

ЛИТЕРАТУРА

1. ART 2: Self-organization of stable category recognition codes for analog input patterns / Gail A. Carpenter and Stephen Grossberg // *Applied Optics*. 1987. Vol. 26, № 23.
2. ART 3: Hierarchical Search Using Chemical Transmitters in Self-Organizing Pattern Recognition Architectures / Gail A. Carpenter and Stephen Grossberg // *Neural Networks*. 1990. Vol 3. P. 129–152.
3. Fuzzy ARTMAP: A Neural Network Architecture for Incremental Supervised Learning of Analog Multidimensional Maps / Gail A. Carpenter, Stephen Grossberg, Natalya Markuzon, John H. Reynolds, David B. Rosen // *Transactions on neural networks*. 1992. Vol 3, № 5.
4. ARTMAP: Supervised Real-Time Learning and Classification of Nonstationary Data by a Self-Organizing Neural Network / Gail A. Carpenter, Stephen Grossberg, John H. Reynolds // *Neural Networks*. 1991. Vol 4. P. 565–588.
5. Fuzzy ART: Fast Stable Learning and Categorization of Analog Patterns by an Adaptive Resonance System / Gail A. Carpenter, Stephen Grossberg, David B. Rosen // *Neural Networks*. 1991. Vol. 4. P. 759–771.
6. Терехов С.А. Лекции по теории и приложениям искусственных нейронных сетей / Лаборатория Искусственных Нейронных Сетей НТО-2. Снежинск : ВНИИТФ, 1998. 66 с.
7. Постарнак Д.В. Критический анализ моделей нейронных сетей // *Вестник ТюмГУ*. 2012. № 4. С. 162–167.

DIRECTION CHOICE METHOD FOR INCREASING SPEED OF CLASSIFICATION USING PROBABILISTIC ART MODEL

D.V. Postarnak

Postgraduate Student, e-mail: dmi1852@yandex.ru

Tumen State University

Abstract. This article presents a new method that increases recognition speed using a class of adaptive resonance theory neural networks. The method implementation is possible within a new adaptive resonance model named as "Probabilistic ART model". The method is useful in the tasks demanding the fastest decision-making, e.g., in recognition of chemical compounds in the moving liquids.

Keywords: neuron, recognition layer, synaptic weight, adaptive resonance theory.

К РЕАЛИЗАЦИИ МОДИФИЦИРОВАННОЙ ВЕРОЯТНОСТНОЙ НЕЙРОННОЙ СЕТИ АДАПТИВНОГО РЕЗОНАНСА

Д.В. Постарнак

аспирант, e-mail: dmi1852@yandex.ru

В.А. Шапцев

д.т.н., профессор кафедры информационных систем, e-mail: VAShaptsev@yandex.ru

Тюменский государственный университет

Аннотация. Предлагается вариант эффективного распределения производительности между вычислительными компонентами при выполнении распознавания образов методом выбора направления в модифицированной вероятностной нейронной сети адаптивного резонанса.

Ключевые слова: нейрон, теория адаптивного резонанса, слой распознавания, синаптический вес, процессор, параллельные вычисления.

Введение

В реализации метода распознавания образов, описанного в [1], производительность и надёжность системы процесса можно повысить применением схемы параллельных вычислений. Ниже представлены модель искусственной нейронной сети, отличающаяся включением выбора направления поиска, и вариант распределения производительности при параллельных вычислениях в процессе распознавания.

1. О модели нейронной сети

Рассматривается модель АРТ (от англ. Adaptive Resonance Theory) нейронной сети, разработанная Стивеном Гроссбергом [2]. Основные черты, отличающие её от остальных: самостоятельное обучение и повышение качества распознавания за счёт наполнения слоя распознающих нейронов [3]. Важны следующие особенности этой модели:

- конкурентный механизм выбора сработавшего нейрона;
- определение степени различия между идеальным образом, закреплённым за текущим распознающим нейроном, и текущим входным образом. На основании этого выносится решение о том, считать состояние текущего нейрона, участвующего в распознавании, результатом классификации или нет;

- если ни один из распознающих нейронов не обеспечивает необходимого уровня соответствия входному образу, то происходит «адаптивный резонанс» - включение в слой распознавания нового нейрона, несущего идеальный образ, полностью соответствующий текущему входному.

В [1] предложена вероятностная модель АРТ, названная там PART (Probabilistic ART). Одной из особенностей этой модификации является использование нескольких связанных структур в качестве слоя распознавания. Это позволяет учесть нюансы входных образов, упускаемые другими типами моделей, и использовать эти нюансы для определения вероятностного показателя успешности пути обхода дерева [1]. Теперь классификация ведётся не по одному нейрону, а по набору сработавших нейронов (порядок расположения которых неважен). Процесс сопровождается набором результатов классификации, каждый (набор) из которых представлен порядковыми номерами нейронов распознавания:

$$B = \langle C, D \rangle, \quad (1)$$

где C — набор порядковых номеров распознавших нейронов, которые формируют успешную текущую классификацию B ; D — число успешных классификаций с данным результатом. Вводим множество

$$A = \{B_1, B_2, \dots, B_n\}, \quad n \geq 0, \quad (2)$$

где B_i — i -й результат классификации; n — число результатов классификации в системе распознавания.

Безуспешная классификация определяется как пустое множество *сработавших* нейронов: $C = \emptyset$. В этом случае, при использовании *скалярного вида образа* (СВО) [1], нецелесообразно переходить к следующему узлу ветвления. Можно вернуться к предыдущему нейрону и по другим СВО попробовать построить *полосу классификации* с другим окончанием. Если от предыдущего нейрона ни один из ближайших (допустимых) СВО не даёт результата, то спускаемся к ещё более раннему распознающему нейрону и проделываем то же самое. Если вплоть до самого корня классификация не удалась, то возвращаемся к первоначальному виду невыполненной классификации и повторяем процесс постепенного перехода к предыдущим нейронам без учёта СВО. Если и это не помогло, то делается вывод о том, что классификация от текущего корня не состоялась.

2. Адаптация к параллельным вычислениям

Алгоритм параллельных вычислений в рамках PART по своей сути не меняется. Разница состоит в том, что теперь процесс ведётся не только от корня к листьям, но и в обратном направлении. Совокупность компонентов, закреплённая за распознающим нейроном, меняет вид:

$$Q = \langle G, I, E_1, E_2, J, M \rangle, \quad (3)$$

где G — набор порядковых номеров распознающих нейронов, связанных с текущим. Он используется при движении от корня к листьям; I — вектор, содержащий набор порядковых номеров распознающих нейронов, связанных с текущим. Используется при движении от листа к корню; E_1 — СВО при движении от корня к листьям; E_2 — СВО при движении от листа к корню; M — маркер участия нейрона в текущей классификации.

Вектор J определяет набор входов, весовые коэффициенты связи нейрона слоя распознавания со входом и элементы образа, на которые настроен нейрон (идеального образа) по отношению к соответствующим входам:

$$J = (l_1, b_1, t_1; l_2, b_2, t_2; \dots; l_n, b_n, t_n), n \geq 0, \quad (4)$$

где l_i ($i = 1, 2, \dots, n$) — порядковый номер входа; b_i ($i = 1, 2, \dots, n$) — весовой коэффициент связи нейрона слоя распознавания со входом; t_i ($i = 1, 2, \dots, n$) — элемент образа, на который настроен нейрон; n — число входов. Наборы могут отличаться по размеру при условии, что они не перекрываются и полностью охватываются конкурирующими нейронами.

Теперь кроме набора порядковых номеров нейронов со связью вверх, то есть от корня к листу, появляется такой же набор со связью вниз. Скалярный вид образа (СВО) также представлен как для верхнего хода, так и для нижнего. Вычисляется СВО аналогично как вверх, так и вниз. То есть суть учёта нюансов не меняется. Есть ещё «маркер», указывающий на то, был ли использован конкретный нейрон в текущем процессе распознавания, т.е. маркер может быть равен единице в одном случае, а в следующий раз — нулю.

Система распознавания PART с применением параллельных вычислений предполагает использование дополнительных структур данных. Листья хранятся отдельно

$$O = (P_1, P_2, \dots, P_n), n \geq 0, \quad (5)$$

где P_i — набор порядковых номеров нейронов распознавания, используемых в качестве листьев; i — номер корня, которому соответствует набор.

Корни (возможные начала классификаций) должны быть также отмечены во внешнем векторе:

$$N = (m_1, m_2, \dots, m_n), n \geq 0, \quad (6)$$

где m_i — порядковый номер нейрона распознавания, используемого в качестве корня; i — номер корня в общем наборе.

Хранение данных о корнях и листьях необходимо для релевантного распределения производительности во избежание ситуации, когда выбирается несколько нейронов, идущих из одного корня и, как следствие, идёт проверка одного и того же образа несколько раз подряд, сводя на нет смысл параллельной работы.

Всякий раз, когда процесс классификации доходит до листа, у которого имеется только обратный набор ссылок, а прямой набор ссылок пуст, этот последний элемент (лист) заносится в вектор листьев (получается, что при условии его отсутствия в нём).

3. Работа модели при параллельных вычислениях

Параллелизм работы модели заключается в следующем. Выбирается любой корень из (6). По виду вектора (6) можно определить, какие листья соответствуют выбранному корню (порядок следования наборов листьев P и корней m совпадает). Запускается классификация от выбранного корня к листьям на первом процессоре CPU1 (рис. 1). По крайним листьям с двух сторон запускается классификация на двух процессорах (два процессора идут по листьям друг к другу) (верхняя часть рис. 1). Если один процессор CPU3 «входит» в дерево, то второй CPU2 не останавливается. Его точка входа может быть ближе к краю дерева. Нормально, что оба процессора CPU2 и CPU3 включились в работу. Главное, чтобы они начали вычисления с разных листьев.

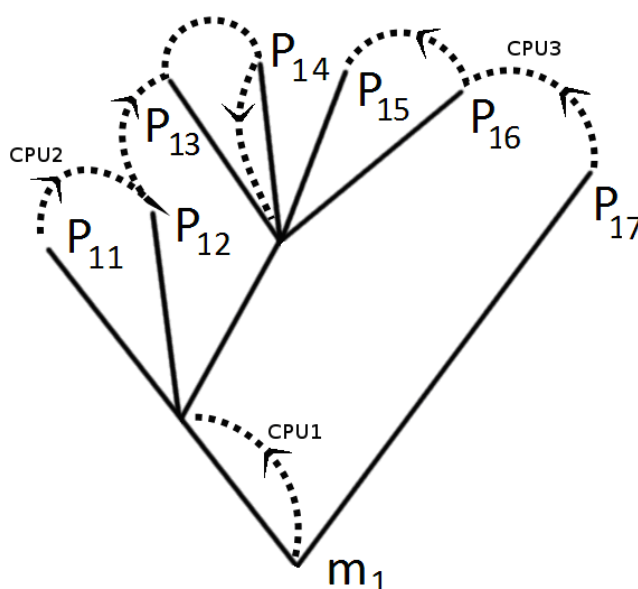


Рис. 1. Вход в дерево

Назовём *ведомыми процессорами* те, которые обходят листья, как описано выше, то есть — CPU2 ($P_{11}, P_{12}, P_{13}, P_{14}$) и CPU3 (P_{17}, P_{16}, P_{15}). Они при погружении должны использовать информацию о СВО.

Если *решающая полоса* (последовательность переходов от нейрона к нейрону, формирующая результат классификации) *ведущего процессора* (того, который ведёт обработку от корня к листьям, в нашем случае CPU1) не соединилась ни с одной из *решающих полос ведомых процессоров* (смотрим по маркеру), то в самом худшем варианте у нас есть три (по общему числу процессоров) завершённых классификации (которые были пройдены по правилам АРТ-концепции) (рис. 2).

В этом случае берём ту полосу ($m_1 - n_1 - P_{11}$ или $m_1 - n_1 - n_2 - P_{14}$ или $P_{17} - P_{16} - P_{15}$), у которой самое большое число срабатываний. Число

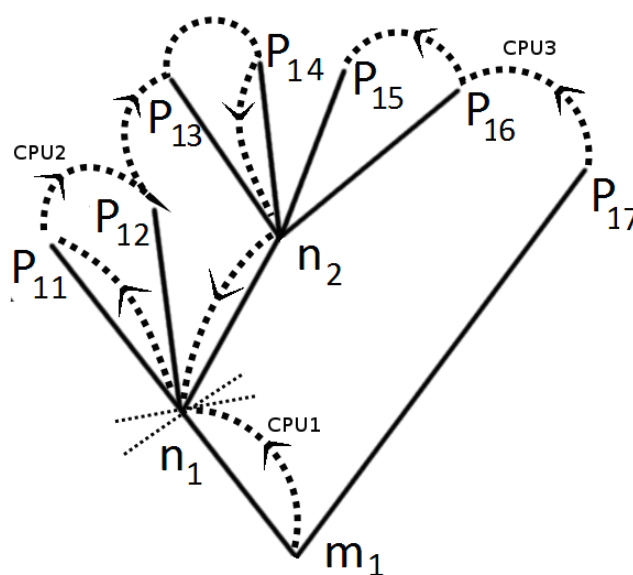


Рис. 2. Расхождение параллельных процессов

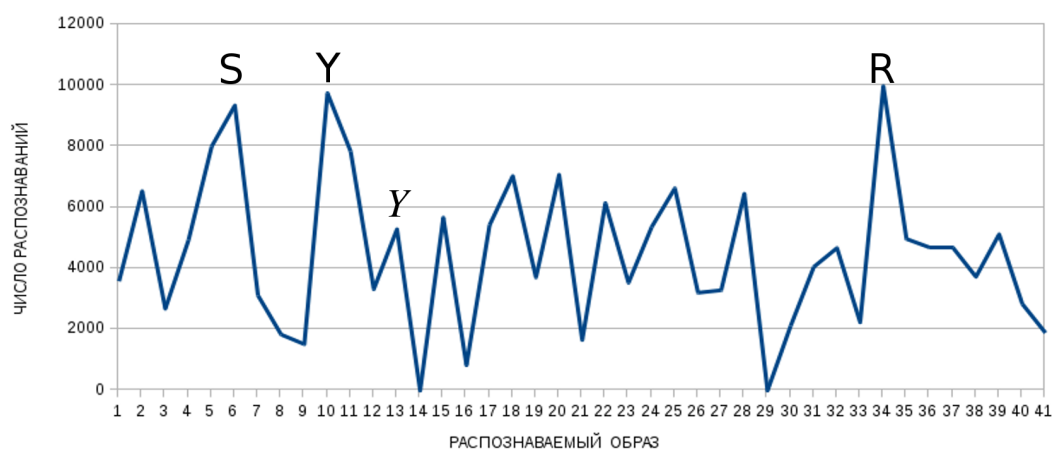


Рис. 3. Число распознаваний образов

срабатываний (в рамках модели A) (2) увеличивается только тогда, когда есть все сработавшие полосы с одинаковыми показателями по срабатыванию (рост только у одной из полос B) (1). Следовательно, на пик по числу срабатываний выходят те образы, которые между собой не соперничали (или соперничали менее интенсивно, чем остальные). Это показано на рис. 3: два варианта написания символа «Y» могли конкурировать до того момента, когда количество символов «Y», написанных курсивом, заметно убавилось.

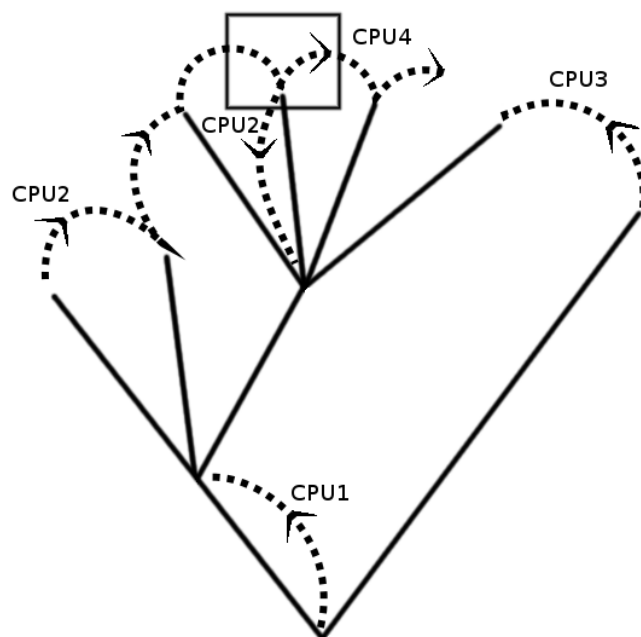


Рис. 4. Продолжение прохода по листьям

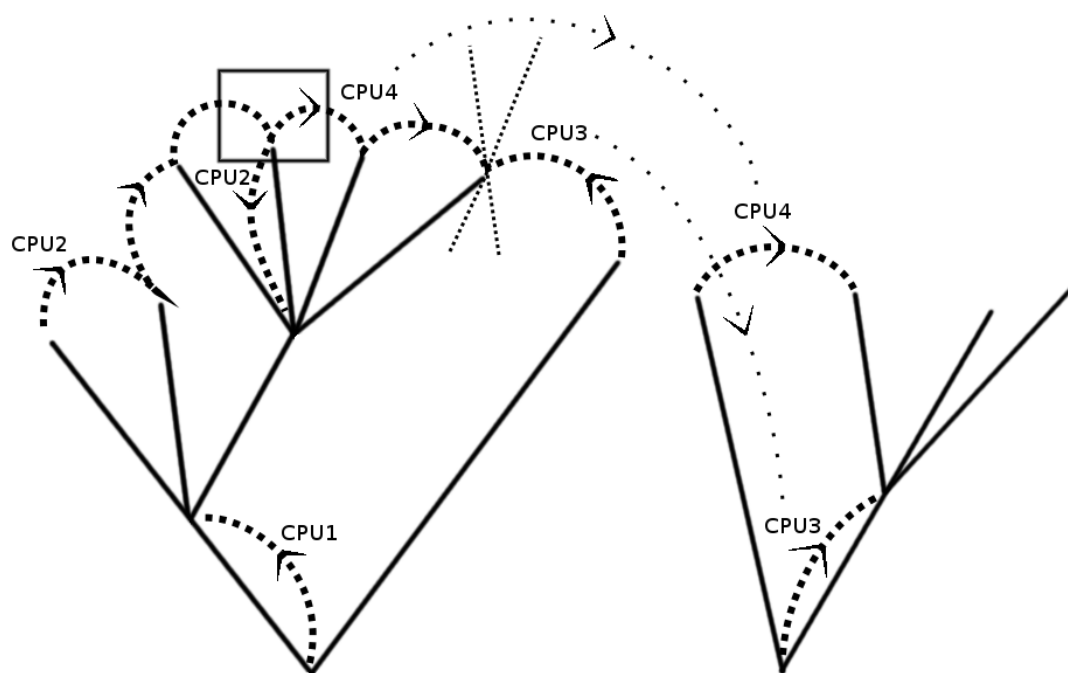


Рис. 5. Параллельное следование от корня

Если один из ведомых процессоров «входит» в дерево, то можно продолжить проход по листьям в том же направлении другими процессорами (рис. 4).

При встрече идущих по листьям процессоров некоторые из них могут освободиться. Из освободившихся процессоров можно сформировать новую группу «ведущий — два ведомых» или «ведущий — ведомый», или вообще один ведущий, лишь бы процессоры не простаивали (рис. 5).

Заключение

Выше представлен вариант эффективного (в смысле ускорения вычислений) распределения производительности между процессорами при выполнении распознавания образов методом выбора направления. Приведённая методика сохраняет эффективность системы при частичном выходе из строя вычислительных компонентов, то есть система может менять число используемых процессоров (по мере их доступности) во время непрерывной работы. Производительность системы должна возрасти в число процессоров за минусом времени, затраченного на управление распараллеливанием.

ЛИТЕРАТУРА

1. Постарнак Д.В. Метод выбора направления для ускорения классификации с помощью вероятностной нейронной сети адаптивного резонанса // Математическое и информационное моделирование: сб. науч. тр. 2013. № 13. С. 218–229.
2. Grossberg S. Competitive Learning: From Interactive Activation to Adaptive Resonance // Cognitive science. 1987. № 11. Р. 23–63.
3. Постарнак Д.В. Критический анализ моделей нейронных сетей // Вестник ТюмГУ. 2012. № 4. С. 162–167.

TOWARDS IMPLEMENTING THE MODIFIED PROBABILISTIC ADAPTIVE RESONANCE NEURAL NETWORK

D.V. Postarnak

Postgraduate Student, e-mail: dmi1852@yandex.ru

V.A. Shaptsev

Professor, Doctor of Engineering, e-mail: VAShaptsev@yandex.ru

Tumen State University

Abstract. We propose a method for an efficient allocation of performance between computational components when carrying out the pattern recognition by choosing direction in the modified probabilistic adaptive resonance neural network.

Keywords: neuron, adaptive resonance theory, recognition layer, synaptic weight, processor, parallel computing.

Научный журнал

Математические структуры и моделирование

№2(34)

Главный редактор

А.К. Гуц

Выпускающий редактор

Д.Н. Лавров

Технический редактор

Н.Ф. Богаченко

Корректор:

И.Н. Баловнева

Корректоры переводов:

Е.А. Илюшечкин

А.Н. Кабанов

Адрес научной редакции

Россия, 644077, Омск, пр. Мира, 55А

Омский государственный университет

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

Электронная версия журнала:

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

<http://msm.omsu.ru>



Подписано в печать 29.06.2015. Формат 60 × 84 1/8.

Усл. печ. л. 13,02. Тираж 100 экз. Заказ № 133.

Отпечатано на полиграфической базе издательства ОмГУ им. Ф.М. Достоевского
644077, г. Омск, пр. Мира, 55А

ISSN 2222-8772



9 772222 877005



15034 >